

# FONDAMENTI DI STATISTICA APPLICATA ALL'ANALISI E ALLA GESTIONE DELL'AMBIENTE

(edizione settembre 2001)

*Lamberto Soliani*

Email: [soliani@dsa.unipr.it](mailto:soliani@dsa.unipr.it) Tel.0521/905662 Fax 0521/905402

## Presentazione

Con intensità crescente negli ultimi anni, collegata alla diffusione dell'informatica e all'emergenza dei problemi ambientali, i tecnici, i ricercatori e i responsabili della gestione ambientale chiedono gli strumenti per una conoscenza operativa delle metodologie statistiche. I corsi di laurea in Biologia, Scienze Naturali e Geologia, quelli di Chimica e di Ingegneria del territorio, dai quali provengono molti degli attuali operatori dell'ambiente, sovente forniscono una preparazione teorica che si dimostra insufficiente, per affrontare i problemi della professione. Anche agli studenti dei corsi di diploma, di laurea, dei master e dei dottorati finalizzati alla ricerca ecologica ed ambientale, appare utile una spiegazione semplice dei concetti fondamentali e dei metodi più importanti. Per essi hanno un'importanza relativa le dimostrazioni matematiche sofisticate, mentre è necessaria un'illustrazione chiara e semplice dei test più diffusi, sovente spiegati in modo troppo sommario e poco comprensibile nei manuali relativi.

L'autore ha cercato di rispondere a questa domanda così diversificata, con una presentazione elementare ma completa dei concetti e dei metodi dell'inferenza statistica, con numerose applicazioni. Nei casi più semplici, sono stati esposti i calcoli manuali; in quelli più complessi, sono stati forniti gli elementi essenziali, per comprendere i risultati dei programmi informatici più utilizzati nella ricerca internazionale.

La riforma degli ordinamenti didattici, in atto nelle Università italiane, esige programmi strutturati in moduli, per una preparazione culturale più completa da fornire nei corsi di laurea, per un approccio rapidamente operativo nei diplomi. Il manuale, che fornisce argomenti ed esempi per oltre 100 ore di lezione, può essere scisso con facilità in parti indipendenti, autonome ed ugualmente di semplice comprensione: i concetti più importanti sono ripetuti in molti esempi, in modo volutamente ridondante, con estrema attenzione alle applicazioni.

Ideato e scritto per la formazione e l'aggiornamento degli operatori dell'ambiente, questo manuale vuole essere utile sia per la preparazione di base sia per l'approfondimento su molti temi di statistica univariata e bivariata, parametrica e non parametrica di uso corrente.

Lamberto Soliani

Professore ordinario di "Fondamenti di analisi dei sistemi ecologici"  
e docente di "Statistica applicata"  
(Dipartimento di Scienze Ambientali, Università di Parma)  
Viale delle Scienze 43100 Parma

## INDICE

### 1. ELEMENTI DI STATISTICA DESCRITTIVA PER DISTRIBUZIONI UNIVARIATE

#### 1.1. La statistica nella ricerca ambientale

- 1.2. Il disegno sperimentale ed il campionamento
- 1.3. Tipi di dati e scale di misurazione
  - 1.3.1 La scala nominale o classificatoria
  - 1.3.2 La scala ordinale o per ranghi
  - 1.3.3 La scala ad intervalli
  - 1.3.4 La scala di rapporti
- 1.4. Classificazione in tabelle
- 1.5. Rappresentazioni grafiche di distribuzioni univariate
- 1.6. Le misure di tendenza centrale
  - 1.6.1 Le misure di tendenza centrale o posizione
  - 1.6.2 La mediana
  - 1.6.3 La moda
- 1.7. Misure di dispersione o variabilità
  - 1.7.1 Intervallo di variazione
  - 1.7.2 La differenza interquartile
  - 1.7.3 Lo scarto medio assoluto dalla media
  - 1.7.4 Lo scarto medio assoluto dalla mediana
  - 1.7.5 La devianza
  - 1.7.6 La varianza
  - 1.7.7 La deviazione standard
  - 1.7.8 L'errore standard
  - 1.7.9 Il coefficiente di variazione
  - 1.7.10 La varianza in dati raggruppati: correzione di Sheppard
- 1.8. Indici di forma: simmetria e curtosi
- 1.9. Metodi per calcolare un generico quantile da una serie di dati
- 1.10. Rappresentazione semi-grafica delle distribuzioni: Box-and-Wisker, diagrammi Stem-and-Leaf
- 1.11. Esercizi sulle misure di tendenza centrale, dispersione, simmetria e curtosi

## 2. DISTRIBUZIONI E LEGGI DI PROBABILITA'

- 2.1. Elementi di calcolo combinatorio semplice
  - 2.1.1 Permutazioni semplici
  - 2.1.2 Disposizioni semplici
  - 2.1.3 Combinazioni semplici
  - 2.1.4 Risposte alle domande del paragrafo 2.1
- 2.2. Definizioni di probabilità: matematica, frequentista, bayesiana
- 2.3. Alcune distribuzioni discrete

2.3.1 Distribuzione binomiale

2.3.2 Distribuzione multinomiale

2.3.3 Distribuzione poissoniana

2.3.4 Distribuzione geometrica e distribuzione di Pascal

2.3.5 Distribuzione ipergeometrica

2.3.6 Distribuzione binomiale negativa

2.3.7 Distribuzione uniforme o rettangolare

2.4. Alcune distribuzioni continue

2.4.1 Distribuzione normale o di Gauss

2.4.2 Distribuzioni asintoticamente normali, con approssimazioni e trasformazioni

2.4.3 Dalla disuguaglianza di Tchebycheff all'uso della distribuzione normale

2.4.4 Approssimazioni e correzioni per la continuità

2.4.5 Distribuzione rettangolare

2.4.6 Distribuzione esponenziale negativa

2.4.7 Le curve di Pearson

2.4.8 La distribuzione gamma

2.5. Distribuzioni campionarie derivate dalla normale ed utili per l'inferenza:  di Pearson, t di Student e F di Fisher

2.5.1 La distribuzione 

2.5.2 La distribuzione t di Student

2.5.3 La distribuzione F di Fisher

### 3. CONFRONTI TRA TASSI E PROPORZIONI

3.1. Confronti tra distribuzioni osservate e distribuzioni te attese

3.2. Condizioni di validità del  e correzione di Yates

3.3. Il metodo di Kolmogorov-Smirnov per un campione

3.4. Il confronto tra due distribuzioni osservate, per test di indipendenza: le tabelle di contingenza 2 x 2 (fourfold tables)

3.5. Confronti tra frequenze relative con la distribuzione normale e sua correzione per la continuità

3.6. Confronto tra test  per tabelle 2 x 2 e test z, senza e con le varie correzioni per la continuità

3.7. Confronto di una proporzione osservata con una attesa:

il test z per grandi campioni; la distribuzione binomiale per piccoli campioni

- 3.8. Tabelle di contingenza 2 x 2 in piccoli campioni: il metodo esatto di Fisher
- 3.9. Le tabelle 2 x N con la formula generale e quella di Brandt-Snedecor. Le tabelle M x N
- 3.10. Il log-likelihood ratio o metodo G
- 3.10.1 Confronto tra una distribuzione osservata ed una attesa con la correzione di Williams
- 3.10.2 Tabelle 2 x 2, con la correzione di Williams e quella di Mantel-Haenszel
- 3.10.3 Tabelle M x N con la correzione di Williams
- 3.11. Il confronto tra due distribuzioni osservate: il metodo di Kolmogorov-Smirnov per 2 campioni indipendenti
- 3.12. Il chi quadro con il metodo di Cochran e di Mantel-Haenszel
- 3.13. Esercizi svolti per dati in tabelle di contingenza

#### 4. VERIFICA DELLE IPOTESI

- 4.1. Risultati significativi e non-significativi
- 4.2. Procedura di verifica delle ipotesi: vero o falso? utile o dannoso?
- 4.3. Potenza di un test
- 4.4. Numero di dati necessari in rapporto alla potenza, alla significatività del test e alla direzionalità dell'ipotesi. Il criterio di Cohen per la scelta di a e b
- 4.5. Dimensioni (n) del campione, nel caso di proporzioni
- 4.6. Le quattro proprietà che deve avere uno stimatore: correttezza, efficienza, consistenza, sufficienza
- 4.7. Intervallo fiduciale di una media
- 4.8. Valutazione del rischio aggiuntivo (f) e intervallo fiduciale di una proporzione
- 4.9. Intervallo di confidenza di una varianza

#### 5. INFERENZA SU UNA O DUE MEDIE CON IL TEST t DI STUDENT

- 5.1. La distribuzione t di Student
- 5.2. Confronto tra una media osservata e una media attesa; calcolo dei limiti di confidenza di una media, con  $\sigma$  ignota
- 5.3. Confronto tra una osservazione e la media di un campione
- 5.4. Il confronto tra le medie di due campioni
- 5.5. Il test t per 2 campioni dipendenti o per dati appaiati ed intervallo fiduciale della media delle differenze
- 5.6. Il test t per 2 campioni indipendenti o per dati non appaiati
- 5.7. Test F, test di Levene e test di Bartlett per ipotesi bilaterali e unilaterali sull'uguaglianza di due varianze
- 5.8. Significatività e intervallo fiduciale di una differenza
- 5.9. Potenza a priori e a posteriori del test t, con un campione e con due campioni dipendenti o indipendenti

- 5.10. Dimensione del campione e precisione nella stima sia di una media, sia della differenza tra due medie
- 5.11. Il bilanciamento di due campioni indipendenti: vantaggi e costi
- 5.12. Potenza a priori e a posteriori del test F per l'omoschedasticità
- 5.13. Correzione per campionamento in una popolazione finita e il concetto di superpopolazione
- 5.14. Test per la differenza tra due coefficienti di variazione con distribuzione z oppure distribuzione t di Student

## 6. METODI NON PARAMETRICI PER UN CAMPIONE

- 6.1. Caratteristiche dei test non parametrici
- 6.2. Il test delle successioni per un campione
- 6.3. Il test dei segni per un campione
- 6.4. Intervallo di confidenza di una probabilità o frequenza relativa, secondo il metodo di Clopper e Pearson
- 6.5. Intervalli di confidenza non parametrici e intervalli di tolleranza
- 6.6. Il test dei segni per ranghi di Wilcoxon
- 6.7. Caratteristiche distintive dei test sulla bontà di adattamento rispetto a quelli su un parametro
- 6.8. Il test di Lilliefors e il test di Cramér e Von Mises

## 7. METODI NON PARAMETRICI PER DUE CAMPIONI

- 7.1. Test per 2 campioni dipendenti o per dati appaiati
- 7.2. Il test di McNemar
- 7.3. Il test dei segni
- 7.4. Il test dei segni per ranghi: test T di Wilcoxon
- 7.5. Test di casualizzazione per 2 campioni dipendenti
- 7.6. Test per 2 campioni indipendenti
- 7.7. Il test della mediana o test di Mood
- 7.8. Il test di Wilcoxon-Mann-Whitney della somma dei ranghi
- 7.9. Il test u di Mann-Whitney o dell'ordine robusto dei ranghi
- 7.10. Cenni del test S di Kendall e rapporti con T ed U; potenza-efficienza dei tre test
- 7.11. Test di casualizzazione per 2 campioni indipendenti
- 7.12. Il test delle successioni per due campioni o test di Wald-Wolfowitz
- 7.13. Test di Siegel-Tukey per l'uguaglianza della varianza: cenni del test di Freund-Ansari-Bradley e del test di Conover
- 7.14. Il test dei ranghi equivalenti di Moses per le differenze nella dispersione o variabilità

## 8. Analisi della varianza (ANOVA I) a UN criterio di classificazione E CONFRONTI TRA PIU' MEDIE

- 8.1. Analisi della varianza ad un criterio di classificazione o a campionamento completamente randomizzato
- 8.2. Confronto tra analisi della varianza con due trattamenti e test t di Student per 2 campioni indipendenti
- 8.3. Test per l'omogeneità della varianza tra più campioni: test di Hartley, Cochran, Bartlett, Levene
- 8.4. I confronti tra più medie
  - 8.4.1 Confronti a priori o pianificati od ortogonali
  - 8.4.2 Test per confronti multipli o a posteriori: Bonferroni, snk di Neuman-Keuls, lsd di Fisher, step-up di Welsch, hsd di Tukey, Scheffé, Dunnett, Duncan
- 8.5. Stima della dimensione N di k gruppi campionari per l'ANOVA

## 9. ANALISI DELLA VARIANZA A PIU' CRITERI DI CLASSIFICAZIONE

- 9.1. Analisi della varianza a due criteri di classificazione o a blocchi randomizzati, con una sola osservazione per casella
- 9.2. Confronto tra analisi della varianza a due criteri e test t di Student per 2 campioni dipendenti
- 9.3. Analisi della varianza a tre o più criteri
- 9.4. Quadrati latini e greco-latini
- 9.5. Dati mancanti o anomali in disegni a più fattori
- 9.6. Efficienza relativa di due disegni sperimentali
- 9.7. Potenza a priori e a posteriori nell'ANOVA, con grafici di Pearson e Hartley
- 9.8. Appendice al capitolo: lettura di tabulati dell'analisi della varianza effettuata con un pacchetto statistico

## 10. ANALISI FATTORIALE, DISEGNI COMPLESSI: TRASFORMAZIONI DEI DATI

- 10.1 Analisi fattoriale ed interazione
- 10.2. Interazione tra due fattori a più livelli
- 10.3. Rappresentazione grafica per l'interpretazione dell'interazione a due fattori
- 10.4. Analisi della varianza a due fattori con repliche ineguali
- 10.5. Il test T di Tukey per il confronto tra le medie in disegni a due fattori con repliche
- 10.6. Esperimenti fattoriali  $2 \times 2$  e  $2^3$  con i confronti ortogonali
- 10.7. Esperimenti fattoriali con P fattori a k livelli
- 10.8. Quadrati latini con repliche
- 10.9. Analisi gerarchica (Nested ANOVA)
- 10.10. Assunzioni di validità e trasformazioni dei dati
- 10.11. La scelta della trasformazione idonea: il metodo di Box-Cox

10.12. Lettura di due tabulati di programmi informatici

## 11. TEST NON PARAMETRICI PER PIU' CAMPIONI

11.1. Test non parametrici analoghi all'analisi della varianza

11.2. Estensione del test della mediana

11.3. Cenni sul test di Nemenyi

11.4. Analisi della varianza per ranghi ad un criterio di classificazione: il test di Kruskal-Wallis

11.5. Confronti multipli nell'analisi della varianza per ranghi, con k campioni indipendenti

11.6. Test per l'eterogeneità della varianza con k campioni

11.7. Il test Q di Cochran

11.8. Estensione del test di McNemar a una tabella quadrata N x N: il test di Bowker.

11.9. Analisi della varianza per ranghi, a 2 criteri di classificazione: test di Friedman con una e con K repliche

11.10. I confronti multipli tra medie di ranghi nell'analisi della varianza non parametrica, a due criteri di classificazione

## 12. REGRESSIONE LINEARE SEMPLICE

12.1. Regressione o correlazione?

12.2. Descrizione di distribuzioni bivariate

12.3. Modelli di regressione

12.4. La regressione lineare semplice

12.5. Valore predittivo della regressione

12.6. Significatività dei parametri b e a della retta di regressione

12.7. Potenza a priori e a posteriori nella regressione lineare

12.8. Intervallo di confidenza dei parametri b e a .

12.9. Limiti di confidenza per i valori medi di  $\bar{x}_i$  stimati

12.10. Limiti di confidenza per singoli valori di  $x_i$  stimati

12.11. Errori delle variabili e limiti di tolleranza

12.12. Il coefficiente di determinazione:  $R^2$  e  $R^2$  adj (aggiustato)

12.13. La predizione inversa

12.14. Confronto tra due o più rette di regressione

12.15. Confronti multipli tra più coefficienti angolari

12.16. Analisi della relazione dose-effetto nel caso di Y ripetute: test per la linearità e calcolo della retta di regressione

12.17. Condizioni di validità della regressione, analisi dei residui e trasformazioni

### 13. CORRELAZIONE E COVARIANZA

13.1. La correlazione

13.2. Condizioni di validità e significatività di  $r$  con  $r = 0$  e con  $r$  diverso da 0

13.3. Significatività della retta con  $R^2$  ?

13.4. Intervallo di confidenza di  $r$

13.5. Potenza a priori e a posteriori per la significatività di  $r$

13.6. Differenza tra due coefficienti di correlazione in campioni indipendenti e calcolo del coefficiente comune

13.7. Potenza a priori e a posteriori del test per la significatività della differenza tra due coefficienti di correlazione

13.8. Test per la differenza tra più coefficienti di correlazione; coefficiente di correlazione comune  $r_w$  e sua significatività

13.9. Cenni sui confronti multipli tra più  $r$

13.10. La correlazione parziale o netta di primo ordine e di ordine superiore; la correlazione semiparziale

13.11. Analisi della covarianza (ancova)

13.12. Lettura di tre tabulati di programmi informatici

### 14. MISURE DI TENDENZA NON PARAMETRICA E DI ASSOCIAZIONE TRA VARIABILI

14.1. Il test di Cox e Stuart

14.2. Test di Jonckheere o Jonckheere-Terpstra per alternative ordinate in  $k$  campioni indipendenti

14.3. Il test di Page per alternative ordinate

14.4. Le misure d'associazione o d'indipendenza

14.5. Associazione in tabelle  $2 \times 2$  o fra variabili dicotomiche: il  $Q$  e l' $Y$  di Yule, il  $j$ , il  $D_{sim}$  e il  $d_{xy}$  di Somers, il  $t_c$ ; il  $f$  e il  $c$  di Pearson, il  $v$  di Cramer, il  $d_t$  di Tschuprow

14.6. Il cross-product ratio (CPR)

14.7. Associazione per variabili categoriali in tabelle  $M \times N$ : la PRE, il  $I$  simmetrico ed asimmetrico di Goodman e Kruskal, le misure fondate sul chi quadrato

14.8. Cograduazione per variabili ordinali in tabelle  $M \times N$ : il  $g$  di Goodman e Kruskal, il  $t_k$  di Kendall, il  $d_{ba}$  e il  $d_{ab}$  di Somers

14.9. Stima dell'accordo con scala nominale: il kappa di Cohen

14.10. Lettura dei tabulati di un pacchetto statistico

### 15. TEST NON PARAMETRICI PER CORRELAZIONE, CONCORDANZA E REGRESSIONE LINEARE

15.1. La correlazione non parametrica  $\rho_s$  di Spearman con la distribuzione di Hotelling-Pabst e il test di Daniels

- 15.2. Il coefficiente di correlazione  $\tau$  di Kendall, con  $t_a$  e  $t_b$ . Il  $c^2$  di Mantel-Haenszel
- 15.3. Il coefficiente di correlazione parziale  $\tau$  di Kendall
- 15.4. Cenni su misure di concordanza tra più valutatori: la W, la u di Kendall
- 15.5. Misure per dati d'intervallo con classi costanti: il coefficiente di correlazione r di Pearson e il coefficiente eta
- 15.6. Odds ratio: rapporto di probabilità o tra rischi
- 15.7. Regressione non parametrica
- 15.8. Calcolo della retta di regressione non parametrica con il metodo di Theil
- 15.9. Confronto fra la retta parametrica e la retta di Theil
- 15.10. Test di Theil per la significatività di b
- 15.11. Il test di Hollander

## 16. IL DISEGNO SPERIMENTALE NELLA RICERCA AMBIENTALE; METODI DI RICAMPIONAMENTO PER L'INFERENZA

- 16.1. I motivi del disegno sperimentale
- 16.2. Concetti sull'analisi sequenziale
- 16.3. Campione e superpopolazione
- 16.4. Stimatori e loro proprietà
- 16.5. Test di normalità, simmetria e curtosi: il test di Lilliefors e il  $c^2$  per l'adattamento di un campione ad una distribuzione normale
- 16.6. Campioni non probabilistici e campioni probabilistici
- 16.7. Tipo di campionamento e varianza
- 16.8. La scelta del test
- 16.9. La perdita di soggetti nel corso dell'indagine
- 16.10. Metodi di ricampionamento per l'inferenza: il jackknife e il bootstrap

Lettere dell'alfabeto greco antico

<b>A</b>	<b>α</b>	alpha
<b>B</b>	<b>β</b>	beta
<b>Γ</b>	<b>γ</b>	gamma
<b>Δ</b>	<b>δ</b>	delta
<b>E</b>	<b>ε</b>	epsilon
<b>Z</b>	<b>ζ</b>	zeta
<b>H</b>	<b>η</b>	eta
<b>Θ, θ</b>	<b>ϑ, ϐ</b>	theta
<b>I</b>	<b>ι</b>	iota
<b>K</b>	<b>κ</b>	kappa
<b>Λ</b>	<b>λ</b>	lambda
<b>M</b>	<b>μ</b>	mu (my, mi)
<b>N</b>	<b>ν</b>	nu (ny, ni)
<b>Ξ</b>	<b>ξ</b>	xi (csi)
<b>O</b>	<b>ο</b>	omicron
<b>Π</b>	<b>π</b>	pi
<b>P</b>	<b>ρ</b>	rho
<b>Σ</b>	<b>σ, ς</b>	sigma
<b>T</b>	<b>τ</b>	tau
<b>Υ</b>	<b>υ</b>	upsilon (ypsilon)
<b>Φ</b>	<b>φ</b>	phi
<b>X</b>	<b>χ</b>	chi
<b>Ψ</b>	<b>ψ</b>	psi
<b>Ω</b>	<b>ω</b>	omega

# FONDAMENTI DI STATISTICA APPLICATA ALL'ANALISI E ALLA GESTIONE DELL'AMBIENTE

## CAPITOLO I

### ELEMENTI DI STATISTICA DESCRITTIVA PER DISTRIBUZIONI UNIVARIATE

#### 1.1. LA STATISTICA NELLA RICERCA AMBIENTALE.

Come in tutta la ricerca scientifica sperimentale, anche nelle scienze ambientali ed in ecologia è indispensabile la conoscenza dei concetti e dei metodi statistici, sia per i problemi di gestione che di indagine. Per pubblicare i risultati di una ricerca, tutte le riviste scientifiche del settore richiedono che la **presentazione dei dati** e la **loro elaborazione** seguano criteri riconosciuti come universalmente validi. Il comportamento nella **fase di raccolta, la descrizione, l'analisi ed il riepilogo dei dati** sono in buona parte codificati, fino nei dettagli. Inviare ad una rivista uno studio od una relazione che denotino una conoscenza sommaria della statistica comporta generalmente una critica dei metodi seguiti, che può giungere fino al rifiuto delle conclusioni o almeno ad una dichiarazione esplicita sulla loro ridotta attendibilità.

**Con una raccolta di dati non corretta, una loro presentazione inadeguata o un'analisi statistica non appropriata diviene impossibile la verifica dei risultati da parte di altri studiosi ed il confronto con altre ricerche ed analisi del settore.** Per il progresso di qualsiasi disciplina sperimentale, anche la semplice possibilità di sommare le esperienze e confrontare i risultati di ricerche diverse, sia in condizioni simili che volutamente differenti, è una finalità importante per l'accumulo delle conoscenze, per formulare nuove ipotesi o verificare teorie già divulgate.

Al fine di facilitare ai lettori la corretta comprensione dei risultati, per la divulgazione delle ricerche presentate le riviste internazionali e quelle di maggior prestigio richiedono tassativamente agli autori di seguire uno schema preciso che, in linea di massima, è fondato sullo sviluppo di quattro elementi.

1) Una **introduzione**, che presenti in modo accurato sia l'**argomento** affrontato, sia le **finalità** della ricerca, mediante citazione dei lavori scientifici pregressi e della **letteratura** specifica.

2) La descrizione di **materiali e metodi**, in cui devono essere definiti

- (a) il tipo di **scala** utilizzato,

- (b) le modalità del **campionamento** o di raccolta dei dati,
- (c) le misure sintetiche delle caratteristiche più importanti della **distribuzione dei dati**, come media e varianza (più raramente, anche simmetria, curtosi, coefficiente di variazione). Spesso, soprattutto per argomenti nuovi o quando siano stati pubblicati solo pochi dati, è prassi richiedere la distribuzione tabellare completa e dettagliata; tabelle e frequenze sono preferite alle rappresentazioni grafiche, che possono essere anche particolareggiate, ma raramente permettono di risalire ai dati originari, indispensabili per verificare i calcoli.

3) I **risultati**, che devono comprendere espressamente la citazione, con bibliografia nei casi meno noti, dei **test di inferenza** utilizzati, allo scopo di permettere alla comunità scientifica di valutare se la loro scelta è appropriata, in funzione

- (a) delle ipotesi che si intendono verificare,
- (b) del tipo di scala con cui sono state misurate le variabili analizzate,
- (c) delle caratteristiche statistiche della distribuzione dei dati.

4) La **discussione**, che deve riportare l'**interpretazione** dei risultati ottenuti con i test applicati, oltre ad eventuali confronti, concordanze o discordanze con analisi già pubblicate. **L'interpretazione deve non solo comprendere l'analisi statistica, ma essere estesa al significato ecologico od ambientale dei risultati ottenuti**; infatti non sempre un risultato statisticamente rilevante assume anche un importante significato ecologico od ambientale. Ne deriva che per analizzare i dati ambientali non è possibile scindere l'analisi statistica dalla interpretazione ecologica.

## 1.2. IL DISEGNO SPERIMENTALE ED IL CAMPIONAMENTO

Per condurre in modo corretto una ricerca scientifica, ottenere tutti i dati richiesti ed applicare le analisi secondo i criteri descritti, occorre seguire alcuni passaggi metodologici, riassumibili in 4 fasi:

- il disegno sperimentale,
- il campionamento,
- la descrizione statistica,
- la scelta dei test per l'inferenza.

1 - Il **disegno sperimentale**, per fare in modo che le osservazioni in natura e le ripetizioni in laboratorio non siano raccolte a caso, ma scelte e programmate in funzione della ricerca e delle ipotesi esplicative. Già nella programmazione dell'esperimento, nella fase chiamata con termine tecnico "disegno sperimentale" dall'inglese *experimental design* (tradotto più correttamente in

italiano con **programmazione dell'esperimento**), occorre avere chiara a priori la formulazione dell'ipotesi che si intende verificare, alternativa all'ipotesi nulla.

Con essa si deve rispondere alle domande: *“Le eventuali differenze riscontrate tra due o più gruppi di dati, oppure di una serie di osservazioni con quanto era atteso, possono essere imputabili a fattori causali specifici o solamente a fattori casuali ignoti? Le differenze riscontrate sono generate dalla naturale variabilità delle misure e del materiale utilizzato oppure probabilmente esiste una causa specifica che le ha determinate?”*

2 - Il **campionamento**, che permette di raccogliere i dati in funzione dello scopo della ricerca, rispettando le caratteristiche della **popolazione o universo** dei dati.

Il **problema fondamentale della statistica** è come raccogliere solamente un numero limitato di dati (per motivi di risparmio, di tempo, di dati effettivamente disponibili), **ma attraverso la loro analisi pervenire ugualmente a conclusioni generali**, che possano essere estese a tutta la popolazione.

3 - La **descrizione** delle caratteristiche statistiche **dell'insieme dei dati raccolti**, in modo che tutti possano verificare sia l'adeguatezza del disegno sperimentale e del campionamento, sia la correttezza delle analisi attuate e dei risultati ottenuti.

4 - L'**utilizzo dei test** già programmati nel disegno sperimentale, in funzione dei quali è stato effettuato il campionamento. Si tratta di un processo logico-matematico che, mediante il calcolo di probabilità specifiche, porta alla conclusione di non poter respingere oppure di respingere l'ipotesi della casualità.

Chiamata **ipotesi nulla** ed indicata con  $H_0$ , di norma tale ipotesi afferma che le differenze tra gruppi o le tendenze riscontrate siano imputabili essenzialmente al caso. Per giungere a queste conclusioni si deve ricorrere all'**inferenza**, che può essere definita come **la capacità di trarre conclusioni generali (sulla popolazione od universo) utilizzando solo un numero limitato di dati variabili (campione)**.

**Il disegno sperimentale ed il campionamento sono le due fasi preliminari sia alla raccolta dei dati in natura sia ad una corretta impostazione degli esperimenti in laboratorio**, per quasi tutte le ricerche a carattere ambientale ed ecologico.

Tuttavia, la presentazione didattica e la corretta comprensione di questi argomenti richiedono concetti complessi e metodologie sofisticate, non sempre facili né intuitivi. Per questi motivi, il disegno

sperimentale ed il campionamento sono sempre trattati nella fase finale del corso, quando è già stata raggiunta sufficiente familiarità con la terminologia statistica, i concetti ed i metodi fondamentali dell'inferenza.

Nell'apprendimento e nell'uso della statistica, il primo passo è comprendere come solamente con una corretta applicazione del campionamento e dei test di confronto statistico, scelti ed organizzati aprioristicamente nel disegno sperimentale, sia possibile rispondere alla **domanda inferenziale** di verifica dell'**ipotesi nulla**. Con essa si pone il seguente quesito:

*" Nell'ipotesi che le differenze fra gruppi di osservazioni empiriche siano dovute a fattori esclusivamente casuali, quale è la probabilità che fra tutte le alternative possibili si presenti proprio la situazione descritta dai dati raccolti o ancora più estrema?"*

Se la probabilità risulta alta, convenzionalmente uguale o superiore al 5%, si imputeranno le differenze a fattori puramente casuali; al contrario, se la probabilità risulta bassa, inferiore al valore prefissato, si accetta come più verosimile che le differenze siano dovute a fattori non casuali, rientranti tra i criteri con cui i dati sono stati raggruppati.

La procedura dell'inferenza statistica è semplice, nelle linee logiche generali. Tuttavia, le analisi e le conclusioni trovano complicazioni per l'**elevata variabilità dei dati ambientali**, a motivo **fondamentalmente di tre cause che, in ordine crescente d'importanza, sono:**

- gli **errori di misurazione**, generati da strumenti e da differenze nell'abilità dei ricercatori;
- l'**operare su campioni**, per cui i dati utilizzati in una ricerca non sono mai identici a quelli rilevati in qualsiasi altra;
- la presenza di vari **fattori contingenti di disturbo** che, come il tempo e la località, possono incidere diversamente sul fenomeno in osservazione, con intensità e direzioni ignote.

Pure se espressi in modo sintetico, questi concetti definiscono il **contenuto della statistica moderna**: *la raccolta, la presentazione e la elaborazione numerica delle informazioni, per agevolare l'analisi dei dati ed i processi decisionali.*

Inscindibile nella conduzione di una ricerca, sotto l'aspetto didattico la **statistica moderna viene da tempo distinta in due parti**:

- 1 - la **statistica descrittiva**, che comprende *l'insieme dei metodi che riguardano la raccolta, la presentazione e la sintesi di un insieme di dati per descriverne le caratteristiche essenziali;*

2 - la **statistica inferenziale**, che comprende *l'insieme dei metodi con cui si possono elaborare i dati dei campioni per dedurre omogeneità o differenze nelle caratteristiche analizzate, al fine di estendere le conclusioni alla popolazione.*

In altri termini, **la statistica inferenziale permette di trarre conclusioni su tutti i dati di una popolazione, quando si conoscono solamente pochi dati**, raggruppati in uno o più campioni. Supponiamo di voler conoscere la velocità d'accrescimento somatico di una determinata specie animale o vegetale. E' ovvio che non è possibile rintracciare e misurare tutti gli individui di quella specie, la popolazione od universo; se non altro per il tempo e le risorse necessari ed il suo continuo rinnovamento per nascite e morti. E' possibile utilizzare per la ricerca solamente alcune unità, una frazione limitatissima della popolazione, in termini tecnici un campione.

Quando poi si trattasse di misurare rapporti tra organi di una specie animale, è ovvio che non è possibile sezionare tutti gli individui della specie. Nello stesso modo, per contare i globuli rossi o quelli bianchi di una persona, non è possibile estrarre tutto il sangue per un conteggio totale, ma si effettua un prelievo limitato a pochi centimetri cubici.

Tuttavia **le conclusioni non devono essere limitate ai pochi (o anche molti) casi realmente raccolti, misurati ed analizzati; ma devono essere generali**, estese a tutti gli individui della specie o a tutto l'organismo.

**Ricoprono effettivo interesse non le conclusioni che restano limitate ai casi del campione utilizzato, ma quelle che sono estese a tutta la popolazione o universo.** Solo in questo modo, la ricerca riveste una importanza generale e contribuisce alla costruzione di teorie scientifiche, di modelli o semplicemente di ipotesi che possono essere universalmente validi.

Una condizione essenziale e preliminare all'uso dei metodi di statistica inferenziale è che **il campione sia corretto, che non riporti in modo distorto od alterato la frequenza delle caratteristiche presenti nella popolazione.**

La **teoria della probabilità** permette poi di verificare la **verosimiglianza** che i risultati del campione non si discostino dagli eventuali risultati raggiunti, analizzando tutta la popolazione o l'universo dei dati.

In un corso completo di statistica applicata alla ricerca e alla gestione dell'ambiente, è importante avere in ogni momento una visione complessiva degli argomenti già discussi e di quelli ancora da affrontare. Il loro elenco è utile anche per comprendere le diverse parti in cui viene distinta la statistica:

- 1 - la **statistica descrittiva** che tratta come i dati raccolti devono essere riportati in tabella, rappresentati in grafici, sintetizzati in indici matematici per individuare le caratteristiche fondamentali della distribuzione dei dati;
- 2 - la **statistica matematica** o presentazione delle distribuzioni teoriche, per evidenziare quali siano le caratteristiche fondamentali di alcune fondamentali distribuzioni teoriche sia discrete sia continue, quali relazioni esistano tra esse, quali usi abbiano nella ricerca ambientale;
- 3 - **l'inferenza statistica**, la parte nettamente prevalente del corso, che può essere distinta in varie sezioni, in rapporto alle caratteristiche dei dati (se permettono o meno il ricorso alla distribuzione normale) e al fatto di considerare una sola variabile (distribuzioni univariate) oppure di analizzare le relazioni tra due variabili (distribuzioni bivariate) o tra più variabili contemporaneamente (statistica multivariata).

La prima parte dell'inferenza è la **statistica univariata parametrica**, che comprende il test t di Student e il test F di Fisher-Snedecor o analisi della varianza: il primo serve sia per confrontare la media di un campione con una media attesa o teorica, sia per confrontare le medie di due campioni; il secondo rappresenta la sua generalizzazione e permette il confronto simultaneo tra più medie, considerando contemporaneamente sia uno sia più fattori di variabilità. Appunto perché fondati sulla distribuzione normale, richiedono condizioni di validità restrittive (discusse nei capitoli successivi), che non sempre i dati raccolti e la misura utilizzata permettono di rispettare.

E' una situazione che si presenta con frequenza elevata nella ricerca ecologia e in quella ambientale, a causa della estrema variabilità dei dati e la presenza di valori anomali; si ricorre allora alla **statistica univariata non parametrica**, che è formata da una serie innumerevole di test, di norma raggruppati sulla base dei campioni ai quali viene applicata: test per un campione, per due campioni dipendenti e indipendenti, test per k campioni dipendenti e indipendenti.

Quando per ogni individuo o situazione si raccolgono informazioni relative a due variabili, è possibile analizzare le relazioni che intercorrono tra esse, mediante sia la regressione e la correlazione parametriche sia la regressione e la correlazione non parametriche. Si parla allora di **statistica bivariata parametrica** e di **statistica bivariata non parametrica**. Quando i dati raccolti sono relativi a più variabili, si deve ricorrere alla **statistica multivariata**, che per molte analisi è solamente **parametrica**.

In questo corso, verranno presentati i metodi relativi alla statistica univariata e bivariata sia parametrica che non parametrica. La serie completa degli argomenti ed il loro ordine sono riportati nell'indice del volume.

### 1.3. TIPI DI DATI E SCALE DI MISURAZIONE

Nell'analisi statistica, già al primo approccio occorre porre sempre molta attenzione alle caratteristiche dei dati, poiché da esse dipendono sia i metodi di descrizione, sia i test da applicare.

**Esistono fondamentalmente due tipi di variabili casuali, alle quali sono associati due tipi di dati: qualitativi e quantitativi.**

I **dati qualitativi** sono generati da **risposte categoriali** (es.: con un test sulla tossicità, le cavie muoiono o sopravvivono; con un farmaco, entro un tempo prefissato i pazienti guariscono o restano ammalati; con esperimenti sulle leggi dell'ereditarietà di Mendel, si possono ottenere fiori rossi o fiori bianchi).

I **dati quantitativi** sono il risultato di **risposte numeriche** (es.: per un'analisi del dimorfismo animale, le dimensioni di organi o il peso di alcuni maschi e di alcune femmine).

I **dati quantitativi** possono essere **discreti o continui**: i primi derivano da un conteggio (es.: quante foglie sono attaccate ad un ramoscello); i secondi da un processo di misurazione con uno strumento (es.: la **lunghezza** di un ramoscello; il **peso** una cavia; il **tempo** di reazione alla somministrazione di una sostanza tossica).

Questa suddivisione, ormai storica nella presentazione ed elaborazione statistica dei dati, è stata resa più chiara e funzionale dalla classificazione delle scale di misurazione proposta dallo psicologo S.S. Stevens nel 1946, aggiornata nel 1951 con le operazioni statistiche "ammissibili" e in seguito divulgata soprattutto da S. Siegel, nel suo manuale di "*Statistica non parametrica*" del 1956.

Le **misure possono essere raggruppate in 4 tipi di scale**, che godono di proprietà formali differenti; di conseguenza, esse ammettono operazioni differenti. Come per le altre discipline, **una scala di misurazione** dei fenomeni ecologici ed ambientali può essere:

- 1) **nominale o classificatoria;**
- 2) **ordinale o per ranghi;**
- 3) **ad intervalli;**
- 4) **di rapporti.**

1.3.1 - **La scala nominale o classificatoria** è il livello più basso di misurazione; viene utilizzata quando i risultati possono essere classificati o raggruppati in categorie qualitative, nominali, eventualmente identificati con simboli. I caratteri nominali, detti anche "sconnessi", costituiscono variabili le cui modalità o attributi non assumono alcun ordine precostituito. In una popolazione animale si possono distinguere gli organismi in maschi e femmine e contare quanti appartengono ai

due gruppi; oppure possono essere suddivisi e contati secondo la loro specie, con una classificazione a più voci.

Nella scala **nominale o qualitativa**, esiste **una sola relazione**, quella di **identità**: **gli individui attribuiti a classi diverse sono tra loro differenti, mentre tutti quelli della stessa classe sono tra loro equivalenti**, rispetto alla proprietà utilizzata nella classificazione.

Un caso particolare è quello dei caratteri dicotomi o dicotomici che possono assumere solo due modalità, spesso indicate in modo convenzionale con 0 e 1.

L'attribuzione di numeri per identificare le varie categorie nominali, come avviene per individuare i giocatori nei giochi di squadra, è solamente un artificio che non può certamente autorizzare ad elaborare quei numeri come se fossero reali, ad esempio calcolandone la media. Quando per la classificazione dei gruppi al posto di nomi vengono usati numeri, si utilizza solo la funzione di identificazione degli elementi numerici come se fossero simboli; ma con tale trasformazione non si determina una informazione diversa.

**L'operazione ammessa è il conteggio** degli individui o dei dati presenti in ogni categoria. I quesiti statistici che possono essere posti correttamente riguardano le frequenze, sia assolute che relative. Sono possibili confronti tra frequenze osservate (es.: Una classe è significativamente più numerosa dell'altra? Le varie classi hanno tutte lo stesso numero di individui, escludendo le variazioni casuali?) oppure tra le frequenze osservate e le rispettive frequenze attese sulla base di leggi biologiche, ipotesi od altro (es.: I risultati ottenuti da un esperimento sulle leggi di Mendel sono in accordo con la sua distribuzione teorica?).

1.3.2 - La **scala ordinale o per ranghi** rappresenta una misurazione che contiene una quantità di informazione immediatamente superiore a quella nominale, assumendo modalità logicamente sequenziali, non importa se in ordine crescente o decrescente; alla proprietà precedente di **equivalenza tra gli individui della stessa classe, si aggiunge una gradazione tra le classi o tra individui con misure diverse**.

Con la precedente scala nominale, si ha la sola informazione che gli individui appartenenti a gruppi differenti sono tra loro diversi, ma non è possibile **stabilire un ordine**. Con la scala per ranghi, le differenti classi possono essere ordinate sulla base dell'intensità del fenomeno. (es.: Si supponga che il risultato di un reagente sia di colorare in verde una serie di provette, secondo la quantità di sostanza contenuta. E' possibile mettere in ordine le provette secondo l'intensità del colore, per avere una stima approssimata della quantità di sostanza contenuta. Se si confrontano tre o più provette con intensità di colore differente, è facile stabilirne l'ordine; rimane impossibile confrontare e misurare la quantità di differenza esistente tra esse).

**In una scala ordinale, non è possibile quantificare le differenze di intensità tra le osservazioni.**

Alcune risposte, apparentemente definite a livello qualitativo o nominale, in realtà possono contenere una scala ordinale o di rango, seppure con molte ripetizioni. E' il caso della suddivisione in giovane, adulto ed anziano per l'età, oppure della classificazione in insufficiente, sufficiente, discreto, buono ed ottimo in valutazioni di merito. Contengono una scala ordinale anche misure che sono rappresentate con simboli, come

--, -, =, +, ++.

Resta l'impossibilità di valutare quanto sia la distanza tra insufficiente e sufficiente, oppure se sia inferiore o superiore alla distanza tra buono ed ottimo.

**La scala ordinale o per ranghi è pertanto una scala monotonica**. Alle variabili così misurate è possibile applicare una serie di **test non parametrici**, ma non quelli parametrici. In questi casi, non sarebbe possibile utilizzare quei test che fanno riferimento alla distribuzione normale, i cui parametri essenziali sono la media e la varianza, poiché si fondano sulle differenze di ogni osservazione dalla media.

Tuttavia, questa indicazione di massima è spesso superata dall'osservazione che variabili discrete o nominali tendono a distribuirsi in modo approssimativamente normale, quando il numero di dati è sufficientemente elevato. Per coloro che sono nella fase iniziale delle applicazioni statistiche, permane sempre molta incertezza sulla scelta dei test più appropriati; infatti, tra gli studiosi permane un'ampia varietà di opinioni su quando il numero di osservazioni sia sufficientemente elevato per ottenere una distribuzione normale. Nel seguito del volume, l'argomento sarà discusso in molte situazioni.

1.3.3 - La **scala ad intervalli** alle due caratteristiche della scala ordinale aggiunge quella di **misurare le distanze o differenze** tra tutte le coppie di valori. La scala di intervalli si fonda su una misura oggettiva e costante, anche se il punto di origine e l'unità di misura sono arbitrari. Esempi classici di scale ad intervalli sono la temperatura (misurata in gradi Celsius o Fahrenheit) ed il tempo (misurato secondo calendari differenti). Valori di temperatura, oltre a poter essere facilmente ordinati secondo l'intensità del fenomeno, godono della proprietà che **le differenze tra loro sono direttamente confrontabili e quantificabili**; le date in un calendario gregoriano, islamico, ebraico o cinese possono essere tra loro ordinate dalla più antica a quella più recente e le **differenze** temporali sono **misurate con precisione oggettiva**.

Ma la scala ad intervalli ha alcuni limiti, non gode di altre proprietà.

Ad esempio, una temperatura di 80 gradi non è il doppio di una di 40 gradi, quando riferita alla temperatura corporea: se una persona ponesse la mano destra in una bacinella con acqua a 80 gradi e

la mano sinistra in una con acqua a 10 gradi, non direbbe certamente che la prima scotta 8 volte più della seconda, ma solo che la prima è bollente e la seconda è fredda.

In una **scala ad intervalli, solo le differenze tra i valori sono quantità continue ed isomorfe** alla struttura dell'aritmetica. **Solo per le differenze sono permesse tutte le operazioni:** possono essere tra loro sommate, elevate a potenza oppure divise, determinando le quantità che stanno alla base della statistica parametrica.

Da una scala d'intervalli è possibile scendere ad una scala di ranghi (es.: utilizzando solo l'informazione dell'ordine dei valori) oppure ad una scala nominale (es.: suddividendo in misure alte e basse, sopra o sotto un valore prefissato). Pertanto, la scala d'intervalli gode anche delle proprietà definite per le due scale precedenti.

**Nella presentazione dei test non parametrici verranno discusse le situazioni in cui è conveniente scendere nel tipo di scala,** pur avendo dati misurati su scale d'intervalli o di rapporti, con una apparente perdita d'informazione.

1.3.4 - **La scala di rapporti ha il vantaggio di avere un'origine reale.** Sono tipiche scale di rapporti l'altezza, la distanza, l'età, il peso, il reddito, più in generale tutte quelle misure in cui 0 (zero) significa quantità nulla. Non solo le differenze, ma **gli stessi valori possono essere moltiplicati o divisi** per quantità costanti, senza che l'informazione di maggiore importanza, il rapporto tra essi, ne risulti alterata.

Alle variabili misurate con una scala di rapporti, il tipo di misurazione più sofisticato e completo, può essere applicato qualsiasi test statistico. Possono essere utilizzati anche la media geometrica ed il coefficiente di variazione, i quali richiedono che il punto 0 (zero) sia reale e non convenzionale.

Pure con una scala di rapporti è possibile scendere nella scala di misurazione, trasformandola in una scala di rango o addirittura qualitativa. Ovviamente, si ha una perdita ancor più rilevante della quantità d'informazione, rispetto alle scale precedenti; di conseguenza, rappresenta un'operazione che deve essere evitata, quando non imposta da altre condizioni dell'analisi statistica o dalle caratteristiche della distribuzione dei dati.

Riassumendo i concetti fondamentali esposti,

- nella **scala nominale**, esistono solo relazioni di *equivalenza*;
- in quella **ordinale**, si aggiungono relazioni di *minore o maggiore di*;
- in quella **ad intervalli** si aggiunge la relazione di *rapporto tra ogni coppia d'intervalli*;
- nella **scala di rapporti** si ha anche la relazione di *rapporto conosciuto tra ogni coppia di valori*.

Come sarà più volte discusso nei prossimi capitoli, anche nella ricerca e nella gestione ambientali occorre porre estrema attenzione al reale significato da attribuire ai valori numerici che vengono utilizzati. Si possono avere numeri che apparentemente hanno le stesse caratteristiche; ma in realtà essi richiedono elaborazioni diverse ed impongono il ricorso a test differenti, per rispondere ai medesimi quesiti. Per esempio, i grammi di una determinata sostanza inquinante sciolta in un litro d'acqua, la percentuale di questa sostanza sul peso complessivo, il punteggio della qualità dell'acqua sono misure che utilizzano scale diverse. Nel primo caso, si ha una classica scala di rapporti ed è possibile usare test parametrici, se la distribuzione dei dati è normale; nel secondo caso, è possibile utilizzare le stesse procedure statistiche e gli stessi test parametrici solamente dopo apposita trasformazione dei valori; nel terzo, si ha una scala di ranghi, poiché la reale informazione fornita da questa serie di punteggi è solo quella di una graduatoria della qualità, nella quale non hanno reale significato né i rapporti né le differenze tra loro.

#### 1.4. CLASSIFICAZIONE IN TABELLE

Un insieme di misure è detto **serie statistica** o **serie dei dati**. Quando la serie non è ordinata, si ha un insieme disordinato di numeri che non evidenzia le caratteristiche fondamentali del fenomeno.

Una sua prima ed elementare elaborazione può essere una distribuzione ordinata di tutti i valori, in modo crescente o decrescente, detta **seriazione**.

Il **valore minimo** e il **valore massimo** insieme permettono di individuare immediatamente il **campo** (od **intervallo**) di **variazione**.

Successivamente, la serie può essere raggruppata in classi, contando quanti valori od **unità statistiche** appartengono ad ogni gruppo o categoria.

Si ottiene una **distribuzione di frequenza o di intensità**, detta anche semplicemente **distribuzione**.

Come prima applicazione, è utile considerare un caso semplice: una **variabile discreta** ottenuta da un conteggio del numero di foglie, germogliate su 45 giovani rami di lunghezza uguale.

Tabella 1. Numero di foglie contate su 45 rami.

5	6	3	4	7	2	3	2	3	2	6	4	3	9	3
2	0	3	3	4	6	5	4	2	3	6	7	3	4	2
5	1	3	4	3	7	0	2	1	3	1	5	0	4	5

Il primo passaggio, quasi intuitivo in una distribuzione discreta, consiste nel **definire le classi**:

- è sufficiente identificare il **valore minimo (0)**, nei dati della tabella) e **quello massimo (9)**,
- contando quante volte compare ogni **modalità di espressione** (cioè quanti sono i rami con un numero di foglie uguali).

Queste informazioni di norma sono presentate in una tabella impostata come la seguente:

Tabella 2. Distribuzione di frequenze assolute e relative delle foglie in 45 rami.

Classe	x	0	1	2	3	4	5	6	7	8	9
<b>freq. assoluta</b>	<b>n</b>	3	3	7	12	7	5	4	3	0	1
<b>freq. relativa</b>	<b>f</b>	0,0 7	0,0 7	0,1 5	0,2 7	0,1 5	0,1 1	0,0 9	0,0 7	0,0 0	0,0 2
<b>freq. cumulata</b>		0,0 7	0,1 4	0,2 9	0,5 6	0,7 1	0,8 2	0,9 1	0,9 8	0,9 8	1,0 0

in cui:

- la **classe** è una modalità di espressione (in questo caso un valore o conteggio);
- la **frequenza assoluta della classe** è il numero di volte con la quale compare ogni valore;
- la **frequenza relativa della classe** è la sua frequenza assoluta divisa per il numero totale;
- la **frequenza cumulata di una classe** (che può essere stimata con quelle assolute e/o con quelle relative) è la somma di tutte le frequenze delle classi minori con quella della classe stessa.

La trasformazione da frequenza assoluta a frequenza relativa risulta utile quando si vogliono confrontare due o più distribuzioni, che hanno un differente numero complessivo di osservazioni.

La frequenza cumulata offre informazioni importanti quando si intende stimare il numero totale di osservazioni inferiore (o superiore) ad un valore prefissato (ad es.: il 71% dei rami ha meno di 5 foglie; il 56% ha un massimo di 3 foglie).

La distribuzione dei dati e la distribuzione delle frequenze cumulate forniscono informazioni non dissimili, essendo possibile passare con facilità dall'una all'altra. Sono diverse nella loro forma, come si vedrà con maggiore evidenza nelle rappresentazioni grafiche. **La prima ha una forma a campana, la seconda una forma a S, di tipo asintotico; si prestano ad analisi differenti** e la scelta è fatta sulla base del loro uso statistico.

**La distribuzione di frequenza offre una lettura rapida delle caratteristiche più importanti della serie di dati.** Nella tabella precedente, il ramo “tipico” ha 3 foglie; se dovessimo sintetizzare con un solo valore il numero di foglie presenti sui rami raccolti diremmo 3, che rappresenta la **tendenza centrale**. Altra caratteristica importante è il numero minimo e il numero massimo, 0 e 9, che insieme forniscono il campo di variazione, una indicazione della **variabilità o dispersione**. La distribuzione del numero di foglie tende ad diminuire in modo simile allontanandosi da 3, seppure mantenga frequenze più alte nelle classi con un numero maggiore di foglie: sono indicazioni sulla **forma** della distribuzione, che in questo esempio non è simmetrica (ma asimmetrica) rispetto alla tendenza centrale, a causa di un eccesso dei valori più alti.

Nella costruzione di tabelle sintetiche (come la tabella 2 rispetto alla 1) uno dei problemi più rilevanti è **quante classi di frequenza costruire**. La scelta dipende strettamente dal numero totale **N** di osservazioni e, in misura minore, dalla variabilità dei dati.

Se, in riferimento alla dimostrazione precedente, i dati fossero stati in numero inferiore ai 45 presentati (ad esempio i 15 valori della prima riga), il campo di variazione sarebbe stato più ridotto (non più da 0 a 9, ma da 2 a 9). Le classi non sarebbero state 10 come prima, ma solamente 8. Tuttavia, come si può osservare dai dati, 8 classi per 15 osservazioni sarebbero ugualmente un numero troppo alto, per riuscire ad evidenziare e rappresentare in modo corretto le caratteristiche principali e la forma reale della distribuzione.

**Le distribuzioni di frequenza tendono a mostrare la distribuzione reale del fenomeno solo quando è possibile utilizzare un numero sufficientemente elevato di osservazioni.**

**L'esperienza ha insegnato che il numero di classi abitualmente varia da un minimo di 4-5 (con  $N = 10-15$ ) ad un massimo di 15-20 (con  $N > 100$ ), in funzione del numero complessivo di osservazioni.**

Un numero troppo basso di classi, raggruppando eccessivamente i dati, determina una perdita di informazione sulle caratteristiche della distribuzione e la rende non significativa; è intuitivo che una o due sole classi determinano l'impossibilità di evidenziare qualunque caratteristica della distribuzione. Inversamente, ma con un risultato finale simile, un numero troppo elevato di classi disperde i valori e non rende manifesta la forma della distribuzione.

Per stimare in modo oggettivo il numero di classi, sono stati proposti vari metodi; tra essi è utile ricordarne due:

1 - quello di **H. Sturges** che nel 1926, sulla base del numero di osservazioni **N**, ha indicato il **numero ottimale di classi** **C** con

$$C = 1 + \frac{10}{3} \cdot \log_{10}(N)$$

2 - quello di **D. Scott** che nel 1979 ha determinato l'**ampiezza ottimale h delle classi** (dalla quale ovviamente dipende direttamente anche il numero di classi **C**), mediante la relazione

$$h = \frac{3,5 \cdot s}{\sqrt{N}}$$

dove

- **s** è la deviazione standard,

che sarà presentata più avanti tra le misure di variabilità dei dati.

Nella costruzione di distribuzioni di frequenza, non è strettamente obbligatorio utilizzare intervalli uguali, anche se è prassi consolidata per una lettura più semplice. Nel caso di classi di ampiezza diversa, la rappresentazione grafica ed il calcolo dei parametri fondamentali esigono alcune avvertenze, non sempre intuitive (di seguito presentate).

Nel caso di una **variabile continua**, il raggruppamento in classi richiede alcuni accorgimenti ulteriori rispetto a quelli utilizzati per una variabile discreta. Si supponga che sia stata misurata l'altezza in cm. di 40 giovani piante della stessa specie, arrotondata all'unità per semplificazione.

Tabella 3. Altezza in cm. di 40 giovani piante.

107	83	100	128	143	127	117	125	64	119
98	111	119	130	170	143	156	126	113	127
130	120	108	95	192	124	129	143	198	131
163	152	104	119	161	178	135	146	158	176

E' evidente come non sia conveniente fare una classe per ogni cm., in analogia a quanto fatto con i dati della tabella 1. In questo caso, il numero di modalità sarebbe nettamente superiore al numero di osservazioni, anche se il campione avesse un numero di osservazioni doppio o triplo. Di conseguenza, si impone la necessità di un raggruppamento in classi, che comprendano più modalità di espressione.

Una volta individuato il valore minimo e quello massimo (64 e 198), si stabilisce l'intervallo di variazione ( $198 - 64 = 134$ ). Nella formazione delle classi, il limite inferiore della prima classe ed il limite superiore dell'ultima classe non devono essere necessariamente i valori osservati, ma li devono ovviamente comprendere. E quindi possibile costruire un campo di variazione, ad esempio di 140 cm. (sempre più ampio di quello calcolato), partendo da cm. 60 e arrivando a cm. 199 compresi. Sulla base del numero di dati (40), si decide il numero di classi. Nel caso specifico, potrebbero essere 7 classi, con un'ampiezza di 20 cm. ognuna.

E' necessario **definire con precisione il valore minimo e quello massimo di ogni classe**, onde evitare incertezze nell'attribuzione di un singolo dato tra due classi contigue. Con i dati dell'esempio, le classi possono essere 60-79 la prima, 80-99 la seconda, 100-119 la terza e così via fino a 180-199 per l'ultima.

E' da evitare la suddivisioni in classi come 60-80, 80-100, 100-120, ...

Poiché la scala è continua, i cm. riportati devono essere intesi con almeno 2 cifre decimali, per cui nella classe 60-79 il primo numero deve essere inteso come 60,00 cm. e 79 come 79,99; nello stesso modo la classe 180-199 deve essere intesa tra i cm. 180,00 e 199,99.

Nonostante le indicazioni di massima presentate, la determinazione dei valori estremi, del numero di classi e dell'intervallo di ogni classe è ampiamente soggettiva. Nella costruzione di una tabella, **la scelta soggettiva di una particolare serie o di un'altra può tradursi in una rappresentazione completamente diversa degli stessi dati**. Per piccoli campioni, l'alterazione e le differenze possono essere sensibili; ma all'aumentare del numero di osservazioni, gli effetti delle scelte soggettive, quando non siano estreme, incidono sempre meno sulla concentrazione dei valori e sulla forma della distribuzione.

Tra le altre avvertenze importanti, è da ricordare che la classe iniziale e quella terminale non devono essere **classi aperte** (come  $< 80$  quella iniziale e  $\geq 180$  quella finale). Con classi estreme aperte, si perde l'informazione del loro valore minimo o massimo e quindi del valore centrale di quella classe; la conseguenza è la perdita di un dato indispensabile, per calcolare la media della classe e quella totale, nonché tutti gli altri parametri da essa derivati. Come verrà successivamente chiarito, con tabelle in cui le classi estreme sono aperte viene impedita o resa soggettiva anche la loro

rappresentazione grafica, per la quale è indispensabile conoscere con precisione il valore iniziale e quello terminale.

I dati della tabella 3 possono essere riportati in modo più schematico e più comprensibile, come nella seguente tabella 4.

Tabella 4. Distribuzione di frequenza assoluta e relativa (in %) dell'altezza di 40 giovani piante.

classe	$x_i$	60-79	80-99	100-19	120-39	140-59	160-79	180-99
freq. ass.	$n_i$	1	3	10	12	7	5	2
freq. rel.	$f_i$	2,5	7,5	25,0	30,0	17,5	12,5	5,0
freq. cumulata		2,5	10,0	35,0	65,0	82,5	95,0	100,0

Rispetto all'elenco grezzo dei dati, la tabella di distribuzione delle frequenze fornisce in modo più chiaro le indicazioni elementari contenute, in particolare la loro

- **posizione o dimensione** (già chiamata anche tendenza centrale) e
- **la variabilità o dispersione.**

Per evidenziare sia queste che altre caratteristiche della distribuzione dei dati raccolti, sovente è di aiuto una rappresentazione grafica che mostra in modo sintetico soprattutto

- la **forma**, come la **simmetria** e la **curtosi**, quando si tratti di grandi gruppi di dati.

Ritornando al problema della rappresentazione tabellare dei dati riportati in tabella 3, secondo le indicazioni di **Sturges** il **numero di classi C** avrebbe dovuto essere

$$C = 1 + \frac{10}{3} \cdot \log_{10}(N) = 1 + \frac{10}{3} \cdot \log_{10}(40) = 6,34$$

uguale a 6,34

dal quale si deduce anche un'ampiezza  $h = \frac{140}{6,34} \cong 22$

circa 22 centimetri.

Secondo le indicazioni di **Scott**, l'**ampiezza h delle classi** avrebbe dovuto essere

$$h = \frac{3,5 \cdot s}{\sqrt{N}} = \frac{3,5 \cdot 28,618}{6,3246} = 15,837$$

uguale a circa 16,

dalla quale si deduce un numero di classi  $C = \frac{140}{15,837} = 8,84$

uguale a 9 (8,84).

Ovviamente, il numero di classi calcolato ( $C = 8,84$ ) deve essere arrotondato all'unità.

Secondo i due metodi proposti, con i dati della tabella 3 il numero di classi può ragionevolmente variare da 6 a 9; si evidenzia la correttezza della scelta di fare 7 classi, suggerita dalla semplicità di formare classi con un'ampiezza di 20 cm.

La rappresentazione dei dati in una tabella di frequenza offre i vantaggi descritti; ma soffre anche di alcune controindicazioni. Lo svantaggio maggiore deriva dal **non poter conoscere come sono distribuiti i dati entro ogni classe.**

Per stimare i parametri della distribuzione (media, varianza, simmetria, curtosi), viene quindi usato **il valore centrale** di ogni classe, **nell'ipotesi che in quell'intervallo i dati siano distribuiti in modo uniforme.** Rispetto alla distribuzione delle singole osservazioni, questa procedura comporta un'approssimazione, poiché **tale ipotesi operativa implicita non è vera** (il concetto sarà ripreso e dimostrato in paragrafi successivi).

### **1.5. RAPPRESENTAZIONI GRAFICHE DI DISTRIBUZIONI UNIVARIATE**

Le rappresentazioni grafiche servono per evidenziare in modo semplice, *a colpo d'occhio*, le quattro caratteristiche fondamentali di una distribuzione di frequenza (tendenza centrale, variabilità, simmetria e curtosi). Insieme con i vantaggi di fornire una visione sintetica e di essere di facile lettura, hanno però **l'inconveniente fondamentale di mancare di precisione e soprattutto di essere soggettive**, quindi di permettere letture diverse degli stessi dati. Pertanto, ai fini di una elaborazione mediante i test e di un confronto dettagliato dei parametri, è sempre preferibile la tabella, che riporta i dati esatti.

Le **rappresentazioni grafiche** proposte sono numerose. Esse debbono essere scelte in rapporto al tipo di dati e quindi alla scala utilizzata.

Per **dati quantitativi**, riferiti a variabili continue misurate su scale ad intervalli o di rapporti, di norma si ricorre a **istogrammi o poligoni**.

Gli istogrammi sono grafici a barre verticali (per questo detti anche diagrammi a rettangoli accostati). Le misure della variabile casuale sono riportate lungo l'asse orizzontale, mentre l'asse verticale rappresenta il numero assoluto, oppure la frequenza relativa o quella percentuale, con cui compaiono i valori di ogni classe.

I lati dei rettangoli sono costruiti in corrispondenza degli estremi di ciascuna classe.

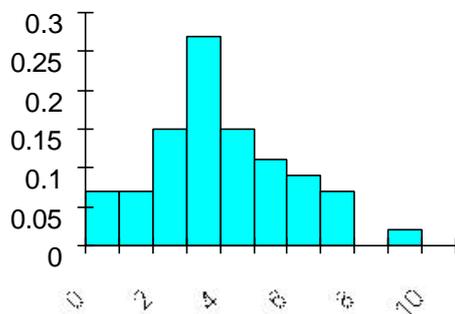


Figura 1. Istogramma dei dati di Tab. 2  
(frequenze relative)

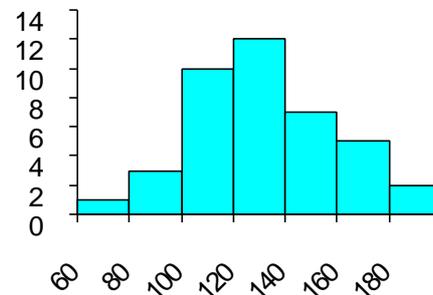


Figura 2. Istogramma dei dati di Tab. 4  
(Valore iniz. =60; Valore finale =199; Passo =20; Classi=7)

**Un istogramma** deve essere inteso come una **rappresentazione areale**: sono le superfici dei vari rettangoli che devono essere proporzionali alle frequenze corrispondenti. Quando le classi hanno la stessa ampiezza, le basi dei rettangoli sono uguali; di conseguenza, le loro altezze risultano proporzionali alle frequenze che rappresentano. Solo quando le basi sono uguali, è indifferente ragionare in termini di altezze o di aree di ogni rettangolo; ma se le ampiezze delle classi sono diverse, bisogna ricordare il concetto generale che le frequenze sono rappresentate dalle superfici e quindi è necessario rendere l'altezza proporzionale. Tale proporzione è facilmente ottenuta dividendo il numero di osservazioni per il numero di classi contenute nella base, prima di riportare la frequenza sull'asse verticale.

Per esempio, con i dati della precedente figura 2, si supponga di avere raggruppato in una classe sola le frequenze della classe da 80 a 99 e da 100 a 119, per un totale di 13 osservazioni (3 + 10)

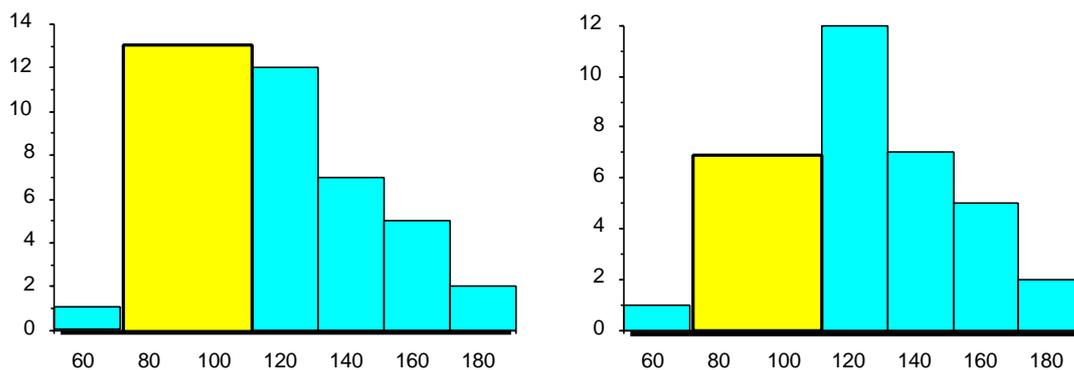


Figura 3. Istogrammi dei dati di Tab. 4

Somma errata di due classi : 2<sup>a</sup> e 3<sup>a</sup>  
della figura precedente

Somma corretta di due classi : 2<sup>a</sup> e 3<sup>a</sup>  
della figura precedente

Nella figura 3, tale somma è rappresentata nel primo caso con un grafico errato e nel secondo caso nella sua versione corretta, che richiede il valore medio delle classi raggruppate.

Un'altra avvertenza importante nella costruzione degli istogrammi è che l'asse verticale, che riporta le frequenze, deve mostrare lo zero reale od "origine", onde non distorcere o travisare le caratteristiche dei dati ed i rapporti tra essi. In relazione alle caratteristiche della distribuzione dei dati, **la larghezza o base del rettangolo non ha alcun significato** e può essere scelta a piacimento; dipende solamente dal numero di classi che si vogliono rappresentare sull'asse delle ascisse.

Anche **il rapporto tra l'altezza dell'asse delle ordinate e la lunghezza delle ascisse può essere scelto a piacimento e non ha alcun significato statistico**. Tuttavia, sulla prassi di costruire figure graficamente eleganti, le dimensioni utilizzate dai programmi informatici seguono uno schema che è ormai uguale per tutti. E' quasi sempre praticato un accorgimento che ha una finalità esclusivamente **estetica**: per costruire una **relazione armonica tra gli elementi del grafico**, è uso corrente che tutto il disegno dell'istogramma debba essere contenuto in un rettangolo virtuale, in cui **l'altezza sia i 2/3 della base** o, come riportano altri testi per fornire lo stesso concetto, **la base sia 1,5 volte l'altezza**. (Gli esempi riportati sono stati costruiti con programmi informatici a grande diffusione internazionale, che seguono questa regola estetica).

La rappresentazione grafica permette di valutare con immediatezza se il numero di classi costruite è adeguato alle caratteristiche della distribuzione originale dei dati. Con poche eccezioni, le variabili quantitative di fenomeni ecologici od ambientali evidenziano una **distribuzione normale**, con

caratteristiche specifiche di addensamento verso i valori centrali e di dispersione più o meno simmetrica, ma con declino regolare verso i due estremi.

**La rappresentazione grafica deve essere in grado di non alterare od interrompere la regolarità della distribuzione**, come può avvenire in particolare quando il numero di classi è troppo alto rispetto al numero di dati.

L'istogramma che segue è una chiara dimostrazione di una suddivisione in classi eccessiva: uno o più gruppi di misure (due nell'esempio) comprese entro gli estremi hanno frequenza zero ed alterano la rappresentazione di una distribuzione normale. La frequenza delle classi e l'altezza dei rettangoli ad essa proporzionali tendono a decrescere in modo relativamente regolare; una forte alterazione, che scompare con suddivisioni in classi meno frammentate, è una indicazione di un possibile errore tecnico di rappresentazione dei dati.

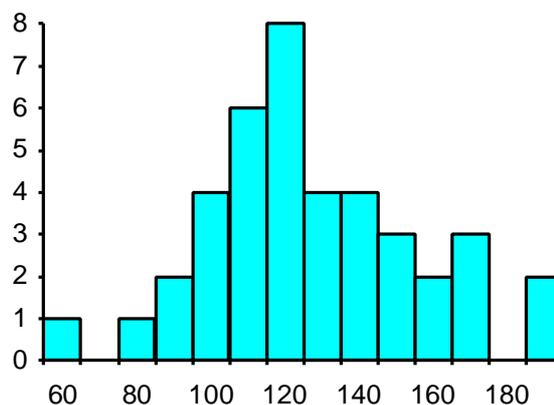


Figura 4. Istogramma dei dati di Tab. 4

(Valore iniziale = 60; Valore finale = 199; Passo = 10; Classi = 14)

(Rappresentazione grafica non adeguata, per eccessiva suddivisione in classi)

**I poligoni** sono figure simili agli istogrammi e sono utilizzati di norma per la rappresentazione di **valori relativi o di percentuali**, in quanto è implicito che l'area totale sottesa sia uguale a 1 o 100%. Come nel caso degli istogrammi, l'asse orizzontale rappresenta il fenomeno, mentre l'asse verticale rappresenta la proporzione o percentuale di ogni classe.

**Un poligono può essere ottenuto a partire dal relativo istogramma**, unendo con una linea spezzata i punti centrali di ogni classe. La linea spezzata deve essere unita all'asse orizzontale, sia all'inizio sia alla fine, per racchiudere l'area della distribuzione. Questo procedimento viene ottenuto con un artificio, utilizzando un istogramma come punto di partenza. Si unisce il valore centrale della

prima classe con il valore centrale di una precedente classe fittizia di valore 0; l'ultimo segmento viene ottenuto unendo il valore centrale dell'ultima classe reale con il valore centrale di una classe successiva, fittizia, di valore 0.

Il primo poligono di seguito riportato (figura 5) corrisponde all'istogramma della figura 2 ed è stato costruito con i dati della tabella 4, spostando le classi sull'asse delle ascisse per comprendere i nuovi estremi della distribuzione.

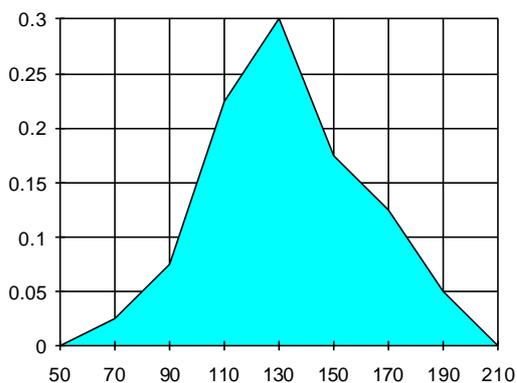


Figura 5. Poligono dei dati di Tab. 4

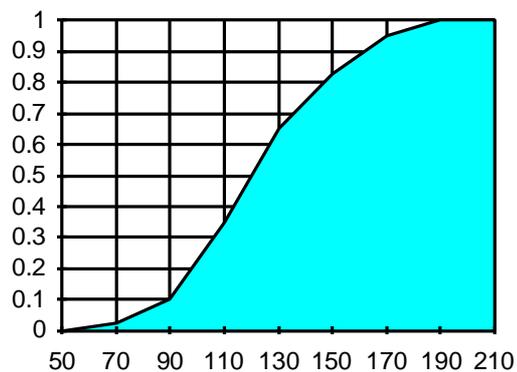


Figura 6. Poligono cumulato di Tab. 4

Le distribuzioni cumulate sono rappresentate sia con istogrammi cumulati sia con **poligoni cumulati**. Non forniscono informazioni differenti da quelle dei relativi istogrammi e poligoni già descritti, poiché è possibile passare con facilità da una distribuzione di frequenza alla sua cumulata con semplici operazioni di somme o di sottrazioni tra classi.

La figura 6 rappresenta il poligono cumulato corrispondente al poligono della figura 5.

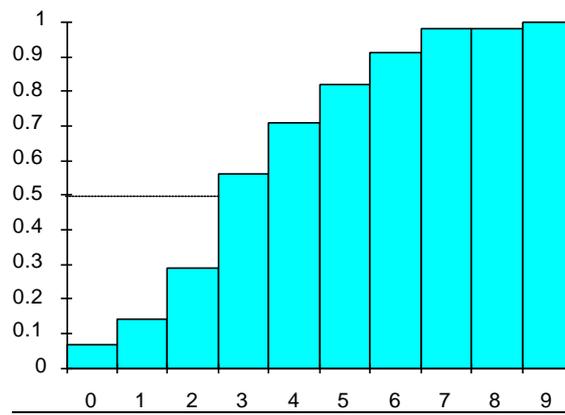


Figura 7. Istogramma cumulato dei dati di Tab. 2

Tuttavia, per la diversa prospettiva che essi offrono a partire dagli stessi dati, gli istogrammi ed i poligoni cumulati sono un altro metodo utile sia per presentare le caratteristiche di dati quantitativi riportati in tabelle, sia per facilitare l'interpretazione e l'analisi. Servono soprattutto per evidenziare, con lettura immediata, quante sono in totale le misure che sono inferiori o superiori ad un certo valore. Il valore dell'asse orizzontale che corrisponde al 50% dei valori identifica la mediana (riportato come linea tratteggiata nella figura 7 che rappresenta un istogramma cumulato); è un parametro di tendenza centrale estremamente importante, quando la distribuzione non è simmetrica (il suo uso e le sue caratteristiche saranno descritte in modo dettagliato nei prossimi paragrafi).

Per le distribuzioni di frequenza di **dati qualitativi**, le rappresentazioni grafiche più frequenti sono

- i **diagrammi a rettangoli distanziati**,
- **gli ortogrammi**,
- i **diagrammi a punti**,
- **gli areogrammi** (tra cui i **diagrammi circolari**),
- i **diagrammi a figure** (o **diagrammi simbolici**).

I **diagrammi a rettangoli distanziati**, detti anche grafici a colonne, sono formati da rettangoli con basi uguali ed altezze proporzionali alle intensità (o frequenze) dei vari gruppi considerati. A differenza degli istogrammi, i rettangoli non sono tra loro contigui, ma distaccati; di conseguenza, sull'asse delle ascisse non vengono riportati misure ordinate ma nomi, etichette o simboli, propri delle classificazioni qualitative.

**Con dati qualitativi o nominali, le basi dei rettangoli sono sempre identiche avendo solo un significato simbolico.**

Si può ricorrere quindi sia a **diagrammi a punti** o *line plot*, in cui i punti sono disposti uno sopra l'altro fino ad un'altezza proporzionale alla frequenza della classe, sia a **diagrammi a barre**, che sono un'altra rappresentazione frequente, in cui al posto di rettangoli o colonne di punti vengono usate linee continue più o meno spesse (figura 8).

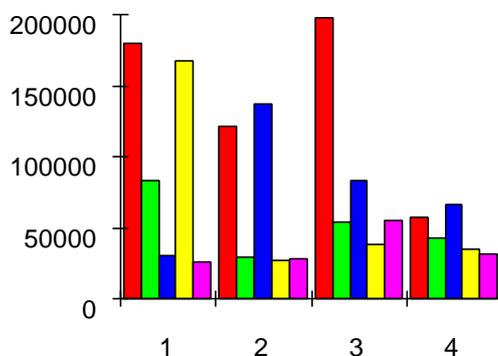


Figura 8. Rettangoli distanziati

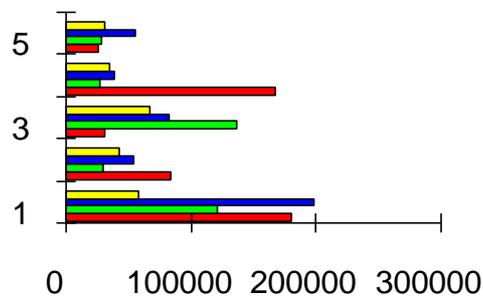


Figura 9. Ortogramma

Nel caso di **dati qualitativi o nominali**, non esiste una logica specifica nell'ordine delle classi.

**Per convenzione, i rettangoli o le colonne sovente (ma non obbligatoriamente) vengono disposti in modo ordinato dal maggiore al minore o viceversa.**

Se le classi qualitative sono composte da **sottoclassi**, è possibile una rappresentazione grafica più articolata, dividendo ogni rettangolo in più parti, con altezze proporzionali alle frequenze delle sottoclassi (figura 10). Avendo basi uguali, le aree sono proporzionali alle altezze; pertanto, anche i diagrammi a rettangoli distanziati sono rappresentazioni areali.

Gli **ortogrammi o grafici a nastri** sono uguali ai rettangoli distanziati; l'unica differenza è che gli assi sono scambiati, per una lettura più facile (figura 9 e figura 11).

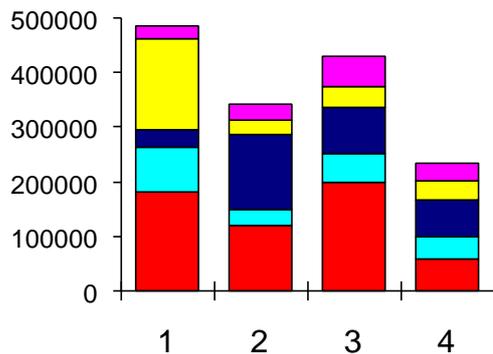


Figura 10. Rettangoli distanziati

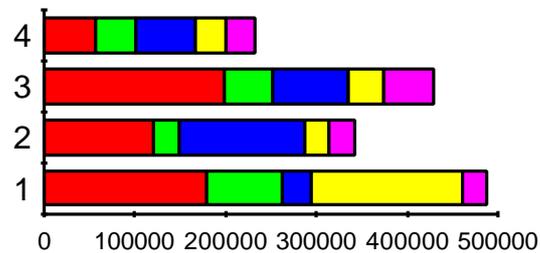


Figura 11. Ortogramma

Anche in questo caso è possibile sostituire ai rettangoli una linea, eventualmente punteggiata; si ottengono **diagrammi a barre o a punti** e l'intensità o frequenza delle varie classi viene letta con una proiezione sull'asse delle ascisse. Secondo alcuni esperti di percezione dei grafici, queste figure vengono lette con maggiore facilità rispetto ai rettangoli distanziati (l'occhio leggerebbe con maggiore facilità la proiezione verticale e di quella orizzontale) e quindi meglio rappresentano le informazioni contenute in distribuzioni di frequenza di dati qualitativi.

**Gli areogrammi** sono grafici in cui le frequenze o le quantità di una variabile qualitativa sono rappresentate da superfici di figure piane, come quadrati, rettangoli o, più frequentemente, cerchi oppure loro parti. La rappresentazione può essere fatta sia con più figure dello stesso tipo, aventi superfici proporzionali alle frequenze o quantità, sia con un'unica figura suddivisa in parti proporzionali.

Nel caso dei **diagrammi circolari o a torta**, si divide un cerchio in parti proporzionali alle classi di frequenza.

Gli areogrammi vengono usati soprattutto per rappresentare **frequenze percentuali**; hanno il vantaggio di fare capire con immediatezza che la somma di tutte le classi è uguale all'unità (1 o 100%); hanno l'inconveniente che evidenziano con estrema difficoltà le differenze che non sono molto marcate. Per differenze piccole, si dimostrano meno efficaci degli ortogrammi.

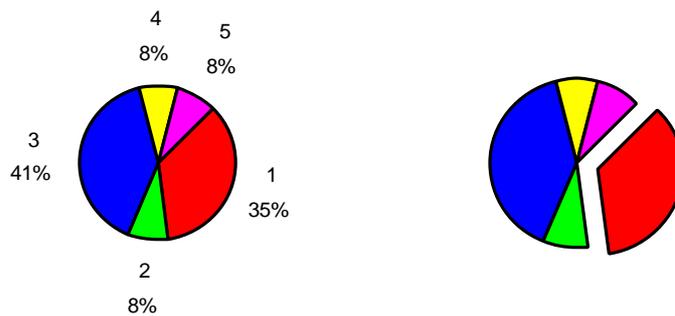


Figura 12. Diagrammi circolari

**I diagrammi circolari sono utilizzati per distribuzioni di variabili nominali, al fine di evitare di stabilire anche involontariamente un ordine, che non esiste tra variabili qualitative.** Mettono in evidenza come sono distribuite le singole parti, rispetto all'intero: il cerchio rappresenta l'intero fenomeno ed i componenti sono rappresentati da settori che sono distinti da tratteggi, colori o gradazioni di colore differenti.

Gli angoli (**a**, nella formula successiva) devono essere proporzionali alle percentuali (**Y in %**) che vogliono rappresentare,

in accordo con la relazione

$$a : 360 = Y \text{ in } \% : 100$$

Con i **diagrammi a figure**, detti anche **diagrammi simbolici** o **pittogrammi**, la frequenza di ogni carattere qualitativo viene rappresentata da una figura, sovente stilizzata, oppure da simboli che ricordano facilmente l'oggetto. E' una specie di istogramma costruito con figure, dove **l'altezza della figura deve essere proporzionale alla frequenza, quando le basi sono uguali.**

Questi diagrammi a figure **hanno tuttavia il grave inconveniente di prestarsi a trarre in inganno con facilità il lettore inesperto di statistica, quando sono stati costruiti con malizia; è pure facile che un ricercatore non sufficientemente esperto li costruisca in modo sbagliato, generando non volutamente il sospetto che egli abbia voluto fornire una impressione di differenza tra i gruppi a confronto non supportata dai dati.**

Per esempio, una popolazione con un numero triplo di persone rispetto ad un'altra spesso è rappresentata da una figura umana proporzionata, di altezza tripla rispetto all'altra.

L'occhio coglie complessivamente non l'altezza di ogni figura ma la superficie che essa occupa, che è il quadrato del valore che si intende rappresentare: se ne ricava l'impressione distorta di un rapporto di 9 a 1 e non di 3 a 1, come dicono in realtà i dati.

E' possibile ovviare all'inconveniente, costruendo non una figura improbabile di altezza variabile e con base uguale (poiché risulterebbe una figura alterata ed una rappresentazione forse incomprensibile), ma ricorrendo all'artificio di **figure identiche, ripetute tante volte quante sono le proporzioni**. Per esempio, se l'unità di misura convenuta è 20 individui, 50 persone possono essere rappresentate in modo corretto da due figure umane e mezza e 105 persone da 5 figure intere più un quarto di figura.

In questo modo si rispetta la regola per costruire gli istogrammi in modo corretto: l'altezza delle due figure è proporzionale al numero di dati dei due gruppi

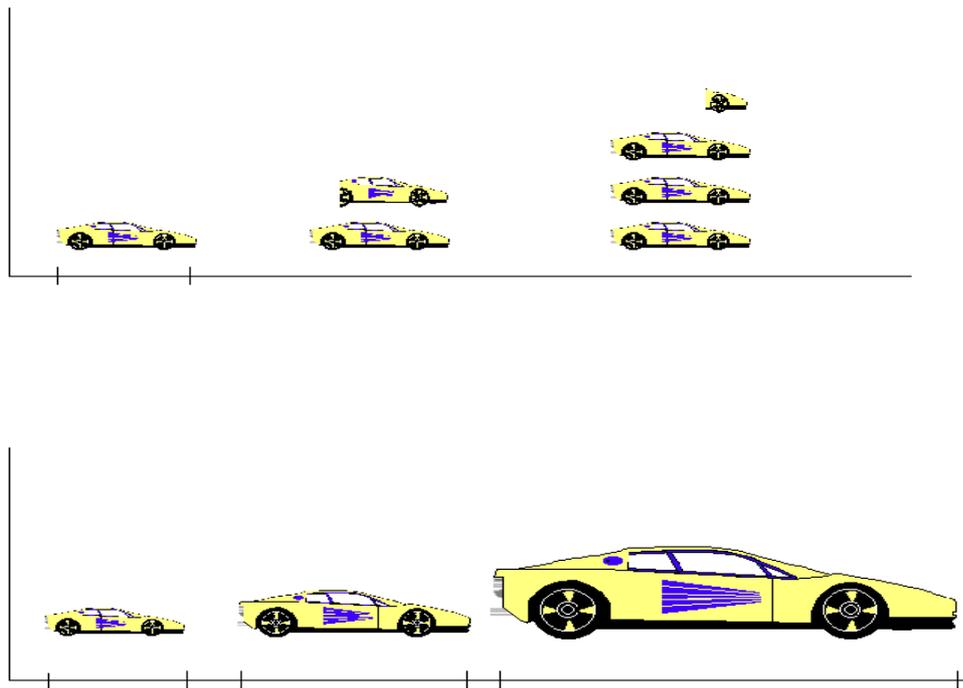


Figura 13. Pittogramma della produzione mensile di auto di 3 case automobilistiche: la prima ha prodotto 100 mila auto, la seconda 180 mila e la terza 320 mila. La parte superiore della figura fornisce una rappresentazione corretta. La parte inferiore, fondata sulla proporzione della lunghezza, fornisce una rappresentazione errata: è la superficie coperta dalla figura che deve essere proporzionale, non la lunghezza

A causa degli inconvenienti, i diagrammi simbolici o a figure sono usati molto raramente nelle pubblicazioni specializzate e mai in quelle scientifiche. Sono riservati a pubblicazioni divulgative, quando è più importante l'impressione della precisione, cioè occorre evidenziare l'importanza del fenomeno a persone che non conoscono esattamente il problema. Gli specialisti preferiscono i dati, poiché da essi sanno valutare il fenomeno.

Molte discipline ricorrono a rappresentazioni grafiche specifiche, che possono essere utili all'ambientalista. Per rappresentare il numero di soggetti presenti in vari località, in geografia si ricorre al cartogramma.

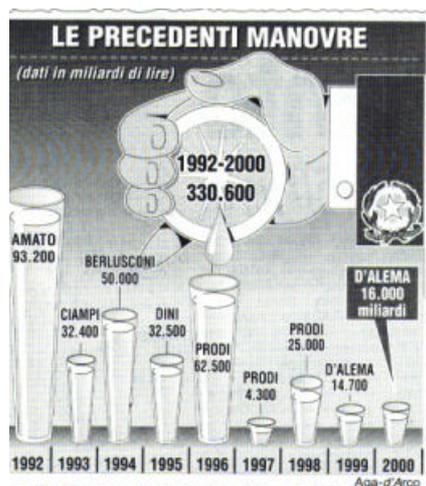


Figura 13b. Esempio tratto da un quotidiano sulle manovre finanziarie in Italia, dal 1992 al 2000. Notare come i numeri danno un'informazione differente dai volumi rappresentati.

**Il cartogramma** evidenzia distribuzioni territoriali mediante carte geografiche, in cui nelle località interessate sono riportati cerchi proporzionali alle frequenze. E' il caso delle città segnate su carte geografiche con cerchi di dimensioni proporzionali al numero di abitanti. Questi cerchi spesso sono solo simboli, illustrati nelle didascalie: per cui un solo cerchio bianco spesso indica una quantità di base (es.: 1.000 persone), due cerchi concentrici indicano una quantità maggiore, anche in modo non proporzionale (es.: 20.000 persone), tre cerchi o un cerchio annerito una quantità ancora superiore (es.: 100.000 persone) e così via. E' importante ricordare che, quando costruiti in modo proporzionale, anche queste rappresentazioni grafiche sono essenzialmente areogrammi e quindi possono trarre in inganno se ad essere proporzionale alle frequenze è il raggio.

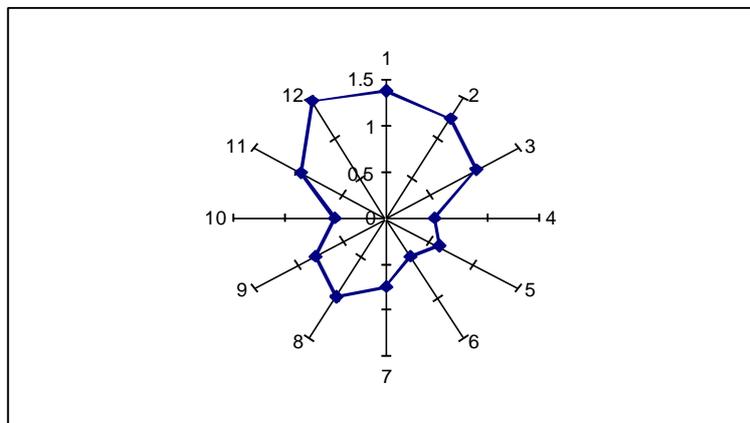
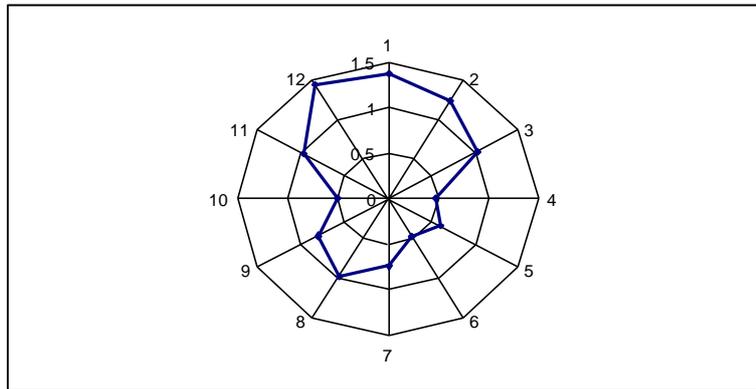
Un'altra rappresentazione grafica che ha un uso specifico per alcuni argomenti è il **diagramma polare o diagramma a coordinate polari**. Serve per rappresentare le **variabili cicliche** (mensili, settimanali, giornaliere), come la quantità di pioggia e la temperatura media mensile; oppure la quantità di inquinanti presenti nell'aria in un ciclo di 24 ore. A partire da un punto centrale, chiamato polo, si traccia una serie di cerchi concentrici, la cui distanza dal centro misura l'intensità del fenomeno. Per rappresentare la variabile ciclica, si divide l'angolo giro in tante parti quante sono le modalità (es.: 12 per i mesi, 24 per le ore). Si devono poi collocare punti nei vari cerchi concentrici, per individuare insieme la modalità (es.: il mese o l'ora) e l'intensità del fenomeno (es.: la quantità di pioggia, la temperatura, la misura d'inquinamento atmosferico o di un corso d'acqua). Il diagramma polare è ottenuto congiungendo i vari punti e l'intensità del fenomeno è rappresentata dalla distanza dal centro.

Le figure relative riportano due differenti impostazioni grafiche di costruire un diagramma polare sui valori medi mensili in Italia della radioattività beta totale nell'anno 1993.

Per la rappresentazione di **dati numerici**, è possibile ricorrere anche a **diagrammi cartesiani**. Essi saranno illustrati nel capitolo dedicato ai dati bivariati; sono utilizzati quando per ogni individuo sono rilevati contemporaneamente 2 variabili, come il peso e l'altezza. Ma possono essere usati anche per una sola variabile, collocando i punti di una distribuzione cumulata su un piano cartesiano: la perpendicolare sull'asse delle ascisse coincide con il valore della variabile e quella sull'asse delle ordinate fornisce le corrispondenti quantità o frequenze; i punti sono uniti da segmenti secondo l'ordine stabilito dal valore riportato in ascissa. E' di particolare utilità il **diagramma quantile**, che risulta graficamente simile al diagramma cumulato, soprattutto quando si dispone di poche unità e la variabile è di tipo continuo: vengono eliminate le anomale presenze di classi nulle entro gli estremi.

Figura 14. Valori medi mensili della radioattività beta totale nell'aria a livello del suolo in Italia nell'anno 1993 (mBq per metro cubo).

Mese	mBq
1 Gennaio	1.37
2 Febbraio	1.24
3 Marzo	1.03
4 Aprile	0.47
5 Maggio	0.60
6 Giugno	0.48
7 Luglio	0.74
8 Agosto	0.98
9 Settembre	0.81
10 Ottobre	0.50
11 Novembre	0.97
12 Dicembre	1.45



Per la scelta del metodo grafico con il quale presentare i dati, si deve prendere in considerazione il tipo di dati (qualitativi o quantitativi), la misura (discreta o continua), il dettaglio che si vuole ottenere nella forma della distribuzione. I metodi non aggiungono alcuna informazione che già non sia contenuta nei dati; ma garantiscono una rappresentazione più efficace, in particolare a persone non esperte dell'argomento trattato.

## 1.6. LE MISURE DI TENDENZA CENTRALE

**Le rappresentazioni grafiche forniscono una sintesi visiva** delle caratteristiche fondamentali delle distribuzioni di frequenza. Rispetto alle cifre, **le figure forniscono impressioni** che sono percepite con maggiore facilità; ma nel contempo hanno il limite di essere **meno precise** e **meno ricche di particolari**.

Per i **caratteri qualitativi**, la tabella e le rappresentazioni grafiche esauriscono quasi completamente gli aspetti descrittivi, quando sia possibile leggere con esattezza le frequenze delle varie classi.

Per i **caratteri quantitativi**, si pone il problema di **sintesi oggettive** che possano essere elaborate matematicamente e quindi che siano **numeriche**, al fine di **un'analisi obiettiva che deve condurre tutti i ricercatori, con gli stessi dati, alle medesime conclusioni**.

Una serie di dati numerici è compiutamente descritta da **3 proprietà principali**:

- 1) la **tendenza centrale o posizione**;
- 2) la **dispersione o variabilità**;
- 3) la **forma**.

Queste misure descrittive sintetiche, riassuntive dei dati tabellari, sono chiamate

- **statistiche, quando sono calcolate su un campione di dati,**
- **parametri, quando descrivono la popolazione od universo dei dati.**

I ricercatori in ecologia e nelle scienze ambientali molto raramente conoscono tutta la popolazione; di conseguenza, i metodi statistici di norma utilizzati sono riferiti quasi esclusivamente alla descrizione, all'analisi e al confronto di **campioni**.

1.6.1 Le **misure di tendenza centrale o posizione** servono per individuare il valore intorno al quale i dati sono raggruppati; la tendenza centrale è la misura più appropriata per sintetizzare l'insieme delle osservazioni, se una distribuzione di dati dovesse essere descritta con un solo valore; è la prima indicazione della dimensione del fenomeno.

Le misure proposte sono essenzialmente 3: **la media, la moda e la mediana**. Più raramente ed in discipline specifiche si utilizzano altre misure, come l'intervallo medio.

La scelta della misura di tendenza centrale di una serie di dati dipende dalle caratteristiche della distribuzione e dal tipo di scala.

La **media aritmetica semplice** è la misura di tendenza centrale più comunemente utilizzata. Quando si parla solo di **media**, si intende la media aritmetica semplice. E' definita come la somma del valore di tutte le osservazioni, diviso il numero di unità. Con simboli, è

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

e, con una notazione più generale, diventa

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

dove:

$\bar{x}$  = media del campione

$x_i$  = i-esima osservazione della variabile X

**n** = numero di osservazioni del campione

$\sum_{i=1}^n$  = sommatoria di tutti gli  $x_i$  del campione.

La media può essere vista come il baricentro della distribuzione campionaria, quando ogni singola osservazione è rappresentata da un peso convenzionale, identico per tutte, lungo l'asse che riporta i valori su una scala di intervalli o di rapporti.

Per dimostrare graficamente che **la media aritmetica corrisponde al punto di bilanciamento o di equilibrio dei dati**, si supponga di avere 5 misure: 10,9 11,5 12,3 12,8 15,4.

La loro media

$$\bar{X} = \frac{10,9 + 11,5 + 12,3 + 12,8 + 15,4}{5} = 12,58$$

è uguale a 12,58.

La rappresentazione grafica dei dati e della media, riportata nella figura seguente, mostra otticamente come la somma della distanza dalla media dei valori collocati prima sia uguale alla somma della distanza dei valori collocati dopo.

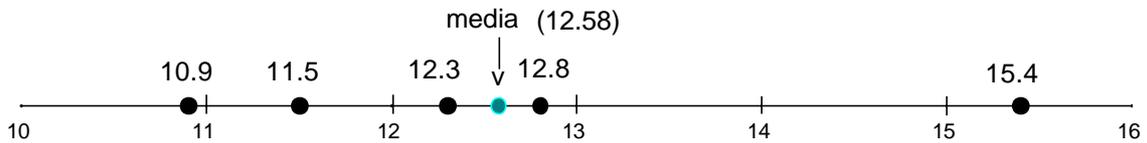


Figura 15. Rappresentazione grafica di 5 dati e della loro media aritmetica.

In una distribuzione di frequenza raggruppata in classi, come valore rappresentativo di ogni classe si prende il dato centrale, nell'assunzione che entro ognuna i dati siano distribuiti in modo uniforme.

La **media aritmetica di distribuzioni di frequenza** raggruppate in classi, detta **media aritmetica ponderata**, è calcolata più rapidamente con

$$\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}$$

dove:

- $\bar{x}$  = media della distribuzione in classi,
- $x_i$  = valore medio della  $i$ -esima classe di intervallo,
- $f_i$  = numero di osservazioni della classe  $i$ -esima classe,
- $n$  = numero di classi,
- $\sum$  = sommatoria per tutte le  $n$  classi.

ESEMPIO. Da un gruppo di 25 dati, raggruppati nella seguente distribuzione in classi

Classe	$x_i$	150-159	160-169	170-179	180-189	190-199
Frequenza	$f_i$	3	5	8	6	3

calcolare la media.

Risposta.

Con la formula della media ponderata

$$(\text{media})\bar{x} = \frac{(155 \cdot 3) + (165 \cdot 5) + (175 \cdot 8) + (185 \cdot 6) + (195 \cdot 3)}{3 + 5 + 8 + 6 + 3} = \frac{4385}{25} = 175,4$$

risulta uguale a 175,4.

Le applicazioni della media aritmetica semplice e di quella ponderata sono numerose e derivano da alcune loro proprietà:

- le **grandezze additive sono le più frequenti in natura**;
- la **media aritmetica effettua la correzione degli errori accidentali** d'osservazione, per cui essa è la stima più precisa di misure ripetute;
- la **media aritmetica è la più semplice** delle medie algebriche.

Quando le quantità od i fattori causali non sono additivi oppure i dati sono ottenuti da rapporti, si ricorre ad altri tipi di medie; in questi casi, trovano un uso relativamente frequente nelle scienze ambientali la **media geometrica**, la **media armonica** e la **media quadratica**.

La **media geometrica semplice** è utilizzata quando le variabili non sono rappresentate da valori lineari, ma ottenuti da prodotti o da rapporti di valori lineari. Serve per il confronto di superfici o volumi, di tassi di accrescimento o di sopravvivenza, per quei valori appunto che sono espressi da rapporti.

Per il calcolo della media geometrica, è condizione necessaria che le quantità siano tutte positive. Se alcune fossero negative, si deve ricorrere al valore assoluto.

**La media geometrica di n dati è uguale alla radice di ordine n (solo positiva) del prodotto degli n dati:** con simbologia matematica è

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

e può essere scritta anche come

$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^n x_i}$$

Una proprietà importante è che

**il logaritmo della media geometrica ( $\log \bar{x}$ )**

**è uguale alla media aritmetica dei logaritmi dei dati**  $\left( \frac{\sum_{i=1}^n \log x_i}{n} \right)$ :

$$\log \bar{x} = \frac{1}{n} \sum_{i=1}^n \log x_i$$

E' una proprietà che risulta utile quando si deve ricorrere alla trasformazione dei dati nei loro logaritmi, allo scopo di normalizzare la distribuzione ed applicare in modo corretto i test di inferenza. (Le trasformazioni dei dati sono discusse ampiamente nel capitolo finale dell'analisi della varianza).

La **media armonica** è la stima più corretta della tendenza centrale, per distribuzioni di dati in cui devono essere usati gli inversi. E' utilizzata quando i valori di X sono espressi come rapporti di un totale costante od in misure di tempi di reazione.

La media armonica è data da

$$m_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

La **media quadratica** è la radice quadrata della media aritmetica dei quadrati:

$$m_q = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$$

Sotto l'aspetto matematico può essere calcolata per valori positivi, nulli o negativi; ma essa ha senso come misura di tendenza centrale solamente se i valori sono positivi o nulli. E' un indice che trova applicazioni quando si analizzano superfici.

**1.6.2 La mediana** è il valore che *occupa la posizione centrale in un insieme ordinato di dati*. E' una **misura robusta**, in quanto **poco influenzata dalla presenza di dati anomali**. La sua utilizzazione è **indispensabile nel caso di scale ordinali o di ranghi**.

Le sue caratteristiche più importanti sono due:

- è calcolata sul numero di osservazioni; si ricorre al suo uso quando si vuole attenuare l'effetto di valori estremi o comunque prendere in considerazione solo l'informazione fornita dai ranghi;
- in una distribuzione o serie di dati, ogni valore estratto a caso ha la stessa probabilità di essere inferiore o superiore alla mediana.

**Come la media è la misura di tendenza centrale nella statistica parametrica, la mediana è la misura di posizione o tendenza centrale utilizzata in quasi tutti i test non parametrici.**

Per **calcolare la mediana di un gruppo di dati**, occorre

- 1 - disporre i valori in una fila ordinata in modo crescente oppure decrescente e contare il numero totale **n** di dati;
- 2 - se il numero (**n**) di dati è dispari, la mediana corrisponde al valore numerico del dato centrale, quello che occupa la posizione **(n+1)/2**;
- 3 - se il numero (**n**) di dati è pari, la mediana è stimata utilizzando i due valori centrali che occupano le posizioni **n/2** e **n/2+1**; con poche osservazioni, come mediana viene assunta la media aritmetica di queste due osservazioni intermedie; con molte osservazioni raggruppate in classi, si ricorre talvolta alle proporzioni.

ESEMPIO. Calcolare la mediana nella serie di 6 dati: 10,1 10,8 13,1 13,9 14,2 14,5 .

Risposta: Il numero di osservazioni è pari e i due valori centrali sono 13,1 e 13,9; la mediana è individuata dalla loro media aritmetica e quindi è uguale a 13,5.

**Per meglio comprendere le differenze tra media aritmetica e mediana**, con la stessa serie di 6 dati (10,1 10,8 13,1 13,9 14,2 14,5 ) in cui

- la media è 12,85 e
- la mediana 13,5

la rappresentazione grafica evidenzia come la media sia il baricentro della distribuzione e la mediana sia collocata tra i valori più addensati.

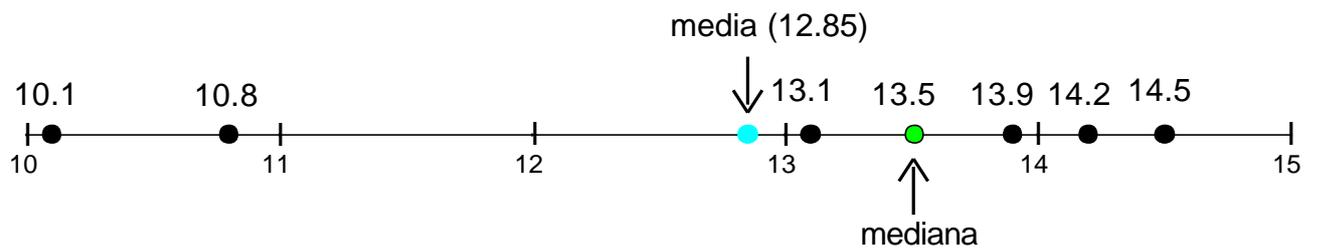


Figura 16. Rappresentazione grafica della media e della mediana di 6 dati.

Nella precedente figura 16, il grafico mostra come, nel caso di dati distribuiti in modo non simmetrico, la mediana rappresenti in modo più adeguato della media l'addensamento dei dati, il valore "normale o tipico" della serie. La media infatti è maggiormente influenzata dalla presenza dei due valori più distanti, che la allontanano dal gruppo dei valori più frequenti e la rendono diversa da essi. Se i due valori anomali fossero più vicini (o più lontani) rispetto agli altri 4, la media cambierebbe mentre la mediana rimarrebbe invariata.

1.6.3 La **moda** (detta più raramente anche **dato prevalente**) è il **valore più frequente di una distribuzione**. Essa non è influenzata dalla presenza di nessun valore estremo; tuttavia viene utilizzata solamente a scopi descrittivi, perché è **meno stabile e meno oggettiva delle altre misure di tendenza centrale**. Può infatti differire nella stessa serie di dati, quando si formano classi di distribuzione con ampiezza differente. Per individuare la moda entro una classe di frequenza, non conoscendo come i dati sono distribuiti, si ricorre all'ipotesi della **uniforme ripartizione**.

Oltre alle distribuzioni di frequenza che hanno una sola moda e che si chiamano **distribuzioni unimodali**, si trovano distribuzioni di frequenza che presentano due o più mode; sono denominate **distribuzioni bimodali o plurimodali**.

Le distribuzioni plurimodali possono essere il risultato della scarsità di osservazioni o dell'arrotondamento dei dati; di norma, sono dovute alla sovrapposizione di più distribuzioni con tendenza centrale differente. Per esempio, misurando le altezze di un gruppo di giovani in cui la parte maggiore sia formata da femmine e la minore da maschi si ottiene una distribuzione bimodale, con una moda principale ed una secondaria, come la seguente.

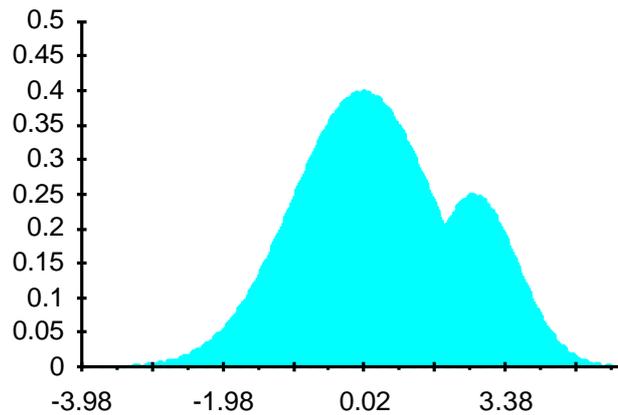


Figura 17. Distribuzione bimodale

**Quando la distribuzione dei dati evidenzia due o più mode**, il ricercatore deve quindi sospettare che i dati non siano omogenei, ma formati da altrettanti gruppi con differenti tendenze centrali. E' pertanto **errato fondare le analisi sulla media generale della distribuzione, poiché non è vera l'assunzione fondamentale che siano dati tratti dallo stesso universo o popolazione con una sola tendenza centrale.**

La media di una distribuzione bimodale, formata in quota pari da maschi e da femmine, sarebbe un valore "assurdo" che non descrive né i maschi né le femmine, ma un individuo inesistente, non essendo né maschio né femmina.

L'**intervallo medio** è semplicemente la media aritmetica tra il valore minimo e quello massimo. Ha il grande vantaggio di essere calcolato molto rapidamente, anche con un numero molto elevato di dati. Deve essere utilizzato con estrema cautela e solamente quando non esistono valori erratici o anomali: la presenza di un solo dato che si differenzia sensibilmente da tutti gli altri determina un valore dell'intervallo medio molto distorto, come misura della tendenza centrale.

In questi casi, può essere usata con maggiore correttezza la **media interquartile**, definita come la media fra il 1° e il 3° quartile, che risente in misura molto più ridotta della presenza di valori estremi. Nelle scienze che studiano l'ambiente, l'intervallo medio era utilizzato in alcune discipline come la meteorologia. Poteva essere utile nel caso di una serie di dati sulla temperatura, ove non esistono mai valori anomali; infatti supponendo che in una giornata la temperatura minima sia stata di 10 gradi e quella massima di 20 gradi, il calcolo della media è rapidissimo (15) ed il valore si avvicina

notevolmente alla media aritmetica, che richiederebbe un numero elevato di osservazioni e un disegno sperimentale accurato.

Per analogia, in meteorologia sovente questo metodo è stato utilizzato anche per il calcolo della precipitazione media mensile. E' un procedimento criticabile, addirittura errato: in questo caso si tratta di un fenomeno con elevatissima variabilità, con la presenza di valori che possono essere anomali e che influenzano fortemente sia l'intervallo medio che la media interquartile.

Oltre alla media, alla mediana e alla moda, insieme all'intervallo medio e alla media interquartile tra le misure di tendenza centrale può essere ricordata anche **la trimedia, proposta da Tuckey** e calcolata come

$$T = (Q_1 + 2Q_2 + Q_3)/4$$

dove  $Q_2$  è la mediana,  $Q_1$  e  $Q_3$  sono rispettivamente le mediane della prima metà e della seconda metà dei dati ordinati, detti anche primo e terzo interquartile.

E' un metodo che potrebbe essere utile quando si dispone di materiale molto variabile o con una distribuzione molto asimmetrica. Per esempio, le misure dell'inquinamento atmosferico presentano vari picchi anomali; la tendenza centrale potrebbe essere espressa dalla trimedia di Tuckey. Ma anche questa misura rientra tra le proposte che hanno avuto scarso seguito.

Le misure classiche, presenti in quasi tutte le discipline ed utilizzate senza sollevare obiezioni, sono media (aritmetica), mediana e moda.

## **1.7. MISURE DI DISPERSIONE O VARIABILITA'**

La **dispersione o variabilità** è la **seconda importante caratteristica** di una distribuzione di dati. Essa definisce la forma più o meno raccolta della distribuzione intorno al valore centrale e fornisce indicazioni sul tipo di test da applicare; nei capitoli successivi verrà dimostrato come per confrontare le medie di due o più campioni sia richiesta l'omogeneità della varianza.

1.7.1 La prima misura ad essere stata storicamente utilizzata per descrivere la dispersione o variabilità dei dati è **il campo o intervallo di variazione**, definito come **la differenza tra il valore massimo e quello minimo**.

$$\text{Intervallo di variazione} = \text{Valore massimo} - \text{valore minimo}$$

Ha il grande vantaggio di essere un metodo intuitivo e molto semplice, in particolare quando i dati sono ordinati.

Tra gli inconvenienti di questa misura sono da prendere in considerazione:

- l'**incapacità di sapere come i dati sono distribuiti** entro l'intervallo, in particolare di dedurre la presenza di valori anomali;
- la sua **dipendenza dal numero di osservazioni**. All'aumentare del numero dei dati, cresce anche la probabilità di trovare un valore minore del minimo precedente ed uno maggiore di quello massimo precedente.

L'intervallo di variazione è quindi una misura poco efficiente della dispersione dei dati: per un confronto omogeneo tra distribuzioni, sarebbe necessario avere campioni delle stesse dimensioni, una condizione operativa eccessivamente limitante per la ricerca e l'analisi dei dati.

1.7.2 **La differenza interquartile** (figura 18), la differenza tra il 3° ( $Q_3$ ) ed il 1° ( $Q_1$ ) quartile ha il vantaggio di eliminare i valori estremi, ovviamente collocati nelle code della distribuzione. Tuttavia le proprietà di questa semi-differenza, chiamata anche **scarto interquartile**, non sono sostanzialmente differenti da quelle del campo di variazione.

$$\frac{0}{\text{---}} \quad \frac{\frac{1}{4}}{Q_1} \quad \frac{\frac{1}{2}}{Q_2} \quad \frac{\frac{3}{4}}{Q_3} \quad \frac{1}{\text{---}}$$

Figura 18. Differenza interquartile = 3°quartile ( $Q_3$ ) - 1°quartile ( $Q_1$ )

Come **misure di posizione non-centrale**, ma con finalità esclusivamente descrittive, sono spesso usati i **quantili**, chiamati anche **frattili**, in quanto ogni sottogruppo contiene la stessa frazione di osservazioni. Quelli più comunemente usati sono i **decili**, che classificano i dati ordinati in decine, ed i **percentili**, che li suddividono in centesimi. Con i quantili, si possono individuare quali sono i valori che delimitano, nel margine inferiore o superiore della distribuzione, una percentuale o frazione stabilita di valori estremi. Per esempio, nello studio dell'inquinamento, come di qualunque altro fenomeno, può essere utile vedere quali sono le zone o i periodi che rientrano nell'1, nel 5 o nel 10 per cento dei valori massimi o minimi. *A valori così rari, facilmente corrispondono cause anomale, che di norma è interessante analizzare in modo più dettagliato. Nello studio di qualunque fenomeno biologico od ecologico, le misure particolarmente piccole o eccezionalmente grandi rispetto ai valori normali quasi sempre evidenziano cause specifiche, meritevoli di attenzione.*

Quando la forma della distribuzione è ignota o risulta fortemente asimmetrica, l'uso dei quantili fornisce indicazioni operative semplici e robuste per individuare i valori **più frequenti**, da ritenersi **“normali”** e quelli **meno frequenti od “anomali”**.

Gli **scarti dalla media** sono la misura più appropriata della variabilità di un insieme di dati. Ma poiché la loro somma è sempre nulla per definizione, in quanto la media è il baricentro della distribuzione, è necessaria una trasformazione che potrebbe essere attuata in due modi:

- a) **gli scarti assoluti dalla media;**
- b) **i quadrati degli scarti dalla media.**

1.7.3 Lo **scarto medio assoluto** ( $S_m$ ) **dalla media** ( $\bar{x}$ ) per dati semplici è dato da

$$S_m = \frac{\sum |x_i - \bar{x}|}{n}$$

e per raggruppamenti in classi è ottenuto con

$$S_m = \frac{\sum |x_i - \bar{x}| \times n_i}{n}$$

dove

$x_i$  = valore dell'**i**-esimo dato in una distribuzione semplice,

$\bar{x}$  = valore centrale della classe in una distribuzione di frequenza,

**n** = numero totale di dati,

**n<sub>i</sub>** = numero di dati della classe **i** in una distribuzione di frequenza.

Un indice analogo, usato nelle discipline sociali ed economiche per valutare la diversità tra due distribuzioni di frequenze relative, è

**l'indice semplice di dissomiglianza (D)**

$$D = \frac{\sum_{i=1}^k |f_{1i} - f_{2i}|}{2}$$

dove **1** e **2** sono i due gruppi e **k** sono le classi.

D è uguale a 0 quando le due distribuzioni di frequenza relativa sono identiche e uguale a 1 quando la prima distribuzione è tutta concentrata in una classe e l'altra distribuzione in una classe diversa.

1.7.4 In alcuni **test di statistica non parametrica**, come misura di dispersione è utilizzato **lo scarto medio assoluto dalla mediana**, che è la media degli scarti assoluti dei singoli dati dalla loro mediana; le formule sono uguali alle due precedenti, sostituendo la mediana alla media.

**E' proprietà specifica della mediana rendere minima la somma degli scarti assoluti.** Di conseguenza, lo scarto medio assoluto dalla mediana è sempre inferiore allo scarto medio assoluto dalla media; i due valori sono uguali solamente quando la distribuzione è simmetrica e quindi media e mediana coincidono.

1.7.5 La **Devianza o Somma dei Quadrati (SQ)** degli scarti dalla media (**SS = Sum of Squares**, in inglese) **è la base delle misure di dispersione dei dati, utilizzate in tutta la statistica parametrica.** Tutta la statistica parametrica è fondata sulla devianza e sulle misure da essa derivate.

$$(1) \quad \text{devianza (SQ)} = \sum (x_i - \bar{x})^2$$

L'equazione precedente è la **formula di definizione od euristica**. Spesso è poco pratica, in particolare quando la media è un valore frazionale, con vari decimali. Diviene allora conveniente ricorrere a un'altra formula, algebricamente equivalente, che permette di effettuare i calcoli manuali in tempi più brevi e con una sola approssimazione finale, chiamata **formula empirica od abbreviata**:

$$(2) \quad \text{devianza (SQ)} = \sum x^2 - \frac{(\sum x)^2}{n}$$

dove:

$\sum x^2$  = sommatoria dei valori dopo che ogni osservazione è stata elevata al quadrato,

$(\sum x)^2$  = sommatoria di tutti i dati, elevata al quadrato,

**n** = numero di osservazioni sulle quali è stata calcolata la somma.

**ESEMPIO.** Calcolare con la formula euristica (1) e con quella abbreviata (2) la devianza (SQ) dei 6 numeri seguenti: 5, 6, 7, 7, 8, 10.

Risposta.

1. Con la formula euristica, si deve calcolare dapprima la media:

$$\bar{x} = \frac{5+6+7+7+8+10}{6} = \frac{43}{6} = 7,1\bar{6}$$

ed in seguito la devianza (SQ),

intesa come Somma dei Quadrati degli scarti di ogni valore dalla media:

$$\begin{aligned} \text{devianza (SQ)} &= \sum (x_i - \bar{x})^2 = \\ &= (5 - 7,1\bar{6})^2 + (6 - 7,1\bar{6})^2 + (7 - 7,1\bar{6})^2 + (7 - 7,1\bar{6})^2 + (8 - 7,1\bar{6})^2 + (10 - 7,1\bar{6})^2 = \\ &= 4,665 + 1,3456 + 0,0256 + 0,0256 + 0,7056 + 8,0656 = 14,8356 \end{aligned}$$

2. Con la formula abbreviata, calcolare direttamente il valore della devianza (SQ), dopo aver fatto sia la somma dei dati precedentemente elevati al quadrato, sia il quadrato della somma dei dati, secondo l'annotazione algebrica seguente

$$\begin{aligned} \text{devianza (SQ)} &= \sum x^2 - \frac{(\sum x)^2}{n} = \\ &= (25 + 36 + 49 + 49 + 64 + 100) - \frac{43^2}{6} = 323 - \frac{1849}{6} = 323 - 308,1\bar{6} = 14,84 \end{aligned}$$

I due valori della devianza spesso non risultano identici, in particolare quando stimati con più cifre decimali, a causa dell'approssimazione con la quale è calcolata la media, se non risulta un valore esatto. In questi casi, è da ritenersi corretta la stima fornita dalla formula abbreviata, che non richiede approssimazioni nei calcoli intermedi.

E' utile ricordare che, **per distribuzioni di dati raggruppati in classi, la formula euristica diventa**

$$\text{devianza (SQ)} = \sum (\bar{x}_i - \bar{\bar{x}})^2 n_i$$

dove

$\bar{x}_i$  è il valore centrale di ogni classe e

$\bar{\bar{x}}$  è la media generale della distribuzione.

Il valore della devianza dipende da 2 caratteristiche della distribuzione: gli scarti di ogni valore dalla media ed il numero di dati. La prima è una misura della dispersione o variabilità dei dati ed è l'effetto che si intende stimare; la seconda è un fattore limitante per l'uso della devianza, in quanto un confronto tra 2 o più devianze richiederebbe campioni con lo stesso numero di dati. Pertanto, per una misura di dispersione dei dati che sia indipendente dal numero di osservazioni, si ricorre alla varianza.

1.7.6 **La varianza o Quadrato Medio** (QM, in italiano; MS da Mean Square, in inglese) è **una devianza media** o devianza rapportata al numero di osservazioni.

La **varianza di una popolazione** (1), il cui simbolo è  $\sigma^2$ , è ottenuta dividendo la devianza per **n**, il numero di osservazioni.

$$(1) \quad \sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

La **varianza di un campione** (2), il cui simbolo è  $s^2$ , è ottenuta dividendo la devianza per **n-1**, il numero di gradi di libertà.

$$(2) \quad s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Ovviamente, quando **n** è grande le differenze tra varianza della popolazione e varianza del campione sono minime; quando **n** è piccolo, le differenze sono sensibili.

E' importante ricordare che quando si parla di **inferenza**, cioè quando si utilizzano i **dati di un campione per conoscere le caratteristiche della popolazione**, si usa sempre la **varianza campionaria**.

Le giustificazioni logiche dell'uso di **dividere la devianza per n-1, detta anche correzione di Student**, sono lunghe e complesse: la più semplice si basa sul fatto che **n-1 è il numero di osservazioni indipendenti, chiamato gradi di libertà**, abbreviato abitualmente in **gdl** (degree freedom in inglese, abbreviato in **df**). Poiché la somma degli scarti dalla media è uguale a 0, **l'ultimo valore di una serie è conosciuto a priori, non è libero di assumere qualsiasi valore, quando siano già noti i precedenti n-1 valori.**

Come concetto generale introduttivo, si può dire che **il numero di gradi di libertà è uguale al numero di dati meno il numero di costanti che sono già state calcolate o di informazioni che siano già state estratte dai dati.** Nel caso specifico della varianza, la costante utilizzata per calcolare gli scarti è la media: quindi i gradi di libertà sono **n-1**.

Mentre la media è un valore lineare, la varianza è un valore al quadrato; per stime associate alla media o per confronti con essa, è necessario ricondurla a un valore lineare.

1.7.7 Lo **scarto quadratico medio o deviazione standard**, il cui simbolo è **s** nel caso della popolazione ed **s** nel caso di un campione, **è la radice quadrata della varianza**. Il termine **standard deviation** e il suo simbolo **s** (la lettera greca sigma minuscola) sono attribuiti al grande statistico inglese Karl Pearson (1867 – 1936) che l'avrebbe coniato nel 1893; in precedenza era chiamato **mean error**. In alcuni testi di statistica è abbreviato anche con **SD** e ed è chiamato **root mean square deviation** oppure **root mean square**,

E' una misura di distanza dalla media e quindi ha sempre un valore positivo. E' una misura della dispersione della variabile casuale attorno alla media.

Nel caso di **un campione**, a partire da una serie di dati la **deviazione standard**, il cui simbolo è **s**, può essere calcolata come:

$$\text{deviazione standard (s)} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 n_i}{n-1}}$$

dove

$x_i$  = valore del dato in una distribuzione semplice,

$\bar{x}$  = valore centrale della classe in una distribuzione di frequenza,

$n_i$  = numero di dati della classe  $i$  in una distribuzione di frequenza,

$n$  = numero totale di dati.

1.7.8 L'**errore standard** (*standard error*, in inglese) è indicato con **es** e misura la dispersione delle medie calcolate su **n** dati, come la deviazione standard serve per la dispersione dei dati.

L'errore standard **es** è

$$es = \frac{s}{\sqrt{n}}$$

ESERCIZIO. Calcolare media, devianza, varianza e deviazione st. e errore st. di : 9 6 7 9 8 8.

Risposta:

media = 7,833; devianza = 6,8333; varianza = 1,367; deviazione st. = 1,169; errore standard = 0,477

Per l'uso della varianza, che sarà fatto nei capitoli dedicati all'inferenza, è importante comprendere che **la varianza tra una serie di dati rappresenta una misura di mutua variabilità tra di essi.**

Essa può essere calcolata in tre modi:

- 1 - come **la metà della media aritmetica del quadrato di tutti gli  $n(n-1)/2$  scarti possibili tra coppie di osservazioni,**
- 2 - **mediante gli scarti tra i dati e la loro media,**
- 3 - **mediante la formula abbreviata.**

1 - Il primo metodo utilizza gli scarti tra tutte le possibile coppie di dati; è una procedura molto lunga, che serve per comprendere il **reale significato della varianza tra dati o tra medie**:

$$s^2 = \frac{1}{2} \cdot \frac{\sum_{i=1}^{J-1} \sum_{j=i+1}^J (x_i - x_j)^2 \cdot f_i \cdot f_j}{\frac{n(n-1)}{2}}$$

2 - Il secondo metodo rappresenta la **formula euristica**, quella che definisce la varianza, come confronto con il valore medio:

$$s^2 = \frac{\sum_{j=1}^J (x_j - \bar{x})^2 \cdot f_j}{n - 1}$$

3 - Il terzo metodo è una delle varie versioni della **formula abbreviata**, quella che serve per semplificare i calcoli manuali e ridurre i tempi per il calcolo

$$s^2 = \frac{\sum_{j=1}^n x_j^2 f_j - \frac{\left(\sum_{j=1}^n x_j f_j\right)^2}{n}}{n - 1}$$

ESEMPIO. Calcolare la varianza di 6 dati (5, 6, 7, 7, 8, 10) mediante le 3 formule proposte, per dimostrare empiricamente la loro equivalenza (ricordando che, in questo esempio,  $f_i = 1$ ).

Risposta:

1. Utilizzando **gli scarti assoluti (j - i) tra tutte le possibili coppie di dati**, riportati nella matrice triangolare sottostante:

j\i	5	6	7	7	8	10
5	0					
6	1	0				
7	2	1	0			
7	2	1	0	0		
8	3	2	1	1	0	
10	5	4	3	3	2	0

si ottiene

$$s^2 = \frac{1}{2} \cdot \frac{1^2 + 2^2 + 2^2 + 3^2 + 5^2 + 1^2 + 1^2 + 2^2 + 4^2 + 0^2 + 1^2 + 3^2 + 1^2 + 3^2 + 2^2}{15} = \frac{90}{30} = 3$$

2. Mediante **gli scarti dalla media** ( $\bar{X} = 7,1\bar{6}$ )

si ottiene

$$s^2 = \frac{(5 - 7,1\bar{6})^2 + (6 - 7,1\bar{6})^2 + \dots + (10 - 7,1\bar{6})^2}{6 - 1} = \frac{15}{5} = 3$$

3. Ricorrendo alla **formula ridotta**

si ottiene

$$s^2 = \frac{(5^2 + 6^2 + \dots + 10^2) - \frac{43^2}{6}}{6 - 1} = \frac{323 - 308,1\bar{6}}{5} = \frac{14,83}{5} = 2,9\bar{6}$$

**1.7.9 Il coefficiente di variazione** (*coefficient of variation* oppure *coefficient of variability*) è una misura relativa di dispersione, mentre le precedenti erano tutte misure assolute. E' quindi particolarmente utile ricorrere ad esso, quando si intende confrontare la variabilità di due o più gruppi con medie molto diverse oppure con dati espressi in scale diverse.

Consideriamo come esempio il confronto tra la variabilità di due specie animali con dimensioni medie sensibilmente diverse, come tra i cani e i cavalli. La varianza tra cavalli di razze diverse è superiore a quella esistente tra i cani, perché gli scarti assoluti dalla media della specie sono maggiori. Ma spesso

il problema consiste nel fare un confronto relativo tra variabilità e dimensioni medie delle due specie; allora il rapporto tra il cane di dimensioni maggiori e quello di dimensioni minori risulta superiore a quello esistente nei cavalli.

**Il Coefficiente di Variazione** (CV oppure semplicemente con **V** in molti testi recenti) **misura la dispersione percentuale in rapporto alla media** .

$$\text{Per una popolazione: } CV = \left( \frac{s}{m} \right) \cdot 100$$

dove

s = deviazione standard della popolazione

m = media della popolazione

$$\text{Per un campione: } CV = \left( \frac{s}{\bar{X}} \right) \cdot 100$$

dove

s = deviazione standard del campione

$\bar{X}$  = media del campione

Quando è calcolato su dati campionari, in particolare se il numero di osservazioni è limitato, il coefficiente di variazione CV deve essere corretto di una quantità  $1/4N$ , dove N è il numero di osservazioni del campione. Di conseguenza,

**il coefficiente di variazione corretto V'** diventa

$$CV' = CV \left( 1 + \frac{1}{4N} \right)$$

La figura successiva (tratta da pag. 16 di George W. Snedecor, William G. Cochran, 1974, *Statistical Methods*, Iowa University Press Ames, Iowa, U.S.A. sixth edition , seventh printing, pp. XIV + 593) è utile per spiegare i concetti già presentati:

- con la linea tratteggiata descrive l'**altezza media** di gruppi di ragazze da 1 a 18 anni d'età, che varia da circa 70 cm. ad oltre 170 cm.; la scala di riferimento è riportata sulla sinistra e varia da 60 a 200 centimetri;

- con la linea formata da tratti e punti alternati descrive la **deviazione standard dell'altezza** di ogni gruppo d'età; la scala è riportata sulla destra (parte superiore) e i valori variano da 0 a 7;
- con la linea continua descrive il **coefficiente di variazione**; la scala è riportata a destra nella parte inferiore più esterna ed i valori variano da 2 a 5.

E' importante osservare come la media e la sua deviazione standard aumentino in modo quasi correlato, mentre il CV resta costante intorno al 4%:

Il coefficiente di variazione è un numero puro, svincolato da ogni scala di misura e dalla tendenza centrale del fenomeno studiato. Secondo molti, appunto perché un rapporto, avrebbe significato solamente se calcolato per variabili misurate con una scala di rapporti.

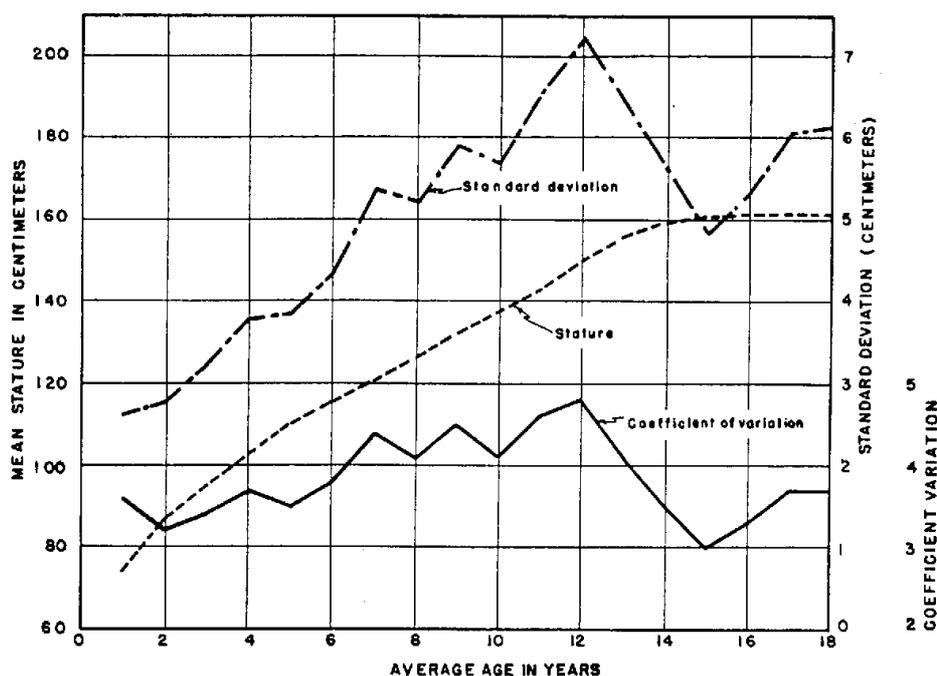


FIG. 2.17.1—Graph of 3 time series; stature, standard deviation, and coefficient of variation of girls from 1 to 18 years of age. See reference (1).

In natura, il coefficiente di variazione tende ad essere costante per ogni fenomeno, con valori che abitualmente oscillano tra il 5% e il 15%. Valori esterni a questo intervallo possono fare sorgere il

sospetto di essere in presenza di un errore di rilevazione o di calcolo; si tratta comunque di situazioni non usuali che occorrerebbe spiegare, individuandone la causa. Nell'esempio precedente, si tratta di individui della stessa età.

Se il materiale biologico in esame ha un CV troppo basso (2-3 %), si può sospettare l'esistenza di un fattore limitante che abbassa notevolmente od elimina la variabilità, come la presenza di omogeneità genetica congiunta ad una situazione ambientale uniforme; viceversa, un CV molto alto (50%) è indice della presenza di condizioni anomale o molto differenti contemporaneamente per più fattori.

Per l'uomo, il coefficiente di variazione dell'altezza è stato calcolato tra il 40% e il 45%, testimoniando l'esistenza nella specie di grandi differenze, dovute sia a cause genetiche che ambientali (alimentazione, condizioni sanitarie, ecc.).

**Quando per misurare lo stesso fenomeno si utilizzano scale differenti**, ad esempio l'altezza misurata in centimetri o in pollici, **la media e la deviazione standard cambiano, ma il CV resta uguale**.

Esso può essere calcolato anche per campioni; ma quando il numero di dati è limitato, la sua stima può indurre in errore.

In laboratorio per valutare la qualità dei reagenti, spesso si ricorre al C.V.: i reagenti che determinano il CV minore sono quelli di qualità superiore, poiché forniscono risposte meno variabili in rapporto ai valori medi.

#### **1.7.10 La varianza in dati raggruppati: correzione di Sheppard**

La varianza calcolata in una distribuzione di frequenza di misure continue è approssimata; la sua stima è fondata sull'ipotesi di distribuzione uniforme entro ogni classe e quindi si presume che il valore centrale di ogni classe corrisponda alla sua media. In realtà, la varianza calcolata sui dati reali e quella stimata a partire dal raggruppamento in classi non sono uguali.

**Quando la distribuzione dei dati è normale, entro ogni classe i valori più vicini alla media generale sono sempre più numerosi di quelli più distanti, collocati verso gli estremi.**

Come già evidenziato, per il calcolo della media, le approssimazioni nella parte sinistra del valore centrale compensano le approssimazioni fatte nella parte destra: la media calcolata direttamente dai dati e quella calcolata con il raggruppamento in classi hanno solo differenze casuali, di solito di entità ridotta.

Per il calcolo della varianza, le approssimazioni di segno opposto nelle due parti della media sono elevate al quadrato: di conseguenza, non si compensano, ma si sommano. In una popolazione con un

numero molto alto di dati, **la varianza calcolata dal raggruppamento in classi è sistematicamente maggiore di quella reale, quella calcolata direttamente dai dati originari. Le differenze crescono all'aumentare della misura dell'intervallo di ogni classe, poiché aumenta l'imprecisione.**

Pertanto si deve apportare una correzione, detta

**correzione di Sheppard**, proposta appunto da W. F. Sheppard nel 1898 sulla rivista *Proceeding London Mathematical Society* e riportata in vari testi, tra cui *Statistical Methods* di George W. Snedecor e William G. Cochran (1967, Iowa State University Press).

Consiste nel **sottrarre alla varianza calcolata un valore pari a**

$$\frac{h^2}{12}$$

per cui

$$\sigma^2_{\text{reale}} = \sigma^2_{\text{calcolata}} - \frac{h^2}{12}$$

dove

- **h** è l'ampiezza delle classi e
- **12** è una costante.

ESEMPIO. In una distribuzione di frequenza, in cui le classi hanno ampiezza costante con intervallo  $h = 10$ , è stata calcolata una varianza  $s^2 = 50$ ; la varianza corretta, quella che si sarebbe ottenuta utilizzando i singoli valori, secondo Sheppard dovrebbe essere

$$\sigma^2_{\text{reale}} = 50 - \frac{10^2}{12} = 50 - 8,3\bar{3} = 41,6\bar{6}$$

uguale a 41,66 come risulta dal calcolo mostrato.

**Questa relazione è ritenuta valida per le popolazioni.**

Con campioni formati da pochi dati, non è facile, spesso non è possibile, verificare se la distribuzione sperimentale utilizzata rispetti **le tre condizioni** fissate da **Sheppard** per applicare la correzione:

- **essere continua;**
- **avere un intervallo di ampiezza finito;**

- **tendere a zero in modo graduale nelle due code della distribuzione.**

Quando si dispone solo di piccoli campioni, la correzione potrebbe essere non adeguata alla forma della distribuzione e determinare un errore maggiore.

Di conseguenza, per piccoli campioni come quelli usati nella ricerca ambientale, **la quasi totalità dei ricercatori preferisce non applicare questa correzione, ma usare direttamente la varianza calcolata dalla distribuzione di frequenza**, in qualunque modo sia stato fatto il raggruppamento in classi.

### **1.8. INDICI DI FORMA: SIMMETRIA E CURTOSI**

Gli **indici di forma di una distribuzione riguardano 2 caratteristiche: la simmetria e la curtosi.**

A differenza di quanto avvenuto nello studio della variabilità, nell'analisi della forma di una distribuzione statistica le misure elaborate sono rimaste rudimentali e le stesse definizioni sono sovente equivocate. Inoltre **l'uso degli indici di forma** non rientra nei test d'inferenza, ma **è limitato alla semplice descrizione della forma della distribuzione.**

**Nelle distribuzioni unimodali si ha simmetria quando media, moda e mediana coincidono;** se la distribuzione è bimodale, possono essere coincidenti solamente la media aritmetica e la mediana.

Di norma, le distribuzioni dei dati sono unimodali; pertanto, l'analisi della simmetria è accentrata su di esse.

In una distribuzione,

- l'**asimmetria** è detta **destra** (più correttamente, **a destra**) quando i valori che si allontanano maggiormente dalla media sono quelli più elevati, collocate a destra dei valori centrali (figura 19); nell'asimmetria destra, la **successione** delle 3 misure di tendenza centrale da sinistra a destra è: **moda, mediana, media**;

- l'**asimmetria** è detta **sinistra** (o **a sinistra**) quando i valori estremi, quelli più distanti dalla media, sono quelli minori (figura 20). nell'asimmetria sinistra, la **successione** delle 3 misure di tendenza centrale da sinistra a destra è invertita rispetto all'ordine precedente: **media, mediana, moda.**

Quando media, mediana e moda non coincidono, la distribuzione è asimmetrica; ma quando queste tre misure coincidono non sempre la distribuzione è simmetrica. **Per avere una distribuzione simmetrica, la perfetta coincidenza delle tre misure di tendenza centrale è condizione solo necessaria, non sufficiente.**

Infatti, supponendo di analizzare una distribuzione come

4 16 20 20 20 30 30

troviamo che

- la media ( $140/7 = 20$ ),
- la mediana (su 7 valori è il 4° = 20) e
- la moda (il valore più frequente è 20)

sono coincidenti (20); ma, come si evidenzia dalla semplice lettura dei dati, la sua forma non è simmetrica poiché i dati non declinano in modo regolare ed identico dalla tendenza centrale verso i due estremi.

Un **altro metodo proposto per valutare la simmetria** utilizza la distanza delle classi di frequenza dalla mediana: **una distribuzione è simmetrica, se i valori che sono equidistanti dalla mediana hanno la stessa frequenza**. Ma è possibile dimostrare che si tratta di una condizione che si realizza sia in distribuzioni unimodali che plurimodali; è quindi una definizione che non caratterizza la distribuzione simmetrica in modo biunivoco, non è vera esclusivamente in una distribuzione normale.

I grafici di seguito riportati evidenziano la forma di una distribuzione simmetrica (Fig. 18),

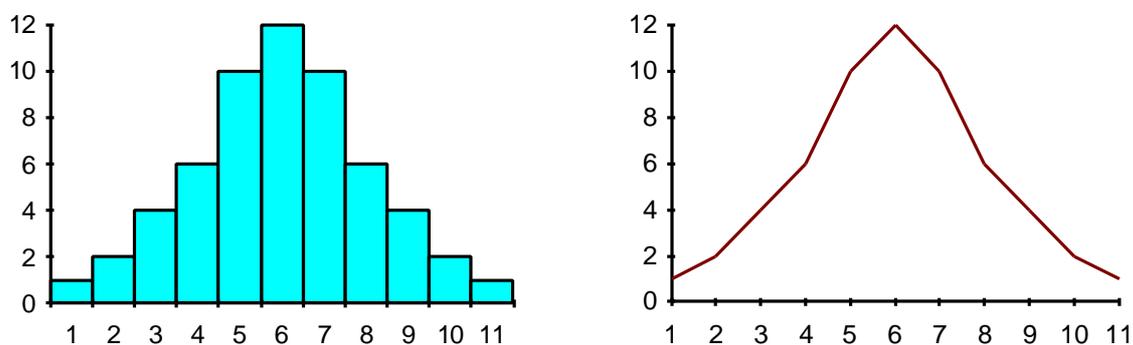


Figura 18. Distribuzioni simmetriche, con istogrammi e con poligoni

quella di una distribuzione destra o positiva (Fig. 19)

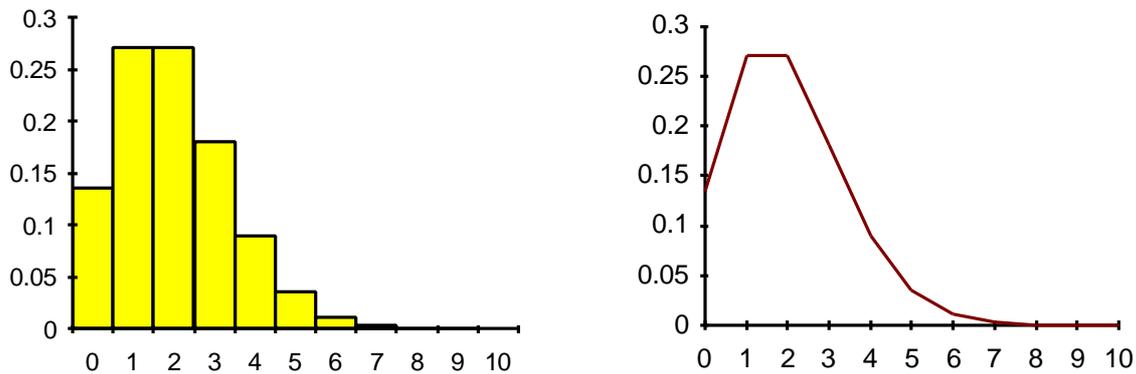


Figura 19. Distribuzioni con asimmetria a destra, con istogrammi e con poligoni

e quella di una distribuzione sinistra o negativa (Fig. 20).

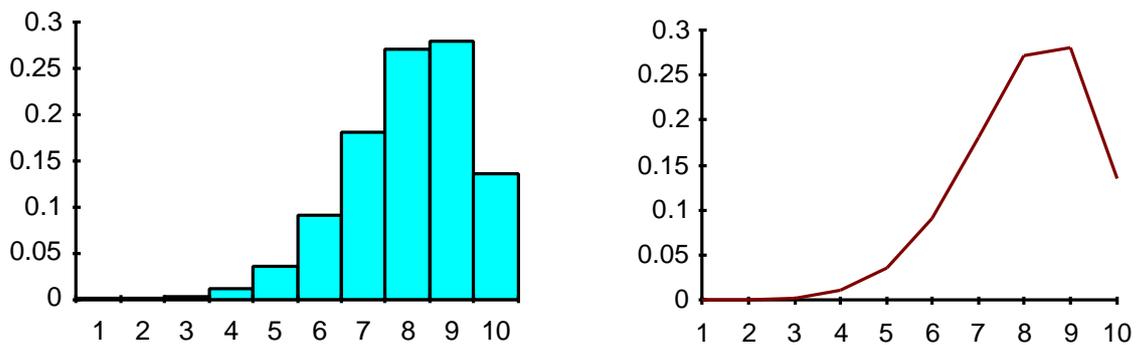


Figura 20. Distribuzioni con asimmetria a sinistra, con istogrammi e con poligoni.

Analizzando la distribuzione in classi dei dati di un campione, è possibile osservare un'**asimmetria causata da un numero ridotto di osservazioni** oppure da **modalità non adeguate nel raggruppamento in classi**, di solito eccessivo; in questi casi, si parla di **asimmetria falsa**, da distinguere dalla **asimmetria vera, che può esistere solo nella distribuzione reale di una popolazione**.

Una proprietà essenziale degli indici di asimmetria di una distribuzione è che essi dovrebbero essere uguali a zero quando, e solamente quando, la distribuzione è simmetrica. E' una proprietà che si realizza per gli indici abituali di variabilità o dispersione, come la devianza, la varianza e le misure derivate; esse sono nulle quando, e solamente quando, tutti i valori sono uguali,

e quindi non esiste variabilità; quando non sono nulle, esiste una variabilità, che cresce all'aumentare del valore dell'indice.

Gli indici di simmetria non godono della stessa proprietà: quando la distribuzione è simmetrica sono nulli; ma possono essere nulli anche per distribuzioni non simmetriche.

Per valutare l'asimmetria di una distribuzione, si possono usare

**- misure dell'asimmetria assoluta**

**misure di asimmetria relativa.**

Gli indici di **asimmetria assoluta** si esprimono con le distanze tra la media e la moda o la mediana.

Una misura assoluta, usata frequentemente, è **la differenza (d) tra la media e la moda**:

$$d = \text{media} - \text{moda}$$

La differenza è:

**d = 0**, se la curva è simmetrica;

**d > 0**, se la curva ha asimmetria positiva (o destra : media > mediana > moda);

**d < 0**, se la curva ha asimmetria negativa (o sinistra : media < mediana < moda).

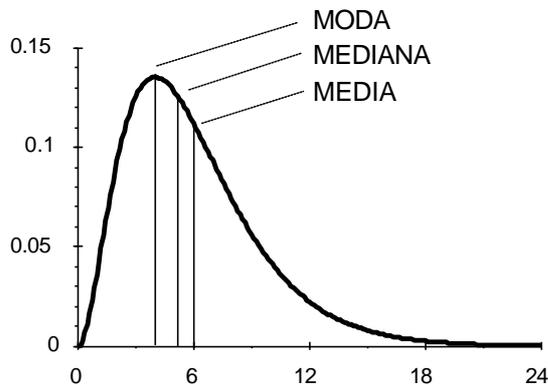


Figura 21. Asimmetria destra o positiva (d>0)

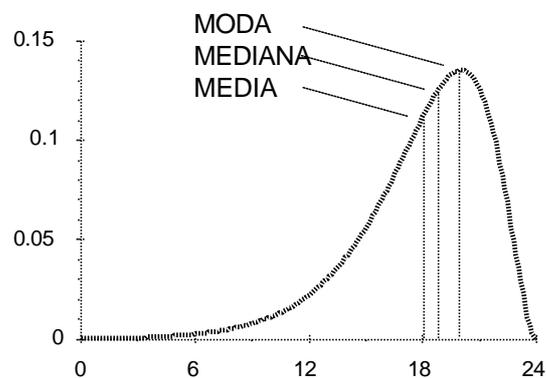


Figura 22. Asimmetria sinistra o negativa (d<0)

E' possibile valutare in modo molto semplice ed empirico **il grado d'asimmetria** di una distribuzione; essa è ritenuta **moderata** se

$$\text{Moda} = \text{Media} - 3(\text{Media} - \text{Mediana})$$

ed è ritenuta **forte** se è sensibilmente maggiore di tale valore.

Per ottenere una misura del grado di asimmetria, che possa essere confrontato con quello di qualsiasi altra distribuzione in quanto indipendente dalle dimensioni delle misure, occorre utilizzare **indici relativi**, quali

**skewness di Pearson;**

**$g_1$  di Fisher;**

**$b_1$  di Pearson.**

L'indice **skewness di Pearson (sk)** è un rapporto: la differenza (**d**) tra la media e la moda è divisa per la deviazione standard (**s**) o scarto quadratico medio. Nel caso di una distribuzione campionaria, dove la deviazione standard è indicata con **s**, è

$$sk = \frac{d}{s}$$

Come per il valore **d** precedente, **sk** può essere nullo, positivo o negativo secondo la forma della distribuzione.

Essendo **un rapporto tra misure statistiche della stessa distribuzione**, è divenuto una **misura adimensionale, indipendente dal valore assoluto degli scarti dalla media**; quindi può essere utilizzato per il confronto tra due o più distribuzioni.

Un altro indice di simmetria, proposto da A. L. **Bowley** nel 1920 ( vedi il testo *Elements of Statistics*, Charles Scribner's Sons, New York), chiamato appunto **Bowley coefficient** e riproposto in alcuni programmi informatici, utilizza i quartili (

$$\text{Skewness} = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

dove

- $Q_2$  = valore della mediana o del secondo Quartile
- $Q_1$  = valore del primo quartile,
- $Q_3$  = valore del terzo quartile.

Il valore ottenuto è uguale a zero se la distribuzione è perfettamente simmetrica, mentre è negativo o positivo in rapporto al tipo di asimmetria, in modo analogo alle formule precedenti.

Anche la **curtosi**, la concavità della distribuzione di dati (più ampiamente spiegata nel prosieguo del paragrafo), può essere misurata con i quantili o meglio gli **ottili**, come proposto da J. J. A. **Moors** nel 1988 (nell'articolo *A quantile alternative for kurtosis*. *Statistician* 37, pp.25-32)

$$\text{Kurtosis} = \frac{(O_7 - O_3) + (O_3 - O_1)}{O_6 - O_2}$$

dove

- $O_1$  è la metà di un quartile, cioè il valore che occupa il 12,5 percentile,
- $O_2, O_3, \dots$  sono multipli di esso fino a  $O_7$  corrispondente al 87,5 percentile.

Se la popolazione è

- esattamente normale, chiamata mesocurtica, il valore risulta uguale a 1233
- platicurtica, varia dal valore precedente a zero
- leptocurtica, varia dal valore precedente a infinito.

Gli indici relativi della forma di una distribuzione attualmente più diffusi sono derivati dai **momenti**.

**I momenti (m) di ordine k rispetto ad un punto c** sono calcolati con

$$m_k = \frac{\sum (x_i - c)^k}{n}$$

per una serie di dati

e con

$$m_k = \frac{\sum (x_i - c)^k \cdot f_i}{n}$$

per una distribuzione di frequenza suddivisa in classi.

Abitualmente, con **c** si indica l'origine (**c = 0**) oppure la media (**c = media**). Nel primo caso, si parla di **momento rispetto all'origine**; nel secondo, di **momento centrale**.

**Il momento di ordine 1 (k = 1) rispetto all'origine dei valori (c = 0) è la media;**

il momento centrale (c = m) di ordine 1 (k = 1) è uguale a 0 (è la somma degli scarti dalla media).

**Il momento centrale (c = m) di ordine 2 (k = 2) è la varianza.**

Nello stesso modo del momento centrale di secondo ordine (**m<sub>2</sub>**), si possono calcolare i momenti centrali di ordine terzo (**m<sub>3</sub>**), quarto (**m<sub>4</sub>**), quinto (**m<sub>5</sub>**),...ennesimo (**m<sub>n</sub>**).

**I momenti centrali di ordine dispari (m<sub>3</sub>, m<sub>5</sub>,...) sono utilizzati per indici di simmetria.**

Essi sono **nulli per distribuzioni simmetriche e differiscono da zero quando le distribuzioni non sono simmetriche**; quanto maggiore è l'asimmetria, tanto più il valore del momento centrale di ordine dispari è grande. Inoltre, in distribuzioni con **asimmetria destra ha un valore positivo** ed in quelle con **asimmetria sinistra ha un valore negativo**.

Per queste sue caratteristiche, il momento centrale di terzo ordine ( $m_3$ ) è adeguato per valutare la simmetria o asimmetria di una distribuzione; ma esiste il limite che il suo valore dipende dalla scala utilizzata.

Per ottenere una **misura relativa, adimensionale**, che permetta i confronti tra più distribuzioni, bisogna dividere  $m_3$  per il cubo dello scarto quadratico medio.

E' **l'indice  $g_1$  di Fisher**

$$\gamma_1 = \frac{m_3}{\sigma^3}$$

detto anche il **momento standardizzato di terzo ordine** e che mantiene le proprietà precedentemente descritte.

I momenti centrali di ordine dispari sono nulli, quando la distribuzione è simmetrica; sono positivi o negativi rispettivamente quando vi è asimmetria destra o sinistra.

Per **valutare il grado di asimmetria**, è convenzione che si abbia una distribuzione ad asimmetria

**forte**, quando  $|g_1| > 1$ ;

**moderata**, quando  $1/2 < |g_1| < 1$ ;

**trascurabile**, quando  $0 < |g_1| < 1/2$ .

L'indice di asimmetria  **$\beta_1$  di Pearson**, storicamente antecedente al  $\gamma_1$  di Fisher, è stato definito come

$$\beta_1 = \left( \frac{m_3}{\sigma^3} \right)^2$$

**Fisher ha solo semplificato l'indice di Pearson**, mediante la relazione semplice

$$\gamma_1 = \sqrt{\beta_1}$$

**ma nella pratica della statistica si è affermato il suo metodo.**

**E' utile ricordare quanto già ripetuto alcune volte: nel caso di distribuzioni simmetriche i 3 indici  $s_k$ ,  $g_1$ ,  $\beta_1$  danno un risultato uguale a 0; ma non sempre vale l'inverso, non sempre un indice di asimmetria uguale a 0 caratterizza la simmetria perfetta di una distribuzione di dati.**

Quando si descrive la forma delle curve unimodali simmetriche, con il termine **curtosi** (dal greco kurtos, che significa curvo o convesso) si intende il **grado di appiattimento, rispetto alla curva normale o gaussiana** (le cui caratteristiche saranno discusse in modo più approfondito nel capitolo dedicato alle distribuzioni teoriche).

Nella valutazione della curtosi, una **distribuzione unimodale simmetrica** è detta:

- **mesocurtica** , quando ha forma uguale alla distribuzione normale;
- **leptocurtica** (figura 23), quando ha un eccesso di frequenza delle classi centrali, una frequenza minore delle classi intermedie ed una presenza maggiore delle classi estreme; è quindi una distribuzione più alta al centro e agli estremi e più bassa ai fianchi; la caratteristica più evidente è l'eccesso di frequenza dei valori centrali;
- **platicurtica** (figura 24), quando rispetto alla normale presenta una frequenza minore delle classi centrali e di quelle estreme, con una frequenza maggiore di quelle intermedie; è quindi una distribuzione più bassa al centro e agli estremi mentre è più alta ai fianchi; la caratteristica più evidente è il numero più ridotto di valori centrali.

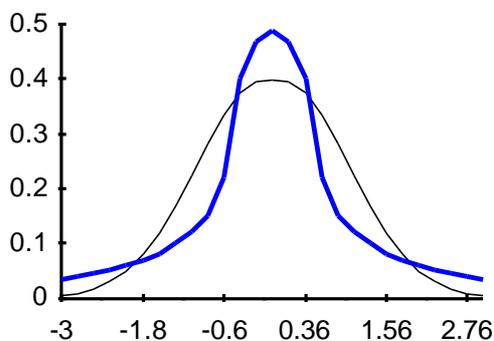


Figura 23. Distribuzione leptocurtica rispetto alla mesocurtica

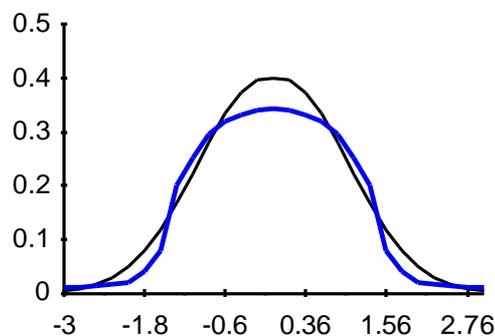


Figura 24. Distribuzione platicurtica rispetto alla mesocurtica

L'**indice di curtosi** è il risultato di un confronto, è un rapporto; quindi, **è una misura adimensionale**.

I due indici di curtosi più utilizzati sono analoghi a quelli di asimmetria:

- l'indice  $g_2$  di Fisher e

- l'indice  $b_2$  di Pearson.

L'indice  $g_2$  di Fisher è fondato sul rapporto

$$\gamma_2 = \frac{m_4}{\sigma^4}$$

**Se la distribuzione è perfettamente normale, il risultato del calcolo è uguale a 3; è maggiore di 3 se la distribuzione è leptocurtica, mentre è minore di 3 se la distribuzione è platicurtica.**

Per spostare la variazione attorno allo 0, l'indice di curtosi di Fisher è scritto come

$$\gamma_2 = \frac{m_4}{\sigma^4} - 3$$

Ovviamente, **il risultato diviene**

**0, se la distribuzione è normale o mesocurtica,  
positivo, se la distribuzione è leptocurtica o ipernormale,  
negativo, se la distribuzione è platicurtica o iponormale**

Mentre l'indice  $g_1$  può variare tra  $\pm \infty$ , l'indice  $g_2$  può variare tra  $-2$  e  $+\infty$ ; non è quindi possibile associare ad esso una gradazione in valore assoluto che valuti l'intensità della curtosi.

Come già precedentemente discusso, le condizioni che  $g_1$  e  $g_2 = 0$  sono necessarie ma non sufficienti, affinché la curva sia simmetrica e mesocurtica.

L'indice  $\beta_2$  di Pearson è il rapporto fra il momento centrale di quarto ordine e la deviazione standard, elevato alla quarta potenza:

$$\beta_2 = \frac{m_4}{\sigma^4}$$

Il suo legame con  $\gamma_2$  di Fisher è semplice, di tipo lineare:

$$\beta_2 = \gamma_2 + 3$$

Come l'indice  $g_2$  varia attorno a 0,  $\beta_2$  varia attorno a 3.

**Tutti gli indici presentati, dalle misure di tendenza centrale a quelle di dispersione e di forma, sono validi sia per variabili discrete che continue, con l'ovvia approssimazione data dal raggruppamento in classi.**

Quando simmetria e curtosi sono stimate non sulla popolazione ( $g_1$  e  $g_2$ ) ma su un campione (quindi indicate con i corrispondenti simboli latini  $g_1$  e  $g_2$ ),

**$g_1$  in valore assoluto tende a sottostimare  $g_1$  ( $|g_1| < |g_1|$ );** infatti è possibile dimostrare che, in un campione di dimensioni  $n$ , non supera il valore della radice di  $n$

$$|g_1| \leq \sqrt{n}$$

Problemi simili esistono per la stima di  $g_2$  in piccoli campioni con forte curtosi.

I limiti di  $g_2$  sono

$$\frac{-2(n-1)}{n-3} \leq g_2 \leq N$$

Con dati campionari, simmetria e curtosi sono ovviamente calcolati da distribuzioni di frequenza raggruppate in classi.

Definendo

$k$  = numero di classi di frequenza

$f_i$  = frequenza della classe  $i$ ,

$\bar{x}_i$  = valore centrale della classe  $i$

$\bar{\bar{x}}$  = media generale del campione

$s$  = deviazione standard del campione

e da essi avendo ricavato

$$\sum x^3 = \sum_{i=1}^k f_i (\bar{x}_i - \bar{\bar{x}})^3$$

e

$$\sum x^4 = \sum_{i=1}^k f_i (\bar{x}_i - \bar{\bar{x}})^4$$

si calcola  $g_1$  con

$$g_1 = \frac{n \cdot \sum x^3}{(n-1) \cdot (n-2) \cdot s^3}$$

e  $g_2$  con

$$g_2 = \frac{(n+1) \cdot n \cdot \sum x^4}{(n-1) \cdot (n-2) \cdot (n-3) \cdot s^4} - \frac{3 \cdot (n-1)^2}{(n-2) \cdot (n-3)}$$

I valori di  $g_1$  e  $g_2$  sono adimensionali: in altri termini, il risultato è identico, qualunque sia la misura utilizzata o la trasformazione applicata alla variabile  $X$ .

### 1.9. METODI PER CALCOLARE UN GENERICO QUANTILE DA UNA SERIE DI DATI

Nella statistica ambientale, è diffuso l'uso dei quantili, per due scopi:

- individuare la collocazione di un dato entro una serie di rilevazioni,
- stimare il valore di uno specifico percentile, come per la mediana o il terzo quartile.

Disponendo di uno o più dati, è utile conoscere la loro collocazione entro una distribuzione sperimentale precedente, come indicazione approssimata e descrittiva della sua posizione. In questo caso, la soluzione è semplice: il percentile è la posizione o rango occupata da quel valore, rapportata appunto a 100.

Più complesso è stimare il valore esatto di un determinato percentile, con differenze metodologiche tra una serie limitata di dati o una distribuzione di frequenza.

Il concetto di stima di un quartile appare semplice; ma i metodi di calcolo non sono così banali ed unanimi come possono apparire. Non esiste un metodo unico, con risultati universalmente accettati, in quanto le varie proposte rintracciabili nei testi mostrano tutte inconvenienti od illogicità di tipo differente. Quando il numero di osservazioni è alto, tutti i metodi forniscono risposte simili, spesso coincidenti; ma quando il numero di dati è limitato e sono presenti valori anomali, i risultati differiscono anche in modo sensibile.

A dimostrazione di questi concetti, si supponga di disporre di una serie di dati fortemente asimmetrica e con valori anomali, quale

1, 7, 4, 2, 50, 51,

in cui  $n = 6$ ,

per calcolare un generico quantile  $P_x$ .

Dopo aver ordinato gli  $n$  dati in modo crescente, ottenendo

1, 2, 4, 7, 50, 51,

**un primo metodo** richiede di

1 - Calcolare  $R$ , che è dato da

$$R = ((n - 1) \cdot P_x) + 1$$

Con  $n = 6$  dati e  $P_x$  supposto uguale al 3° quartile, (3/4 oppure 75/100, espresso nell'intervallo 0-1) e quindi  $P_x = 0,75$

$$R = ((6 - 1) \cdot 0,75) + 1 = 3,75 + 1 = 4,75$$

si ottiene  $R = 4,75$ .

Il valore di  $R$  (che nell'esempio è uguale a 4,75) indica che il quantile da stimare si trova tra il 4° e il 5° valore nella serie ordinata dei dati ed esattamente nella posizione 0,75 della distanza tra i valori di rango 4 e rango 5. Per l'individuazione di tale valore, il metodo qui presentato (valido anche per la mediana con  $P_x = 0,5$ ) chiede ulteriori passaggi, quali

2 – Prendere  $I$ , la parte intera di  $R$ ,

$$I = \text{Int} ( R )$$

per cui, nell'esempio,

$$I = \text{Int} (4,75) = 4$$

$I$  risulta uguale a 4.

3 – Calcolare  $D$  per differenza tra  $R$  e  $I$

$$D = R - I$$

che, sempre con i 6 dati dell'esempio

$$D = 4,75 - 4 = 0,75$$

risulta uguale a 0,75.

4 – Individuare nella serie ordinata dei dati  $X_{(I)}$  e  $X_{(I+1)}$

cioè (con  $I = 4$ ) i valori che occupano il rango 4° e 5°, per cui, con i dati dell'esempio,

$$X_{(4)} = 7 \quad \text{e} \quad X_{(5)} = 50$$

5 - La stima del quantile ( $Q$ ) è determinata dalla relazione

$$= (1 - D) \cdot X_{(I)} + D \cdot X_{(I+1)}$$

Con i dati dell'esempio, il 3° quartile ( $Q_{0,75}$ ) è

$$Q_{0,75} = (1 - 0,75) \times 7 + 0,75 \times 50 = 1,75 + 37,5 = 39,25$$

uguale a 39,25.

Dopo aver calcolato che il quantile ( $Q_{0,75}$ ) desiderato si trova in posizione 4,75 su 6 dati, **una variante del primo metodo** appena descritto è fondata sull'interpolazione lineare a 0,75 tra il valore che occupa il 4° rango ( $X_{(4)} = 7$ ) e quello che occupa il 5° rango ( $X_{(5)} = 50$ ). Dopo averne stimato la differenza  $d$

$$d = X_{(I+1)} - X_{(I)} = 50 - 7 = 43$$

si calcola la quota dovuta alla proporzione  $P$  (0,75) che eccede il rango  $I$  mediante la proporzione

$$P = 43 \times 0,75 = 32,25$$

e viene sommata al valore del rango  $I$

$$Q_{0,75} = 7 + 32,25 = 39,25$$

per ottenere un valore (39,25) uguale al precedente.

**Un secondo metodo** calcola il quantile  $P_x$  mediante la relazione

$$R_x = n \cdot P_x + 0,5$$

per cui il 75° percentile o terzo quartile con  $n = 6$  dati è

$$R_{0,75} = 6 \times 0,75 + 0,5 = 5,0$$

esattamente il 5° valore.

Con i 6 dati dell'esempio precedente  $Q_{0,75}$  risulta uguale a 50.

Per la quota eccedente l'intero  $I$ , quando esiste, si può usare **l'interpolazione come calcolata prima**, fra il valore  $X_{(I)}$  e  $X_{(I+1)}$ .

Altri autori, con **un terzo metodo**, definiscono il valore  $Q_x$  del quantile  $P_x$  nei termini della relazione

$$R_x = P_x \cdot (n + 1)$$

per cui il 75° percentile o terzo quartile con  $n = 6$  dati è

$$R_x = 0,75 \times (6 + 1) = 5,25$$

il valore che occupa la posizione 5,25.

Di conseguenza  $Q_{0,75}$  può essere stimato per interpolazione, tra il 5° e il 6° valore, risultando

$$Q_{0,75} = 50 + 0,25 (51 - 50) = 50,25$$

uguale a 50,25.

Anche **questo metodo presenta varianti**, fondate sulla logica di non voler stimare un valore che pretende di essere molto più preciso di quanto siano oggettivamente i dati:

- una prima è l'arrotondamento all'intero più vicino, per cui è il 5° valore e  $Q_{0,75}$  risulta uguale a 50,
- una seconda è l'interpolazione come media tra i due valori, calcolando quindi  $Q_{0,75}$  uguale a 50,5.

Anche il primo metodo, al quale ricorrono vari programmi informatici a grande diffusione, presenta inconvenienti logici, come evidenzia l'esempio seguente.

Le misure dell'inquinamento idrico spesso sono fornite come medie mensili; in Italia spesso manca il dato di agosto, coincidente con il mese di ferie. Calcolare il 9° decile della serie di 11 valori

$$12, 10, 8, 7, 14, 27, 29, 21, 14, 11, 9$$

Dopo aver ordinato per rango i valori

$$7, 8, 9, 10, 11, 12, 14, 14, 21, 27, 29$$

il 90° percentile

$$R_{0,9} = (11 - 1) \times 0,9 + 1 = 9 + 1 = 10$$

risulta il 10° valore, per cui  $Q_{0,9}$  è uguale a 27.

Se è corretto che la mediana o  $R_{0,5}$  sia uguale esattamente al sesto valore, è indubbiamente una stima approssimata che tutti i decili da 1 a 9, come indicano i calcoli, risultino esattamente i valori che occupano le posizioni dalla seconda alla decima.

E' utile ricordare quanto affermato da **Peter Armitage** e **Geoffry Berry** (in *Statistica Medica, metodi statistici per la ricerca in medicina*, 3a edizione, in italiano, McGraww-Hill Libri Italia srl, Milano 1996, a pag. 33):” **Si noti che non esiste un'unica procedura standard nel calcolo dei quartili (e dei quantili). Le diverse convenzioni conducono, comunque, a piccole e insignificanti differenze tra i risultati finali**”.

Con eccezione della sola mediana, non appare possibile definire quale sia il metodo migliore. Le differenze tra i diversi risultati, come nel caso di dati fortemente anomali, possono anche essere di quantità rilevanti, contrariamente a quanto affermato da Armitage; ma è un'incertezza insita nella variabilità delle osservazioni campionarie e nel numero limitato di osservazioni. Di conseguenza, è evidente la difficoltà di pervenire a conclusioni generali e condivise attraverso analisi fondate sui quantili.

## 1.10. RAPPRESENTAZIONE SEMI-GRAFICA DELLE DISTRIBUZIONI:

### BOX-AND-WHISKER, DIAGRAMMI STEM-AND-LEAF

I diagrammi Box-and-Whisker (scatola e baffi), chiamati anche semplicemente boxplot, sono un metodo grafico diffuso recentemente con i programmi informatici, per rappresentare visivamente alcuni indici sintetici delle distribuzioni statistiche e quindi le loro caratteristiche fondamentali.

Essi evidenziano 3 caratteristiche di una serie statistica:

- **il grado di dispersione o variabilità dei dati**, rispetto alla mediana e/o alla media;
- **la simmetria**;
- **la presenza di valori anomali.**

Sono utili per analizzare la distribuzione, senza ricorrere obbligatoriamente alle misure della statistica parametrica, come la varianza e gli indici fondati sui momenti.

Servono per evidenziare la presenza di dati che si discostano in modo rilevante dagli altri; essi potrebbero essere valori reali, ma non raramente sono generati da errori di misura o di scrittura.

E' quindi sempre importante utilizzare tecniche che li evidenziano, per una ulteriore verifica della correttezza della rilevazione e della trascrizione.

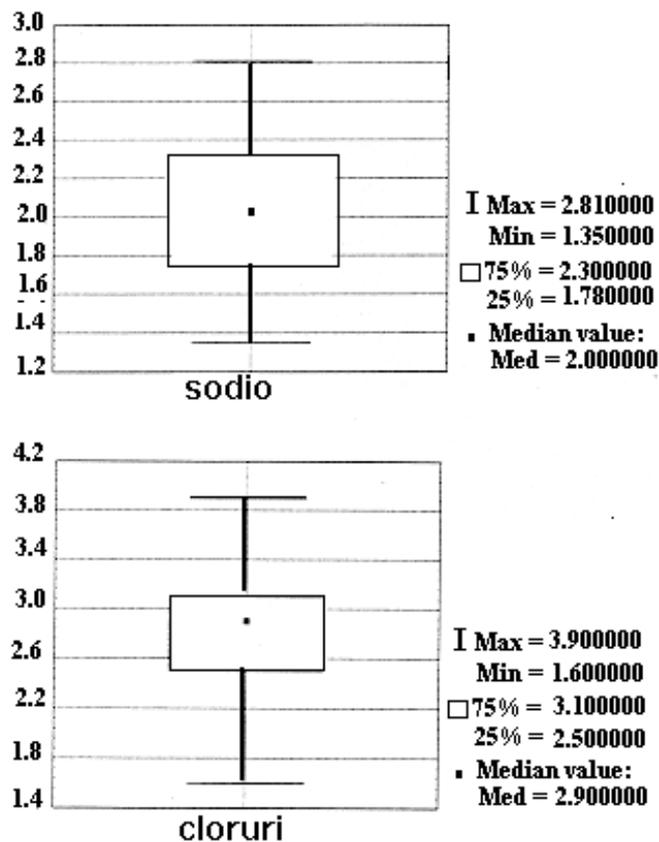


Figura 25. Box-and-Whisker con misure non parametriche.

Secondo il metodo originale proposto da **Tukey nel 1977** e da **McGill et al. nel 1978**, la costruzione di un diagramma **Box-and-Whisker** o **boxplot** ha origine da una linea orizzontale, che rappresenta la mediana; vicino ad essa è collocata una croce, che rappresenta la media; attorno ad esse è costruita una scatola, i cui margini inferiore e superiore indicano rispettivamente il valore del primo e del terzo quartile; infine, all'esterno della scatola si estendono due linee verticali, che terminano con un breve segmento orizzontale (i baffi), per un valore uguale a 1,5 volte la distanza interquartile. I valori che si discostano dalla mediana tra 1,5 e 3 volte la distanza interquartile possono essere considerati nella norma; quelli che si discostano oltre 3 volte dovrebbero essere molto rari e meritano una verifica ulteriore, per escludere con sicurezza banali errori di misura o

trascrizione. Quando la media è distante dalla mediana e/o il primo ed il terzo interquartile distano diversamente dalla mediana, la distribuzione è asimmetrica. La distanza del primo e del terzo interquartile dalla mediana è una misura della dispersione o variabilità dei dati.

Sul lato sinistro del grafico Box-and-Whisker è riportata una linea verticale, con i corrispondenti valori di X.

I diagrammi **Box-and-Whisker** hanno avuto una serie di adattamenti ed evoluzioni; quello descritto è solo un tipo, anche se tra quelli più frequentemente utilizzati. Tra le versioni più diffuse nei programmi informatici internazionali, sono da ricordare due tipi: quelli che impiegano la mediana come valore di tendenza centrale ed utilizzano la distribuzione dei quartili o dei percentili; quelli che riportano la media, insieme con l'errore standard e la deviazione standard.

I primi forniscono una descrizione non parametrica della forma della distribuzione, evidenziando dispersione e simmetria; i secondi rappresentano indici parametrici, presuppongono una distribuzione normale ed evidenziano sia la dispersione dei dati sia quella della media campionaria (questi argomenti saranno trattati in modo dettagliato quando si discuterà l'intervallo fiduciale o di confidenza).

Nei primi 2 Box-and-Whisker (figura 25) il valore di riferimento centrale è la mediana, la scatola delimita il primo ed il terzo quartile, mentre i baffi individuano il valore minimo e quello massimo. Le due distribuzioni non sono perfettamente simmetriche: la loro mediana non è equidistante dal 1° e dal 3° quartile, individuato dall'altezza della scatola, né dal valore minimo e massimo, rappresentato dai baffi.

La distribuzione dei dati del sodio ha una asimmetria positiva o destra, mentre la distribuzione dei valori dei cloruri ha una asimmetria sinistra o negativa. La rappresentazione in istogrammi e la misura del grado di asimmetria descrivono una lieve alterazione rispetto ad una distribuzione perfettamente normale.

Negli altri 2 boxplot (figura 26), il valore di riferimento è la media, la scatola riporta la distanza di 1 errore standard ed i baffi una distanza di 1 deviazione standard. Sono misure parametriche di dispersione rispettivamente della media e delle singole osservazioni, che saranno discusse dopo la presentazione della distribuzione normale e del test t di Student.

I baffi (Whisker) esterni riportano gli estremi che comprendono circa i 2/3 della distribuzione dei dati, mentre la scatola (box) fornisce gli estremi che comprendono i 2/3 delle medie che hanno identica variabilità e numerosità del campione raccolto.

Come sarà chiarito con l'uso della distribuzione normale, la frazione o percentuale di valori compresi nell'intervallo dipende da quante volte è riportato il valore della deviazione standard o dell'errore standard.

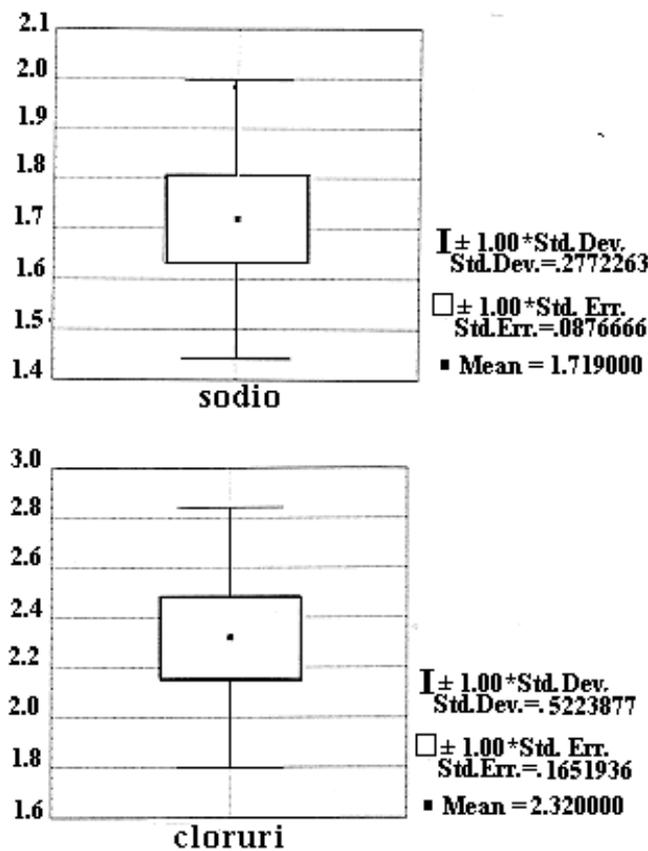


Figura 26. Box-and-Whisker con misure parametriche.

Si tratta di concetti che saranno sviluppati con l'uso della distribuzione normale, quando è nota la varianza della popolazione; si ricorrerà alla distribuzione t di Student, quando si utilizza la varianza del campione. I metodi servono per valutare sia la dispersione dei dati intorno alla media, sia la distribuzione delle medie di campioni con  $n$  osservazioni.

Il diagramma **stem-and-leaf** (*ramo e foglia*) è una tecnica semi-grafica, che può essere descritta come un incrocio tra un istogramma e una tabella di frequenza. E' riportata in molti programmi informatici dell'ultima generazione ed è utile per una prima descrizione di una distribuzione di dati; la sua conoscenza può essere di aiuto per decidere il livello di precisione con il quale raccogliere i dati. Si assuma di aver effettuato 30 rilevazioni della concentrazione di CO (mg/mc) lungo una strada con traffico, riportati in ordine crescente nella tabella seguente:

6,7	8,5	8,7	8,8	9,1	9,1	9,3	10,2	10,2	10,3
10,5	10,5	10,9	11,2	11,4	11,5	11,7	11,7	11,7	12,8
13,2	13,3	13,5	14,0	14,1	14,2	20,0	20,5	21,5	22,0

Per costruire un diagramma **stem-and leaf**,

- dapprima si devono individuare i valori **stem** e i valori **leaf**: i primi sono quelli che danno una misura approssimata del fenomeno, in questo caso la parte intera; i secondi quelli che rendono la stima più precisa, in questo caso i valori decimali;
- successivamente i valori **stem** sono ordinati modo crescente, lungo un'asse verticale, riportando anche le classi vuote
- mentre i valori **leaf** sono riportati in ordine crescente lungo l'asse orizzontale, costruito lateralmente ai valori **stem**.

Come nella figura successiva, si ottiene una specie di istogramma, con un numero di classi molto più elevato di quello utile ad una rappresentazione grafica corretta; in esso l'altezza di ogni classe è fornito dal numero di decimali riportati di fianco alla parte intera; la somma di **leave** (plurale di *leaf*) corrisponde al numero totale di dati.

Stem	Leaves
6	7
7	
8	5 7 8
9	1 1 3
10	2 2 3 5 5 7
11	2 4 5 7 7 7
12	8
13	2 3 5
14	0 1 2
15	
16	
17	
16	
18	
20	0 5
21	5
22	0

Da questa rappresentazione semigrafica, si ricava:

- l'**intervallo di variazione** del fenomeno: da 6 a 22;
- gli **stim modali**: i valori 10 e 11;
- la **mediana**: tra 10,5 e 10,7 trattandosi di 30 dati;
- i **quantili** utili ad una descrizione più dettagliata: l'80% dei valori è compreso tra 8,7 (il 10° percentile) e 20,5 (il 90° percentile);
- la **forma della distribuzione**: fortemente asimmetrica a destra e forse bimodale;

- la presenza di **outliers** (valori anomali rispetto alla distribuzione; se essi distano molto dagli altri stem, non è necessario riportare tutti i valori stem intermedi);
- la **precisione** con la quale i dati sono stati rilevati: gli ultimi 4 valori sembrano arrotondati alla mezza unità, mentre i primi sembrano stimati con una precisione al decimale.

La costruzione di un diagramma **stem-and-leaf** deve essere adattata alla dimensione del fenomeno. Ad esempio, sempre nella misura della qualità dell'aria, i valori guida o livelli di attenzione sono da 100-150 mcg/mc come valore medio di 24 ore per SO<sub>2</sub>. Le misure possono quindi essere approssimate all'unità; di conseguenza, gli **stem** possono essere dati dalle decine e le **leaf** dalle unità. Se i dati sono stati raccolti con troppa approssimazione, ad esempio i dati di CO rilevati con arrotondamento dell'unità, non è più possibile costruire un diagramma come quello presentato. **Non sempre a posteriori è possibile costruire questo diagramma.** Simmetricamente, se i dati sono raccolti con precisione eccessiva rispetto alla loro variabilità, ad esempio i valori di CO alla seconda cifra decimale, risulta necessario arrotondarli.

**1.11. ESERCIZI SULLE MISURE DI TENDENZA CENTRALE, DISPERSIONE,  
SIMMETRIA E CURTOSI**

ESEMPIO 1. In 36 laghi degli Appennini settentrionali è stato prelevato un campione d'acqua e sono state misurate le concentrazioni di Sodio e di Cloruri, espresse in mg/l, (vedi tabella seguente):

<b>Lago</b>	<b>Sodio</b>	<b>Cloruri</b>
1	1,78	1,60
2	1,63	1,80
3	1,85	2,90
4	2,10	2,90
5	1,35	2,90
6	1,40	2,90
7	1,82	2,00
8	1,35	2,00
9	2,06	2,00
10	1,85	2,20
11	1,51	2,30
12	2,00	2,30
13	2,02	2,80
14	1,90	2,80
15	1,60	2,80
16	2,18	2,50
17	1,82	2,50
18	1,90	2,50
19	1,75	2,60
20	2,11	2,60

21	2,30	2,60
22	1,95	2,70
23	2,60	2,90
24	2,44	2,90
25	2,18	3,00
26	2,51	3,10
27	2,37	3,10
28	2,54	3,30
29	2,06	3,30
30	2,77	3,40
31	2,31	3,40
32	2,81	3,60
33	2,33	3,70
34	1,45	3,80
35	1,78	3,80
36	2,09	3,90

Calcolare le misure della tendenza centrale, della variabilità e degli indici di forma;  
 - rappresentare graficamente i dati in istogrammi.

Risposta.

Le statistiche calcolate dai programmi informatici comprendono varie misure di tendenza centrale, di dispersione, di simmetria e di curtosi. Quelle di seguito riportate presuppongono una distribuzione normale e sono fondate sulla media e sui momenti della distribuzione. Esistono programmi che utilizzano la mediana come misura della tendenza centrale e ricorrono ai quantili per descrivere la dispersione e la simmetria, come nel caso dei primi 2 boxplot riportati in figura 25.

I programmi informatici forniscono una serie di valori, che descrivono compiutamente i dati campionari, come la tabella seguente (tra parentesi è riportato il termine inglese):

	Sodio	Cloruri
Numero di dati (Count, N. of data)	36	36
Somma (Sum)	72,87	101,4
Minimo (Minimum)	1,35	1,6
Massimo (Maximum)	2,81	3,9
Intervallo (Range)	1,46	2,3
Media (Mean)	2,024	2,817
Media geometrica (Geometric Mean)	1,987	2,756
Media armonica (Harmonic Mean)	1,949	2,692
Devianza (Sum of Squares)	5,29	11,76
Varianza (Variance, Mean Square)	0,151	0,336
Deviazione standard (Standard Deviation)	0,389	0,58
Errore standard (Standard Error)	0,065	0,097
Curtosi (Kurtosis)	-0,655	-0,53
Asimmetria (Skewness)	0,084	-0,015

Per valutare in modo più dettagliato e completo le caratteristiche delle 36 misure di sodio e cloruri presenti nei laghi campionati, è utile anche la loro rappresentazione in istogrammi. Quasi sempre sono forniti dai medesimi programmi informatici che calcolano anche gli indici già presentati.

Nei due istogrammi, i valori riportati sull'asse delle ascisse individuano la media della classe di riferimento. Nel primo grafico, sono riportati in modo alternato per evitare una eccessiva densità di numeri che renderebbe poco agevole la lettura. Sull'asse delle ordinate sono riportate le frequenze assolute.

Notare come i rapporti tra l'altezza e la lunghezza dell'istogramma rispondano ai criteri di eleganza grafica, già presentati.

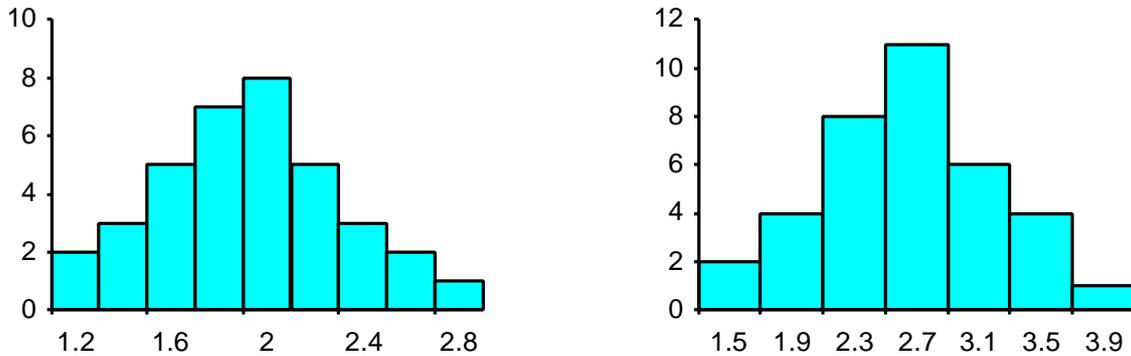


Figura 27

Istogramma delle concentrazioni del Sodio

Istogramma delle concentrazioni dei Cloruri

Le due serie di valori hanno una distribuzione normale molto vicino alla normale, con curtosi negativa ed una leggerissima asimmetria, negativa per il sodio e positiva per i cloruri. Per analisi e confronti, possono essere applicati i test parametrici.

ESEMPIO 2. In 4 laghi degli Appennini settentrionali, sono state stimate le densità dei principali taxa fitoplanctonici riportati in tabella:

	Clorophyceae	Cryptophyceae	Crysophyceae	Diatomophyceae	Dinophyceae
Lago 1	179857	83497	30891	166861	25600
Lago 2	120893	29000	136791	27500	28000
Lago 3	198043	54454	82770	38712	54734
Lago 4	57496	42980	66440	34356	31270

Rappresentare i dati nelle forme grafiche di uso più comune.

Risposta.

Sono distribuzioni di caratteri qualitativi. Le rappresentazioni grafiche adeguate sono i rettangoli distanziati e i diagrammi a torta: può essere scelta una delle varie versioni, secondo la caratteristica da evidenziare.

La figura 28 riporta le 5 specie presenti nei 4 laghi con le due differenti versioni di rettangoli distanziati; potrebbero essere utilmente applicati anche gli ortogrammi in una delle due forme equivalenti.

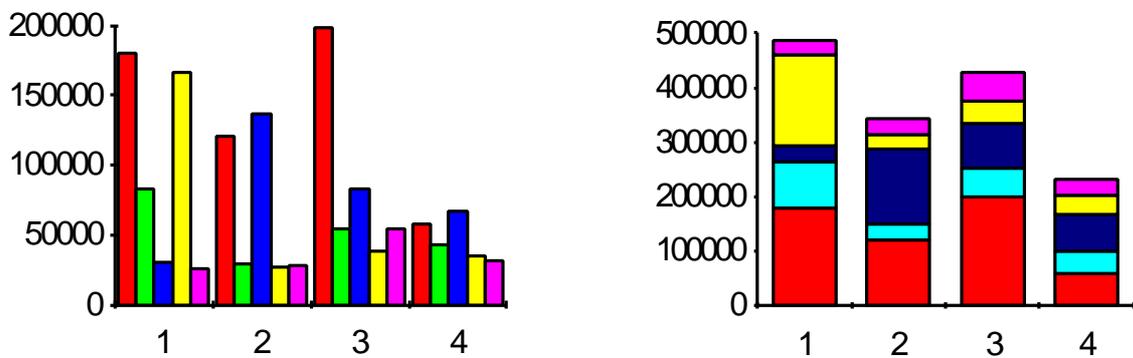


Figura 28. Istogrammi

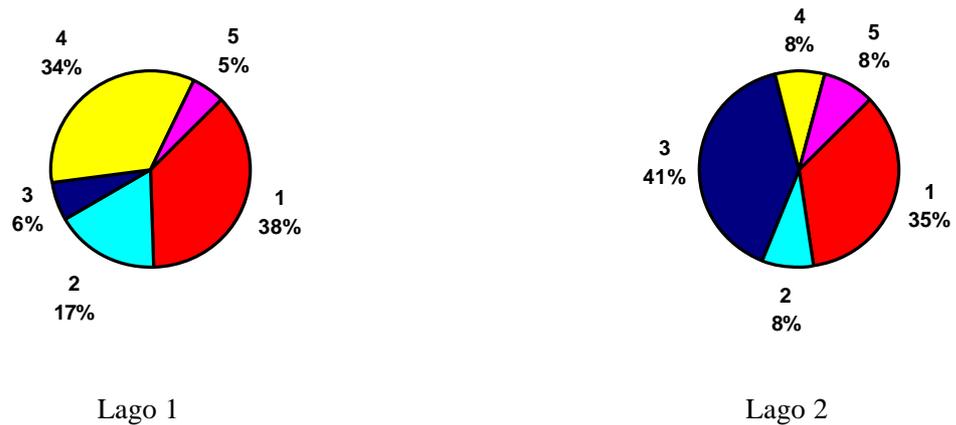


Figura 29. Diagrammi a torta delle specie presenti nei laghi 1 e 2

I diagrammi a torta dimostrano di essere di più difficile lettura. Senza le percentuali, diventa difficile un confronto tra il numero di individui delle diverse specie presenti (ad esempio nel lago 1 è impossibile distinguere se è più alta la frequenza della specie 4 o della specie 1).\

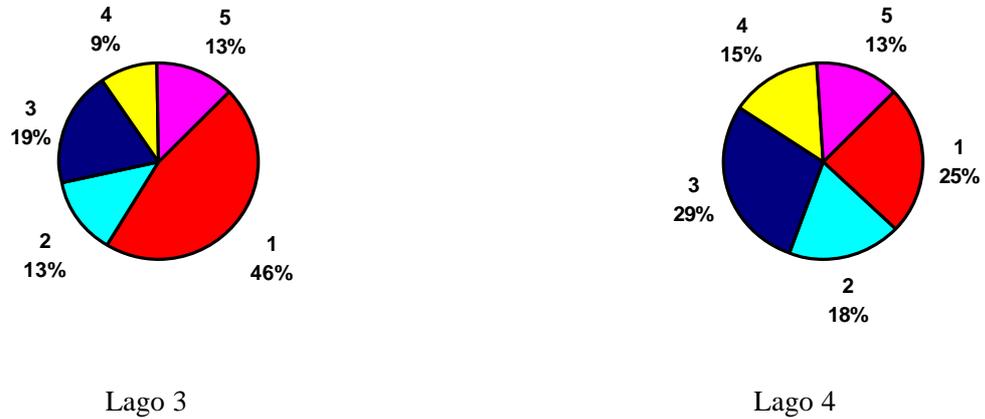


Figura 30. Diagrammi a torta delle specie presenti nei laghi 3 e 4.

Il confronto visivo tra i 4 diagrammi a torta mostra quanto sia difficile cogliere le differenze tra le 4 distribuzioni, meglio evidenziate dai grafici della figura 28.

ESEMPIO 3. Date due serie di dati relative ai campioni A e B

A: 5 7 2 4 3  
 B: 15 11 9 8 10 12

calcolare per ognuna di esse

- le misure della tendenza centrale,
- le misure della dispersione,
- gli indici di forma.

Sono stati proposti 2 esercizi con pochi dati, per evitare che i calcoli manuali richiedano troppo tempo.

Si chiede di

- calcolare la devianza secondo la formula euristica e quella abbreviata,

- calcolare la varianza sia con la formula euristica che con quella che considera tutti i possibili scarti tra coppie di dati.

Risposta.

Sono riportati i risultati di un programma informatico:

Statistiche dei 2 campioni	A	B
Numero di dati (Count, N. of data)	5	6
Somma (Sum)	21	65
Minimo (Minimum)	2	8
Massimo (Maximum)	7	15
Intervallo (Range)	5	7
Media (Mean)	4,2	10,833
Media geometrica (Geometric Mean)	3,845	10,60
Media armonica (Harmonic Mean)	3,506	10,398
Devianza (Sum of Squares)	14,7	30,83
Varianza (Variance, Mean Square)	3,7	6,167
Deviazione standard (Standard Deviation)	1,924	2,483
Errore standard (Standard Error)	0,86	1,014
Curtosi (Kurtosis)	-1,005	-0,605
Asimmetria (Skewness)	0,084	0,636

## CAPITOLO II

### DISTRIBUZIONI E LEGGI DI PROBABILITÀ'

#### 2.1. ELEMENTI DI CALCOLO COMBINATORIO SEMPLICE

La stima della probabilità di un evento è uno strumento fondamentale della statistica. Nelle sue forme più semplici, si fonda sul calcolo combinatorio; è evidente ed intuitiva la sua immediata applicazione ai giochi d'azzardo, ai quali effettivamente fu associata alla sua origine. Anche se il risultato di ogni singolo tentativo è imprevedibile, **con un numero elevato di ripetizioni si stabiliscono regolarità** che possono essere previste e calcolate. Dal punto di vista didattico, l'associazione del concetto di probabilità al calcolo combinatorio è un aspetto importante: serve per **collegare una scelta alla probabilità con la quale l'evento atteso può avvenire**, nel contesto di tutti gli eventi alternativi possibili. E' la base dell'inferenza statistica, della scelta scientifica in tutti i casi d'incertezza.

I concetti e i metodi del calcolo combinatorio possono essere spiegati in modo semplice, con una serie di esempi, tra loro collegati, per coglierne somiglianze e differenze.

In una corsa con 10 concorrenti, che abbiano le medesime possibilità di vittoria, è possibile porsi molti quesiti, tra i quali:

- a) quanti differenti ordini d'arrivo sono possibili?
- b) quale è la probabilità di indovinare i primi 3 al traguardo, secondo l'ordine?
- c) quale la probabilità di indovinare i primi 3, senza considerare il loro ordine?
- d) è conveniente scommettere 10 mila lire per guadagnarne 500 mila, se si indovinassero i primi 2 nell'ordine?
- e) è conveniente senza stabilire l'ordine?

Per calcolare le probabilità richieste, occorre prestare attenzione alle 4 caratteristiche fondamentali di questi **eventi**:

- (1) *si escludono a vicenda*,
- (2) sono *tutti ugualmente possibili*,
- (3) sono *casuali*,
- (4) sono *indipendenti*.

Il calcolo combinatorio di **raggruppamenti semplici** o **senza ripetizione**, così definiti in quanto ogni elemento compare una volta sola (in altri termini, lo stesso oggetto deve presentarsi in ciascun gruppo una volta sola), permette di calcolare la probabilità con cui può avvenire ogni evento possibile, che rispetti le 4 condizioni citate. Se le condizioni fossero differenti, si dovrebbe ricorrere ad altri metodi.

Per esempio, quando lo stesso oggetto può essere presente più volte in uno stesso gruppo (come l'estrazione ripetuta di una carta rimessa ogni volta nel mazzo), si devono utilizzare i raggruppamenti con ripetizione (o calcolo combinatorio con ripetizioni).

Nel **calcolo combinatorio semplice**, i raggruppamenti possibili possono essere distinti in **permutazioni, disposizioni, combinazioni**.

### 2.1.1 **Permutazioni semplici**.

**Dato un insieme di  $n$  oggetti differenti  $a_1, a_2, a_3, \dots, a_n$ , si chiamano permutazioni semplici tutti i sottoinsiemi che si possono formare, collocando gli  $n$  elementi in tutti gli ordini possibili.**

Alcuni esempi di permutazione delle 4 lettere a, b, c, d sono: abcd, abdc, acbd, adcb, cabd, cdab, dbac, cbda, ecc.

Il numero di permutazioni di  $n$  elementi è

$$P_n = n!$$

dove  **$n!$**  ( **$n$**  fattoriale) è il prodotto degli  **$n$**  elementi:  **$n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n$** .

Esempio 1: le permutazioni delle 4 lettere ( $P_4$ ) a, b, c, d, sono 4!

$$P_4 = 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$$

Esempio 2: le permutazioni di 3 elementi abc sono: abc, acb, bca, bac, cba, cab;  
cioè

$$P_3 = 3! = 6$$

Per i calcoli che saranno proposti durante il corso, è utile ricordare che, per definizione,

$$0! = 1$$

e che

$$1! = 1$$

**Tabella dei fattoriali di interi** (per facilitare i calcoli).

$n$	$n!$	$n$	$n!$
1	1	26	$4.03291 \times 10^{26}$
2	2	27	$1.08889 \times 10^{28}$
3	6	28	$3.04888 \times 10^{29}$
4	24	29	$8.84176 \times 10^{30}$
5	120	30	$2.65253 \times 10^{32}$
6	720	31	$8.22284 \times 10^{33}$
7	5040	32	$2.63131 \times 10^{35}$
8	40320	33	$8.68332 \times 10^{36}$
9	362880	34	$2.95233 \times 10^{38}$
10	$3.62880 \times 10^6$	35	$1.03331 \times 10^{40}$
11	$3.99168 \times 10^7$	36	$3.71993 \times 10^{41}$
12	$4.79002 \times 10^8$	37	$1.37638 \times 10^{43}$
13	$6.22702 \times 10^9$	38	$5.23023 \times 10^{44}$
14	$8.71783 \times 10^{10}$	39	$2.03979 \times 10^{46}$
15	$1.30767 \times 10^{12}$	40	$8.15915 \times 10^{47}$
16	$2.09228 \times 10^{13}$	41	$3.34525 \times 10^{49}$
17	$3.55687 \times 10^{14}$	42	$1.40501 \times 10^{51}$
18	$6.40327 \times 10^{15}$	43	$6.04153 \times 10^{52}$
19	$1.21645 \times 10^{17}$	44	$2.65827 \times 10^{54}$
20	$2.43290 \times 10^{18}$	45	$1.19622 \times 10^{56}$
21	$5.10909 \times 10^{19}$	46	$5.50262 \times 10^{57}$
22	$1.12400 \times 10^{21}$	47	$2.58623 \times 10^{59}$
23	$2.58520 \times 10^{22}$	48	$1.24139 \times 10^{61}$
24	$6.20448 \times 10^{23}$	49	$6.08282 \times 10^{62}$
25	$1.55112 \times 10^{25}$	50	$3.04141 \times 10^{64}$

### 2.1.2 Disposizioni semplici.

**Dato un insieme di  $n$  oggetti differenti  $a_1, a_2, a_3, \dots, a_n$  si chiamano disposizioni semplici i sottoinsiemi di  $p$  elementi che si diversificano almeno per un elemento o per il loro ordine.**

Le disposizioni delle 4 lettere a,b,c,d, raggruppate 3 a 3 sono: abc, acb, bac, dba, bda, abd, ecc.

Il numero di disposizioni semplici di  $n$  elementi  $p$  a  $p$  è

$$D_n^p = \frac{n!}{(n-p)!}$$

Esempio 1: le disposizioni di 4 elementi 3 a 3 sono:

$$D_4^3 = \frac{4!}{(4-3)!} = \frac{24}{1} = 24$$

Derivato dalla semplificazione di questa formula, un altro modo per calcolare le disposizioni semplici di  $n$  elementi  $p$  a  $p$  è

$$D_n^p = n(n-1)(n-2)\dots(n-p+1)$$

Le disposizioni di 4 elementi 3 a 3 possono quindi essere calcolate anche mediante

$$D_4^3 = 4(4-1)(4-2) = 4 \cdot 3 \cdot 2 = 24$$

Esempio 2: le disposizioni di 7 elementi 3 a 3 sono:

$$D_7^3 = 7(7-1)(7-2) = 7 \cdot 6 \cdot 5 = 210$$

### 2.1.3 Combinazioni semplici

**Dato un insieme di  $n$  oggetti differenti  $a_1, a_2, a_3, \dots, a_n$ , si chiamano combinazioni semplici di  $n$  elementi  $p$  a  $p$  i sottoinsiemi che si diversificano almeno per un elemento, ma non per il loro ordine.**

Le combinazioni semplici delle 4 lettere a,b,c,d, 3 a 3 sono: abc, abd, acd, bcd.

Il numero di combinazioni semplici di  $n$  elementi  $p$  a  $p$  è

$$C_n^p = \frac{n!}{(n-p)! p!}$$

Sotto l'aspetto del calcolo e dal punto di vista concettuale, il numero di combinazioni di  $n$  elementi  $p$  a  $p$  corrisponde al rapporto tra il numero di disposizioni di  $n$  elementi  $p$  a  $p$  ed il numero di permutazioni di  $p$  elementi.

Esempio 1: le combinazioni di 4 elementi 3 a 3 sono

$$C_4^3 = \frac{4!}{(4-3)! 3!} = 4$$

Per le applicazioni, è utile ricordare tre **casi particolari**:

$$a) \quad C_n^n = \frac{n!}{n! \cdot 0!} = 1$$

Il numero di combinazioni di  $n$  elementi presi  $n$  ad  $n$  è  $1$ : c'è un solo sottoinsieme formato da tutti gli elementi.

$$b) \quad C_n^1 = \frac{n!}{1!(n-1)!} = n$$

Il numero di combinazioni di  $n$  elementi presi  $1$  a  $1$  è uguale a  $n$ : il numero di sottoinsiemi con  $1$  solo elemento è  $n$ .

$$c) \quad C_n^0 = \frac{n!}{0!n!} = 1$$

Il numero di combinazioni di  $n$  elementi  $0$  a  $0$  è  $1$ : c'è un solo sottoinsieme vuoto.

Come è impostato per il calcolo, **il numero di combinazioni è solo apparentemente frazionario**: risulta sempre  $n$ , numero intero, che si indica con il simbolo  $\binom{p}{n}$  chiamato **coefficiente binomiale** e si legge  **$p$  su  $n$** .

#### 2.1.4 Risposte alle domande del paragrafo 2.1

Si è ora in grado di fornire le risposte ai cinque quesiti introdotti nel paragrafo 2.1

a) In una corsa con 10 concorrenti, i possibili ordini d'arrivo sono le permutazioni di 10 elementi. Il loro numero è

$$P_{10} = 10! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 = 3.628.800$$

b) In una corsa di 10 concorrenti, il numero dei possibili gruppi differenti formati dai primi 3 all'arrivo, tenendo conto anche del loro ordine, sono le disposizioni di 10 elementi 3 a 3, cioè

$$D_{10}^3 = \frac{10!}{(10-3)!} = 720$$

La probabilità di indovinare i primi 3 concorrenti secondo l'ordine d'arrivo è  $1/720 = 0,001389$ .

c) In una corsa di 10 concorrenti, i possibili gruppi dei primi 3 concorrenti, senza distinzioni interne di ordine, sono le combinazioni di 10 elementi 3 a 3, cioè

$$C_{10}^3 = \frac{10!}{(10-3)! 3!} = 120$$

La probabilità di indovinare i primi 3 concorrenti, senza stabilirne l'ordine, è  $1/120 = 0,008333$ ; è 6 (3!) volte più alta di quella in cui si chiede di indovinare anche il loro ordine.

d) Il numero di possibili gruppi formati dai primi 2 concorrenti, stabilendo chi sarà il primo e chi il secondo, in un gruppo di 10 è determinato dalle disposizioni di 10 elementi 2 a 2, cioè

$$D_{10}^2 = \frac{10!}{(10-2)!} = 90$$

La probabilità di indovinare chi saranno i primi 2 è uguale a  $1/90$ . È un rapporto più sfavorevole del rapporto di 1 a 50 fissato nella scommessa. Per chi scommette non è conveniente vincere 50 volte la posta, quando la probabilità di vincere è  $1/90$ .

e) Il numero di possibili gruppi formati dai primi 2 concorrenti, senza stabilire l'ordine, in un gruppo di 10 è dato dalle combinazioni di 10 elementi 2 a 2, cioè

$$C_{10}^2 = \frac{10!}{(10-2)! 2!} = 45$$

La probabilità di indovinare i primi 2 senza dover stabilire l'ordine uguale a  $1/45$ ; è più favorevole del rapporto di 1 a 50 fissato dalla scommessa. Per chi scommette è conveniente, perché l'eventuale guadagno è superiore al rischio. Una scommessa tra i due giocatori è in parità, solamente quando il prodotto tra la probabilità d'indovinare e il moltiplicatore della posta è uguale a 1.

## ESERCIZI

1. In un esperimento sulla fertilità del terreno, si vogliono studiare in modo sistematico gli equilibri binari tra i seguenti elementi: Ca, Mg, Na, N, P, K.

A. Quante coppie di elementi occorrerà prendere in considerazione?

$$\text{(Risposta : } C_6^2 = \frac{6!}{(6-2)!2!} = \frac{5 \cdot 6}{2} = 15 \text{)}$$

B. Se si intende valutare tutti gli equilibri ternari, quanti gruppi diversi formati da tre elementi occorrerà formare?

$$\text{(Risposta : } C_6^3 = \frac{6!}{(6-3)!3!} = 20 \text{)}$$

2. Nel timore che, durante una settimana con molto traffico, il tasso d'inquinamento dell'aria in una città fosse in costante aumento, è stata effettuata una serie di rilevazioni.

A. Quale è la probabilità che, senza un reale aumento dell'inquinamento e solo per caso, i valori osservati siano tutti in ordine crescente, se sono state fatte 4 rilevazioni?

$$\text{(Risposta : } P_4 = 4! = 24; \text{ la probabilità è } 1/24 = 0,04166 \text{ o } 4,166\%)$$

B. E se durante la settimana è stata fatta una rilevazione per giorno, quale è la probabilità che solo per caso siano tutti in ordine decrescente?

$$\text{(Risposta : } P_7 = 7! = 5.040; \text{ la probabilità è } 1/5.050 = 0,000198 \text{ o } 0,0198\%)$$

3. Per una serie di misure, da un gruppo di animali di dimensioni differenti sono stati estratti alcuni individui.

A. Se si estraggono casualmente 3 individui da un gruppo di 15, quale è la probabilità che essi siano i 3 con dimensioni maggiori?

$$\text{(Risposta : } C_{15}^3 = 15! / 3! 12! = 455; \text{ la probabilità è } 1/455 = 0,00219 \text{ o } 0,219\%)$$

B. Se si estraggono 4 animali da un gruppo di 20, quale è la probabilità che per caso possano essere i 4 con dimensioni minori?

$$\text{(Risposta: } C_{20}^4 = 20! / 4! 16! = 4.845; \text{ la probabilità è } 1/4.845 = 0,00021 \text{ o } 0,021\%)$$

## 2.2. DEFINIZIONI DI PROBABILITA': MATEMATICA, FREQUENTISTA, BAYESIANA

In un gioco fondato su **eventi casuali, mutuamente esclusivi, ugualmente possibili ed indipendenti**, il risultato di ogni singolo tentativo è imprevedibile. Ma su un elevato numero di tentativi, si stabiliscono regolarità e leggi. A lungo termine, l'esito è prevedibile. Le probabilità dei successi e quelle d'ogni evento alternativo possono essere calcolate con precisione, che è crescente all'aumentare del numero di osservazioni.

La natura del concetto di probabilità è chiarita dal **teorema di Bernoulli**, pubblicato postumo nel 1713. Chiamato **legge dei grandi numeri** e meno frequentemente **convergenza in probabilità**, può essere così enunciato: **un evento, che abbia probabilità costanti P, in una serie di prove tende a P, al crescere del numero di tentativi**. Un tipo d'incertezze e regolarità simile a quello dei giochi si verifica anche in tutte le scienze sperimentali, tra cui l'ecologia, la biologia e le scienze ambientali.

Il concetto di **probabilità classica**, fondato su una **probabilità matematica** o **a priori**, è stato il primo ad essere definito. E' il caso in cui si deve calcolare la probabilità di ottenere testa o croce con una moneta, di avere un numero da 1 a 6 con un dado, di prevedere i possibili ordini d'arrivo in una gara con diversi concorrenti, che rispettino le 4 condizioni citate. Non si richiede nessun dato sperimentale; i risultati sono conosciuti a priori, senza attendere alcuna rilevazione od osservazione, poiché è **sufficiente il solo ragionamento logico per calcolare con precisione le probabilità**. Se la moneta, il dado e la gara non sono truccate, le verifiche sperimentali si allontaneranno da questi dati attesi solo per quantità trascurabili, determinate da eventi casuali o da errori di misura.

Queste probabilità non sono determinate solamente da leggi matematiche. Una volta compresi i meccanismi della natura, molte discipline evidenziano regolarità, sovente chiamate leggi, che permettono di stimare in anticipo e con rilevante precisione i risultati di esperimenti od osservazioni. In biologia è il caso sia delle leggi di Mendel, che permettono di calcolare le frequenze genotipiche e fenotipiche attese nell'incrocio tra ibridi, sia della legge di Hardy-Weinberg, utile per definire le frequenze genetiche in una popolazione mendeliana panmittica.

Fermo restando il numero di casi possibili, la probabilità di un evento aumenta quando cresce il numero dei casi favorevoli; oppure quando, fermo restando il numero di casi favorevoli, diminuisce quello dei casi possibili. La **definizione di probabilità classica** è attribuita sia a **Bernouilli** (1654-1705) sia a **Laplace** (1749-1827): *la probabilità di un evento casuale è il rapporto tra il numero di casi favorevoli ed il numero di casi possibili, purché siano tutti equiprobabili.*

La stima di una **probabilità a priori** ha limitazioni gravi nella ricerca sperimentale: per calcolare la probabilità di un evento, è necessario conoscere preventivamente le diverse probabilità di tutti gli eventi possibili. Pertanto, questo approccio non può essere utilizzato sempre. Con la probabilità matematica non è assolutamente possibile rispondere a quesiti che per loro natura richiedono un approccio empirico, che possono essere fondati solo su osservazioni sperimentali ripetute, poiché i diversi risultati ottenibili non sono tutti ugualmente possibili, né casuali. Per esempio, come rispondere alla domanda: "Con un dado truccato, quali sono le probabilità che esca il numero 5?". Occorre fare esperimenti con lanci ripetuti, per sapere come quel dado è truccato e quindi quali siano le effettive probabilità di ogni numero.

Nello stesso modo, se nella segregazione di un diibrido o nella distribuzione di un carattere ereditario intervengono fenomeni di selezione, con un'intensità ignota, quali saranno le probabilità dei vari fenotipi di essere presenti in una data generazione? Ogni alterazione dei rapporti di equiprobabilità o dei fenomeni di casualità, i soli possono essere stimati a priori sulla base della logica, richiede l'esperienza di almeno una serie di osservazioni ripetute.

Come stima della **probabilità di un evento sperimentale** può essere utilizzata la sua **frequenza**, quando essa nelle varie ripetizioni si mantiene all'incirca costante. Una definizione chiara di questo concetto è stata data nel 1920 da **von Mises**: *la probabilità di un evento casuale è il limite a cui essa tende al crescere del numero delle osservazioni, in una serie di esperienze fatte nelle stesse condizioni.*

Se  $f$  è la frequenza relativa di un evento in una popolazione, generalmente si può osservare che all'aumentare del numero di osservazioni  $n$  del campione la frequenza tende a diventare sempre più simile a quella della popolazione. Questa affermazione non può essere dimostrata né con gli strumenti della matematica, in quanto si riferisce a dati osservazionali, né in modo empirico, poiché un esperimento non può essere ripetuto infinite volte. Tuttavia è una **regolarità statistica, genericamente chiamata legge empirica del caso**, che costituisce la base sperimentale sia di ogni teoria statistica che del ragionamento matematico.

In questi casi, si parla di **probabilità frequentista** o **frequentistica**, di **probabilità a posteriori**, di **legge empirica del caso** o di **probabilità statistica**.

Nella ricerca ambientale ed ecologica, si applica a tutti i casi in cui le leggi dei fenomeni studiati non sono note a priori, ma possono essere determinate solo a posteriori, sulla base dell'osservazione statistica. Per calcolare la probabilità di trovare un numero stabilito di individui di una certa specie in una popolazione formata da specie diverse, deve essere nota una legge o regola sulla sua presenza percentuale, che può essere stabilita solo con una serie di osservazioni. Per stimare la probabilità di trovare un individuo oltre una certa dimensione, è necessario disporre di una distribuzione sperimentale delle misure presenti in quella popolazione. **Solamente in un modo è possibile rispondere a molti quesiti empirici: concepire una serie di osservazioni od esperimenti, in condizioni uniformi o controllate statisticamente, per rilevarne la frequenza relativa.**

I due tipi di probabilità presentati, quella classica e quella frequentista, hanno una caratteristica fondamentale in comune: entrambe richiedono che **i vari eventi possano essere ripetuti e verificati in condizioni uniformi o approssimativamente tali**. In altri termini, si richiede che quanto avvenuto nel passato possa ripetersi in futuro. Ma esistono anche fenomeni che non possono assolutamente essere ridotti a queste condizioni generali, perché considerati **eventi unici od irripetibili**. Per esempio, come è possibile rispondere alle domande: "Quale è la probabilità che avvenga una catastrofe o che entro la fine dell'anno scoppi la terza guerra mondiale? Quale è la probabilità che una specie animale o vegetale a rischio effettivamente scompaia? Quale è la probabilità che un lago, osservato per la prima volta, sia effettivamente inquinato?". Sono assunzioni che presuppongono il **giudizio di numerosi individui sullo stesso fenomeno o la stima personale di un solo individuo** sulla base di un suo pregiudizio o della sua esperienza pregressa; non possono assolutamente fondarsi né sulla logica matematica né su una serie di esperimenti. Nella teoria della

probabilità si sono voluti comprendere anche questi **fenomeni non ripetibili**. Da questa scelta, deriva un'altra concezione della **probabilità**: quella **soggettiva o personalistica**.

L'obiezione fondamentale a questa **probabilità logica** è come misurare un grado di aspettativa, quando è noto che individui diversi attribuiscono probabilità differenti allo stesso fenomeno. E' una critica che viene superata dall'approccio soggettivo, secondo il quale *la probabilità è una stima del grado di aspettativa di un evento, secondo l'esperienza personale di un individuo*.

La probabilità nell'impostazione **soggettivista**, detta anche "**bayesiana**", viene intesa come una misura della convinzione circa l'esito di una prova o che accada un certo evento. E' un approccio che ha vaste ed interessanti applicazioni nelle scienze sociali ed in quelle economiche, dove la sola attesa di un fenomeno o la convinzione di una persona influente sono in grado di incidere sui fenomeni reali, come la svalutazione, i prezzi di mercato, i comportamenti sociali; in medicina, è il caso della decisione del medico sulla cura da prescrivere al paziente. Per la statistica, il problema fondamentale consiste nell'indicare come si debba modificare la probabilità soggettiva di partenza, in dipendenza dei successivi avvenimenti oggettivi, quando non si hanno repliche. Per coloro che ritengono che il mondo esterno sia una realtà oggettiva, conoscibile ed indipendente da loro, la conoscenza obiettiva non può derivare da convinzioni personali o da preferenze individuali; pertanto, l'approccio soggettivo non sarebbe attendibile, in quanto non permetterebbe la conoscenza oggettiva del reale.

Questa varietà di concetti, sinteticamente esposti in modo elementare, sul significato più esteso e comprensivo di probabilità, si riflettono sulla interpretazione delle stime ottenute da dati sperimentali; ma gli aspetti formali del calcolo variano solo marginalmente. Nel contesto delle scienze ambientali e sperimentali, **esistono casi di applicazione della probabilità soggettiva**; in particolare, **quando si tratta di scegliere una strategia o prendere una decisione** (es.: E' necessario un provvedimento immediato di divieto del traffico urbano, perché nei prossimi giorni l'inquinamento dell'aria supererà i limiti di legge?).

Nella ricerca ecologica ed ambientale, di norma predominano i casi in cui si studiano eventi ripetibili, in condizioni almeno approssimativamente uguali o simili. Pertanto, **quasi esclusivamente si fa ricorso all'impostazione frequentista della probabilità**, trascurando l'impostazione soggettivista. Nell'ultimo capitolo saranno comunque presentati alcuni concetti e metodi fondamentali della statistica bayesiana, che ha ampie prospettive di applicazioni in tutti i settori in cui occorre prendere decisioni razionali, come nella gestione dell'ambiente.

### **2.3. ALCUNE DISTRIBUZIONI DISCRETE**

Le variabili casuali hanno **distribuzioni di probabilità di due tipi: discrete o continue**. Negli esercizi precedenti, con il calcolo combinatorio, si sono stimate distribuzioni di probabilità discrete,

che possono essere calcolate per un numero definito di casi. Nelle **variabili casuali discrete**, i valori argomentali sono i numeri naturali: 0, 1, 2, ..., n. Servono per calcolare la probabilità di eventi che hanno un numero discreto di ricorrenze.

Una **variabile casuale è continua** quando la sua distribuzione è continua. Con tale variabile continua, si stima la probabilità di estrarre non un singolo valore ma valori ad esso uguali o maggiori (oppure minori). Una distribuzione continua non permette la stima della probabilità di estrarre un particolare valore, ma solo quelli compresi in un dato intervallo. Per esempio, nella distribuzione delle altezze di una popolazione di studenti, non è possibile stimare la probabilità di avere un individuo alto esattamente 176,000 cm ma quella di avere un individuo tra 180 e 190 centimetri.

### 2.3.1 DISTRIBUZIONE BINOMIALE

La **binomiale è una distribuzione teorica discreta e finita**, per eventi classificati con una **variabile binaria**. E' denominata anche **distribuzione di Bernoulli** o **distribuzione bernoulliana**, in onore del matematico svizzero J. Bernoulli (1654-1705), che ha fornito importanti contributi alla teoria della probabilità.

In un collettivo di **n** unità che possono essere ripartite solo in due classi A e B, con frequenze assolute **n<sub>a</sub>** e **n<sub>b</sub>**, le cui frequenze relative sono **p** e **q** con

$$p = \frac{n_a}{n} \quad \text{e} \quad q = \frac{n_b}{n} \quad \text{tali che} \quad p + q = 1$$

**la probabilità di avere i volte l'evento A** (e quindi **n - i** volte l'evento alternativo B) è data da

$$P_i = C_n^i p^i q^{n-i}$$

ricordando, dalle combinazioni semplici, che

$$C_n^i = \frac{n!}{i! (n-i)!}$$

**La distribuzione binomiale o bernoulliana fornisce le risposte al problema delle prove ripetute: stima le probabilità che un evento, con probabilità a priori o frequentista p, avvenga rispettivamente 0, 1, 2,...i,...n volte, nel corso di n prove identiche ed indipendenti.** Le prove possono essere successive oppure simultanee, purché siano tra loro indipendenti, non si influenzino reciprocamente e quindi le probabilità dei singoli eventi si mantengano costanti.

Le variabili casuali di tipo binario sono numerose: maschio/femmina, successo/insuccesso, malato/sano, inquinato/non inquinato, alto/basso, negativo/positivo. Inoltre tutte le variabili, sia le multinomiali sia quelle continue, possono sempre essere ridotte alla più semplice variabile dicotomica o binaria, seppure con perdita d'informazione. Per esempio, una popolazione classificata in individui

di specie diverse (A, B, C, D, E, ...) può sempre essere ricondotta ad una classificazione binaria in specie A e specie non-A; una serie di misure con scala discreta o continua, non importa se di ranghi, d'intervalli o di rapporti, può sempre essere ricondotta ad una classificazione binaria di valori superiori (+) od inferiori (-) ad un limite prefissato.

#### ESEMPIO 1

E' statisticamente dimostrato, anche se non è stata ancora trovata una spiegazione esauriente, che in tutte le popolazioni umane nascono più maschi che femmine, con un rapporto di 105-106 maschi ogni cento femmine. Possiamo quindi stabilire, a posteriori e sulla base di queste analisi, che la probabilità frequentista della nascita di un maschio è approssimativamente  $p = 0,52$  e che quella di una femmina è, di conseguenza,  $q = 0,48$  ( $q = 1 - p$ ).

Usando la distribuzione binomiale, possiamo calcolare le specifiche probabilità  $P$  di avere 0, 1, 2, 3, 4 figli maschi nelle famiglie con 4 figli:

$$P_0 = C_4^0 p^0 q^4 = 1 \cdot 1 \cdot (0,48)^4 = 0,05$$

$$P_1 = C_4^1 p^1 q^3 = 4 \cdot (0,52) \cdot (0,48)^3 = 0,23$$

$$P_2 = C_4^2 p^2 q^2 = 6 \cdot (0,52)^2 \cdot (0,48)^2 = 0,37$$

$$P_3 = C_4^3 p^3 q^1 = 4 \cdot (0,52)^3 \cdot (0,48) = 0,28$$

$$P_4 = C_4^4 p^4 q^0 = 1 \cdot (0,52)^4 \cdot 1 = 0,07$$

Se **gli eventi sono casuali ed indipendenti** e le **probabilità** di avere un maschio od una femmina **sono costanti**, in famiglie con 4 figli la probabilità

- $P_0$  di avere 0 figli maschi è 0,05
- $P_1$  di avere 1 figlio maschio è 0,23
- $P_2$  di avere 2 figli maschi è 0,37
- $P_3$  di avere 3 figli maschi è 0,28
- $P_4$  di avere 4 figli maschi è 0,07.

Non esistono altri eventi possibili oltre quelli calcolati; di conseguenza, il totale delle probabilità stimate deve necessariamente essere uguale a 1 ( $0,05 + 0,23 + 0,37 + 0,28 + 0,07 = 1,00$ ); è prassi che gli arrotondamenti siano tali da dare sempre una somma uguale a 1,00.

La rappresentazione grafica di queste probabilità

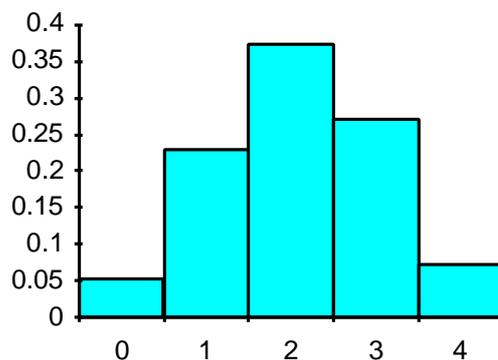


Figura 1. Probabilità del numero di maschi in famiglie con 4 figli.

mediante istogramma mostra con evidenza una distribuzione leggermente asimmetrica. La causa è il differente valore dei due eventi alternativi ( $p = 0,52$ ;  $q = 0,48$ ) e del numero basso di eventi ( $n = 4$ ). Se le probabilità  $p$  e  $q$  fossero state uguali (ovviamente entrambe  $0,5$ ) la distribuzione sarebbe stata simmetrica; con  $p$  e  $q$  diversi, diventa simmetrica all'aumentare del numero di dati, come sarà di seguito dimostrato empiricamente.

ESEMPIO 2. Applicando la stessa legge, in eventuali famiglie con 10 figli le probabilità  $P(i)$  di avere  $i$  figli é

$i$	$P(i)$
0	0.000649
1	0.007034
2	0.034289
3	0.099056
4	0.187793
5	0.244131
6	0.220396
7	0.136436
8	0.055427
9	0.013344
10	0.001446

come riportato nella tabella 1

La sua rappresentazione grafica in istogrammi (Fig. 2)

Tabella 1. Probabilità di avere  $i$  maschi in famiglie con 10 figli.

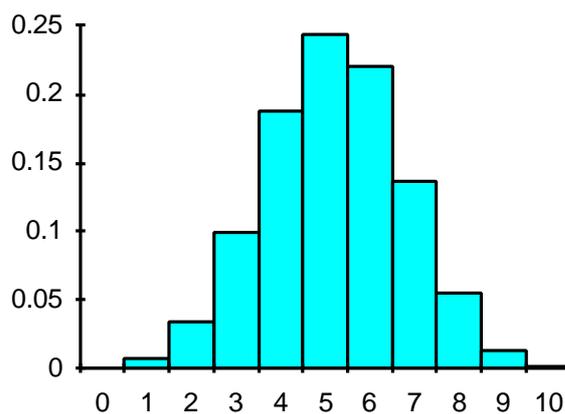


Figura 2. Probabilità del numero di maschi in famiglie con 10 figli

evidenzia meglio la leggera asimmetria.

La distribuzione di probabilità della binomiale dipende da 2 parametri:  $p$  e  $n$ .

Se  $p$  e  $q$  sono uguali a 0,5 la distribuzione è sempre simmetrica, indipendentemente da  $n$ . Se  $p$  è molto più grande o più piccolo di  $q$ , la distribuzione è asimmetrica, ma l'asimmetria tende a diminuire al crescere di  $n$ .

Di norma, si ricorre alla distribuzione binomiale per valori di  $p$  che variano da 0,1 a 0,9. Per valori di  $p$  esclusi da questo intervallo si ricorre alla distribuzione poissoniana, quando  $n$  non è grande. Quando  $n$  è così grande che anche  $n \times p$  è grande, si ricorre comunque alla normale, per qualsiasi valore di  $p$ . Quando un campione è di grandi dimensioni, la stima delle probabilità è ottenuta dalla distribuzione normale. A tal fine è intuitivo che in una distribuzione binomiale la media della popolazione se calcolata come **frequenza assoluta** è  $n \times p$

$$m = n \times p$$

mentre, se calcolata come **frequenza relativa**, è

$$m = p$$

Per esempio, in famiglie di 10 figli il numero medio di maschi è  $10 \times 0,52 = 5,2$  in frequenza assoluta e 0,52 in frequenza relativa.

Senza ricorrere alla dimostrazione, che sarebbe lunga, è ugualmente importante ricordare che la varianza  $s^2$  della popolazione in **frequenza assoluta** è data dal prodotto di  $n \times p \times q$

$$s^2 = n \times p \times q$$

mentre in una **frequenza relativa** è

$$s^2 = p \times q / n$$

I rapporti tra media e varianza offrono indicazioni importanti, quando dai dati sperimentali sia necessario risalire alla più probabile legge di distribuzione che li hanno determinati.

**Nella distribuzione binomiale la varianza è inferiore alla media:**

- con una media uguale a  $n \times p$  e una varianza uguale a  $n \times p \times q$ ,
- poiché  $p + q$  è uguale a 1 ed il valore di  $q$  è inferiore a 1,
- il valore di  $n \times p \times q$  è inferiore a  $n \times p$ .

**ESERCIZIO 1.** La distribuzione binomiale è utile anche nei casi in cui le probabilità sono note a priori, come nel lancio di dadi non truccati. Lanciando un dado 5 volte, quale è la probabilità di avere 3 volte il numero 1?

Risposta: ( $n = 5$ ;  $i = 3$ ;  $p = 1/6$ ;  $q = 5/6$ )

$$P_3 = C_5^3 p^3 q^2 = \frac{5!}{3!2!} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^2 = 0,03215$$

ESERCIZIO 2. In un'urna contenente un numero elevatissimo, praticamente infinito, di biglie nel 70% nere e per il rimanente 30% bianche, quale è la probabilità di estrarre 4 biglie tutte nere?

Risposta: (n = 4; i = 4; p = 0,7; q = 0,3)

$$P_4 = C_4^4 p^4 q^0 = \frac{4!}{4!0!} \cdot 0,7^4 \cdot 0,3^0 = 0,2401$$

ESERCIZIO 3. Un esperimento di laboratorio di solito risulta positivo nel 20% dei casi. Con 10 tentativi quale è la probabilità che 9 risultino positivi e 1 solo negativo?

Risposta: (n = 10; i = 9; p = 0,2; q = 0,8)

$$P_9 = C_{10}^9 p^9 q^1 = \frac{10!}{9!1!} \cdot 0,2^9 \cdot 0,8^1 = 0,000004096$$

ESERCIZIO 4. In un lago, la specie A rappresenta il 33% degli individui presenti, la specie B e C entrambi il 25% e la specie D il 17%; estraendo a caso 15 individui, quale è la probabilità che

- a) nessuno sia della specie A,
- b) tutti siano della specie A
- c) almeno 10 siano della specie A
- d) meno di 7 siano della specie A

Risposte :

- a) la probabilità che nessuno sia della specie A è 0,002461: vedi  $P_{(0)}$  nella tabella sottostante;
- b) la probabilità che tutti siano della specie A è minore di 1 su 1 milione di casi : vedi  $P_{(15)}$
- c) la probabilità complessiva che almeno 10 dei 15 individui estratti a caso siano della specie A è data dalla somma delle probabilità calcolate per  $P_{(10)}$ ,  $P_{(11)}$ ,  $P_{(12)}$ ,  $P_{(13)}$ ,  $P_{(14)}$  e  $P_{(15)}$ ;
- d) la probabilità complessiva che meno di 7 individui siano della specie A è data dalla somma delle singole probabilità da  $P_{(0)}$  a  $P_{(6)}$  compresi.

<b>i</b>	<b>P(i)</b>
0	0.002461
1	0.018183
2	0.062689
3	0.133798
4	0.197702
5	0.214226
6	0.175858
7	0.111364
8	0.054851
9	0.021013
10	0.00621
11	0.00139
12	0.000228
13	0.000026
14	0.000002
15	0.000000

Tabella 2. Distribuzione binomiale con  $n = 15$  e  $p = 0.33$

L'istogramma delle probabilità da  $P_{(0)}$  a  $P_{(15)}$  mostra come la distribuzione sia già approssimativamente normale, benché la probabilità di ogni singolo evento si discosti da  $1/2$ .

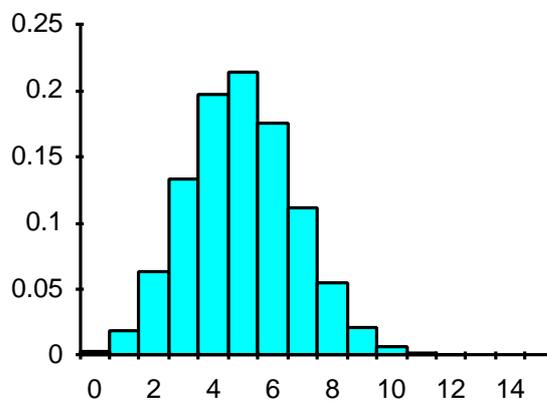


Figura 3. Iistogramma della distribuzione binomiale con  $n = 15$  e  $p = 0.33$

ESERCIZIO 5. Per dimostrare come all'aumentare del numero di osservazioni la distribuzione diventi simmetrica e pertanto come essa possa essere bene approssimata da una distribuzione normale, chi è in grado di usare e programmare un calcolatore stimi tutte le probabilità possibili con  $p = 1/6$  e  $n = 100$ , ovviamente con  $i$  che varia a 0 a 100.

Risposta.

Le probabilità associate ai diversi tipi di estrazione possono essere espresse anche dai termini dello sviluppo del binomio  $(p+q)^n$ . La loro rappresentazione grafica, riportata nel grafico sottostante, evidenzia come la distribuzione abbia forma molto simile alla normale, causa dell'alto valore di  $n$  benché  $p$  sia lontano da 0,5.

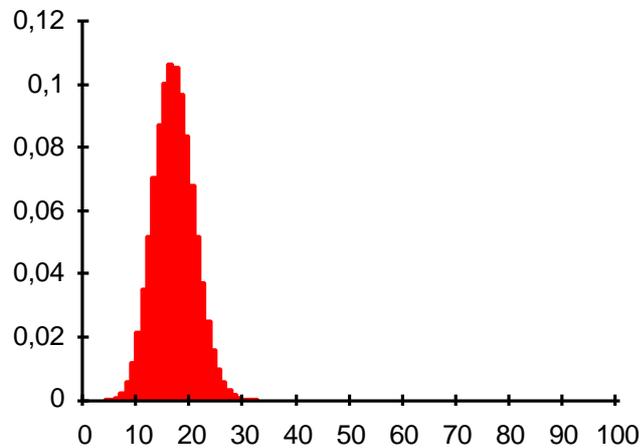


Figura 4. Istogramma della distribuzione binomiale con  $n = 100$  e  $p = 1/6$

### 2.3.2 DISTRIBUZIONE MULTINOMIALE

La distribuzione multinomiale rappresenta una estensione di quella binomiale; si applica a  $k$  eventi indipendenti di probabilità  $p_1, p_2, \dots, p_i, \dots, p_k$  (la cui somma è uguale a 1) che possono comparire nel corso di  $N$  prove indipendenti, successive o simultanee. Permette di calcolare la probabilità di ogni evento possibile, quando determinato solo dal caso.

La probabilità che si realizzino congiuntamente tutti gli eventi indicati è determinata dallo sviluppo del multinomio:

$$P_{(n_1 \ n_2 \ \dots \ n_k)} = \frac{N!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$$

ESEMPIO. Si supponga che in un lago, con un numero teoricamente infinito di pesci, il 10% ( $p_1 = 0,10$ ) siano della specie A, il 40% ( $p_2 = 0,40$ ) siano della specie B, il 20% ( $p_3 = 0,20$ ) siano di quella C ed il 30% ( $p_4 = 0,30$ ) della specie D;

a) estraendo 10 pesci, quale è la probabilità di avere 2 A, 3 B, 2 C e 3 D?

b) estraendo 8 pesci quale è la probabilità di avere 4 B e 4 D? (naturalmente con 0 A e 0 C).

Risposte:

a) La probabilità di estrarre 2 individui della specie A, 3 della specie B, 2 della specie C e 3 della specie D, in un lago in cui le quattro specie hanno le frequenze relative della domanda, è calcolata con

$$P_{(2A, 3B, 2C, 3D)} = \frac{10!}{2!3!2!3!} (0,10)^2 (0,40)^3 (0,20)^2 (0,30)^3 = 0,017418$$

e risulta uguale a 0,011612 o 11,612 per mille.

b) La probabilità di estrarre per caso, dallo stesso lago, 4 individui della specie B e 4 della specie D con 8 estrazioni casuali è

$$P_{(0A, 4B, 0C, 4D)} = \frac{8!}{0!4!0!4!} (0,10)^0 (0,40)^4 (0,20)^0 (0,30)^4 = 0,01452$$

uguale a 0,04587 o 45,87 per mille.

Non esiste un limite al numero di fattori che possono essere considerati insieme.

La distribuzione binomiale è utilizzata con frequenza nella statistica non parametrica. Se il numero di dati è ridotto, molti test non parametrici ricorrono ad essa per il calcolo delle probabilità.

Seppure apparentemente più generale, la distribuzione multinomiale ha un uso limitato, circoscritto alla stima della probabilità complessiva di più eventi indipendenti, per ognuno dei quali sia nota la probabilità  $p_i$ .

### 2.3.3 DISTRIBUZIONE POISSONIANA

Quando il numero di dati  $n$  è molto grande e la probabilità  $p$  è molto piccola, la distribuzione binomiale presenta vari inconvenienti pratici, importanti soprattutto prima dell'introduzione del calcolo automatico: l'innalzamento di frequenze molto basse a potenze elevate e il calcolo di fattoriali per numeri grandi rendono il calcolo manuale praticamente impossibile.

**Per  $n$  che tende all'infinito e  $p$  che tende a 0, in modo tale che  $np$  sia costante**, il matematico francese S. D. **Poisson** (1781-1840) nel 1837 **ha dimostrato che**

$$P_i = \frac{\mu^i}{i!} e^{-\mu} \quad \text{se} \quad \begin{array}{l} n \rightarrow \infty \\ p \rightarrow 0 \\ n \cdot p = \text{cost.} \end{array}$$

( $e = 2,718281828$ )

**E' una distribuzione teorica discreta, totalmente definita da un solo parametro, la media.**

Anche nella distribuzione poissoniana la media attesa  $m$  è data dal prodotto  $n \times p$ . Affinché le frequenze o probabilità calcolate con la legge di Poisson siano esatte,  $m$  deve essere un parametro costante per tutta la distribuzione. Anche in questa distribuzione  $s^2 = n \times p \times q$ . Si può facilmente verificare come la varianza sia uguale alla media: applicando le tre condizioni appena enunciate, avremo che la varianza della distribuzione di Poisson è data da:

$$\sigma^2 = \lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0}} npq = \lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0}} (np)q = \lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0}} c(1-p) = c = \mu$$

In termini discorsivi,

**se  $p$  tende a 0,  $q$  tende a 1; pertanto  $n \times p = n \times p \times q$ .**

La legge di distribuzione poissoniana è detta anche **legge degli eventi rari**, poiché la probabilità che l'evento si verifichi è estremamente bassa. E' chiamata pure **legge dei piccoli numeri**, in quanto la frequenza assoluta di questi eventi è espressa da un numero piccolo, anche in un numero elevato di prove.

La distribuzione poissoniana ha una forma molto asimmetrica e la classe più frequente o più probabile è zero, quando  $\mu$  è inferiore a 1; è ancora asimmetrica per valori di  $\mu$  inferiori a 3; ma una media uguale a 6-7 determina una distribuzione delle probabilità simmetrica ed è bene approssimata dalla distribuzione normale o gaussiana.

La poissoniana può essere usata al posto della binomiale, per  $p$  inferiore a 0,05 e  $n$  superiore a 100.

La distribuzione poissoniana è utilizzata per eventi che si manifestano sia nello spazio che nel tempo; è il caso del numero di insetti in uno spazio di piccole dimensioni o il numero di eventi che possono avvenire in un periodo ridotto. In molti testi, il numero medio di eventi è indicato non con  $m$  ma con  $\lambda$ , in particolare quando si tratta di eventi temporali. Per avere una distribuzione poissoniana, **una variabile casuale deve avere tre requisiti: stazionarietà, non-multiplicità, indipendenza.**

Una variabile casuale

- è **stazionaria**, quando la probabilità di ogni evento in un intervallo di tempo  $(t, t + h)$  o in uno spazio infinitesimo è approssimativamente costante, pari a  $\lambda h$  per ogni  $t$ ;
- gode del requisito della **non-multiplicità**, quando la probabilità che due o più eventi avvengano nello stesso intervallo infinitesimo di tempo o di spazio  $h$  di quello di un solo evento è molto minore, cioè la probabilità di avere due o più eventi è molto minore di quella di un evento;
- è **indipendente** se, in un intervallo ed in uno spazio finito, gli eventi si presentano in modo indipendente uno dall'altro ed il numero di eventi avvenuti in un intervallo è indipendente da quello di ogni altro.

**ESEMPIO 1.** Tassi elevati di inquinamento atmosferico possono determinare reazioni allergiche gravi. Durante un mese (30 giorni), nel pronto soccorso di un ospedale si sono avuti 27 ricoveri urgenti: la media giornaliera è 0,9 (27/30). Se non esistono correlazioni con uno o più fattori

ambientali che possono variare le medie giornaliere (es. inquinamento più elevato in certi giorni, maggiore presenza di polline in giornate serene), è ipotizzabile che gli eventi abbiano una distribuzione giornaliera costante, in accordo con la legge di Poisson.

Calcolare la probabilità (**Pi**) di avere **i** casi di allergia al giorno, per **i** che va da 0 a 8.

Risposta.

La tabella 3 riporta, per ogni evento **i**, la probabilità **P(i)** di avvenire con media uguale a 0,9.

<b>i</b>	<b>P(i)</b>
0	0.40657
1	0.365913
2	0.164661
3	0.049398
4	0.011115
5	0.002001
6	0.000300
7	0.000039
8	0.000004

Tabella 3. Distribuzione di Poisson con  $\mu = 0.9$ .

Nel 40% dei giorni non si dovrebbe avere nessun caso; si dovrebbe avere 1 solo ricoverato nel 36,6%; 2 ricoverati nel 16,5% dei giorni, ecc. fino a 8 ricoverati con probabilità 4 su un milione.

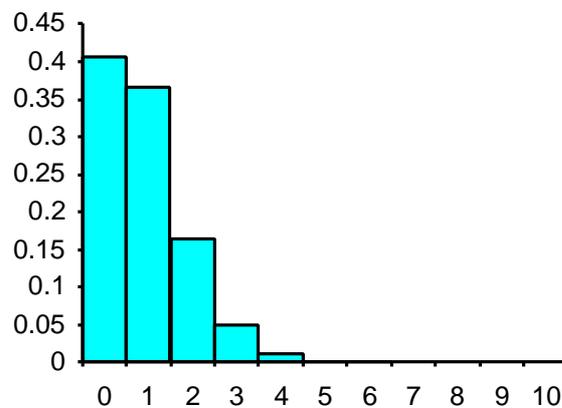


Figura 5. Distribuzione di Poisson con  $\mu = 0.9$ .

Come evidenzia il grafico, la distribuzione delle probabilità con media 0,9 è fortemente asimmetrica, con asimmetria destra.

ESEMPIO 2. E' stato ipotizzato che gli individui della specie A abbiano una distribuzione perfettamente casuale sul terreno. Dividendo il terreno in appezzamenti della stessa dimensione, si è osservato che mediamente in essi vengono trovati 2 individui.

Calcolare la frequenza attesa ( $P_i$ ) di avere appezzamenti con  $i$  individui, per  $i$  che va da 0 a 5, in una distribuzione poissoniana con media uguale a 2.

Risposta:

$i$	$P(i)$
0	0.135335
1	0.270671
2	0.270671
3	0.180447
4	0.090224
5	0.036089

Tabella 4. Distribuzione di Poisson con  $\mu = 2$

Come mostra il grafico successivo, la forma della distribuzione poissoniana con media uguale a 2 è ancora asimmetrica, seppure in modo molto meno accentuato della distribuzione precedente, che aveva una media inferiore. L'asimmetria è sempre destra o positiva, fino a quando la distribuzione diviene normale e simmetrica.

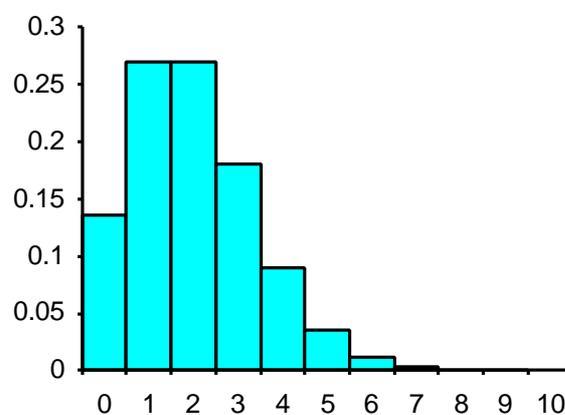


Figura 6. Distribuzione di Poisson,  $\mu = 2$

ESEMPIO 3. Nella ricerca ambientale ed ecologica, la distribuzione poissoniana serve per stimare l'insorgenza di nuovi casi entro un certo periodo di tempo, la frequenza di piante infestanti per unità di superficie, quella di trovare colonie batteriche od animali entro un volume unitario, quando gli eventi sono rari. Per rendere vera questa condizione, è sufficiente che lo spazio sia limitato ed il tempo sia breve. Inoltre è richiesta la **condizione di indipendenza**: il presentarsi di un evento, in una particolare unità di tempo o di spazio, non deve influenzare la probabilità che l'evento si ripresenti in un altro istante o luogo. La seconda condizione necessaria è la **individualità** (chiamata anche **non-multiplicità**): gli eventi avvengono singolarmente, non a coppie o a gruppi. La terza è quella dell'**omogeneità** (detta anche **stazionarietà**): la media (o la probabilità) dell'evento deve essere costante su tutto il periodo considerato o l'area in osservazione.

Per esempio, se si analizza la distribuzione di una certa specie d'insetti sulle foglie di un albero, occorre che la specie non abbia colonie solo su alcuni rami o che per il suo insediamento preferisca la parte inferiore o superiore dell'albero. La media generale di insetti presenti su una foglia deve essere generalmente valida per ogni zona dell'albero.

Tutti questi concetti sono sinteticamente compresi nell'espressione che i fenomeni non devono essere contagiosi.

In una distribuzione poissoniana con media uguale a 12, calcolare la probabilità (**Pi**) di avere esattamente 8 volte l'evento atteso, di averlo meno di 5 volte, di averlo più di 10 volte.

Risposta:

<b>i</b>	<b>P(i)</b>
0	0.000006
1	0.000074
2	0.000442
3	0.001770
4	0.005309
5	0.012741
6	0.025481
7	0.043682
8	0.065523
9	0.087364
10	0.104837
11	0.114368
12	0.114368
13	0.105570
14	0.090489
15	0.072391
16	0.054293
17	0.038325
18	0.025550
19	0.016137
20	0.009682
21	0.005533
22	0.003018
23	0.001574
24	0.000787
25	0.000378

Tabella 5. Distribuzione di Poisson con  $\mu = 12$ .

- La probabilità di averlo esattamente 8 volte è il 6,55%;
- la probabilità di averlo meno di 5 volte è dato dalla somma delle probabilità **P(i)**, con **i** che varia da 0 a 4;
- la probabilità di averlo più di 10 volte è dato dalla somma delle probabilità **P(i)** da 11 a tutte le successive; in pratica fino ad **i** uguale a 25, dove la probabilità è già molto bassa, inferiore a 4 su 10.000.

La distribuzione delle probabilità con  $\mu = 12$  ha una forma quasi perfettamente normale: le probabilità maggiori sono relative agli eventi intorno alla media 12. Da essi, le probabilità diminuiscono in modo quasi approssimativamente simmetrico, con differenze dalla normale che sono trascurabili come evidenzia la rappresentazione grafica.

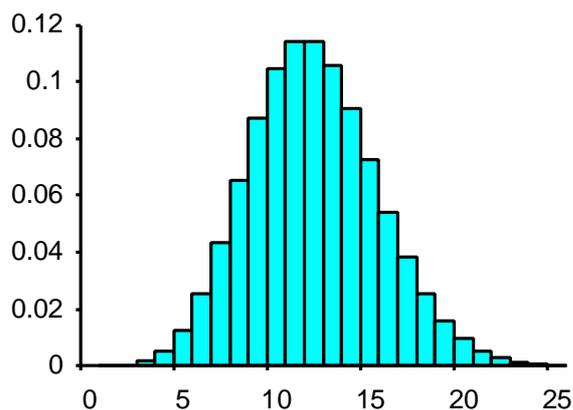


Figura 7. Distribuzione di Poisson con  $\mu = 12$ .

Con una media così elevata è più adatta la distribuzione normale, con  $\mu = 12$  e  $\sigma^2 = 12$ .

Il suo uso sarà spiegato nei paragrafi successivi.

ESEMPIO 4. In letteratura come esempio storico della distribuzione poissoniana, tratta da dati sperimentali, è famosa la rilevazione di **Bortkewitch**. Questo veterinario dell'armata prussiana del XIX secolo per 20 anni ha contato in 10 corpi d'armata il numero di soldati che ogni anno morivano a causa di un calcio di mulo; ha quindi classificato i decessi nei 200 eventi (20 corpi d'armata per 10 anni), ottenendo la tabella sottostante:

numero di decessi <b>i</b>	0	1	2	3	4
eventi osservati <b>r</b>	109	65	22	3	1

Tabella 6. Morti per corpo d'armata e per anno, rilevati da Bortkewitch.

Come riportato in essa, in 109 casi non si è avuto nessun morto, in 65 casi si è avuto 1 morto, in 22 casi sono stati contati 2 morti, in 3 casi 3 morti ed in 1 caso 4 morti. In totale, nei 200 ( $109 + 65 + 22 + 3 + 1 = 200$ ) casi esaminati il numero di morti è stato di 122 ( $109 \times 0 + 65 \times 1 + 22 \times 2 + 3 \times 3 + 1 \times 4 = 122$ ).

Il calcolo della media e della varianza, secondo le formule presentate nel primo capitolo sulle distribuzioni di frequenza, da i seguenti risultati:

$$\text{media} = \mu = 122/200 = \mathbf{0,6100}$$

$$\text{varianza} = \sigma^2 = \mathbf{0,6079}$$

E' importante osservare che la varianza di questa distribuzione sperimentale è quasi identica alla sua media, come atteso in una distribuzione poissoniana teorica. E' una buona indicazione che questi

eventi seguono la legge di Poisson; ma la dimostrazione più completa è fornita dal confronto tra la distribuzione attesa e quella osservata.

Applicando la distribuzione di Poisson, si determinano le probabilità teoriche di avere ogni anno per corpo d'armata **0** morti, **1** morto, ..., **n** morti, eseguendo i calcoli sottostanti.

$$P_0 = \frac{0,61^0}{0!} \cdot \frac{1}{2,71^{0,61}} = \frac{1}{1} \cdot \frac{1}{1,837} = 0,5440$$

$$P_1 = \frac{0,61^1}{1!} \cdot \frac{1}{2,71^{0,61}} = \frac{0,61}{1} \cdot \frac{1}{1,837} = 0,3318$$

$$P_2 = \frac{0,61^2}{2!} \cdot \frac{1}{2,71^{0,61}} = \frac{0,3721}{2} \cdot \frac{1}{1,837} = 0,1010$$

$$P_3 = \frac{0,61^3}{3!} \cdot \frac{1}{2,71^{0,61}} = \frac{0,2270}{6} \cdot \frac{1}{1,837} = 0,0203$$

$$P_4 = \frac{0,61^4}{4!} \cdot \frac{1}{2,71^{0,61}} = \frac{0,1385}{6} \cdot \frac{1}{1,837} = 0,0029$$

Le probabilità stimate sono riferite ad ogni corpo d'armata in un anno. Si ottengono i relativi eventi attesi rapportandole a 200 (20 corpi d'armata per 10 anni).

Numero di decessi	0	1	2	3	4
<b>Eventi osservati</b>	<b>109</b>	<b>65</b>	<b>22</b>	<b>3</b>	<b>1</b>
Frequenze relative attese	0,5440	0,3318	0,1010	0,0203	0,0029
<b>Eventi attesi (su 200)</b>	<b>108,80</b>	<b>66,36</b>	<b>20,20</b>	<b>4,06</b>	<b>0,58</b>

Tabella 7. Eventi osservati ed eventi attesi per corpo d'armata e per anno.

E' evidente il ridottissimo scarto tra gli **eventi osservati** (seconda riga) e quelli attesi (quarta riga) per 0 morti (109 osservati contro 108,80 attesi); parimenti ridotte sono le differenze per 1 morto (65 osservati contro 66,36 attesi), 2 morti (22 osservati contro 20,20 attesi), 3 morti (3 osservati contro 4,06 attesi) e 4 morti (1 osservato contro 0,58 atteso), anche in considerazione del fatto che **mentre gli eventi attesi sono derivati da probabilità o frequenze relative in una scala continua, i valori reali sono conteggi e quindi valori discreti.**

Anche quando una distribuzione sperimentale è totalmente determinata da una distribuzione teorica, i valori osservati non sono quasi mai identici a quelli attesi.

**Il problema reale del confronto tra frequenze attese e frequenze osservate consiste nel capire se le differenze sono di entità trascurabile, puramente imputabili al caso, oppure se sono di entità tale da lasciare presupporre l'intervento di leggi o fattori diversi da quelli ipotizzati nella distribuzione teorica.**

**Per valutare l'accordo tra distribuzione degli eventi osservati e la distribuzione degli eventi attesi, non è sufficiente la semplice descrizione precedente. Occorre ricorrere all'inferenza statistica, come sarà compiutamente esposto nel capitolo IV, per quanto riguarda il confronto tra distribuzioni di frequenza.**

ESERCIZIO - In una popolazione di plancton, la specie Y ha una presenza del 2%.

Campionando 200 individui quale è la probabilità di non trovarla?

Campionando 100 individui quale è la probabilità di trovarla 4 volte?

E se la specie avesse una presenza del 5%, quali sarebbero le probabilità precedenti?

Risposte:

1) Campionando 200 individui, se la frequenza relativa è uguale a 0,02 la media attesa di eventi è 4  
( $200 \times 0,02 = 4$ ).

La probabilità di trovare 0 individui è  $P_0 = \frac{4^0}{0!} e^{-4} = 0,0183$

2) Campionando 100 individui con frequenza relativa 0,02 la media attesa è 2.

La probabilità di trovare 4 individui è  $P_4 = \frac{2^4}{4!} e^{-2} = 0,0902$

3) Campionando 200 individui, se la frequenza relativa è uguale a 0,05 la media attesa di eventi è 10  
( $200 \times 0,05 = 10$ ).

La probabilità di trovare 0 individui è  $P_0 = \frac{10^0}{0!} e^{-10} = 0,0000454$

4) Campionando 100 individui con frequenza relativa 0,05 la media attesa è 5.

La probabilità di trovare 4 individui è  $P_4 = \frac{5^4}{4!} e^{-5} = 0,1755$

### **2.3.4 DISTRIBUZIONE GEOMETRICA E DISTRIBUZIONE DI PASCAL**

La distribuzione binomiale serve per valutare, entro un numero complessivo **n** di eventi, la probabilità **P** di avere **i** eventi favorevoli ( $P_i$ ), ognuno dei quali ha una probabilità **p** costante. Con **n** prefissato (esempio: famiglie di 4 figli) e **p** costante (esempio: la probabilità  $p = 0.52$  di avere un figlio maschio),

permette di stimare la probabilità di avere **i** volte l'evento atteso (esempio: avere 0 oppure 1 oppure 2, 3, 4 figli maschi). Nel calcolo, **n e p sono costanti, mentre i varia.**

La **distribuzione geometrica** serve per valutare la probabilità di avere **un** evento favorevole (**i = 1**). Con **i** prefissato (uguale a 1) e **p** costante, permette di stimare la probabilità (**P<sub>n</sub>**) di avere l'evento desiderato all'aumentare del numero **n** di osservazioni. Nel calcolo, **i e p sono costanti mentre n varia.**

Ad esempio, si supponga che, in un'analisi complessa di laboratorio, un esperimento abbia il 30% di probabilità di dare una risposta positiva. Quante prove occorre fare per avere la probabilità del 90 o del 95 per cento di avere una risposta positiva?

La probabilità che il primo tentativo sia positivo è p. Scritto in modo formale:

$$P(N = 1; p) = p$$

La probabilità che il secondo tentativo dia un risultato positivo è (1 - p) x p:

$$P(N = 2; p) = (1 - p) \cdot p$$

La probabilità che sia il terzo a dare il risultato positivo è

$$P(N = 3; p) = (1 - p)^2 \cdot p$$

**e la probabilità che sia l'n-esimo è**

$$P(N = n; p) = (1 - p)^{n-1} \times p$$

ESEMPIO. Con p = 0,3 calcolare la probabilità di avere l'evento desiderato alle prove da 1 a n.

Risposta:

n	P(n) con p = 0,3		Σ P(n)
1	0,3	0,3	0,300000
2	0,7 · 0,3	0,21	0,510000
3	0,7 <sup>2</sup> · 0,3	0,147	0,657000
4	0,7 <sup>3</sup> · 0,3	0,1029	0,752900
5	0,7 <sup>4</sup> · 0,3	0,07203	0,83193
6	0,7 <sup>5</sup> · 0,3	0,050421	0,882351
7	0,7 <sup>6</sup> · 0,3	0,035295	0,917646
8	0,7 <sup>7</sup> · 0,3	0,024706	0,942352
9	0,7 <sup>8</sup> · 0,3	0,017294	0,959646
10	0,7 <sup>9</sup> · 0,3	0,012106	0,971752

Mediante la formula sopra riportata, si calcola la probabilità che l'esperimento riesca al primo tentativo, al secondo e così via. Sono probabilità esatte: forniscono la probabilità che ogni singolo tentativo dia la risposta desiderata.

Se è necessario sapere quanti tentativi occorre fare, per avere una determinata probabilità che entro essi sia presente la risposta positiva, si deve utilizzare la sommatoria dei singoli valori esatti. L'elenco riportato nella tabella precedente mostra che per avere una probabilità non inferiore al 90% di avere il primo risultato positivo occorre fare 7 tentativi, mentre per una probabilità non inferiore al 95% occorre fare 9 tentativi, sempre che la probabilità sia costante e gli eventi siano indipendenti.

La **distribuzione di Pascal** (Blaise Pascal matematico e filosofo francese, nato nel 1623 e morto nel 1662) è la **generalizzazione della distribuzione geometrica: supponendo una probabilità  $p$  costante per ogni evento atteso, serve per valutare la probabilità che entro  $n$  osservazioni sequenziali sia compreso  $i$  volte l'evento atteso** (con  $i$  che può essere uguale a 1, 2, 3, ...,  $n$ ).

**La distribuzione geometrica è un caso della distribuzione di Pascal, con  $i = 1$ .**

Se ogni evento ha una probabilità  $p$  costante ed indipendente, la probabilità che esso avvenga  $i$  volte all'aumentare del numero  $n$  di osservazioni è

$$P(N = n; i; p) = \frac{(n-1)!}{(i-1)!(n-i)!} \cdot p^i \cdot (1-p)^{n-i}$$

Dove, ovviamente,  $n \geq i$  (la cui interpretazione è: per avere 3 risultati simultaneamente positivi, occorre fare almeno 3 prove).

ESEMPIO. Si supponga che, in una popolazione animale allo stato selvatico, il 40% degli individui ( $p = 0,4$ ) sia colpito da una malattia; per una serie di analisi più approfondite e proporre la cura servono 3 animali ammalati. Quanti occorre catturarne, **con un campionamento sequenziale**, per avere  $i = 3$  individui ammalati ad una probabilità prefissata?

Risposta:

Con i dati dell'esempio,  $i = 3$  e  $p = 0,4$

la formula generale divent

$$P(N = n; 3; 0,4) = \frac{(n-1)!}{(3-1)!(n-3)!} \cdot 0,4^3 \cdot 0,6^{n-3}$$

con **n** che aumenta progressivamente a partire da 3.

Con  $n = 7$ , il calcolo della probabilità di trovare tra essi 3 ammalati è

$$\frac{6!}{2!4!} \cdot 0,4^3 \cdot 0,6^4 = 15 \cdot 0,064 \cdot 0,1296 = 0,124416$$

uguale a 0,124416 o 12,44%

Nella tabella è stato riportato il calcolo delle probabilità con  $n$  che varia da 3 a 12:

<b>n</b>	<b>P(n) con i = 3 e p = 0,4</b>	<b>S P(n)</b>
3	0,064	0,064000
4	0,1152	0,179200
5	0,13824	0,317440
6	0,13824	0,455680
7	0,124416	0,580096
8	0,104509	0,684605
9	0,083607	0,768212
10	0,064497	0,832709
11	0,048373	0,881082
12	0,035473	0,916555

Se si catturano 3 animali, la probabilità che tutti e tre siano ammalati è uguale a 6,4 per cento.

Catturando un quarto animale, la probabilità che esso possa essere il terzo ammalato è di 11,52% e la probabilità complessiva di avere i tre ammalati che servono diventa 17,92 %.

Con il dodicesimo animale catturato, la probabilità di averne tre ammalati è superiore al 90%, esattamente uguale a 91,6555 per cento.

L'esempio serve anche per risolvere altri tipi di problemi, quale la frequenza reale di ammalati nella popolazione da cui si estrae il campione.

Se, con un numero elevato di animali catturati (esempio 30), non è stato possibile selezionare i 3 individui ammalati, deve sorgere il sospetto che la percentuale di ammalati sia il realtà nettamente inferiore al 40 per cento stimato. Per esempio, si supponga che con 15 animali la probabilità cumulata

superi il 95%; se con 15 animali catturati non si fosse ancora raggiunto il numero di 3 ammalati, si potrebbe ragionevolmente pensare che la frequenza  $p$  di ammalati sia minore di 0,4. Tale affermazione avrebbe una probabilità di oltre il 95% di essere vera e quindi inferiore al 5% di essere falsa.

Se con la distribuzione di **Pascal** si conta il numero di insuccessi avuti prima di arrivare al numero desiderato di casi positivi, si ottiene una distribuzione simile alla **distribuzione binomiale negativa** (che sarà presentata in un paragrafo successivo).

### 2.3.5 DISTRIBUZIONE IPERGEOMETRICA

Nella distribuzione binomiale (presa come riferimento di tutte le distribuzioni teoriche presentate), la probabilità  $p$  di un evento si mantiene sempre costante. Quando la probabilità  $p$  varia in funzione degli eventi precedenti, come succede nell'estrazione di alcuni oggetti da un campione di piccole dimensioni, si ha la distribuzione ipergeometrica.

Un esempio semplice è fornito dal gioco delle carte, con il calcolo delle probabilità nell'estrazione di un secondo re da un mazzo di 40 carte. Se il gioco avviene con reimmissione della carta già estratta, la probabilità di estrarre un re è sempre uguale a 4/40. Ma se il gioco avviene senza reintroduzione della carta estratta, la probabilità che la seconda carta sia un re varia in rapporto all'estrazione della prima: se la prima era un re, la probabilità per la seconda è 3/39; se la prima era una carta diversa, la probabilità che la seconda sia un re è 4/39.

In questa ultima condizione, le probabilità dei vari eventi sono stimate dalla distribuzione ipergeometrica, che viene espressa come rapporto tra combinazioni:

$$P_{(r/n)} = \frac{C_n^r \cdot C_{N-n}^{n_1-r}}{C_N^{n_1}}$$

dove:

- N** è il numero totale degli individui del campione (è un conteggio, quindi è un numero intero positivo);
- n** è il numero di individui estratti dal campione (è un numero intero non negativo, che al massimo può essere uguale a **N**);
- n<sub>1</sub>** il numero degli individui del campione che possiedono il carattere in oggetto (è un intero positivo, al massimo uguale a **N**);
- r** è il numero degli individui che presentano il carattere in oggetto tra quelli estratti (è un numero intero non negativo, che al massimo può essere uguale a **n<sub>1</sub>**).

La formula presentata può essere spiegata con un ragionamento logico, fondato sul calcolo combinatorio.

Si supponga che un'urna contenga  $N$  biglie, delle quali  $n_1$  bianche e  $N - n_1$  nere. Si estraggono dall'urna  $n$  biglie (con  $n \leq N$ ) senza reintroduzione; si vuole determinare la probabilità  $P(r/n)$  che delle  $n$  biglie estratte  $r$  siano bianche (con  $r \leq n$ ).

Il calcolo delle varie probabilità richiede 4 passaggi logici:

1 - delle  $N$  biglie,  $n$  possono essere estratte in  $\binom{N}{n}$  modi differenti;

2 - delle  $n_1$  biglie bianche,  $r$  possono essere estratte in  $\binom{n_1}{r}$  modi differenti;

3 - delle  $N - n_1$  biglie nere,  $n - r$  possono essere estratte in  $\binom{N - n_1}{n - r}$  modi differenti;

4 - ognuna delle  $\binom{n_1}{r}$  diverse possibilità di estrazione delle biglie bianche si combina con ognuna delle

$\binom{N - n_1}{n - r}$  possibilità d'estrazione di biglie nere;

dai quali è possibile derivare la formula riportata

$$P_{(r/n)} = \frac{C_n^r \cdot C_{N-n}^{n-r}}{C_N^n}$$

**ESEMPIO 1.** In una riserva di piccole dimensioni sono cresciuti 9 cinghiali, di cui 3 sono femmine (e ovviamente 6 sono maschi). Per ridurre il loro numero, è stata decisa una battuta di caccia, durante la quale sono stati abbattuti 5 animali, senza che vi sia stata la possibilità di fare attenzione al sesso. E' possibile stimare le diverse probabilità che nel gruppo degli animali uccisi siano compresi animali dei due sessi nei vari rapporti possibili.

Nel caso dell'esempio, le domande possono essere:

- Quale è la probabilità che vengano uccise tutte le femmine?
- Quale è la probabilità che resti 1 sola femmina?
- Quale quella che sia uccisa 1 sola femmina?
- Quale quella che sopravvivano tutte e 3 le femmine?

Risposte:

Per impostare il calcolo combinatorio, è necessario attribuire i 4 gruppi in modo corretto:

1 - totale animali presenti:  $N = 9$

2 - totale animali uccisi:  $n = 5$

3 - femmine presenti:  $n_1 = 3$

4 - femmine uccise:  $r$  a) = 3; b) = 2; c) = 1; d) = 0.

Da essi è possibile dedurre gli altri 2 gruppi:

5 - animali che sopravvivono:  $N - n$

6 - femmine che sopravvivono:  $n_1 - r$ .

<b>r</b>	<b>P</b>
0	0.047619
1	0.357143
2	0.47619
3	0.119048

Tabella 8. Probabilità di eliminare le 3 femmine su 9 cinghiali, dai quali ne sono stati uccisi 5.

Nella tabella precedente, sono riportate le probabilità **P** per i vari valori di **r**, calcolate nel modo seguente:

$$\begin{aligned} a) P_{(3/5)} &= \frac{C_5^3 \cdot C_{9-5}^{3-3}}{C_9^3} = \frac{\frac{5!}{3!2!} \cdot \frac{4!}{0!4!}}{\frac{9!}{3!6!}} = 0,119 \text{ (11,9\%)} \\ b) P_{(2/5)} &= \frac{C_5^2 \cdot C_{9-5}^{3-2}}{C_9^3} = \frac{\frac{5!}{2!3!} \cdot \frac{4!}{1!3!}}{\frac{9!}{3!6!}} = 0,4762 \text{ (47,62\%)} \\ c) P_{(1/5)} &= \frac{C_5^1 \cdot C_{9-5}^{3-1}}{C_9^3} = \frac{\frac{5!}{1!4!} \cdot \frac{4!}{2!2!}}{\frac{9!}{3!6!}} = 0,3572 \text{ (35,72\%)} \\ d) P_{(0/5)} &= \frac{C_5^0 \cdot C_{9-5}^{3-0}}{C_9^3} = \frac{\frac{5!}{0!5!} \cdot \frac{4!}{3!1!}}{\frac{9!}{3!6!}} = 0,0476 \text{ (4,76\%)} \end{aligned}$$

Non esistono altri eventi possibili oltre a quelli stimati; di conseguenza, la somma delle loro probabilità è uguale a 1(o 100%).

Le probabilità sono ovviamente uguali, se gli stessi eventi sono riferiti ai maschi. Nelle condizioni espresse nell'esempio, possono variare da 2 (quando tra i cinque animali uccisi sono presenti 3 femmine) a 5 (quando tra i cinque animali uccisi non è compresa nessuna femmina).

La rappresentazione grafica delle probabilità mostra l'effetto del diverso rapporto tra maschi e femmine nel campione, per cui la distribuzione non è simmetrica

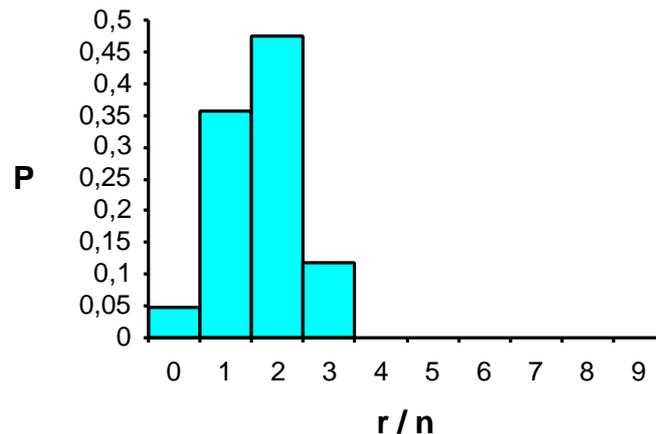


Figura 8. Probabilità di uccidere  $r$  femmine (da 0 a 3) sui 9 cinghiali.

La distribuzione ipergeometrica è **definita da 3 parametri** ( $N$ ,  $n_1$  ed  $n$ , che rappresentano nell'ordine il numero totale di individui che formano la popolazione, il numero degli oggetti del gruppo considerato, il numero di individui estratti) **in funzione del quarto** ( $r$ , il numero di individui estratti appartenenti al gruppo considerato).

**Per  $N$  che tende all'infinito, la distribuzione ipergeometrica converge verso la distribuzione binomiale**, poiché le probabilità restano praticamente costanti. Di conseguenza, **la ipergeometrica è una distribuzione tipica di eventi che riguardano i gruppi formati da poche unità**.

Nella distribuzione ipergeometrica, la media  $m$  è  $\frac{n_1}{N} \cdot n$ ; poiché  $\frac{n_1}{N} = p$ , essa risulta uguale a  $n \cdot p$  e quindi alla media della distribuzione binomiale corrispondente.

La varianza  $s^2$  è uguale a  $n \cdot p \cdot q \cdot \frac{N-n}{N-1}$ ; pertanto, (poiché  $\frac{N-n}{N-1}$  è minore di 1) è inferiore alla varianza della binomiale, per gli stessi valori di  $N$  e di  $p$ .

ESEMPIO 2. In un lago sono presenti 12 pesci di grandi dimensioni di specie diverse, di cui il 50% della stessa specie A. Estraendo 4 pesci a caso, quale è la probabilità che nessuno sia della specie A?

Risposta.

Dopo aver individuato i 4 parametri della distribuzione ipergeometrica

$$N = 12; n_1 = 6; n = 4; r = 0$$

la probabilità risulta uguale

$$P_{(0/4)} = \frac{28}{294} = 0,0303$$

a 0,0303 o 3,03 per cento.

In questo caso specifico, relativamente semplice, è possibile arrivare alla soluzione con un altro approccio in 4 passaggi logici, utile a comprendere come variano le probabilità nella distribuzione ipergeometrica:

- a) la probabilità di non prendere un pesce della specie A alla prima estrazione è 6/12;
- b) alla seconda estrazione i pesci non-A sono rimasti 5 su un totale di 11; pertanto la probabilità di estrarre un pesce non-A è 5/11;
- c) alla terza estrazione la probabilità è scesa a 4/10;
- d) alla quarta estrazione diventa 3/9;

di conseguenza, la probabilità complessiva è:

$$P = 6/12 \times 5/11 \times 4/10 \times 3/9 = 360/11880 = 0,0303$$

esattamente coincidente con il valore calcolato in precedenza con la formula generale.

### 2.3.6 DISTRIBUZIONE BINOMIALE NEGATIVA

Le distribuzioni teoriche sono numerose. Tra quelle discrete, nella ricerca ambientale è importante anche la distribuzione binomiale negativa, della quale si illustrano solo alcune caratteristiche, data la sua complessità per un corso elementare di statistica.

Nella distribuzione binomiale  $p + q = 1$ , dove  $p$  indica la probabilità di successi e  $q$  la probabilità dell'evento alternativo. Con  $n$  prove, le probabilità dei diversi eventi sono determinate dallo sviluppo di  $(p + q)^n$ ; la varianza ( $n \times p \times q$ ) è inferiore alla media ( $n \times p$ ), dato che  $q < 1$ .

La distribuzione binomiale negativa si applica quando la varianza ha un valore superiore a quello della media. E' possibile partire da una considerazione ovvia: in una distribuzione di dati sperimentali, come quella di animali che vivono in colonia contati su unità di superficie, è possibile che la varianza ( $n \times p \times q$ ) risulti superiore alla media ( $n \times p$ ). Da questa semplice osservazione sperimentale si deduce:

- a) poiché  $n \cdot p \cdot q > n \cdot p$ , deve essere  $q > 1$ ; quindi, necessariamente  $p$  (uguale a  $1 - q$ ) ha un valore negativo;
- b) con  $p$  negativo, poiché la media ( $n \cdot p$ ) deve essere positiva (in quanto media di eventi), anche  $n$  è negativo. Ma  $n$  è un conteggio e non può essere negativo. Quindi, ponendo  $n = -k$ , con  $k$  intero positivo, la potenza del binomio assume la forma

$$c) \quad [q + (-p)]^{-k} = (q - p)^{-k}$$

Come quella binomiale e quella poissoniana, la distribuzione binomiale negativa si presta al calcolo delle probabilità di eventi misurati mediante un conteggio. E' impiegata nelle scienze ambientali soprattutto nei conteggi di popolazioni animali e negli studi di epidemiologia. E' il caso in cui, (ad es., nello studio di distribuzioni di afidi) si contano le foglie che hanno la presenza di 0 animali, di 1 animale, di 2 animali. E' ugualmente utile nell'analisi della letalità di una malattia, per stimare quanti sono i periodi (giorni, settimane o mesi) con 0 morti, 1 morto, 2 morti, ...,  $n$  morti, che sono ovviamente contati con numeri interi positivi.

Per  $n$  grande e probabilità basse, quando la media è unica le frequenze attese sono fornite dalla distribuzione poissoniana. Ma quando il fenomeno è complesso, la distribuzione di frequenza è spesso determinata da due o più fattori, ognuno con media diversa; ne deriva che la variabilità aumenta e la varianza è superiore alla media. In questi casi, la distribuzione delle frequenze può essere stimata in modo appropriato dalla distribuzione binomiale negativa.

**La distribuzione binomiale negativa può essere vista come una mistura o combinazione di distribuzioni poissoniane.**

La probabilità  $P_i$  che l'evento atteso si verifichi  $i$  volte, con  $i$  che varia da  $0$  a  $k$ , è data da

$$P_i = \frac{(k + i - 1)! \left(\frac{p}{q}\right)^k}{i!(k - 1)!q^k}$$

I parametri essenziali, poiché gli altri possono essere derivati matematicamente da questi come precedentemente dimostrato, sono 2: la media  $n \cdot p$  e l'esponente  $k$  ( $-n$ ).

Dopo aver calcolato la media e la varianza dalla distribuzione sperimentale, il valore di  $k$  può essere ottenuto in vari modi; il più semplice, ma non sempre applicabile, è

$$k = \frac{(n \cdot p)^2}{n \cdot p \cdot q - n \cdot p}$$

Per calcolare le probabilità di una distribuzione binomiale negativa servono due parametri: la media (che corrisponde a  $n \cdot p$ ) e la probabilità ( $p$ ). Nell'istogramma sottostante è rappresentata una distribuzione binomiale negativa con media uguale a 6,66 e probabilità uguale a 0,6. Anche essa tende rapidamente alla normalità, quando la probabilità è prossima a 0,5.

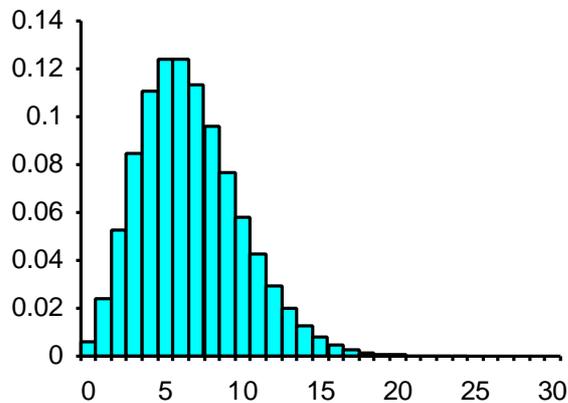


Figura 9. Distribuzione binomiale negativa ( $\mu = 6.66$ ,  $p = 0.6$ )

### 2.3.7 DISTRIBUZIONE UNIFORME O RETTANGOLARE

La più semplice distribuzione discreta è quella uniforme; la sua caratteristica fondamentale è l'identica possibilità del verificarsi di tutti i risultati possibili. Per esempio, la probabilità che esca un numero da 1 a 6 con il lancio di un dado non truccato è uguale per ognuno dei 6 possibili risultati.

Ad essa si ricorre con frequenza quando, in assenza di ipotesi specifiche più dettagliate o di una conoscenza precisa del fenomeno, la verifica di una distribuzione uniforme rappresenta un punto di partenza minimo, che è quasi sempre accettabile.

L'espressione matematica con cui calcolare la probabilità che una variabile discreta  $X$ , che segue la distribuzione uniforme, assuma un particolare valore è stabilita da

$$P(x) = \frac{1}{(b-a) + 1}$$

dove:

$b$  = risultato maggiore possibile di  $X$

$a$  = risultato minore possibile di  $X$ .

Nell'esempio dei dadi, è semplice verificare che con  $b = 6$  e  $a = 1$  si ottiene una probabilità

$$P(x) = \frac{1}{(6-1) + 1} = \frac{1}{6}$$

pari a 1/6.

Nella distribuzione discreta uniforme la media è

$$\mu = \frac{a + b}{2}$$

e la deviazione standard  $\sigma$  è

$$s = \sqrt{\frac{[(b - a) + 1]^2 - 1}{12}}$$

La utilizzazione della distribuzione rettangolare è limitata quasi esclusivamente all'analisi di probabilità a priori. Un caso che ricorre con frequenza nella ricerca ambientale è l'analisi della frequenza di animali in varie aree, per verificare un'omogeneità di dispersione. L'ipotesi alternativa è l'esistenza di una associazione tra presenza (o assenza) della specie e la condizione ambientale dell'area.

## 2.4. ALCUNE DISTRIBUZIONI CONTINUE

Tutti i modelli precedenti forniscono la distribuzione teorica di variabili casuali discrete. Quando si devono descrivere **variabili casuali continue e positive**, come peso, altezza, reddito, tempo, i modelli più utili sono quelli di seguito riportati. Tra essi, la distribuzione più frequente e di maggiore utilità nella ricerca sperimentale, **fondamento della statistica parametrica, è la distribuzione normale o gaussiana.**

### 2.4.1 DISTRIBUZIONE NORMALE O DI GAUSS

La più importante distribuzione continua è la curva normale. E' stata individuata per la prima volta da De Moivre (1733) ed è stata proposta da Gauss (1809) nell'ambito della teoria degli errori. Nella letteratura francese è attribuita anche a Laplace (1812), che ne avrebbe definito le proprietà principali prima della trattazione più completa fatta da Gauss in varie riprese, a partire dal 1809.

Il nome di **curva normale** deriva dalla convinzione, non sempre corretta, che molti fenomeni, da quelli biologici e quelli fisici, normalmente si distribuiscano secondo la **curva gaussiana**. La sua denominazione di **curva degli errori accidentali**, diffusa soprattutto nelle discipline fisiche, deriva dall'osservazione sperimentale che la distribuzione degli errori, commessi quando si misura ripetutamente la stessa grandezza, è molto bene approssimata da tale curva.

Sotto l'aspetto matematico, la distribuzione gaussiana può essere considerata come il limite della distribuzione binomiale

- per **n** che tende all'infinito,

- mentre né **p** né **q** tendono a 0 (condizione che la differenzia dalla poissoniana).

Se **n** tende all'infinito e **p** resta costante, la media (**n×p**) a sua volta si approssima all'infinito e rende la distribuzione senza applicazioni pratiche. Per contro, la variabile considerata, che nel caso di pochi

dati era quantificata per unità discrete, può essere espressa in unità sempre minori, tanto che diventa accettabile esprimerla come una grandezza continua.

La distribuzione gaussiana può essere considerata il limite anche della distribuzione poissoniana, quando  $i$  e  $m$  diventano molto grandi.

Quando  $n$  tende all'infinito (in realtà quando  $n$ ,  $i$  e  $m$  sono molto grandi) a condizione che né  $p$  né  $q$  tendano a 0, secondo il teorema di De Moivre (1833) la probabilità  $P_i$  della distribuzione binomiale è sempre meglio approssimata da

$$P(i) = \frac{1}{\sqrt{2\pi \cdot n \cdot p \cdot q}} e^{-\frac{(i - n \cdot p)^2}{2 \cdot n \cdot p \cdot q}}$$

Sostituendo  $n \cdot p$  con la media  $m$  della popolazione,

$n \cdot p \cdot q$  con la varianza  $s^2$  della popolazione

e il conteggio  $i$  (indice di un valore discreto) con  $x$  (indice di una misura continua),

si ottiene

$$y = f(x) = \frac{1}{\sqrt{2\pi \cdot s^2}} e^{-\frac{(x - m)^2}{2s^2}}$$

che è l'espressione della **funzione di densità di probabilità** (o delle frequenze relative) della distribuzione normale. In termini meno matematici, **permette di stimare il valore di Y (il valore dell'ordinata o altezza della curva) per ogni valore di X (il valore della ascissa).**

La curva normale ha la forma rappresentata nella figura 10.

La distribuzione normale con media  $\mu$  e varianza  $s^2$  è indicata con  $N(\mu, s)$ ; al variare di questi due parametri che la definiscono compiutamente, si possono avere infinite curve normali.

Le caratteristiche più importanti della normale sono una frequenza relativamente più elevata dei valori centrali e frequenze progressivamente minori verso gli estremi. La funzione di densità è simmetrica rispetto alla media: cresce da zero fino alla media e poi decresce fino a  $+\infty$ . Ha due flessi: il primo, ascendente, nel punto  $\mu - \sigma$ ; il secondo, discendente, nel punto  $\mu + \sigma$ .

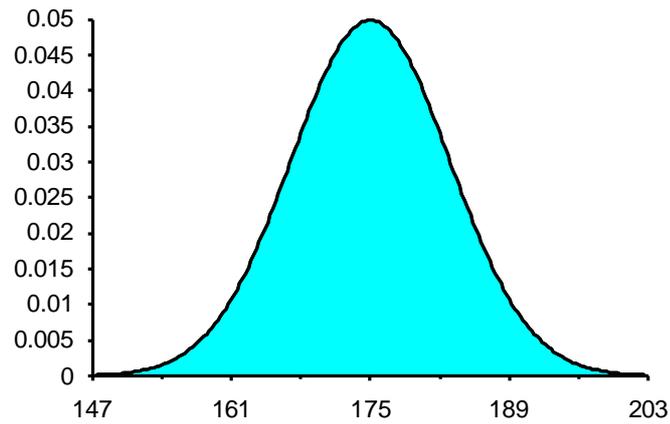


Figura 10. Distribuzione normale ( $\mu=175$ ,  $\sigma=8$ ).

In ogni curva normale, la media, la moda e la mediana sono coincidenti.

Se  $\mu$  varia e  $\sigma$  rimane costante, si hanno infinite curve normali con la stessa forma e la stessa dimensione, ma con l'asse di simmetria in un punto diverso. Quando due distribuzioni hanno media differente, è possibile ottenere l'una dall'altra mediante traslazione o trasformazione lineare dei dati.

Se invece  $\mu$  rimane costante e  $s$  varia, tutte le infinite curve hanno lo stesso asse di simmetria; ma hanno forma più o meno appiattita, secondo il valore di  $s$ .

Le due curve della figura 11 hanno media  $m$  identica e deviazione standard  $s$  differente.

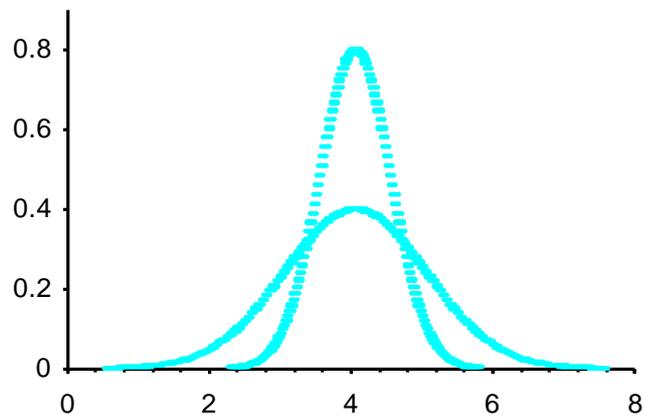


Figura 11. Curve normali con  $\mu$  uguale e  $s$  diversa.

Una seconda serie di curve con medie diverse e dispersione uguale è quella riportata nella figura successiva (Fig.12).

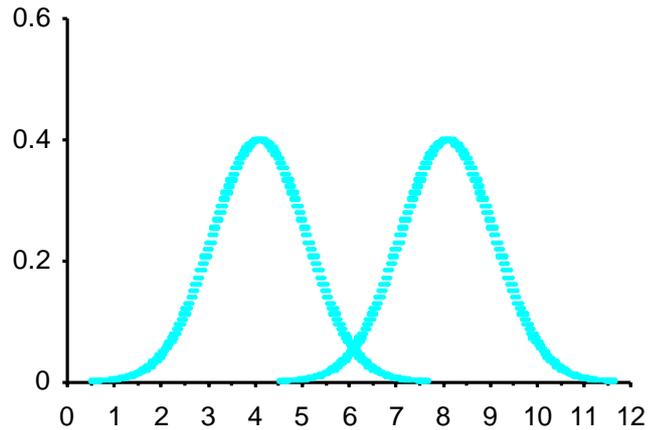


Figura 12. Curve normali con  $\mu$  diversa e  $s$  uguale.

Una terza serie è quella delle distribuzioni normali che differiscono sia per la media sia per la dispersione dei dati (figura 13).

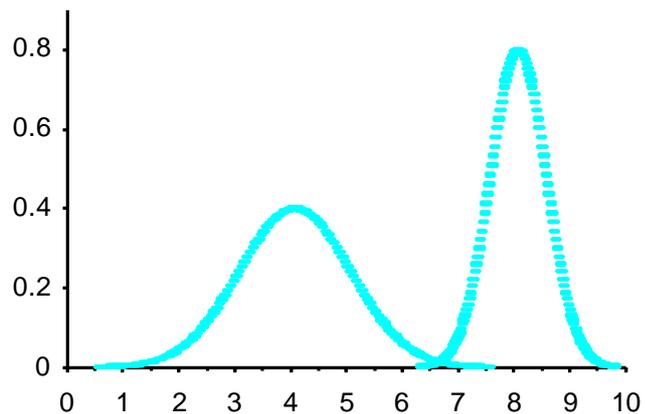


Figura 13. Curve normali  $\mu$  e  $s$  diverse.

**I momenti e gli indici di forma** della distribuzione normale (già presentati nel primo capitolo sulla statistica descrittiva) valutano in modo sintetico e mediante una rappresentazione numerica le sue caratteristiche essenziali, rese visibili e di più facile lettura, ma non quantificate, nella rappresentazione grafica. Poiché la distribuzione teorica normale è simmetrica, tutti i momenti di ordine dispari dalla media sono nulli. Per i momenti di ordine pari, è bene ricordare che

- il momento di secondo ordine è uguale alla varianza ( $\mu_2 = s^2$ )
- ed il momento di quarto ordine è uguale a 3 volte la varianza al quadrato ( $\mu_4 = 3s^4$ ).

L'indice di **simmetria di Pearson** risulta  $\beta_1 = 0$ .

L'indice di **curtosi di Pearson** è  $b_2 = \frac{m_4}{s^4} = \frac{3s^4}{s^4} = 3$ .

L'indice di **simmetria di Fisher** è  $\gamma_1 = 0$ .

L'indice di **curtosi di Fisher** è  $g_2 = \frac{\mu_4}{s^4} - 3 = 0$

Le infinite forme della distribuzione normale, determinate dalle combinazioni di differenze nella media e nella varianza, possono essere tutte ricondotte alla medesima forma. E' la **distribuzione normale standardizzata** o **normale ridotta**, che è ottenuta mediante il cambiamento di variabile dato da

$$z = \frac{x - m}{s}$$

La **standardizzazione** è una trasformazione che consiste nel:

- rendere la media nulla ( $m = 0$ ), poiché ad ogni valore viene sottratta la media;
- prendere la deviazione standard  $s$  come unità di misura ( $s = 1$ ) della nuova variabile.

Come conseguenza, si ottiene anche una trasformazione degli scarti  $x - \mu$  in scarti ridotti,

$$z' = \frac{x - \mu}{\sigma}$$

La distribuzione **normale ridotta** viene indicata con **N(0,1)**., che indica appunto una distribuzione normale con media 0 e varianza uguale a 1.

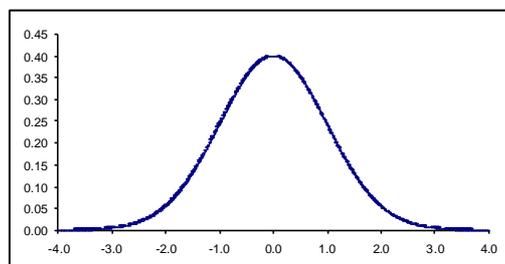
Dopo il cambiamento di variabile, nella normale ridotta la densità di probabilità è data da

$$y = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$$

dove

**z** il valore sull'asse delle ascisse, misurato in unità di deviazioni standard dalla media.

Tale relazione evidenzia come **la forma della distribuzione non dipenda più né dalla sua media**



**né dalla sua varianza: è sempre identica, qualunque sia la distribuzione gaussiana considerata.**

Y = Ordinata della curva normale standardizzata in z.

<b>z</b>	<b>.00</b>	<b>.01</b>	<b>.02</b>	<b>.03</b>	<b>.04</b>	<b>.05</b>	<b>.06</b>	<b>.07</b>	<b>.08</b>	<b>.09</b>
<b>0.0</b>	.3989	.3989	.3989	.3988	.3986	.3984	.3982	.3980	.3977	.3973
<b>0.1</b>	.3970	.3965	.3961	.3956	.3951	.3945	.3939	.3932	.3925	.3918
<b>0.2</b>	.3910	.3902	.3894	.3885	.3876	.3867	.3857	.3847	.3836	.3825
<b>0.3</b>	.3814	.3802	.3790	.3778	.3765	.3752	.3739	.3725	.3712	.3697
<b>0.4</b>	.3683	.3668	.3653	.3637	.3621	.3605	.3589	.3572	.3555	.3538
<b>0.5</b>	.3521	.3503	.3485	.3467	.3448	.3429	.3410	.3391	.3372	.3352
<b>0.6</b>	.3332	.3312	.3292	.3271	.3251	.3230	.3209	.3187	.3166	.3144
<b>0.7</b>	.3123	.3101	.3079	.3056	.3034	.3011	.2989	.2966	.2943	.2920
<b>0.8</b>	.2897	.2874	.2850	.2827	.2803	.2780	.2756	.2732	.2709	.2685
<b>0.9</b>	.2661	.2637	.2613	.2589	.2565	.2541	.2516	.2492	.2468	.2444
<b>1.0</b>	.2420	.2396	.2371	.2347	.2323	.2299	.2275	.2251	.2227	.2203
<b>1.1</b>	.2179	.2155	.2131	.2107	.2083	.2059	.2036	.2012	.1989	.1965
<b>1.2</b>	.1942	.1919	.1895	.1872	.1849	.1826	.1804	.1781	.1758	.1736
<b>1.3</b>	.1714	.1691	.1669	.1647	.1626	.1604	.1582	.1561	.1539	.1518
<b>1.4</b>	.1497	.1476	.1456	.1435	.1415	.1394	.1374	.1354	.1334	.1315
<b>1.5</b>	.1295	.1276	.1257	.1238	.1219	.1200	.1182	.1163	.1145	.1127
<b>1.6</b>	.1109	.1092	.1074	.1057	.1040	.1023	.1006	.0989	.0973	.0957
<b>1.7</b>	.0940	.0925	.0909	.0893	.0878	.0863	.0848	.0833	.0818	.0804
<b>1.8</b>	.0790	.0775	.0761	.0748	.0734	.0721	.0707	.0694	.0681	.0669
<b>1.9</b>	.0656	.0644	.0632	.0620	.0608	.0596	.0584	.0573	.0562	.0551
<b>2.0</b>	.0540	.0529	.0519	.0508	.0498	.0488	.0478	.0468	.0459	.0449
<b>2.1</b>	.0440	.0431	.0422	.0413	.0404	.0396	.0387	.0379	.0371	.0363
<b>2.2</b>	.0355	.0347	.0339	.0332	.0325	.0317	.0310	.0303	.0297	.0290
<b>2.3</b>	.0283	.0277	.0270	.0264	.0258	.0252	.0246	.0241	.0235	.0229
<b>2.4</b>	.0224	.0219	.0213	.0208	.0203	.0198	.0194	.0189	.0184	.0180
<b>2.5</b>	.0175	.0171	.0167	.0163	.0158	.0154	.0151	.0147	.0143	.0139
<b>2.6</b>	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110	.0107
<b>2.7</b>	.0104	.0101	.0099	.0096	.0093	.0091	.0088	.0086	.0084	.0081
<b>2.8</b>	.0079	.0077	.0075	.0073	.0071	.0069	.0067	.0065	.0063	.0061
<b>2.9</b>	.0060	.0058	.0056	.0055	.0053	.0051	.0050	.0048	.0047	.0046
<b>3.0</b>	.0044	.0043	.0042	.0040	.0039	.0038	.0037	.0036	.0035	.0034
<b>3.1</b>	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026	.0025	.0025
<b>3.2</b>	.0024	.0023	.0022	.0022	.0021	.0020	.0020	.0019	.0018	.0018
<b>3.3</b>	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014	.0013	.0013
<b>3.4</b>	.0012	.0012	.0012	.0011	.0011	.0010	.0010	.0010	.0009	.0009
<b>3.5</b>	.0009	.0008	.0008	.0008	.0008	.0007	.0007	.0007	.0007	.0006
<b>3.6</b>	.0006	.0006	.0006	.0005	.0005	.0005	.0005	.0005	.0005	.0004
<b>3.7</b>	.0004	.0004	.0004	.0004	.0004	.0004	.0003	.0003	.0003	.0003
<b>3.8</b>	.0003	.0003	.0003	.0003	.0003	.0002	.0002	.0002	.0002	.0002
<b>3.9</b>	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001

Figura 14 con tabella delle ordinate per ascisse (z) della distribuzione normale standardizzata

La tabella riportata serve solamente per rendere veloce la stima dell'ordinata. Una volta fissato il valore di z, è possibile calcolare il valore dell'ordinata, come nell'esempio successivo per il punto individuato in ascissa da z uguale a 1.

Utilizzando la formula

$$y = \frac{1}{\sqrt{2p}} \cdot e^{-\frac{z^2}{2}}$$

con i dati dell'esempio si ottiene

$$y = \frac{1}{\sqrt{2 \times 3,14}} \cdot \frac{1}{2,71828} = \frac{1}{2,5060} \cdot \frac{1}{1,6488} = 0,399 \times 0,6065 = 0,2420$$

un valore di Y uguale a 0,2420.

La tabella precedente riporta i valori della ordinata della curva normale standardizzata in **z**, per **z** che varia da 0 a 3.99.

In essa si trovano i valori delle ordinate, per valori di **z** riportati sommando la prima colonna (in grassetto) con la prima riga (in grassetto).

Ad esempio, per **z** uguale a 1.00 si deve

- individuare nella prima colonna (in grassetto) il valore 1.0
- individuare nella prima riga (in grassetto) il valore 0.00,

perché la somma 1.0 + 0.00 è uguale a 1.00. Nel loro incrocio, è riportato .2420 che indica appunto il valore dell'ordinata.

Per **z** uguale a 1,85 si deve

- individuare nella prima colonna (in grassetto) il valore di 1.8
- individuare nella prima riga (in grassetto) il valore .05

e leggere il valore riportato al loro incrocio, uguale a .0721 che indica appunto l'ordinata per **z** uguale a 1,85.

#### **2.4.2. DISTRIBUZIONI ASINTOTICAMENTE NORMALI, CON APPROSSIMAZIONI E TRASFORMAZIONI**

L'interesse per le applicazioni della distribuzione normale dipende dal fatto che molte variabili sono distribuite secondo questa legge; inoltre, è accresciuto dal fatto che varie distribuzioni, che non sono rigorosamente normali o addirittura lontane dalla normalità, possono divenirle od essere ritenute tali quando

- 1) certi loro parametri tendono all'infinito (**leggi asintoticamente normali**),
- 2) sono quasi normali (**approssimazioni**),
- 3) oppure mediante trasformazioni appropriate, che conducono a variabili distribuite normalmente almeno in modo approssimato (**trasformazioni**).

1) Come esempi di **leggi asintoticamente normali** si possono ricordare 3 casi già presentati nei paragrafi precedenti.

a - **La distribuzione binomiale  $(p + q)^n$  tende alla legge di distribuzione normale**, quando **n** tende all'infinito.

b - **La distribuzione poissoniana tende alla distribuzione gaussiana quando la media è elevata**; in pratica superiore a 6.

c - **La media di n variabili aleatorie indipendenti**, che singolarmente seguono una legge qualunque, **segue la legge normale quando n è grande**.

Non sono distribuzioni esattamente normali, ma sono considerate approssimativamente tali. E' il **teorema centrale** della statistica, noto anche come **teorema fondamentale della convergenza stocastica** o **teorema di Laplace-Tchebycheff-Liapounoff** (spesso citato solo come teorema di Laplace): **“Qualunque sia la forma della distribuzione di n variabili casuali indipendenti  $(x_i)$ , la loro somma  $X$  (con  $X = x_1 + x_2 + x_3 + \dots + x_n$ ) è asintoticamente normale, con media generale uguale alla somma delle singole medie e varianza generale uguale alla somma delle singole varianze”**.

Una dimostrazione semplice può essere fornita dal lancio dei dadi. Con un solo dado, i 6 numeri hanno la stessa probabilità ed hanno una distribuzione uniforme; ma con due o più dadi, la somma dei loro numeri tende ad essere sempre più simile alla normale, all'aumentare del numero dei dadi. Se invece della somma si considera la media di **n** lanci, si ottengono i medesimi risultati.

2) Come esempio di **approssimazione alla normale di una distribuzione non normale**, è possibile ricordare che nelle popolazioni animali e vegetali abitualmente la distribuzione normale viene usata sia nello studio della massa o volume, come nello studio dell'altezza o delle dimensioni di singoli individui; ma tra essi il rapporto non è lineare e quindi queste variabili non potrebbero essere tutte contemporaneamente rappresentate con la stessa legge di distribuzione. **Nella pratica, sono tutte ugualmente bene approssimate dalla distribuzione normale**; quindi, per tutte indistintamente è prassi ricorrere all'uso della normale.

3) Quando i dati hanno una distribuzione differente dalla normale, **spesso una semplice trasformazione conduce ad una distribuzione normale**. E' il caso delle trasformazioni con la **radice quadrata o cubica**, oppure con il **reciproco**, l'**elevamento a potenza** o con i **logaritmi**. Oltre agli indici statistici sulla forma (varianza, simmetria e curtosi), che misurano come la distribuzione trasformata modifichi i suoi parametri, nella scelta del tipo di trasformazione occorre considerare anche la legge biologica o naturale che determina il fenomeno di dispersione dei dati.

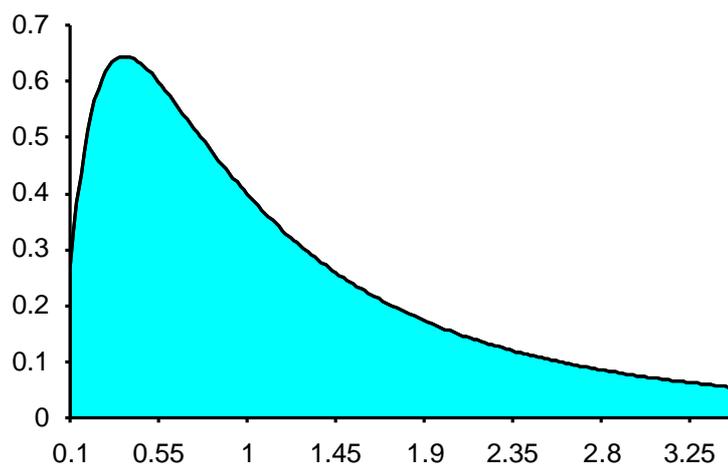


Figura 15. Distribuzione lognormale ( $\mu = 0,2$ ;  $\sigma = 1$ )

Il caso di trasformazione che ricorre forse con frequenza maggiore in biologia e nelle scienze ambientali è quella logaritmica

$$X' = \log X$$

dove

$X'$  diviene una serie di valori distribuiti in buon accordo con la normale.

Quando la distribuzione di una variabile  $X$  ha una forma simile a quella rappresentata nella precedente figura 15, con la trasformazione logaritmica in  $X'$  assume appunto una forma molto simile alla distribuzione normale.

Le trasformazioni dei dati, qui citate in modo estremamente sintetico con l'unico scopo di presentare il concetto di distribuzioni normali dopo trasformazione, sono numerose. Saranno presentate in modo più approfondito alla fine del secondo capitolo sull'analisi della varianza, quando si discuterà sulle condizioni di validità dei test di statistica parametrica.

La distribuzione normale di **Gauss** e **Laplace**, derivata dai loro studi sulla teoria della distribuzione degli errori d'osservazione, per lungo tempo ha occupato un posto di assoluta preminenza nella statistica. E' stato considerato assiomatico che, con un numero elevato di osservazioni, la variabile casuale avesse una distribuzione normale. Più recentemente sono stati sollevati dubbi e critiche; alcuni critici e autori di testi di statistica sono giunti ad affermare che tale assunzione era accettata universalmente solo perché **“gli statistici sperimentali pensavano che essa derivasse da un teorema, mentre i matematici pensavano che essa fosse un fatto sperimentale”**.

Benché la sua importanza sia stata ridimensionata, attualmente è diffusamente accettato che **moltissime distribuzioni sono approssimativamente normali**.

### 2.4.3 DALLA DISUGUAGLIANZA DI TCHEBYCHEFF ALL'USO DELLA DISTRIBUZIONE NORMALE

Nella pratica statistica, **le proprietà più utili della distribuzione normale** non sono i rapporti tra ascissa ed ordinata, presentati in precedenza, **ma le relazioni tra la distanza dalla media e la densità di probabilità sottesa dalla curva**. In modo più semplice, è possibile definire quanti sono i dati compresi tra la media ed un determinato valore, misurando la distanza dalla media  $m$  in unità di deviazioni standard  $s$ .

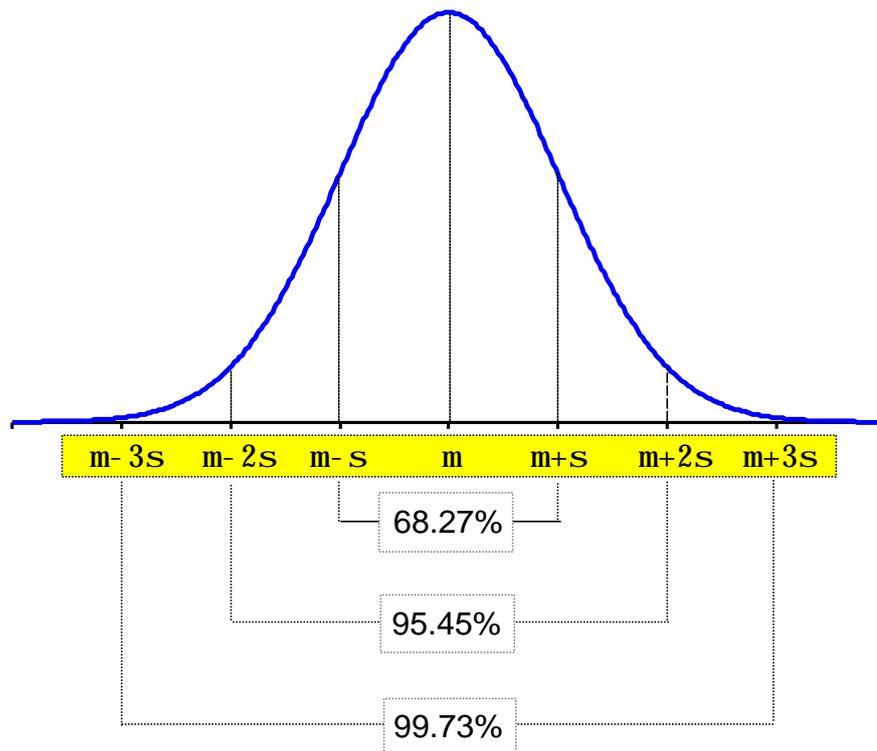


Figura 16. Relazioni tra distanza dalla  $m$  in  $s$  e densità di probabilità sottesa dalla curva.

La frazione dei casi compresi

- fra  $\mu+\sigma$  e  $\mu-\sigma$  è uguale al 68,27% (in cifra tonda o in valore approssimato  $\frac{2}{3}$ ),
- quella fra  $\mu+2\sigma$  e  $\mu-2\sigma$  è uguale 95,45% (in cifra tonda 95%),
- quella fra  $\mu+3\sigma$  e  $\mu-3\sigma$  è esattamente uguale al 99,73% (circa il 99,9%).

In pratica, nella curva normale la quasi totalità dei dati è compresa nell'intorno della media di ampiezza  $3s$ .

La relazione tra la percentuale di dati sottesi dalla curva e le dimensioni dell'intervallo tra due valori è una caratteristica di rilevante importanza nella statistica applicata: **se la distribuzione è normale, è sufficiente conoscere due parametri di una serie di dati, la media  $m$  e la varianza  $s^2$**  (o altro

parametro da esso derivato come la deviazione standard  $s$ ), **per conoscere anche la sua distribuzione.**

Più di un secolo fa, a partire da dati sperimentali due matematici **Bienaymé e Tchebycheff** (qui scritto secondo la pronuncia inglese, in altri testi anche Cebycev secondo quella francese) avevano enunciato: **in un gruppo di dati comunque distribuito, la percentuale (P) di osservazioni comprese entro la distanza di  $k$  deviazioni standard ( $s$ ) intorno alla media  $m$  sono almeno**

$$P \left( k \frac{|X - m|}{s} \right) \leq \left( 1 - \frac{1}{K^2} \right) \cdot 100 \quad (\text{in\%})$$

Per **dati distribuiti in qualsiasi modo**, secondo questo teorema noto come **disuguaglianza di Tchebycheff**, nell'intervallo compreso

tra  $\pm 2s$  rispetto alla media  $m$  si ha

$$\left( 1 - \frac{1}{2^2} \right) \cdot 100 = 75$$

almeno il 75% delle osservazioni,

mentre tra  $m \pm 3s$  si trova

$$\left( 1 - \frac{1}{3^2} \right) \cdot 100 = 88,89$$

almeno l'88,89% dei dati.

e nell'intervallo  $m \pm 4s$  si trova

$$\left( 1 - \frac{1}{4^2} \right) \cdot 100 = 93,75$$

almeno il 93,75% dei dati.

Con questa relazione, non è possibile calcolare la quantità di osservazioni compreso nell'intervallo  $m \pm 1s$ .

Questo teorema, che **come principio è di notevole importanza nella statistica, in quanto giustifica perché la media e la varianza bastano nella maggior parte dei casi per descrivere la distribuzione di una variabile statistica**, fornisce una stima molto approssimata.

Nel 1946 **Cramèr** ha dimostrato che, **se la distribuzione è simmetrica e unimodale**, la stima può essere molto più accurata. La relazione tra le dimensioni dell'intervallo intorno alla media e la distribuzione di frequenza o delle probabilità **P** diviene

$$P \left( k \frac{|X - m|}{s} \right) \leq \left[ 1 - \left( \frac{4}{9} \cdot \frac{1}{k^2} \right) \right] \cdot 100 \quad (\text{in\%})$$

**La migliore è la distribuzione normale standardizzata** (presentata in precedenza), che **permette i calcoli più precisi** e viene appunto utilizzata per questa sua caratteristica, quando la distribuzione dei dati ha tale forma.

**Tabelle specifiche** forniscono le frequenze sottese alla curva per ogni valore di **z**, ottenuto con la trasformazione in normale ridotta di qualsiasi distribuzione normale, mediante la relazione

$$z = \frac{x - \mu}{\sigma}$$

Conoscendo **z**, con esse è possibile stimare le frequenze o probabilità; con il percorso inverso, è possibile anche stimare il valore di **z** partendo dalle frequenze. Per comprenderne l'uso, più di spiegazioni teoriche sono utili dimostrazioni pratiche, come quelle riportate negli esercizi seguenti.

**Alla fine del capitolo ne sono state riportate 4, anche se i testi di statistica di norma ne riportano una sola. I modi con cui i valori della distribuzione normale sono pubblicati sono numerosi, ma tutti forniscono le stesse informazioni, permettono di derivare con facilità l'una dalle altre, servono per i medesimi scopi.**

In tutte quattro le tabelle riportate, il valore di **z** è fornito con la precisione di 2 cifre decimali. Nella prima colonna è riportato la quota intera con la prima cifra decimale; spesso il valore si ferma a 3,0 e quasi mai supera 4,0.

Nella prima riga è riportato il secondo decimale, che ovviamente varia da 0 a 9 e spesso è indicato con 3 cifre da 0,00 a 0,09.

Entro la tabella, all'incrocio del valore dato dalla somma della prima colonna con la prima riga, è riportata la quota di probabilità o frequenza relativa sottesa alla curva entro un intervallo prestabilito, stimato in modo differente nelle 4 tabelle.

La **prima tabella** riporta la quota dell'area in una coda della curva normale standardizzata, non importa se destra o sinistra. Il primo valore, collocato in alto a sinistra, è uguale a 0.500 e corrisponde ad un valore di **z** uguale a 0.00. Significa che la quota di probabilità sottesa alla curva normale, a destra di un valore che si discosta dalla media **m** di una quantità pari a 0,00 **s** (quindi coincidente con la media stessa), è pari a 0.500 oppure 50% dei valori, se espresso in percentuale. In altri termini, i dati con valore superiore alla media sono il 50%.

Per **z** uguale a 1.00, la probabilità è 0.1587: significa che, considerando solo una coda della distribuzione, il 15,87 % dei dati ha un valore che si discosta dalla media di una quantità superiore a 1.00 volte **s**. I valori di **z** usati con frequenza maggiore sono:

- 1.645 perché delimita il 5% dei valori maggiori,

- 1.96 per la probabilità del 2,5% sempre in una coda della distribuzione,
- 2.328 per l'1%,
- 2.575 per il 5 per mille.

La **seconda tabella** riporta l'area sottostante la distribuzione normale standardizzata nell'intervallo compreso tra la media  $m$  e il valore di  $z$ . Considera le probabilità sottese in una coda della distribuzione, considerando la curva compresa tra la media e il valore che si discosta da essa di  $z$  volte  $s$ . In termini molto semplici, è la differenza a .5000 della tabella precedente: mentre la prima tabella variava da .5000 a 0, questa varia simmetricamente da 0 a .5000.

Per  $z$  uguale a 0.00 è uguale a .0000, poiché nell'intervallo tra  $z$  e  $z + 0.00s$  non è compreso alcun valore. Per  $z$  uguale a 1 è 0.3413. Considerando i valori di  $z$  più frequentemente utilizzati, si hanno le seguenti relazioni:

- 1.645 per il 45%,
- 1.96 per il 47,5%,
- 2.328 per il 49%,
- 2.575 per il 49,5%.

La **terza tabella** fornisce la probabilità di ottenere un valore dello scarto standardizzato minore di  $z$ . Poiché i valori inferiori alla media sono il 50%, parte da 0.5000 e tende a 1. Rispetto alla prima tabella, può essere visto come il complemento ad 1; rispetto alla seconda, ha aggiunto .5000 ad ogni valore.

Per  $z$  uguale a 0.00 è .5000, per  $z$  uguale a 1.0 è uguale a 0.8413. Considerando i valori di  $z$  più frequentemente utilizzati si hanno le seguenti relazioni:

- 1.645 per il 95%,
- 1.96 per il 97,5%,
- 2.328 per il 99%,
- 2.575 per il 99,5%.

La **quarta tabella** fornisce le probabilità nelle due code della distribuzione. Per ogni  $z$ , il valore riportato è il doppio di quello della prima tabella. Parte da 1.000 e tende a 0 all'aumentare di  $z$ , con le seguenti relazioni:

- 1.645 per il 10%,
- 1.96 per il 5%,
- 2.328 per il 2%,
- 2.575 per l'1%.

Con una serie di esercizi, è possibile dimostrare in modo semplice e facilmente comprensibile l'utilizzazione pratica della distribuzione normale standardizzata. I calcoli utilizzano quasi esclusivamente la prima tabella, maggiormente diffusa nei testi di statistica applicata.

ESERCIZIO 1. In una popolazione di pesci, la lunghezza media ( $\mu$ ) è di cm 35 e la deviazione standard ( $\sigma$ ) di cm 5. Quale è la probabilità di trovare:

- a) pesci di lunghezza superiore a cm 40;
- b) pesci di lunghezza inferiore a cm 40;
- c) pesci di lunghezza inferiore a cm 25;
- d) pesci di lunghezza compresa tra cm 40 e cm 50;
- e) pesci di lunghezza tra cm 30 e cm 40.

Risposte:

- a) la probabilità sottostante l'area tra la media e  $z = 1$  è 0,3413; la probabilità di trovare individui di lunghezza superiore a 40 cm è pertanto 0,1587 corrispondente a 15,87%;
- b) la probabilità di trovare valori inferiori a 40 cm è 0,8413 corrispondente a 84,13%;
- c) con  $z = 2$  l'area esclusa a sinistra della media è 0,0228 equivalente a 2,28%;
- d) determinano valori compresi tra  $z = 1$  e  $z = 3$  ai quali corrisponde, dalla media, un'area sottesa rispettivamente di 0,3413 e 0,49865; pertanto l'area sottesa dall'intervallo è data dalla loro differenza che è uguale a 0,15735 (o 15,735%);
- e) sono i valori compresi nell'intervallo  $z = -1$  e  $z = 1$ ; a sinistra e a destra della media l'area sottesa è in entrambi i casi pari a 0,3413 determinando un'area totale pari a 0,6826 (o 68,26%).

ESERCIZIO 2. In una specie di pesci adulti, maschi e femmine sono distinguibili dalle dimensioni medie: misurate in cm le femmine hanno  $\mu = 37,5$  e  $\sigma = 3,8$  mentre i maschi hanno  $\mu = 34,5$  e  $\sigma = 3,2$ .

Determinare.

- a) Sono più rari i maschi di dimensioni superiori a 40 cm o le femmine superiori a 41 cm?
- b) Quale è la lunghezza minima del 5% delle femmine con dimensioni maggiori?
- c) Quale è la lunghezza massima del 5% dei maschi con dimensioni minori?
- d) Quale percentuale di maschi avranno le stesse dimensioni del 30% delle femmine di dimensioni maggiori?
- e) Quale percentuale di maschi avranno le stesse dimensioni del 20% delle femmine di dimensioni minori?

Risposte:

- a) Per i maschi di 40 cm  $z = 1,72$  che esclude a destra un'area equivalente a 4,26 %; per le femmine di 41 cm  $z = 0,92$  che esclude a destra un'area equivalente a 17,88%. Di conseguenza, sono molto più rari i pesci maschi con dimensioni uguali o superiori a 40 cm delle femmine con dimensioni uguali o superiori a 41 centimetri.
- b) Il 5% delle femmine di dimensioni maggiori sono alla destra della media di una quantità minima pari a  $1,645\sigma$ ; in centimetri è equivalente a  $1,645 \times 3,8 = 6,251$ . La lunghezza minima del 5% delle femmine di dimensioni maggiori è pertanto  $cm\ 37,5 + 6,251 = 43,751$ .
- c) Il 5% dei maschi di dimensioni minori sono alla sinistra della media di una quantità pari a  $1,645\sigma$ ; in centimetri è equivalente a  $1,645 \times 3,2 = 5,264$ . La lunghezza massima del 5% dei maschi di dimensioni minori è pertanto  $cm\ 34,5 - 5,264 = 29,236$ .
- e) Il valore di  $z$  che esclude il 30% della popolazione è 0,525. A destra della media delle femmine corrisponde a  $\mu + 0,525\sigma$ ; tradotti in centimetri sono uguali a  $37,5 + 0,525 \times 3,8 = 39,495$ . Ai maschi con tali dimensioni minime corrisponde un valore di  $z = (39,495 - 34,5) / 3,2 = 1,56$  pari a una frequenza di probabilità uguale a 5,94%.
- e) Il valore di  $z$  che esclude il 20% della popolazione è 0,842. A sinistra della media delle femmine corrisponde a  $\mu - 0,842\sigma$ ; tradotti in centimetri sono uguali a  $37,5 - 0,842 \times 3,8 = 34,3004$ . Ai maschi con tali dimensioni massime corrisponde un valore di  $z = (34,3004 - 34,5) / 3,2 = -0,0623$  pari ad una frequenza di probabilità uguale a 47,5%

#### 2.4.4 APPROSSIMAZIONI E CORREZIONI PER LA CONTINUITA'

Molte distribuzioni discrete, quali **la binomiale, l'ipergeometrica e la normale, sono bene approssimate dalla distribuzione normale, per campioni sufficientemente grandi**. L'uso della normale è giustificata anche dalla impossibilità pratica di effettuare i calcoli con le formule delle distribuzioni discrete, a causa di elevamenti a potenze alte e del calcolo di fattoriali per numeri grandi. Nonostante la corrispondenza dei risultati, tra distribuzioni discrete e distribuzioni continue esistono differenze nel calcolo delle probabilità. Le prime forniscono le probabilità per **singoli valori** della variabile casuale, cioè la probabilità di ottenere **esattamente**  $i$  volte un determinato evento; le seconde forniscono la probabilità cumulata da un certo valore fino all'estremo della distribuzione. Per calcolare la probabilità di un singolo valore, con la distribuzione normale si deve calcolare l'area sottesa all'intervallo  $X \pm 0,5$ . In altri termini, per individuare un valore discreto  $i$  in una scala continua, occorre prendere non il valore esatto  $X$  ma l'intervallo unitario  $X \pm 0,5$ .

ESEMPIO In una popolazione planctonica, la specie A ha una presenza del 10%; in un campionamento casuale di 120 individui quale è la probabilità di:

- a) trovarne **esattamente** 15 della specie A?
- b) trovarne **almeno** 15 della specie A?

c) trovarne **meno** di 15 della specie A?

Risposte: (ricordando che  $n = 120$ ;  $x = 15$ ;  $\mu = n \cdot p = 120 \times 0,10 = 12$ ;

$$\sigma^2 = n \cdot p \cdot q = 120 \times 0,10 \times 0,90 = 10,8; \quad \sigma = 3,29)$$

a) Per stimare la probabilità di avere esattamente 15, che è un valore discreto, in una scala continua si deve calcolare la probabilità compresa nell'intervallo tra 14,5 e 15,5. Poiché la tabella della distribuzione normale fornisce la probabilità cumulata da quel punto verso l'estremo, il calcolo richiede 3 passaggi:

1- la stima della probabilità da 15,5 verso destra

$$z_1 = \frac{(x + 0,5) - \mu}{\sigma} = \frac{(15 + 0,5) - 12}{\sqrt{10,8}} = \frac{3,5}{3,29} = 1,06$$

corrispondente a  $\mu + 1,06\sigma$  ed equivalente al **35,54%** delle osservazioni;

2- la stima della probabilità da 14,5 verso l'estremo nella stessa direzione precedente:

$$z_2 = \frac{(x - 0,5) - \mu}{\sigma} = \frac{(15 - 0,5) - 12}{\sqrt{10,8}} = \frac{2,5}{3,29} = 0,76$$

corrispondente a  $\mu + 0,76\sigma$  ed equivalente al **27,64%** delle osservazioni;

3 - la sottrazione della prima probabilità dalla seconda, per cui la probabilità di trovare **esattamente** 15 individui della specie A è

$$35,54 - 27,64 = 7,90\%$$

**uguale a 7,9 %.**

Con la distribuzione binomiale, per risolvere lo stesso esercizio il calcolo della probabilità avrebbe dovuto essere

$$P_{(15)} = C_{120}^{15} (0,10)^{15} (0,90)^{105}$$

che rappresenta un'operazione difficile da risolvere manualmente.

b) La probabilità di trovare almeno 15 individui della specie A potrebbe teoricamente essere calcolata con la distribuzione binomiale, sommando le singole probabilità esatte di trovare 15, 16, 17, ecc. fino a 120 individui. Alla precedente difficoltà di calcolo, si aggiunge quella del tempo richiesto per stimare queste 106 probabilità

$$P(X, n) = \sum_{x=0}^n C_n^x p^x q^{n-x}$$

da sommare per ottenere quella complessiva.

Con l'utilizzazione della distribuzione normale il calcolo diviene molto più semplice e rapido:

$$z = \frac{(x - 0,5) - \mu}{\sigma} = \frac{14,5 - 12}{\sqrt{10,8}} = \frac{2,5}{3,29} = 0,76$$

Si stima un valore della distribuzione normale standardizzata con z uguale a 0,76 equivalente ad una probabilità totale verso l'estremo uguale a 22,36%. Pertanto, la probabilità di trovare **almeno** 15 individui (15 o più individui) è uguale a 22,36%.

c) La probabilità di trovare **meno** di 15 individui, con la distribuzione binomiale è data dalla somma delle probabilità esatte di trovare da 0 a 14 individui:

$$C_{120}^0 (0,1)^0 (0,9)^{120} + C_{120}^1 (0,1)^1 (0,9)^{119} + \dots + C_{120}^{15} (0,1)^{15} (0,9)^{105}$$

Con la distribuzione normale il calcolo diviene:

$$z = \frac{(x - 0,5) - \mu}{\sigma} = \frac{14,5 - 12}{\sqrt{10,8}} = \frac{2,5}{3,29} = 0,76$$

Nella tavola della distribuzione normale standardizzata a  $z = 0,76$  corrisponde una probabilità, a partire dalla media, uguale a 27,64%. Ad essa va sommato 50 %, che corrisponde alla probabilità di trovare da 0 a 12 individui (la media). Pertanto, la probabilità complessiva di trovare **meno** di 15 individui è

$$50\% + 27,64\% = 77,64\%$$

uguale al 77,64%.

Allo stesso modo della distribuzione binomiale, è possibile **approssimare la distribuzione ipergeometrica** con la formula

$$z = \frac{(X - n \cdot p) \pm 0,5}{\sqrt{n \cdot p \cdot q} \cdot \sqrt{\frac{N - n}{N - 1}}}$$

dove:

$N$  = numero totale di individui del campione,

$n$  = numero di individui estratti dal campione.

Per la **distribuzione di Poisson** il calcolo del valore di  $z$  diviene

$$z = \frac{(X - \mu) \pm 0,5}{\sqrt{\mu}}$$

dove

$m$  è il valore della media che, come dimostrato, coincide con quello della varianza  $s^2$ .

#### 2.4.5 DISTRIBUZIONE RETTANGOLARE

Come nelle distribuzioni discrete, anche tra le distribuzioni continue la più semplice è la distribuzione rettangolare, detta anche **distribuzione uniforme continua**.

La **distribuzione rettangolare continua**, compresa nell'intervallo tra  $x_1 = \alpha$  e  $x_2 = \beta$ , come densità di frequenze relative ha la funzione

$$f(x) = \frac{1}{\beta - \alpha} \quad \text{con } (\alpha < x < \beta)$$

ed è pertanto caratterizzata da una densità costante in tutto l'intervallo da  $\alpha$  a  $\beta$ .

Nella rappresentazione grafica questa distribuzione ha la forma di un rettangolo, figura che giustifica il suo nome.

La media è

$$\mu = \frac{\alpha + \beta}{2}$$

e la varianza

$$\sigma^2 = \frac{(\beta - \alpha)^2}{12}$$

Ovviamente questa distribuzione è l'equivalente dell'uniforme discreta, considerata nel continuo.

#### 2.4.6 DISTRIBUZIONE ESPONENZIALE NEGATIVA

La **esponenziale negativa è una distribuzione continua** con funzione

$$f(x) = \alpha e^{-\alpha x} \quad \text{con } (\alpha > 0, x > 0)$$

che prende il nome dall'esponente negativo che compare nella relazione. E' una funzione positiva o nulla continuamente decrescente, che tende a 0 per x che tende all'infinito. Nel discreto ha l'equivalente nella distribuzione geometrica decrescente.

Media e varianza sono rispettivamente:

media  $\mu$

$$\mu = \frac{1}{\alpha}$$

varianza  $\sigma^2$

$$\sigma^2 = \frac{1}{\alpha^2} = \mu^2$$

E' di estremo interesse pratico, **per dedurre la curva sottostante una determinata distribuzione di dati sperimentali**, notare che in questa distribuzione **la varianza è uguale al quadrato della media**.

#### 2.4.7 LE CURVE DI PEARSON

Karl Pearson ha proposto **non una curva sola ma una famiglia di curve**, utili per descrivere con elevata approssimazione molte distribuzioni empiriche, modificando i suoi parametri. Nonostante gli ottimi risultati ottenuti nell'approssimazione, ha il grave limite che i parametri che la definiscono non sono esplicativi del fenomeno e quindi si prestano male ad usi predittivi.

La forma esplicita della funzione può essere espressa come

$$\frac{dy}{dx} = \frac{y(x+c)}{b_0 + b_1x + b_2x^2}$$

che dipende dalle radici dell'espressione quadratica del denominatore, cioè dai valori dei parametri  $b_0$ ,  $b_1$  e  $b_2$ , e dove  $y$  e  $x$  sono i valori degli assi e  $c$  è una costante.

Il sistema gode della proprietà di rappresentare molte curve, come quelle di seguito disegnate.

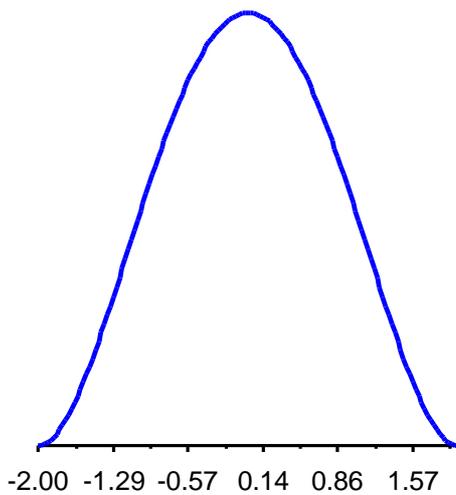


Fig. 17. Curva di Pearson : forma a campana simmetrica

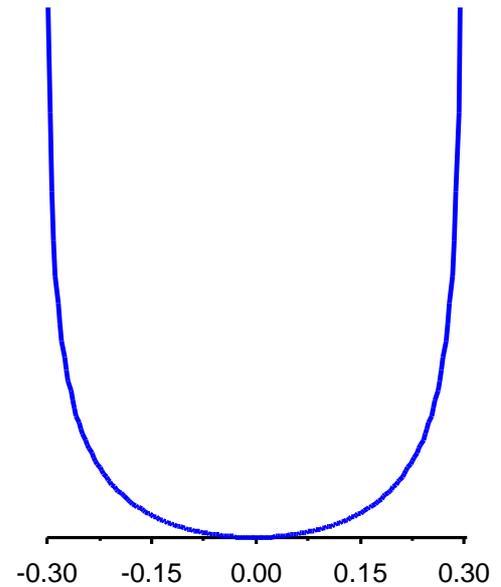


Fig. 18. Curva di Pearson: forma a U simmetrica

Variando i parametri prima definiti, si passa dall'una all'altra forma di curva.

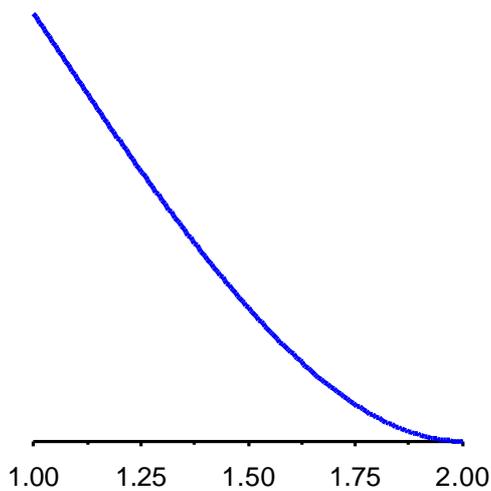


Figura 19. Curva di Pearson : forma a J rovesciato.

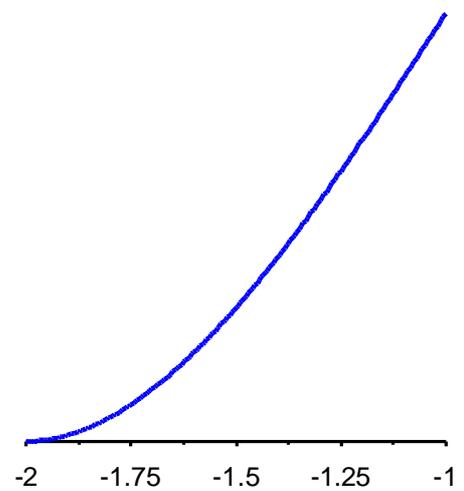


Figura 20. Curva di Pearson : forma a J .

Anche la distribuzione normale e le sue approssimazioni, con curtosi ed asimmetria variabili, possono essere rappresentate come una delle possibili curve di Pearson.

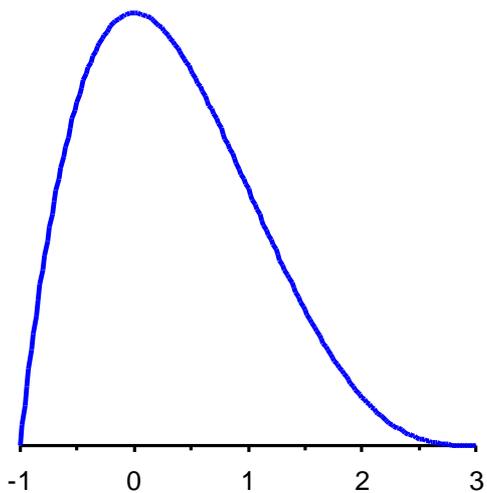


Fig. 21 Curva di Pearson a campana asimmetrica

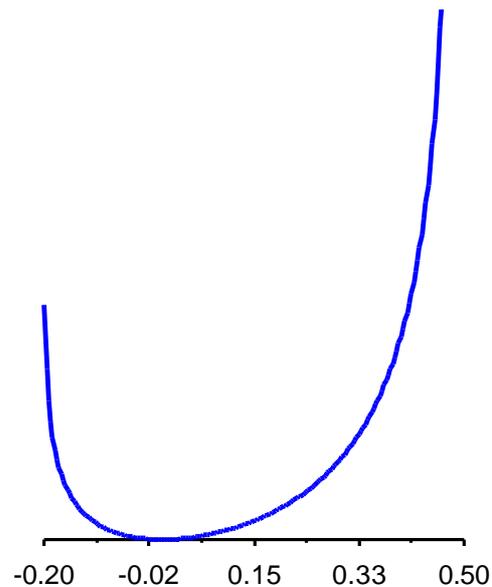


Fig. 22 Curva di Pearson: forma a U asimmetrica.

Lo studio dettagliato delle diverse curve richiederebbe una trattazione complessa, che non viene affrontata in modo più dettagliato in questo corso, per le sue ridotte applicazioni nelle scienze ambientali ed ecologiche.

## 2.4.8 LA DISTRIBUZIONE GAMMA

Un altro modello per descrivere la distribuzione di variabili casuali continue e positive, come altezza o lunghezza, peso, tempo, densità e concentrazione, è la distribuzione Gamma ( $\Gamma$ ). La sua funzione di densità di probabilità è

$$f(x) = K \cdot (x^{\nu-1} / \mu^{\nu}) \exp(-vx / \mu)$$

per  $x > 0$

e dove sia  $m$  che  $n$  sono maggiori di 0,

mentre  $K$  è una costante che rende unitaria l'area sottesa dalla curva.

Quando  $x \leq 0$  la funzione è uguale a 0.

I parametri che determinano la funzione di densità della curva  $\Gamma$  ( indicata sui testi anche G da Gamma)

sono la media  $m$  ed  $n$  (chiamato indice della distribuzione),

mentre la costante  $K$  è data da

$$K = \nu^{\nu} / \Gamma(\nu), \quad \text{con} \quad \Gamma(\nu) = \int_0^{\infty} x^{\nu-1} e^{-x} dx$$

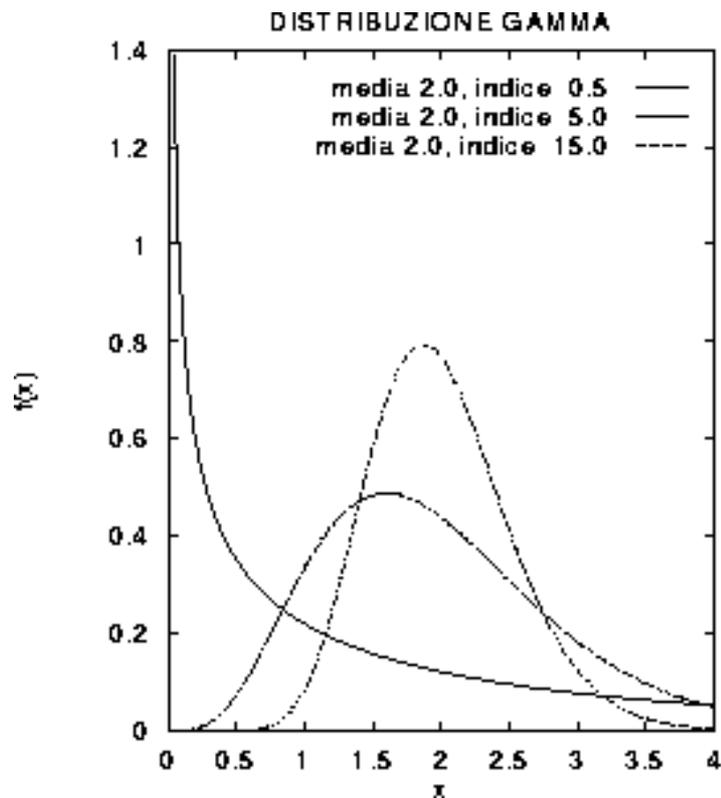


Figura 22. Alcune distribuzioni Gamma

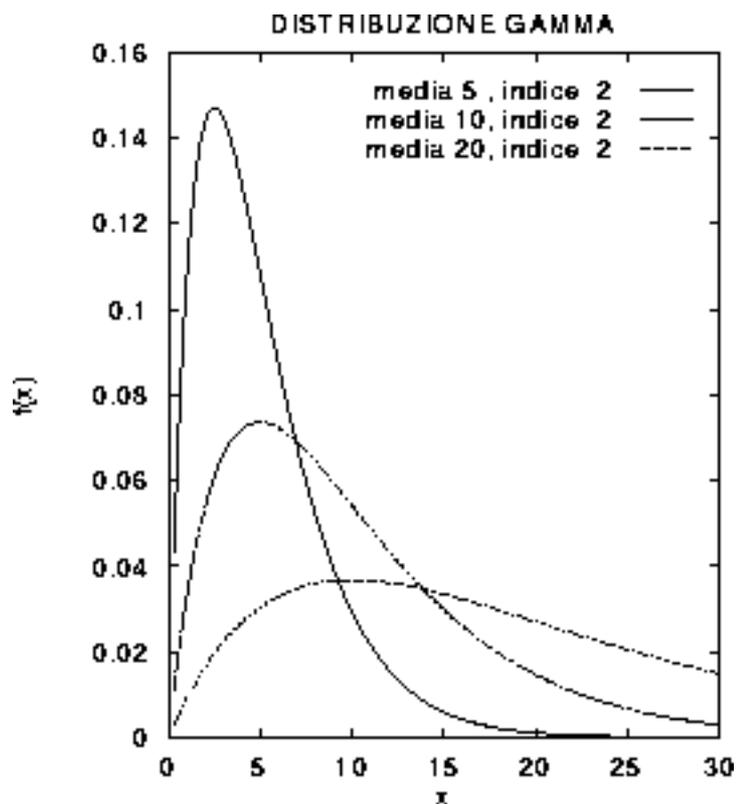


Figura 23. Altre forme della distribuzione Gamma.

Per il calcolo di  $\Gamma(v)$  sono disponibili tavole apposite.

Per  $n$  intero positivo, è possibile calcolare  $\Gamma(v)$  mediante

$$\Gamma(v) = (v - 1) !$$

per i valori interi e

$$\Gamma(v + \frac{1}{2}) = \sqrt{\pi} \{(1 \cdot 3 \cdot 5 \cdot 7 \cdot \dots \cdot (2v-1)) / 2^v\}$$

per valori interi + 0,5.

I casi particolari più importanti della distribuzione Gamma sono la distribuzione Esponenziale e la distribuzione Chi-quadrato ( $\chi^2$ ).

La funzione di densità di probabilità della **distribuzione esponenziale** è

$$f(x) = (1 / \mu) \exp(-x / \mu) \quad \text{per } x > 0$$

e dove  $\mu > 0$ .

La esponenziale è utile per descrivere la distribuzione delle durate di intervalli di tempo, in cui si manifesta un fenomeno biologico od ambientale, calcolando la distribuzione degli intervalli di tempo tra il manifestarsi di due eventi successivi.

E' in relazione con la distribuzione di Poisson, che fornisce la distribuzione degli eventi in un determinato intervallo di tempo. Se il numero di eventi  $i$ , che avvengono in un determinato intervallo di tempo  $t$ , segue la legge di Poisson con media  $l$ , il tempo di attesa  $X$  intercorrente tra due eventi segue la legge esponenziale con parametro  $m = 1 / l$ . Il tempo medio di attesa  $\mu$  tra due eventi è costante ed è pari al reciproco della media utilizzata nella distribuzione binomiale (in epidemiologia,

chiamato tasso d'incidenza ed uguale al numero di nuovi eventi contati in un periodo unitario, come ora, giorno, mese, anno).

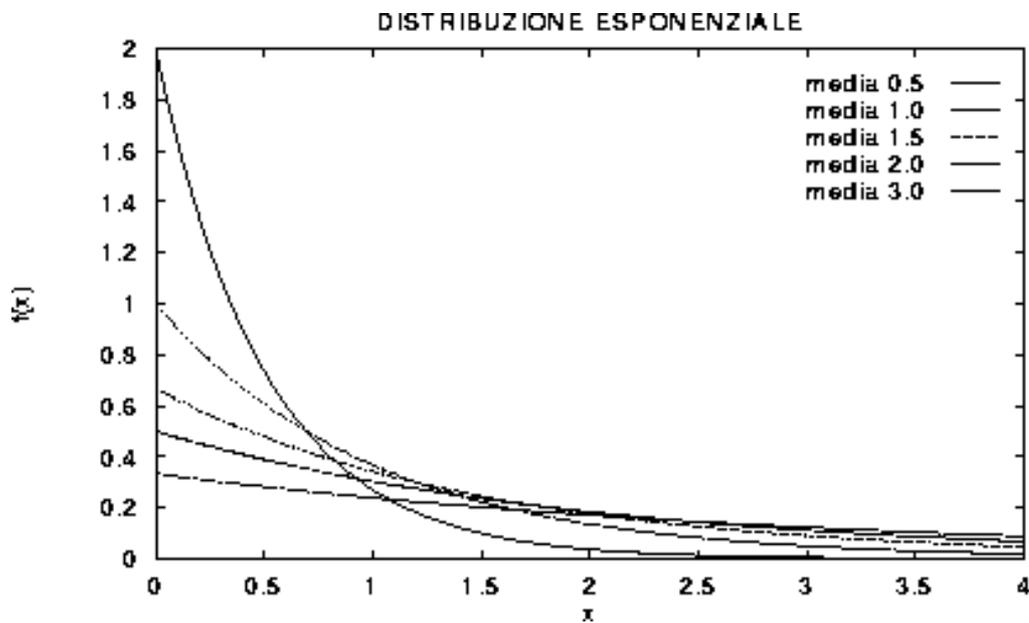


Figura 24. Distribuzione esponenziale (negativa).

## 2.5. DISTRIBUZIONI CAMPIONARIE DERIVATE DALLA NORMALE ED UTILI PER L'INFERENZA: $\chi^2$ DI PEARSON, $t$ DI STUDENT E $F$ DI FISHER

La distribuzione normale è valida per campioni molto numerosi, teoricamente infiniti. Spesso è possibile disporre nella statistica economica e sociale, in cui si analizzano i dati personali di una regione o una nazione. Nella pratica della ricerca statistica biologica, naturalistica ed ambientale, per l'inferenza sono disponibili alcune decine, al massimo poche centinaia di osservazioni. In molti settori della ricerca applicata, **molto spesso i campioni hanno dimensioni ancora minori e la loro numerosità gioca un ruolo importante, nel determinare la forma della distribuzione.** Essa non può più essere considerata normale od approssimativamente tale, ma se ne discosta quanto più il campione è piccolo.

**Per l'inferenza, nella statistica parametrica l'ipotesi fondamentale è che questi campioni siano estratti da una popolazione normalmente distribuita.** E' un'ipotesi limitativa, ma basilare per le distribuzioni **t di Student** e **F di Fisher**, che insieme rappresentano le distribuzioni fondamentali dell'inferenza statistica parametrica. E' importante comprendere come le 3 distribuzioni più utilizzate nell'inferenza statistica, la distribuzione  $\chi^2$  di Pearson in quella non parametrica, la distribuzione **t di Student** e la **F di Fisher**, per la statistica parametrica, **siano legate logicamente e matematicamente con la distribuzione normale e tra loro.**

### 2.5.1 LA DISTRIBUZIONE $c^2$

La distribuzione **Chi-quadrato** ( $c^2$ ), il cui uso è stato introdotto dallo statistico inglese Karl Pearson (1857–1936), può essere fatta derivare dalla distribuzione normale. **Date n variabili casuali indipendenti  $x_1, x_2, \dots, x_n$ , normalmente distribuite con  $m = 0$  e  $s = 1$ , il  $c^2$  è una variabile casuale data dalla somma dei loro quadrati.**

La funzione di densità di probabilità della distribuzione  $\chi^2$  è

$$f(x) = K \cdot x^{(v/2)-1} \exp(-x/2)$$

dove  $n = 1, 2, \dots$  e  $K = 2^{-v/2} / \Gamma(v/2)$ .

**La funzione di densità del  $c^2$  è determinata solo dal parametro n, il numero di gradi di libertà, pertanto viene scritta come  $C^2_{(n)}$ .**

La distribuzione  $\chi^2$  **parte da n uguale a 1 e al suo aumentare assume forme sempre diverse, fino ad una forma approssimativamente normale per  $n = 30$ . Con n molto grande (oltre 200, per**

$$\sqrt{2c^2}$$

alcuni autori di testi) è possibile dimostrare che si ottiene una nuova variabile casuale

- normalmente distribuita,
- con media m uguale a 0 e

$$\sqrt{2n-1}$$

- deviazione standard s uguale a 1.

La distribuzione chi quadrato e le sue relazioni con la normale possono essere spiegate in modo semplice attraverso alcuni passaggi.

Supponendo di avere una popolazione di valori X, distribuita in modo normale,

$$z = \sqrt{2c^2} - \sqrt{2n-1}$$

la media m di questa distribuzione è

$$E(X) = \mu$$

e la varianza  $s^2$  è

$$E(X - \mu)^2 = \sigma^2$$

Se da questa distribuzione si estrae un valore  $x$  alla volta, per ogni singolo valore estratto si può stimare un **punteggio  $z^2$  standardizzato** attraverso la relazione

$$z^2 = \frac{(x - m)^2}{s^2}$$

Questo valore al quadrato, che a differenza della  $z$  può essere solo positivo e variare da 0 all'infinito,

$$c_{(1)}^2 = z^2$$

coincide con il chi quadrato con un grado di libertà.

Nella distribuzione  $z$ , il 68% dei valori è compreso nell'intervallo tra  $-1$  e  $+1$ ; di conseguenza il chi quadrato con 1 gdl calcolato con

$$c_{(1)}^2 = \frac{(x - m)^2}{s^2}$$

ha una quantità equivalente di valori (il 68% appunto) tra 0 e 1.

Analizzando **non un caso solo ma due casi**, con la formula

$$z_1^2 = \frac{(x_1 - m)^2}{s^2} \quad \text{e} \quad z_2^2 = \frac{(x_2 - m)^2}{s^2}$$

si calcola un chi quadrato con 2 gradi di libertà

$$c_{(2)}^2 = \frac{(x_1 - m)^2}{s^2} + \frac{(x_2 - m)^2}{s^2} = z_1^2 + z_2^2$$

fondato su due osservazioni indipendenti, che avrà una forma meno asimmetrica del caso precedente e una quantità minore di valori compresi tra 0 e 1.

Con  $n$  osservazioni  $x_i$  indipendenti, estratte casualmente da una popolazione normale con media  $m$  e varianza  $s^2$ , si stima una variabile casuale chi quadrato

$$c_{(n)}^2 = \frac{\sum_{i=1}^n (x_i - m)^2}{s^2} = \sum_{i=1}^n z_i^2$$

con  $n$  gradi di libertà e uguale alla somma degli  $n$  valori  $z^2$ .

La v. c.  $c^2$  **gode della proprietà additiva**: se due o più chi-quadrato, ognuno con i propri gdl sono indipendenti, dalla loro somma si ottiene un nuovo chi-quadrato con gdl uguale alla somma dei gdl.

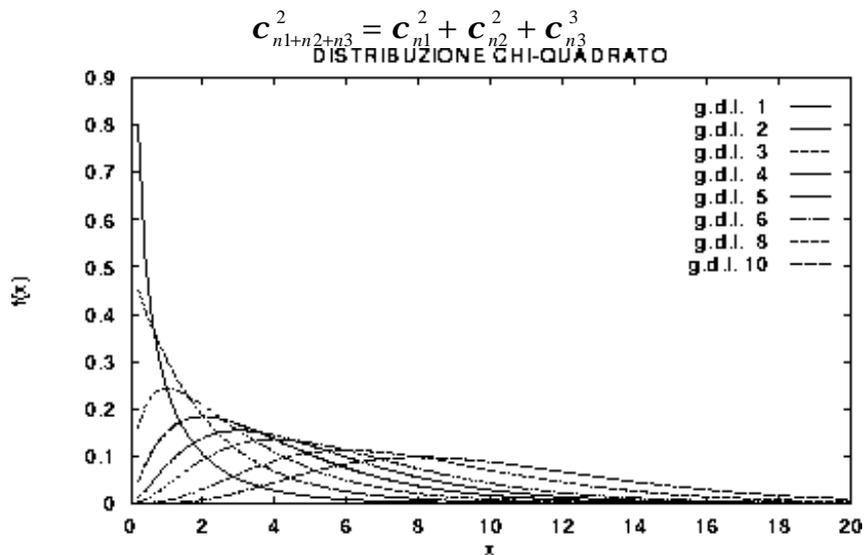


Figura 25. Alcune distribuzioni  $\chi^2_{(v)}$ , con gdl che variano da 1 a 10

Anche la **varianza campionaria  $s^2$  ha una distribuzione chi quadrato**, come verrà in seguito approfondito. Il  $c^2$  può servire per valutare se la varianza  $S^2$  di una popolazione, dalla quale sia stato estratto un campione con varianza  $s^2$ , sia uguale o diversa da un valore predeterminato  $S_0^2$ . Questi concetti sono espressi nell'ipotesi nulla  $H_0$

$$H_0: S^2 = S_0^2$$

con ipotesi alternativa  $H_1$

$$H_1 = S^2 \neq S_0^2$$

Per decidere alla probabilità  $\alpha$  tra le due ipotesi, si stima un valore del chi quadrato

$$c_{(n-1)}^2 = \frac{(n-1) \cdot s^2}{S_0^2}$$

determinato dal rapporto tra il prodotto degli **n-1** gradi di libertà con il valore sperimentale  $s^2$  e la varianza  $S_0^2$  attesa o predeterminata.

Per ogni grado di libertà, si dovrebbe avere una tabella dei valori del  $\chi^2_{(v)}$ , come già visto per la distribuzione normale. Per evitare di stampare tante pagine quante sono i gradi di libertà, di norma viene utilizzata una **tabola sinottica**, una pagina riassuntiva, che per ogni grado di libertà riporta solo i valori critici più importanti corrispondenti alla probabilità  $\alpha$  del 5% (0.05), 1% (0.01), 5 per mille (0.005) e 1 per mille (0.001).

**All'uso del  $c^2_{(n)}$  è dedicato il successivo capitolo 3.**

Non disponendo delle tabelle relative al chi quadrato e alla distribuzione normale, è possibile passare dall'una all'altra.

Per passare dai valori del  $c^2$  al valore  $z$ , ricordando che, con  $n$  grande, la distribuzione  $\chi^2_{(v)}$  è approssimativamente normale, è possibile ricorrere alla relazione

$$z_a = \frac{c^2_{(n,a)} - n}{\sqrt{2n}}$$

poiché quando i gradi di libertà sono molto più di 100

la media  $m$  della distribuzione  $c^2_{(n)}$  è uguale a  $n$  e

e la varianza  $s^2$  è uguale a  $2n$ .

Per esempio, si abbia con  $n = 100$ , alla probabilità  $a = 0.05$  il valore di  $c^2 = 124,342$ ; mediante la relazione

$$\frac{124,342 - 100}{\sqrt{2 \cdot 100}} = \frac{24,342}{14,142} = 1,72$$

si ottiene un valore di  $z$  uguale a 1,72 mentre il valore corretto è 1,6449. L'errore relativo è del 4,5%.

Inversamente, dal valore di  $z$  è possibile ricavare quello del  $c^2_{(n)}$  alla stessa probabilità  $a$ . Quando  $n$  è grande, maggiore di 100, per stimare il valore del chi quadrato che esclude una quota  $a$  di valori in una coda della distribuzione si ricorre alla relazione

$$c^2_{(n,a)} = \frac{1}{2} \cdot [z_a + \sqrt{2n-1}]^2$$

in cui  $z_a$  è il valore di  $z$  alla probabilità  $a$  prefissata.

Per esempio, con  $n = 100$ , alla probabilità  $a = 0.05$  con il valore di  $z = 1,6449$  mediante la relazione

$$c_{(100,0.05)}^2 = \frac{1}{2} \left[ 1,6449 + \sqrt{2 \cdot 100 - 1} \right]^2 = \frac{1}{2} (1,6449 + 14,1467)^2 = \frac{1}{2} \cdot 15,7516^2 = 124,056$$

si calcola un valore di  $c_{(100,0.05)}^2$  uguale a 124,056 mentre il valore corretto alla terza cifra decimale, riportato nelle tabelle, è 124,342. Il processo inverso permette una stima migliore.

Una approssimazione ancora migliore, che fornisce una stima accurata anche con pochi gradi di libertà ( $n$ ), è stata proposta da **Wilson** e **Hilferty** nel 1931 con la relazione

$$c_{(n,a)}^2 = n \cdot \left[ z_a \sqrt{\frac{2}{9n}} + 1 - \frac{2}{9n} \right]^3$$

Per esempio, con  $n = 10$ , alla probabilità  $a = 0.05$  con il valore di  $z = 1,6449$  mediante la prima relazione, valida per  $n$  grande

$$c_{(10,0.05)}^2 = \frac{1}{2} \left[ 1,6449 + \sqrt{2 \cdot 10 - 1} \right]^2 = \frac{1}{2} (1,6449 + 4,3559)^2 = \frac{1}{2} \cdot 6,0008^2 = 18,0048$$

si trova un valore di  $c_{(10,0.05)}^2$  uguale a 18,0048

mentre con la seconda relazione

$$c_{(10,0.05)}^2 = 10 \cdot \left[ 1,6449 \cdot \sqrt{\frac{2}{9 \cdot 10}} + 1 - \frac{2}{9 \cdot 10} \right]^3 = 10 \cdot (1,6449 \cdot 0,1491 + 1 - 0,0222)^3 = 10 \cdot 1,2231^3$$

$$c_{(10,0.05)}^2 = 10 \times 1,8297 = 18,297$$

si trova un valore di  $c_{(10,0.05)}^2$  uguale a 18,297 che è molto più vicino al valore 18,3070 riportato nelle tabelle, appunto per  $n = 10$ , alla probabilità  $a = 0.05$ .

Nelle 2 tabelle successive, sono riportati i valori di  $z$  alle varie probabilità  $a$  per trovare il valore corrispondente del  $c^2$  per i gradi di libertà  $n$  prefissati

(la tabella del chi quadrato è riportata alla fine del terzo capitolo).

a	0.995	0.990	0.975	0.950	0.900	0.750	0.500
Z	-2,5758	-2,3263	-1,9600	-1,6449	1,2816	-0,6745	0,0000

a	0.250	0.100	0.050	0.025	0.010	0.005	0.001
Z	+0,6745	+1,2816	+1,6449	+1,9600	+2,3263	+2,5758	+3,0902

Occorre ricordare che anche la distribuzione chi quadrato è normale, quando  $n$  è molto grande. Ciò spiega, in modo semplice ed intuitivo, perché in tale situazione quando  $z$  è uguale a  $0$ , alla probabilità  $\alpha$  corrispondente al **50%**, si abbia un valore del chi quadrato uguale alla sua media  $n$ .

La tabella dei valori critici mostra che con gradi di libertà  $n = 100$ , la media (corrispondente alla probabilità  $\alpha = 0.500$ ) non è esattamente 100 ma 99,3341 a dimostrazione del fatto che non è una distribuzione perfettamente normale.

### 2.5.2 LA DISTRIBUZIONE $t$ DI STUDENT

La distribuzione  $t$  di Student (pseudonimo del chimico inglese Gosset che ne propose l'applicazione al confronto tra medie campionarie) considera le relazioni tra media e varianza, **in campioni di piccole dimensioni**, quando si utilizza la varianza del campione. La scelta tra l'uso della normale o della distribuzione  $t$  di Student nel confronto tra medie deriva appunto dalla conoscenza della varianza  $s^2$  della popolazione o dal fatto che essa sia ignota e pertanto che, in sua vece, si debba utilizzare la varianza campionaria  $s^2$ .

Se una serie di medie campionarie ( $\mathbf{x}$ ) è tratta da una distribuzione normale ridotta ( $m = 0$ ,  $s = 1$ ) e la varianza del campione è  $s^2$ , con distribuzione  $c^2$  e  $n$  gdl, è possibile derivare la v.c.  **$t$  di Student**, tramite la relazione

$$t^2 = \frac{z^2}{c^2/n}$$

dove i gdl  $n$  corrispondono a  $N - 1$ , con  $N$  uguale al numero totale di dati.

La curva corrispondente è simmetrica, **leggermente più bassa della normale e con frequenze maggiori agli estremi**, quando il numero **di gdl ( $n$ ) è molto piccolo**.

$$f(t) = f_0 \left( 1 + \frac{t^2}{n} \right)^{-\frac{n+1}{2}}$$

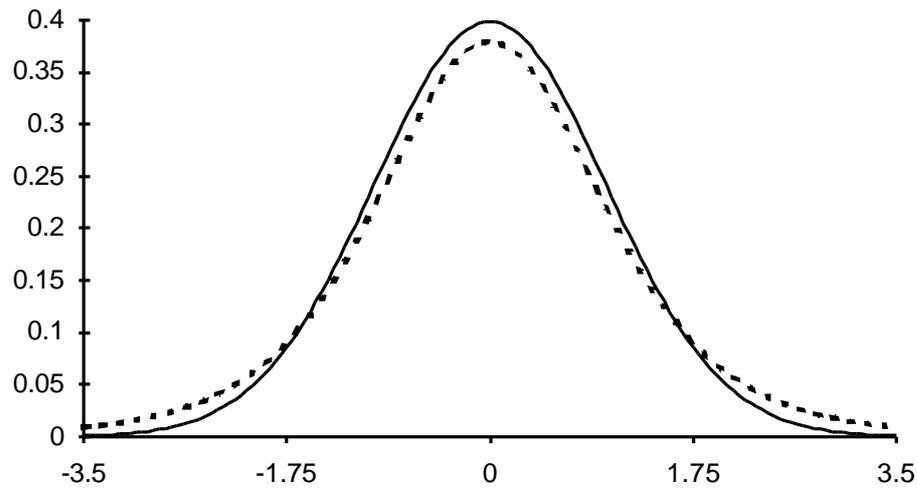


Figura 26. Confronto tra la densità di probabilità della v.c.  $t$  di Student con gdl 5 (linea tratteggiata) e la distribuzione normale corrispondente, con stessa media e stessa varianza (linea continua).

Per  $n$  che tende all'infinito, la curva tende alla normale.

### 2.5.3 LA DISTRIBUZIONE F DI FISHER

Un'altra distribuzione di notevole interesse pratico, sulla quale è fondata l'inferenza di molta parte della statistica parametrica, è la **distribuzione F**. Essa corrisponde alla distribuzione del **rapporto di 2 variabili casuali chi-quadrato indipendenti** ( $A$  e  $B$ ), **divise per i rispettivi gradi di libertà** ( $m$  e  $n$ ).

$$F = (A/m) / (B/n)$$

$$f(F) = f_0 \left( n_2 F^{\frac{n_1}{2}-1} + n_1 F^{-\frac{n_2}{2}-1} \right)$$

Questo **rapporto F** è definito tra  $0$  e  $+\infty$ .

La curva dipende sia dal valore di  $n_1$  e  $n_2$ , tenendo conto delle probabilità  $\alpha$ ; di conseguenza, in quanto definita da tre parametri, la distribuzione dei valori di  $F$  ha tre dimensioni.

Il problema della rappresentazione dei valori di  $F$  è risolto praticamente con 2-4 pagine sinottiche, che riportano solo i valori più utilizzati, quelli che fanno riferimento in particolare alle probabilità 0.05, 0.01 e, più raramente, 0.005 e 0.001.

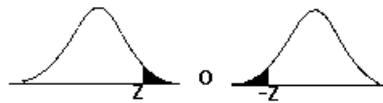
L'ordine con il quale sono riportati i due numeri che indicano i gradi di libertà è importante: la densità della distribuzione F non è simmetrica rispetto ad essi. Per convenzione, le tavole sono calcolate per avere F uguale o maggiore di 1.

Per primo si riporta sempre il numero di gradi di libertà del numeratore, che è sempre la varianza maggiore, e per secondo quello del denominatore, che è sempre la varianza minore. **Il valore di F in teoria può quindi variare da 1 a +∞. In realtà sono molto rari i casi in cui supera 10; avviene solo quando i gradi di libertà sono pochi.**

Storicamente, la distribuzione F è stata proposta dopo la distribuzione t e ne rappresenta una generalizzazione. Tra esse esistono rapporti precisi. **Il quadrato di una v.c. t di Student con n gradi di libertà è uguale ad una distribuzione F di Fisher con gradi di libertà 1 e n.**

$$t_{(n)}^2 = F_{(1,n)} \quad \text{oppure} \quad t_{(n)} = \sqrt{F_{(1,n)}}$$

## Aree in una coda della curva normale standardizzata



La tabella riporta la probabilità nell'area annerita

<b>Z</b>	<b>0,00</b>	<b>0,01</b>	<b>0,02</b>	<b>0,03</b>	<b>0,04</b>	<b>0,05</b>	<b>0,06</b>	<b>0,07</b>	<b>0,08</b>	<b>0,09</b>
<b>0,0</b>	0.500	0.496	0.492	0.488	0.484	0.480	0.476	0.472	0.468	0.464
<b>0,1</b>	0.460	0.456	0.452	0.448	0.444	0.440	0.436	0.433	0.429	0.425
<b>0,2</b>	0.421	0.417	0.413	0.409	0.405	0.401	0.397	0.394	0.390	0.386
<b>0,3</b>	0.382	0.378	0.374	0.371	0.367	0.363	0.359	0.356	0.352	0.348
<b>0,4</b>	0.345	0.341	0.337	0.334	0.330	0.326	0.323	0.319	0.316	0.312
<b>0,5</b>	0.309	0.305	0.302	0.298	0.295	0.291	0.288	0.284	0.281	0.278
<b>0,6</b>	0.274	0.271	0.268	0.264	0.261	0.258	0.255	0.251	0.248	0.245
<b>0,7</b>	0.242	0.239	0.236	0.233	0.230	0.227	0.224	0.221	0.218	0.215
<b>0,8</b>	0.212	0.209	0.206	0.203	0.200	0.198	0.195	0.192	0.189	0.187
<b>0,9</b>	0.184	0.181	0.179	0.176	0.174	0.171	0.169	0.166	0.164	0.161
<b>1,0</b>	0.159	0.156	0.154	0.152	0.149	0.147	0.145	0.142	0.140	0.138
<b>1,1</b>	0.136	0.133	0.131	0.129	0.127	0.125	0.123	0.121	0.119	0.117
<b>1,2</b>	0.115	0.113	0.111	0.109	0.107	0.106	0.104	0.102	0.100	0.099
<b>1,3</b>	0.097	0.095	0.093	0.092	0.090	0.089	0.087	0.085	0.084	0.082
<b>1,4</b>	0.081	0.079	0.078	0.076	0.075	0.074	0.072	0.071	0.069	0.068
<b>1,5</b>	0.067	0.066	0.064	0.063	0.062	0.061	0.059	0.058	0.057	0.056
<b>1,6</b>	0.055	0.054	0.053	0.052	0.051	0.049	0.048	0.048	0.046	0.046
<b>1,7</b>	0.045	0.044	0.043	0.042	0.041	0.040	0.039	0.038	0.037	0.037
<b>1,8</b>	0.036	0.035	0.034	0.034	0.033	0.032	0.031	0.030	0.029	0.029
<b>1,9</b>	0.029	0.028	0.027	0.027	0.026	0.026	0.025	0.024	0.024	0.023
<b>2,0</b>	0.023	0.022	0.022	0.021	0.021	0.020	0.020	0.019	0.019	0.018
<b>2,1</b>	0.018	0.017	0.017	0.017	0.016	0.016	0.015	0.015	0.015	0.014
<b>2,2</b>	0.014	0.014	0.013	0.013	0.013	0.012	0.012	0.012	0.011	0.011
<b>2,3</b>	0.011	0.010	0.010	0.010	0.010	0.009	0.009	0.009	0.009	0.008
<b>2,4</b>	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.006
<b>2,5</b>	0.006	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.005
<b>2,6</b>	0.005	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
<b>2,7</b>	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
<b>2,8</b>	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
<b>2,9</b>	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001
<b>3,0</b>	0.001									

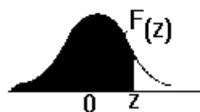
### Valori della distribuzione normale standardizzata.



La parte annerita rappresenta l'area sottostante la distribuzione normale standardizzata dalla media aritmetica a z.

<b>z</b>	<b>0,00</b>	<b>0,01</b>	<b>0,02</b>	<b>0,03</b>	<b>0,04</b>	<b>0,05</b>	<b>0,06</b>	<b>0,07</b>	<b>0,08</b>	<b>0,09</b>
<b>0,0</b>	00000	00399	00792	01197	01595	01994	02392	02790	03188	03586
<b>0,1</b>	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
<b>0,2</b>	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
<b>0,3</b>	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
<b>0,4</b>	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
<b>0,5</b>	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
<b>0,6</b>	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
<b>0,7</b>	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
<b>0,8</b>	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
<b>0,9</b>	31594	31859	32121	32381	32639	32944	33147	33398	33646	33891
<b>1,0</b>	34134	34375	34614	34849	35083	35314	35543	35769	35993	36214
<b>1,1</b>	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
<b>1,2</b>	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
<b>1,3</b>	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
<b>1,4</b>	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
<b>1,5</b>	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
<b>1,6</b>	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
<b>1,7</b>	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
<b>1,8</b>	46407	46485	46562	46637	46712	46784	46856	46926	46995	47062
<b>1,9</b>	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
<b>2,0</b>	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
<b>2,1</b>	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
<b>2,2</b>	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
<b>2,3</b>	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
<b>2,4</b>	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
<b>2,5</b>	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
<b>2,6</b>	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
<b>2,7</b>	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
<b>2,8</b>	49745	49752	49760	49767	49774	49781	49788	49795	49801	49807
<b>2,9</b>	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
<b>3,0</b>	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900
<b>3,1</b>	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929
<b>3,2</b>	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
<b>3,3</b>	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
<b>3,4</b>	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
<b>3,5</b>	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983
<b>3,6</b>	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989
<b>3,7</b>	49989	49990	49990	49990	49991	49991	49991	49992	49992	49992
<b>3,8</b>	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995
<b>3,9</b>	49995	49995	49995	49996	49996	49996	49996	49996	49997	49997

## Valori dell'integrale di probabilità della distribuzione normale standardizzata



L'area annerita rappresenta la probabilità di ottenere un valore dello scarto standardizzato minore di z.

<b>z</b>	<b>0,00</b>	<b>0,01</b>	<b>0,02</b>	<b>0,03</b>	<b>0,04</b>	<b>0,05</b>	<b>0,06</b>	<b>0,07</b>	<b>0,08</b>	<b>0,09</b>
<b>0,0</b>	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
<b>0,1</b>	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
<b>0,2</b>	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
<b>0,3</b>	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
<b>0,4</b>	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
<b>0,5</b>	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
<b>0,6</b>	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
<b>0,7</b>	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
<b>0,8</b>	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
<b>0,9</b>	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
<b>1,0</b>	0.84134	0.84375	0.84614	0.84850	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
<b>1,1</b>	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
<b>1,2</b>	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
<b>1,3</b>	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
<b>1,4</b>	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92786	0.92922	0.93056	0.93189
<b>1,5</b>	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
<b>1,6</b>	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
<b>1,7</b>	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
<b>1,8</b>	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
<b>1,9</b>	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
<b>2,0</b>	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
<b>2,1</b>	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
<b>2,2</b>	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
<b>2,3</b>	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
<b>2,4</b>	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
<b>2,5</b>	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
<b>2,6</b>	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
<b>2,7</b>	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
<b>2,8</b>	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
<b>2,9</b>	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
<b>3,0</b>	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99897	0.99900
<b>3,1</b>	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
<b>3,2</b>	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
<b>3,3</b>	0.99952	0.99953	0.99957	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
<b>3,4</b>	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976

### Area nelle due code della distribuzione normale standardizzata



La tabella riporta le probabilità nelle aree annerite.

<b>z</b>	<b>0,00</b>	<b>0,01</b>	<b>0,02</b>	<b>0,03</b>	<b>0,04</b>	<b>0,05</b>	<b>0,06</b>	<b>0,07</b>	<b>0,08</b>	<b>0,09</b>
<b>0,0</b>	1.000	0.992	0.984	0.976	0.968	0.960	0.952	0.944	0.936	0.928
<b>0,1</b>	0.920	0.912	0.904	0.897	0.889	0.881	0.873	0.865	0.857	0.849
<b>0,2</b>	0.841	0.834	0.826	0.818	0.810	0.803	0.795	0.787	0.779	0.772
<b>0,3</b>	0.764	0.757	0.749	0.741	0.734	0.726	0.719	0.711	0.704	0.697
<b>0,4</b>	0.689	0.682	0.674	0.667	0.660	0.653	0.646	0.638	0.631	0.624
<b>0,5</b>	0.617	0.610	0.603	0.596	0.589	0.582	0.575	0.569	0.562	0.555
<b>0,6</b>	0.549	0.542	0.535	0.529	0.522	0.516	0.509	0.503	0.497	0.490
<b>0,7</b>	0.484	0.478	0.472	0.465	0.459	0.453	0.447	0.441	0.435	0.430
<b>0,8</b>	0.424	0.418	0.412	0.407	0.401	0.395	0.390	0.384	0.379	0.373
<b>0,9</b>	0.368	0.363	0.358	0.352	0.347	0.342	0.337	0.332	0.327	0.322
<b>1,0</b>	0.317	0.312	0.308	0.303	0.298	0.294	0.289	0.285	0.280	0.276
<b>1,1</b>	0.271	0.267	0.263	0.258	0.254	0.250	0.246	0.242	0.238	0.234
<b>1,2</b>	0.230	0.226	0.222	0.219	0.215	0.211	0.208	0.204	0.201	0.197
<b>1,3</b>	0.194	0.190	0.187	0.184	0.180	0.177	0.174	0.171	0.168	0.165
<b>1,4</b>	0.162	0.159	0.156	0.153	0.150	0.147	0.144	0.142	0.139	0.136
<b>1,5</b>	0.134	0.131	0.129	0.126	0.124	0.121	0.119	0.116	0.114	0.112
<b>1,6</b>	0.110	0.107	0.105	0.103	0.101	0.099	0.097	0.095	0.093	0.091
<b>1,7</b>	0.089	0.087	0.085	0.084	0.082	0.080	0.078	0.077	0.075	0.073
<b>1,8</b>	0.072	0.070	0.069	0.067	0.066	0.064	0.063	0.061	0.060	0.059
<b>1,9</b>	0.057	0.056	0.055	0.054	0.052	0.051	0.050	0.049	0.048	0.047
<b>2,0</b>	0.046	0.044	0.043	0.042	0.041	0.040	0.039	0.038	0.038	0.037
<b>2,1</b>	0.036	0.035	0.034	0.033	0.032	0.032	0.031	0.030	0.029	0.029
<b>2,2</b>	0.028	0.027	0.026	0.026	0.025	0.024	0.024	0.023	0.023	0.022
<b>2,3</b>	0.021	0.021	0.020	0.020	0.019	0.019	0.018	0.018	0.017	0.017
<b>2,4</b>	0.016	0.016	0.016	0.015	0.015	0.014	0.014	0.014	0.013	0.013
<b>2,5</b>	0.012	0.012	0.012	0.011	0.011	0.011	0.010	0.010	0.010	0.010
<b>2,6</b>	0.009	0.009	0.009	0.009	0.008	0.008	0.008	0.008	0.007	0.007
<b>2,7</b>	0.007	0.007	0.007	0.006	0.006	0.006	0.006	0.006	0.005	0.005
<b>2,8</b>	0.005	0.005	0.005	0.005	0.005	0.004	0.004	0.004	0.004	0.004
<b>2,9</b>	0.004	0.004	0.004	0.003	0.003	0.003	0.003	0.003	0.003	0.003
<b>3,0</b>	0.003									

**CAPITOLO III**  
**CONFRONTI TRA TASSI E PROPORZIONI**

**3.1. CONFRONTI TRA DISTRIBUZIONI OSSERVATE E DISTRIBUZIONI ATTESE**

Nella pratica sperimentale, è frequente la necessità di verificare se esiste accordo tra una distribuzione osservata e la corrispondente distribuzione attesa o teorica. Il test viene definito **test per la bontà dell'adattamento** (in inglese, *goodness of fit test*). Sia per dati qualitativi che possono essere classificati in categorie nominali, sia per dati quantitativi distribuiti in classi di frequenza, nella ricerca ambientale è spesso necessario **saggiare la concordanza tra fatto ed ipotesi. E' lo scopo per il quale storicamente è stato costruito il test  $\chi^2$  (g. d. l.)** (chi-quadro o chi-quadrato).

E' un metodo di **inferenza statistica** che **non richiede ipotesi "a priori" sul tipo e sulle caratteristiche della distribuzione**, come invece avviene per la statistica parametrica che fa riferimento alla distribuzione normale. E' uno dei **metodi non parametrici** (detti anche "distribution free"), con i quali è **possibile stabilire se una serie di dati, raccolti in natura od in laboratorio, è in accordo con una specifica ipotesi sulla loro distribuzione o sulla loro frequenza relativa.**

Tabella 1 A. Distribuzioni fenotipiche (osservate ed attese) di *Pisum sativum* in alcuni esperimenti di Mendel per un carattere.

A			
<b>Segregazione di un ibrido:</b>			
Carattere	dominante	recessivo	Totale
<i>a) colore del fiore</i>	<i>rossi</i> 705	<i>bianchi</i> 224	929
distribuzione attesa (3:1)	696,75	232,25	929
<i>b) lunghezza del fusto</i>	<i>alte</i> 787	<i>basse</i> 277	1064
distribuzione attesa (3:1)	798	266	1064
<i>c) colore del seme</i>	<i>gialli</i> 6022	<i>verdi</i> 2001	8023
distribuzione attesa (3:1)	6017,25	2005,75	8023
<i>d) forma del seme</i>	<i>lisci</i> 5474	<i>rugosi</i> 1850	7324
distribuzione attesa (3:1)	5493	1831	7324

Il test  $\chi^2$  (g. d. l.) serve anche per il **confronto tra 2 o più distribuzioni osservate**; in queste condizioni, il suo uso più frequente è per la **verifica dell'associazione tra le varie modalità di due o**

**più caratteri qualitativi.** Risulta particolarmente utile nella fase iniziale dell'analisi, quando si ricercano le variabili più significative e le relazioni di associazione tra esse.

Per l'applicazione di questo tipo di inferenza, le distribuzioni di frequenze osservate delle classi fenotipiche e quelle attese secondo le leggi di Mendel forniscono un esempio classico.

Nella loro analisi, si pone il problema di verificare se la distribuzione della progenie degli ibridi rispetta la distribuzione teorica attesa di 3 a 1 per un solo carattere (Tabella 1A), oppure quella di 9:3:3:1 quando si seguono due caratteri (Tabella 1/B).

Tabella 1 B. Distribuzioni fenotipiche (osservate ed attese) di *Pisum sativum* in un esperimento di Mendel per due caratteri.

B		
<b>Segregazione di un diibrido per colore e forma del seme:</b>		
	Osservati	Attesi
<i>gialli-lisci</i>	<b>315</b>	9/16 = <b>312,75</b>
<i>gialli-rugosi</i>	<b>101</b>	3/16 = <b>104,25</b>
<i>verdi-lisci</i>	<b>108</b>	3/16 = <b>104,25</b>
<i>verdi-rugosi</i>	<b>32</b>	1/16 = <b>34,75</b>
<i>Totale</i>	<b>556</b>	<b>556,00</b>

E' evidente (come mostra con chiarezza la tabella 1B) che tra distribuzioni osservate e distribuzioni attese non si ha mai una perfetta coincidenza, anche quando si possono constatare valori molto simili. In tutti i casi in cui si fanno prove ripetute per verificare una legge di distribuzione, è quasi impossibile ottenere esattamente i medesimi risultati sperimentali. Tra l'altro, mentre ogni classe di una distribuzione osservata è un conteggio ed è sempre formata da numeri interi, una distribuzione attesa segue una legge teorica di distribuzione dell'ammontare totale ed è spesso formata da classi con numeri frazionari. E' ovvio, con il semplice buon senso, che **differenze piccole possono essere ritenute accidentali e quindi non sono tali da negare un sostanziale accordo tra osservato ed atteso, mentre differenze grandi lasciano supporre che non siano state ottenute per caso, ma che siano presenti fattori differenti da quelli ipotizzati.**

Il problema statistico è di poter **dedurre scientificamente ed in modo universalmente accettato se le differenze sono trascurabili** e quindi probabilmente dovute solo al caso (**ipotesi nulla**, indicata con  $H_0$ ); oppure se sono di dimensioni tali da fare più **ragionevolmente supporre una distribuzione**

**realmente diversa da quella attesa (ipotesi alternativa**, indicata con  $H_1$ ), anche se le cause sono ignote. La scelta tra le due ipotesi avviene sulla base della probabilità che lo scarto trovato e scarti ancora maggiori possano essere stati determinati dal caso.

L'importanza dell'inferenza consiste nella possibilità di **trarre conclusioni generali dal singolo esperimento**; in altri termini, nel **conoscere la probabilità con cui le differenze tra una distribuzione osservata e quella attesa possono riprodursi per caso, in una serie di esperimenti analoghi**.

Per affrontare questo problema di inferenza statistica, è possibile ricorrere al test  $\chi^2_{(g. d. l.)}$  (chi-quadrato), proposto da Pearson nel 1900, che fa ipotesi sulla distribuzione di tassi e proporzioni, ma per la stima della probabilità utilizza le **frequenze assolute**, secondo la formula

$$c^2_{(g.d.l.)} = \sum_{i=1}^n \frac{(f_i^{oss} - f_i^{att})^2}{f_i^{att}}$$

dove:

$f_i^{oss}$  = frequenza osservata i-esima

$f_i^{att}$  = frequenza attesa i-esima

**g.d.l.** = numero di gruppi meno uno

e la sommatoria S è estesa a tutti gli **n** gruppi.

**La distribuzione della densità di probabilità del  $c^2_{(g. d. l.)}$  dipende dai suoi gradi di libertà, abbreviati in g.d.l.** (in inglese, degrees of freedom, abbreviato in d.f.). **Conteggiati nel calcolo delle frequenze attese, per definizione i gradi di libertà sono il numero di classi che restano indipendenti, conoscendo il numero totale dei dati.** Nell'esempio delle classi fenotipiche, i gdl del chi quadrato sono **n-1**, dove **n** è il numero di gruppi o classi.

**Il numero di g.d.l. viene riportato tra parentesi, ai piedi del simbolo:** corrisponde al numero di osservazioni indipendenti. Infatti **i valori attesi di ogni gruppo, che sono calcolati dal totale ed attribuiti ad ogni gruppo secondo la legge di distribuzione che si vuole verificare, sono liberi di assumere qualsiasi valore, con eccezione del valore atteso nell'ultimo gruppo**, la cui frequenza è totalmente determinata dalla differenza tra la somma di tutti i gruppi precedenti, già definiti, ed il totale.

Negli esempi fino ad ora presentati, il numero di gradi di libertà corrisponde al numero di gruppi meno uno. Ma quando tra n variabili casuali sussistono k vincoli lineari, cioè relazioni che riducono il numero di osservazioni indipendenti, i gradi di libertà del corrispondente  $\chi^2$  diminuiscono di un numero pari a k.

Il numero dei gradi di libertà è determinato dai vincoli, di qualsiasi natura, che esistono fra le frequenze dei vari gruppi.

Per esempio, in genetica delle popolazioni le frequenze attese fenotipiche dei gruppi sanguigni **A**, **B**, **AB** e **O** sono calcolate dalle frequenze relative **p**, **q**, ed **r** (il cui totale è sempre uguale a 1) dei geni **I<sup>A</sup>**, **I<sup>B</sup>** ed **i**, mediante lo sviluppo di

$$(p + q + r)^2 = 1;$$

pertanto, i 4 gruppi fenotipici attesi, calcolati da 3 frequenze geniche, hanno 2 gradi di libertà.

Per la stessa legge, anche i 6 gruppi genotipici (**I<sup>A</sup>I<sup>A</sup>**, **I<sup>A</sup>i**, **I<sup>B</sup>I<sup>B</sup>**, **I<sup>B</sup>i**, **I<sup>A</sup>I<sup>B</sup>**, **ii**) hanno 2 gdl.

Secondo uno schema valido per tutti i test statistici, **il procedimento logico che deve essere seguito** nell'applicazione del  $\chi^2$  comprende diverse fasi, che possono essere riassunte in 7 passaggi:

- 1 - stabilire l'ipotesi nulla ( $H_0$ ) e l'eventuale ipotesi alternativa ( $H_1$ );
- 2 - scegliere il test più appropriato per saggiare l'ipotesi nulla  $H_0$ , secondo le finalità della ricerca e le caratteristiche statistiche dei dati (in questo caso, ovviamente, è il test chi quadrato);
- 3 - specificare il livello di significatività (i cui criteri saranno discussi nel capitolo 4), l'ampiezza del campione e i gradi di libertà;
- 4 - trovare la distribuzione di campionamento del test statistico nell'ipotesi nulla  $H_0$ , di norma fornita da tabelle;
- 5 - stabilire la zona di rifiuto (che negli esercizi di norma sarà prefissata al 5% o  $\alpha = 0.05$ );
- 6 - calcolare il valore del test statistico sulla base dei dati sperimentali, stimando il valore di probabilità ad esso associato;
- 7 - sulla base della probabilità, trarre le conclusioni: se la probabilità risulta superiore a quella prefissata, concludere che non è possibile rifiutare l'ipotesi nulla  $H_0$ ; se la probabilità risulta inferiore a quella prefissata, rifiutare l'ipotesi nulla e quindi implicitamente accettare l'ipotesi alternativa  $H_1$ .

ESEMPIO 1. Utilizzando i dati sulla segregazione mendeliana della precedente tabella 1B, il calcolo del  $\chi^2$  è semplice

$$c_{(3)}^2 = \frac{(315 - 312,75)^2}{312,75} + \frac{(101 - 104,25)^2}{104,25} + \frac{(108 - 104,25)^2}{104,25} + \frac{(32 - 34,75)^2}{34,75}$$

$$c_{(3)}^2 = \frac{(2,25)^2}{312,75} + \frac{(-3,25)^2}{104,25} + \frac{(3,75)^2}{104,25} + \frac{(-2,75)^2}{34,75} = 0,47$$

e si stima un chi quadrato con 3 gdl uguale a 0,47.

Con l'aiuto delle tavole, è possibile stimare con precisione la probabilità di trovare differenze uguali o superiori a quelle riscontrate tra distribuzione osservata e distribuzione attesa, nell'ipotesi nulla ( $H_0$ ) che le differenze siano dovute esclusivamente a fattori casuali.

Nella tavola a 2 entrate della distribuzione dei valori critici del  $\chi^2$  per 3 gradi di libertà (indicato sulla riga) e per probabilità 0.05 (indicato sulla colonna), il valore del  $\chi^2$  approssimato alla seconda cifra decimale risulta uguale a 7,81.

Il valore calcolato nell'esercizio è sensibilmente minore ( $\chi^2_{(3)} = 0,47$ ) di quello tabulato. La probabilità che le differenze siano imputabili solo al caso è alta, superiore al valore prefissato del 5%; di conseguenza, non si può rifiutare l'ipotesi nulla, secondo la quale le differenze riscontrate tra distribuzione osservata e distribuzione attesa sono dovute esclusivamente a fattori casuali. Si afferma che le differenze tra distribuzione osservata e distribuzione attesa non sono significative.

**Per la comprensione dell'inferenza statistica con il test chi quadrato, è utile ricordare che quanto più le differenze tra osservato ed atteso sono grandi, tanto più il valore del  $c^2$  sarà elevato; di conseguenza, la probabilità che tali differenze siano dovute solo al caso sarà bassa e si rifiuterà l'ipotesi nulla, accettando implicitamente l'ipotesi alternativa  $H_1$ . Al contrario, quando le differenze tra osservato ed atteso sono ridotte, ugualmente basso sarà il valore del  $c^2$ ; pertanto, sarà elevata la probabilità che esse siano imputabili esclusivamente al caso e si accetterà l'ipotesi nulla  $H_0$ .**

ESEMPIO 2 . In una popolazione di *Mixodiatomus kupelwieseri* (Copepode, Calanoide) campionate in una pozza temporanea (Lagastro - Val d'Aveto) sono state osservate le seguenti frequenze di 4 alleli del locus MPI (Mannoso fosfato isomerasi)

Tipo di allele	Freq. osservata
<b>allele 1</b>	<b>26</b>
<b>allele 2</b>	<b>38</b>
<b>allele 3</b>	<b>62</b>
<b>allele 4</b>	<b>118</b>
Totale	244

Le differenze riscontrate fra le frequenze dei vari alleli possono essere imputate al caso ( $H_0$ ) oppure è possibile pensare ragionevolmente che esistano uno o più fattori che li rendono effettivamente differenti ( $H_1$ )?

Risposta.

Se fosse vera l'ipotesi nulla espressa (i quattro gruppi hanno la stessa frequenza), la frequenza attesa per ogni allele sarebbe  $244/4 = 61$ .

Il valore del chi quadrato con 3 gradi di libertà per saggiare tale ipotesi

$$C_{(3)}^2 = \frac{(26-61)^2}{61} + \frac{(38-61)^2}{61} + \frac{(62-61)^2}{61} + \frac{(118-61)^2}{61} = \frac{1225}{61} + \frac{529}{61} + \frac{1}{61} + \frac{3249}{61} = 82,03$$

risulta uguale a 82,03.

Consultando la tabella del chi-quadrato per 3 gradi di libertà, alla probabilità 0.05 corrisponde un valore di 7,82 mentre alla probabilità  $\alpha = 0.01$  corrisponde un valore critico di 11,34 e alla probabilità  $\alpha = 0.001$  un valore critico di 16,27.

Il valore del chi quadrato calcolato sui dati sperimentali è molto più grande. La probabilità che le differenze tra le frequenze riscontrate e quelle attese secondo l'ipotesi nulla siano imputabili esclusivamente al caso è molto piccola, inferiore non solo al 5% ma addirittura al 0,1%; di conseguenza, si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa.

Il test permettere di arrivare alla conclusione che, con probabilità inferiore a 0,1% di commettere un errore, si può sostenere che i 4 alleli hanno frequenze tra loro molto differenti.

### 3.2. CONDIZIONI DI VALIDITA' DEL $\chi^2$ E CORREZIONE DI YATES

Fissata la probabilità, **il valore critico del chi quadrato è totalmente determinato dai suoi gradi di libertà** e quindi dal numero di gruppi. **Il numero di osservazioni, in totale ed entro ogni gruppo, sono a questo riguardo ininfluenti.** Eppure è logico pensare che il risultato sia tanto più attendibile quanto più elevato è il numero di osservazioni o di casi utilizzati nell'esperimento.

**Nel test  $C_{(g, d. l.)}^2$  il numero di osservazioni, sia in totale che entro ogni classe, è considerato la condizione essenziale di validità.**

**Il  $\chi^2$  è valido solamente quando è applicato a grandi campioni.**

Definito il principio, sotto l'aspetto pratico esiste scarsa concordanza su quando un campione possa essere universalmente ritenuto di grandi dimensioni. La maggioranza dei testi di statistica pone limiti, seppure non delimitati chiaramente, con i quali si possono formare 3 classi.

1- **Il test è valido quando il numero totale di osservazioni è superiore a 100;** più recentemente, tuttavia si pretende un numero maggiore, **circa 200;** altri addirittura richiedono 500 osservazioni.

2 - Il test è meno attendibile ed ha bisogno di **una correzione, detta correzione di Yates,** quando il numero di osservazioni è **tra 200 (o 100) e 30 circa.**

3 - Il test **perde ogni attendibilità quando il numero di osservazioni è inferiore a 30**; diversi autori pongono questo limite a 40, mentre altri tendono ad abbassarlo fino a 25-20 osservazioni.

Questo primo criterio di suddivisione definisce una condizione necessaria ma non sufficiente: **occorre considerare anche il numero di osservazioni entro ogni classe.**

Una condizione aggiuntiva di validità del test  $\chi^2$  è che ogni gruppo o classe abbia, **per le frequenze attese, un numero minimo di 5 osservazioni.**

Anche su questo secondo criterio non esiste unanimità, come normalmente avviene per soluzioni approssimate. Alcuni testi abbassano questo limite fino a 4 osservazioni, molto raramente fino a 3; altri autori ritengono ancora accettabile il risultato, quando non più di 2 -3 gruppi abbiano meno di 5 osservazioni attese. Tuttavia, la maggioranza degli studiosi definisce queste ultime frequenze come insufficienti.

**Quando ha un numero abbastanza alto di gradi di libertà, il chi quadrato è meno sensibile agli errori determinati da piccole frequenze attese; in questi casi, si accettano frequenze attese minime di 2-3.**

Con questi **doppi limiti, che interessano sia il numero totale di osservazioni sia la loro frequenza entro ogni gruppo,** anche il numero di suddivisioni che è possibile considerare è ovviamente limitato e dipende dal numero totale di dati. Per esempio, con un totale di 50 osservazioni non è possibile avere più di 10 gruppi, perché in una distribuzione uniforme la media sarebbe di 5 e, in una distribuzione non uniforme, in alcuni gruppi le frequenze attese sarebbero necessariamente inferiori.

**Per campioni di dimensioni inferiori a 200 (100) ma superiori a 30, si deve apportare la correzione di Yates o correzione per la continuità.**

Essa consiste in una operazione molto semplice: **si sottrae 0,5 allo scarto** (tra frequenza osservata e frequenza attesa) **maggiore in valore assoluto e si aggiunge 0,5 allo scarto minore in valore assoluto.** Con due sole classi, si abbassa di 0,5 in valore assoluto sia lo scarto negativo che quello positivo.

ESEMPIO. In un conteggio sulla distribuzione di una specie vegetale, in 3 appezzamenti della stessa superficie sono state contate rispettivamente 15, 21 e 24 unità. Verificare se la distribuzione osservata sia in accordo con la distribuzione teorica di 20 in ogni classe, attesa nell'ipotesi di distribuzione uniforme.

Risposta

Senza correzione, il valore del  $\chi^2$  sarebbe:

$$c_{(2)}^2 = \frac{(5)^2}{20} + \frac{(1)^2}{20} + \frac{(4)^2}{20} = 2,100$$

Con la correzione di Yates diventa:

$$c_{(2)}^2 = \frac{(4,5)^2}{20} + \frac{(1,5)^2}{20} + \frac{(4)^2}{20} = 1,925$$

**La correzione di Yates riduce il valore del  $\chi^2$ .**

Il suo merito principale è quello di **rendere le conclusioni più prudenti**, in modo tanto più sensibile quanto più basso è il numero di osservazioni, perché maggiore è l'effetto della variazione  $\pm 0,5$ . Con la correzione, si abbassa la probabilità di rifiutare l'ipotesi nulla. Infatti, **quando il numero di osservazioni è limitato, le variazioni casuali tendono ad aumentare la loro incidenza relativa.**

Gli scarti tra frequenze osservate e frequenze attese non risentono solamente delle differenze realmente esistenti tra le due distribuzioni a confronto, ma anche delle variazioni casuali. In questa ottica, si può intuitivamente comprendere come il test  $\chi^2$  non abbia validità per rifiutare l'ipotesi nulla, in termini più tecnici sia troppo poco potente, quando il numero di osservazioni è troppo ridotto: le variazioni casuali tendono ad essere così elevate, da non permettere più di evidenziare in modo significativo l'esistenza di differenze reali tra osservato ed atteso.

Per indicare l'effetto delle variazioni casuali in un campione di dimensioni troppo ridotte, in alcuni testi si usa il termine **rumore o rumore bianco** (*white noise*), per distinguerlo dalle altre cause che possono generare variabilità (rumore). Quando il numero di osservazioni è basso, il "rumore" (le variazioni casuali) diventa così forte da non permettere di evidenziare le reali tendenze di fondo delle distribuzioni.

Tuttavia, non vi è accordo generale sull'utilità di questa correzione; in alcuni testi non viene né citata né riportata, appunto perché riduce la potenza del test. Più recentemente la maggioranza richiede la sua applicazione o il ricorso ad altri test, come il test di Kolmogorov-Smirnov, il metodo esatto di Fisher, il test G che saranno presentati nei paragrafi successivi.

### 3.3. IL METODO DI KOLMOGOROV-SMIRNOV PER UN CAMPIONE

Quando il numero di osservazioni è ridotto, convenzionalmente inferiore a 30, oppure le frequenze attese entro due o più gruppi sono inferiori a 5, non è possibile utilizzare il test  $\chi^2$ .

E' tuttavia ugualmente possibile verificare la concordanza tra fatto ed ipotesi, ricorrendo al **test di Kolmogorov-Smirnov per un campione**. Ma per la sua applicazione è indispensabile una **condizione aggiuntiva, rispetto al test chi quadrato**:

**i gruppi non possono essere qualitativi, ma devono essere ordinati secondo una scala di tipo almeno ordinale o di rango.**

Per gruppi qualitativi, come gli esperimenti di Mendel o la distribuzione dei gruppi sanguigni, in cui si abbiano meno di 30 osservazioni, non esiste alcun test per verificare la concordanza tra distribuzione osservata e una qualsiasi distribuzione attesa.

Il motivo per cui l'uso del test di Kolmogorov-Smirnov richiede che i gruppi siano ordinati secondo una scala ordinale, quindi in modo prefissato ed uguale per tutti, può essere facilmente compreso mediante il confronto della sua metodologia con quella del chi quadrato:

- mentre per il calcolo del chi quadrato l'ordine dei vari gruppi è ininfluenza sul risultato, trattandosi di una sommatoria,
- la metodologia del test di Kolmogorov-Smirnov si fonda sulle distribuzioni cumulate, che variano in funzione dell'ordine degli addendi.

In questo test, il confronto tra distribuzione osservata ed attesa viene realizzato mediante il valore di massima divergenza tra le due distribuzioni cumulate. Successivamente, **la tabella dei valori critici indica la probabilità di trovare una divergenza pari o superiore a quella calcolata, qualora fosse vera l'ipotesi nulla.**

Il test si fonda sulla logica che, se un campione fosse estratto da una determinata distribuzione teorica, la sommatoria della distribuzione osservata dovrebbe discostarsi dalla sommatoria della distribuzione attesa solo per fattori casuali, di piccola entità.

Indicando con  $O(X_i)$  ogni valore della sommatoria dei dati osservati e con  $A(X_i)$  ogni valore della sommatoria dei dati attesi, la deviazione massima  $D'$  è

$$D' = \max | O(X_i) - A(X_i) |$$

Per l'uso delle tabelle, alcuni testi propongono di rendere  $D'$  indipendente dal numero di osservazioni mediante la trasformazione

$$D = D' / N$$

benché la sua significatività dipenda dalla dimensione ( $N$ ) del campione.

La tavola, riportata alla fine del capitolo, presenta i valori critici per un test a 2 code in rapporto alla dimensione ( $N$ ) del campione e al livello di significatività  $\alpha$  prefissata.

Ad esempio, per apprendere il suo uso, alla probabilità  $\alpha = 0.05$

- con 5 dati ( $N = 5$ ) è significativa una differenza uguale o superiore ( $D^{\geq}$ ) a 0,565;
- con 6 dati è significativa una differenza di 0,521
- con 7 dati una differenza di 0,486, ecc.

Alla probabilità  $\alpha = 0.01$  sono rispettivamente significativi valori di  $D$  uguali o superiori

- a 0,669 (per  $N = 5$ )
- a 0,618 (per  $N = 6$ )
- a 0,577 (per  $N = 7$ ).

La tavola riporta i valori critici fino ad  $N$  uguale a 35.

Con  $N$  maggiore di 35, alla probabilità  $\alpha = 0.05$  sono significativi, sempre per test bilaterali, valori di

$$D^{\geq} 1,36 / \sqrt{N}$$

e alla probabilità 0.01 valori di

$$D^{\geq} 1,63 / \sqrt{N}$$

Con campioni di dimensioni uguali o superiori a 35, si potrebbe utilizzare anche il test chi-quadrato.

**Ma il ricorso al test di Kolmogorov-Smirnov permette di avere un numero molto alto di gruppi, ognuno con poche osservazioni attese, mentre il test chi-quadrato impone un loro raggruppamento per non avere frequenze attese inferiori a 5.**

Il test di Kolmogorov-Smirnov è più potente del test  $\chi^2_{(g.d.l.)}$ , in particolare quando il campione non è grande.

(Il concetto di potenza sarà trattato nel capitolo IV).

Quando la numerosità del campione è grande, i due test hanno potenza simile e forniscono probabilità simili.

ESEMPIO. In dieci ore di osservazione, dalle 7 alle 17, un ricercatore ha avvistato 15 uccelli della stessa specie dal suo luogo di appostamento, con la seguente cadenza oraria:

Ore	7-8	9-10	11-12	13-14	15-16
Uccelli avvistati	0	1	1	9	4

Il ricercatore intende verificare se si tratta di una distribuzione approssimativamente uniforme; in altri termini, se le differenze osservate rispetto a tale ipotesi possono essere considerate entro i limiti delle variazioni accidentali ( $H_0$ ), oppure se sia più attendibile pensare ad una incidenza effettiva dell'ora sul

numero di avvistamenti ( $H_1$ ) e quindi le frequenze nelle diverse fasce orarie seguono una legge diversa (non nota, ma differente da quella di uniformità).

Risposta.

Se l'ora non incidesse sulla frequenza di volo degli uccelli, il ricercatore avrebbe dovuto avvistarne un numero costante per ogni intervallo unitario di tempo; con 15 uccelli osservati in 5 intervalli di tempo, il ricercatore avrebbe dovuto osservarne 3 ogni 2 ore, con la seguente cadenza (distribuzione attesa, nella condizione che l'ipotesi nulla  $H_0$  sia vera):

Ore	7-8	9-10	11-12	13-14	15-16
Distribuzione attesa	3	3	3	3	3

Il confronto a coppie tra le due **distribuzioni cumulate** permette di calcolare le differenze tra coppie di frequenze e di trovare facilmente la differenza assoluta massima:

Ore	7-8	9-10	11-12	13-14	15-16
Distribuzione osservata cumulata	<b>0</b>	<b>1</b>	<b>2</b>	<b>11</b>	<b>15</b>
Distribuzione attesa cumulata	<b>3</b>	<b>6</b>	<b>9</b>	<b>12</b>	<b>15</b>
Differenze assolute	3	5	<b>7</b>	1	0

La differenza assoluta massima è 7.

E' intuitivo pensare che essa possa essere tanto più grande quanto maggiore è lo scarto tra osservato ed atteso. Ma questo valore dipende anche dal numero totale di osservazioni: la differenza massima tende ad aumentare in valore assoluto al crescere delle dimensioni del campione. Anche se la sua significatività è strettamente legata ad esso, è possibile rendere lo scarto massimo assoluto indipendente dal numero totale di osservazioni, mediante il rapporto

$$D[\text{deviazione massima}] = \frac{\text{scarto massimo}}{\text{numero totale di osservazioni}}$$

che, nel caso dell'esempio,

$$D = \frac{7}{15} = 0,466$$

è  $D = 0,466$ .

Sulla tabella dei valori critici di **D** nel test di Kolmogorov-Smirnov per un campione,

con un'ampiezza del campione (**N**) pari a 15 osservazioni,

il valore critico della deviazione massima **D** è

- 0,338 per la probabilità  $\alpha = 0.05$
- 0,404 per la probabilità  $\alpha = 0.01$ .

Il valore calcolato nell'esempio (uguale a 0,466) risulta maggiore di quelli tabulati sia alla probabilità  $\alpha = 0.05$  che a quella  $\alpha = 0.01$ .

Di conseguenza, si rifiuta l'ipotesi nulla **H<sub>0</sub>**, secondo la quale le variazioni sono solamente accidentali; pertanto, implicitamente si accetta l'ipotesi alternativa **H<sub>1</sub>**, secondo la quale le differenze tra distribuzione osservata e distribuzione attesa sono troppo grandi per poter essere, ragionevolmente, ritenuti casuali.

Traducendo il risultato dell'analisi statistica in termini biologici, si può affermare che la specie di uccelli oggetto dell'osservazione si alzi in volo in modo preferenziale durante alcune ore del giorno.

Se nell'esperimento durante le 12 ore d'osservazione si fossero rilevati 30 individui, il campione sarebbe stato sufficientemente grande per ricorrere all'uso del test  $\chi^2$ . Tuttavia, per rispettare anche la condizione di avere in ogni casella un valore minimo di osservazioni non inferiore a 5, sarebbe stato necessario raggruppare le osservazioni di classi adiacenti. E' un'operazione che permette di rispettare le condizioni di validità; ma che determina una perdita d'informazione sulle differenze tra le varie ore. Il test di Kolmogorov-Smirnov permette di suddividere il tempo in un numero molto più grande di gruppi: **con un'analisi più dettagliata, permette di evidenziare differenze che con il raggruppamento tendono a scomparire**. Inoltre, il raggruppamento in classi di dimensioni maggiori, per ridurre il numero di gruppi, può contenere una dose elevata di soggettività, determinando differenze tra le due cumulate che variano in rapporto al numero e al tipo di classi formate.

### **3.4. IL CONFRONTO TRA DUE DISTRIBUZIONI OSSERVATE, PER TEST DI**

#### **INDIPENDENZA: LE TABELLE DI CONTINGENZA 2 X 2 (FOURFOLD TABLES)**

Quando si confrontano **le frequenze relative di risposte binarie in due popolazioni indipendenti**, si può costruire una tabella a doppia entrata, chiamata tabella di contingenza (dal latino *contingo*, *vengo a contatto*, in quanto i risultati sono prodotti dall'incontro di due serie di fattori o caratteristiche). Per ognuno dei due gruppi, deve essere riportato il conteggio di risposte binarie, quali il numero di successi e quello di insuccessi oppure di quelli che presentano la caratteristica X o quella alternativa Y.

Trattandosi del **confronto tra due differenti campioni con risposte alternative di tipo binario, la tabella costruita con i dati sperimentali è chiamata tabella 2 x 2** (in inglese, *fourfold tables*).

**Il test chi quadrato permette di verificare se le proporzioni di successi e di insuccessi nei due gruppi sono indipendenti dal trattamento al quale sono sottoposti, oppure se esiste associazione tra essi.**

Per esempio, si supponga di voler analizzare il problema se vivere in una zona ad alto inquinamento atmosferico incida sulla frequenza di malattie polmonari. A questo scopo, in una zona con tassi elevati d'inquinamento dell'aria ed in una senza inquinamento, o comunque con livelli molto bassi, sono stati analizzati alcune decine d'individui, residenti da alcuni anni, contando quanti sono coloro che presentano malattie polmonari.

DISTRIBUZIONE OSSERVATA IN TABELLA 2 X 2

	Persone <u>con</u> malattie	Persone <u>senza</u> malattie	Totale
Zona ad <u>alto</u> inquinamento	<b>32</b> a	<b>48</b> b	80 n <sub>1</sub>
Zona a <u>basso</u> inquinamento	<b>13</b> c	<b>57</b> d	70 n <sub>2</sub>
Totale	45 n <sub>3</sub>	105 n <sub>4</sub>	150 N

La tabella precedente riporta i risultati dell'analisi epidemiologica. Con questi dati, si può concordare con la teoria precedentemente espressa?

**Il test chi quadrato nella metodologia qui presentata utilizza i casi effettivamente contati, non le frequenze relative o percentuali, anche se su di esse vengono formulate le ipotesi.**

Per comprendere esattamente il metodo statistico necessario per rispondere, è bene seguire alcuni passaggi logici.

- 1- Se fosse vera l'ipotesi nulla ( $H_0$ : vivere in una zona ad alto inquinamento atmosferico non cambia la frequenza di malattie polmonari, rispetto ad una zona a basso inquinamento), la frequenza relativa di persone con malattie polmonari nei 2 gruppi a confronto sarebbe uguale; le differenze riscontrate sarebbero da interpretare come variazioni casuali.
- 2- La stima migliore di questa frequenza relativa o incidenza percentuale, valida nella condizione che l'ipotesi nulla sia vera, è data dalla somma delle persone con malattie polmonari nei 2 gruppi ( $32 + 13 = 45$ ) rapportate al numero totale di persone osservate ( $45 / 150 = 0,3$ ).

- 3- Considerando che i due campioni a confronto hanno un numero differente di osservazioni, sempre nel caso che sia vera l'ipotesi nulla, nel primo campione (che è composto da 80 individui) dovremmo aspettarci di trovare 24 persone ( $0,3 \times 80 = 24$ ) con malattie polmonari e nel secondo campione (composto da 70 individui) di trovarne 21 ( $0,3 \times 70 = 21$ ).

I quattro valori attesi possono essere presentati in una tabella 2 x 2, come i valori osservati.

Per la sua costruzione, è utile riportare dapprima i 4 totali marginali ed il totale generale.

Successivamente, si calcola ognuno dei 4 valori attesi, moltiplicando il totale di riga per il totale di colonna, diviso per il totale generale:

$$a = n_1 \times n_3 / N;$$

$$b = n_1 \times n_4 / N;$$

$$c = n_2 \times n_3 / N$$

$$d = n_2 \times n_4 / N$$

DISTRIBUZIONE ATTESA IN TABELLA 2 X 2

	Persone <u>con</u> malattie	Persone <u>senza</u> malattie	Totale
Zona ad <u>alto</u> inquinamento	24 <b>a</b>	56 <b>b</b>	80 <b>n<sub>1</sub></b>
Zona a <u>basso</u> inquinamento	21 <b>c</b>	49 <b>d</b>	70 <b>n<sub>2</sub></b>
Totale	45 <b>n<sub>3</sub></b>	105 <b>n<sub>4</sub></b>	150 <b>N</b>

**E' utile osservare come non sia necessario calcolare tutte e quattro le frequenze attese. Poiché i totali marginali sono già fissati, è sufficiente calcolare uno solo dei 4 valori attesi, qualunque esso sia, per poter dedurre gli altri 3 per differenza.** Per esempio, una volta calcolato 24 per la casella a è ovvio che nella casella b possiamo trovare solo 56, perché il totale della riga è 80, e nella casella c possiamo trovare solo 21, perché il totale è già prefissato a 45; da uno qualsiasi di questi 3 dati già scritti deriva che la casella d può riportare solamente 49, affinché i vari totali possano coincidere.

Il ragionamento non varia, partendo da qualsiasi casella.

Queste considerazioni spiegano perché **una tabella attesa 2 x 2 ha solamente 1 grado di libertà.**

Colui che propose questo metodo per primo, **Karl Pearson**, attribuì erroneamente un numero maggiore di gradi di libertà. Fu il suo allievo **R. A. Fisher**, allora molto giovane, che mostrò il procedimento esatto.

Stimata la distribuzione attesa nell'ipotesi che sia vera l'ipotesi nulla, dalle differenze tra osservato ed atteso si calcola il valore del chi quadrato, mediante la formula generale già presentata:

$$c_{(g.d.l.)}^2 = \sum_{i=1}^n \frac{(f_i^{oss} - f_i^{att})^2}{f_i^{att}}$$

dove:

$f_i^{oss}$  = frequenza osservata i-esima

$f_i^{att}$  = frequenza attesa i-esima

ed estendendo la sommatoria ai dati di tutte quattro le caselle.

Con i dati dell'esempio

$$\begin{aligned} \chi_{(1)}^2 &= (32 - 24)^2 / 24 + (48 - 56)^2 / 56 + (13 - 21)^2 / 21 + (57 - 49)^2 / 49 = \\ &= 2,666 + 1,143 + 3,048 + 1,306 = 8,163 \end{aligned}$$

si ottiene un valore del chi quadrato, con 1 gdl, uguale a 8,163

La tavola sinottica del  $\chi_{(1)}^2$  riporta il valore critico di **3,84** alla probabilità  $\alpha = 0.05$  ed il valore critico di **6,64** alla probabilità  $\alpha = 0.01$ . Il valore calcolato (8,163) è superiore; di conseguenza, si rifiuta l'ipotesi nulla ed implicitamente si accetta l'ipotesi alternativa.

Con i dati dell'esempio, lo statistico arriva alla conclusione che, con probabilità inferiore a 0.01 di commettere un errore (cioè che sia vera l'ipotesi nulla), la percentuale di persone con malattie polmonari è significativamente differente nei due gruppi a confronto.

Questa procedura è utile per capire il reale significato del test chi quadrato in tabelle di contingenza 2 x 2. Inoltre, **il confronto tra distribuzione osservata e distribuzione attesa mostra in quali caselle si trovano le differenze più importanti**. Nell'esempio, tale confronto mostra che le persone con malattie polmonari (riportate nella tabella delle frequenze osservate) sono più frequenti nella zona con maggior inquinamento e sono meno frequenti nella zona senza inquinamento atmosferico, rispetto all'ipotesi nulla che esse abbiano la stessa frequenza percentuale (riportate nella tabella delle frequenze attese).

Si può ottenere lo stesso risultato ed evitare il lungo calcolo delle frequenze attese, con il ricorso alla **formula per il calcolo rapido del chi quadrato per le tabelle di contingenza 2 x 2**

$$c_{(1)}^2 = \frac{(a \cdot d - b \cdot c)^2 \cdot N}{n_1 \cdot n_2 \cdot n_3 \cdot n_4}$$

dove, con la simbologia riportata nelle tabelle precedenti,

**a, b, c, d** sono le frequenze osservate nei due campioni a confronto

$n_1, n_2, n_3, n_4$  sono i totali marginali

$N$  è il totale generale di osservazioni.

Il calcolo, con i dati sperimentali dell'esempio precedentemente utilizzato, fornisce

$$\chi^2_{(1)} = [(32 \times 57 - 48 \times 13)^2 \times 150] / (80 \times 70 \times 45 \times 105) = \\ = (1824 - 624)^2 \times 150 / 26460000 = 1440000 \times 150 / 26460000 = 8,163$$

un  $\chi^2_{(1)} = 8,163$ .

E' un valore identico a quello calcolato in precedenza, con la formula estesa.

L'equivalenza tra le due formule potrebbe essere dimostrata con una serie di passaggi matematici; ma per l'utente della statistica applicata è sufficiente ricordare le due formule, da usare nelle differenti condizioni.

ESEMPIO. Si vuole controllare l'effetto di due tossici su due gruppi di animali. Il tossico A, somministrato a 70 animali, ha causato la morte di 22 individui (e naturalmente 48 sono sopravvissuti), mentre il tossico B in un esperimento con 50 animali ha causato la morte di 24 individui. Si vuole sapere se i due tossici hanno gli stessi effetti sulla mortalità o sopravvivenza ( $H_0$ ); oppure se i due tossici hanno probabilmente effetti letali differenti ( $H_1$ ).

Per meglio evidenziare i dati del problema, i conteggi devono essere riportati in una tabella a due entrate

	Animali morti	Animali sopravv.	Totale
Tossico A	<b>22</b>	<b>48</b>	70
Tossico B	<b>24</b>	<b>26</b>	50
Totale	46	74	120

che fornisce in modo chiaro tutte le informazioni utili.

Nella condizione che l'ipotesi nulla sia vera (i farmaci hanno lo stesso effetto e le variazioni riscontrate sono dovute solo al caso), le frequenze attese possono essere calcolate dai totali: se i due tossici producessero lo stesso effetto, in ognuno dei due gruppi morirebbero il 38% (46/120) e sopravviverebbero il 62% (74/120), ovviamente rapportati ai totali dei due gruppi di 70 e 50 animali.

E' quindi possibile costruire una tabella delle frequenze attese.

Dapprima, si riportano i totali marginali e quello generale; successivamente, da essi si stimano i dati attesi in ognuna delle 4 caselle con la relazione

$$f_{\text{frequenza attesa}} = \frac{\text{totale riga} \times \text{totale colonna}}{\text{totale generale}}$$

che devono essere collocati nella tabella 2 x 2:

	Animali morti	Animali sopravv.	Totale
Tossico A	<b>26,83</b>	<b>43,17</b>	70
Tossico B	<b>19,17</b>	<b>30,83</b>	50
Totale	46	74	120

E' importante osservare che una volta che è stata riportata la prima frequenza attesa

$$26,83 = \frac{70 \cdot 46}{120}$$

le altre possono essere ottenute anche per differenza dai totali rispettivi di riga o di colonna:

$$43,17 = \frac{70 \cdot 74}{120} \quad \text{oppure} \quad 43,17 = 70 - 26,83$$

$$19,17 = \frac{50 \cdot 46}{120} \quad \text{oppure} \quad 19,17 = 46 - 26,83$$

L'ultima frequenza attesa (30,83) può essere calcolata sia dai suoi due totali marginali che dal totale generale.

La tabella di contingenza  $2 \times 2$  ha un solo valore che può assumere qualsiasi valore, quindi ha 1 gdl.

Dal confronto tra tabella dei dati osservati e tabella dei dati attesi si può calcolare il valore del  $\chi^2$  mediante la formula generale:

$$c_{(1)}^2 = \sum_{i=1}^4 \frac{(f_i^{oss} - f_i^{att})^2}{f_i^{att}}$$

Con i dati dell'esempio

$$\begin{aligned} c_{(1)}^2 &= \frac{(22 - 26,83)^2}{26,83} + \frac{(48 - 43,17)^2}{43,17} + \frac{(24 - 19,17)^2}{19,17} + \frac{(26 - 30,83)^2}{30,83} = \\ &= \frac{23,33}{26,83} + \frac{23,33}{43,17} + \frac{23,33}{19,17} + \frac{23,33}{30,83} = 0,87 + 0,55 + 1,24 + 0,76 = 3,42 \end{aligned}$$

si ottiene un valore uguale a 3,42.

Oppure, ricordando la simbologia,

	X	x	Totale
Y	a	b	n <sub>1</sub>
y	c	d	n <sub>2</sub>
Totale	n <sub>3</sub>	n <sub>4</sub>	N

e la formula abbreviata

$$c_{(1)}^2 = \frac{(a \cdot d - b \cdot c)^2 \cdot N}{n_1 \cdot n_2 \cdot n_3 \cdot n_4}$$

è possibile arrivare al risultato in modo molto più rapido

$$c_{(1)}^2 = \frac{(22 \cdot 26 - 48 \cdot 24)^2 \cdot 120}{70 \cdot 50 \cdot 46 \cdot 74} = \frac{(572 - 1152)^2 \cdot 120}{11914000} = \frac{336400 \cdot 120}{11914000} = \frac{40368000}{11914000} = 3,389$$

ottenendo un chi quadrato uguale a 3,389.

I due risultati 3,42 e 3,389 non coincidono perfettamente anche per effetto degli arrotondamenti, ma risultano sempre molto simili. **La formula abbreviata è da preferire per il calcolo** (richiede meno tempo), mentre **la formula generale è di aiuto nell'interpretazione dei risultati** (mediante il confronto tra frequenze osservate e frequenze attese).

Nella tabella dei valori critici della distribuzione  $\chi^2$  per 1 grado di libertà alla probabilità 5% è riportato 3,84; il valore del chi quadrato che è stato calcolato è inferiore. Di conseguenza, la probabilità che le differenze riscontrate tra i due tossici siano dovute al caso è da ritenersi elevata: non si può respingere l'ipotesi nulla.

In termini biologici, con questi dati non è possibile dimostrare che i due tossici abbiano effetti significativamente differenti.

**Anche per le tabelle 2 x 2 valgono le stesse condizioni di validità, legate al numero di osservazioni raccolte: il chi quadrato fornisce risultati attendibili per grandi campioni (superiori a 100 osservazioni) e non è attendibile per campioni eccessivamente ridotti (inferiori a 30 osservazioni).**

Quando si debbono confrontare campioni che hanno un numero totale di osservazioni tra 100 e 30, con la formula generale si deve aggiungere 0,5 allo scarto minore e togliere 0,5 allo scarto maggiore.

Nella formula abbreviata, questa **correzione per la continuità di Yates** comporta una riduzione del valore del chi quadrato, mediante la sottrazione di  $N/2$  al valore assoluto dello scarto tra le due diagonali.

Con la correzione di Yates, la formula del chi quadrato per tabelle  $2 \times 2$  diviene:

$$c_{(1)}^2 = \frac{\left( |a \cdot d - b \cdot c| - \frac{N}{2} \right)^2 \cdot N}{n_1 \cdot n_2 \cdot n_3 \cdot n_4}$$

**Gli effetti della correzione sono relativamente tanto maggiori quanto più basso è il numero di osservazioni.** Quando il campione è di grandi dimensioni, la correzione diviene trascurabile. Sulla base di queste osservazioni, alcuni testi come formula abbreviata riportano solamente questa ultima: è utile in campioni con meno di 100 osservazioni e diviene superflua, non comporta alterazioni sensibili, quando il campione è molto grande .

ESEMPIO. Per valutare gli effetti di due diserbanti, si confronta il numero di piante cresciute normalmente e di quelle non cresciute nei rispettivi appezzamenti di terreno. Nella zona in cui è stato utilizzato il diserbante A, su un totale di 18 pianticelle presenti 12 sono cresciute e 6 sono seccate; nella zona dove è stato utilizzato il diserbante B, 26 sono cresciute e 9 sono morte.

I due diserbanti hanno effetti significativamente differenti sulla crescita delle pianticelle?

Risposta

Per presentare i dati in modo completo e chiaro, in molti testi è ritenuto indispensabile riportare sempre la tabella di contingenza  $2 \times 2$ .

	Piante cresciute	Piante non cresciute	Totale
Diserbante A	<b>12</b>	<b>6</b>	18
Diserbante B	<b>26</b>	<b>9</b>	35
Totale	38	15	53

Secondo l'ipotesi nulla ( $H_0$ ), i due diserbanti hanno effetti uguali: le differenze riscontrate sono imputabili al caso.

Secondo l'ipotesi alternativa ( $H_1$ ), i due diserbanti hanno effetti significativamente differenti.

E' un confronto tra due campioni indipendenti con un numero di osservazioni (53) che è ritenuto sufficiente per l'uso del test  $\chi^2$ , ma con la correzione di Yates. E' possibile ricorrere alla formula per il calcolo rapido, apportando la correzione per la continuità.

$$c_{(1)}^2 = \frac{\left( |12 \cdot 9 - 6 \cdot 26| - \frac{53}{2} \right)^2 \cdot 53}{18 \cdot 35 \cdot 38 \cdot 15} = \frac{(|108 - 156| - 26,5)^2 \cdot 53}{359100} = \frac{462,25 \cdot 53}{359100} = \frac{24499,25}{359100} = 0,0945$$

Il valore calcolato del chi quadrato con un grado di libertà (0,0945) è particolarmente basso, inferiore a quello tabulato alla probabilità 0.05. Con una lettura della tavola dei valori critici più dettagliata, è possibile osservare che è inferiore a quello corrispondente alla probabilità 0.90.

Di conseguenza, esiste una probabilità elevatissima di trovare per caso scarti uguali o superiori a quelli riscontrati: non si può rifiutare l'ipotesi nulla, secondo la quale le differenze osservate tra gli effetti dei due diserbanti sono dovute solamente a variazioni casuali.

Nella sua pubblicazione dell'anno 1900 (*On the criterion that a given system of deviations from the probable in the case of a correlated system of variables in such that it can be reasonably supposed to have arisen from random sampling*. Philos. Mag., 50 (5), pp. 157-175), **Karl Pearson** propone una formula fondata sul calcolo delle proporzioni

$$c^2 = N \cdot \sum_{i=1}^2 \sum_{j=1}^2 \frac{\left( |p_{ij} - p_{i.} \cdot p_{.j}| - \frac{1}{2N} \right)^2}{p_{i.} \cdot p_{.j}}$$

che ingloba anche la correzione per la continuità, dove

	X	x	Totale
Y	$p_{11}$	$p_{12}$	$p_{1.}$
y	$p_{21}$	$p_{22}$	$p_{2.}$
Totale	$p_{.1}$	$p_{.2}$	<b>1</b>

$p_{ij}$  è la frequenza relativa osservata di una casella

$p_i$  e  $p_j$  è la somma di riga e di colonna delle frequenze relative, per cui

il loro prodotto  $p_i \cdot p_j$  è la frequenza attesa in ogni casella collocata al loro incrocio e

$N$  è il numero assoluto di osservazioni.

Riprendendo l'ultima tabella con valori osservati, in cui  $N = 53$ , le proporzioni diventano

	Piante cresciute	Piante non cresciute	Totale
Diserbante A	<b>0,226</b>	<b>0,113</b>	<b>0,339</b>
Diserbante B	<b>0,491</b>	<b>0,170</b>	<b>0,661</b>
Totale	<b>0,717</b>	<b>0,283</b>	<b>1</b>

e il calcolo del chi quadrato

$$\chi^2 = 53 \cdot \left[ \frac{\left( \left| 0,226 - 0,339 \cdot 0,717 \right| - \frac{1}{2 \cdot 53} \right)^2}{0,339 \cdot 0,717} + \frac{\left( \left| 0,113 - 0,339 \cdot 0,283 \right| - \frac{1}{2 \cdot 53} \right)^2}{0,339 \cdot 0,283} + \frac{\left( \left| 0,491 - 0,661 \cdot 0,717 \right| - \frac{1}{2 \cdot 53} \right)^2}{0,661 \cdot 0,717} + \frac{\left( \left| 0,170 - 0,661 \cdot 0,283 \right| - \frac{1}{2 \cdot 53} \right)^2}{0,661 \cdot 0,283} \right]$$

$$\chi^2 = 53 \cdot \left[ \frac{\left( \left| 0,226 - 0,243 \right| - 0,0094 \right)^2}{0,243} + \frac{\left( \left| 0,113 - 0,0959 \right| - 0,0094 \right)^2}{0,0959} + \frac{\left( \left| 0,491 - 0,474 \right| - 0,0094 \right)^2}{0,474} + \frac{\left( \left| 0,170 - 0,187 \right| - 0,0094 \right)^2}{0,187} \right]$$

$$\chi^2 = 53 \cdot \left( \frac{0,0076^2}{0,243} + \frac{0,0077^2}{0,096} + \frac{0,0076^2}{0,474} + \frac{0,0076^2}{0,187} \right)$$

$$\chi^2 = 53 \cdot (0,00024 + 0,00062 + 0,00012 + 0,00031) = 53 \cdot 0,00129 = 0,0684$$

stima un valore (uguale a 0,0684) diverso dal precedente (0,0945) a causa degli arrotondamenti e dell'assenza della correzione per la continuità.

E' semplice osservare come, senza la correzione per la continuità, questa formula sia uguale a quella generale

$$\chi^2 = (\text{Freq. Osservata} - \text{Freq. Attesa})^2 / \text{Freq. Attesa}$$

### 3.5. CONFRONTI TRA FREQUENZE RELATIVE CON LA DISTRIBUZIONE NORMALE E SUA CORREZIONE PER LA CONTINUITA'

Per il teorema del limite centrale, in campioni abbastanza numerosi la distribuzione della frequenza relativa  $p$  di una popolazione è approssimativamente normale, con media campionaria  $p$  e deviazione standard della popolazione  $\sigma_p$  (dove  $\sigma^2 = p \cdot q$ ). L'assunzione rimane valida anche per le percentuali, che tuttavia devono essere trasformate in frequenze relative, per utilizzare le formule proposte.

Questa **approssimazione** della distribuzione chi quadrato alla distribuzione normale non è ritenuta corretta, quando il numero totale di osservazioni  $N$  è piccolo oppure  $p$  e  $q$  sono molto distanti da 0,5. Si ha un uso corretto della **distribuzione normale nel confronto tra rapporti, quando  $Np$  e  $Nq$  sono entrambi maggiori di 5.**

Anche questa norma non trova tutti i testi unanimi, poiché alcuni accettano frequenze assolute inferiori; è quindi da **considerare solamente indicativa non tassativa.**

In grandi campioni, se  $p_1$  e  $p_2$  sono le **proporzioni osservate** di casi con la caratteristica in esame in due campioni indipendenti, è possibile verificare la significatività della loro differenza con **un test z,**

$$z = \frac{p_1 - p_2}{\sqrt{p^*(1-p^*) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

dove

$p^*$  è la proporzione media ponderata dei 2 gruppi a confronto, ottenuta con

$$p^* = \frac{m_1 + m_2}{n_1 + n_2}$$

in cui  $m_1$  e  $m_2$  sono i casi positivi nei gruppi **1** e **2** a confronto, composti rispettivamente da  $n_1$  e  $n_2$  casi.

A volte, si pone il problema di verificare se le due proporzioni differiscono di una quantità predeterminata  $p$ . Per esempio, per sostituire un farmaco già sperimentato, si potrebbe esigere che il nuovo farmaco proposto determini una frequenza di guarigione superiore almeno di una certa quantità  $p$ , senza accontentarsi del fatto che abbia solo una frequenza di guarigioni significativamente maggiore del precedente. In queste condizioni, nella formula sopra riportata **il numeratore diviene  $(p_1 - p_2) - p$ .**

Inoltre, l'effetto dovuto alle numerosità dei due campioni, stimato con  $(1/n_1 + 1/n_2)$ , può essere calcolato anche con  $(n_1 + n_2) / (n_1 \times n_2)$ , esistendo la relazione

$$\frac{1}{n_1} + \frac{1}{n_2} = \frac{n_1 + n_2}{n_1 \times n_2}$$

Pertanto, **la formula per il calcolo del valore di z nel confronto tra due proporzioni può essere scritta** pure come

$$z = \frac{|p_1 - p_2| - p}{\sqrt{p^* (1 - p) \cdot \frac{n_1 + n_2}{n_1 \times n_2}}}$$

dove

$p_1$  è la proporzione di casi in esame nel campione **1** con numerosità  $n_1$ ,

$p_2$  è la proporzione di casi in esame nel campione **2** con numerosità  $n_2$ ,

$p$  è la differenza tra le due proporzioni di cui si intende verificare la significatività,

$p^*$  è la proporzione della popolazione, stimata come media ponderata dai due campioni.

Anche nell'uso dell'approssimazione normale, **per campioni di dimensioni medie si deve apportare la correzione per la continuità, proposta da F. Yates** nel 1934 (con l'articolo *Contingency tables involving small numbers and the  $\chi^2$  test* sulla rivista *J. roy. statist. Soc. Suppl., 1, pp. 217-235*). Infatti, le proporzioni derivano da dati discreti (quali sono i conteggi) e possono assumere solo alcuni valori, mentre la distribuzione normale standardizzata è continua. **La correzione può essere sempre utilizzata, ma la sua importanza diminuisce al crescere del numero d'osservazioni; deve essere apportata quando il numero di osservazioni ( $N = n_1 + n_2$ ) è approssimativamente tra 30 e 100 (o 200).**

Come nei casi precedenti, la correzione riduce il valore calcolato e rende le conclusioni del test più prudentziali: abbassa la probabilità di rifiutare l'ipotesi nulla.

**Nel confronto tra due proporzioni, la correzione di Yates consiste nel diminuire la differenza tra le due proporzioni campionarie di una quantità uguale a**

$$\frac{1}{2} (1/n_1 + 1/n_2)$$

che ha un effetto trascurabile quando i due campioni hanno molte osservazioni.

**Quando l'ipotesi nulla è che i due campioni siano stati estratti da popolazioni con la stessa proporzione**

$$H_0: p_1 = p_2$$

con la correzione per la continuità di **Yates** la formula estesa del test z diviene

$$z = \frac{|p_1 - p_2| \cdot \frac{1}{2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{p^* (1 - p^*) \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Mentre il test chi quadrato fornisce solo valori positivi, poiché gli scarti sono elevati al quadrato, **il test z fornisce risultati sia positivi che negativi, secondo il segno della differenza tra le due proporzioni.**

Di conseguenza, **nel formulare l'ipotesi alternativa, i due test offrono possibilità differenti.**

La **tabella del  $c^2$  fornisce la probabilità per un test a due code o bilaterale.** In altri termini, è possibile formulare solo una ipotesi alternativa: le due proporzioni a confronto appartengono a popolazioni differenti. Con i simboli, si scrive

$$\mathbf{H_1: p_1 \neq p_2}$$

Nel caso di tabelle 2 x 2, con il test chi quadrato è solo possibile dimostrare che le 2 percentuali a confronto sono differenti, quando si è in grado di rifiutare l'ipotesi nulla.

**Con la distribuzione normale applicata alle proporzioni o percentuali, sono possibili due diverse impostazioni dell'ipotesi alternativa  $H_1$ .** E' possibile verificare:

1 - se esiste una differenza nelle frequenze relative tra i due gruppi, senza predeterminare quale dei due debba essere il maggiore (o il minore): si tratta di un **test bilaterale o a due code, come già per il test  $c^2$ :**

$$\mathbf{H_1: p_1 \neq p_2}$$

2 - se un gruppo ha una frequenza relativa significativamente maggiore (oppure minore): è un **test unilaterale o a una coda:**

$$\mathbf{H_1: p_1 < p_2 \quad oppure \quad H_1: p_1 > p_2}$$

In ognuno di questi ultimi 2 casi ad una coda, viene a priori rifiutata come non accettabile od illogica la possibilità alternativa a quella proposta.

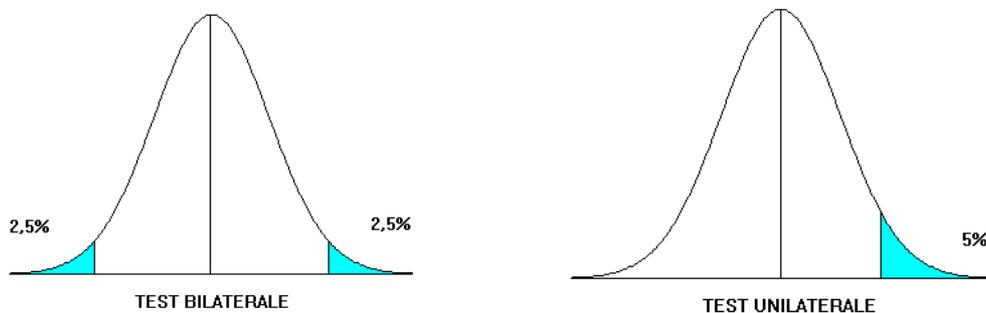
Alcuni esempi chiariscono meglio questi concetti.

Si supponga di confrontare l'effetto di 2 farmaci, per scegliere il migliore sulla base della percentuale di persone guarite. Se la risposta è accettata qualunque sia il farmaco, è un test a 2 code o bilaterale.

Ma quando si confronta l'effetto di un farmaco con il placebo (sostanza senza principio attivo; spesso acqua distillata nel caso di iniezioni oppure farina nel caso di pillole), in realtà non si vuole sapere se esiste differenza tra le percentuali di guariti nei due gruppi a confronto, ma solo se nel campione di persone alle quali è stato somministrato il farmaco la percentuale di guariti è effettivamente superiore a

quella riscontrata nel gruppo di persone alle quali è stato somministrato il placebo: è un test ad una coda o unilaterale. In questo secondo caso, sarebbe strano che i dati dell'esperimento evidenziassero che la percentuale di persone guarite fosse maggiore nel gruppo del placebo, rispetto a quello cui è stato somministrato il farmaco. Il risultato dovrebbe essere interpretato come privo di significato reale; l'esperimento verrebbe messo in discussione, nel ragionevole dubbio che sia stato condotto in modo errato.

**La distinzione tra test a due code e test a una coda non è solamente una questione di logica. Ha effetti pratici importanti: da essa dipende la distribuzione delle probabilità ed il valore critico per rifiutare l'ipotesi nulla, come chiarisce il grafico.**



Scegliendo la probabilità del 5%,

- in un test a due code, si hanno due zone di rifiuto collocate ai due estremi, ognuna con un'area di 2,5%
- in un test a una coda, si ha una sola zona di rifiuto, con un'area di 5 %.

**Esistono maggiori probabilità di rifiutare l'ipotesi nulla quando si effettua un test ad una coda, che quando si effettua un test a due code.** Anche nella rappresentazione grafica, risulta evidente in modo visivo che, alla stessa probabilità totale, in un test unilaterale il valore critico è minore di quello in un test bilaterale. Come verrà più ampiamente discusso nel capitolo 4, **il test unilaterale è più potente del test bilaterale** (definizione: la potenza di un test è la capacità di rifiutare l'ipotesi nulla quando essa è falsa).

ESEMPIO 1. Un metodo di ricattura ha dimostrato una efficienza del 20% su una popolazione di 200 individui, mentre un secondo metodo ha dimostrato un'efficienza del 26 % su 150 individui. Tra i due metodi esiste una differenza significativa nella percentuale di ricattura?

Risposta.

Si tratta di un test bilaterale o a due code, perché il quesito a priori accetta che possa risultare migliore sia il primo che il secondo metodo.

**L'ipotesi nulla** è che la probabilità di ricattura nelle due popolazioni sia uguale;  $H_0: p_1 = p_2$

**L'ipotesi alternativa** è che le due proporzioni siano diverse;  $H_1: p_1 \neq p_2$

I due campioni sono di dimensioni sufficientemente grandi per poter utilizzare la distribuzione  $z$ , mediante

$$z = \frac{p_1 - p_2}{\sqrt{p^*(1-p^*) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

dove

$$p_1 = 0,20; \quad p_2 = 0,26; \quad n_1 = 200; \quad n_2 = 150;$$

$$m_1 = (0,20 \times 200) = 40; \quad m_2 = (0,26 \times 150) = 39; \quad p^* = \frac{40 + 39}{200 + 150} = 0,226$$

Il valore di  $z$  è

$$Z = \frac{0,20 - 0,26}{\sqrt{0,226 \cdot 0,774 \cdot \left(\frac{1}{200} + \frac{1}{150}\right)}} = \frac{-0,06}{\sqrt{0,17492 \cdot 0,01166}} = \frac{-0,06}{0,04516} = -1,32$$

uguale a  $-1,32$ .

Nella tabella della distribuzione normale, a  $z = 1,32$  corrisponde una probabilità di 9,34% nella coda sinistra. Trattandosi di un test a due code, ad essa occorre sommare anche la probabilità di 9,34% nella coda destra, per un totale di 18,68%. E' una probabilità molto alta, superiore al limite del 5% che di solito viene considerato il valore soglia tra probabilità alte e probabilità basse; non è quindi possibile rifiutare l'ipotesi nulla, secondo la quale la differenza riscontrata tra i due campioni è imputabile solamente al caso.

Tra i due metodi di ricattura i dati raccolti non dimostrano una differenza significativa.

**ESEMPIO 2.** Su una rivista scientifica, viene descritto un metodo di ricattura che ha dimostrato un'efficienza del 15% su una popolazione di 400 animali. Alcuni ricercatori hanno apportato alcuni miglioramenti e dopo alcune prove pensano di aver aumentato l'efficienza del metodo di almeno il 2%. In un esperimento successivo, questo secondo metodo permette la ricattura di 25 individui su 100. Il secondo metodo è effettivamente migliore del precedente di almeno il 2% ?

Risposta.

Si tratta di un test unilaterale o a una coda.

L'ipotesi nulla è che la probabilità di ricattura del secondo metodo sia uguale a quello del primo aumentato del 2%;  $H_0: p_1 + 0,02 = p_2$

L'ipotesi alternativa è che le probabilità di ricattura del secondo metodo sia significativamente superiore a quella del primo aumentata del 2%;  $H_1: p_2 > p_1 + 0,02$

I due campioni sono di dimensioni sufficientemente grandi per poter utilizzare la distribuzione z, mediante

$$z = \frac{|p_1 - p_2| - P}{\sqrt{p^*(1-p^*) \cdot \frac{n_1 + n_2}{n_1 \times n_2}}}$$

dove

$$p_1 = 0,15; \quad n_1 = 400; \quad m_1 = 60; \quad p_2 = 0,25; \quad n_2 = 100; \quad m_2 = 25;$$

e

$$\pi = 0,02; \quad p^* = \frac{60 + 25}{400 + 100} = 0,17$$

Pertanto, il valore di z è

$$Z = \frac{|0,25 - 0,15| - 0,02}{\sqrt{0,17 \cdot 0,83 \cdot \left(\frac{400 + 100}{400 \times 100}\right)}} = \frac{0,08}{\sqrt{0,1411 \cdot 0,0125}} = \frac{0,08}{0,04199} = 1,90$$

uguale a 1,90.

Ad un valore di  $Z = 1,90$  nella tabella della distribuzione normale standardizzata corrisponde una probabilità .0287, pari al 2,87% in un test ad una coda; è una probabilità molto bassa, inferiore al 5%.

Si rifiuta l'ipotesi nulla  $H_0$  e quindi si accetta l'ipotesi  $H_1$  alternativa, secondo la quale il secondo metodo è migliore del primo metodo, con una percentuale di ricattura significativamente più elevata di almeno il 2%.

Per verificare solamente se il secondo metodo sia migliore, con ipotesi nulla  $H_0: p_1 = p_2$

ed ipotesi alternativa  $H_1: p_1 < p_2$  oppure  $p_2 > p_1$

il calcolo finale per il test z sarebbe stato

$$z = \frac{0,10}{0,04199} = 2,38$$

Con un valore di  $z = 2,38$  in un test unilaterale si sarebbe stimata una probabilità di 0,87% che la differenza riscontrata sia imputabile solamente al caso. E' una probabilità molto bassa, ovviamente inferiore a quella stimata precedentemente.

A maggior ragione, si sarebbe dovuto rifiutare l'ipotesi nulla ed accettare l'ipotesi alternativa: il secondo metodo di ricattura mostra una efficienza significativamente superiore al primo.

Per esprimere gli stessi concetti con altri termini, si può affermare che esiste associazione tra il metodo di ricattura ed il numero di animali catturati.

**ESEMPIO 3.** Si vuole confrontare l'effetto di due sostanze tossiche, diluite nell'acqua alla stessa concentrazione, sulla sopravvivenza di una specie di pesci. Con la sostanza A dopo 48 ore sono stati contati 4 morti su un campione di 80 individui; con la sostanza B i morti sono stati 10 su un campione di 50.

Esiste una differenza significativa tra la tossicità delle due sostanze?

Risposta.

E' un test a due code, dove l'**ipotesi nulla** è  $H_0: p_1 = p_2$

e l'**ipotesi alternativa** è  $H_1: p_1 \neq p_2$

Almeno un campione, il primo, è di dimensioni relativamente piccole, con un numero di decessi (n·p) uguale a 4.

E' conveniente **utilizzare la distribuzione normale standardizzata z, con la correzione di Yates** con la relazione

$$z = \frac{|p_1 - p_2| - \frac{1}{2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{p^* (1 - p^*) \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

dove

$$p_1 = (4/80) = 0,05; \quad n_1 = 80; \quad m_1 = 4;$$

$$p_2 = (10/50) = 0,2; \quad n_2 = 50; \quad m_2 = 10; \quad p^* = \frac{4 + 10}{80 + 50} = 0,108$$

Sostituendo ai simboli i dati del problema, si ottiene un valore di **z**

$$z = \frac{|0,05 - 0,20| - \frac{1}{2} \left( \frac{1}{80} + \frac{1}{50} \right)}{\sqrt{\left[ 0,108 \cdot 0,892 \cdot \left( \frac{1}{80} + \frac{1}{50} \right) \right]}} = \frac{|0,15| - 0,01625}{\sqrt{[0,096 \cdot 0,0325]}} = \frac{0,13375}{0,05596} = 2,39$$

uguale a 2,39.

In un test bilaterale, ad un valore di **z = 2,39** corrisponde una probabilità pari a circa 0,017 equivalente a 1,7%. E' una probabilità bassa, inferiore al limite del 5%; quindi, si rifiuta l'ipotesi nulla (tra i due

tossici esistono solo differenze casuali) e si accetta l'ipotesi alternativa dell'esistenza di una differenza significativa tra gli effetti letali dei due tossici.

Senza correzione di Yates, il valore di  $z$  sarebbe stato uguale a 2,69; per un test bilaterale, si sarebbe avuta una probabilità uguale a 0,7%.

**Senza correzione, in campioni di piccole dimensioni è più facile rifiutare l'ipotesi nulla. Ma le conclusioni sono viziate da una distorsione eccessiva della distribuzione normale standardizzata; le conclusioni raggiunte con il test possono essere discusse e rifiutate, in quanto il test non rispetta le condizioni di validità. E' conveniente utilizzare la formula con la correzione, che è più prudentiale.**

### **3.6. CONFRONTO TRA TEST $\chi^2$ PER TABELLE 2 X 2 E TEST Z, SENZA E CON LE VARIE CORREZIONI PER LA CONTINUITA'**

**Il test z ed il test  $\chi^2_{(1)}$  per tabelle 2 x 2 possono essere applicati negli stessi casi.**

Il confronto tra due percentuali con il **test z** ed il confronto tra due distribuzioni osservate in tabelle di contingenza 2 x 2 con il **test  $\chi^2$**  forniscono la stessa risposta. Tra test **t** e test **c<sup>2</sup>** sono differenti la **presentazione dei dati ed ovviamente i valori calcolati; ma le ipotesi da verificare sono le stesse e dovrebbero coincidere anche le probabilità stimate**, a meno delle approssimazioni dei calcoli nei due differenti metodi. Le probabilità stimate divergono quando i campioni hanno poche osservazioni; ma in questi casi, **con campioni piccoli, è in discussione la validità dei due test.**

Come è possibile dedurre osservando la formula di Pearson, la distribuzione dei valori **c<sup>2</sup>** con 1 g.d.l. è la distribuzione di quadrati di variabili casuali normali standardizzate, dove la standardizzazione è ottenuta dividendo la differenza (al quadrato) tra osservato ed atteso per la frequenza assoluta attesa. Si ottiene un rapporto che, come tale, è indipendente dal numero assoluto di osservazioni.

In termini matematici, si può scrivere che

$$z^2 @ c^2_{(1)}$$

Per **n** variabili casuali normali standardizzate ed indipendenti, come sono i gruppi a confronto in un test chi quadrato, la somma dei quadrati segue la distribuzione  $\chi^2$  con **n** gradi di libertà, secondo la relazione

$$\sum_{i=1}^n z_i^2 = c^2_{(n)}$$

ESEMPIO 1. I popolamenti zooplanctonici dei laghi artificiali sono in prevalenza composti da Cladoceri e Rotiferi; mentre nei grandi laghi naturali predominano i Copepodi.

Cladoceri e Rotiferi trovano nei laghi artificiali, di recente formazione e con instabilità idrologica, le condizioni migliori per sfruttare la loro strategia r di colonizzazione; i Copepodi, con la loro strategia k, sono avvantaggiati in ambienti caratterizzati da elevata stabilità, come i grandi laghi naturali.

In due campioni di popolamenti zooplanctonici sono stati contati i Copepodi ed insieme Cladoceri e Rotiferi.

Nel campione 1 i Copepodi erano 31, mentre Cladoceri e Rotiferi insieme erano 154; nel campione 2 sono risultati presenti 39 Copepodi contro 110 Cladoceri e Rotiferi.

Si può affermare che i Copepodi siano più facilmente associati ad uno dei due laghi?

Risposta.

E' un test bilaterale (d'altronde con il test  $\chi^2$  sono possibili solo test bilaterali).

L'ipotesi nulla è  $H_0: p_1 = p_2$

e l'ipotesi alternativa è  $H_1: p_1 \neq p_2$

**Per applicare il test  $c^2$  con 1 g.d.l., è utile presentare i dati in una tabella di contingenza 2 x 2, completando le informazioni sui dati. Spesso alcuni di essi non sono espressamente riportati nel testo del problema, in quanto inutili alla comprensione dell'argomento e facilmente deducibili dagli altri. Per il calcolo è indispensabile che siano riportati tutti con la stessa evidenza.**

Con i dati presentati, la tabella 2 x 2 diviene

	Copepodi	Cladoceri e Rotiferi	Totale
Lago 1	<b>31</b>	<b>154</b>	185
Lago 2	<b>39</b>	<b>110</b>	149
Totale	70	264	<b>334</b>

E' un campione di grandi dimensioni e conviene utilizzare la formula generale per il calcolo rapido,

$$c_{(1)}^2 = \frac{(a \cdot d - b \cdot c)^2 \cdot N}{n_1 \cdot n_2 \cdot n_3 \cdot n_4}$$

senza correzione per la continuità

$$c_{(1)}^2 = \frac{(31 \cdot 110 - 154 \cdot 39)^2 \cdot 334}{185 \cdot 149 \cdot 70 \cdot 264} = \frac{(3410 - 6006)^2 \cdot 334}{509401200} = \frac{6739216 \cdot 334}{509401200} = \frac{2250898144}{509401200} = 4,42$$

dal quale risulta un  $\chi^2_{(1)}$  uguale a 4,42.

Nella tabella sinottica del  $\chi^2_{(1)}$  ad un valore di 4,42 corrisponde **una probabilità compresa tra 0.05** (il cui valore critico esatto è 3,84) e **0.025** (il cui valore critico esatto è 5,02).

Per lo stesso problema è possibile utilizzare **il test z**, ricorrendo alla formula generale

$$z = \frac{p_1 - p_2}{\sqrt{p^*(1-p^*) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

con i dati richiesti;

dove:  $p_1 = 0,168$ ;  $p_2 = 0,262$ ;  $n_1 = 185$ ;  $n_2 = 149$ ;

$m_1 = 31$ ;  $m_2 = 39$ ;

$$p^* = \frac{31 + 39}{185 + 149} = 0,210$$

Il valore di **z**, ottenuto dal calcolo,

$$z = \frac{0,168 - 0,262}{\sqrt{0,21 \cdot 0,79 \cdot \left(\frac{1}{185} + \frac{1}{149}\right)}} = \frac{-0,094}{\sqrt{0,1659 \cdot 0,01166}} = \frac{-0,094}{0,0448} = -2,10$$

risulta uguale a -2,10. Nella tabella della distribuzione normale standardizzata ad una coda, al valore 2,10 in una coda corrisponde una probabilità uguale a 0.018; **per un test bilaterale, la probabilità è uguale a 0.036** (quindi coincidente con la risposta del chi quadrato che dava una probabilità compresa tra 5% e 2,5%).

E' una misura più precisa di quella possibile con il test chi quadrato, soltanto perché nella tabella sinottica (riassuntiva) dei valori critici del chi quadrato sono riportati solamente alcuni di quelli ritenuti più importanti. Esistono tante distribuzioni  $\chi^2$  quanti sono i gradi di libertà; servirebbero quindi tante pagine, quanti sono i gradi di libertà.

Di conseguenza, si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa. Secondo l'interpretazione ecologica del risultato, la presenza percentuale di Copepodi (e simmetricamente di Cladoceri più Rotiferi) nei due laghi è significativamente differente.

Con il test  $z$ , il segno negativo della differenza indica che è maggiore la presenza di Cladoceri nel lago 2. Con il test chi quadrato, per evidenziare le frequenze osservate che hanno determinato la significatività è utile costruire la tabella delle frequenze attese.

**Con il test  $z$  è stata posta l'attenzione sulla proporzione dei Copepodi. Si fosse impostato il problema sulla proporzione di Cladoceri e Rotiferi, il risultato sarebbe stato identico: identica sarebbe stata la differenza tra le due proporzioni, riportata al numeratore; identico sarebbe stato l'errore standard della differenza, riportato al denominatore.**

**Come verifica della relazione**

$$z^2 @ c^2_{(1)}$$

anche con i dati di questo esempio è **possibile osservare che** il valore del test  $z$  (uguale a - 2,14) elevato al quadrato ( $-2,14^2 = 4,58$ ) è approssimativamente uguale al valore calcolato con il test  $c^2_{(1)}$  (uguale a 4,63)

$$- 2,10^2 @ 4,42$$

ESEMPIO 2. In una seconda serie di conteggi, nel campione pescato nel lago 1 sono stati classificati 6 Copepodi e 34 tra Cladoceri e Rotiferi, mentre nel campione del lago 2 sono stati classificati 10 Copepodi e 19 tra Cladoceri e Rotiferi.

Si può sostenere che la percentuale di Copepodi, oppure quella di Cladoceri e Rotiferi insieme, riscontrata nei due laghi sia significativamente differente?

Risposta.

E' un test bilaterale, dove l'ipotesi nulla è  $H_0: p_1 = p_2$  e l'ipotesi alternativa è  $H_1: p_1 \neq p_2$

Riportati in una tabella 2 x 2, i dati dei due campioni a confronto sono:

	Copepodi	Cladoceri e Rotiferi	Totale
Lago 1	6	34	40
Lago 2	10	19	29
Totale	16	53	69

Il numero totale di osservazioni è 69: non è un campione di grandi dimensioni e si richiede la correzione di Yates. La formula con la quale abbreviare i tempi necessari per i calcoli manuali è

$$c_{(1)}^2 = \frac{\left( |a \cdot d - b \cdot c| - \frac{N}{2} \right)^2 \cdot N}{n_1 \cdot n_2 \cdot n_3 \cdot n_4}$$

Con i dati del problema,

$$c_{(1)}^2 = \frac{\left( |6 \cdot 19 - 34 \cdot 10| - \frac{69}{2} \right)^2 \cdot 69}{40 \cdot 29 \cdot 16 \cdot 53} = \frac{(|114 - 340| - 34,5)^2 \cdot 69}{983680} = \frac{36672,25 \cdot 69}{983680} = \frac{2530385,25}{983680} = 2,57$$

fornisce un valore di  $\chi^2_{(1)}$  uguale a 2,57.

Nella tabella sinottica del  $\chi^2_{(1)}$  al valore 2,57 **corrisponde una probabilità compresa tra 0.10 ( il cui valore critico è 2,706) e 0.25 (il cui valore critico è 1,323)**; purtroppo, è una stima molto approssimata, a causa dei pochi valori critici di norma riportati nelle tavole sinottiche.

Per il medesimo problema, con gli stessi dati è possibile ricorrere al **test z** per il confronto tra percentuali, ricorrendo alla formula **con la correzione per la continuità di Yates**

$$z = \frac{|p_1 - p_2| - \frac{1}{2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{p^* (1 - p^*) \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

dove:

$$p_1 = 0,150; \quad p_2 = 0,345; \quad n_1 = 40; \quad n_2 = 29;$$

$$m_1 = 6; \quad m_2 = 10; \quad p^* = \frac{6 + 10}{40 + 29} = 0,232$$

Il valore di **z**

$$z = \frac{|0,150 - 0,345| - \frac{1}{2} \left( \frac{1}{40} + \frac{1}{29} \right)}{\sqrt{0,23 \cdot 0,77 \cdot \left( \frac{1}{40} + \frac{1}{29} \right)}} = \frac{0,195 - 0,03}{\sqrt{0,1771 \cdot 0,0594}} = \frac{0,165}{0,1026} = 1,608$$

risulta uguale a 1,608.

Nella tabella della distribuzione normale standardizzata ad una coda, il valore 1,61 (arrotondamento di 1,608) esclude una densità di probabilità uguale a 0.055; per un test bilaterale, la probabilità è uguale a 0.11 (11,0%).

E' semplice osservare che 1.608 è approssimativamente uguale a 2.58: il test  $c^2$ , con un grado di libertà, ed il test  $z$ , per il confronto tra due percentuali, forniscono risultati molto simili, nonostante le approssimazioni nei calcoli determinate dalle due diverse formule e le distorsioni determinate dal numero ridotto di dati.

Comunque, la probabilità stimata con i due calcoli è alta, nettamente superiore al 5%. Pertanto, la differenza non è significativa e non si può rifiutare l'ipotesi nulla: la percentuale di Copepodi (e simmetricamente di Cladoceri e Rotiferi) riscontrate nei due laghi non è significativamente diversa e differisce solo per fattori casuali.

Per meglio comprendere i concetti fondamentali dell'inferenza statistica, è importante osservare come, **con i dati di questo secondo esercizio, la differenza tra le due percentuali sia maggiore rispetto a quella del primo esercizio ( $p_1-p_2 = 0,126$  rispetto a  $0,094$ )**; eppure, diversamente dal caso precedente, non si è in grado di rifiutare l'ipotesi nulla. La causa è il più ridotto numero di osservazioni, per cui le variazioni casuali sono notevolmente maggiori. **Con pochi dati, il test è poco potente** (la potenza del test sarà discussa nel successivo capitolo 4): **non si è in grado di rifiutare l'ipotesi nulla, anche quando è evidentemente falsa.**

**Non essere in grado di rifiutare l'ipotesi nulla non significa che le due percentuali siano uguali.** Se si fosse convinti della reale esistenza di una differenza tra le due percentuali, **per rendere il test significativo sarebbe sufficiente aumentare il numero di osservazioni. Quando il numero di osservazioni è molto grande, risultano significative anche differenze molto piccole.**

### 3.7. CONFRONTO DI UNA PROPORZIONE OSSERVATA CON UNA ATTESA:

#### IL TEST Z PER GRANDI CAMPIONI;

#### LA DISTRIBUZIONE BINOMIALE PER PICCOLI CAMPIONI

La distribuzione  $z$  (presentata nel capitolo 2) permette anche il confronto tra la proporzione osservata in un singolo esperimento e la corrispondente proporzione attesa o teorica.

La formula può essere derivata da quella già utilizzata per **la distribuzione di una osservazione campionaria  $x$  rispetto alla media della popolazione  $m$ , quando sia nota la varianza  $s^2$  della popolazione, attraverso la relazione**

$$z = \frac{\bar{x} - m}{\sqrt{s^2}}$$

poiché **la varianza di una proporzione è totalmente definita dal suo valore medio  $p$  e dal numero totale di osservazioni** essendo  $s^2 = n \times p \times (1-p)$

**Nel caso di una proporzione, il test  $z$  diventa**

$$z = \frac{x - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}}$$

ricordando che:

$p$  = proporzione attesa o teorica,

$n$  = numero totale di osservazioni o dati dell'esperimento;

$x$  = numero di individui osservati con la caratteristica in esame;

$n \cdot p$  = numero atteso di individui con la caratteristica in esame.

**Nel test  $z$ , la distribuzione delle probabilità è simmetrica ed il risultato evidenzia se la differenza è positiva oppure negativa. L'ipotesi alternativa  $H_1$  può essere non solo bilaterale ma anche unilaterale.**

ESEMPIO 1. In un esperimento di incroci nel pisello, da una pianta eterozigote sono state ottenuti 65 semi lisci e 15 semi rugosi. Secondo le leggi di Mendel, il rapporto atteso di  $3/4$  e  $1/4$  avrebbe dovuto dare 60 semi della prima caratteristica e 20 della seconda.

Il numero di semi lisci si discosta dal valore atteso di  $3/4$ ? Questa distribuzione sperimentale segue la legge di segregazione di Mendel?

Risposta.

L'**ipotesi nulla  $H_0$**  afferma che il numero osservato (65) è in sostanziale accordo con quello atteso ( $60 = 80 \times 3/4$ ), stimato sulla base della legge di Mendel; la differenza riscontrata tra frequenza osservata e frequenza attesa è un fenomeno altamente probabile, imputabile solamente a fattori casuali.

L'**ipotesi alternativa  $H_1$**  afferma che il numero osservato si discosta molto da quello atteso; la differenza riscontrata non è dovuta a fattori casuali, ma al fatto che la distribuzione osservata non segue la legge di Mendel. **E' un test bilaterale.**

La probabilità complessiva di trovare scarti tra osservato ed atteso che siano uguali o superiori a quello riscontrato nell'esperimento può essere calcolata mediante il test  $z$

$$z = \frac{x - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}}$$

dove:  $x = 65$ ;  $n = 80$ ;  $p = 3/4$ .

Con i dati dell'esempio, si può stimare il valore di  $z$ ,

$$Z = \frac{65 - (80 \cdot \frac{3}{4})}{\sqrt{80 \cdot \frac{3}{4} \cdot \frac{1}{4}}} = \frac{5}{3,87} = 1,29$$

che risulta uguale a 1,29.

Il valore calcolato di  $z$  è inferiore a quello tabulato per la distribuzione normale alla probabilità del 5% (1,96), in un test a due code. Di conseguenza, non si può rifiutare l'ipotesi nulla: le differenze riscontrate tra osservato ed atteso sono imputabili solamente a variazioni casuali. La proporzione osservata ( $65/80 = 0,8125$ ) è in sostanziale accordo con la proporzione di  $3/4$  della legge di Mendel.

**La domanda avrebbe potuto essere impostata sulla frequenza del fenomeno alternativo**, quella del numero di semi rugosi, che sono risultati 15 su 80 ( $15/80 = 0,1875$ ) quando la proporzione attesa sarebbe stata di  $1/4$ .

**Il valore ottenuto con il test  $z$  sarebbe stato identico, ovviamente con segno opposto**, come mostra il calcolo

$$z = \frac{15 - (80 \cdot \frac{1}{4})}{\sqrt{80 \cdot \frac{1}{4} \cdot \frac{3}{4}}} = \frac{-5}{3,87} = -1,29$$

ESEMPIO 2. Per il materiale dei loro esperimenti a carattere biologico con cavie, semi o uova, l'ecologo e l'ambientalista spesso si riforniscono da ditte specializzate. Queste, a testimonianza della qualità del loro prodotto, assicurano la sopravvivenza di una certa percentuale o proporzione di individui per un periodo definito, in condizioni standard.

Da 150 semi, di cui era stata garantita una riuscita del 90%, ad un ricercatore sono spuntate 122 pianticelle. E' una percentuale di successi di poco superiore a 81%, contro il 90% atteso.

Si può sostenere che non esiste una differenza significativa con l'atteso oppure che il numero di pianticelle è effettivamente minore di quanto era stato assicurato?

Risposta.

E' un **test ad una coda**, in cui l'ipotesi nulla afferma che la differenza è trascurabile, mentre l'ipotesi alternativa sostiene che la percentuale trovata è significativamente minore di quella attesa.

Il test  $z$ , dove  $x = 122$ ;  $p = 0,9$ ;  $n = 150$

$$z = \frac{122 - (150 \cdot 0,9)}{\sqrt{150 \cdot 0,9 \cdot 0,1}} = \frac{122 - 135}{\sqrt{13,5}} = \frac{-13}{3,674} = -3,53$$

stima un valore uguale a -3,53. Nella tabella della distribuzione normale standardizzata, in una coda corrisponde ad una probabilità inferiore a .00023.

Si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa: la percentuale di riuscita è significativamente inferiore a quella attesa. Le pianticelle non hanno avuto la riuscita che era stata garantita.

**Il test  $\chi^2$  ed il test z non sono gli unici metodi per testare la significatività di una differenza tra proporzioni, quando il campione non è piccolo. E' anche possibile usare i limiti di confidenza di una proporzione** (presentati nel capitolo 4).

**Nel caso di una frequenza relativa con campioni piccoli, come quando  $np$  oppure  $n(1-p)$  è inferiore a 5 (e a maggior ragione quando entrambi sono inferiori a 5), occorre risolvere il problema in modo esatto con la distribuzione binomiale.**

La distribuzione binomiale (presentata nel capitolo 2)

$$P_i = C_i^n \cdot p^i \cdot q^{n-i}$$

permette di stimare **la probabilità esatta di avere la risposta  $i$  in oggetto.**

A differenza del test **z**, che fornisce un valore al quale nella distribuzione normale ridotta è associata la probabilità di avere valori uguali o superiori a quello calcolato, **con la distribuzione binomiale si ottiene direttamente la probabilità che si realizzi una specifica risposta.** Per rifiutare l'ipotesi nulla, **questa probabilità deve essere cumulata con tutte quelle associate alle possibili risposte più estreme.**

Quando si ha un **test unilaterale**, per ottenere la probabilità complessiva che permette di rifiutare l'ipotesi nulla occorre **sommare la probabilità associata all'evento osservato con tutti i possibili eventi più estremi nella stessa direzione.**

Quando si ha un **test bilaterale**, occorre **moltiplicare per 2 il valore precedentemente calcolato**, perché le possibili risposte più estreme sono collocate in entrambe le direzioni.

ESEMPIO. E' possibile fornire una prima stima dell'attività biologica di una sostanza, mediante test di tossicità acuta condotti in tempi brevi, di norma tra le 24 e le 96 ore di esposizione. Si ottengono informazioni sugli effetti tossici immediati di immissioni, più o meno accidentali, in ambiente acquatico. Le sostanze impiegate in questi test vanno dai pesticidi a quelle contenute in rifiuti industriali e domestici. Le specie animali utilizzate vanno dai pesci agli invertebrati, mentre le specie vegetali utilizzate di solito sono microalghe.

Con numerosi esperimenti, è stato dimostrato che le immissioni in ambiente acquatico di un corso d'acqua determina, in esperimenti di laboratorio, la mortalità del 40% dei pesci dopo una esposizione di 48 ore.

Nell'ultimo monitoraggio, le sostanze immesse determinavano un colore del corso d'acqua più vistoso ed un odore più acuto, lasciando supporre una maggiore concentrazione di sostanze tossiche. L'esperimento di laboratorio ha determinato la mortalità di 8 animali sui 9 esposti al rischio.

Si può statisticamente sostenere che le ultime emissioni contengono sostanze con un effetto tossico significativamente superiore a quello di norma riscontrato?

Risposta.

E' un **test ad 1 coda**; ma **l'ipotesi deve essere formulata a priori**, rispetto ai risultati dell'esperimento. Non è corretto guardare i risultati dell'esperimento e chiedersi se è significativamente superiore alla media, solamente dopo aver osservato che la percentuale di letalità supera il risultato medio.

Nell'esempio, **l'ipotesi nulla  $H_0$**  afferma che la percentuale di letalità riscontrata nell'ultima rilevazione ( $8/9 = 0.89$ ) è una variazione casuale della percentuale di solito riscontrata, uguale a 0.4; **l'ipotesi alternativa  $H_1$**  afferma che la letalità delle sostanze campionate nell'ultimo monitoraggio è significativamente superiore alla media.

La scelta tra le due ipotesi dipende dalla probabilità complessiva stimata.

Per calcolare la probabilità complessiva che permetta di decidere tra le due ipotesi, dapprima si deve calcolare la probabilità di ottenere la risposta osservata di 8 decessi ( $P_8$ ), con i dati del problema, dove  $n = 9$ ;  $i = 8$ ;  $p = 0.4$

$$P_8 = C_8^9 \cdot 0.4^8 \cdot 0.6^1 = 0.00354$$

ma essa non è la risposta più estrema. Nella stessa direzione, si sarebbe avuto un effetto ancora più letale, se fossero morti tutti 9 gli animali utilizzati nell'esperimento.

Con  $n = 9$ ;  $i = 9$ ;  $p = 0.4$ , la probabilità di ottenere per caso anche questo risultato ( $P_9$ ) è

$$P_9 = C_9^9 \cdot 0.4^9 \cdot 0.6^0 = 0.00026$$

La probabilità complessiva di ottenere la risposta osservata oppure una risposta più estrema nella stessa direzione è data dalla somma di queste due singole probabilità

$$P_{8+9} = 0.00354 + 0.00026 = \mathbf{0.0038} \text{ (0,38\%)}$$

Una probabilità uguale a 0,38% deve essere considerata molto piccola: si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa.

Come indicavano il colore più evidente e l'odore più sgradevole, la letalità (0.89 con 9 osservazioni) delle sostanze tossiche presenti nell'immissione è significativamente più alta della letalità media (0.4) degli esperimenti precedenti.

Se il quesito avesse avuto un'ipotesi alternativa bilaterale, per ottenere la probabilità relativa si sarebbe dovuto moltiplicare questa probabilità per 2.

### **3.8. TABELLE DI CONTINGENZA 2 X 2 IN PICCOLI CAMPIONI: IL METODO ESATTO DI FISHER**

**Il  $\chi^2$  derivato dalla distribuzione normale, è valido solo per grandi campioni. Se il numero di frequenze attese è piccolo, nel caso di tabelle 2 x 2 si deve ricorrere al metodo esatto di Fisher, derivato dalla distribuzione ipergeometrica, come nel caso di una sola proporzione si ricorre alla distribuzione binomiale.**

Per passare da indicazioni di principio a raccomandazioni pratiche sulla scelta appropriata del test, è consigliato utilizzare il metodo esatto di Fisher in sostituzione del chi quadrato quando

- il campione ha un numero totale di osservazioni inferiore a circa 30,
- e/o almeno una frequenza attesa è inferiore a 5.

Sono criteri identici alle raccomandazioni precedenti, che consigliavano di evitare l'uso del  $\chi^2$  quando il valore di  $n \times p$  oppure quello di  $n \times (1-p)$  sono inferiori a 5.

**Il metodo delle probabilità esatte di Fisher, dal nome del suo proponente in un articolo del 1935** (Fisher R. A. *The logic of scientific inference*, su *J. Roy. Stat. Soc.*, 98, pp. 39-54), è di estrema utilità sotto l'aspetto didattico, perché spiega con chiarezza la logica dell'inferenza statistica.

L'uso di questo metodo richiede l'impiego dei fattoriali; di conseguenza, è di semplice e rapida applicazione solo quando il numero di osservazioni è molto piccolo. Il metodo potrebbe essere applicato anche nel caso di campioni di dimensioni medie; ma con un numero più alto di dati, diviene possibile stimare la probabilità solamente con l'uso di calcolatori.

**Il metodo permette di stimare la specifica probabilità (Pi) di ottenere una tabella 2 x 2 uguale a quella osservata.**

Usando la medesima simbologia dei precedenti paragrafi 4 e 6, riportata nella tabella seguente

	Risposta X	Risposta x	Totale
Campione Y	<b>a</b>	<b>b</b>	<b>n<sub>1</sub> = a + b</b>
Campione y	<b>c</b>	<b>d</b>	<b>n<sub>2</sub> = c + d</b>
Totale	<b>n<sub>3</sub> = a + c</b>	<b>n<sub>4</sub> = a + d</b>	<b>N = a+b+c+d</b>

con la distribuzione ipergeometrica la probabilità **Pi** è calcolata con la formula

$$Pi = \frac{C_{a+c}^a \cdot C_{b+d}^b}{C_N^{a+b}} = \frac{\frac{(a+c)!}{a! \cdot c!} \cdot \frac{(b+d)!}{b! \cdot d!}}{\frac{N!}{(a+b)! \cdot (c+d)!}} = \frac{(a+b)! \cdot (c+d)! \cdot (a+c)! \cdot (b+d)!}{a! \cdot b! \cdot c! \cdot d! \cdot N!}$$

In modo più rapido, lo stesso risultato è ottenuta con

$$Pi = \frac{n_1! \cdot n_2! \cdot n_3! \cdot n_4!}{a! \cdot b! \cdot c! \cdot d! \cdot N!}$$

Con questa **formula abbreviata**, (abbrevia i tempi richiesti dal calcolo manuale) la probabilità (Pi) di trovare quel particolare insieme dei dati osservati è determinata dal rapporto tra il prodotto dei fattoriali dei quattro totali marginali ed il prodotto dei fattoriali delle quattro frequenze osservate moltiplicato il numero totale di osservazioni.

**Il metodo di Fisher si fonda sul concetto che, tenendo fissi i totali, i numeri riportati nelle 4 caselle possano assumere per caso qualsiasi valore.** Sulla base di questo presupposto, si può calcolare la probabilità di ottenere ognuna delle risposte possibili.

Per stabilire se esiste una differenza significativa tra le due distribuzioni osservate dei campioni Y e y, non è sufficiente calcolare la probabilità della distribuzione osservata. Come con la precedente distribuzione binomiale, nel caso di metodi esatti **si deve stimare la probabilità totale di osservare una combinazione di dati così estrema oppure più estrema.**

A questo fine, si riduce di 1 il numero di osservazioni nella casella con il numero minore, modificando i valori delle altre caselle per mantenere uguali i totali marginali; successivamente, si calcola la probabilità di ottenere ognuna di queste risposte. E' necessario elencare tutte le possibili combinazioni delle osservazioni più estreme e quindi calcolare le probabilità esatte associate ad ognuna di queste possibili combinazione dei dati.

Per poter decidere tra le due ipotesi, **la probabilità che occorre stimare è data dalla somma della probabilità della distribuzione osservata e di quelle delle risposte più estreme nella stessa direzione.**

**La probabilità così stimata corrisponde ad un test ad una coda; per un test a due code, si deve moltiplicare per due questa probabilità.**

In modo più dettagliato, i passaggi per calcolare la probabilità che permette di rifiutare l'ipotesi nulla sono:

- 1 - calcolare la probabilità associata ai dati osservati;
- 2 - individuare la casella con il numero minore; se è zero, è sufficiente questa probabilità, perché la risposta osservata è quella più estrema;
- 3 - se è diverso da zero, ridurre il valore di 1, modificando le frequenze nelle altre tre caselle, in modo che i totali marginali (e quindi quello totale) restino immutati;
- 4 - calcolare la probabilità associata alla nuova tabella;
- 5 - ripetere le operazioni 3 e 4, finché il valore minore diventa zero;
- 6 - per un test ad una coda, sommare tutte queste probabilità;
- 7 - per un test a due code, moltiplicare per 2 il risultato della precedente operazione 6;
- 8 - se la probabilità totale calcolata è inferiore al valore di probabilità prefissato come limite critico (di solito 0,05), si rifiuta l'ipotesi nulla  $H_0$  ed implicitamente si accetta l'ipotesi alternativa  $H_1$ , che può essere sia bilaterale che unilaterale.

ESEMPIO. Su un numero ridotto di cavie, si sono sperimentati due pesticidi per verificare se quello di nuova produzione (B) ha effetti più letali di quello usato in precedenza (A).

Il risultato dell'esperimento è riportato nella tabella sottostante:

DATI OSSERVATI	Animali sopravvissuti	Animali morti	Totale
Pesticida A	<b>7</b>	<b>1</b>	8
Pesticida B	<b>3</b>	<b>6</b>	9
Totale	10	7	17

Il pesticida B, di nuova produzione, è più efficace del precedente pesticida A?

Risposta.

Si tratta di un **test ad una coda**.

La probabilità di avere per caso la risposta osservata nell'esperimento è

$$P_i = \frac{8! \cdot 9! \cdot 10! \cdot 7!}{7! \cdot 1! \cdot 3! \cdot 6! \cdot 17!} = 0.03455$$

Ma la risposta osservata non è quella più estrema, nell'ipotesi che B sia più letale di A. E' semplice osservare che è possibile un'altra risposta più estrema: quella in cui con il pesticida A sopravvivano tutti 8 gli animali e di conseguenza, per mantenere fissi i totali come richiede il metodo proposto da Fisher, con il pesticida B sopravvivano non 3 ma solo 2 animali sui 9 del campione.

Le frequenze previste in una tabella 2 x 2 che evidenzi una differenza maggiore tra gli effetti del pesticida A rispetto a quelli del pesticida B sono riportate nella tabella seguente:

RISPOSTA PIU' ESTREMA	Animali sopravvissuti	Animali morti	Totale
Pesticida A	<b>8</b>	<b>0</b>	8
Pesticida B	<b>2</b>	<b>7</b>	9
Totale	10	7	17

La probabilità di avere per caso questa possibile risposta, mantenendo fissi i totali, è data da

$$P_i = \frac{8! \cdot 9! \cdot 10! \cdot 7!}{8! \cdot 0! \cdot 2! \cdot 7! \cdot 17!} = 0.00185$$

e risulta uguale a 0.00185.

La probabilità totale di ottenere la risposta osservata oppure una di quelle più estreme (in questo caso solo una è più estrema) nella stessa direzione è data dalla somma dei due valori già calcolati:

$$P = 0.03455 + 0.00185 = 0.0364$$

La probabilità complessiva 0.0364 (o 3,64%), stimata nella condizione che sia vera l'ipotesi nulla (i valori della risposta sono determinati dal caso, poiché non esiste una differenza reale nella letalità dei due pesticidi) è bassa, inferiore al 5%. Di conseguenza, si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa: il nuovo pesticida B è più letale del pesticida A.

Se la domanda fosse stata se esiste differenza tra l'effetto dei due pesticidi, si dovevano ipotizzare come possibili anche le risposte in cui il numero di morti fosse più alto con il pesticida A, collocate quindi nell'altra coda della distribuzione. Ma spesso, come avviene con i dati di questo esercizio, le risposte non sono simmetriche. Per questo motivo, **nei test a due code la probabilità complessiva è ottenuta moltiplicando per 2 quella precedentemente calcolata ad una coda:**

$$P = 0.0364 \times 2 = 0.0728$$

La probabilità 0.0728 (o 7,28%) sarebbe stata superiore al 5%; di conseguenza, non si sarebbe rifiutata l'ipotesi nulla.

**Anche in questo caso, si dimostra come il test ad una coda sia più potente del test a due code: è maggiore la probabilità di rifiutare l'ipotesi nulla, quando essa è falsa.**

Un modo semplice, ma più lungo, **per capire esattamente il funzionamento del test di Fisher è quello di verificare tutte le possibili risposte**, tenendo costanti i totali marginali. Individuata la frequenza osservata minore, occorre sostituire ad essa il valore 0 e simmetricamente le altre 3; successivamente si deve aumentare progressivamente quel valore di 1 unità, fino a quando compare 0 in un'altra casella. A quel punto non sono più possibili altre risposte.

Con i dati dell'esempio, è possibile osservare che le risposte differenti che si sarebbero potute ottenere sono le 8 seguenti

(1)	(2)	(3)	(4)																
<table border="1" style="border-collapse: collapse; width: 40px; height: 40px;"> <tr><td style="padding: 5px;">8</td><td style="padding: 5px;">0</td></tr> <tr><td style="padding: 5px;">2</td><td style="padding: 5px;">7</td></tr> </table>	8	0	2	7	<table border="1" style="border-collapse: collapse; width: 40px; height: 40px;"> <tr><td style="padding: 5px;">7</td><td style="padding: 5px;">1</td></tr> <tr><td style="padding: 5px;">3</td><td style="padding: 5px;">6</td></tr> </table>	7	1	3	6	<table border="1" style="border-collapse: collapse; width: 40px; height: 40px;"> <tr><td style="padding: 5px;">6</td><td style="padding: 5px;">2</td></tr> <tr><td style="padding: 5px;">4</td><td style="padding: 5px;">5</td></tr> </table>	6	2	4	5	<table border="1" style="border-collapse: collapse; width: 40px; height: 40px;"> <tr><td style="padding: 5px;">5</td><td style="padding: 5px;">3</td></tr> <tr><td style="padding: 5px;">5</td><td style="padding: 5px;">4</td></tr> </table>	5	3	5	4
8	0																		
2	7																		
7	1																		
3	6																		
6	2																		
4	5																		
5	3																		
5	4																		
(5)	(6)	(7)	(8)																
<table border="1" style="border-collapse: collapse; width: 40px; height: 40px;"> <tr><td style="padding: 5px;">4</td><td style="padding: 5px;">4</td></tr> <tr><td style="padding: 5px;">6</td><td style="padding: 5px;">3</td></tr> </table>	4	4	6	3	<table border="1" style="border-collapse: collapse; width: 40px; height: 40px;"> <tr><td style="padding: 5px;">3</td><td style="padding: 5px;">5</td></tr> <tr><td style="padding: 5px;">7</td><td style="padding: 5px;">2</td></tr> </table>	3	5	7	2	<table border="1" style="border-collapse: collapse; width: 40px; height: 40px;"> <tr><td style="padding: 5px;">2</td><td style="padding: 5px;">6</td></tr> <tr><td style="padding: 5px;">8</td><td style="padding: 5px;">1</td></tr> </table>	2	6	8	1	<table border="1" style="border-collapse: collapse; width: 40px; height: 40px;"> <tr><td style="padding: 5px;">1</td><td style="padding: 5px;">7</td></tr> <tr><td style="padding: 5px;">9</td><td style="padding: 5px;">0</td></tr> </table>	1	7	9	0
4	4																		
6	3																		
3	5																		
7	2																		
2	6																		
8	1																		
1	7																		
9	0																		

e che non esistono altri possibili risultati, che diano gli stessi totali di riga e di colonna.

Con la formula del metodo esatto di Fisher, è possibile **calcolare le probabilità di ottenere ognuna di queste risposte teoricamente possibili**; ovviamente, se non sono stati commessi errori od

approssimazioni eccessive, **il loro totale darà 1 se espresso in valore unitario**, oppure 100 se espresso in percentuale.

Dalla prima all'ottava possibile risposta si passa da un estremo di una maggiore letalità del pesticida B (con B si hanno 7 morti su 9, mentre con A sono sopravvissuti tutti 8), all'altro estremo di un effetto maggiore del pesticida A (con A ne muoiono 7 su 8, mentre con B sopravvivono tutti 9).

Per poter stabilire se esiste una differenza significativa tra i due pesticidi, occorre sommare alla probabilità calcolata per la risposta 2 (che coincide con quella sperimentale) la probabilità di ottenere le risposte più estreme nella stessa direzione (nell'esempio è una sola, la risposta 1). Se la loro somma supera il 5%, non è possibile rifiutare l'ipotesi nulla. Essa sarà rifiutata quando la somma delle probabilità di ottenere la risposta in oggetto o quelle più estreme non supera il valore prefissato.

Le probabilità complessive calcolate con il metodo esatto di Fisher possono essere estese in una sola direzione, per test ad una coda; ma possono essere estese ad ambedue le direzioni, per test a due code.

In questo ultimo caso, se il campione fosse di grandi dimensioni la probabilità complessiva coinciderebbe con quanto è possibile calcolare con il test  $\chi^2$ , che è un test a due code.

**Considerare fissi i totali marginali appare ad alcuni autori una scelta impropria ed arbitraria.** Sono state proposte **modifiche al metodo di Fisher, come quella di Tocher.** Queste modifiche in questo manuale sono tralasciate, perché **nella ricerca ambientale quasi sempre il metodo di Fisher è usato come proposto originariamente dall'autore**, che ne raccomandò l'applicazione a tutti i dati dicotomici.

### **3.9. LE TABELLE 2 x N CON LA FORMULA GENERALE E QUELLA DI BRANDT-SNEDECOR. LE TABELLE M x N**

Il metodo del  $\chi^2$  per tabelle 2 x 2, con 1 grado di libertà, può essere esteso al caso generale di tabelle a due entrate, ognuna con classificazioni multiple anziché dicotomiche, con più gradi di libertà. **Con l'applicazione dei medesimi concetti ed il ricorso a formule analoghe, è possibile il confronto tra M popolazioni indipendenti, per verificare l'ipotesi nulla che tutte le N percentuali o proporzioni a confronto siano uguali.**

Sono le tabelle M x N in cui l'ipotesi nulla è

$$H_0: p_1 = p_2 = p_3 = \dots = p_M$$

e l'ipotesi alternativa è

$$H_1 = \text{almeno una delle } \pi \text{ è diversa dalle altre.}$$

Il caso più semplice di tabelle  $M \times N$  è la tabella di contingenza  $2 \times N$ , per **risposte dicotomiche di N gruppi a confronto**. Essa ha  **$N - 1$  gradi di libertà**, derivati dalla formula generale

$$(N-1) \times (2-1)$$

Anche in queste tabelle, è bene **evitare di avere caselle con frequenze teoriche od attese inferiori a 5**, per non avere una eccessiva perdita di potenza del test. Tuttavia, la tolleranza in merito a queste condizioni di validità diviene maggiore: si accettano frequenze attese di 1 o 2, oppure un numero più alto di frequenze uguali a 4-5, poiché le variazioni casuali tendono a compensarsi.

**Il  $\chi^2$  con parecchi gradi di libertà è meno sensibile agli errori determinati da frequenze attese piccole.**

ESEMPIO. Si vuole confrontare l'effetto di 5 pesticidi, dispersi in 5 aree diverse, sulla sopravvivenza di una stessa specie animale. I risultati ottenuti sono riportati nella tabella sottostante.

#### DISTRIBUZIONE OSSERVATA

	Pesticida A	Pesticida B	Pesticida C	Pesticida D	Pesticida E	Totale
Morti	8	10	14	11	7	50
Sopravvissuti	12	6	20	22	10	70
Totale	20	16	34	33	17	120

Esiste una differenza significativa tra le percentuali (o le proporzioni) di animali morti con i vari pesticidi sperimentati?

Risposta.

L'ipotesi nulla è che tutti 5 i pesticidi a confronto determinino la stessa frequenza percentuale  $\pi$  di animali morti  $H_0: p_1 = p_2 = p_3 = \dots = p_M$

L'ipotesi alternativa  $H_1$  è che almeno 1 di esse sia significativamente differente dalle altre.

Dopo la formulazione delle ipotesi, il primo passo è il calcolo delle frequenze attese, nella condizione che l'ipotesi nulla sia vera.

Dopo aver riportato i totali marginali e quello generale, è possibile calcolare la distribuzione attesa in ogni casella, con il prodotto

**totale colonna x totale riga / totale generale.**

ottenendo la tabella seguente

DISTRIBUZIONE ATTESA O TEORICA SECONDO L'IPOTESI NULLA

	Pesticida A	Pesticida B	Pesticida C	Pesticida D	Pesticida E	Totale
Morti	<b>8,33</b>	<b>6,67</b>	<b>14,17</b>	<b>13,75</b>	<b>7,08</b>	50
Sopravvissuti	<b>11,67</b>	<b>9,33</b>	<b>19,83</b>	<b>19,25</b>	<b>9,92</b>	70
Totale	20	16	34	33	17	120

Nel calcolo della distribuzione attesa, risulta evidente che **il numero di g. d. l. è (5-1) x (2-1) = 4**, poiché, dopo che sono stati fissati i totali, solo 4 valori sono liberi di assumere qualsiasi valore.

Successivamente, mediante la formula generale

$$c_{(g.d.l.)}^2 = \sum_{i=1}^{M \cdot N} \frac{(f_i^{oss} - f_i^{att})^2}{f_i^{att}}$$

si calcola il valore del  $\chi^2_{(4)}$ , estendendo la sommatoria a tutte 10 le caselle

$$c_{(4)}^2 = \frac{(8-8,33)^2}{8,33} + \frac{(10-6,67)^2}{6,67} + \dots + \frac{(10-9,92)^2}{9,92} = 3,9266$$

Il valore del  $c_{(4)}^2$  calcolato (3,926) è inferiore al valore critico (9,49) riportato nella tabella alla probabilità 0.05 e per 4 gradi di libertà.

Di conseguenza, non si può rifiutare l'ipotesi nulla: le differenze riscontrate tra valori osservati e valori attesi sono imputabili solo a variazioni casuali di campionamento.

In termini biologici, si afferma che la letalità dei 5 pesticidi a confronto non è significativamente diversa.

Anche per il calcolo del  $c^2$  in tabelle 2 x N sono stati proposti procedimenti abbreviati. Una formula frequentemente proposta nei testi di statistica applicata è quella di **Brandt e Snedecor**

$$c_{g.d.l.}^2 = \frac{C \cdot 100}{\bar{p} \cdot (1 - \bar{p})}$$

con C uguale a

$$C = \sum_{i=1}^k p_i \cdot n_i - \bar{p} \cdot \sum_{i=1}^k n_i$$

e dove

**k** = numeri di gruppi a confronto,

**p<sub>i</sub>** = frequenza percentuale del carattere in esame nel gruppo o campione i,

**n<sub>i</sub>** = frequenza assoluta del carattere in esame nel gruppo o campione i,

**N** = numero totale di osservazioni

$\bar{p}$  = frequenza percentuale media di tutti i gruppi per il carattere in esame

Applicata ai dati dell'esempio, è indispensabile calcolare dapprima le percentuali di ogni gruppo e la percentuale media totale, in riferimento ai morti oppure ai sopravvissuti. Nell'esempio, in analogia all'interpretazione precedente, è stata utilizzata la prima riga che riporta le frequenze degli animali morti

	Pesticida A	Pesticida B	Pesticida C	Pesticida D	Pesticida E	Totale
Morti $n_i$	8	10	14	11	7	50
$P_i$ in %	<b>40,0</b>	<b>62,5</b>	<b>41,2</b>	<b>33,3</b>	<b>41,2</b>	<b>41,66</b>
Sopravvissuti	12	6	20	22	10	70
Totale	20	16	34	33	17	120

ed in riferimento a questi dati sono calcolati i parametri richiesti dalla formula

$$\sum_{i=1}^5 p_i \cdot n_i = 40,0 \times 8 + 62,5 \times 10 + 41,2 \times 14 + 33,3 \times 11 + 41,2 \times 7 = 2176,5$$

$$\bar{p} = 41,66; \quad N = 120$$

$$\bar{p} \cdot \sum_{i=1}^5 n_i = 41,7 \times 50 = 2085$$

$$C = 2176,5 - 2085 = 91,5$$

Da essi si stima un valore del  $c_{(4)}^2$

$$c_{(4)}^2 = \frac{91,5 \times 100}{41,66 \times 58,34} = 3,765$$

che risulta uguale a 3,765

Si può osservare come, con la formula di Brandt e Snedecor, si ottenga un risultato (3,765) simile a quello della formula generale (3,9266), a meno delle approssimazioni necessarie nei calcoli.

La formula abbreviata semplifica i calcoli e riduce i tempi richiesti; ma per l'interpretazione del risultato è sempre utile disporre anche della distribuzione attesa, poiché i confronti tra le caselle con le frequenze assolute osservate e quelle corrispondenti con le frequenze, permettono di individuare le cause della significatività complessiva.

**Nel caso più generale di una tabella di contingenza  $M \times N$ , il  $\chi^2$  è più frequentemente utilizzato come test per l'indipendenza tra i caratteri riportati in riga madre (di norma, i Trattamenti) e quelli riportati nella prima colonna (le Categorie).** L'ipotesi nulla è che vi sia indipendenza tra tali variabili, mentre l'ipotesi alternativa bilaterale è che esista associazione.

Anche in questo caso, in molti test di statistica applicata è sconsigliato avere caselle con frequenze attese inferiori a 5. In altri testi, si sostiene che la maggiore robustezza del chi quadrato con più gradi di libertà permette risultati attendibili anche quando si dispone di frequenze minori. Tuttavia, qualora si avessero alcune frequenze molto basse, è bene riunire questi gruppi in un numero inferiore di categorie, aggregando ovviamente in modo logico le variabili che sono tra loro più simili.

**In una tabella di contingenza  $M \times N$ , i gradi di libertà sono**

$$(M-1) \times (N-1)$$

**dove  $M$  è il numero di colonne e  $N$  è il numero di righe.**

Il valore del chi quadrato può essere ottenuto con la formula generale, fondata sullo scarto tra frequenze osservate e frequenze attese.

Anche per le tabelle  $M \times N$  sono state proposte formule rapide, come il metodo di Skory. In realtà, sono metodi più complessi di quelli già illustrati e non presentano vantaggi apprezzabili nel tempo richiesto e nelle approssimazioni dei calcoli, rispetto alla formula generale. Inoltre, nell'interpretazione dei risultati hanno lo svantaggio di evidenziare la differenza complessiva, ma non ogni singola differenza tra la distribuzione attesa e quella osservata.

Quando si analizzano e si interpretano i risultati in tabelle M x N dopo il calcolo del  $\chi^2$ , se **si è rifiutata l'ipotesi nulla** non è semplice **individuare con precisione a quali caselle**, a quali associazioni positive o negative, **sia imputabile in prevalenza il risultato complessivo**. A questo scopo esistono due metodi.

Il più semplice consiste nel riportare in una tabella M x N il contributo al valore del chi quadrato fornito da ogni casella; ma è utile solo per la descrizione. Il secondo si fonda sulla scomposizione e sull'analisi dei singoli gradi di libertà, come verrà di seguito schematicamente illustrata nei suoi concetti fondamentali.

Il contributo al valore totale dato da ogni casella è evidenziato riportando per ognuna di essa, in una tabella M x N, il valore del rapporto

$$\left( \frac{f_{i,j}^{oss} - f_{i,j}^{att}}{f_{i,j}^{att}} \right)^2$$

ESEMPIO. Si vuole verificare se esiste associazione tra tipo di coltivazione del terreno e presenza di alcune specie d'insetti. In 4 diversi appezzamenti di terreno, con coltivazioni differenti, è stata contata la presenza di 5 specie differenti di insetti, secondo le frequenze riportate nella tabella sottostante.

#### DISTRIBUZIONE OSSERVATA

	Specie A	Specie B	Specie C	Specie D	Specie E	Totale
Coltivazione I	<b>12</b>	<b>8</b>	<b>21</b>	<b>5</b>	<b>4</b>	50
Coltivazione II	<b>15</b>	<b>10</b>	<b>5</b>	<b>20</b>	<b>8</b>	58
Coltivazione III	<b>9</b>	<b>6</b>	<b>10</b>	<b>17</b>	<b>11</b>	53
Coltivazione IV	<b>23</b>	<b>12</b>	<b>12</b>	<b>31</b>	<b>17</b>	95
Totale	59	36	48	73	40	256

Con questi dati raccolti in natura, si può sostenere che esiste associazione tra specie d'insetti e tipo di coltivazione del terreno?

Risposta.

L'ipotesi nulla afferma che il tipo di coltivazione del terreno non influisce sulla presenza delle specie d'insetti. L'ipotesi alternativa sostiene che esiste associazione tra tipo di coltivazione e presenza d'insetti, ricordando che può esistere associazione sia quando il tipo di coltivazione aumenta la frequenza di alcune specie sia quando la riduce significativamente.

E' un test bilaterale (i test in tabelle M x N possono essere solo bilaterali).

Se non esistesse associazione e la distribuzione delle 5 specie d'insetti nella 4 zone con coltivazioni differenti fosse uniforme, si avrebbe la seguente distribuzione attesa che ha 12 g.d.l (4 x 3).

#### DISTRIBUZIONE ATTESA O TEORICA, SECONDO L'IPOTESI NULLA

	Specie A	Specie B	Specie C	Specie D	Specie E	Totale
Coltivazione I	<b>11,5</b>	<b>7,0</b>	<b>9,4</b>	<b>14,3</b>	<b>7,8</b>	50
Coltivazione II	<b>13,4</b>	<b>8,2</b>	<b>10,9</b>	<b>16,5</b>	<b>9,1</b>	58
Coltivazione III	<b>12,2</b>	<b>7,5</b>	<b>9,9</b>	<b>15,1</b>	<b>8,3</b>	53
Coltivazione IV	<b>21,9</b>	<b>13,3</b>	<b>17,8</b>	<b>27,1</b>	<b>14,8</b>	95
Totale	59	36	48	73	40	256

Con la formula generale, considerando le frequenze osservate e quelle attese nelle 20 caselle si ottiene un valore del chi quadrato con 12 g.d.l. uguale a 32,251

$$c_{(12)}^2 = \frac{(12 - 11,5)^2}{11,5} + \frac{(8 - 7)^2}{7} + \dots + \frac{(17 - 14,8)^2}{14,8} = 32,251$$

Sia per ottenere il risultato complessivo che per la successiva interpretazione, è utile calcolare il contributo di ogni casella al valore del chi quadrato totale, utilizzando la sua proprietà additiva, come evidenzia la tabella successiva

(OSSERVATO - ATTESO)<sup>2</sup> / ATTESO

	Specie A	Specie B	Specie C	Specie D	Specie E	Totale
Coltivazione I	<b>0,022</b>	<b>0,143</b>	<b>14,315</b>	<b>6,048</b>	<b>1,851</b>	22,379
Coltivazione II	<b>0,191</b>	<b>0,395</b>	<b>3,194</b>	<b>0,742</b>	<b>0,133</b>	4,655
Coltivazione III	<b>0,839</b>	<b>0,300</b>	<b>0,001</b>	<b>0,239</b>	<b>0,878</b>	2,257
Coltivazione IV	<b>0,055</b>	<b>0,127</b>	<b>1,890</b>	<b>0,561</b>	<b>0,327</b>	2,960
Totale	1,107	0,965	19,400	7,590	3,189	<b>32,251</b>

Alla probabilità 0.05 con 12 g.d.l., la tabella del  $\chi^2$  dà un valore critico uguale a 21.03.

Il valore del  $\chi^2_{(12)}$  calcolato (uguale a 32,251) è significativo. Si rifiuta l'ipotesi nulla ed implicitamente si accetta l'ipotesi alternativa: esistono specie che hanno una frequenza maggiore ed altre una frequenza minore in rapporto al tipo di coltivazione.

Per entrare nella interpretazione più fine di questa significatività complessiva, la tabella che riporta il contributo di ogni casella al valore complessivo del chi quadrato mostra in quali associazioni tra righe e colonne si trovano gli **scarti relativi maggiori tra osservati ed attesi**.

Nell'esempio, una sola casella fornisce quasi la metà del valore totale: è la specie C nella coltivazione I, con un contributo di 14,315 al valore totale di 32,251. Il confronto tra valori osservati ed attesi mostra che la significatività è imputabile ad una presenza maggiore dell'atteso di individui della specie C nella coltivazione I.

La specie C e la coltivazione I formano un'**associazione positiva**.

Contribuisce sensibilmente al valore del chi quadrato anche una presenza più ridotta della specie D nella coltivazione I; è un'**associazione negativa**, che contribuisce al chi quadrato complessivo con un valore di 6,048.

**La scomposizione dei gradi di libertà di queste tabelle complesse** è un altro modo che permette di avere informazioni più dettagliate, sugli effetti di ogni particolare gruppo di dati.

**La proprietà additiva del  $\chi^2$  e dei relativi gradi di libertà consente la scomposizione di una tabella  $M \times N$  in tanti test  $2 \times 2$ , ognuno con 1 g.d.l., quanti sono i gradi di libertà totali della matrice.**

Quando si è interessati ad individuare la causa di una significativa deviazione dall'ipotesi nulla, è possibile costruire i test che ne spiegano le quote maggiori.

Prendendo come schema di riferimento una teorica tabella  $3 \times 3$  con la relativa simbologia

	TRATT. I	TRATT. II	TRATT. III	Totali
Blocco A	$a_1$	$a_2$	$a_3$	$n_1$
Blocco B	$b_1$	$b_2$	$b_3$	$n_2$
Blocco C	$c_1$	$c_2$	$c_3$	$n_3$
Totali	$n_4$	$n_5$	$n_6$	$N$

con 9 dati si ottiene un  $\chi^2$  che ha 4 gradi di libertà. Se risulta significativo, è utile scomporre questa valutazione globale, per conoscere quali confronti singoli  $2 \times 2$  siano la causa di questa differenza tra frequenze osservate e frequenze attese.

Con 4 gradi di libertà è possibile fare solamente 4 confronti. Se impostati correttamente, **la somma dei valori di questi 4  $\chi^2_{(1)}$  con 1 grado di libertà deve essere uguale al valore complessivo del  $\chi^2_{(4)}$  con 4 g.d.l. calcolato su tutti i dati.**

La ripartizione deve essere eseguita in modo gerarchico: stabilita una prima suddivisione, **le ripartizioni successive devono essere attuate sempre all'interno della precedente.** E' il modo per **rendere i confronti ortogonali: la conclusione precedente non deve dare informazioni sul test successivo.**

Con la tabella  $3 \times 3$  presentata, una possibile partizione dei 4 gradi di libertà è quella di seguito riportata:

1)

$a_1$	$a_2$
$b_1$	$b_2$

2)

$a_1 + a_2$	$a_3$
$b_1 + b_2$	$b_3$

3)

$a_1 + b_1$	$a_2 + b_2$
$c_1$	$c_2$

4)

$a_1 + a_2 + b_1 + b_2$	$a_3 + b_3$
$c_1 + c_2$	$c_3$

Anche dalla semplice osservazione risulta evidente che esistono molte possibilità differenti di suddivisione della medesima tabella.

La scelta dipende dal ricercatore, che è totalmente libero di scegliere i raggruppamenti di caselle che gli sembrano più logici ed utili per spiegare la significatività ottenuta; ma **tale scelta deve essere fatta “a priori” non “a posteriori”**, per non alterare la probabilità di scegliere una distribuzione casualmente significativa. Scelta a priori significa che essa deve essere fatta in modo totalmente indipendente dai dati rilevati; non è corretto individuare quali gruppi hanno le frequenze maggiori e quali le frequenze minori e successivamente pianificare la suddivisione, sulla base delle differenze osservate, scegliendo quelle che danno valori del chi quadrato maggiori.

**L’argomento è complesso e l’applicazione richiede altre conoscenze**, oltre a questi concetti fondamentali, adeguati al livello di conoscenze del presente manuale.

Per la scomposizione dei gradi di libertà del  $\chi^2$  in tabelle M x N, non molto frequente nella letteratura attuale, si rinvia a testi specifici.

### 3.10. IL LOG-LIKELIHOOD RATIO O METODO G

Il test  $c^2$  rappresenta il metodo classico. Più recentemente, vari testi e programmi informatici a grande diffusione hanno riproposto il **Log-likelihood ratio** o test **G**, indicato nel passato anche con **G<sup>2</sup>**, che utilizza i logaritmi naturali.

Rispetto al test  $c^2$  questo metodo, che affronta gli stessi problemi inferenziali, ha il vantaggio di richiedere calcoli semplici quando il disegno diventa complesso, come in matrici a più di due dimensioni. Sotto l’aspetto teorico, il vantaggio principale di questo metodo è di essere ritenuto un **metodo più “robusto” del  $c^2$  e a volte più potente, nel caso di frequenze piccole**.

**Larntz K.** nel 1978 (nell’articolo *Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics*, pubblicato su *Journal Amer. Stat. Ass.* vol. 73, pp. 253-263) ha dimostrato che, quando le frequenze attese variano tra 1,5 e 4, alla significatività del 5% permette di rifiutare l’ipotesi nulla molto più spesso del  $c^2$ , mentre non ha trovato differenze quando le frequenze attese sono 0 oppure 1. Questi confronti tuttavia non hanno considerato gli effetti delle **correzioni proposte da Williams D. A. nel 1976** (con l’articolo *Improved likelihood ratio test for complete contingency tables*, pubblicato su *Biometrika* vol. 63, pp. 33-37), che forniscono un risultato più conservativo e che riportano il G a valori più vicini a quello del  $c^2$  corrispondente.

Il **Maximum Likelihood Estimate (MLE)** di un parametro è il suo possibile valore, supponendo che esso coincida con quello del campione.

Il **log likelihood ratio**, chiamato in italiano **logaritmo del rapporto di verosimiglianza** o, con termine meno tecnico, **log del rapporto di probabilità**, è fondato appunto sul logaritmo naturale o

neperiano di tale rapporto. Per comprendere il significato insito nel metodo, è utile rifarsi ad un esempio semplice di confronto tra una distribuzione binomiale osservata ed una distribuzione attesa.

Si supponga di avere ottenuto da un esperimento di Mendel, su un totale di 104 individui, 89 con il fenotipo **A** e 15 con il fenotipo **a**, mentre l'atteso rapporto di 3 a 1 avrebbe dovuto dare 78 individui **A** e 26 individui **a**.

Il calcolo del *log-likelihood ratio test for goodness of fit* si sviluppa nei quattro passaggi logici:

1 – con la distribuzione binomiale, calcolare la probabilità **P** esatta di trovare 89 individui di tipo **A** e 15 individui di tipo **a**, nell'ipotesi che la probabilità **p** vera di avere un individuo **A** sia quella sperimentale di **89/104** e, ovviamente, che la probabilità **q** di avere **a** sia **15/104**:

$$C_{104}^{89} \cdot \left(\frac{89}{104}\right)^{89} \cdot \left(\frac{15}{104}\right)^{15} = 0.11071$$

2 – sempre con la distribuzione binomiale calcolare la probabilità **P** esatta di trovare 89 individui di tipi **A** e 15 di tipo **a**, questa volta nell'ipotesi che la probabilità **p** vera di avere un individuo **A** sia quella attesa o teorica di **3/4** e quella di avere un individuo **a** sia **1/4**:

$$C_{104}^{89} \cdot \left(\frac{3}{4}\right)^{89} \cdot \left(\frac{1}{4}\right)^{15} = 0.00337$$

che può essere scritto anche come

$$C_{104}^{89} \cdot \left(\frac{78}{104}\right)^{89} \cdot \left(\frac{26}{104}\right)^{15} = 0.00337$$

3 - la prima probabilità è sempre maggiore della seconda poiché si fonda sui dati osservati; la seconda è tanto più vicina alla prima quanto più l'atteso è vicino al valore osservato; il test è fondato sul rapporto **L**

$$L = 0.110071 / 0.00337 = 32,85$$

4 – la distribuzione del valore **L** è complessa e poco conosciuta; è invece nota **la distribuzione G derivata da L**, data da

$$G = 2 \ln L$$

(dove **ln** indica il logaritmo naturale).

Nel caso di grandi campioni è bene approssimata dalla distribuzione  $\chi^2$  con gli stessi gdl (che in questo esempio è uguale a 1).

Con i dati dell'esempio

$$G = 2 \ln L = 2 \ln 32,85 = 2 \times 3,49 = 6,98$$

G risulta uguale a 6,98; poiché il valore del chi-quadrato alla probabilità 0.05 e per 1 gdl è uguale a 3,84 si conclude che la presenza di individui A è significativamente maggiore dell'atteso.

Spesso al posto di **G** si usa la simbologia **2I**, per specificare che G è il doppio dell'informazione contenuta nel campione: infatti il rapporto L è calcolato mediante l'informazione contenuta sia nel campione sia nell'atteso, cioè sia nell'ipotesi  $H_1$  alternativa (il campione) sia nell'ipotesi nulla  $H_0$  (l'atteso).

Quanto spiegato nell'esempio svolto contiene anche le informazioni per derivare e comprendere la formula base del test G.

Indicando con

- n1 la frequenza osservata del primo gruppo,
- n2 la frequenza osservata del secondo gruppo,
- N il numero totale di osservazioni ( $N = n1 + n2$ ),
- $p_o$  la probabilità osservata del primo gruppo,
- $q_o$  la probabilità osservata del secondo gruppo,
- $p_a$  la probabilità attesa del primo gruppo,
- $q_a$  la probabilità attesa del secondo gruppo,

la formula utilizzata può essere scritta

$$L = \frac{C_N^{n_1} \cdot p_o^{n_1} \cdot q_o^{n_2}}{C_N^{n_1} \cdot p_a^{n_1} \cdot q_a^{n_2}} = \left( \frac{p_o}{p_a} \right)^{n_1} \cdot \left( \frac{q_o}{q_a} \right)^{n_2}$$

Poiché

$$n1 \text{ osservato} = Np_o \quad \text{e} \quad n2 \text{ osservato} = Nq_o$$

e similmente

$$n1 \text{ atteso} = Np_a \quad \text{e} \quad n2 \text{ atteso} = Nq_a$$

$$L = \left( \frac{n_o^1}{n_a^1} \right)^{n_o^1} \cdot \left( \frac{n_o^2}{n_a^2} \right)^{n_o^2}$$

da cui

$$\ln L = n_o^1 \ln \left( \frac{n_o^1}{n_a^1} \right) + n_o^2 \ln \left( \frac{n_o^2}{n_a^2} \right)$$

Per la significatività, **il valore di G non ha una sua tabella, ma utilizza la stessa distribuzione dei valori critici del  $\chi^2$  e ha gli stessi gradi di libertà.**

Il test **likelihood ratio** può essere applicato nei 3 casi già descritti:

a – il confronto tra una distribuzione osservata e la corrispondente distribuzione attesa;

b – il confronto tra due campioni indipendenti, in tabelle 2 x 2;

c – il confronto tra più campioni indipendenti in tabelle M x N.

A questi può essere aggiunto il caso di tabelle a più dimensioni (quando i fattori sono più di 2 ognuno con **p** modalità), che è molto difficile analizzare con il metodo  $\chi^2$ .

### **3.10.1 Confronto tra una distribuzione osservata ed una attesa con la correzione di Williams**

Nel caso del confronto tra una distribuzione osservata e la corrispondente distribuzione attesa, il calcolo di **G** può essere ottenuto con una delle due formule sottoriportate, di cui la prima risulta più rapida per i calcoli manuali

$$\mathbf{G \text{ o likelihood ratio}} = 2 \sum_{i=1}^k [Oss. \ln(Oss./Att.)] = 2 \left[ \sum_{i=1}^k Oss. (\ln Oss.) - \sum_{i=1}^k Oss. (\ln Att.) \right]$$

dove,

la sommatoria  $\sum$  è estesa a tutte le **k** caselle,

**ln** = logaritmo naturale,

Oss. = frequenze osservate,

Att. = frequenze attese, in accordo con l'ipotesi nulla.

Quando il campione è **inferiore alle 200 unità**, **Williams** ha proposto la seguente **correzione q**

$$q = 1 + \frac{k^2 - 1}{6Nn}$$

dove

**k** = il numero di gruppi;

**n** = il numero di gradi di libertà;

**N** = il numero totale di osservazioni.

Quando il numero di g.d.l. è uguale al numero di gruppi meno 1 (come è nella norma di gruppi tra loro indipendenti) e quindi **n = k - 1**,

la formula può essere semplificata in

$$q = 1 + \frac{k + 1}{6N}$$

Nel caso di due gruppi, la formula può essere scritta come

$$q = 1 + \frac{1}{2N}$$

Il valore di G corretto (*adjusted*), simboleggiato con  $G_{adj}$ , è ottenuto con il rapporto

$$G_{adj} = \frac{G}{q}$$

ESEMPIO 1. Per valutare la ricchezza in specie di 4 zone (A, B, C, D) sono state contate le specie presenti, con il seguente risultato:

ZONE	A	B	C	D
Specie presenti	55	28	37	43

Esiste una differenza significativa tra le 4 zone?

Valutare la risposta con il test  $c^2$  e con il metodo G, senza e con la correzione.

Risposta.

Dopo aver calcolato il totale e la frequenza attesa in ogni classe nella condizione che sia vera l'ipotesi nulla,

ZONE	A	B	C	D	Totale
Specie presenti	55	28	37	43	163
Specie attese	40,75	40,75	40,75	40,75	163,00

si calcola il valore del  $c^2$  che ha 3 gdl

$$c_{(3)}^2 = \frac{(55 - 40,75)^2}{40,75} + \frac{(28 - 40,75)^2}{40,75} + \frac{(37 - 40,75)^2}{40,75} + \frac{(43 - 40,75)^2}{40,75}$$

$$c_{(3)}^2 = \frac{14,25^2}{40,75} + \frac{12,75^2}{40,75} + \frac{3,75^2}{40,75} + \frac{2,25^2}{40,75} = 4,983 + 3,989 + 0,345 + 0,124 = 9,441$$

e risulta uguale a 9,441; con la correzione di Yates il valore diminuisce a

$$c_{(3)}^2 = \frac{13,75^2}{40,75} + \frac{12,75^2}{40,75} + \frac{3,75^2}{40,75} + \frac{2,75^2}{40,75} = 4,640 + 3,989 + 0,345 + 0,186 = 9,160$$

9,160.

Usando il metodo G

$$G = 2 \left( 55 \ln \frac{55}{40,75} + 28 \ln \frac{28}{40,75} + 37 \ln \frac{37}{40,75} + 43 \ln \frac{43}{40,75} \right)$$

$$G = 2(16,493 - 10,507 - 3,572 + 2,311) = 2 \times 4,725 = 9,450$$

si ottiene un valore di 9,450; con la **correzione di Williams** il valore q, stimato con la formula

$$q = 1 + \frac{k+1}{6N}$$

dove

k = 4 e N = 163

$$q = 1 + \frac{4+1}{6 \times 163} = 1 + \frac{5}{978} = 1,0051$$

risulta uguale a 1,0051

e quindi il valore di  $G_{adj}$

$$G_{adj} = \frac{9,450}{1,0051} = 9,402$$

risulta uguale a 9,402.

In questo esempio, il valore di  $G$  risulta leggermente superiore a quello del corrispondente  $c^2$ . Poiché il valore critico del  $c^2$  con 3 gdl alla probabilità 0,05 è uguale a 7,815 è possibile rifiutare l'ipotesi nulla.

**ESEMPIO 2.** Nell'incrocio tra due ibridi **Aa x Aa**, sono stati contati 95 individui con fenotipo **A** e 41 con fenotipo **a**. E' una distribuzione in accordo con l'atteso di 3 a 1?

Risposta.

Dopo aver calcolato le due frequenze attese

Fenotipi	A	a	TOTALE
Freq. osservate	95	41	136
Freq. attese	102,0	34,0	136,0

si calcola il valore di **G**

$$G = 2 \cdot \left( 95 \ln \frac{95}{102} + 41 \ln \frac{41}{34} \right) = 2 \cdot [95 \cdot (-0,071) + 41 \cdot (+0,187)] = 2 \cdot (-6,745 + 7,662) = 1,844$$

che risulta uguale a 1,844 da confrontare con quelli riportati nella tabella del  $\chi^2$  con i gdl.

La **correzione q di Williams** è

$$q = 1 + \frac{1}{2 \cdot 136} = 1,0037$$

uguale a 1,0037

per cui il valore di **G** aggiustato per le dimensioni non grandi del campione è

$$G_{adj} = \frac{1,844}{1,0037} = 1,837$$

uguale a 1,837.

### 3.10.2 Tabella 2 x 2, con la correzione di Williams e quella di Mantel-Haenszel.

Nel caso di tabelle 2 x 2, per verificare l'indipendenza tra il fattore riportato in riga e quello riportato in colonna, entrambi con variabili binarie, usando la consueta simbologia per le **frequenze assolute**

	Risp. X	Risp. x	Totale
Camp. Y	a	b	<b>n<sub>1</sub></b>
Camp. y	c	d	<b>n<sub>2</sub></b>
Totale	<b>n<sub>3</sub></b>	<b>n<sub>4</sub></b>	N

il valore di **G** o **Log-likelihood ratio** è dato da

$$G = 2[(a \ln a + b \ln b + c \ln c + d \ln d) - (n_1 \ln n_1 + n_2 \ln n_2 + n_3 \ln n_3 + n_4 \ln n_4) + (N \ln N)]$$

**Sempre da impiegare nel caso di campioni con meno di 200 osservazioni**, in tabelle 2 x 2 il coefficiente di **correzione q di Williams** è calcolato con

$$q = 1 + \frac{\left(\frac{N}{n_1} + \frac{N}{n_2} - 1\right) \cdot \left(\frac{N}{n_3} + \frac{N}{n_4} - 1\right)}{6N}$$

ed il valore di  $G_{adj}$  dato dal rapporto

$$G_{adj} = G/q$$

Nel caso di tabelle 2 x 2 è diffusa anche la **correzione di Mantel-Haenszel, utilizzata pure in altri casi, come per il calcolo del  $c^2$** .

Consiste nell'aggiungere o togliere 0,5 ad ognuna delle 4 frequenze osservate, sulla base del confronto dei prodotti delle due diagonali:  $a \times d$  contro  $b \times c$ .

Se il prodotto  $a \times d$  è maggiore di  $b \times c$ ,

- si toglie 0,5 sia ad  $a$  che a  $d$  e
- si aggiunge 0,5 sia a  $b$  che a  $c$ .

Se il prodotto  $a \times d$  è minore di  $b \times c$ ,

- si aggiunge 0,5 sia ad  $a$  che a  $d$  e
- si toglie 0,5 sia a  $b$  che a  $c$ .

Con tali correzioni, **sia i totali marginali  $n_1, n_2, n_3, n_4$ , sia il totale generale N restano invariati**.

ESEMPIO . In due appezzamenti di terreno (A e B) con suoli di natura diversa sono stati messi a dimora alberi della stessa specie. Solo una parte di essi ha avuto una crescita normale (+).

	Risp. +	Risp. -	Totale
Camp. A	35	18	53
Camp. B	23	26	49
Totale	58	44	102

Si può affermare che la diversa natura dei due suoli incide sulla crescita normale di tale specie?

Confrontare i risultati ottenuti (A) dal test G, con le due (B e C) correzioni proposte, ed il test  $\chi^2$  senza (D) e con la correzione (E) per campioni non grandi.

Risposta.

A) Con il metodo G, che ha 1 gdl, scindendo l'operazione in 3 parti,

I - quella che riguarda i 4 valori osservati a, b, c, d,

II - quella che riguarda i 4 totali marginali  $n_1, n_2, n_3, n_4$ ,

III - quella che riguarda il totale generale N,

si ottiene:

$$I = 35 \ln 35 + 18 \ln 18 + 23 \ln 23 + 26 \ln 26 = 124,44 + 52,06 + 72,12 + 84,71 = 333,33$$

$$II = 53 \ln 53 + 49 \ln 49 + 58 \ln 58 + 44 \ln 44 = 210,43 + 190,70 + 235,51 + 166,50 = 803,14$$

$$III = 102 \ln 102 = 471,75$$

da cui si stima

$$G = 2 \cdot (333,33 - 803,14 + 471,75) = 2 \cdot 1,94 = 3,88$$

un valore di G uguale a 3,88.

B) Con la correzione di Williams

$$q = 1 + \frac{\left(\frac{102}{53} + \frac{102}{49} - 1\right) \cdot \left(\frac{102}{58} + \frac{102}{44} - 1\right)}{6 \cdot 102} = 1 + \frac{(1,925 + 2,082 - 1) \cdot (1,759 + 2,318 - 1)}{612} = 1,015$$

$$G_{adj} = \frac{3,88}{1,015} = 3,82$$

si ottiene un valore di G aggiustato uguale a 3,82.

C) Con la correzione di Mantel-Haenszel,

poiché il prodotto di 35 x 26 è maggiore di quello dato da 18 x 23 e quindi i dati osservati devono essere modificati in

	Risp.+	Risp. -	Totale
A	34,5	18,5	53
B	23,5	25,5	49
Totale	58	44	102

si ottiene

$$I = 34,5 \ln 34,5 + 18,5 \ln 18,5 + 23,5 \ln 23,5 + 25,5 \ln 25,5 = 122,16 + 53,98 + 74,19 + 82,59 = 332,92$$

da cui

$$G_{adj} = 2 \cdot (332,92 - 803,14 + 471,75) = 2 \cdot 153 = 3,06$$

si ottiene un valore di **G** aggiustato uguale a 3,06.

D) **Il  $c^2$  con la formula abbreviata per grandi campioni**

$$c_{(1)}^2 = \frac{(35 \cdot 26 - 18 \cdot 23)^2 \cdot 102}{53 \cdot 49 \cdot 58 \cdot 44} = \frac{(910 - 414)^2 \cdot 102}{6627544} = \frac{25093632}{6627544} = 3,78$$

da un valore uguale a 3,78 con 1 gdl.

E) **Il  $c^2$  con la correzione di Yates** per campioni non grandi

$$c_{(1)}^2 = \frac{\left( |35 \cdot 26 - 18 \cdot 23| - \frac{102}{2} \right)^2 \cdot 102}{53 \cdot 49 \cdot 58 \cdot 44} = \frac{(|910 - 414| - 51)^2 \cdot 102}{6627544} = \frac{20198550}{6627544} = 3,05$$

da un valore uguale a 3,05.

### 3.10.3 **Tabelle M x N con la correzione di Williams**

Nel caso di una tabella **M x N**, **la formula per calcolare G** è un'estensione di quella già utilizzata per tabelle 2 x 2.

Indicando con **f** le frequenze osservate e spezzando le operazioni in 3 passaggi, con

**I** = à **f ln f** di ogni casella,

**II** = à **f ln f** di ogni totale marginale sia di riga che di colonna,

**III** = à **N ln N**, con **N** uguale al totale generale del numero di osservazioni,

si ottiene il valore di **G** con

$$G = 2 \times (I - II + III)$$

In questo caso, la **correzione q di Williams** in una formula semplice è data da

$$q = 1 + \frac{(m+1) \cdot (n+1)}{6N}$$

dove

**m** e **n** sono il numero di righe e il numero di colonne della matrice,

**N** è il numero totale di osservazioni.

Con la simbologia utilizzata nella tabella seguente, applicata al caso di una tabella 3 x 3,

	TRATT. I	TRATT. II	TRATT. III	Totali
Blocco A	<b>a<sub>1</sub></b>	<b>a<sub>2</sub></b>	<b>a<sub>3</sub></b>	<b>n<sub>1</sub></b>
Blocco B	<b>b<sub>1</sub></b>	<b>b<sub>2</sub></b>	<b>b<sub>3</sub></b>	<b>n<sub>2</sub></b>
Blocco C	<b>c<sub>1</sub></b>	<b>c<sub>2</sub></b>	<b>c<sub>3</sub></b>	<b>n<sub>3</sub></b>
Totali	<b>n<sub>4</sub></b>	<b>n<sub>5</sub></b>	<b>n<sub>6</sub></b>	<b>N</b>

il modo per calcolare il valore del **likelihood ratio** è

$$G = [ ( a_1 \ln a_1 + a_2 \ln a_2 + a_3 \ln a_3 + b_1 \ln b_1 + b_2 \ln b_2 + b_3 \ln b_3 + c_1 \ln c_1 + c_2 \ln c_2 + c_3 \ln c_3 ) - ( n_1 \ln n_1 + n_2 \ln n_2 + n_3 \ln n_3 + n_4 \ln n_4 + n_5 \ln n_5 + n_6 \ln n_6 ) + ( N \ln N ) ]$$

I gradi di libertà sono **(m - 1) x (n - 1)**, uguali a 2 x 2 = 4 nel caso della tabella precedente.

ESEMPIO. In tre zone di una città (chiamate A, B, C) sono state rilevate varie misure d'inquinamento, da quello acustico a quello atmosferico. Successivamente, dai diversi valori d'inquinamento sono stati derivati punteggi o indici, suddivisi in tre categorie: bassi, medi, alti.

Nella tabella sottostante, sono riportate le frequenze delle tre categorie di indici, per ognuna delle tre zone chiamate rispettivamente A, B e C.

### FREQUENZE OSSERVATE

	Punteggi Bassi	Punteggi Medi	Punteggi Alti	Totali
Zona A	5	2	1	8
Zona B	3	10	1	14
Zona C	5	2	4	11
Totali	13	14	6	33

Esiste una differenza significativa nella distribuzione dei punteggi delle tre zone? Punteggi bassi, medi e alti hanno la stessa distribuzione percentuale nelle tre zone?

(Calcolare il valore del  $\chi^2$  ed il valore del **likelihood ratio**, confrontando i risultati.)

Risposta.

L'ipotesi nulla **H<sub>0</sub>** sostiene che la distribuzione dei valori bassi, medi e alti nelle 3 zone è uguale e che le differenze riscontrate sono imputabili solamente a variazioni casuali; in altri termini **i livelli d'inquinamento sono indipendenti dalla zona**.

L'ipotesi alternativa **H<sub>1</sub>** afferma che tra le tre zone esiste una distribuzione significativamente differente dei valori bassi, medi e alti. In altri termini, le tre zone hanno una percentuale differente di valori bassi, medi o alti: **esiste associazione tra livelli d'inquinamento e zona**

E' un test bilaterale (nei confronti multipli e con i test proposti non sono possibili test unilaterali).

### FREQUENZE ATTESE

	Punteggi Bassi	Punteggi Medi	Punteggi Alti	Totali
Zona A	3,15	3,39	1,46	8
Zona B	5,52	5,94	2,54	14
Zona C	4,33	4,67	2,00	11
Totali	13	14	6	33

Per il calcolo del  $\chi^2$  e per la successiva interpretazione è utile calcolare la tabella delle frequenze attese, nella condizione che l'ipotesi nulla sia vera.

Per i calcoli successivi, è necessario costruire la tabella delle differenze tra valori osservati ed attesi. Serve anche come verifica dei calcoli già effettuati, poiché sia **i totali di riga che quelli di colonna ed il totale generale devono essere uguali a zero**, ricordando che le frequenze attese sono calcolate a partire dai totali marginali, secondo la relazione

$$\text{Frequenza attesa in ogni casella} = \text{Totale di riga} \times \text{Totale di colonna} / \text{Totale generale}$$

FREQUENZE OSSERVATE - FREQUENZE ATTESE

	Punteggi Bassi	Punteggi Medi	Punteggi Alti	Totali
Zona A	<b>+1,85</b>	<b>-1,39</b>	<b>-0,46</b>	0
Zona B	<b>-2,52</b>	<b>+4,06</b>	<b>-1,54</b>	0
Zona C	<b>+0,67</b>	<b>-2,67</b>	<b>+2,00</b>	0
Totali	0	0	0	0

Il passo successivo è la stima del contributo di ogni casella al valore complessivo del  $\chi^2$ . Ogni casella indica la differenza tra frequenza osservata e frequenza attesa, in rapporto alla frequenza attesa.

$$(\text{FREQUENZE OSSERVATE} - \text{FREQUENZE ATTESE})^2 / \text{FREQUENZE ATTESE}$$

	Punteggi Bassi	Punteggi Medi	Punteggi Alti	Totali
Zona A	<b>1,087</b>	<b>0,570</b>	<b>0,145</b>	1,802
Zona B	<b>1,150</b>	<b>2,775</b>	<b>0,934</b>	4,859
Zona C	<b>0,104</b>	<b>1,527</b>	<b>2,000</b>	3,631
Totali	2,341	4,872	3,079	<b>10,292</b>

Si ottiene un valore complessivo del  $\chi^2_{(4)}$  uguale a 10,292; la tabella sinottica alla probabilità 0.05 fornisce un valore critico uguale a 9,49.

Si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa: fra le tre zone esiste una differenza significativa nella distribuzione degli indici d'inquinamento e quindi associazione tra zone e livelli d'inquinamento. La lettura dei valori del chi quadrato in ogni casella ed il confronto tra distribuzione osservata e distribuzione attesa evidenziano che i contributi più importanti alla significatività sono dati sia da una presenza di valori medi osservati maggiore dell'atteso nella zona **B**, sia da una presenza di valori bassi nella zona **B** minori dell'atteso.

**L'obiezione più importante che si può rivolgere alla attendibilità delle conclusioni raggiunte con il  $c^2$  deriva dalle ridotte dimensioni del campione: sono solamente 33 osservazioni in totale, distribuite in 9 caselle delle quali 5 hanno frequenze attese inferiori a 4.**

Pertanto può essere richiesto l'uso del metodo **likelihood ratio** proposto come metodo robusto, da applicare correttamente anche nel caso di campioni piccoli.

Applicato ai dati dell'esempio, il valore del **log likelihood ratio** è ottenuto mediante

$$G = 2 \left[ \begin{array}{l} (5 \ln 5 + 2 \ln 2 + 1 \ln 1 + 3 \ln 3 + 10 \ln 10 + 1 \ln 1 + 5 \ln 5 + 2 \ln 2 + 4 \ln 4) - \\ - (8 \ln 8 + 14 \ln 14 + 11 \ln 11 + 13 \ln 13 + 14 \ln 14 + 6 \ln 6) + \\ + (33 \ln 33) \end{array} \right]$$

$$G = 2 \times \{ [(5 \times 1,609) + (2 \times 0,693) + (1 \times 0) + (3 \times 1,099) + (10 \times 2,303) + (1 \times 0) + (5 \times 1,609) + \\ + (2 \times 0,693) + (4 \times 1,386) ] - \\ - [(8 \times 2,079) + (14 \times 2,639) + (11 \times 2,398) + (13 \times 2,565) + (14 \times 2,639) + 6 \times 1,792] + \\ + [(33 \times 3,497)] \}$$

$$G = 2 \times [(8,045 + 1,386 + 0 + 3,297 + 23,030 + 0 + 8,045 + 1,386 + 5,544) - \\ - (16,632 + 36,946 + 26,378 + 33,345 + 36,946 + 10,752) + (115,409)]$$

$$G = 2 \times (50,733 - 160,999 + 115,409) = 2 \times 5,143 = 10,286$$

e fornisce un valore di **G** uguale a 10,286.

**La probabilità è fornita dalla medesima tabella dei valori critici del  $c^2$ , per gli stessi gradi di libertà.**

La correzione q di Williams è

$$q = 1 + \frac{(3+1) \cdot (3+1)}{6 \times 33} = 1 + \frac{16}{198} = 1,081$$

e pertanto il valore di  $G$  aggiustato è

$$G_{adj} = \frac{10,286}{1,081} = 9,515$$

uguale a 9,515 per 4 gdl.

Il valore critico alla probabilità  $\alpha = 0.05$  è sempre uguale 9,49.

Per quanto riguarda le conclusioni, è importante ricordare che quando il valore calcolato è vicino a quello critico non si deve decidere in modo netto, accettando o rifiutando l'ipotesi nulla con certezza. La risposta del test non è significativa se è appena superiore o appena inferiore al valore critico; si tratta di probabilità e il risultato non è molto differente se la probabilità è appena superiore o appena inferiore al livello critico prescelto: in entrambi i casi, si deve parlare di **risposte tendenzialmente significative**.

### 3.11. IL CONFRONTO TRA DUE DISTRIBUZIONI OSSERVATE:

#### IL METODO DI KOLMOGOROV-SMIRNOV PER 2 CAMPIONI INDIPENDENTI

**Una tabella  $2 \times N$ , in cui siano riportate le classi di frequenza di due distribuzioni osservate, può essere analizzata mediante il test di Kolmogorov-Smirnov. A differenza del test  $\chi^2$ , non richiede un numero minimo di osservazioni, come condizione preliminare di validità.**

Il test di Kolmogorov-Smirnov, necessario per piccoli campioni, è utile anche per grandi campioni, quando si abbia un numero elevato di classi od intervalli, poiché in tali condizioni ha una potenza maggiore del test chi quadrato.

Nella fase di raccolta dei dati, è sempre consigliato usare il maggior numero possibile di intervalli o classi. Solo successivamente, se alcune avranno frequenze molto piccole o addirittura nulle, si potranno fare raggruppamenti, che rappresentano sempre una perdita d'informazione e rendono il test meno potente.

Il test di Kolmogorov-Smirnov serve per il confronto tra due distribuzioni osservate, formate da più di 2 classi di frequenza. Non diversamente dal caso di un solo campione, **l'unica condizione di validità è che i gruppi o le classi siano ordinate secondo una scala di tipo almeno ordinale**, come d'altronde è necessario per avere una distribuzione ordinata di frequenze senza l'intervento soggettivo del ricercatore.

**Le distribuzioni di frequenza possono riguardare qualunque variabile o tipo di misura, da quelle classiche di peso ed altezza, per le quali possono essere applicati anche test parametrici,**

**a quelle espresse da rapporti, percentuali, valori angolari, colorazioni, indici, punteggi assoluti o relativi, ecc., purché tali misure possano essere tradotte in ranghi, cioè ordinate per dimensioni od intensità, anche se con molti valori uguali.**

Il test di Kolmogorov-Smirnov per due campioni indipendenti è utilizzato per verificare l'ipotesi alternativa  $H_1$  se le distribuzioni di frequenza di due campioni appartengano a popolazioni **aventi una diversa distribuzione di frequenza**. A differenza del chi quadrato, esso serve per impostare sia un test a due code sia un test a una coda.

**Come test a due code, analizza la concordanza tra due distribuzioni e quindi è sensibile ad ogni tipo di differenza tra due distribuzioni: fornisce risultati significativi sia per differenze nella tendenza centrale che nella dispersione, nella simmetria e, sebbene in modo meno evidente, nella curtosi.** Non è quindi un test specifico per nessuno di questi fattori, anche se è più sensibile alle differenze nelle tendenze centrali; per individuare con esattezza quale caratteristica della distribuzione determini la differenza maggiore riscontrata, occorre di conseguenza ricorrere anche all'uso di altri test, che ne saggiavano solamente una, e dedurre la causa principale dalle differenti risposte.

**Come test a una coda, serve per verificare l'ipotesi che una delle due distribuzioni sia formata da punteggi più elevati: quindi è sensibile soprattutto a differenze tra le tendenze centrali.**

Nello stesso modo del test per un campione, la metodologia di questo test per 2 campioni richiede dapprima la trasformazione delle frequenze assolute in frequenze relative entro ogni campione, mediante il rapporto della frequenza di ogni classe con il numero totale di osservazioni; successivamente, di confrontare le frequenze cumulate entro gli stessi intervalli, per individuare la deviazione o differenza massima. Sulla base dell'ipotesi precedente formulata, essa può essere considerata con il segno oppure in valore assoluto.

Indicando con  $O_1(X_i)$  ogni valore della sommatoria dei **dati osservati nel primo campione** e con  $O_2(X_i)$  ogni valore della sommatoria dei **dati osservati nel secondo campione**,

**nel caso di un test ad una coda si deve calcolare la deviazione massima D con il segno**

$$D = \text{diff. mass. } ( O_1(X_i) - O_2(X_i) )$$

**mentre per un test a due code la direzione della differenza non è importante e lo scarto massimo viene calcolato in valore assoluto**

$$D = \text{diff. mass. } | O_1(X_i) - O_2(X_i) |$$

Per valutare la significatività della differenza massima tra le due cumulate esistono condotte che si diversificano in rapporto alle dimensioni dei due campioni e al tipo d'ipotesi alternativa, se bilaterale o unilaterale.

Nel **caso di piccoli campioni**, quando le due distribuzioni hanno al massimo 25 osservazioni (altri testi definiscono i campioni come piccoli fino ad un massimo di 40 osservazioni), si può ricorrere a tabelle specifiche per verificare se la differenza massima tra le cumulate delle frequenze relative supera il valore critico ed è quindi significativa. **Il valore da confrontare con la tabella è ottenuto moltiplicando la differenza massima  $D$  per le dimensioni dei due campioni  $n_1$  e  $n_2$ .**

$$D \times n_1 \times n_2$$

Le tabelle, riportate alla fine del capitolo, sono differenti per test a una coda e per test a due code.

Entrambe riportano sulla prima riga e sulla prima colonna il numero di osservazioni del primo e del secondo campione (rispettivamente  $n_1$  e  $n_2$ ); alla loro intersezione sono riportati tre valori critici in colonna, associati rispettivamente dall'alto al basso alla probabilità 0.10, alla probabilità 0.05 e a quella 0.01. **Il valore calcolato con  $(D \times n_1 \times n_2)$  indica una differenza significativa quando è uguale o superiore a quello critico riportato nella tabella.**

Per esempio nel caso di un test ad una coda, con 10 osservazioni nel campione 1 e con 12 osservazioni nel campione 2 la differenza tra le due distribuzioni cumulate è significativa alla probabilità  $\alpha = 0.10$  quando il valore è  $\geq 52$ , alla probabilità  $\alpha = 0.05$  è significativa quando è  $\geq 60$  e alla probabilità  $\alpha = 0.01$  quando è  $\geq 74$ . **E' possibile osservare che la distribuzione dei valori critici è simmetrica, per cui alle stesse probabilità sono identici quando si hanno 12 osservazioni nel campione 1 e 10 osservazioni nel campione 2.**

Nel **caso di grandi campioni**, si devono seguire due metodologie differenti in rapporto alla considerazione che l'ipotesi alternativa  $H_1$  sia a una coda oppure a due code.

Se il test è a una coda, secondo **Goodman** il valore critico viene determinato mediante

$$\chi^2_{(2)} = 4D^2 \frac{n_1 \cdot n_2}{n_1 + n_2}$$

che ha una distribuzione bene approssimata da quella del chi quadrato con 2 gradi di libertà.

Rifiutando l'ipotesi nulla, si accetta l'ipotesi alternativa che i dati del campione prescelto hanno una tendenza centrale maggiore di quella dell'altro gruppo.

Se il test è a due code, è possibile definire il **valore critico** per le probabilità  $\alpha$  prescelte in rapporto alle dimensioni  $n_1$  e  $n_2$  dei due campioni:

- alla probabilità  $\alpha = 0.05$  è uguale a

$$1,36 \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}$$

- alla probabilità  $\alpha = 0.01$  è uguale a

$$1,63 \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}$$

- alla probabilità  $\alpha = 0.005$  è uguale a

$$1,73 \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}$$

e alla probabilità  $\alpha = 0.001$  è uguale a

$$1,95 \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}$$

Si rifiuta l'ipotesi nulla quando la differenza massima osservata  $D$  nel confronto tra i due campioni è uguale o superiore al valore critico stimato. **L'ipotesi alternativa  $H_1$  è che le due distribuzioni a confronto sono differenti, senza distinguere tra gli effetti della tendenza centrale, della dispersione o variabilità, della simmetria e della curtosi.**

ESEMPIO 1. Mediante le cartine al tornasole è possibile misurare il pH di alcuni campioni d'acqua; metodi analoghi di colorazione vengono usati per confrontare la quantità di fosfati e di nitrati.

Su una scala ordinale con intensità crescente, suddivisa in 8 livelli, sono state riportate le frequenze osservate durante una giornata di rilevazioni in due serie differenti di campioni, raccolti all'ingresso ed all'uscita di un depuratore.

All'ingresso sono stati raccolti 10 campioni ed all'uscita 12 campioni, secondo la distribuzione riportata nella tabella sottostante.

DISTRIBUZIONE OSSERVATA

	Intensità della colorazione							
	I	II	III	IV	V	VI	VII	VIII
Ingresso	0	0	0	3	6	0	1	0
Uscita	1	4	4	1	1	1	0	0

Questi dati dimostrano che la quantità di sostanza inquinante contenuta nell'acqua all'uscita del depuratore è significativamente minore di quella all'entrata?

Risposta.

**E' il confronto di 2 piccoli campioni, con un test ad una coda.**

Dapprima si trasformano le frequenze assolute in frequenze relative e si calcolano le cumulate; successivamente, mediante la serie di differenze tra le due cumulate, si individua lo scarto massimo.

CALCOLO DELLA DIFFERENZA MASSIMA TRA LE DISTRIBUZIONI CUMULATE

	Intensità della colorazione							
	I	II	III	IV	V	VI	VII	VIII
<b>Ingresso</b> freq. relativa.	0,000	0,000	0,000	0,300	0,600	0,000	0,100	0,000
<b>Distr.cumul.</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,300</b>	<b>0,900</b>	<b>0,900</b>	<b>1,000</b>	<b>1,000</b>
<b>Uscita</b> freq. relativa	0,084	0,333	0,333	0,084	0,083	0,083	0,000	0,000
<b>Distr.cumul.</b>	<b>0,084</b>	<b>0,417</b>	<b>0,750</b>	<b>0,834</b>	<b>0,917</b>	<b>1,000</b>	<b>1,000</b>	<b>1,000</b>
<b>Differenza tra cumulate</b>	0,084	0,417	<b>0,750</b>	0,534	0,017	0,100	0,000	0,000

Sottraendo le frequenze cumulate all'ingresso a quelle in uscita, la differenza massima è uguale a 0,750.

**Il valore da confrontare con la tabella è**

$$0,750 \cdot 10 \cdot 12 = 90$$

La tabella sinottica che riporta i valori critici per test ad una coda in piccoli campioni, per  $n_1 = 10$  e  $n_2 = 12$

**(anche se è indifferente, perché la tabella dei valori critici è simmetrica)**

come valore massimo riporta 74 alla probabilità  $\alpha = 0.01$ .

Il valore calcolato, uguale a 90, è superiore; quindi, la probabilità che la differenza riscontrata sia imputabile al caso è inferiore a 0.01.

**Si rifiuta l'ipotesi nulla**, accettando l'ipotesi alternativa che la tendenza centrale della distribuzione dei dati raccolti in uscita dal depuratore sia significativamente minore.

**Il carico inquinante all'uscita del depuratore è significativamente minore di quello all'entrata.**

ESEMPIO 2. Sovente, all'ambientalista si pone il problema di analizzare la distribuzione territoriale di specie animali o vegetali, per rispondere al quesito se sono più concentrate in alcune zone e rare in altre oppure se sono distribuite in modo uniforme. Altro quesito importante è se due specie che vivono sullo stesso territorio hanno distribuzione simile o differente, se occupano le stesse aree o aree diverse. In un'area approssimativamente circolare, è stata misurata la presenza degli individui della specie A e della specie B suddividendo il cerchio in 8 zone, corrispondenti ad angoli di 45 gradi. Si sono raccolte le osservazioni riportate nella tabella sottostante, con il campione della specie A che ha 19 osservazioni e il campione della specie B della quale si sono contati 17 individui.

DISTRIBUZIONE OSSERVATA

	Zone							
	I	II	III	IV	V	VI	VII	VIII
Specie A	4	5	0	2	0	0	1	7
Specie B	1	4	4	6	1	1	0	0

Si può sostenere che le due specie occupano aree differenti?

Risposta.

**E' un test a 2 code e si dispone di piccoli campioni.**

Dapprima si calcola la differenza massima tra le due cumulate senza considerare il segno, dopo trasformazione delle frequenze assolute in frequenze relative.

CALCOLO DELLA DIFFERENZA MASSIMA TRA LE DISTRIBUZIONI CUMULATE

(in valore assoluto)

	Zone							
	I	II	III	IV	V	VI	VII	VIII
<b>Specie A</b> freq. relativa	0,211	0,263	0,000	0,105	0,000	0,000	0,053	0,368
<b>Distr.cumul.</b>	<b>0,211</b>	<b>0,474</b>	<b>0,474</b>	<b>0,579</b>	<b>0,579</b>	<b>0,579</b>	<b>0,632</b>	<b>1,000</b>
<b>Specie B</b> freq. relativa	0,059	0,235	0,235	0,353	0,059	0,059	0,000	0,000
<b>Distr.cumul.</b>	<b>0,059</b>	<b>0,294</b>	<b>0,529</b>	<b>0,882</b>	<b>0,941</b>	<b>1,000</b>	<b>1,000</b>	<b>1,000</b>
<b>Differenza tra cumulate</b>	0,152	0,180	0,055	0,303	0,362	<b>0,421</b>	0,368	0,000

La differenza massima **D** è uguale a 0,421.

Il valore da confrontare con le tabelle è dato da

$$D \times n_1 \times n_2$$

dove **D = 0,421**; **n<sub>1</sub> = 19** e **n<sub>2</sub> = 17**

dai quali si ottiene

$$0,421 \cdot 19 \cdot 17 = 135,98$$

un risultato uguale a 135,98.

La tabella dei valori critici per un **test a due code**

con **n<sub>1</sub> = 19** e **n<sub>2</sub> = 17** alla probabilità **a = 0.05** riporta 141.

Il valore calcolato è inferiore: la probabilità che le differenze siano imputabili al caso è superiore a 0.05 e quindi non si può rifiutare l'ipotesi nulla.

ESEMPIO 3. Con un primo esperimento si è voluto valutare l'effetto di un tossico alla concentrazione del 2%, immettendo in un acquario 150 dafnie per 10 giorni. In un secondo momento, si è valutato l'effetto della stessa sostanza alla concentrazione 3% e sono state immesse 200 dafnie.

Nella tabella sottostante, sono riportati i decessi contati ogni giorno nei due differenti esperimenti.

NUMERO OSSERVATO DI DECESSI PER GIORNO

	Giorno I	Giorno II	Giorno III	Giorno IV	Giorno V	oltre V G.
Concentr. 2%	<b>22</b>	<b>43</b>	<b>15</b>	<b>18</b>	<b>16</b>	<b>36</b>
Concentr. 3%	<b>19</b>	<b>39</b>	<b>31</b>	<b>52</b>	<b>59</b>	<b>0</b>

Dai due esperimenti di laboratorio è dimostrato che la concentrazione maggiore abbia una letalità significativamente superiore, come appare logico attendersi?

Risposta.

**E' un test ad una coda, con due campioni di grandi dimensioni.** Occorre verificare se la concentrazione al 3% ha una frequenza relativa maggiore nei valori bassi e quindi una frequenza relativa minore nei valori alti.

Si deve calcolare la differenza massima tra le due distribuzioni cumulate osservate, dopo trasformazione nelle frequenze relative.

FREQUENZE RELATIVE DI DECESSI PER GIORNO

	Giorno I	Giorno II	Giorno III	Giorno IV	Giorno V	Oltre V
<b>Concentr. 2%</b>						
freq. relativa	0,147	0,287	0,100	0,120	0,106	0,240
<b>distr. cumulata</b>	<b>0,147</b>	<b>0,434</b>	<b>0,534</b>	<b>0,654</b>	<b>0,760</b>	<b>1,000</b>
<b>Concentr. 3%</b>						
freq. relativa	0,095	0,195	0,155	0,260	0,295	0,000
<b>distr. cumulata</b>	<b>0,095</b>	<b>0,290</b>	<b>0,445</b>	<b>0,705</b>	<b>1,000</b>	<b>1,000</b>
<b>Differenza</b>	0,052	0,144	0,089	-0,051	<b>-0,240</b>	0,000

Per un **test ad una coda**, la differenza massima nella direzione dell'ipotesi alternativa è 0,240. La sua significatività è stimata dalla distribuzione  $\chi^2$  con 2 gradi di libertà dopo il calcolo del valore

$$\chi^2_{(2)} = 4 \cdot D^2 \frac{n_1 \cdot n_2}{n_1 + n_2}$$

dove:

$$D = -0,240; \quad n_1 = 150; \quad n_2 = 200$$

dai quali si ottiene

$$\chi^2_{(2)} = 4 \cdot 0,240^2 \frac{150 \cdot 200}{150 + 200} = 0,2304 \frac{30000}{350} = 19,75$$

un risultato uguale a 19,75.

Alla probabilità  $\alpha = 0.001$  il valore critico per 2 gradi di libertà riportato nella tabella sinottica del  $\chi^2$  è uguale a 13,82. Il valore calcolato con i dati dell'esempio è superiore: si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa di una maggiore letalità del tossico alla concentrazione 3%.

Se il confronto fosse stato tra due tossici diversi A e B per un **test a due code**, in cui verificare le differenze nella distribuzione del numero di cavie decedute per giorno, la stima della significatività della differenza massima (uguale a 0,240) avrebbe dovuto essere confrontata, per la probabilità  $\alpha = 0.05$ , con il valore critico ottenuto dalla relazione

$$1,36 \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}} = 1,36 \cdot \sqrt{\frac{150 + 200}{150 \cdot 200}} = 0,147$$

che è uguale a 0,147

e alla probabilità  $\alpha = 0.001$  con il valore critico

$$1,95 \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}} = 1,95 \cdot \sqrt{\frac{150 + 200}{150 \cdot 200}} = 0,211$$

che è uguale a 0,211.

Per entrambe le formule,  $n_1 = 150$  e  $n_2 = 200$  osservazioni.

La differenza massima riscontrata tra le due distribuzioni cumulate sarebbe risultata significativa.

### 3.12. IL CHI QUADRATO CON IL METODO DI COCHRAN E DI MANTEL-HAENSZEL

Quando si dispone di una **serie di tabelle 2 x 2, quindi una distribuzione di frequenze a tre dimensioni**, un metodo per valutare l'associazione tra due variabili categoriali, raccomandata da F. Yates nel 1955 (*The use of transformations and maximum likelihood in the analysis of quantal*

experiments involving two treatments, pubblicato su *Biometrika*, 42, pp. 382-403) e da J. L. Fleiss nel 1970 (*On the asserted invariance of the odds ratio*, pubblicato su *Brit. J. prev. soc. Med.*, 24, pp. 45-46) è la cosiddetta **differenza standardizzata**

$$d = \frac{p_1 - p_2}{\bar{p} \cdot (1 - \bar{p})}$$

dove

$p_1$  e  $p_2$  sono rispettivamente le frequenze relative dei due campioni a confronto,  $\bar{p}$  è la media ponderata delle due frequenze relative (vedi cap. III).

Estesa a più coppie di gruppi a confronto, come possono essere  $k$  tabelle  $2 \times 2$ , con misure di associazione la **differenza standardizzata** diventa

$$d_k = \frac{p_{k1} - p_{k2}}{\bar{p}_k \cdot (1 - \bar{p}_k)}$$

ed il suo errore standard e.s.( $d_k$ ) è

$$\text{e.s.}(d_k) = \sqrt{\frac{1}{\bar{p}_k \cdot (1 - \bar{p}_k)} \cdot \frac{n_k}{n_{k1} \cdot n_{k2}}}$$

dal quale si ricava

$$w_k = \frac{\bar{p}_k \cdot (1 - \bar{p}_k) \cdot n_{k1} \cdot n_{k2}}{n_k}$$

Derivato dal metodo che, per grandi campioni, utilizza la distribuzione  $z$  e sulla base della relazione

$$c_{(n)}^2 = \sum_{i=1}^n z_i^2$$

il **metodo del  $c^2$  proposto da W. G. Cochran** nel 1954 (*Some methods of strengthening the common  $c^2$  tests*, pubblicato da *Biometrics*, 10, pp. 417-451)

nel caso di **due campioni** è

$$c_{(1)}^2 = \frac{(p_1 - p_2)^2}{\bar{p} \cdot (1 - \bar{p}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

e nel caso di **k campioni** diventa

$$c_{(k)}^2 = \sum \frac{(p_{k1} - p_{k2})^2}{\bar{p}_k \cdot (1 - \bar{p}_k) \cdot \left(\frac{1}{n_{k1}} + \frac{1}{n_{k2}}\right)}$$

Lo **stesso risultato** è ottenuto con la formula abbreviata

$$c_{(k)}^2 = w_k \cdot d_k^2$$

dove  $w_k$  e  $d_k$  sono calcolati con le formule appena presentate.

ESEMPIO. Un'applicazione di questi concetti può essere tratta dal volume di Joseph L. Fleiss del 1973 (*Statistical Methods for Rates and Proportions*, John Wiley & Sons, New York, p. IX + 223). L'impostazione dell'esempio è stato leggermente variata, affinché la sua utilizzazione nella ricerca ambientale fosse meglio compresa.

La percentuale di persone con asma o disturbi respiratori in genere è ritenuta un indicatore dell'inquinamento atmosferico; ma questi sintomi possono essere determinati anche da allergie ai pollini, la cui diffusione dipende dalla stagione. Per valutare la significatività delle differenze tra due zone, tenendo in considerazione la variabilità stagionale, con visite dell'ufficiale sanitario nelle scuole elementari di due zone della stessa città è stata valutata la frequenza di alunni con malattie polmonari, ripetendo la verifica in tre stagioni diverse (autunno, inverno, primavera).

I risultati possono essere presentati con **tre tabelle 2 x 2**, ognuna relativa ad una rilevazione:

#### Rilevazione I

##### Diagnosi

Z		Sani	Mal.	Tot
O	A	67	38	105
N	B	72	33	105
A	Tot	139	71	210

#### Rilevazione II

##### Diagnosi

Z		Sani	Mal.	Tot
O	A	108	44	152
N	B	113	61	174
A	Tot	221	105	326

#### Rilevazione III

		Diagnosi		
Z		Sani	Mal.	Tot
O	A	102	43	145
N	B	112	33	145
A	Tot	214	76	290

Risposta:

I - **I dati** di queste tre tabelle **possono essere presentati in modo più schematico**, ma con la stessa quantità d'informazione, in una tabella unica:

**Proporzioni di individui affetti da malattie polmonari sul numero di individui campionati**

Rilevazione	Zona A		Zona B	
	$p_{kA}$	$n_{kA}$	$P_{kB}$	$N_{kB}$
I	0,362	105	0,314	105
II	0,289	152	0,351	174
III	0,297	145	0,228	145

Essa offre il vantaggio di evidenziare meglio i dati, per quantificare i diversi valori del chi-quadrato con **il metodo di Cochran**.

II - A questo scopo, con le formule precedenti, **si calcolano le quantità riportate** nella tabella

	1	2	3	4	5	6	7
Rilevazione	$p_{kA} - p_{kB}$	$\bar{p}_k$	$d_k$	$n_k$	$w_k$	$w_k d_k$	$w_k d_k^2$
I	0,048	0,338	0,215	210	11,75	2,53	0,54
II	-0,062	0,322	-0,284	326	17,69	-5,02	1,43
III	0,069	0,263	0,356	290	14,06	5,00	1,78
Totale	-----	-----	-----	-----	43,50	2,51	3,75

Limitando la dimostrazione dei vari passaggi solo alla II rilevazione

$$1) \quad p_{kA} - p_{kB} = 0,289 - 0,351 = -0,062$$

$$2) \quad \bar{p}_k = \frac{(0,289 \cdot 152) + (0,351 \cdot 174)}{152 + 174} = \frac{54 + 61}{326} = 0,322$$

$$3) \quad d_k = \frac{0,289 - 0,351}{0,322 \cdot 0,678} = \frac{-0,062}{0,218} = -0,284$$

$$4) \quad n_k = 326$$

$$5) \quad w_k = \frac{0,322 \cdot 0,678 \cdot 152 \cdot 174}{326} = \frac{0,218 \cdot 26448}{326} = 17,69$$

$$6) \quad w_k d_k = 17,69 \cdot (-0,284) = -5,02$$

$$7) \quad w_k d_k^2 = 17,69 \cdot (-0,284)^2 = 1,43$$

III - L'analisi dei risultati (ultima colonna dell'ultima tabella) permette di concludere che non sono significativi

- nessuno dei **tre valori di chi quadrato** (0,54; 1,43; 1,78), **ognuno con 1 gdl**,
- né il **chi quadrato totale** (3,75) **con 3 gdl**.

**Il test di Cochran permette di scomporre questo chi quadrato con k gdl**, quanti sono le tabelle 2 x 2. Con il test detto per l'omogeneità delle differenze standardizzate, è possibile verificare **se le tre rilevazioni hanno dato risposte omogenee** o significativamente differenti tra le tre stagioni, mediante la formula

$$c_{omog.}^2 = \sum_{k=1}^3 w_k d_k^2 - \frac{\left( \sum_{k=1}^3 w_k d_k \right)^2}{\sum_{k=1}^3 w_k}$$

IV - Con i dati dell'esempio

$$c_{omog.}^2 = 3,75 - \frac{2,51^2}{43,5} = 3,75 - 0,145 = 3,605$$

il chi quadrato per l'omogeneità risulta uguale a 3,605 con 2 gdl; di conseguenza, non è significativo (infatti per  $\alpha = 0.05$  il valore critico del  $\chi^2$  è uguale a 5,99).

V - La significatività della **differenza d complessiva tra le due zone** nella proporzione di persone ammalate è ottenuta attraverso la stima di d

$$d = \frac{\sum_{k=1}^3 w_k d_k}{\sum_{k=1}^3 w_k} = \frac{2,51}{43,5} = 0,058$$

e del suo errore standard e.s.(d)

$$e.s.(d) = \frac{1}{\sqrt{\sum_{k=1}^3 w_k}} = \frac{1}{\sqrt{43,5}} = 0,152$$

Il valore del chi quadrato per verificare **la significatività di questa differenza media standardizzata**,

cioè dell'**associazione tra malattia e zona** in complesso è

$$c_{assoc.}^2 = \left( \frac{d}{e.s.(d)} \right)^2 = \left( \frac{0,058}{0,152} \right)^2 = 0,146$$

risulta uguale a **0,146 con 1 gdl**.

La conclusione è che non esiste una differenza significativa nella frequenza relativa di ammalati nelle due zone. **Tenendo presente i diversi gdl, si può sostenere che è maggiore la differenza tra stagioni che tra zone, seppure nessuna delle due sia risultata significativa.**

Il **metodo del  $c^2$  proposto da N. Mantel e W. Haenszel** nel 1959 (*Statistical aspects of the analysis of data from retrospective studies of disease* pubblicato su *J. natl. Cancer Inst.*, 22, pp. 719-748), reso più generale da N. Mantel nel 1963 (con l'articolo *Chi-square tests with one degree one freedom. Extension of the mantel-Haenszel procedure*, pubblicato su *J. Amer. statist. Assoc.*, 58, pp.690-700) può essere visto come **la correzione di quello di Cochran per piccoli campioni**.

Utilizzando la stessa impostazione di Cochran, nella parte finale

$$c_{(k)}^2 = w_k \cdot d_k^2$$

apporta una correzione

- sia a  $w_k$

$$w_k = \frac{\bar{p}_k \cdot (1 - \bar{p}_k) \cdot n_{k1} \cdot n_{k2}}{n_k - 1}$$

- sia a  $d_k$

$$d_k = \frac{p_{k1} - p_{k2}}{\bar{p}_k \cdot (1 - \bar{p}_k)} \cdot \frac{n_k - 1}{n_k}$$

togliendo **1** a  $n_k$ , (il numero di osservazioni del gruppo **k**).

**Quando il numero di dati è grande, la differenza dal metodo precedente è assolutamente trascurabile**; diviene relativamente importante, quando il campione è di piccole dimensioni.

Utilizzando gli stessi dati dell'esempio precedente, si ottiene la correzione delle colonne

$d_k$  e  $w_k$

Rilevazione	$p_{kA} - p_{kB}$	$\bar{p}_k$	$d_k$	$n_k$	$w_k$	$w_k d_k$	$w_k d_k^2$
I	0,048	0,338	<b>0,214</b>	210	<b>11,80</b>	2,53	<b>0,54</b>
II	-0,062	0,322	<b>-0,283</b>	326	<b>17,72</b>	-5,02	<b>1,43</b>
III	0,069	0,263	<b>0,355</b>	290	<b>14,09</b>	5,00	<b>1,78</b>
Totale	-----	-----	-----	-----	<b>43,61</b>	2,51	<b>3,75</b>

Poiché il campione utilizzato può essere considerato grande, se il calcolo è effettuato con due sole cifre decimali, il valore del chi quadrato di ogni rilevazione e quello totale (3,78) restano invariati.

### 3.13. ESERCIZI SVOLTI PER DATI IN TABELLE DI CONTINGENZA

ESERCIZIO 1. Per lo studio di frequenze alleliche del marcatore genetico ossidasi in popolazioni naturali di faggio di 4 differenti località dell'Appennino (Abetone, Pisanino, Pradarena, Pradaccio), sono state rilevate le frequenze di 3 alleli.

Le frequenze osservate sono quelle riportate nella tabella sottostante.

FREQUENZE OSSERVATE

	Allele 1	Allele 2	Allele 3	Totale
Abetone	<b>7</b>	<b>244</b>	<b>49</b>	300
Pisanino	<b>8</b>	<b>156</b>	<b>24</b>	188
Pradarena	<b>22</b>	<b>231</b>	<b>31</b>	284
Pradaccio	<b>143</b>	<b>185</b>	<b>116</b>	444
Totale	180	816	220	1216

Esiste una differenza significativa nella distribuzione delle 3 frequenze alleliche? Si può sostenere che esiste associazione tra località e frequenze alleliche, per cui i 3 alleli non hanno la stessa distribuzione percentuale nelle 4 località?

Risposta.

Dai totali marginali delle frequenze osservate, si calcolano le frequenze che dovremmo attenderci se fosse vero che non esiste differenza tra le percentuali dei 3 alleli nelle 4 località (distribuzione attesa, nella condizione che  $H_0$  sia vera).

FREQUENZE ATTESE SECONDO L'IPOTESI NULLA

	Allele 1	Allele 2	Allele 3	Totale
Abetone	<b>44,4</b>	<b>201,3</b>	<b>54,3</b>	300
Pisanino	<b>27,8</b>	<b>126,2</b>	<b>34,0</b>	188
Pradarena	<b>42,0</b>	<b>190,6</b>	<b>51,4</b>	284
Pradaccio	<b>65,8</b>	<b>297,9</b>	<b>80,3</b>	444
Totale	180	816	220	1216

Con la formula generale  $(\text{Oss.} - \text{Att})^2 / \text{Att}$ , si calcola il valore del  $\chi^2$ , che avrà 6 gradi di libertà.

$$\begin{aligned}
c_{(6)}^2 = & \frac{(7-44,4)^2}{44,4} + \frac{(244-201,3)^2}{201,3} + \frac{(49-54,3)^2}{54,3} + \frac{(8-27,8)^2}{27,8} + \frac{(156-126,2)^2}{126,2} + \\
& + \frac{(24-34)^2}{34} + \frac{(22-42)^2}{42} + \frac{(231-190,6)^2}{190,6} + \frac{(31-51,4)^2}{51,4} + \frac{(143-65,8)^2}{65,8} + \\
& + \frac{(185-297,9)^2}{297,9} + \frac{(116-80,3)^2}{80,3} = 240,571
\end{aligned}$$

Si ottiene un chi quadrato con 6 gradi  $c_{(6)}^2$  di libertà uguale a 240,571.

Il valore risulta molto alto; la probabilità  $\alpha$  che sia casuale è molto piccola, inferiore a 0,0001.

Si conclude che esiste una differenza altamente significativa, fra le distribuzioni percentuali dei 3 alleli nelle 4 località.

Per una lettura più dettagliata del test, sarebbe utile valutare quanto ogni casella contribuisce al valore complessivo del chi quadrato. Il confronto tra la distribuzione osservata e quella attesa evidenzia che, rispetto alla media delle 4 zone, all'Abetone e al Pisanino si ha un eccesso dell'allele 2 e una carenza dell'allele 1 e dell'allele 3, mentre al Pradaccio e a Pradarena si ha un eccesso degli alleli 1 e 3 e una carenza dell'allele 2.

ESERCIZIO 2. Si sono sottoposti dei cloni di *Daphnia magna* a quattro diversi trattamenti o regimi alimentari. Dopo 39 giorni si è fatto un bilancio complessivo di quanti sono stati i morti (e i sopravvissuti) in ogni campione.

Si intende verificare se il tasso di mortalità è uguale per i 4 diversi trattamenti.

Le differenze riscontrate sono dovute al caso, oppure sono imputabili al diverso trattamento alimentare?

#### FREQUENZE OSSERVATE

	Cloni morti	Cloni sopravvissuti	Totale cloni
Trattamento I	<b>6</b>	<b>23</b>	29
Trattamento II	<b>2</b>	<b>26</b>	28
Trattamento III	<b>8</b>	<b>22</b>	30
Trattamento IV	<b>3</b>	<b>20</b>	23
Totale	19	91	110

Risposta.

Si calcolano le frequenze attese

FREQUENZE ATTESE SECONDO L'IPOTESI NULLA

	Cloni morti	Cloni sopravvissuti	Totale cloni
Trattamento I	<b>5,0</b>	<b>24,0</b>	29
Trattamento II	<b>4,8</b>	<b>23,2</b>	28
Trattamento III	<b>5,2</b>	<b>24,8</b>	30
Trattamento IV	<b>4,0</b>	<b>19,0</b>	23
Totale	19	91	110

e successivamente, attraverso la formula generale, il valore del chi quadrato con 3 gradi di libertà

$$\begin{aligned} \chi_{(3)}^2 &= \frac{(6-5)^2}{5} + \frac{(23-24)^2}{24} + \frac{(2-4,8)^2}{4,8} + \frac{(26-23,2)^2}{23,2} + \\ &+ \frac{(8-5,2)^2}{5,2} + \frac{(22-24,8)^2}{24,8} + \frac{(3-4)^2}{4} + \frac{(20-19)^2}{19} = 4,02325 \end{aligned}$$

Si ottiene un valore  $\chi_{(3)}^2$  uguale a 4,023 che, per 3 gradi di libertà, è associato ad una probabilità superiore al 25%. La probabilità che sia vera l'ipotesi nulla è molto elevata, superiore al 25%.

Non è possibile rifiutare l'ipotesi nulla: non è dimostrato che i 4 trattamenti determinino una mortalità significativamente differente.

ESERCIZIO 3. Nella tabella seguente, sono riportati i risultati di un esperimento sulla schiusa di uova di *Heterocypris incongruens*, mantenute a diverse condizioni di temperatura.

FREQUENZE OSSERVATE

Temperatura	Uova schiuse	Uova non-schiuse	Totale
16°C	<b>131</b>	<b>32</b>	163
24°C	<b>100</b>	<b>64</b>	164
28°C	<b>90</b>	<b>91</b>	181
Totale	320	188	508

Trarre le conclusioni.

Risposta.

Si calcolano le frequenze attese

FREQUENZE ATTESE SECONDO L'IPOTESI NULLA

Temperatura	Uova schiuse	Uova non-schiuse	Totale
16°c	<b>102,7</b>	<b>60,3</b>	163
24°c	<b>103,3</b>	<b>60,7</b>	164
28°c	<b>114,0</b>	<b>67,0</b>	181
Totale	320	188	508

e attraverso la formula generale si può computare il valore complessivo del chi quadrato,

$$\begin{aligned}
 \mathbf{c}_{(2)}^2 &= \frac{(131 - 102,7)^2}{102,7} + \frac{(32 - 60,3)^2}{60,3} + \frac{(100 - 103,3)^2}{103,3} + \\
 &+ \frac{(64 - 60,7)^2}{60,7} + \frac{(90 - 114)^2}{114} + \frac{(91 - 67)^2}{67} = 35,004
 \end{aligned}$$

con il contributo fornito da ogni casella

(OSS. - ATT.)<sup>2</sup> / ATT PER OGNI CASELLA

Temperatura	Uova schiuse	Uova non-schiuse	Totale
16° C	<b>7,798</b>	<b>13,282</b>	
24° C	<b>0,105</b>	<b>0,179</b>	
28° C	<b>5,053</b>	<b>8,587</b>	
Totale			<b>35,004</b>

Il valore del  $\mathbf{c}_{(2)}^2$  (uguale a 35,004) è molto elevato; per 2 gradi di libertà, è associato ad una probabilità  $\alpha$  estremamente bassa, inferiore a 0.001. Si può affermare che le percentuali di uova schiuse alle 3 diverse temperature sono significativamente differenti.

I contributi maggiori sono forniti dalle due temperature estreme. Alla temperatura più bassa (16° C) il numero di uova che non si sono schiuse è molto maggiore, mentre alla temperatura più alta (28° C) è molto minore delle frequenze attese, se fosse stata vera l'ipotesi nulla che la temperatura non influisce. Alla temperatura intermedia (24° C), la percentuale di uova schiuse è molto vicina alla media dei 3 esperimenti.

ESERCIZIO 4. Si sono mantenute per 24 ore in due soluzioni acquose, contenenti uguale concentrazione di rame, due gruppi formati da 48 individui ciascuno, di 2 specie di protozoi: *Euplotes patella* e *Paramecium caudatum*.

Alla fine dell'esperimento, si è contato il numero di individui morti in ognuno dei due campioni. I risultati sono riportati nella tabella sottostante.

Verificare se la percentuale di decessi è diversa per le 2 specie, sottoposte alle medesime condizioni d'inquinamento.

FREQUENZE OSSERVATE

	Morti	Sopravvissuti	Totale
<i>Euplotes p.</i>	<b>11</b>	<b>37</b>	48
<i>Paramecium c.</i>	<b>15</b>	<b>33</b>	48
Totale	26	70	96

Risposta.

E' una tabella di contingenza 2 x 2 .

Calcolando il chi-quadro, che avrà 1 g.d.l., con la formula rapida

$$c_{(1)}^2 = \frac{(11 \cdot 33 - 37 \cdot 15)^2 \cdot 96}{48 \cdot 48 \cdot 26 \cdot 70} = 0,843$$

si ottiene un valore uguale a 0, 843.

Una obiezione al risultato può essere che il numero totale di osservazioni non può essere considerato grande. In queste condizioni, a scopo cautelativo e quindi per una maggiore attendibilità delle conclusioni, è conveniente ricorrere al calcolo del chi quadrato con la correzione di YATES per la continuità

$$c_{(1)}^2 = \frac{\left( |11 \cdot 33 - 37 \cdot 15| - \frac{96}{2} \right)^2 \cdot 96}{48 \cdot 48 \cdot 26 \cdot 70} = 0,4747$$

che (con un valore uguale a 0,4747) si dimostra più conservativa.

Con questi dati, è possibile utilizzare anche il metodo del **log likelihood ratio** o  $G^2$ :

$$G^2 = 2 \cdot \{ [ 11 \cdot \ln(11) + 37 \cdot \ln(37) + 15 \cdot \ln(15) + 33 \cdot \ln(33) + 96 \cdot \ln(96) ] - [ 48 \cdot \ln(48) + 48 \cdot \ln(48) + 26 \cdot \ln(26) + 70 \cdot \ln(70) ] \} =$$

$$G^2 = 2 \cdot \{ [(11 \times 2,398) + (37 \times 3,611) + (15 \times 2,708) + (33 \times 3,497) + (96 \times 4,564)] - [(48 \times 3,871) + (48 \times 3,871) + (26 \times 3,258) + (70 \times 4,248)] \} =$$

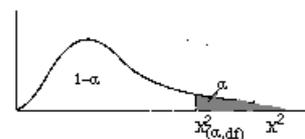
$$G^2 = 2 \cdot (754,104 - 753,742) = 2 \times 0,362 = 0,724$$

Il valore del  $\chi^2$  è piccolo anche con questo metodo. E' simile a quelli prima calcolati; è nettamente inferiore al valore critico di 3,83 corrispondente alla probabilità  $\alpha = 0.05$ .

Le variazioni riscontrate rientrano quindi fra quelle imputabili solo al caso.

Non è possibile dimostrare che le due specie hanno reagito in modo differente alla stessa concentrazione di rame.

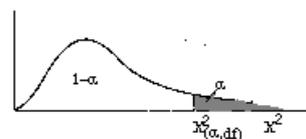
VALORI CRITICI DELLA DISTRIBUZIONE  $\chi^2$  (con gdl da 1 a 30)



Gradi di libertà	Area della coda superiore											
	.995	.99	.975	.95	.90	.75	.25	.10	.05	.025	.01	.005
1			0.001	0.004	0.016	0.102	1.323	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.575	2.773	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	1.213	4.108	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	1.923	5.385	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	2.675	6.626	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	3.455	7.841	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	4.255	9.037	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	5.071	10.219	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	5.899	11.389	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	6.737	12.549	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	7.584	13.701	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	8.438	14.845	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	9.299	15.984	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	10.165	17.117	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	11.037	18.245	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	11.912	19.369	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	12.792	20.489	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	13.675	21.605	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	14.562	22.718	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	15.452	23.828	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	16.344	24.935	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	17.240	26.039	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	18.137	27.141	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	19.037	28.241	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	19.939	29.339	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	20.843	30.435	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	21.749	31.528	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	22.657	32.620	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	23.567	33.711	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	24.478	34.800	40.256	43.773	46.979	50.892	53.672

(segue)

VALORI CRITICI DELLA DISTRIBUZIONE  $\chi^2$  (con gdl da 31 a 60)



Gradi di libertà	Area della coda superiore											
	.995	.99	.975	.95	.90	.75	.25	.10	.05	.025	.01	.005
31	14.458	15.655	17.539	19.281	21.434	25.390	35.887	41.422	44.985	48.232	52.191	55.003
32	15.134	16.362	18.291	20.072	22.271	26.304	36.973	42.585	46.194	49.480	53.486	56.328
33	15.815	17.074	19.047	20.867	23.110	27.219	38.058	43.745	47.400	50.725	54.776	57.648
34	16.501	17.789	19.806	21.664	23.952	28.136	39.141	44.903	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	24.797	29.054	40.223	46.059	49.802	53.203	57.342	60.275
36	17.887	19.233	21.336	23.269	25.643	29.973	41.304	47.212	50.998	54.437	58.619	61.581
37	18.586	19.960	22.106	24.075	26.492	30.893	42.383	48.363	52.192	55.668	59.892	62.881
38	19.289	20.691	22.878	24.884	27.343	31.815	43.462	49.513	53.384	56.896	61.162	64.181
39	19.996	21.426	23.654	25.695	28.196	32.737	44.539	50.660	54.572	58.120	62.428	65.476
40	20.707	22.164	24.433	26.509	29.051	33.660	45.616	51.805	55.758	59.342	63.691	66.766
41	21.421	22.906	25.215	27.326	29.907	34.585	46.692	52.949	56.942	60.561	64.950	68.053
42	22.138	23.650	25.999	28.144	30.765	35.510	47.766	54.090	58.124	61.777	66.206	69.336
43	22.859	24.398	26.785	28.965	31.625	36.436	48.840	55.230	59.304	62.990	67.459	70.616
44	23.584	25.148	27.575	29.787	32.487	37.363	49.913	56.369	60.481	64.201	68.710	71.893
45	24.311	25.901	28.366	30.612	33.350	38.291	50.985	57.505	61.656	65.410	69.957	73.166
46	25.041	26.657	29.160	31.439	34.215	39.220	52.056	58.641	62.830	66.617	71.201	74.437
47	25.775	27.416	29.956	32.268	35.081	40.149	53.127	59.774	64.001	67.821	72.443	75.704
48	26.511	28.177	30.755	33.098	35.949	41.079	54.196	60.907	65.171	69.023	73.683	76.969
49	27.249	28.941	31.555	33.930	36.818	42.010	55.265	62.038	66.339	70.222	74.919	78.231
50	27.991	29.707	32.357	34.764	37.689	42.942	56.334	63.167	67.505	71.420	76.154	79.490
51	28.735	30.475	33.162	35.600	38.560	43.874	57.401	64.295	68.669	72.616	77.386	80.747
52	29.481	31.246	33.968	36.437	39.433	44.808	58.468	65.422	69.832	73.810	78.616	82.001
53	30.230	32.018	34.776	37.276	40.308	45.741	59.534	66.548	70.993	75.002	79.843	83.253
54	30.981	32.793	35.586	38.116	41.183	46.676	60.600	67.673	72.153	76.192	81.069	84.502
55	31.735	33.570	36.398	38.958	42.060	47.610	61.665	68.796	73.311	77.380	82.292	85.749
56	32.490	34.350	37.212	39.801	42.937	48.546	62.729	69.919	74.468	78.567	83.513	86.994
57	33.248	35.131	38.027	40.646	43.816	49.482	63.793	71.040	75.624	79.752	84.733	88.236
58	34.008	35.913	38.844	41.492	44.696	50.419	64.857	72.160	76.778	80.936	85.950	89.477
59	34.770	36.698	39.662	42.339	45.577	51.356	65.919	73.279	77.931	82.117	87.166	90.715
60	35.534	37.485	40.482	43.188	46.459	52.294	66.981	74.397	79.082	83.298	88.379	91.952
70	43.275	45.441	48.757	51.739	55.329	61.698	77.576	85.527	90.531	95.023	100.43	104.22
80	51.172	53.540	57.153	60.391	64.277	71.144	88.130	96.578	101.88	106.63	112.33	116.32
90	59.196	61.754	65.466	69.126	73.291	80.624	98.649	107.57	113.15	118.14	124.12	128.30
100	67.327	70.064	74.221	77.929	82.358	90.133	109.14	118.50	124.34	129.56	135.81	140.17

Valori critici di D nel test di Kolmogorov-Smirnov per un campione

Dimensione del campione ( <i>N</i> )	Livello di significatività per $D = \max  F_0(x) - S_N(x) $				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.381	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.21	.22	.24	.27	.32
30	.19	.20	.22	.24	.29
35	.18	.19	.21	.23	.27
Oltre 35	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$



Valori critici nel test, a due code, di Kolmogorov-Smirnov per 2 campioni indipendenti. - Per essere significativo,  $D \cdot n_1 \cdot n_2$  deve essere uguale o maggiore del valore tabulato. - Il valore superiore è associato alla probabilità 0.10; quello centrale alla probabilità 0.05 e quello inferiore alla probabilità 0.01, per ogni coppia di  $n_1$  e  $n_2$ .

$n_1$	$n_2$																									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
1																										
2					10	12	14	16	18	18	20	22	24	24	26	28	30	32	32	34	36	38	38	40	42	42
3			9	12	15	15	18	21	21	24	27	27	30	33	33	36	36	39	42	45	45	48	48	51	54	57
4			12	16	16	18	21	24	27	28	29	36	35	38	40	44	44	46	49	52	52	56	57	60	63	63
5		10	15	16	20	24	25	27	30	35	35	36	46	42	50	48	50	52	56	60	60	63	65	67	75	
6		12	15	18	24	30	28	30	33	36	38	48	46	48	51	54	56	66	64	66	69	70	73	78	78	
7		14	18	21	25	28	35	34	36	40	44	46	50	56	56	59	61	65	69	72	77	77	80	84	86	
8		16	21	24	27	30	34	40	40	44	48	52	54	58	60	72	68	72	74	50	81	84	89	96	104	
9		18	21	27	30	33	36	40	54	50	52	57	59	63	69	69	74	81	80	84	90	91	94	99	101	
10		18	24	28	35	36	40	44	50	60	57	60	64	68	75	76	79	82	85	100	95	98	101	106	110	
11		20	27	29	35	38	44	48	52	57	66	64	67	73	76	80	85	88	92	96	101	110	108	111	117	
12		22	27	36	36	48	46	52	57	60	64	72	71	78	84	88	90	96	99	104	108	110	113	132	120	
13		24	30	35	40	46	50	54	59	64	67	71	91	78	87	91	96	99	104	108	113	117	120	125	131	
14		24	33	38	42	48	56	58	63	68	73	78	78	98	92	96	100	104	110	114	126	124	127	132	136	
15		26	33	40	50	51	56	60	69	75	76	84	87	92	105	101	105	111	114	125	126	130	134	141	145	
16		28	36	44	48	54	59	72	69	76	80	88	91	96	101	112	109	116	120	128	130	136	141	152	149	
17		30	36	44	50	56	61	68	74	79	85	90	96	100	105	109	136	118	126	132	136	142	146	151	156	
18		32	39	46	52	66	65	72	81	82	88	96	99	104	111	116	123	128	133	162	142	152	159	160	180	
19	19	32	42	49	56	64	69	74	80	85	92	99	104	110	114	120	126	133	152	144	147	152	159	164	168	
20	20	34	42	52	60	66	72	80	84	100	96	104	108	114	125	128	132	136	144	160	154	160	164	172	180	
21	21	36	45	52	60	69	77	81	90	95	101	108	113	126	126	130	136	144	147	154	168	163	171	177	182	
22	22	38	48	56	63	70	77	84	91	98	110	110	117	124	130	136	142	148	152	160	163	198	173	182	189	
23	23	38	48	57	65	73	80	89	94	101	108	113	120	127	134	141	146	152	159	164	171	173	207	183	195	
24	24	40	51	60	67	78	84	96	99	106	111	132	125	132	141	152	151	162	164	172	177	182	193	216	204	
25	25	42	54	63	75	78	86	95	101	110	117	120	131	136	145	149	156	162	168	180	182	189	195	204	225	

## CAPITOLO IV

### VERIFICA DELLE IPOTESI

#### 4.1. RISULTATI SIGNIFICATIVI E NON-SIGNIFICATIVI

Nel capitolo precedente sul chi quadrato, dopo l'elenco delle fasi elementari in cui è possibile scomporre la serie di passaggi logici richiesti dall'uso corretto di un test statistico, è stata applicata la procedura per la **verifica delle ipotesi**. Sono stati spiegati in modo sommario ed applicati i concetti di

- **ipotesi nulla  $H_0$  e ipotesi alternativa  $H_1$ ,**
- **test unilaterali o a una coda e test bilaterali o a due code,**
- **valori critici e livello di significatività,**
- **zona o regione di rifiuto e zona o regione di accettazione o di non-rifiuto.**

Oltre a rendere familiari i termini, con varie applicazioni in condizioni differenti è stato dimostrato in modo sperimentale come una loro utilizzazione elementare sia relativamente semplice, quasi intuitiva. Tuttavia, per una esatta comprensione dei metodi e al fine di procedere nell'approfondimento dei concetti verso livelli più sofisticati ed applicazioni più complesse, è indispensabile **conoscere con chiarezza:**

- **le convenzioni abitualmente usate nell'applicazione dei test statistici,**
- **alcune nozioni teoriche fondamentali sull'inferenza.**

Si può definire **test statistico** una procedura che, sulla base di dati campionari e con un certo grado di probabilità, consente di decidere se è ragionevole respingere l'ipotesi nulla  $H_0$  (ed accettare implicitamente l'ipotesi alternativa  $H_1$ ) oppure se non esistono elementi sufficienti per respingerla.

**La scelta sulle due ipotesi ( $H_0$  e  $H_1$ )** è fondata sulla **probabilità di ottenere per caso il risultato osservato nel campione o un risultato più estremo, nella condizione che l'ipotesi nulla  $H_0$  sia vera**. Quanto più tale probabilità è piccola, tanto più è improbabile che  $H_0$  sia vera.

**La probabilità** può essere calcolata direttamente nei test esatti (come nell'uso della distribuzione binomiale o nel metodo di Fisher per tabelle 2 x 2, per ricordare solamente test già utilizzati); ma abitualmente dipende dal valore stimato con il test, per il quale sono state costruite **tabelle di densità di probabilità** (come nel caso di test che ricorrono al valore di  $z$  o del  $c^2$ ).

L'insieme di valori ottenibili con il test formano la distribuzione campionaria dell'indice statistico e può essere diviso in due zone:

1 - la **zona di rifiuto** dell'ipotesi nulla, detta anche **regione critica**, che **corrisponde ai valori collocati agli estremi della distribuzione, quelli che hanno una probabilità piccola di verificarsi per caso, quando l'ipotesi nulla  $H_0$  è vera;**

2 - la **zona di accettazione** di  $H_0$ , che comprende **i restanti valori.**

Se il valore dell'indice statistico calcolato cade nella zona di rifiuto, si respinge l'ipotesi nulla  $H_0$ .

Per consolidata **convenzione internazionale, i livelli di soglia delle probabilità** ai quali di norma si ricorre sono tre: 0.05 (5%); 0.01 (1%); 0.001 (0.1%). Nella presentazione sintetica dei risultati e nella discussione conclusiva dei test, quando è possibile solo l'uso di tabelle sinottiche (riassuntive) con i valori critici, i differenti livelli di significatività sono indicati con una **simbologia e con parole chiave**, che hanno significati precisi, non equivoci o generici, nella terminologia statistica.

Le tre probabilità e i valori critici più frequentemente utilizzati sono definiti dalle parole chiave e sono indicati con i simboli mostrati nello schema sottostante:

LIVELLO DI PROBABILITA'	RISULTATO DEL TEST	SIMBOLO
P < 0.05 (livello 5%)	significativo	*
P < 0.01 (livello 1%)	molto significativo	**
P < 0.001 (livello 0,1%)	altamente significativo	***

Riportati di fianco al risultato del test, **uno** oppure **due** o **tre asterischi** indicano in modo simbolico che il risultato è significativo ad una probabilità minore rispettivamente del **5%**, dell'**1%** o dello **0,1%**. Convenzionalmente, in termini discorsivi, si dice che il risultato è **significativo**, **molto significativo** od **altamente significativo**, secondo i tre diversi livelli di probabilità ai quali viene rifiutata l'ipotesi nulla.

Questi **valori critici di probabilità** sono numeri tondi, **puramente orientativi e non possiedono particolari proprietà intrinseche.**

**Valori di probabilità leggermente differenti sia tra loro sia con quella prefissata a priori**, per esempio uguali a 0.0505 oppure a 0.0491, **conducono in realtà alle medesime conclusioni. Queste probabilità non sono tali da indurre, in modo certo o definitivo, ad accettare oppure a rifiutare l'ipotesi nulla perché leggermente inferiori o superiori** al valore soglia di 0.05 eventualmente prescelto: si deve parlare di **risultati tendenzialmente significativi**. I valori critici non devono essere assunti come confini rigidi, ma come indicazioni orientative per l'accettazione o il rifiuto dell'ipotesi nulla.

Sono concetti che sovente inducono alla diffidenza verso la statistica il ricercatore che vi ricorra per le prime volte. Egli vorrebbe una risposta precisa, universale e senza equivoci, attribuendo alla statistica il compito di scegliere senza incertezze o dubbi di qualsiasi natura. In realtà, **il livello di significatività non è sempre uguale nel tempo ed in ogni circostanza**. Anche secondo Fisher, il padre della statistica moderna al quale si deve soprattutto l'analisi della varianza, l'ipotesi deve essere accettata o rifiutata in relazione a ciascun caso particolare, alla sua evidenza e alle idee del ricercatore.

Nel capitolo precedente, con il test statistico  $\chi^2$ , si è potuto verificare se una specifica distribuzione osservata fosse in sostanziale accordo con una determinata ipotesi o teoria ( $H_0$ ) e se le differenze rilevate fossero imputabili al caso. Oppure, se le due distribuzioni fossero significativamente differenti ( $H_1$ ), per la presenza di un fattore noto o ignoto, comunque diversamente incidente sulle distribuzioni e tale da alterare le frequenze assolute delle classi a confronto. In termini generali, **l'ipotesi riguardava la forma di distribuzione della popolazione, che poteva essere diversa per almeno uno dei 4 parametri** (tendenza centrale, variabilità, simmetria, curtosi) **o per un loro effetto combinato**; in questi casi, **si parla di ipotesi funzionale**.

In altri esercizi, si è verificato se le percentuali o proporzioni rilevate in due campioni potessero essere giudicate statisticamente uguali ( $H_0$ ) oppure se esistesse una differenza significativa ( $H_1$ ), anche questa determinata da cause diversamente presenti nei campioni raccolti. **L'ipotesi formulata riguardava un solo parametro specifico della popolazione, che quasi sempre è la media oppure la varianza: è un' ipotesi parametrica**.

**L'ipotesi nulla è, in generale, l'ipotesi che si vorrebbe rifiutare**. Essa afferma che gli effetti osservati nei campioni sono dovuti a fluttuazioni casuali, sempre possibili quando esiste variabilità tra gli individui; si tratta di variazioni che sono relativamente tanto più marcate quanto più ridotto è il numero di osservazioni.

**L'ipotesi nulla  $H_0$  deve essere rifiutata solamente se esiste l'evidenza che la contraddice. E' importante comprendere che l'ipotesi nulla non è necessariamente vera, quando i dati campionari (eventualmente pochi) non sono tali da contraddirla. **L'ipotesi nulla non è mai provata o verificata; è solo possibile negarla o disapprovarla, sulla base di dati sperimentali.****

Contrapposta all'ipotesi nulla  $H_0$  si ha l'ipotesi alternativa  $H_1$ , in alcuni testi indicata con  $H_A$ . Essa, in rapporto al problema e al test utilizzato, può essere di tre tipi, tra loro mutuamente esclusivi: bilaterale, unilaterale in una direzione o unilaterale nell'altra.

In modo più dettagliato, con un test si verifica

$$H_0: q = q_0 \quad \text{contro} \quad H_1: q \neq q_0$$

oppure

$$H_0: q = q_0 \quad \text{contro} \quad H_1: q > q_0$$

oppure

$$H_0: q = q_0 \quad \text{contro} \quad H_1: q < q_0$$

dove

$q$  è il valore del parametro nel campione estratto dalla popolazione studiata e

$q_0$  è il valore dell'effetto teorico atteso o prescelto come confronto.

Secondo Fisher, non tutte le ipotesi possono essere scelte come ipotesi alternative: devono essere scelte sulla base del test e delle conoscenze acquisite prima dell'esperimento (ad esempio, di norma il confronto tra l'effetto di un farmaco e quello del placebo richiede un'ipotesi unilaterale).

Tra le tre ipotesi prima elencate,

- la prima ipotesi alternativa  $H_1$  è chiamata **bilaterale**,
- la seconda  $H_1$  è **unilaterale destra**,
- la terza  $H_1$  è **unilaterale sinistra**.

Mediante il ricorso ai test, nel capitolo precedente si è sempre pervenuti alla **stima di una probabilità complessiva**, che corrisponde a quella di ottenere differenze uguali o superiori a quelle sperimentalmente riscontrate, **nell'ipotesi che i due campioni a confronto fossero estratti dalla stessa popolazione**. Quando la probabilità è risultata inferiore a quella prescelta, si è concluso che **esisteva una differenza statisticamente significativa**.

Per una corretta comprensione dei concetti utilizzati in statistica, è importante evidenziare che, accettando questa conclusione, **è possibile commettere un errore: la differenza riscontrata nell'esperimento in realtà potrebbe non esistere**. Tuttavia, la conclusione è ugualmente corretta, poiché **con il test non si perviene ad una affermazione assoluta, ma ad una probabilità conosciuta di poter commettere un errore**.

Questi concetti possono essere ulteriormente approfonditi in modo chiaro e semplice, con due esempi. Come primo caso, si supponga di dare ad un giocatore una **moneta perfettamente bilanciata, ma di cui egli non conosca le caratteristiche**. Mediante alcuni lanci, egli deve decidere se la moneta è bilanciata ( $H_0$ ) oppure truccata ( $H_1$ ). Si supponga quindi che egli lanci questa moneta 6 volte e che ottenga croce tutte 6 le volte.

Se il giocatore fosse uno statistico ragionerebbe in questo modo. Avere questa risposta di 6 croci su 6 lanci è un evento raro ( $0,5^6 = 0,0156$  o 1,56%), se la moneta non fosse truccata ( $H_0$  vera). E' semplice calcolare che, con ipotesi bilaterale e quindi comprendendo anche la possibilità di avere 6

volte testa, la probabilità è esattamente uguale a 3,12%. Quindi, ottenere 6 volte testa oppure 6 volte croce sono eventi complessivamente poco probabili, seppure possibili. Se egli avesse prefissato come valore soglia la probabilità del 5%, con questo test statistico rifiuterebbe l'ipotesi nulla. Giungerebbe alla conclusione che tra atteso (3 volte teste e 3 volte croce su 6 lanci) ed osservato (6 volte croce) esiste una differenza significativa e che pertanto la moneta è truccata; ma noi sappiamo che in realtà essa non la è.

E' un errore, che in statistica si chiama **errore di I tipo** (in altri testi, di **prima specie**): **consiste nel rifiutare l'ipotesi nulla  $H_0$ , quando in realtà essa è vera.**

Si supponga ora, come secondo caso, che sempre all'insaputa del giocatore questa volta la **moneta sia truccata** e dia solo croce: con 3 lanci, ovviamente otterrebbe 3 volte croce.

In questo caso, se fosse uno statistico seguirebbe questo ragionamento. Se la moneta non fosse truccata ( $H_0$  vera), la probabilità di trovare per caso 3 volte croce è alta (uguale a  $0,5^3 = 0,125$  o 12,5%); pertanto, **non è possibile rifiutare l'ipotesi nulla**. Errando, arriverebbe alla conclusione che la moneta non è truccata. In questo caso, si ha l'**errore di II tipo** (o **seconda specie**).

**In statistica non è possibile eliminare questi due tipi di errore. E' possibile solamente ridurre la loro frequenza al minimo e conoscere con precisione la probabilità con la quale entrambe avvengono. Conoscendo la probabilità di sbagliare od il rischio di commettere errori, è possibile scegliere. La statistica è la scienza che permette di scegliere e prendere decisioni, perché fornisce la probabilità di errare che è associata ad ogni scelta e quindi di conoscere il rischio che si corre, se la scelta si dimostrasse errata.**

#### **4.2. PROCEDURA DI VERIFICA DELLE IPOTESI.**

##### **VERO O FALSO ? UTILE O DANNOSO ?**

Per evidenziare con un test l'effetto di un trattamento, nel **controllo di un'ipotesi statistica** è possibile commettere due tipi di errore:

- l'**errore di primo tipo o errore  $\alpha$  (alfa)**, se si rifiuta l'ipotesi nulla quando in realtà essa è vera;
- l'**errore di secondo tipo o errore  $\beta$  (beta)**, se si accetta l'ipotesi nulla, quando in realtà essa è falsa.

**La probabilità di commettere l'errore di I tipo si chiama livello di significatività ed è indicata convenzionalmente con  $\alpha$  (alfa); essa corrisponde alla probabilità che il valore campionario dell'indice statistico cada nella zona di rifiuto, quando l'ipotesi nulla è vera.**

La **probabilità di commettere l'errore di II tipo**, indicato convenzionalmente con  $\beta$ , è la **probabilità di estrarre dalla popolazione un campione che non permette di rifiutare l'ipotesi nulla, quando in realtà essa è falsa.**

Da questi concetti derivano direttamente anche quelli di livello di **protezione** e di **potenza** di un test, che sono i parametri più importanti per scegliere il test più adatto alle caratteristiche dei dati e al quesito. Sono concetti tra loro legati, secondo lo schema riportato nella tabella seguente, che confronta la realtà con la conclusione del test

		REALTÀ'	
		H <sub>0</sub> vera	H <sub>0</sub> falsa
CONCLUSIONE  DEL  TEST	H <sub>0</sub> vera  (statisticamente non significativo)	Esatto  $P = 1 - \alpha$  <u>Protezione</u>	Errore II tipo  $P = \beta$
	H <sub>0</sub> falsa  (statisticamente significativo)	Errore I tipo  $P = \alpha$  <i>Significatività</i>	Esatto  $P = 1 - \beta$  <u>Potenza</u>

Un test statistico conduce ad una **conclusione esatta in due casi**:

- se non rifiuta l'ipotesi nulla, quando in realtà è vera;
- se rifiuta l'ipotesi nulla, quando in realtà è falsa.

Per aumentare la probabilità  $(1 - \alpha)$  del primo caso, occorre incrementare la **protezione**;

per quella  $(1 - \beta)$  del secondo caso, occorre aumentare la **potenza**.

**Esiste una sorta di concorrenza tra errori di primo tipo (a) ed errori di secondo tipo (b): se si abbassa il livello di significatività, cioè la probabilità di commettere errori di I tipo (a), si accresce quella dell'errore di II tipo (b); e viceversa.**

Si tratta di vedere quale dei due è più dannoso nella scelta che si deve effettuare. **L'unico modo per ridurli entrambi è quello di aumentare il numero dei dati.** Tuttavia non sempre è possibile aumentare le dimensioni del campione, perché già raccolto oppure perché i costi ed il tempo necessari diventano eccessivi, per le disponibilità reali del ricercatore.

Per l'**inferenza statistica** sono stati proposti due approcci, quello classico e quello decisionale.

La soluzione adottata nell'**approccio classico** all'inferenza statistica consiste:

- 1) nel fissare dapprima un livello di significatività  $\alpha$  conveniente basso, per contenere entro il limite prescelto la probabilità di commettere errori di I tipo;
- 2) successivamente nel scegliere la zona di rifiuto, in modo che  $\beta$  sia minimo.

Pertanto, nell'approccio classico si tenta di ridurre soprattutto l'errore di I tipo.

L'**approccio classico all'inferenza statistica**, così detto perché storicamente ha preceduto gli altri (l'approccio decisionale e quello bayesiano), è quello più noto ed applicato. Fa riferimento alla **concezione frequentista** della probabilità ed è rivolto alla pura conoscenza, alla esclusiva finalità scientifica di accettare o respingere un modello teorico. Non considera le iniziative che possono essere intraprese o le scelte da attuare, in seguito alle conclusioni raggiunte. È tipico della ricerca pura, come può essere quella biologica ed ecologica, quando evidenzia leggi o regolarità (come quelle di Mendel o la distribuzione geografica di una specie animale) per le quali non esiste alcun vantaggio o danno derivante da un'azione successiva alla scelta dell'ipotesi  $H_0$  oppure  $H_1$ .

**Nell'approccio classico, l'inferenza è fondata sul principio di ripetizione del campionamento,** per cui i dati sperimentali raccolti in natura o prodotti in laboratorio sono solamente uno degli infiniti possibili campioni, che teoricamente si ottengono ripetendo l'operazione infinite volte nelle stesse condizioni. L'inferenza ottenuta non considera solo il campione raccolto, ma tutti i possibili dati che teoricamente potrebbero essere ottenuti ripetendo l'esperimento.

Nell'**approccio decisionale**, si prendono in considerazione anche le conseguenze derivanti dagli errori e si cerca di valutare le "perdite", che sono determinate da eventuali decisioni sbagliate poiché il risultato del test è stato ritenuto corretto.

L'**approccio decisionale**, proposto per la prima volta in modo completo da Wald nel 1950, intende fornire metodi per decidere in situazioni d'incertezza: **il concetto di base è la perdita o il rischio che derivano da una decisione, se successivamente essa si rivelasse errata.**

L'approccio decisionale ha finalità operative: la conclusione non solo può essere corretta od errata, ma può avere conseguenze più o meno costose, se a posteriori si rivelasse sbagliata. Per l'ambientalista, è frequente la situazione in cui si devono decidere interventi, senza sapere in anticipo con precisione quali possono esserne le conseguenze (es.: nuove norme sulle discariche in un lago, che possono avere conseguenze negative per altri aspetti, come quelli economici sulle aziende vicine; la sospensione del traffico per abbassare i tassi d'inquinamento atmosferico, che può suscitare irritazione nella cittadinanza e quindi la perdita di consensi all'amministrazione). La diffusione della scienza applicata determina un'importanza crescente della teoria delle decisioni.

**La differenza tra approccio classico e approccio decisionale è più scolastica che reale.** Da una parte, sempre più spesso, la conoscenza scientifica è successivamente tradotta in applicazioni, che danno risultati economici (come i principi della selezione genetica applicati a piante da frutto o ad animali domestici); dall'altra, a volte le decisioni amministrative implicano rischi per la vita o la salute delle persone oppure danni permanenti all'ambiente, che non possono essere tradotti in valore economico (come un'eventuale dispersione territoriale di sostanze radioattive, che può incrementare la frequenza di tumori e i decessi).

**L'approccio decisionale ha applicazioni di estrema utilità quando le conseguenze delle scelte possono essere tradotte in conseguenze economiche,** con buona approssimazione. Ma la teoria delle decisioni esula dagli argomenti affrontati in questo corso.

Con una presentazione dei concetti precedenti più formale e meno discorsiva, utile all'apprendimento del linguaggio scientifico, si può affermare che il controllo statistico delle ipotesi ammette una pluralità di procedure, che differiscono sotto il profilo logico-metodologico. In questo corso, si fa riferimento a quelle che trovano le loro premesse

- nella **“teoria della significatività”** e
- nella **“teoria dei test”**.

Sono diverse tra loro, nella impostazione teorica e logica; ma, sotto particolari condizioni, convergono tecnicamente.

Nella **teoria della significatività**, dovuta prevalentemente a **R. A. Fisher**, il controllo, sulla base di una generico test, attiene ad una sola ipotesi detta “ipotesi nulla” (dall'inglese *null hypothesis*) indicata con  $H_0$ . Essa configura **la completa accidentalità dell'esito campionario: ipotizza che tale esito sia giustificabile nei termini dell'errore di campionamento da un insieme più vasto**, detta popolazione o universo. In questo contesto, si conviene di ritenere falsa (quindi di rifiutare) l'ipotesi  $H_0$  quando l'evento che si è verificato (ed eventi di esso più estremi) ha, sotto quella ipotesi, una probabilità di accadimento inferiore a un livello prefissato  $\alpha$  (detto livello di significatività). Non è certo che l'ipotesi nulla sia falsa; ma ad essa è associata una probabilità di errore non superiore ad  $\alpha$ .

**Nella teoria della significatività,  $H_0$  è rifiutata o non rifiutata; mai accettata.**

Nella **teoria dei test d'ipotesi**, dovuta a **J. Neyman e E. Pearson**, sono poste a confronto due ipotesi  $H_1$  e  $H_2$ , sulla base di una generica statistica test  $T$  e il rifiuto dell'una implica necessariamente l'accettazione dell'altra. Si conviene di rifiutare  $H_1$ , quindi accettare  $H_2$ , quando l'esito campionario (o esiti ad esso più estremi) risulta, dato  $H_1$ , meno verosimile che nella condizione in cui sia vera l'ipotesi  $H_2$ . Si può allora incorrere in 2 errori: rifiutare un'ipotesi quando è vera, (errore di I tipo) o accettarla

quando è falsa (errore di II tipo). Le loro rispettive probabilità di accadere (o meglio i massimi valori ammessi per esse) vengono indicate con  $a$  e  $b$ .

Intuitivamente, la regola migliore di decisione sulla sorte di una ipotesi dovrebbe consentire che contemporaneamente sia l'errore di I tipo sia l'errore di II tipo abbiano la minor probabilità possibile di avvenire. Tuttavia una tale regola non esiste, poiché le probabilità  $a$  e  $b$  sono legate da una relazione inversa: se una cresce, l'altra cala.

Pertanto,

- si fissa il valore di  $a$  e
- si cerca di individuare la procedura, o test, che da luogo al  $b$  con il valore minimo.

La quantità  $p = 1 - b$ , definita **potenza della regola di decisione**, misura la probabilità di rifiutare (correttamente) un'ipotesi quando è falsa.

Una delle due ipotesi poste a confronto può essere l'ipotesi di completa accidentalità ( $H_0$ ) e l'ipotesi alternativa  $H_A$  (più spesso  $H_1$ ), una negazione (unilaterale o bilaterale) di  $H_0$ . In questo schema di riferimento, **gli strumenti per il controllo statistico di  $H_0$  nell'ambito della teoria dei test coincidono, da un punto di vista tecnico, con quelli sviluppati nella teoria della significatività.**

#### 4.3. POTENZA DI UN TEST

**Il complemento di  $b$  ( $1-b$ ) misura la potenza** (da *power*; in alcuni testi di statistica in italiano è chiamata anche *forza*) **di un test statistico, definita come la probabilità di rifiutare l'ipotesi nulla quando è falsa; è la probabilità di non commettere l'errore di secondo tipo.**

I fattori che, con modalità ed intensità differente, incidono sulla potenza di un test per la verifica dell'ipotesi nulla sono 6:

- 1 - **il livello di significatività ( $\alpha$ );**
- 2 - **la dimensione della differenza ( $d$  oppure  $d$ ), di cui si vuole verificare la significatività;**
- 3 - **la variabilità ( $s^2$  oppure  $S^2$ ) dei dati (nel caso di confronti tra medie);**
- 4 - **la direzione dell'ipotesi (unilaterale oppure bilaterale);**
- 5 - **la dimensione ( $n$ ) del campione;**
- 6 - **le caratteristiche del test (a grandi linee, parametrico oppure non-parametrico).**

1) Il timore di commettere errori di I tipo tende a far abbassare al ricercatore il **livello di significatività  $\alpha$** . Ma, riducendo il valore di  $\alpha$ , egli diminuisce la probabilità di scoprire delle differenze, anche quando nella realtà esistono; in altri termini, egli aumenta la probabilità  $b$  di commettere errori di II tipo.

Il rischio  $\alpha$  implica che si arrivi alla conclusione che esiste una differenza significativa; permette quindi che si arrivi ad una decisione, che può rivelarsi errata e le cui conseguenze possono anche essere molto gravi. Se è molto importante prendere questa decisione senza correre troppi rischi di sbagliare, abitualmente si abbassa il livello di probabilità  $\alpha$ .

Due esempi possono chiarire meglio i due contrastanti interessi e i criteri per la scelta di  $\alpha$ , già in parte evidenziati nella presentazione delle caratteristiche dell'approccio decisionale.

ESEMPIO 1. Un'azienda pensa di essere in grado di produrre un prodotto nuovo per la cura dei tumori, che assicuri una più alta percentuale di guarigione.

Si supponga che questa azienda sia in ottime condizioni economiche, con una buona immagine, con prodotti di qualità e che il guadagno dato dalla vendita del nuovo prodotto sia percentualmente piccolo, rispetto al fatturato totale. Nel contempo, il rischio che l'azienda corre nell'immettere sul mercato il nuovo prodotto è alto. Può infatti accadere che, dopo un buon periodo di vendite, il farmaco non si manifesti realmente più efficace del precedente. La pubblicità negativa di questa notizia potrebbe rovinare l'immagine della società, su tutti i suoi prodotti. La conseguenza potrebbe essere una forte perdita economica.

In tale situazione, colui che deve prendere per la ditta la decisione se immettere o meno il prodotto sul mercato tenderà a **tenere molto basso il livello  $\alpha$**  di significatività sul test conclusivo, (per esempio,  $\alpha$  uguale a 0.0001 e non certamente 0.05) per ridurre al minimo il pericolo che il prodotto non sia realmente più efficace del precedente.

La conseguenza di questa scelta è **un aumento del rischio  $\beta$  ovvero il rischio di non mettere sul mercato un prodotto realmente più efficace.**

Un  $\alpha$  basso diminuisce il rischio per la società farmaceutica di mettere sul mercato un farmaco non effettivamente migliore del precedente, ma riduce anche la probabilità di trarne gli eventuali vantaggi.

ESEMPIO 2. Si supponga ora che il nuovo farmaco anti-tumorale sia prodotto e sperimentato da un'azienda in grave difficoltà economiche, la quale si trovi nelle condizioni di dover assolutamente aumentare le vendite, se vuole evitare il fallimento. Si supponga inoltre che, per la sua situazione, l'eventuale danno determinato dal fallimento del farmaco per l'azienda sia relativamente molto basso, essendo essa prossima al fallimento nella situazione attuale. Il responsabile della ditta tenderà a **tenere molto elevato il livello di  $\alpha$**  (per esempio,  $\alpha$  uguale a 0.10 o addirittura 0.20 e non certamente 0.05) al fine di aumentare sensibilmente nei test la probabilità di evidenziare un miglioramento significativo nel confronto con il farmaco precedente. Si avrà così un rischio molto elevato di mettere sul mercato un nuovo prodotto che non sia realmente più efficace (errore di I tipo) e dare ulteriore dimostrazione d'incapacità; ma contemporaneamente si avrà una nuova opportunità per sé e un vantaggio maggiore per gli ammalati, che potranno provare ed eventualmente scegliere un nuovo prodotto che potrebbe

dimostrarsi realmente migliore (diminuisce la probabilità di commettere un errore di II tipo). E' una **diminuzione del rischio b di accettare l'ipotesi nulla, quando in realtà il farmaco è più efficace.** (I due esempi sono solo ipotetici, ai fini didattici e con lo scopo di chiarire i concetti. Per i farmaci in molti paesi esistono commissioni, che hanno lo scopo di valutare scientificamente i risultati prima dell'immissione sul mercato).

2) Un secondo fattore che incide sulla potenza di un test è **la dimensione della differenza** tra il valore osservato e il valore atteso nell'ipotesi nulla (di solito, la media). **La potenza di un test statistico è funzione crescente della differenza, considerata in valore assoluto.** E' intuitivo che sia più facile rilevare differenze grandi di quelle piccole. Quando la differenza è grande, la potenza del test è maggiore rispetto al caso in cui la differenza è piccola, perché maggiore è la probabilità di rifiutare con il test l'ipotesi nulla, che nella realtà essa è falsa (appunto perché esiste differenza). Mediante la distribuzione normale, una esemplificazione chiara è fornita dal test per verificare se esiste una differenza significativa tra la media campionaria ( $\bar{x}$ ) e quella della popolazione ( $\mu$ ), quando sia nota la deviazione standard ( $\sigma$ ) della popolazione ed ovviamente la dimensione del campione ( $n$ ). Essa infatti dipende dalla formula

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Da essa si può agevolmente dedurre che il valore di  $z$  sarà tanto più grande e quindi il test risulterà maggiormente significativo quanto più grande è la differenza tra la media campionaria ( $\bar{x}$ ) e quella dell'ipotesi ( $\mu$ ).

Per avere un test significativo, è necessario che la differenza tra media campionaria e media ipotizzata ( $\bar{x} - \mu$ ) sia maggiore od uguale al prodotto del valore di  $z$ , alla probabilità  $\alpha$  prefissata, per l'errore standard della media

$$\bar{x} - \mu \geq z \cdot \frac{s}{\sqrt{n}}$$

dove

$z$  è il valore della distribuzione normale per la probabilità  $\alpha$  prefissata  
(per  $\alpha$  uguale al 5%,  $z$  è uguale a 1,645)

$\frac{s}{\sqrt{n}}$  è la misura della dispersione della media, calcolata su  $n$  dati.

La differenza tra le medie di 2 campioni  $\bar{x}_1 - \bar{x}_2$  può essere maggiore o minore della differenza vera  $\mu_1 - \mu_2$ . E' quindi preferibile fondare le richieste sul valore reale della differenza tra  $\mu_1 - \mu_2$ , che di norma è indicato con  $\delta$  (delta). Con dati campionari, non si può mai essere sicuri di trovare una differenza significativa: le fluttuazioni campionarie della differenza tra due medie ( $\bar{x}_1 - \bar{x}_2$ ) possono essere molto più piccole della differenza reale ( $\mu_1 - \mu_2$ ) e quindi non essere significativamente diverse da zero.

Sovente è utile **scegliere il valore di d in rapporto alla differenza minima ritenuta importante** per il fattore in esame.

3) Un terzo fattore è **la variabilità** dei dati. **La potenza di un test è funzione decrescente della varianza**. Le formule riportate per l'analisi della differenza richiedono che si conosca la deviazione standard  $S$  della popolazione. Nella pratica, è raro conoscere  $S$  prima dell'esperimento, anche se può essere stimata da ricerche precedenti, eventualmente reperibili in letteratura o nell'esperienza del ricercatore. Si ricorre quindi alla deviazione standard del campione ( $s$ ), con la quale occorre utilizzare non la distribuzione normale standardizzata  $z$ , ma la distribuzione  $t$  di Student, che sarà presentata nel capitolo successivo.

Riprendendo la trasformazione della formula precedente con la medesima simbologia,

$$z = \frac{d}{\frac{s}{\sqrt{n}}}$$

è evidente che

- **al diminuire di  $s$ , aumenta il valore di  $z$  e quindi la potenza del test nell'evidenziare un effetto di una determinata grandezza assoluta  $d$ ;**
- viceversa, all'aumentare di  $s$  diminuisce  $z$  e la potenza del test nel rendere significativa la stessa differenza.

Molto spesso, **l'influenza della differenza  $d$  e della deviazione standard  $s$** , derivati dai medesimi dati, **vengono considerate assieme** mediante il rapporto

$$f = d / s$$

Questo indice  $f$  (phi), appunto perché determinato dal rapporto di valori stimati dagli stessi dati, ha il vantaggio di essere **adimensionale e quindi con valori relativamente stabili**, che di norma variano da **0,5 a 1**. Ne consegue che, nel calcolo della funzione di potenza di un test,  $f$  offre il vantaggio pratico di ridurre il numero dei valori da considerare, nettamente inferiore a quello delle innumerevoli combinazioni possibili tra valori diversi della differenza  $d$  e della deviazione standard  $s$ .

4) **L'ipotesi alternativa  $H_1$ , da verificare con un test, può essere bilaterale o unilaterale.**

E' **bilaterale**, quando ci si chiede se tra la media del gruppo **A** e quella del gruppo **B**, a confronto, esiste una differenza significativa, senza sapere a priori quale dei due gruppi potrà dare il risultato migliore.

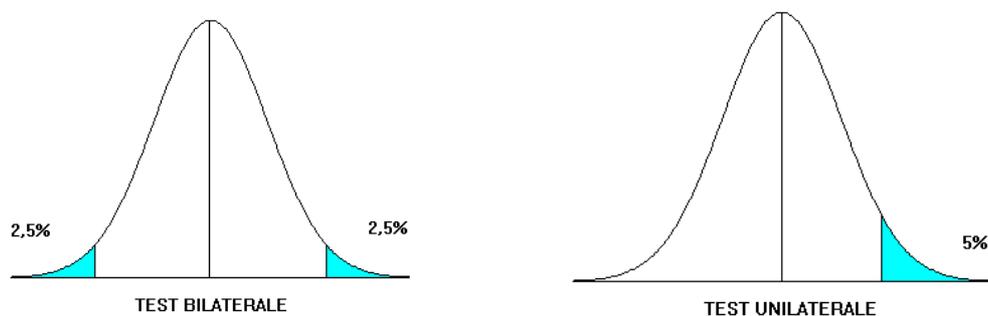
E' **unilaterale** quando è possibile escludere a priori, come privo di significato e risultato solo di errori nella conduzione dell'esperimento, il fatto che la media di **B** possa essere minore o maggiore della **A**. Il test statistico serve per verificare solamente se la media di **B** sia significativamente superiore a quella di **A** o viceversa; non entrambi.

Per esempio, in tossicologia

- si ha un **test bilaterale** quando si confronta l'effetto di due tossici **A** e **B** sull'accrescimento di due gruppi di animali, per valutare quale abbia l'effetto maggiore: sono due risposte alternative, che lo sperimentatore ritiene ugualmente logiche e possibili;

- si ha invece un **test unilaterale** quando si confrontano i risultati di un tossico rispetto al placebo. E' evidente che da questo secondo confronto non ci si può ragionevolmente aspettare che gli animali ai quali è stato somministrato il tossico abbiano risultati migliori nella crescita e nella sopravvivenza di coloro ai quali è stato somministrato il placebo. L'unica domanda razionale è se gli animali ai quali è stato somministrato il tossico abbiano un accrescimento significativamente peggiore di quelli trattati con il placebo. **Se il gruppo al quale è stato somministrato il tossico avesse un risultato medio migliore dell'altro, sarebbe illogico e quindi sarebbe inutile proseguire l'analisi, con qualunque test statistico.**

**I concetti su test bilaterale e test unilaterale spesso sono espressi sinteticamente con un grafico.**



**La differenza tra test unilaterale e test bilaterale non è solamente una questione teorica, ma una scelta con effetti pratici rilevanti sulla potenza del test, poiché è importante per la determinazione della zona di rifiuto dell'ipotesi nulla. In un test unilaterale essa sarà solamente in**

una coda della distribuzione; in un test bilaterale essa sarà equamente divisa nelle due code della distribuzione.

In una distribuzione normale, prendendo come livello di significatività il 5%,

- in un test ad una coda l'area di rifiuto dell'ipotesi nulla inizia dal valore critico di 1,645
- in un test a due code essa inizia dal valore critico di 1,96.

**Allo stesso livello di significatività ( $\alpha$ ), con la stessa deviazione standard ( $\sigma$ ), la medesima differenza in valore assoluto ( $d$ ) ed un uguale numero di dati ( $n$ ), un test unilaterale è sempre più potente del corrispondente test bilaterale, poiché il valore dell'indice al quale si rifiuta l'ipotesi nulla è sistematicamente minore, in valore assoluto.**

Un test unilaterale è quindi senza dubbio preferibile; ma per esso occorre una quantità d'informazione superiore sui possibili risultati dell'esperimento, che non sempre è disponibile.

5) **La dimensione del campione ( $n$ ) è il parametro che ha l'effetto più facilmente misurabile sulla potenza di un test.**

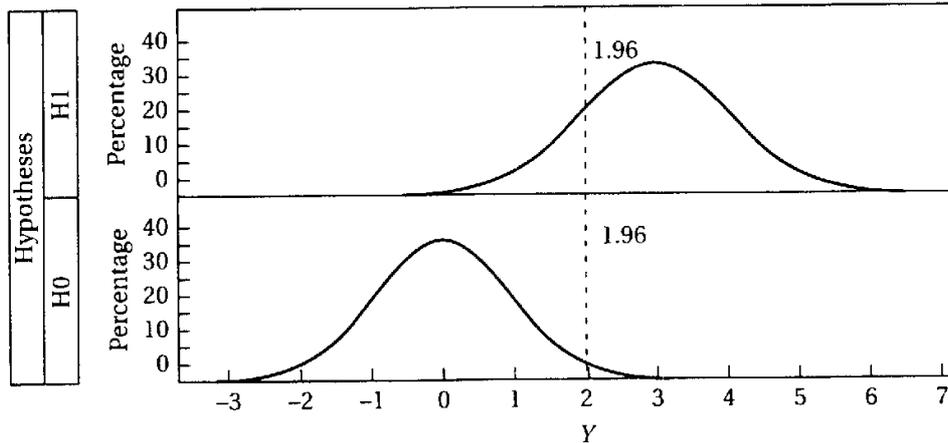
Applicando l'equazione

$$\bar{x} = m - z \cdot \frac{s}{\sqrt{n}}$$

con l'abituale simbologia della distribuzione normale,

è semplice calcolare che la significatività di una differenza media campionaria ( $\bar{x}$ ) è inversamente proporzionale alla radice quadrata del numero di dati del campione ( $\sqrt{n}$ ).

ESEMPIO I concetti, espressi nelle pagine precedenti, sui fattori che determinano la potenza ( $1-\beta$ ) di un test possono essere spiegati con un esempio e la sua illustrazione grafica, come riportata nella figura sottostante.



Per una comprensione più semplice, è utile scomporre i diversi passaggi logici in due fasi.

I - Dapprima si supponga che l'effetto di un fattore sia nullo (ipotesi  $H_0$  vera, come nella distribuzione normale inferiore), quindi che la sua media sia  $\mu = 0$ . Supponendo inoltre che

- la sua deviazione standard sia  $s = 2,8$
- da questa popolazione sia stato estratto un campione di dimensione  $n = 6$
- l'ipotesi alternativa  $H_1$  sia bilaterale,
- con  $\alpha = 0.05$

non si rifiuterà l'ipotesi nulla fino a quando la media campionaria  $\bar{x}$  sarà compresa nell'intervallo fiduciale

$$\bar{x} = \mu \pm z_{\alpha} \frac{s}{\sqrt{n}}$$

Con i dati dell'esempio, risulta

$$0 \pm 1,96 \cdot \frac{2,8}{\sqrt{6}} = \pm 2,24$$

che  $\bar{x}$  può variare tra  $-2,24$  e  $+2,24$  senza essere statisticamente diversa da 0.

II - Ora, si deve passare all'altra distribuzione di probabilità, la probabilità  $\beta$ , ipotizzando che  $H_0$  sia falsa (come nella distribuzione normale superiore).

Se l'effetto reale del trattamento ( $H_1$  vera) in realtà è  $\mu = 3$ , **non si potrà rifiutare l'ipotesi nulla se la media  $\bar{x}$  di un campione estratto da questa popolazione sarà più vicina a 0 di  $+2,24$ . Per la stima di  $\mu$ , in questa fase del calcolo il test è unilaterale.**

Supponendo sempre, come nel caso precedente, che

- la sua deviazione standard sia  $s = 2,8$
- da questa popolazione si estraiga un campione di dimensione  $n = 6$ ,

la probabilità  $b$  di estrarre una media campionaria  $\bar{x}$  minore di **+2,24** sarà data da

$$z_{\beta} = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}}$$

Con i dati dell'esempio

$$z_{\beta} = \frac{2,24 - 3}{\frac{2,8}{\sqrt{6}}} = -0,67$$

si ottiene un valore di  $z_{\beta}$  uguale a  $-0,67$ .

**Questa stima ha significato solo se la media reale di  $H_1$  è dalla stessa parte del campione, rispetto alla media reale dell'ipotesi  $H_0$ . Di conseguenza, la probabilità di  $b$  deve essere stimata su una distribuzione normale unilaterale.**

In una tavola normale unilaterale, al valore di  $z_{\beta} = 0,67$  corrisponde una probabilità  $b = 0,251$ . Di conseguenza, la potenza  $(1-b)$  di questo test (cioè la probabilità di rifiutare l'ipotesi nulla, che noi sappiamo in realtà essere falsa) è uguale a  $(1 - 0,251)$  cioè  $0,749$  o  $74,9\%$ .

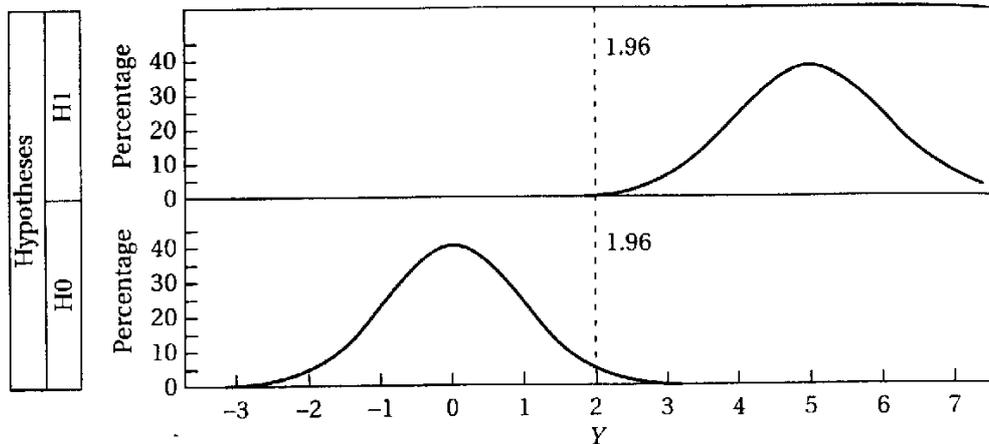
E' semplice dimostrare che

- scegliendo una probabilità  $\alpha$  minore,
- diminuendo  $s$ ,
- aumentando  $n$
- incrementando  $d$ , la distanza tra le medie reali  $m$  delle due ipotesi,

diminuisce anche la probabilità  $b$  e quindi **aumenta la potenza  $(1-b)$  del test**, cioè la probabilità di rifiutare l'ipotesi nulla, con i dati del campione che sarà raccolto.

Se a parità di tutti gli altri fattori considerati, come nella figura successiva, la  $m$  reale dell'ipotesi nulla fosse stata uguale a  $5$ ,

## Hypothesis testing



il valore di  $z_b$  sarebbe risultato

$$z_b = \frac{2,24 - 5}{\frac{2,8}{\sqrt{6}}} = -2,40$$

uguale a 2,40 e quindi si sarebbe ottenuto una probabilità  $b = 0.008$  e una potenza  $(1-b)$  pari al 99,2%.

Questa stima di  $b$  è chiamata **potenza a posteriori**. Di norma, quando il test non risulta significativo, ricorrendo alla distribuzione  $t$  di Student serve per valutare quale poteva essere la probabilità di rifiutare l'ipotesi nulla, con le caratteristiche  $(\bar{x}, s)$  e la quantità  $(n)$  dei dati raccolti.

Spesso è utile stimare **la potenza a priori**.

Al momento della programmazione dell'esperimento, un problema pratico che si pone al ricercatore è quanti dati debba raccogliere, al fine di dimostrare con il test prescelto che una certa differenza è significativa. La **dimensione minima di un campione** può essere determinata dopo aver fissato

- uno specifico livello di significatività (da cui dipende  $z$ ),
- un errore campionario (che dipende dalla variabilità del fenomeno, misurata con  $s$ ).

La dimensione del campione  $n$  può essere determinata per un prefissato livello di significatività  $(\alpha)$  e di potenza del test  $(1-\beta)$  mediante la distribuzione normale standardizzata  $z$  con la formula

$$n = \frac{s^2 \cdot (Z_a - Z_b)^2}{(m_0 - m_1)^2}$$

dove

$\sigma^2$  è la varianza della popolazione,

$z_a$  è il valore di  $z$  per il livello di significatività  $\alpha$ ,

$z_b$  è il valore di  $z$  per il rischio  $\beta$  dell'errore di II tipo,

$m_0$  è la media della popolazione nell'ipotesi nulla  $H_0$ ,

$m_1$  è la media della popolazione nell'ipotesi alternativa  $H_1$ .

Per esempio, in un **test unilaterale**,

- per un livello di significatività del 5% e
- una potenza dell'80% ( $\beta = 20$ )

il valore di  $z_a$  è 1,645 e quello di  $z_b$  è 0,84.

Per un **test bilaterale**,

con gli stessi dati viene modificato solamente il valore di  $z_a$  che alla significatività del 5% risulta uguale a 1,96.

Il numero  $n$  di dati necessari risulta quindi

- direttamente proporzionale al valore della varianza della popolazione ed
- inversamente proporzionale alla differenza tra le medie della popolazione, espresse nell'ipotesi nulla  $H_0$  e nell'ipotesi alternativa  $H_1$

(Le modalità per stimare le dimensioni utili del campione secondo i vari test saranno spiegati nei capitoli successivi).

**L'uso principale del calcolo della potenza è la stima del numero di dati necessari a rilevare un effetto abbastanza grande, da essere rilevante od importante nello studio dei fattori in esame.**

Quando non si è stati in grado di rifiutare l'ipotesi nulla, a posteriori può essere di grande utilità calcolare la potenza del test attuato, al fine di evidenziare e discutere l'eventuale contributo fornito,

- sia dalla differenza analizzata
- sia dalle dimensioni del campione raccolto.

Il calcolo della potenza serve per stimare le dimensioni del campione di una successiva ricerca, affinché la differenza media rilevata possa risultare significativa. Sovente, piuttosto che concentrare l'attenzione sulla alternativa accettazione-rifiuto dell'ipotesi nulla è utile valutare la **potenza** con la quale i dati suggeriscono la presenza di un effetto.

In conclusione, per stimare con precisione quanti dati si devono raccogliere affinché un test risulti significativo, devono essere stimati in anticipo

- l'effetto che si pensa di osservare o che valga la pena rilevare (d),
- il grado di fiducia con il quale si spera di accettare (b) e
- di rifiutare (a) l'ipotesi nulla di inefficacia,
- la deviazione standard (S) della popolazione studiata.

I dati possono essere tratti da risultati già pubblicati, da studi-pilota condotti dal ricercatore od essere fondati su supposizioni più o meno logiche.

6) A partire dagli stessi dati, **non tutti i test hanno la stessa capacità di rifiutare l'ipotesi nulla quando è falsa.**

E' quindi **molto importante scegliere il test più adatto,**

- in rapporto alle **caratteristiche dei dati** (qualitativi o quantitativi),
- al **tipo di scala** o misura (scale di rango, di intervalli o di rapporti),
- alla **variabilità** dei dati,
- alla **simmetria** della distribuzione,
- alla **omoschedasticità** dei gruppi a confronto.

**Test diversi hanno condizioni di validità differenti e sono più o meno robusti** (sopportano in modo differente l'allontanamento dalle condizioni di validità).

E' già stato fatto osservare che, con un numero ridotto di osservazioni suddivise in vari gruppi, il test di Kolmogorov-Smirnov è più potente del test  $\chi^2$ .

Nei test di statistica non parametrica che verranno presentati, si farà sovente un confronto di potenza rispetto agli altri test, soprattutto a quelli corrispondenti di statistica parametrica. E' infatti molto importante **utilizzare il test più potente, in funzione del tipo di scala ed in accordo con le caratteristiche dei dati.** Per esempio, nel confronto tra le tendenze centrali in due campioni dipendenti possono essere utilizzati test differenti: tra quelli non parametrici è possibile scegliere tra il test dei segni, il test T di Wilcoxon, il test di casualizzazione; tra quelli parametrici è possibile scegliere tra l'analisi della varianza a due criteri di classificazione oppure ad un solo criterio.

Errare nella scelta del test, significa non scegliere il più potente per quelle condizioni specifiche; quindi il risultato può essere quello di non rifiutare l'ipotesi nulla (che sappiamo falsa), rendendo nullo il lavoro di raccolta ed analisi dei dati.

Nella scelta del test più appropriato al problema, si deve sempre considerare anche il confronto tra le potenze.

**Di norma, quanto più scarsi o deboli sono i postulati su cui il test è fondato, tanto più le conclusioni hanno un valore generale. E' meno probabile ottenere risultati significativi; ma**

**l'eventuale affermazione di significatività molto difficilmente può essere contestata.** Per questo motivo, con i test di statistica non parametrica solitamente è più difficile rifiutare l'ipotesi nulla, quando essa è falsa; ma questi test hanno il vantaggio di avere condizioni di validità meno restrittive e quindi di sollevare meno obiezioni o dubbi sulle conclusioni raggiunte.

Questi confronti tra test sono validi, quando si utilizzano campioni con lo stesso numero di osservazioni. Infatti, il numero di dati è un parametro che incide direttamente sulla potenza-efficienza di un test.

Il concetto di **potenza-efficienza** del test A rispetto al test B è fondato sul numero di osservazioni necessario al test A per avere la stessa potenza del test B.

$$\text{potenza - efficienza del test A [in \%]} = \frac{N_b}{N_a} \cdot 100$$

dove

$N_a$  e  $N_b$  sono rispettivamente il numero di dati od osservazioni utilizzati nei due test A e B.

Per esempio,

se il test A richiede 30 osservazioni per avere la stessa potenza del test B con 20 osservazioni,

la potenza di A sarà  $20/30 \times 100$  di B e corrisponde al 66%.

Significa che ogni 6,6 osservazioni per il test B occorrono 10 osservazioni per A, se si vuole la stessa potenza.

A volte, è **preferibile avere un test con condizioni di validità meno restrittive (e quindi meno potente), se è possibile aumentare il numero di osservazioni o rifiutare comunque l'ipotesi nulla alla probabilità prefissata, perché le conclusioni saranno più generali e quindi più difficilmente criticabili.**

Secondo le caratteristiche dei dati, in particolare in rapporto alla loro variabilità, quando si analizzano più fattori diventa importante scegliere il disegno sperimentale più adatto, quello che rende massima l'efficienza dell'analisi. Si parla allora di **efficienza relativa** (il concetto verrà ripreso nell'analisi della varianza).

Sperimentazioni progettate e condotte in modo corretto ed analizzate con metodi appropriati possono non evidenziare differenze reali e quantitativamente importanti nella disciplina studiata, a causa di un campione troppo piccolo, non di grado in fornire una potenza sufficiente per rendere l'effetto statisticamente significativo. **L'analisi della potenza permette di valutare in modo critico i risultati** per ripetere l'esperimento con un numero di dati adeguato.

Altre volte, l'interpretazione della stima della potenza può evidenziare la necessità di un campione troppo grande, per essere attuato nelle condizioni reali del ricercatore. La causa principale della non significatività sarebbe allora da ricercare nell'effetto troppo piccolo che si vuole analizzare o nella grande variabilità dei dati.

Comunque se la potenza del test evidenzia la necessità di un campione che superi le possibilità reali del ricercatore, per i tempi richiesti nella raccolta dei dati oppure per i costi dell'esperimento o per l'impossibilità oggettiva di disporre di tanti casi, si raggiungono conclusioni ugualmente importanti per la ricerca.

La situazione più diffusa è quella in cui il numero di dati può variare con relativa facilità, non rappresentando il vincolo fondamentale della ricerca. Negli esperimenti in laboratorio e nella raccolta dei dati in natura, generalmente l'ambientalista dovrebbe fissare il numero di osservazioni prima di iniziare l'esecuzione.

In altri settori della ricerca, ogni singolo dato è molto costoso sotto l'aspetto economico o morale; in altri ancora, si tratta di eventi che avvengono molto raramente, a grandi distanze di tempo. La dimensione della ricerca non è prevista all'inizio come nei casi trattati in precedenza: si cerca decidere in modo definitivo, appena i dati permettono di giungere ad una conclusione. E' **l'analisi sequenziale**, esposta in altro capitolo.

Nell'analisi sequenziale, i dati raccolti sono analizzati ogni volta che si aggiunge un nuovo dato. L'applicazione del test permette 3 risposte:

- accettare definitivamente l'ipotesi nulla;
- rifiutarla definitivamente;
- aggiungere o aspettare un nuovo dato, in quanto con i dati già raccolti non è possibile decidere in nessuna delle due precedenti direzioni.

La procedura è relativamente più sofisticata dei test che verranno proposti nel corso.

Per le analisi sequenziali sono disponibili grafici sui quali riportare i risultati già ottenuti. Alcuni ricercatori li utilizzano al posto dei test classici, per evitare i calcoli richiesti. E' un errore, in quanto i test classici permettono l'utilizzazione migliore dei dati raccolti.

E' pure errato ripetere i test classici, tutte le volte che si aggiunge un dato nuovo ai precedenti: tale procedura determina valori di probabilità  $P$  troppo ottimistici. Sono concetti che verranno approfonditi quando si tratteranno i confronti multipli nell'analisi della varianza e nei test non parametrici per  $k$  campioni.

**4.4. NUMERO DI DATI NECESSARI IN RAPPORTO ALLA POTENZA, ALLA SIGNIFICATIVITA' DEL TEST E ALLA DIREZIONALITA' DELL'IPOTESI. IL CRITERIO DI COHEN PER LA SCELTA DI  $\alpha$  E  $\beta$**

Riprendendo i concetti già presentati nei paragrafi precedenti, è utile ricordare che:

- $\alpha$  indica la probabilità di rifiutare (erroneamente) l'ipotesi nulla  $H_0$  quando essa è vera; è la significatività del test e il risultato è chiamato **falso positivo**, poiché con esso si accetta come significativa una differenza che in realtà non esiste:
- $1 - \alpha$  è la probabilità di accettare l'ipotesi nulla quando è vera (risultato esatto),
- $\beta$  indica la probabilità di non trovare (erroneamente) una differenza significativa quando in realtà essa esiste e il risultato è chiamato anche **falso negativo**, poiché si nega l'esistenza di una differenza che in realtà esiste)
- $1 - \beta$  è la probabilità di trovare significativa una differenza che in realtà esiste (risultato esatto); è la **potenza del test**.

Il valore di  $\alpha$  in genere è fissato al momento dell'analisi. La sua scelta dipende essenzialmente dalle conseguenze negative di una conclusione errata, quando si dichiara significativa una differenza che in realtà non esiste. **L'errore di II Tipo, misurato da  $\beta$ , dipende soprattutto dalla programmazione dell'esperimento**. Ad esempio, come si vedrà nell'analisi della varianza, quando la differenza  $d$  è piccola rispetto alla variabilità  $s$  dei dati, un'analisi ad un criterio di classificazione non permette di rifiutare l'ipotesi nulla, mentre un campionamento a strati, che permette di isolare i fattori di variabilità, rende significativa anche tale differenza  $d$ , attraverso la riduzione della varianza d'errore. E' quindi importante avere un'idea delle dimensioni di  $d$ , in rapporto alla varianza e al numero dei dati raccolti.

Una differenza reale  $d$  o della popolazione risulta significativa alla probabilità  $\alpha$ , quando essa

$$d > z_{\alpha} \frac{s}{\sqrt{n}}$$

- è maggiore del valore di  $z_{\alpha}$
- moltiplicato per l'errore standard  $\frac{s}{\sqrt{n}}$ .

Ma, se non si considera anche  $\beta$ , si ha solo una probabilità del 50% che questo avvenga: infatti, per  $\beta = 0,5$  in una coda della distribuzione normale, il valore di  $z$  è uguale a 0.

Per ottenere un test con potenza maggiore di  $1-\beta$ , il valore della differenza  $d$  essere maggiore di

$$d > z_{\beta} \frac{s}{\sqrt{n}}$$

del prodotto di  $z_b$  per l'errore standard  $\frac{S}{\sqrt{n}}$

Da questi due concetti, **per rispettare entrambe le condizioni** si deriva la relazione

$$z_\alpha \frac{S}{\sqrt{n}} < d - z_\beta \frac{S}{\sqrt{n}}$$

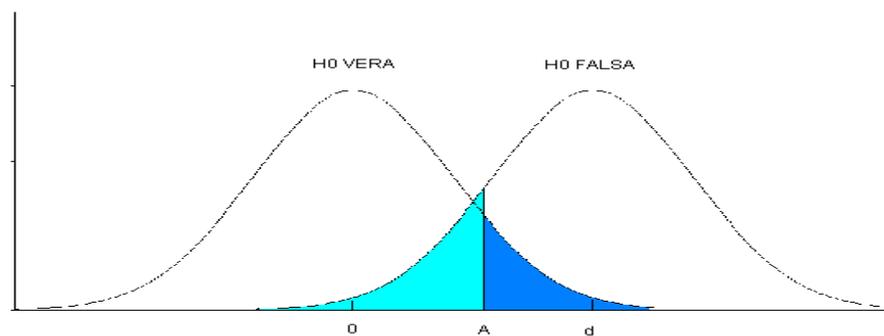
per cui il valore della differenza  $d$  deve essere maggiore di

$$d > (z_\alpha + z_\beta) \frac{S}{\sqrt{n}}$$

Da essa si stima la dimensione minima  $n$  del campione, che deve essere maggiore di

$$n > \left[ \frac{(z_\alpha + z_\beta) \cdot S}{d} \right]^2$$

Le relazioni tra  $a$  e  $b$  sono illustrati nella figura sottostante:



Rappresentazione grafica della distribuzione delle probabilità di  $a$  (a sinistra) e di  $b$  (a destra).

Se il confronto non avviene tra una media campionaria e una media teorica o attesa, ma tra le medie di due campioni indipendenti che abbiano lo stesso numero d'osservazioni (ed ovviamente la stessa varianza  $\sigma^2$ ),

l'errore standard  $es$  della differenza  $d$  diviene

$$es = \sqrt{S^2 (2/n)}$$

In questi test, il numero totale  $n$  di osservazioni, complessivamente per i due campioni, deve essere

$$n > 2 \left[ \frac{(z_a + z_b) \cdot S}{d} \right]^2$$

Con questa ultima formula o con quella per un campione,

è possibile evidenziare le relazioni che esistono tra

- la dimensione  $n$  del campione,
- la significatività  $\alpha$  del test,
- la potenza  $1-\beta$  del test,
- il valore dell'effetto  $d$  che si vuole dimostrare significativo.

Per effettuare gli esercizi, è utile ricordare che

**il valore di  $z_b$  è fornito dalla stessa tabella dei valori critici della distribuzione  $z$  usata per  $\alpha$ .**

Per valutare la **potenza** ( $1-\beta$ ), occorre stimare  $\beta$ .

Nel caso di **due campioni indipendenti** con lo stesso numero  $n$  di osservazioni,

$$z_\beta = \sqrt{\frac{n \cdot d^2}{2 \cdot S^2}} - z_\alpha$$

e per **un solo campione** è

$$z_\beta = \sqrt{\frac{n \cdot d^2}{S^2}} - z_\alpha$$

Nel caso di **due campioni indipendenti con un numero diverso di osservazioni (campioni non bilanciati)**, si ha una perdita di potenza del test. L'argomento verrà trattato in modo più approfondito alla fine del capitolo V sul test  $t$  di Student.

**Secondo alcuni autori, la potenza di un test dovrebbe essere sempre calcolata, in particolare quando non è possibile rifiutare l'ipotesi nulla a causa di un campione troppo piccolo.**

Se la potenza risulta bassa, la conclusione dovrebbe evidenziare che la differenza  $d$  facilmente esiste ugualmente, nonostante il risultato del test; ma che la sua dimostrazione richiede un campione di dimensioni maggiori.

Se la potenza è alta, si può concludere che la differenza  $d$ , se esiste, è comunque di entità trascurabile.

Nello stesso modo, è possibile stimare la **differenza minima  $d$**  che risulta significativa con il numero  $n$  di dati raccolti. Conoscendo la deviazione standard  $s$  della popolazione ed avendo prefissato le probabilità relative ad  $a$  e a  $b$  per **2 campioni bilanciati** è

$$d \geq \sqrt{\frac{(z_a + z_b)^2 \cdot 2s^2}{n}}$$

e **per un solo campione** è

$$d \geq \sqrt{\frac{(z_a + z_b)^2 \cdot s^2}{n}}$$

ESEMPIO 1. Si intende verificare se tra due quartieri della stessa città esiste una differenza significativa nei livelli d'inquinamento superiore a  $d = 40$ .

Con  $s = 60$ , rilevato da studi precedenti, quanti dati occorrerà raccogliere per ogni gruppo, se la probabilità  $\alpha$  per un test bilaterale deve essere uguale a 0.05 e

1 - la potenza del test pari al 90%,

2 - la potenza del test pari al 99%.

Risposta.

Per un test bilaterale si ricorre alla formula

$$n > 2 \left[ \frac{(z_a + z_b) \cdot s}{d} \right]^2$$

in cui,

per  $p = 0.025$  in una coda, il valore di  $z_\alpha$  è uguale a 1,96

per una potenza del 90% (e quindi  $\beta = 0.10$ ) il valore di  $z_\beta$  è uguale a 1,28

$\sigma = 60$  e  $\delta = 40$ .

Da essi

$$n > 2 \left[ \frac{(1,96 + 1,28) \cdot 60}{40} \right]^2 = 2 \left( \frac{3,24 \cdot 60}{40} \right)^2 = 2 \cdot 4,86^2 = 2 \cdot 23,6 = 47,2$$

si ottiene  $n > 47,2$ .

Sono necessarie in totale almeno 48 osservazioni, 24 in ognuno dei 2 campioni.

**b indica la probabilità di non trovare significativa, alla probabilità  $\alpha$ , la differenza reale  $d$  con  $s$  noto estraendo un campione di dimensioni  $n$ .**

Se i calcoli fossero fatti senza considerare  $b$ ,

e quindi

$$n > 2 \left( \frac{1,96 \cdot 60}{40} \right)^2 = 2 \cdot 8,64 = 17,28$$

si sarebbe ottenuto un totale di 18 dati, 9 per gruppo.

Con essi, **l'esperimento avrebbe solo il 50% di probabilità di rifiutare l'ipotesi nulla**, ovviamente a parità di tutte le altre condizioni.

Per diminuire questa probabilità  $\beta$  a 1% (quindi per avere una potenza  $1-\beta$  pari al 99%), al posto di 1,28 dell'esercizio precedente è necessario mettere 2,328.

Dalla relazione

$$n > 2 \left[ \frac{(1,96 + 2,328) \cdot 60}{40} \right]^2 = 2 \left( \frac{4,288 \cdot 60}{40} \right)^2 = 2 \cdot 6,43^2 = 2 \cdot 41,34 = 82,68$$

si stima un campione di almeno 84 osservazioni in totale e quindi composto da 42 osservazioni per ognuno dei 2 gruppi.

**ESEMPIO 2.** L'esempio 1 propone un test bilaterale. Se il test fosse stato unilaterale (dimostrare che nel quartiere A l'inquinamento è minore di quello presente nel quartiere B di una quantità  $d = 40$ ), quanti dati servono per una potenza pari al 99%?

Risposta

Con i dati dell'esempio 1, ultima stima,

al posto di 1,96 (riferito alla probabilità  $\alpha = 0,025$  in una coda della distribuzione)

si deve mettere 1,645 (riferito alla probabilità  $\alpha = 0,05$  in una coda della distribuzione).

Si ottiene quindi un valore di  $n$

$$n > 2 \left[ \frac{(1,645 + 2,328) \cdot 60}{40} \right]^2 = 2 \left( \frac{3,973 \cdot 60}{40} \right)^2 = 2 \cdot 5,96^2 = 2 \cdot 35,52 = 71,02$$

uguale a 71,02: servono 2 campioni con 36 dati ognuno.

Trattandosi di un test unilaterale al posto di uno bilaterale, a parità di tutte le altre condizioni il numero complessivo di dati da raccogliere scende da 42 a 36, per ognuno dei due gruppi a confronto.

Quando **la varianza della popolazione è ignota**, si può effettuare un'indagine "pilota" mediante un campione di numerosità ridotta ed utilizzare la deviazione standard  $s$  come prima approssimazione di quella della popolazione ( $\sigma$ ). In questo caso si dovrà utilizzare la distribuzione  $t$  di Student. Tuttavia, per un primo approccio al problema, sono stati proposti metodi approssimati, che usano la distribuzione normale, quando **non è nota la varianza ma si ha una stima del campo di variazione**. Per esempio, eccetto pochi esperti, nessuno conosce la varianza dell'altezza in ragazze ventenni europee; ma quasi tutti sanno che essa potrà variare approssimativamente tra 155 e 185 centimetri.

Disponendo di una semplice **indicazione del campo di variazione** ( $x_{\max}$  e  $x_{\min}$ ) del fenomeno in oggetto, dalla quantità ( $x_{\max} - x_{\min}$ ) è possibile dedurre un valore massimo della varianza, attraverso la disuguaglianza

$$\sigma^2 \leq (x_{\max} - x_{\min})^2 / 4$$

Più in generale, anche quando la stima di  $\sigma$  è ottenuta mediante un campione "pilota", è sufficientemente corretto utilizzare la distribuzione normale  $z$  al posto della distribuzione  $t$  di Student, se il campione stimato è di alcune decine d'unità. Con oltre 30 osservazioni, le differenze tra il valore di  $z$  e quello di  $t$  sono praticamente trascurabili, per le stessa probabilità di  $a$  e  $b$ .

**La scelta di  $a$  e di  $b$  sono soggettive**. Tuttavia per tale scelta sono state fornite indicazioni operative. Per il valore da attribuire ad  $a$ , il criterio è il costo del rifiuto dell'ipotesi nulla quando essa è vera. Per  $b$ , la probabilità di non trovare una differenza che effettivamente esiste, è conveniente utilizzare **il criterio di Cohen** proposto nel 1969 (a pag. 54 del volume "Statistical power analysis for the behavioral sciences, New York, Academic Press), che ha il solo pregio di apparire **ragionevole, basato sul buon senso pratico, ma che non ha nessuna base teorica**.

Secondo il **criterio di Cohen**, il valore di  **$b$  è legato alla scelta di  $a$** , secondo la relazione:

$$b = 4a$$

o, tradotto in potenza, approssimativamente

$$1 - b = 1 - 4a$$

Così, se l'errore di primo tipo è  $a = 0.01$ , è ragionevole scegliere l'errore di secondo tipo  $b = 0.05$ .

I rapporti utilizzati con frequenza maggiore sono:

- $a = 0.02$  e  $b = 0.10$
- $a = 0.05$  e  $b = 0.20$
- $a > 0.05$  e  $b = 0.25$ .

#### 4.5. DIMENSIONI (n) DEL CAMPIONE, NEL CASO DI PROPORZIONI

Anche nel confronto tra le proporzioni ( $p_1$  e  $p_2$ ) di due campioni indipendenti, per valutare la significatività della loro differenza ( $d = p_1 - p_2$ ), è possibile commettere errori di due tipi.

Il primo, chiamato **errore di I Tipo** (*Type I error*) o di prima specie e che consiste nel **dichiarare che la differenza tra le due proporzioni è significativa, quando in realtà è uguale a 0**, è quello che ha avuto l'attenzione maggiore nei testi di statistica e nell'applicazione dell'inferenza.

Ma, secondo vari autori, è un punto di vista puramente teorico. **Nella realtà è una preoccupazione eccessiva, in quanto tale errore non è mai commesso nella pratica sperimentale.**

Come sottolinea **Joseph L. Fleiss** (nel cap. 3 del volume *Statistical Methods for Rates and Proportion*, 1973, John Wiley & Sons, New York) molti autori di testi di statistica applicata pensano che **quasi mai due popolazioni siano identiche, poiché inevitabilmente esiste sempre una differenza**, per quanto piccola e insignificante essa possa essere.

Quantificare la differenza che ha significato biologico od ecologico, un approccio ritenuto valido anche nella ricerca ambientale. Nella programmazione di un esperimento, un ricercatore dovrebbe **evidenziare solo le differenze che assumono una importanza reale nella sua disciplina**, non astrattamente una differenza di qualsiasi entità, anche effettivamente trascurabile; di conseguenza, **deve rivolgere la sua attenzione all'evitare di raccogliere un campione molto più grande di quanto gli sia necessario per non commettere l'errore di II tipo**. Con un aumento ingiustificato del campione, avrebbe un incremento dei costi e dei tempi oltre quanto gli è utile per ottenere l'esperimento significativo che desidera.

Per verificare la significatività della differenza **tra due proporzioni** in campioni di grandi dimensioni, si può ricorrere alla distribuzione normale. Con la consueta simbologia, già illustrata nel capitolo precedente, si ricorre al calcolo di  $z$  mediante la relazione

$$z = \frac{p_2 - p_1}{\sqrt{\bar{p}(1-\bar{p}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

con  $\bar{p}$  uguale alla media ponderata di  $p_1$  e  $p_2$ .

Nel caso di due campioni delle stesse dimensioni ( $n$ ), detti bilanciati, si può scrivere

$$z = \frac{p_2 - p_1}{\sqrt{\frac{2\bar{p}(1-\bar{p})}{n}}}$$

Per **non commettere un errore di I tipo** alla probabilità  $\alpha$ , in un **test bilaterale** occorre che il valore di  $z$  sia superiore alla quantità riportata nella tabella **sinottica della distribuzione normale alla probabilità  $\alpha/2$** .

Il concetto è scritto come

$$|z| > C_{\alpha/2}$$

e in un **test unilaterale**

$$|z| > C_{\alpha}$$

dove  $C$

è il valore riportato nella tabella della distribuzione  $z$ .

Quando  $\alpha = 0.05$  tale valore è

- uguale a 1,96 per un test bilaterale,
- uguale a 1,645 per un test unilaterale.

Nello stesso tempo, quando si raccolgono i dati del campione, per **non commettere un errore di II tipo** alla probabilità  $\beta$ , **che è sempre unilaterale**, il valore di  $z$  deve essere superiore a quello corrispondente alla probabilità  $\beta$

$$|z| > C_{\beta}$$

per la quale si usa sempre la distribuzione normale, con la semplice sostituzione del simbolo  $\alpha$  con il simbolo  $\beta$ .

Impostato su  $n$  e considerando contemporaneamente le due condizioni espresse,

- in **un test con  $H_1$  bilaterale**

la soluzione per stimare il numero di dati da raccogliere per ognuno dei due campioni, diventa

$$n = \frac{\left( z_{\alpha/2} \cdot \sqrt{2\bar{p}(1-\bar{p})} + z_b \cdot \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right)^2}{(p_2 - p_1)^2}$$

e

- **per un test  $H_1$  unilaterale**

è sufficiente sostituire  $Z_{\alpha/2}$  con  $Z_\alpha$ .

$$n = \frac{\left( z_\alpha \cdot \sqrt{2\bar{p}(1-\bar{p})} + z_b \cdot \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right)^2}{(p_2 - p_1)^2}$$

Quando il **campione è di piccole dimensioni** (poche decine di osservazioni per campione), nel test di verifica dell'ipotesi nulla con il test  $\chi^2$  si apporta la correzione per la continuità o correzione di Yates.

Con la formula già presentata nel capitolo precedente è

$$z = \frac{|p_1 - p_2| - \frac{1}{2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{\bar{p}(1-\bar{p}) \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Questa correzione ha uno scopo cautelativo: abbassa la significatività del test, appunto perché con pochi dati le conclusioni sono meno attendibili e nella logica statistica non si vuole rifiutare l'ipotesi nulla quando la risposta è incerta.

Nella stima della dimensione  $n$  del campione da raccogliere, per rendere il test significativo alla stessa probabilità, è quindi necessario aumentare la quantità  $n$ , calcolata con la formula generale già descritta.

Nel 1959, **M. Kramer e S. W. Greenhouse** (nell'articolo *Determination of sample size in treatment-control comparison for chronic disease studies in which drop-out or non-adherence is a problem*, pubblicato dalla rivista *J. chronic. Dis.*, n. 20, pp. 233-239) hanno proposto  **$n'$ , una stima corretta di  $n$** , che tiene appunto presente la correzione per la continuità.

Dopo avere stimato  $n$ , **per considerare la correzione per la continuità di Yates**, si perviene ad una sua **valutazione corretta  $n'$**  mediante la relazione

$$n' = \frac{n}{4} \cdot \left[ 1 + \sqrt{1 + \left( \frac{8}{n} \cdot \frac{1}{|p_2 - p_1|} \right)} \right]^2$$

Come esempio del calcolo di  $n$ , supponendo

$$p_2 = 0,28 \text{ e } p_1 = 0,12 \quad \text{con} \quad a = 0,05 \text{ e } b = 0,10$$

in un test bilaterale in cui  $z_{a/2} = 1,96$  e  $z_b = 1,28$  con  $\bar{p} = (0,28 + 0,12)/2 = 0,2$

il numero  $n$  di dati per ognuno dei due gruppi è

$$n = \frac{\left(1,96 \cdot \sqrt{2 \cdot 0,2 \cdot 0,8} + 1,28 \cdot \sqrt{(0,28 \cdot 0,72) + (0,12 \cdot 0,88)}\right)^2}{(0,28 - 0,12)^2}$$

$$n = \frac{(1,96 \cdot 0,5657 + 1,28 \cdot 0,5543)^2}{0,16^2} = \frac{(1,1088 + 0,7075)^2}{0,16^2} = \frac{3,299}{0,0256} = 128,9$$

uguale a 128,9; cioè 129 in ognuno dei due gruppi a confronto.

Trattandosi di campioni di poco superiori a 100 osservazioni, vari autori di testi di statistica suggeriscono di usare un valore di  $n$  leggermente superiore,  $n'$  appunto, che in questo caso risulta

$$n' = \frac{N}{4} \cdot \left[1 + \sqrt{1 + \left(\frac{8}{N} \cdot \frac{1}{|p_2 - p_1|}\right)}\right]^2 = \frac{129}{4} \cdot \left[1 + \sqrt{1 + \left(\frac{8}{129} \cdot \frac{1}{0,16}\right)}\right]^2 = 32,25 \times (1 + 1,1779)^2 = 142$$

uguale a 142.

Vari programmi informatici, insieme con la stima di  $n$  (il valore della potenza a priori), nel tabulato della risposta forniscono anche la significatività del test a posteriori, qualora si realizzassero esattamente le condizioni supposte per la stima di  $n$ .

Se prima della raccolta dei dati il valore di  $a$  poteva essere uguale a 0,05, una volta raccolti i dati (a posteriori) la significatività del test è di gran lunga superiore (quindi  $a$  nettamente minore), poiché **il calcolo a priori di  $n$  inglobava il rischio che la differenza ipotizzata tra le due proporzioni fosse, per variazione casuale, minore dell'atteso.** Nella stima a priori di  $n$  è compresa la probabilità  $b$  di commettere un errore di II tipo; a posteriori questo rischio non esiste più.

Come dimostrazione semplice di questi concetti, è utile stimare la significatività della differenza tra le proporzioni dei due campioni, se poi effettivamente avessero le proporzioni  $p_1$  e  $p_2$  ipotizzate. Con la formula generale

$$z = \frac{p_2 - p_1}{\sqrt{\bar{p}(1-\bar{p}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

e i dati già riportati

nell'esempio precedente  $p_1 = 0,20$   $p_2 = 0,28$   $n_1 = n_2 = 129$  si otterrebbe un valore di  $z$

$$z = \frac{0,28 - 0,20}{\sqrt{0,2 \cdot 0,8 \cdot \left(\frac{1}{129} + \frac{1}{129}\right)}} = \frac{0,08}{\sqrt{0,16 \cdot 0,0155}} = \frac{0,08}{0,0498} = 1,61$$

uguale a 1,61.

Ad esso, in un test bilaterale, corrisponde una probabilità  $\alpha = 0,0014$  o 1,4 per mille.

Il test sarebbe quindi molto più significativo del livello  $\alpha$  (5%) definito in precedenza.

Nel caso in cui serva una sola frequenza relativa, e quindi per test che utilizzano la distribuzione binomiale, la stima della dimensione  $n$  del campione è data da

$$n = \frac{z^2 \cdot p \cdot q}{d^2}$$

dove

$z$  è uguale a 1,96 per la probabilità 0,05

$d$  rappresenta l'errore di stima ritenuto accettabile,

mentre  $p$  e  $q$  sono le proporzioni dei due caratteri alternativi ( $p + q = 1$ ).

Anche in questo caso,  $p$  e  $q$  devono essere noti o determinati da un esperimento pilota oppure assunti entrambi uguali a 0,5, poiché per essi si ha la varianza massima e quindi anche il valore massimo di  $n$

Il numero minimo, ma sicuramente sufficiente di dati utili è  $n_{\max}$ , che può quindi essere ottenuto dalla relazione

$$n_{\max} = \frac{z^2}{4d^2}$$

Dopo aver raccolto il numero di osservazioni stimato, è possibile che la proporzione effettiva del campione sia diversa da quella stimata, appunto per **errore** (in statistica, sinonimo di variabilità non di sbaglio) **di campionamento**. L'accuratezza  $d$  della proporzione trovata può essere stimata con

$$d = \sqrt{\frac{z^2 \cdot p \cdot q}{n}}$$

ESEMPIO 1. In un'indagine pilota, su una decina di laghi è stato trovato che la specie X è presente nel 25% dei casi. Con una ricerca di dimensioni maggiori, si intende stimare la vera proporzione di laghi in cui la specie X in oggetto è presente. Quanti laghi occorre analizzare affinché 95 volte su cento la vera proporzione sia compresa in un intervallo  $\pm 7$  rispetto alla proporzione stimata nel campione?

Risposta.

Per stimare la proporzione di laghi in cui è presente (od assente) la specie X si può utilizzare la distribuzione binomiale, anche se il numero n risulta elevato e si ha una buona approssimazione alla normale. Alla probabilità  $\alpha = 0.05$  il valore di  $z$  è uguale a 1,96; poiché  $d$  è richiesto uguale a 7 (uguale a 0,07 in proporzione), il numero di laghi da campionare è

$$\frac{1,96^2 \times 0,25 \times 0,75}{0,07^2} = \frac{3,8416 \times 0,25 \times 0,75}{0,0049} = 147$$

uguale a 147.

ESEMPIO 2. Si supponga che nell'analisi dei 147 laghi, la presenza di un'altra specie (Y) sia stata riscontrata nel 15% dei casi.

Quale è l'accuratezza  $d$  della proporzione 0,15 trovata?

Risposta.

$$d = \sqrt{\frac{3,8614 \times 0,15 \times 0,85}{147}} = 0,058$$

La stima del 15% di presenza della specie Y con il 95% di probabilità ha un'accuratezza di  $\pm 6\%$  (valore arrotondato della proporzione  $p = 0,058$ ). Con 147 analisi, 95 volte su 100 la proporzione stimata di 15% si colloca in un intervallo ( $\pm 6\%$ ) e quindi si troverà tra 9% e 21%.

#### **4.6. . LE QUATTRO PROPRIETA' CHE DEVE AVERE UNO STIMATORE: CORRETTEZZA, EFFICIENZA, CONSISTENZA, SUFFICIENZA**

Una categoria rilevante di problemi dell'inferenza statistica è **la stima del valore di un parametro ignoto della popolazione, quando si dispone solamente di misure tratte da un campione**. I parametri utilizzati con frequenza maggiore sono

- **la media e la varianza** per fenomeni quantitativi misurati con una scala d'intervalli o di rapporti,
- **la frequenza relativa** o proporzione in caso di fenomeni qualitativi,
- **il coefficiente di correlazione lineare**, come indice statistico bidimensionale, quando si abbiano due fenomeni quantitativi.

E' intuitivo che per conoscere l'intensità di un fenomeno o il valore di un fattore si debba raccogliere una serie di rilevazioni e calcolarne la media. E' un **criterio di stima puntuale**, in quanto fa riferimento ad un valore unico, che può essere rappresentato come un punto sull'asse dei valori reali. Questo metodo non è privo d'inconvenienti: quando non si può disporre di tutti i dati dell'universo: **il valore sintetico del campione raccolto non coincide esattamente con il valore reale**; anzi, ne differisce di una quantità ignota.

Per arrivare alla stima migliore del parametro della popolazione o universo partendo dal campione, sono state definite **quattro proprietà, di cui uno stimatore puntuale dovrebbe godere**:

- **correttezza**,
- **efficienza**,
- **consistenza**,
- **sufficienza**.

Si ha **correttezza** di uno stimatore quando, estraendo dal medesimo universo vari campioni con lo stesso numero di osservazioni, **le singole stime risultano maggiori o minori del valore vero (ignoto), per differenze che tendono a compensarsi**. Nell'approccio classico all'inferenza statistica, il **principio del campionamento ripetuto** è fondato sulla correttezza: **estraendo varie volte un campione di n elementi, l'insieme delle differenti stime ha media uguale al valore del parametro dell'universo, se lo stimatore è corretto**.

Uno **stimatore non corretto è detto distorto**; la differenza tra la media generale dei vari campioni e il valore (vero) del parametro della popolazione è detta **distorsione nella stima** o **errore sistematico**.

Si ha **efficienza** di uno stimatore, quando **le varie misure sono vicine al valore reale della popolazione**. Fra più stimatori, il migliore è quello che ha la varianza minore. Correttezza ed efficienza sono concetti diversi e definiscono metodi di valutazione differenti; ma a volte sono in concorrenza per scegliere lo stimatore migliore, che è sempre quello che maggiormente si avvicina al valore dell'universo.

Si ha **consistenza**, quando **all'aumentare del numero di osservazioni la differenza tra la stima ed il valore vero diventa minore di una quantità prefissata**. E' quindi possibile scegliere un campione sufficientemente grande, da assicurare una differenza inferiore alla quantità prefissata.

Si ha **sufficienza**, quando uno stimatore **sintetizza tutte le informazioni presenti nel campione**, che sono importanti per la stima del parametro.

#### 4.7. INTERVALLO FIDUCIALE DI UNA MEDIA

Le tre tecniche d'analisi

- sulla **potenza di un test (1-b)**,
- per la stima delle **dimensioni (n) di un campione** e
- per la stima dell'**intervallo di confidenza**

sono tra loro collegate nei concetti, nelle finalità e nelle formule da applicare.

**Quando si programma un esperimento, le stime della potenza del test e delle dimensioni del campione hanno lo scopo principale di indicare:**

- a) quanti dati campionari è necessario raccogliere, perché le analisi statistiche forniscano risultati accurati e attendibili, evitando che il campione sia troppo piccolo (quindi non significativo) o troppo grande (quindi con costi e tempi eccessivi, rispetto a quanto necessario per la ricerca da condurre);
- b) con quale probabilità i test statistici potranno dimostrare la presenza di effetti di una data dimensione, nella situazione sperimentale prevista.

In questo contesto, **la stima dell'intervallo di confidenza** è utile per

- a) implementare gli obiettivi,
- b) valutare le dimensioni (la media, la proporzione o la varianza) del fenomeno nella realtà.

Inoltre, un numero crescente di testi di statistica suggeriscono di **utilizzare la stima dell'intervallo di confidenza anche per l'inferenza, in sostituzione dei test tradizionali**, nell'analisi di dati sperimentali.

Come per la potenza, anche per capire i concetti sull'intervallo fiduciale occorre rifarsi alla teoria del campionamento. Se, da una popolazione (quindi con un numero infinito di dati) distribuita normalmente e con media reale  $m$ , si estrae a caso un campione di  $n$  oggetti, a causa dell'**errore di campionamento**

- la media campionaria  $\bar{x}$  non avrà un valore identico alla media  $\mu$  della popolazione,
- ogni media  $\bar{x}$  di un nuovo campione avrà un valore diverso,

- la distribuzione di questi valori  $\bar{x}$  avrà forma normale
- intorno alla media reale  $\mu$ ,
- con una dispersione determinata dall'errore standard  $\frac{S}{\sqrt{n}}$ .

Inversamente, ed è la situazione più frequente, **dalla media  $\bar{x}$  di un campione è possibile dedurre quale è la media  $m$  della popolazione.** Anche in questo caso, l'inferenza classica o frequentista non risponde con una sola misura, quella fornita da uno stimatore puntuale, ma fornisce due valori che determinano **un intervallo, entro il quale si trova il valore del parametro alla probabilità  $\alpha$  prescelta.** I due valori estremi sono detti **limiti fiduciali** o **limiti di confidenza** e comprendono l'**intervallo di confidenza.** Questo approccio serve anche per il test d'inferenza e fornisce esattamente le stesse conclusioni.

**Per un test bilaterale, se l'intervallo di confidenza alla probabilità  $1-\alpha$  contiene il valore espresso nell'ipotesi nulla  $H_0$  (di solito uguale a 0), non esistono prove sufficienti per respingerla, ad una probabilità  $P < \alpha$ . Se l'intervallo di confidenza non contiene tale valore, esistono elementi sufficienti per rifiutare l'ipotesi nulla  $H_0$ , con una probabilità  $P < \alpha$ .**

Secondo alcuni autori, l'intervallo di confidenza o dei limiti fiduciali addirittura rappresenta un approccio preferibile per interpretare i risultati di esperimenti, rispetto ai tradizionali test di ipotesi. Infatti fornisce anche un'idea della precisione della stima, che è funzione inversa dell'ampiezza dell'intervallo: a parità di altre condizioni, un intervallo di confidenza minore corrisponde ad una stima più precisa.

Gli intervalli di confidenza vengono utilizzati anche per stimare i limiti entro i quali si colloca una certa frazione dei valori, come i "**limiti di normalità**" nei test di laboratorio. E' possibile individuare quando un singolo dato o una risposta media di una serie di rilevazioni rientra nella norma (i valori più frequentemente riscontrabili nella popolazione) oppure rappresenta un valore anomalo, cioè vicino agli estremi. In questo caso, è indice di una situazione presumibilmente patologica per l'individuo o l'ambiente.

Per calcolare l'intervallo di confidenza della media  $m$  (ignota) di una popolazione con varianza  $S^2$  (nota), a partire dalla media  $\bar{x}$  calcolata da  $n$  osservazioni, si utilizza lo **scostamento standardizzato della media campionaria.**

Il valore  $z$  che ogni singolo campione estratto dalla popolazione assume nella distribuzione normale standardizzata è

$$z = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}}$$

Come la deviazione standard  $s$  valuta la distribuzione di campionamento delle osservazioni, l'**errore standard** ( $es$ ) stima la distribuzione delle medie di  $n$  dati,

$$es = \frac{\sigma}{\sqrt{n}}$$

La distribuzione di campionamento delle medie

- con media  $m$

- errore standard  $es = \frac{\sigma}{\sqrt{n}}$

segue una distribuzione normale standardizzata  $z$

$$P\left(-z \leq \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}} \leq +z\right) = P(z)$$

che può essere usata per determinare i limiti di confidenza.

Alla probabilità del 95%, con  $z = 1,96$  diventa

$$P\left(\bar{x} - 1,96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1,96 \frac{\sigma}{\sqrt{n}}\right) = 0,95$$

e significa:

con probabilità del 95% la media  $m$  della popolazione si trova nell'intervallo compreso tra gli estremi individuati dalla media più o meno il prodotto di  $z$  per l'errore standard

$$\mu = \bar{x} \pm 1,96 \cdot \frac{s}{\sqrt{n}}$$

Per una probabilità del 99% è sufficiente sostituire 1,96 con 2,58

$$\mu = \bar{x} \pm 2,58 \cdot \frac{s}{\sqrt{n}}$$

come applicazione della formula generale

$$\mu = \bar{x} \pm z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

ESEMPIO. Da una popolazione con  $s = 3$  è stato estratto un campione di 10 dati, la cui media risulta 25. Calcolare l'intervallo di confidenza entro il quale si troverà la media della popolazione alla probabilità del 99%.

Risposta.

Alla probabilità  $\alpha = 0.01$  corrisponde un valore di  $z$ , in una distribuzione bilaterale, uguale a 2,58. Di conseguenza, con

$\bar{x} = 25$ ,  $\sigma = 3$ ,  $n = 10$ , si calcolano due valori

$$\mu = 25 \mp 2,58 \cdot \frac{3}{\sqrt{10}} = 25 \mp 2,58 \cdot 0,9487 = 25 \mp 2,45 = \begin{cases} 22,55 \\ 27,45 \end{cases}$$

22,55 e 27,45.

La media reale  $m$  della popolazione con probabilità del 99% si trova nell'intervallo compreso tra 22,55 e 27,45.

Se la probabilità con la quale si desidera che la media sia compreso nell'intervallo è minore, ad esempio del 95%, è sufficiente prendere il valore di  $z$  per

- $\alpha = 0.05$  in una distribuzione bilaterale oppure
- $\alpha = 0.025$  in una distribuzione unilaterale

(sono entrambi uguali a 1,96)

#### 4.8. VALUTAZIONE DEL RISCHIO AGGIUNTIVO (f) E INTERVALLO FIDUCIALE DI UNA PROPORZIONE

Nello studio di una sostanza tossica, si valutano gli effetti letali mediante la proporzione di decessi, contati su un campione di  $n$  cavie dopo un determinato periodo di esposizione. Poiché si hanno decessi di cavie anche quando la sostanza non è tossica, il suo effetto reale è valutato mediante il rapporto tra i decessi avuti con la sostanza tossica e quelli avuti con la sostanza di confronto.

Se con la sostanza priva di tossico, nel tempo  $t$ , muore una proporzione  $p_1$  degli individui che formano il campione di controllo (ovviamente la proporzione di quelli che sopravvivono è  $1-p_1$ ), con la sostanza alla quale è stato aggiunto il tossico la proporzione di decessi sarà

$$p_2 = p_1 + f(1-p_1)$$

Supponendo, come caso di studio, che

$p_1 = 0,3$  e che

la sostanza tossica determini un incremento del 20 per cento ( $f = 0,2$ ) sulla frequenza dei decessi, la proporzione  $p_2$

$$p_2 = 0,3 + 0,2(1 - 0,3) = 0,44$$

risulterà uguale a 0,44 (non 0,50 come si otterrebbe con la somma  $0,3 + 0,2$  cioè  $p_1 + f$ ).

Sotto l'aspetto metodologico, occorre sottolineare che quando la mortalità del controllo è alta (per es.:  $p_1 = 0,70$ ), lo stesso effetto della sostanza tossica ( $f = 0,2$ ) precedente determina nel campione esposto una proporzione  $p_2$  di decessi con un aumento minore in valore assoluto; con i dati dell'esempio risulta

$$p_2 = 0,70 + 0,2(1 - 0,7) = 0,76$$

uguale a 0,76.

E' un incremento di **0,06** che potrebbe apparire determinato da un effetto aggiuntivo minore del **0,14** precedente, quando invece è esattamente della stessa dimensione (0,20).

E' quindi importante calcolare correttamente  $f$ , il **fattore di rischio aggiuntivo** o **la differenza relativa** di  $p_2$  rispetto a  $p_1$ , che è dato da

$$f = \frac{p_2 - p_1}{1 - p_1}$$

Come ulteriore dimostrazione pratica, si supponga che nello studio di un ambiente inquinato (con una quota di decessi, nel tempo  $t$ , pari a **0,28**) sia stato individuato ed eliminato uno dei fattori ritenuti responsabili dell'inquinamento; le analisi successive stimano, per lo stesso tempo  $t$ , una quota di decessi pari a **0,19**.

La **differenza relativa o il fattore di rischio f eliminato** (con  $p_2 = 0,28$  e  $p_1 = 0,19$ ) è

$$f = \frac{0,28 - 0,19}{1 - 0,19} = \frac{0,09}{0,81} = 0,111$$

uguale a 0,111.

Per ottenere i **limiti di confidenza di una proporzione o frequenza relativa** la procedura è analoga a quella per le medie, con la differenza che la deviazione standard  $s$  è determinata direttamente dal valore della media  $p$ .

Anche in questo caso, per utilizzare la distribuzione normale standardizzata  $z$  è richiesto un campione grande (secondo le dimensioni di  $n$  già definite); se il campione è piccolo, occorre fare riferimento alla distribuzione esatta, che nel caso di una frequenza relativa o proporzione è la distribuzione binomiale (l'argomento sarà trattato nel test dei segni, nel capitolo di statistica non parametrica per un campione).

Nel caso di frequenze relative o percentuali, la media generale di più osservazioni campionarie è stimata da tutte le osservazioni sperimentali, con la relazione

$$m = \frac{n_A}{n}$$

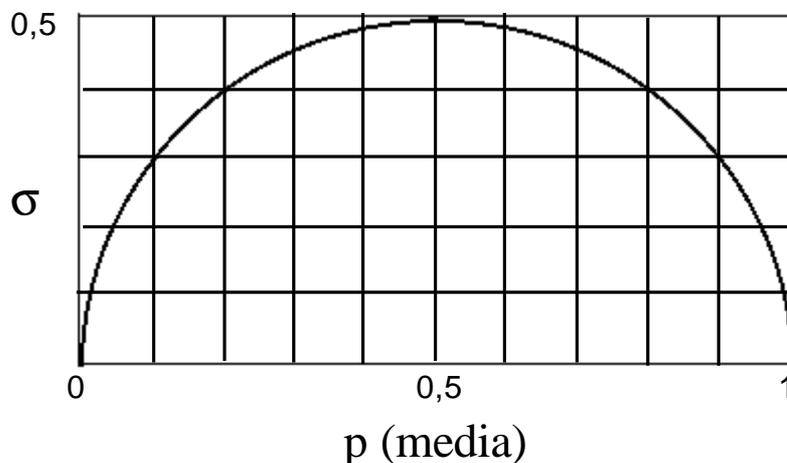
dove

- $n_A$  = il numero totale di oggetti A
- $n$  numero totale di oggetti considerati.

Nella distribuzione binomiale,  $m = np$  e  $s^2 = npq$  con  $q = 1 - p$

Quando la media è espressa come proporzione, quindi  $m = p$ , la sua varianza è  $s^2 = pq$

Questo rapporto è descritto dalla figura seguente, nella quale di norma sull'asse delle ascisse è riportata la media  $p$  e sull'asse delle ordinate la sua deviazione standard  $\sigma = \sqrt{pq}$



Questa relazione stretta tra media e varianza che caratterizza le proporzioni, eventualmente trasformate in percentuali, porrà problemi quando ad esse si applicheranno test parametrici per l'inferenza, quali il test **t** di **Student** e l'analisi della varianza o test **F** di **Fisher-Snedecor**.

Una delle condizioni fondamentali di validità per confrontare due o più medie, è che i vari gruppi a confronto abbiano la stessa varianza. E' una condizione irrealizzabile nel caso di proporzioni, poiché quando le medie sono diverse anche le varianze sono diverse, a causa della relazione che lega media e varianza. Si dovrà quindi ricorrere a trasformazioni dei dati espressi come proporzioni o percentuali; in altri casi, che saranno discussi in seguito, si dovranno utilizzare test non parametrici per 2 o per k campioni.

In analogia all'errore standard  $\frac{s}{\sqrt{n}}$  della media **m** per misure continue,

nel caso di una proporzione **p** (che è una media) **la sua deviazione standard s** (che in realtà, essendo la dispersione della media, è un errore standard) è

$$s = \sqrt{\frac{p \cdot (1-p)}{n}}$$

Da essi è possibile stimare l'intervallo fiduciale di una proporzione

Conoscendo la proporzione reale **p** di una popolazione (quale l'effetto dichiarato di un tossico che dopo una settimana determina la morte di una proporzione **p** di individui) è possibile stimare la distribuzione della proporzione campionaria **p**, in un gruppo di **n** individui, mediante la relazione

$$P\left(p - z_{\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} < p < p + z_{\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}\right) = 1-\alpha$$

Per la probabilità del 95% può essere scritta come

$$P\left(p - 1,96 \cdot \sqrt{\frac{p \cdot (1-p)}{n}} < p < p + 1,96 \cdot \sqrt{\frac{p \cdot (1-p)}{n}}\right) = 0,95$$

Per i calcoli è conveniente utilizzare

$$p = p \pm z_{\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

dove,

in una distribuzione bilaterale (quindi  $\alpha/2$  in ogni coda), il valore di **z**

- per la probabilità del 95% è uguale 1,96

- per la probabilità del 99% è uguale a 2,58.

ESEMPIO 1. Con numerose ricerche è stato dimostrato che un tossico diluito in acqua alla concentrazione standard determina la morte del 30% degli individui della specie A.

Entro quali limiti alla probabilità del 95% sarà compresa la frequenza relativa dei decessi in un campione di 80 individui?

Risposta

Con  $z = 1,96$  associata alla probabilità  $\alpha = 0.05$ ,  $p = 0,3$  e  $n = 80$

$$p = 0,3 \pm 1,96 \cdot \sqrt{\frac{0,3 \cdot 0,7}{80}} = 0,3 \pm 1,96 \cdot 0,051 = 0,3 \pm 0,1$$

si ottiene un valore di  $p$

compreso nell'intervallo tra 0,2 e 0,4.

Con una probabilità del 95% una frequenza relativa  $\pi = 0,3$  in campioni di 80 individui può variare tra 0,2 e 0,4.

ESEMPIO 2. Il tossico X determina la morte del 4% delle cavie utilizzate. Entro quali limiti alla probabilità del 99% sarà compresa la percentuale di decessi in un esperimento con 500 individui?

Risposta.

Con  $z = 2,58$  associata alla probabilità bilaterale  $\alpha = 0.01$  e

con  $p = 0,04$  e  $n = 500$

$$p = 0,04 \pm 2,58 \cdot \sqrt{\frac{0,04 \cdot 0,96}{500}} = 0,04 \pm 2,58 \cdot 0,0088 = 0,04 \pm 0,023$$

per il valore campionario  $p$  si ottiene un intervallo compreso tra 0,017 e 0,063.

Sovente i testi di statistica applicata riportano, in forma grafica oppure in tabelle, il campo di variazione alla probabilità  $1-\alpha$  prefissata di una percentuale campionaria  $p$  estratta da una popolazione con percentuale  $p$ . I valori, indicati come proporzioni, sono quelli della tabella successiva.

INTERVALLO DI VARIAZIONE DI **p** IN RAPPORTO ALLA PROPORZIONE **p**  
 DELLA POPOLAZIONE E ALLA DIMENSIONE **n** DEL CAMPIONE,  
 ALLA PROBABILITÀ DEL 95 % .

$\pi$	DIMENSIONI (n) DEL CAMPIONE					
	20	50	100	200	500	1000
<b>0.050</b>	--- ---	--- ---	.007 - .093	.020 - .080	.031 - .069	.036 - .064
<b>0.100</b>	--- ---	.017 - .183	.041 - .159	.058 - .142	.074 - .126	.081 - .119
<b>0.200</b>	.025 - .375	.089 - .311	.122 - .278	.145 - .255	.165 - .235	.175 - .225
<b>0.300</b>	.099 - .501	.173 - .427	.210 - .390	.236 - .364	.260 - .340	.272 - .328
<b>0.400</b>	.185 - .615	.264 - .536	.304 - .496	.332 - .468	.357 - .443	.370 - .430
<b>0.500</b>	.281 - .719	.361 - .639	.402 - .598	.431 - .569	.456 - .544	.469 - .531
<b>0.600</b>	.385 - .815	.464 - .736	.504 - .696	.532 - .668	.557 - .643	.570 - .630
<b>0.700</b>	.499 - .901	.573 - .827	.610 - .790	.636 - .764	.660 - .740	.672 - .728
<b>0.800</b>	.625 - .975	.689 - .911	.722 - .878	.745 - .855	.765 - .835	.775 - .825
<b>0.900</b>	--- ---	.817 - .983	.841 - .959	.858 - .942	.874 - .926	.881 - .919
<b>0.950</b>	--- ---	--- ---	.907 - .993	.920 - .980	.931 - .969	.936 - .964

La sua lettura è semplice. Per esempio, estraendo da una popolazione con percentuale **p** uguale a 0.300 (30%) un campione di 20 individui (4<sup>a</sup> riga e 1<sup>a</sup> colonna), la percentuale campionaria **p** con probabilità del 95% è compresa nell'intervallo tra .099 e .501 (9,9% e 50,1%).

Mantenendo costante la probabilità, **all'aumentare del numero di osservazioni (n) il campo di variazione della stessa percentuale campionaria p si riduce**: sempre per **p** uguale a 0.300 con

- 50 osservazioni **p** sarà compresa tra .173 e .427 (17,3% e 42,7%);
- 100 osservazioni tra 21,0% e 39,0%;
- 200 osservazioni tra 23,6% e 36,4%;
- 500 osservazioni tra 26,0% e 34,0%;
- 1000 osservazioni tra 27,2% e 32,8%.

La tabella mostra anche che, alla stessa probabilità del 95% e con lo stesso numero (**n**) di osservazioni, **il campo di variazione è massimo quando p è 0,50 e minimo agli estremi** (0.05 oppure 0.95 nella tabella) in modo simmetrico.

Nei casi in cui le dimensioni campionarie (**n**) sono ridotte, non sono stati riportati i valori dell'intervallo fiduciale per le proporzioni **p** vicine a 0 né per quelle vicine a 1. **Con piccoli campioni e p vicino agli estremi, la distribuzione non può essere approssimata alla normale standardizzata**, tanto che i valori dell'intervallo (negativi o superiori a 1) diventano privi di significato per una proporzione.

In queste condizioni, come già ricordato e dimostrato nel capitolo 3, **si deve ricorrere alla distribuzione binomiale.**

Molto spesso, negli esperimenti in laboratorio e nella raccolta dei dati in natura, la situazione è opposta a quella appena illustrata: è la stima della frequenza relativa  $p$  della popolazione ad essere ignota, per cui sulla base dei dati di un campione, la cui media è  $\mathbf{p}$ , si vuole stimare **la frequenza relativa  $p$  dell'universo.**

Rovesciando l'impostazione precedente, per stimare **la proporzione reale  $p$**  a partire dalla conoscenza dalla **proporzione campionaria  $\mathbf{p}$**  stimata su  **$n$  dati**

si ricorre alla relazione

$$P\left(p - z_{\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} < \mathbf{p} < p + z_{\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}\right) = 1-\alpha$$

L'intervallo entro il quale si colloca  $\pi$  è dato da

$$\pi = p \pm z_{\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

dove

- per la probabilità del 95% il valore di  $\mathbf{z}$  è 1,96
- per una probabilità del 99% è uguale a 2,58.

ESEMPIO 1. In un campione di 80 fumatori, il 35% ( $\mathbf{p} = 0,35$ ) ha presentato sintomi di polmonite. Quali sono i limiti di confidenza entro i quali alla probabilità del 95% e del 99% si troverà la media reale ( $p$ ) di individui con sintomi di polmonite, nella popolazione dei fumatori?

Risposta.

$$\text{Per il 95\% : } \mathbf{p} = 0,35 \pm 1,96 \cdot \sqrt{\frac{0,35 \cdot 0,65}{80}} = 0,35 \pm 0,1045 = \left\langle \begin{array}{l} 0,2455 \\ 0,4545 \end{array} \right.$$

$$\text{Per il 99\% : } \mathbf{p} = 0,35 \pm 2,58 \cdot \sqrt{\frac{0,35 \cdot 0,65}{80}} = 0,35 \pm 0,1376 = \left\langle \begin{array}{l} 0,2124 \\ 0,4876 \end{array} \right.$$

Alla probabilità del 95%, la percentuale reale  $p$  di una popolazione dalla quale sia stato estratto un campione di 80 individui con percentuale 35% si trova tra il 24,55% e il 45,45%.

Con probabilità 99% si trova tra 21,24% e il 48,76%.

ESEMPIO 2. E se il campione raccolto fosse stato di 100 fumatori dei quali 35 con sintomi di polmonite, quale sarebbe l'intervallo di variazione di  $p$  sempre alla probabilità 95% e 99%?

Risposta.

$$\text{Per il 95\% : } p = 0,35 \pm 1,96 \cdot \sqrt{\frac{0,35 \cdot 0,65}{100}} = 0,35 \pm 0,9349 = \left\langle \begin{array}{l} 0,2565 \\ 0,4435 \end{array} \right.$$

$$\text{Per il 99\% : } p = 0,35 \pm 2,58 \cdot \sqrt{\frac{0,35 \cdot 0,65}{100}} = 0,35 \pm 0,1231 = \left\langle \begin{array}{l} 0,2269 \\ 0,4731 \end{array} \right.$$

Alla probabilità del 95%, la media reale di una popolazione dalla quale è stato estratto un campione di 100 individui con media 35% si trova tra il 25,65% e il 44,35%.

Con probabilità 99% si trova tra 22,69% e il 47,31%.

Si può notare che con un campione di 100 individui gli intervalli sono leggermente più stretti di quelli stimati con un campione di 80 individui.

**La probabilità associata all'intervallo fiduciale di  $p$  ha un significato identico a quello della probabilità per l'intervallo di confidenza di  $m$ :**

- se dalla popolazione si estraessero tutti i possibili campioni e si costruissero tutti i possibili intervalli di confidenza,
- una frazione uguale a  $1-a$  comprenderebbe il valore reale di  $p$  e
- la rimanente frazione  $a$  non lo comprenderebbe.

#### **4.9. INTERVALLO DI CONFIDENZA DI UNA VARIANZA**

Anche nella ricerca ambientale, in varie situazioni può essere richiesto di stimare la varianza della popolazione  $s^2$  sulla base dei dati campionari. Ad esempio, stime della varianza sono necessarie per la verifica della precisione di uno strumento di misura, per lo studio della variabilità di una caratteristica ereditaria quantitativa, per l'omogeneità del livello d'inquinamento in un'area, poiché quando la variabilità è maggiore è più probabile che alcuni valori superino i limiti di legge, a parità degli altri parametri della distribuzione.

Il valore del  $\chi^2$  riportato nella formula generale sulle frequenze assolute

$$c_{(n-1)}^2 = \sum_{i=1}^n \frac{(Oss_i - Att_i)^2}{Att_i}$$

è una funzione di probabilità che varia da zero all'infinito positivo e serve per valutare la variabilità in funzione dei gradi di libertà.

**In una popolazione normale standardizzata** può essere scritto come

$$c_{(n-1)}^2 = \frac{\sum_{i=1}^n (x_i - m)^2}{s^2}$$

ricordando che

- la **frequenza assoluta osservata** coincide con il singolo valore  $x_i$
- la **frequenza assoluta attesa** in ognuno degli  $n$  gruppi coincide con  $m$  e
- in una distribuzione poissoniana (poiché la probabilità di trovare un individuo del gruppo  $i$  è piccola) la media e la varianza sono uguali  $m = s^2$ .

Di conseguenza, poiché con  $n$  dati campionari

$$\sum_{i=1}^n (x_i - m)^2 = Devianza = s^2 \cdot (n-1)$$

la formula del  $\chi^2$  può essere scritta come

$$c_{(n-1)}^2 = \frac{s^2 \cdot (n-1)}{s^2}$$

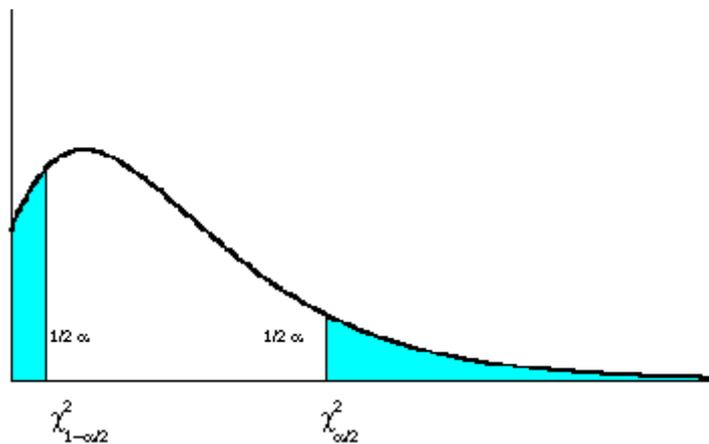
evidenziando ancor meglio come il  $\chi^2$  con gdl  $n-1$  sia un rapporto tra la somma delle varianze osservate e quella attesa, cioè tra la devianza campionaria e la varianza della popolazione.

**L'intervallo di confidenza di una varianza può essere ottenuto con la distribuzione  $c^2$ , poiché il rapporto tra la devianza campionaria di  $n$  dati e la varianza della popolazione seguono una distribuzione chi quadrato con  $n-1$  gradi di libertà.**

Il calcolo dell'intervallo fiduciale di una varianza richiede una procedura più complessa di quella della media, poiché la distribuzione  $c^2$ , a differenza della distribuzione  $z$ , non è simmetrica. Per calcolare l'intervallo di confidenza alla probabilità  $\alpha$ , occorre individuare i valori esatti di  $c^2$  che escludono una quantità  $(\alpha / 2)$  di possibili valori in ognuna delle due parti della distribuzione.

Ad esempio, per una probabilità  $\alpha$  pari a 0.05, si scelgono i due valori di  $\alpha$  in modo che

- uno escluda il 2,5% dei valori a sinistra, cioè il valore corrispondente a 97,5% o 0.975
- l'altro escluda il 2,5% o 0.025 dei valori a destra.



INTERVALLO ATTESO DI  $\chi^2_{(n-1)} = \frac{s^2(n-1)}{S^2}$  ALLA PROBABILITÀ 0.95

PER CAMPIONI ESTRATTI DA UNA POPOLAZIONE NORMALE

La figura descrive l'asimmetria dei due valori, quando il chi quadrato ha pochi gdl.

**L'intervallo di confidenza della varianza  $\sigma^2$**  della popolazione è calcolato con

$$\frac{s^2 \cdot (n-1)}{C^2_{1-\frac{\alpha}{2}}} > S^2 > \frac{s^2 \cdot (n-1)}{C^2_{\frac{\alpha}{2}}}$$

e l'intervallo di confidenza della deviazione standard  $S$  risulta

$$\sqrt{\frac{s^2(n-1)}{C^2_{1-\frac{\alpha}{2}}}} > S > \sqrt{\frac{s^2(n-1)}{C^2_{\frac{\alpha}{2}}}}$$

ESERCIZIO 1. Si vuole conoscere l'intervallo di confidenza al 99% della varianza relativa alla presenza di solventi clorurati totali nell'atmosfera di una città. A parità di media, inferiore al valore soglia, una varianza maggiore indica che potranno essere superati i limiti di legge con frequenza più elevata. L'ambientalista dovrà quindi valutare con maggiore prudenza i valori medi di una serie di rilevazioni che hanno varianza significativamente maggiore e predisporre interventi più radicali per una loro riduzione.

Da un campione di 16 osservazioni, è stata misurata la quantità media d'inquinamento (in  $\mu\text{g}/\text{m}^3$  a  $0^\circ\text{C}$  e a 1013 mbar) e si è calcolata una varianza campionaria  $s^2 = 8210,267$ .

Entro quali valori può trovarsi la varianza reale  $s^2$  al 99% di probabilità?

Risposta.

Dalla tabella dei valori critici del  $\chi^2$ , si devono rilevare i valori critici

- con 15 gdl associati rispettivamente
- alla probabilità 0.995 (1 - 0.005) e 0.005.

Essi risultano uguali rispettivamente a

$$\chi_{0,995; 15}^2 = 4,605 \quad \chi_{0,005; 15}^2 = 32,85$$

Da questi valori si calcola l'intervallo fiduciale della varianza reale  $s^2$

$$\frac{8210,67 \cdot 15}{32,85} < \sigma^2 < \frac{8210,67 \cdot 15}{4,605}$$

$$\mathbf{3749,16 < s^2 < 26744,85}$$

che, con probabilità del 99%, risulta compresa tra 3749,16 e 26744,85.

ESERCIZIO 2. Calcolare l'intervallo di confidenza al 95% della varianza reale  $\sigma^2$  del carapace di *Heterocypris incongruens*. La varianza di un campione di 41 individui, (misurati in mm), è risultata uguale a 0,0412.

Risposta.

Dapprima si devono individuare i valori critici del  $\chi^2$

- con 40 gdl e associati rispettivamente
- alle probabilità 0.975 e 0.025.

essi risultano rispettivamente

$$\chi_{0,975; 40}^2 = 24,44 \quad \chi_{0,025; 40}^2 = 59,20$$

uguali a 24,44 e 59,20.

Successivamente si calcola l'intervallo che, come estremi,

$$\frac{0,0412 \cdot 40}{59,20} < s^2 < \frac{0,0412 \cdot 40}{24,44}$$

$$0,0278 < s^2 < 0,0674$$

ha 0,0278 e 0,0674 mm.

**La distribuzione  $c^2$  non è simmetrica; quindi, a differenza della media campionaria  $\bar{x}$  il valore di  $s^2$  sperimentale non è collocato al centro dell'intervallo.**

E' condizione essenziale di validità che i dati siano distribuiti normalmente; l'assunzione è importante, ma difficile da dimostrare quando  $n$  è piccolo. Di conseguenza, **quando la normalità della distribuzione campionaria non può essere dimostrata, i risultati del calcolo dell'intervallo fiduciale di una varianza vanno applicati con molta cautela.**

Anche nel caso della varianza, **l'intervallo fiduciale serve per confrontare se una stima campionaria  $s^2$  si differenzia in modo significativo da un valore teorico o atteso  $s^2$  di una popolazione.** Se la varianza reale  $s^2$  è compresa nell'intervallo costruito intorno alla varianza campionaria  $s^2$ , non esiste una differenza significativa.

**Il test è identico a quello ottenuto con**

$$c^2_{(n-1)} = \frac{s^2(n-1)}{s^2}$$

**che risulterà**

- **significativo quando  $s^2$  non è compreso nell'intervallo**
- **non significativo quando  $s^2$  è compresa nell'intervallo costruito intorno a  $s^2$ , alla stessa probabilità  $\alpha$ .**

## CAPITOLO V

### INFERENZA SU UNA O DUE MEDIE CON IL TEST $t$ DI STUDENT

#### 5.1. LA DISTRIBUZIONE $t$ DI STUDENT

Nella prassi della ricerca sperimentale, utilizzare un test di inferenza sulla media campionaria ( $\bar{x}$ ) conoscendo  $s^2$ , la varianza della popolazione, come ipotizzato nel capitolo precedente, è un caso più teorico che reale. E' una procedura possibile, ricorrendo a pubblicazioni o ad esperienze personali fondate su molti dati, ma per un ricercatore è un fatto poco frequente. Quando la media della popolazione ( $m$ ) non è nota, di norma anche la sua **varianza ( $s^2$ ) è ignota**; di conseguenza, occorre utilizzare un sostituto della varianza della popolazione e **la varianza del campione ( $s^2$ ) ne rappresenta la stima più logica ed attendibile**.

Con  $s$  **ignota** ed il ricorso all'**uso di  $s$  in sua sostituzione**, la distribuzione delle probabilità non è più fornita dalla distribuzione normale  $z$  ma da quella del  **$t$** , detta  **$t$  di Student**, dallo **pseudonimo di W.S. Gosset**.

Questo chimico inglese, dopo un'iniziale carriera accademica, in assenza di prospettive all'Università lavorò per alcuni anni in una fabbrica di birra, eseguendo analisi statistiche su campioni dei prodotti, per un compito simile a quello che oggi viene chiamato controllo di qualità. Usando campioni necessariamente ridotti per motivi economici e di praticità, studiò lo scarto tra le medie dei campioni estratti dalla stessa popolazione e la media dell'universo, in rapporto all'errore standard. Ne derivò una distribuzione diversa dalla normale, soprattutto nel caso di piccoli campioni. Nella sua situazione contrattuale, non potendo diffondere i risultati delle sue ricerche per il timore dei suoi datori di lavoro che ciò potesse aiutare la concorrenza, pubblicò i suoi studi e descrisse le caratteristiche della nuova distribuzione con il "nom de plume", lo pseudonimo, di **Student** in due articoli su "Biometrika" negli anni 1907 e 1908.

Sono lavori che rappresentano, storicamente, i fondamenti della distribuzione  **$t$** .

In una sperimentazione statistica elementare, nella quale si voglia ripetere l'esperienza di Gosset, la distribuzione  **$t$**  può essere ottenuta, **con un campione costante** di dati ( $n$ ), dalle variazioni determinate dal rapporto

$$t = \frac{\text{media del campione} - \text{media della popolazione}}{\text{errore standard del campione}} = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}}$$

Per una migliore comprensione dei concetti fondamentali ed una corretta applicazione dei test che ne sono derivati, è importante evidenziare le caratteristiche specifiche che differenziano questa distribuzione dalla gaussiana:

- **la distribuzione normale considera la variazione di campionamento della sola media ( $\bar{x}$ ),**
- **mentre la distribuzione t di Student tiene conto anche della variazione di campionamento della deviazione standard (s).**

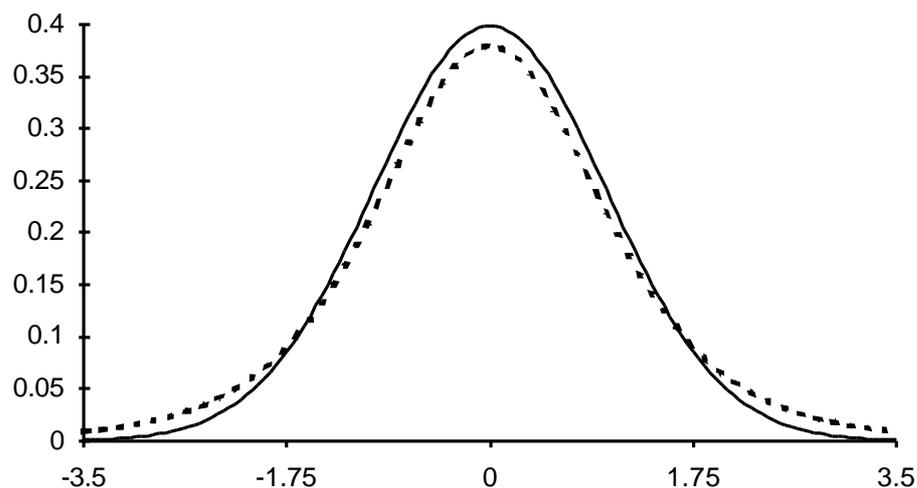
Per attuare una inferenza sulla media di una popolazione partendo da dati campionari, occorre pertanto **considerare nello stesso tempo**

- **sia la variazione di  $\bar{x}$  come stima di  $m$**
- **sia la variazione di s come stima di  $s$ .**

**All'aumento del numero di dati campionari (n), s è una stima sempre migliore di  $s$ .** Quando n è sufficientemente grande (teoricamente infinito, ma in pratica oltre 120; oppure, secondo altri testi, oltre 150 gdl), s e  $s$  hanno valori praticamente identici. Di conseguenza, si può affermare che, **all'aumentare di n, si ha la convergenza dei valori della distribuzione t di Student verso la distribuzione normale standardizzata z.**

Mentre è trascurabile oltre qualche decina di osservazioni, nel caso di piccoli campioni la differenza tra il valore del **t di Student** ed il corrispondente valore di **z** alla stessa probabilità  $\alpha$  è rilevante; pertanto, nella pratica sperimentale

- **i metodi che utilizzano il test t si riferiscono esclusivamente a piccoli campioni, spesso con una decina o meno di osservazioni**, raramente arrivano ad alcune decine.



DISTRIBUZIONE NORMALE STANDARDIZZATA (LINEA CONTINUA)  
E DISTRIBUZIONE t DI STUDENT PER 5 GRADI DI LIBERTÀ (LINEA TRATTEGGIATA).

La distribuzione **t di Student** è una distribuzione di probabilità teoriche, con un'area totale uguale al 100%, se espressa in percentuale, oppure uguale a 1, se espressa in valori unitari.

La sua **forma è simmetrica e a campana come la normale, ma con una dispersione maggiore.**

Analizzando la tabella della distribuzione **t**, si evidenzia che **non vi è una sola curva t** a differenza di quanto osservato per la gaussiana:

**- esiste una intera famiglia di distribuzioni t, una per ogni grado di libertà.**

Per un valore di gdl infinito, in pratica per un numero di dati di poco superiore al centinaio, curva dei valori **t** e curva dei valori **z** coincidono, come si può facilmente osservare dal confronto delle rispettive tabelle di distribuzione (riportate alla fine del capitolo). Sotto l'aspetto matematico, significa che **la distribuzione normale rappresenta il limite della distribuzione t, quando i gdl aumentano.**

Viceversa, al diminuire dei gdl la distribuzione **t** è progressivamente sempre più dispersa.

Il confronto tra la rappresentazione grafica della distribuzione normale e della distribuzione **t** con 5 gradi di libertà evidenzia visivamente ed in modo sintetico le differenze (vedi figura precedente).

I testi pubblicano una sola o al massimo 2 pagine di distribuzione dei valori della gaussiana, mentre è possibile calcolare tante pagine di distribuzione **t** quanti sono i gradi di libertà. In realtà, come per il chi quadrato, anche per la distribuzione **t** **abitualmente si utilizza una sola tavola sinottica**, una pagina ordinata di sintesi, **che riporta i valori critici più importanti.**

Alla fine del capitolo, per favorire la comprensione dei concetti e la stima delle probabilità, sono state riportate 3 differenti tabelle di distribuzione del **t di Student**:

- la prima è per un test bilaterale,
- la seconda è per un test unilaterale,
- la terza permette un confronto rapido tra le probabilità associate ai due casi.

Il modo di lettura delle tre tabelle è identico:

- la parte superiore di ogni colonna indica l'area sottesa nelle rispettive code della distribuzione,
- mentre ogni riga fa riferimento ad uno specifico grado di libertà, riportato nella prima colonna.

**I valori critici per l'area in una coda alla probabilità  $\alpha$  coincidono con quelli della probabilità  $2\alpha$  nella distribuzione a due code; viceversa, quelli associati alla probabilità  $\alpha$  in due code coincidono con i valori associati alla probabilità  $\alpha/2$  nella distribuzione a una coda.**

Per esempio,

- confrontando le rispettive tabelle si osserva che con 10 gdl in un test ad una coda per  $\alpha$  uguale a 0.05 il valore è 1,8125; per un test bilaterale, dove  $\alpha$  è uguale alla probabilità 0.025 nelle due code, sempre per 10 gdl il valore è 2,634;
- nella distribuzione ad una coda, i valori di  $t$  per  $\alpha$  uguale a 0.05 coincidono con la colonna di  $\alpha$  uguale a 0.10 nella distribuzione a due code; inversamente, i valori per  $\alpha$  uguale a 0.05 nella tabella per test a due code coincidono con la colonna dei valori di  $\alpha$  uguale a 0.025 nella tabella per test ad una coda.

**E' condizione di validità della distribuzione t di Student, e quindi dei test che la utilizzano, che la distribuzione dei dati sia normale e che le osservazioni siano raccolte in modo indipendente.**

Ma la distribuzione t è robusta, intendendo con tale termine tecnico che è approssimativamente valida anche per distribuzioni di dati con marcata deviazione dalla normalità.

Un test viene definito robusto quando i risultati possono essere accettati anche quando tutte le assunzioni di validità non sono verificate in modo rigoroso.

Nella statistica applicata, il test t è utilizzato in quattro casi: per il confronto tra

- 1 - la media di un campione e la media dell'universo o una generica media attesa;
- 2 - un singolo dato e la media di un campione, per verificare se possono appartenere alla stessa popolazione;
- 3 - la media delle differenze di due campioni dipendenti con una differenza attesa;
- 4 - le medie di due campioni indipendenti.

Per ognuno dei tre casi sulle medie

- media di un campione,
- **media delle differenze** tra due campioni dipendenti,
- **differenza tra le medie** di due campioni indipendenti,

è possibile calcolare l'intervallo fiduciale.

Come già visto con la distribuzione z, con esso si può conoscere l'intervallo entro il quale è collocato il valore reale della popolazione alla probabilità  $\alpha$ , partendo dalla misura campionaria. Inoltre, il metodo dell'**intervallo fiduciale rappresenta una alternativa ai test inferenziali corrispondenti.**

## **5.2. CONFRONTO TRA UNA MEDIA OSSERVATA E UNA MEDIA ATTESA; CALCOLO DEI LIMITI DI CONFIDENZA DI UNA MEDIA, CON S IGNOTA**

La distribuzione t con n-1 gdl (indicata con  $t_{n-1}$ ) è data dal rapporto

- tra la differenza della media campionaria  $\bar{x}$  e quella attesa  $m_0$  secondo l'ipotesi nulla,
- con il suo errore standard  $s/\sqrt{n}$

come espresso dalla formula

$$t_{(n-1)} = \frac{\bar{x} - m_0}{\frac{s}{\sqrt{n}}}$$

dove

n è il numero di dati,

s è la deviazione standard calcolata sui dati del campione,

E' importante ricordare che, come ampiamente presentato nel capitolo sulla statistica descrittiva, la deviazione standard è ottenuta a partire dalla devianza mediante la formula

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

dove

la parte sotto radice è chiamata **varianza campionaria corretta o stima corretta della varianza**.

Ai fini pratici del calcolo, la devianza  $\sum (x - \bar{x})^2$  è stimata più rapidamente e con maggiore precisione, in quanto non prevede arrotondamenti caratteristici delle medie, con la formula abbreviata

$$\sum (x - \bar{x})^2 = \sum x^2 - \frac{(\sum x)^2}{n}$$

Per verificare l'ipotesi relativa alla media  $\bar{x}$  di un campione rispetto ad una media attesa, **l'ipotesi nulla  $H_0$**  generalmente è scritta come

$$H_0 : \mu = \mu_0$$

dove

$\mu$  è la media della popolazione da cui è estratto il campione con media osservata  $\bar{x}$

$\mu_0$  è la media attesa o di riferimento per il confronto.

Con la medesima simbologia, **l'ipotesi alternativa  $H_1$**  può essere di tre tipi e

- scritta come:

$$H_1 : \mu \neq \mu_0$$

per un **test bilaterale** quando sono accettabili entrambe le risposte,

- oppure

$$H_1 : \mu < \mu_0$$

per un **test unilaterale**, al fine di verificare se la media del campione è significativamente minore della media attesa,

- oppure

$$H_1 : \mu > \mu_0$$

per un **test unilaterale di segno opposto**, al fine di verificare se la media della popolazione  $\mu$  da cui è estratto il campione è maggiore di quella attesa  $\mu_0$ .

Dalla precedente formula da utilizzare per verificare la differenza tra media campionaria e media attesa

$$t_{(n-1)} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

si può derivare quella dell'intervallo di confidenza, entro il quale alla probabilità  $\alpha$  è compresa la media reale  $\mu$  della popolazione dalla quale è estratto il campione. La formula per il calcolo dell'intervallo fiduciale diventa

$$\mu = \bar{x} \pm t_{\frac{\alpha}{2}; n-1} \cdot \frac{s}{\sqrt{n}}$$

dove  $t_{\frac{\alpha}{2}; n-1}$  indica il valore della distribuzione  $t$  con  $n-1$  gradi di libertà alla probabilità  $\frac{\alpha}{2}$ .

Sommando alla media campionaria le due parti, la quota positiva e quella negativa, si ottiene l'intervallo che comprende la media con probabilità prefissata  $\alpha$ .

ESEMPIO 1. In un appezzamento di terreno adibito a vivaio, sono coltivate pianticelle della specie A; una lunga serie di misure ha dimostrato che dopo due mesi dalla semina raggiungono un'altezza media di 25 centimetri. A causa di un incidente, su quel terreno sono state disperse sostanze tossiche; si ritiene che esse incidano negativamente sulla crescita di alcune specie, tra le quali la specie A.

Per una verifica di tale ipotesi, vengono seminate sul terreno inquinato 7 pianticelle che, controllate dopo 2 mesi, raggiungono le seguenti altezze in cm.: 22, 25, 21, 23, 24, 25, 21.

Si intende rispondere a due quesiti.

1 - Si può sostenere che le sostanze tossiche disperse inibiscano la crescita della specie A?

2 - Quale è la media reale dell'altezza delle piante dell'età di due mesi, nella nuova condizione del terreno?

Risposte.

1 - E' un test ad una coda in cui l'ipotesi nulla è

$$H_0 : \mu = \mu_0$$

e l'ipotesi alternativa è

$$H_1 : \mu < \mu_0$$

Infatti, se le sostanze tossiche inibiscono la crescita, la media  $m$  della popolazione da cui è estratto il campione di 7 piante può solo essere inferiore alla media  $m_0$  della popolazione precedente pari a 25.

**Il test assume significato solamente se la media campionaria  $\bar{x}$  è minore della media attesa  $m_0$ : il test serve per verificare se la differenza è casuale oppure significativa.** Se, con l'ipotesi alternativa espressa, la media campionaria fosse risultata superiore alla media attesa, diverrebbe totalmente inutile applicare il test: non si riuscirebbe mai a dimostrare che la media del campione è significativamente minore di quella espressa nell'ipotesi nulla.

Scegliendo una probabilità  $\alpha$  uguale a 0.05 e applicando la formula

$$t_{(n-1)} = \frac{\bar{x} - m_0}{\frac{s}{\sqrt{n}}}$$

dove, sulla base dei 7 dati campionari,

$$\bar{x} = 23,0 \quad s = 1,732 \quad t_{0,025; 6} = 2,447 \quad n = 7 \quad m_0 = 25,0$$

si ottiene un valore di  $t$  con **6 gdl**

$$t_{(6)} = \frac{23,0 - 25,0}{\frac{1,732}{\sqrt{7}}} = \frac{-2}{0,655} = -3,053$$

uguale a - 3,05.

**Il segno negativo indica solamente che la differenza è negativa rispetto al valore atteso; ai fini della significatività, il valore di t viene preso in modulo.**

Per un test ad una coda, il valore critico del **t** alla probabilità 0.05 con 6 gdl è uguale a 1,943.

Il valore calcolato in modulo è superiore a quello riportato nella tabella sinottica della distribuzione t. Pertanto, **con probabilità inferiore a 0.05 (di commettere un errore) si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa:** le sostanze tossiche disperse inibiscono la crescita delle piante della specie A in modo significativo.

2 - L'altezza media reale  $\mu$  della popolazione dalla quale sono stati estratti i 7 dati può essere stimata mediante l'intervallo fiduciale

$$\mu = \bar{x} \pm t_{\frac{\alpha}{2}; n-1} \cdot \frac{s}{\sqrt{n}}$$

Con i dati del campione

$$\bar{x} = 23,0 \quad s = 1,732 \quad n = 7$$

ed il valore del **t** associato alla probabilità 0.05 per un test a due code con 6 gdl  $t_{0,025; 6} = 2,447$

si calcola che

$$\mu = 23 \pm 2,447 \cdot \frac{1,732}{\sqrt{7}} = 23 \pm 1,602$$

Pertanto, alla probabilità complessiva  $\alpha = 0.05$  il limite inferiore  $l_1$  ed il limite superiore  $l_2$  dell'intervallo fiduciale risultano rispettivamente:

$$l_1 = 21,398 \quad l_2 = 24,602$$

In questo caso, non è possibile il confronto diretto tra le due risposte poiché

- nella prima è un test unilaterale (con  $t_{(0,05, 6)} = 1,943$ ),
- mentre nella seconda l'intervallo fiduciale utilizza il valore per un test bilaterale (con  $t_{(0,025,6)} = 2,447$ ).

**ESEMPIO 2.** Disponendo di un campione di 13 individui di *Heterocypris incongruens* pescati in un fiume, dei quali sono riportate le lunghezze (in mm),

Individui	Lunghezza (mm)
1	1,21
2	1,39
3	1,21
4	1,21
5	1,21
6	1,21
7	1,20
8	1,18
9	1,23
10	1,21
11	1,23
12	1,24
13	1,33

si vuole verificare se alla probabilità 0.99 la loro lunghezza media è significativamente differente dalla media di 1,25 mm stimata per la stessa specie nei laghi della regione, in varie ricerche precedenti.

Risposta.

E' possibile fornire una risposta sia mediante l'applicazione del test t per un campione (1), sia attraverso la stima dell'intervallo fiduciale della media campionaria (2).

1 - Dai 13 dati campionari, devono essere calcolati il valore della media e della deviazione standard:

$$\bar{x} = 1,235 \quad s = 0,059 \quad n = 13$$

La domanda dell'esempio richiede un test a due code o bilaterale, poiché prima della raccolta dei dati è ugualmente logico che la media del campione abbia un valore sia significativamente minore sia maggiore della media attesa.

Indicando con  $m$  la media reale del campione estratto dal fiume e con  $m_0$  la media della popolazione che vive nei laghi, l'ipotesi nulla è

$$H_0 : \mu = \mu_0$$

e l'ipotesi alternativa

$$H_1 : \mu \neq \mu_0$$

Mediante il test t

$$t_{(12)} = \frac{1,235 - 1,250}{\frac{0,059}{\sqrt{13}}} = \frac{-0,015}{0,01636} = -0,917$$

si ottiene un valore di  $t_{(12)}$  uguale a -0,917.

Alla probabilità  $\alpha = 0.01$  per un test bilaterale con 12 gdl il valore critico riportato è uguale a 3,055. Il valore calcolato in modulo è nettamente inferiore a quello corrispondente riportato nella tavola sinottica; di conseguenza, non si è in grado di rifiutare l'ipotesi nulla.

La dimensione media dei 13 individui della specie *Heterocypris incongruens* pescati nel fiume non è significativamente diversa da quella degli individui della stessa specie che vivono nei laghi della regione.

2 - Per la stima dell'intervallo fiduciale, dopo il calcolo dei medesimi parametri si deve scegliere il valore del t con 12 gdl

- alla probabilità  $\alpha = 0.01$  per un test a due code oppure
- alla probabilità  $\alpha = 0.005$  per un test a una coda

trovando in entrambi i casi  $t_{0,005; 12} = 3,055$ .

I valori del limite inferiore  $l_1$  e del limite superiore  $l_2$  dell'intervallo fiduciale

$$\mu = 1,235 \pm 3,055 \frac{0,059}{\sqrt{12}} = 1,235 \pm 0,05203$$

risultano uguali rispettivamente a  $l_1 = 1,175$  e  $l_2 = 1,287$ .

La media della popolazione  $\mu_0$  uguale a 1,25 è compresa nell'intervallo fiduciale della media campionaria. Pertanto, non esiste una differenza significativa alla probabilità prefissata di  $\alpha = 0.01$  in un test bilaterale.

**Si può osservare come questo risultato coincida con quello ottenuto nella prima parte della risposta: il valore atteso  $\mu_0$ , risulta non significativamente differente dal valore della media campionaria; pertanto alla stessa probabilità  $\alpha$  risulta compreso nel suo intervallo fiduciale. Se fosse stata rifiutata l'ipotesi nulla, il valore atteso risulterebbe escluso dall'intervallo fiduciale calcolato.**

Con un numero maggiore di osservazioni, la differenza facilmente sarebbe risultata significativa. Per la stima del **numero di dati utili**, è importante osservare che **un aumento del numero di dati campionari agisce doppiamente sulla riduzione dell'intervallo di confidenza e sulla significatività del t:**

- attraverso il **valore del  $t_{(n-1)}$** , che diminuisce al crescere di gdl,
- mediante **la riduzione dell'errore standard**, come evidenzia il rapporto  $\left(\frac{s}{\sqrt{n}}\right)$ .

### 5.3. CONFRONTO TRA UNA OSSERVAZIONE E LA MEDIA DI UN CAMPIONE

Nelle rilevazioni in natura ed in laboratorio, sovente sorge il problema che una certa misura possa essere errata, rispetto alle altre del campione raccolto. Le cause possono essere le più diverse: l'imperizia del nuovo rilevatore rispetto all'esperto che ha raccolto gli altri dati, il funzionamento non corretto dello strumento, una condizione ambientale differente, il materiale o i reagenti di qualità diversa, una procedura nuova o applicata in modo non corretto. In altri casi, come in tassonomia, sorge il problema di verificare se l'individuo che si sta analizzando possa essere di un'altra popolazione rispetto alle unità già raccolte.

La verifica a questi sospetti, che può essere condotta con un test unilaterale o bilaterale secondo le ipotesi formulate,

è fatta mediante un test t:

$$t_{(n_A-1)} = \frac{X_1 - \bar{X}_A}{\sqrt{s_A^2 \frac{n_A + 1}{n_A}}}$$

dove

$X_1$  = rilevazione da verificare;  $\bar{X}_A$  = media del campione;  
 $s_A^2$  = varianza del campione;  $n_A$  = numero di osservazioni del campione.

ESEMPIO. Si vuole verificare se una misura (49,7) può essere ritenuta diversa dalle 6 del campione raccolto (40,3 - 38,8 - 33,5 - 38,6 - 31,9 - 37,6).

Risposta.

La media del campione ( $\bar{X}_A$ ) è uguale a 36,873 e la varianza ( $s_A^2$ ) risulta uguale a 12,206.

Da essi deriva il valore di t con 6-1 gdl

$$t_{(5)} = \frac{49,7 - 36,873}{\sqrt{12,206 \cdot \frac{6+1}{6}}} = \frac{12,917}{3,774} = 3,42$$

che risulta uguale a 3,42.

I valori critici di t con 5 gdl, per un **test a una coda**, sono

- 2,015 alla probabilità  $\alpha = 0.05$ ;
- 3,365 alla probabilità  $\alpha = 0.01$
- 5,893 alla probabilità  $\alpha = 0.001$ .

Per un **test bilaterale**, i valori critici sono

- 2,571 alla probabilità  $\alpha = 0.05$
- 4,032 alla probabilità  $\alpha = 0.01$ .

Se è stato fatto un test bilaterale (il valore è diverso da quelli del campione?) si rifiuta l'ipotesi nulla con probabilità  $\alpha$  compresa tra 0.05 e 0.01.

Se il test era unilaterale (il valore è maggiore di quelli del campione?), si respinge l'ipotesi nulla con probabilità  $\alpha$  compresa tra 0.01 e 0.001.

#### 5.4. IL CONFRONTO TRA LE MEDIE DI DUE CAMPIONI

Nella ricerca ambientale, le situazioni più ricorrenti sono quelle del confronto tra due medie campionarie. In questi casi, la distribuzione **t** di Student può essere derivata dal rapporto tra la differenza delle due medie campionarie ed il suo errore standard:

$$t = \frac{\text{differenza fra 2 medie campionarie}}{\text{errore standard della differenza fra 2 medie campionarie}}$$

Il test di significatività tra due medie campionarie comporta un'ipotesi zero o ipotesi nulla, secondo la quale le due medie a confronto ( $\mu_1$  e  $\mu_2$ ) sono estratte dalla stessa popolazione o comunque sono identiche; di conseguenza, le differenze effettivamente riscontrate nelle medie campionarie  $\bar{x}_1$  e  $\bar{x}_2$  sarebbero imputabili a variazioni casuali, come effetti dovuti al campionamento, cioè all'estrazione casuale di alcuni dati da un universo teoricamente infinito, formato da valori tra loro diversi e con una distribuzione normale intorno alla loro media..

Nell'ipotesi nulla  $H_0$ , le due medie dell'universo o popolazione sono identiche:

$$H_0: \mu_1 = \mu_2 \quad \text{oppure} \quad H_0: \mu_1 - \mu_2 = 0$$

Mediante l'inferenza sulle medie calcolate sui dati di due campioni, si determina la probabilità di ottenere tra loro differenze così grandi o maggiori di quelle sperimentalmente osservate, nella condizione che l'ipotesi nulla  $H_0$  sia vera. Se questa probabilità risulta alta, si accetta l'ipotesi nulla; se la probabilità risulta piccola, convenzionalmente inferiore al 5%, come riportato nei testi a scopi puramente didattici, si inferisce che esiste una ragionevole evidenza per dubitare della validità dell'ipotesi nulla, che quindi è rifiutata. Di conseguenza, alla probabilità  $\alpha$  stimata di commettere un errore, si afferma l'esistenza di una differenza reale tra le due medie, dicendo che appartengono a popolazioni diverse.

L'esempio classico di un test  $t$  è il confronto tra un campione di individui sottoposti a trattamento ed un campione di soggetti che servono come controllo; è tuttavia possibile il confronto tra le medie di due trattamenti diversi.

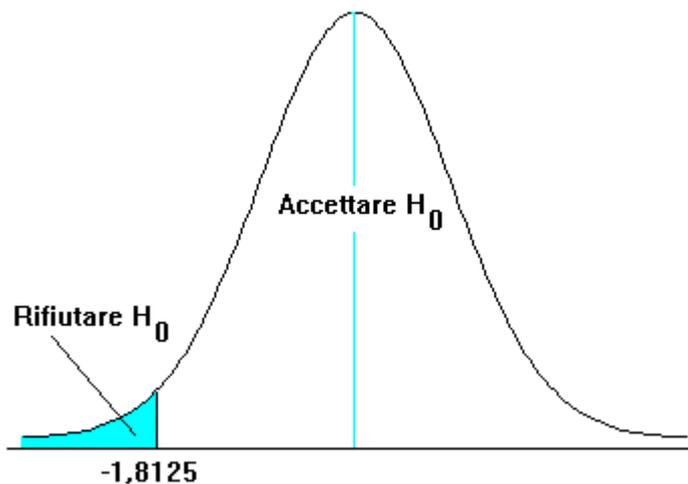
Nel primo caso, di norma si tratta di un test unilaterale o a una coda; nel secondo, di un test bilaterale o a due code. Questa direzionalità del confronto è insita nella natura stessa dell'esperimento; ma è sempre importante che essa sia evidenziata in modo esplicito, poiché da essa deriva la distribuzione delle probabilità a con le quali è possibile rifiutare l'ipotesi nulla.

Un **test è unilaterale o a una coda**, quando il ricercatore si chiede se una media è maggiore dell'altra, escludendo a priori che essa possa essere minore.

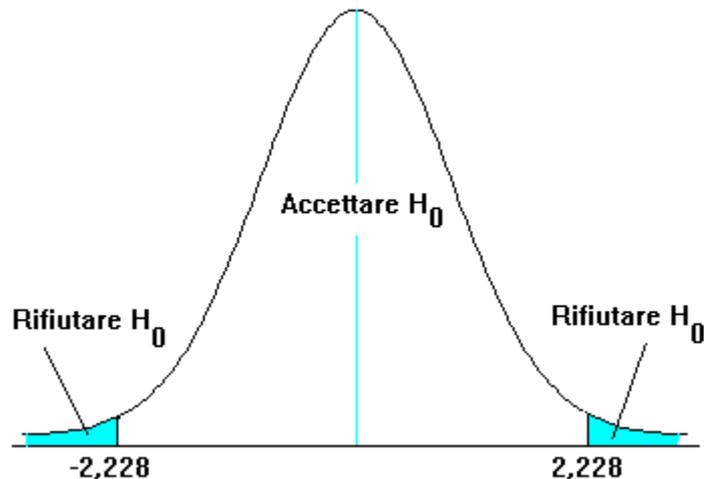
Un **test è bilaterale o a due code**, quando il ricercatore si chiede se tra le due medie esista una differenza significativa, senza che egli abbia indicazioni su quali sia la maggiore o la minore.

**Nel test ad una coda, la zona di rifiuto sarà solamente da una parte della distribuzione (a sinistra quando il segno è negativo, a destra quando è positivo); in un test a due code sarà simmetricamente distribuita dalle due parti.**

**E' maggiore la probabilità di dimostrare l'esistenza di una differenza significativa mediante un test ad una coda che con un test a due code.** Con un termine tecnico, si dice che **il test a due code è più conservativo, mentre quello ad una coda è più potente.**



TEST UNILATERALE PER UNA PROBABILITA' ASSOCIATA AL LIVELLO DI SIGNIFICATIVITÀ DEL 5% CON 10 GDL.



TEST BILATERALE PER UNA PROBABILITÀ ASSOCIATA AL LIVELLO DI SIGNIFICATIVITÀ DEL 5% CON 10 GDL.

Alla probabilità  $\alpha = 0.05$ , il valore di  $t$  con 10 gdl

- in un test unilaterale è uguale a 1,8125
- un test bilaterale è uguale a 2,228.

E' quindi possibile che con gli stessi dati si possa rifiutare l'ipotesi nulla, se il test è unilaterale, mentre non si possa rifiutarla, se il test è bilaterale.

Per il test  $t$  di Student, il **confronto tra 2 medie** può essere attuato

- sia con **2 campioni dipendenti**
- sia con **2 campioni indipendenti**

E' importante distinguere le due situazioni, che dipendono dal modo con cui le due serie di misure a confronto sono ottenute; infatti i due test si differenziano

- sia nelle procedure di applicazione del test  $t$ ,
- sia nel modo di misurare gli effetti della variabilità dei soggetti considerati.

### 5.5. IL TEST $t$ PER 2 CAMPIONI DIPENDENTI O PER DATI APPAIATI ED INTERVALLO FIDUCIALE DELLA MEDIA DELLE DIFFERENZE

**Caratteristica distintiva del confronto tra 2 campioni dipendenti è poter accoppiare ogni osservazione di un campione con una e una sola osservazione dell'altro campione;** necessariamente, i due gruppi hanno sempre lo stesso numero di dati. Le tecniche sono tre:

- 1 – dati **auto-appaiati**,
- 2 – dati **naturalmente appaiati**,
- 3 – dati **artificialmente appaiati**.

1 - La situazione più semplice è quella dell'**auto-accoppiamento**, in cui ogni soggetto serve come controllo di se stesso; si parla anche di **dati auto-appaiati** e si confrontano i valori presi sugli **stessi**

**soggetti in 2 momenti diversi.** Un esempio classico è la comparazione tra i livelli di inquinamento rilevati nello stesso gruppo di località in due momenti differenti, al fine di verificare se vi siano state in media variazioni significative; oppure prima e dopo interventi di risanamento, al fine di valutare la loro efficacia. In medicina e in ecotossicologia si parla di autoappaiamento dei dati quando gli stessi individui sono misurati prima e dopo l'intervento, appunto per valutarne l'efficacia

2 - Tuttavia questa tecnica non sempre è possibile: se si deve analizzare l'effetto di due sostanze tossiche, l'effetto di due interventi chirurgici su cavie non è possibile somministrare entrambe agli stessi individui, ma si devono formare due gruppi distinti, cercando di fare in modo che gli individui che li compongono siano tra loro simili a coppie. E' la seconda situazione, quando le **osservazioni** nei due gruppi sono **naturalmente appaiate**: **le misure non sono tratte dagli stessi individui, ma sono effettuate su coppie di individui scelti appositamente.**

E' il caso in cui, per esaminare l'effetto di due tossici, si deve avere a disposizione alcune nidiate per scegliere coppie di animali entro ognuna; il primo viene assegnato casualmente ad un gruppo ed il secondo all'altro gruppo. Nella scelta degli individui dalla nidiate e nell'attribuzione al gruppo può essere presente un certo grado di soggettività, che tuttavia è in gran parte eliminata con la scelta casuale, la **randomizzazione**, effettuata mediante l'estrazione di numeri casuali.

Un altro esempio di dati naturalmente appaiati può essere tratto da esperimenti di etologia, con il confronto nella cura dei figli tra il comportamento della madre e quello del padre, ovviamente riferiti agli stessi cuccioli.

3 - La terza situazione è **l'appaiamento artificiale**, un procedimento relativamente frequente nella ricerca ambientale. Si supponga di avere alcuni laghi con un tasso elevato d'inquinamento e di non avere raccolto da essi misure in condizioni normali o standard. Per ogni lago inquinato è possibile cercare un lago non inquinato (relativamente simile per vicinanza fisica, altezza sul livello del mare, configurazione del terreno e per altri vari fattori), che si presume possa essere rappresentativo del corrispondente prima dell'inquinamento. Le differenze nel parametro rilevato tra i tassi di ogni coppia di laghi diviene una stima dell'effetto dell'inquinamento. E' ovvio che ricercatori differenti potrebbero scegliere coppie di laghi diverse, pure agendo nelle stesse condizioni sperimentali.

E' una procedura in cui la componente di soggettività è molto elevata.

**Scopo principale della tecnica di appaiamento dei dati è determinare**

- **il massimo di omogeneità entro ogni coppia e**
- **il massimo di eterogeneità tra le coppie.**

**Il confronto** tra trattamento e controllo, fatto su gli stessi individui o tra situazioni simili, **tenta di eliminare alcune sorgenti di variabilità che potrebbero essere in grado di nascondere le reali differenze tra le due serie di misure: l'obiettivo è di esaminare le differenze fra due misurazioni, dopo aver ridotto l'effetto della variabilità dovuta agli elementi o individui, poiché sono gli stessi che compongono i due gruppi.**

Ad esempio, scegliendo dalla stessa nidata due cavie dello stesso sesso, si elimina l'eventuale differenza degli effetti di una sostanza tossica nei due sessi o in individui di età diversa.

Tecnicamente il confronto tra le medie di 2 serie di osservazioni appaiate è semplice: **l'analisi è applicata ad una nuova serie di dati, quelli risultanti dalle differenze tra gli elementi di ciascuna coppia.**

Per il test t di Student, l'ipotesi nulla  $H_0$  è che la media dell'universo delle differenze sia uguale a 0

$$H_0: \mathbf{d} = 0$$

mentre l'ipotesi alternativa  $H_1$  ha formulazioni differenti, in rapporto al tipo di test scelto (se bilaterale o unilaterale).

- In un **test bilaterale**, l'ipotesi alternativa  $H_1$  afferma che la media delle differenze è diversa da un valore prefissato qualunque, che sovente è 0;  $H_1$  spesso è indicata con

$$H_1: \mathbf{d} \neq 0$$

- In un **test unilaterale**, l'ipotesi alternativa  $H_1$  afferma che la differenza è maggiore (oppure minore) della quantità prefissata (spesso 0); può essere scritta come

$$H_1: \mathbf{d} > 0 \quad \text{oppure} \quad H_1: \mathbf{d} < 0$$

Per la scelta tra l'ipotesi nulla  $H_0$  e l'ipotesi alternativa  $H_1$ , formulata a priori, la significatività della media delle differenze è verificata con il rapporto

$$t_{(n-1)} = \frac{\bar{\mathbf{d}} - \mathbf{d}}{\frac{s_d}{\sqrt{n}}}$$

dove

- $\bar{\mathbf{d}}$  è la media della colonna delle differenze;
- $\mathbf{d}$  è la differenza media attesa: spesso, ma non necessariamente, è uguale a 0;
- $s_d$  è la deviazione standard calcolata sulla colonna delle differenze;
- $n$  è il numero di differenze, corrispondente anche al numero di coppie di dati.

**L'intervallo di confidenza della media delle differenze  $\bar{\mathbf{d}}$  tra i due campioni dipendenti**, entro il quale alla probabilità  $\alpha$  è compresa la media reale  $\mathbf{d}$  delle differenze, con la medesima simbologia della formula precedente, è calcolato mediante

$$d = \bar{d} \pm t_{\frac{\alpha}{2}; n-1} \cdot \frac{s_d}{\sqrt{n}}$$

dove

$t_{\frac{\alpha}{2}; n-1}$  indica il valore della distribuzione **t** con **n-1** gradi di libertà alla probabilità  $\frac{\alpha}{2}$ .

ESEMPIO 1. Un programma di disinquinamento, applicato sui dati di una regione, ha abbassato la presenza di sostanze inquinanti di 30 punti. Su 8 laghi è stata applicata una procedura differente; di questi ultimi sono riportati i valori d'inquinamento prima e dopo l'intervento:

Lago	Prima	Dopo
A	180	140
B	175	145
C	165	140
D	195	160
E	180	150
F	180	145
G	200	160
H	190	145

Ai fini di una valutazione dei risultati è necessario rispondere a due quesiti:

- 1- Il secondo programma di disinquinamento è significativamente migliore del primo?
- 2 - Quale è il reale valore di abbattimento del tasso di inquinamento alla probabilità 0.05?

Risposte.

1 - E' un **test ad una coda**, dove

- l'**ipotesi nulla  $H_0$**  assume che la media  $\bar{d}$  delle differenze tra le coppie di dati non è significativamente maggiore della precedente, che era risultata uguale a +30;

$$H_0: d = 30$$

- l'**ipotesi alternativa  $H_1$**  è unilaterale: afferma che la media delle differenze tra prima e dopo l'intervento negli 8 laghi sia significativamente maggiore di 30, la media della sperimentazione precedente.

$$H_1: d > 30$$

Per l'applicazione del **test t di Student** nel caso di due campioni dipendenti, dapprima si deve calcolare la colonna delle differenze (d) tra coppie di dati campionari (riportati in grassetto nella tabella sottostante).

Per tutti i calcoli successivi si utilizzano solamente questi valori, ignorando i dati originali rilevati prima e dopo gli interventi; in questo modo, si elimina l'effetto di avere laghi che hanno livelli medi (tra prima e dopo) d'inquinamento diversi.

Lago	Prima	Dopo	<b>d</b>
A	180	140	<b>40</b>
B	175	145	<b>30</b>
C	165	140	<b>25</b>
D	195	160	<b>35</b>
E	180	150	<b>30</b>
F	180	145	<b>35</b>
G	200	160	<b>40</b>
H	190	145	<b>45</b>

Dalla colonna delle differenze (**d**), si stimano

- la media  $\bar{d}$
- la devianza  $s_d$

ottenendo i valori:

$$\bar{d} = \frac{280}{8} = 35; \quad s_d = \sqrt{\frac{300}{7}} = 6,55; \quad n = 8;$$

necessari per calcolare un **t** con 7 gdl,

$$t_7 = \frac{35 - 30}{\frac{6,55}{\sqrt{8}}} = 2,16$$

Il risultato è uguale a 2,16.

Nella tavola sinottica, **il valore critico di t**

- con **7 gdl**
- per un **test ad una coda**
- **alla probabilità  $\alpha = 0.05$**

è uguale a **1,895**.

Il valore calcolato (2,16) è superiore a quello tabulato: la probabilità  $\alpha$  che questa differenza tra media osservata e media attesa sia dovuta al caso è inferiore a 0.05. Di conseguenza, si rifiuta l'ipotesi nulla  $H_0$  e si accetta l'ipotesi alternativa  $H_1$ .

Il nuovo metodo di disinquinamento è più efficace del precedente: abbatte i valori d'inquinamento di una quantità significativamente superiore a 30 punti.

2 - Per conoscere la media reale ( $d$ ) di riduzione della presenza di liquami con il secondo esperimento, si deve calcolare l'intervallo fiduciale della media delle differenze associato alla probabilità  $\alpha = 0.05$ .

Applicando alla formula generale

$$d = \bar{d} \pm t_{\frac{\alpha}{2}; n-1} \cdot \frac{s_d}{\sqrt{n}}$$

i dati delle differenze campionarie già stimate

$$\bar{d} = \frac{280}{8} = 35; \quad s_d = \sqrt{\frac{300}{7}} = 6,55; \quad n = 8;$$

ed utilizzando il valore di  $t_{(0,025, 7)} = 2,365$  si ottiene un intervallo

$$\delta = 35 \pm 2,365 \cdot \frac{6,55}{\sqrt{8}} = 35 \pm 5,485$$

in cui il limite inferiore è  $l_1 = 29,515$  ed il limite superiore è  $l_2 = 40,485$ .

La media reale ( $\delta$ ) di abbattimento dei liquami con il secondo processo alla probabilità  $\alpha = 0.05$  varia da 29,515 e 40,485.

Il valore di 30 è compreso in questo intervallo. Quindi, in un **test bilaterale** associato alla probabilità  $\alpha = 0.05$  non si sarebbe rifiutata l'ipotesi nulla.

Precedentemente è stato possibile dimostrare un miglioramento significativo, solo perché il test  $t$  verificava un'ipotesi alternativa  $H_1$  ad una coda.

**ESEMPIO 2.** Si vuole verificare se un conservante per alimentazione umana abbia effetti sui fattori di crescita. A questo scopo, un gruppo di 10 cavie adulte è stato sottoposto a un regime di alimentazione contenente la sostanza da testare.

Ogni soggetto è stato pesato sia prima che dopo la nuova dieta, per misurarne le variazioni.

Nella tabella sottostante sono riportati i pesi in grammi prima e dopo la dieta, per ognuna delle 10 cavie:

Cavia	Prima	Dopo
1	180	190
2	175	170
3	150	175
4	158	164
5	174	185
6	187	184
7	172	185
8	157	168
9	164	180
10	165	173

Si vuole sapere:

- 1 - Se la sostanza può essere la causa di variazioni significative di peso.
- 2 - Quale è la reale variazione ( $d$ ) di peso determinata dal conservante, alla probabilità  $\alpha = 0.05$ .

Risposte.

1 - E' un test bilaterale, in cui (come sempre)

- l'ipotesi nulla  $H_0$  afferma che la media reale ( $\delta$ ) delle differenze ( $d$ ) è uguale a 0

$$H_0: \delta = 0$$

- mentre l'ipotesi alternativa  $H_1$  afferma che è diverso da 0:

$$H_1: \delta \neq 0$$

Dalla colonna delle differenze ( $D$ ) tra le 10 coppie di valori osservati

Cavia	Prima	Dopo	<b>D</b>	$(d - \bar{d})^2$
1	180	190	<b>-10</b>	<b>1</b>
2	175	170	<b>5</b>	<b>196</b>
3	150	175	<b>-25</b>	<b>256</b>
4	158	164	<b>-6</b>	<b>9</b>
5	174	185	<b>-9</b>	<b>0</b>
6	187	184	<b>3</b>	<b>144</b>
7	172	185	<b>-13</b>	<b>16</b>
8	157	168	<b>-11</b>	<b>4</b>
9	164	180	<b>-16</b>	<b>49</b>
10	165	173	<b>-8</b>	<b>1</b>
<b>TOTALE</b>			<b>-90</b>	<b>676</b>

si calcolano

- la somma delle  $D$  che risulta uguale a -90 e
- la somma delle  $(d - \bar{d})^2$  (cioè la devianza) che risulta uguale a 676.

Da esse, si ricavano

- la media ( $\bar{d}$ ),

$$\bar{d} = \frac{-90}{10} = -9$$

- la deviazione standard ( $s_d$ )

$$s_d = \sqrt{\frac{676}{10-1}} = 8,66$$

- il numero di **coppie di osservazioni** ( $n$ ) o di differenze sulle quali sono stati effettuati i calcoli

$$n = 10$$

Infine si stima

$$t_9 = \frac{-9}{\frac{8,66}{\sqrt{10}}} = -3,28$$

un valore di **t** con **9** gdl uguale a **-3,28**.

Per un **test a due code**, il valore critico della distribuzione **t** per 9 gdl e  $\alpha = 0.05$  è uguale a 2,262.

Il valore calcolato (-3,28) in modulo è superiore: la probabilità  $\alpha$  che la differenza riscontrata sia dovuta al caso è minore di 0.05.

Si rifiuta l'ipotesi nulla  $H_0$  e si accetta l'ipotesi alternativa  $H_1$ : la nuova dieta determina nelle cavie una differenza ponderale che è significativa.

2 - Il valore reale ( $d$ ) della differenza è ottenuto mediante la stima dell'**intervallo fiduciale della differenza media**, che **per due campioni dipendenti è dato da**

$$d = \bar{d} \pm t_{\frac{\alpha}{2}; n-1} \cdot \frac{s_d}{\sqrt{n}}$$

Con i dati dell'esercizio,

- per la probabilità  $\alpha = 0.05$  alla quale corrisponde
  - un valore con  $t_{0,025}$  in entrambe le code o  $t_{0,05}$  bilaterale uguale a 2,262
- si ottiene

$$\delta = -9 \pm 2,262 \cdot \frac{7,57}{\sqrt{10}} = 9 \pm 5,42$$

dalla quale risulta un intervallo

$$d_1 = -3,58 \quad \text{e} \quad d_2 = -14,42$$

La nuova dieta alla quale sono state sottoposte le cavie determina una riduzione media ( $\delta$ ) di peso uguale a 9; alla probabilità  $\alpha = 0.05$  essa ha come valori estremi - 3,58 e -14,42. Si può osservare che il valore di **d** espresso nell'ipotesi nulla ( $H_0: d = 0$ ) è escluso dall'intervallo fiduciale calcolato; quindi si discosta dal valore medio sperimentale in modo significativo alla probabilità prescelta.

**In un test a due code, alla medesima probabilità con cui nel test t per dati appaiati la differenza media risulta significativa, il valore dell'ipotesi nulla è escluso dall'intervallo fiduciale.**

Per quanto riguarda la significatività, i due approcci forniscono la medesima informazione.

**Secondo vari autori di testi di statistica applicata, quando il numero di coppie di dati appaiati è superiore a 40 (altri indicano 50 o addirittura 150), la distribuzione t di Student è approssimata in modo sufficiente dalla distribuzione normale standardizzata z.** Confrontando i valori riportati nelle tabelle, si può infatti osservare che alla probabilità  $\alpha = 0.05$  il valore critico di t (uguale rispettivamente a 2,02 per 40 gdl e 2,01 per 50 gdl) si discosta di una quantità trascurabile (circa il 2% in percentuale) dal corrispondente valore di z (uguale a 1,96).

**Nella formula per il calcolo della varianza, con una differenza di risultati che da molti ricercatori è ritenuta trascurabile, quando n è di alcune decine è possibile sostituire il valore critico di t con il valore di z associato alla probabilità  $\alpha$  prefissata, sia in test ad una coda che in test a due code.**

Di conseguenza,

- il test z per la significatività della media diventa

$$z = \frac{\bar{d} - d}{\frac{s_d}{\sqrt{n}}}$$

- l'intervallo fiduciale è stimato con

$$d = \bar{d} \pm z_{\frac{\alpha}{2}} \cdot \frac{s_d}{\sqrt{n}}$$

## 5.6. IL TEST t PER 2 CAMPIONI INDIPENDENTI O PER DATI NON APPAIATI

In molti casi, non è tecnicamente possibile formare due campioni dipendenti. Spesso non è possibile misurare gli effetti di due differenti trattamenti sugli stessi individui: è il caso in cui si confrontano misure di accrescimento somatico in animali o piante sottoposte a condizioni ambientali differenti oppure si confrontano livelli d'inquinamento idrico tra due fiumi differenti con rilevazioni in varie stazioni. Altre volte, non è possibile nemmeno fare appaiamenti naturali o artificiali, perché le situazioni non si ripetono a coppie nelle medesime condizioni e la dose di soggettività è ritenuta eccessiva.

L'unica possibile strategia di analisi dei dati è quella di confrontare **due campioni indipendenti**, due campioni formati da individui differenti. **Aumenta la variabilità tra i due gruppi:** nel caso di cavie, in uno possono essere presenti più maschi o più femmine, più individui anziani o più giovani, più sani o ammalati, per cui la variabilità delle risposte aumenta. Ma si ottengono due vantaggi:

- **poter utilizzare un numero differente di osservazioni,**
- **avere dati che più facilmente sono espressione della variabilità casuale,**
- **utilizzare per il confronto con il proprio un campione raccolto da altri.**

Con il test di significatività per due campioni indipendenti, viene verificata la **stessa ipotesi del caso di dati appaiati, seppure espressa in forma diversa.** E' infatti fondamentale comprendere che

- per due campioni dipendenti i calcoli vengono effettuati sulla sola colonna delle differenze, mentre
- nel caso di due campioni indipendenti **i calcoli vengono effettuati sulle due serie di osservazioni.**

L'**ipotesi nulla  $H_0$**  è che i due campioni (indicati con A e B) siano estratti dalla stessa popolazione oppure da due popolazioni differenti ma con media ( $\mu$ ) uguale; tale ipotesi può essere scritta come

$$H_0: \mu_A = \mu_B \quad \text{oppure} \quad H_0: \mu_A - \mu_B = 0$$

L'**ipotesi alternativa  $H_1$**  può essere riferita

- sia a **un test a due code** ed essere scritta come

$$H_1: \mu_A \neq \mu_B \quad \text{oppure} \quad H_1: \mu_A - \mu_B \neq 0$$

- sia a **un test ad una coda** e scritta come

$$H_1: \mu_A > \mu_B \quad \text{oppure} \quad H_1: \mu_A - \mu_B > 0$$

quando in una direzione e

$$H_1: \mu_A < \mu_B \quad \text{oppure} \quad H_1: \mu_A - \mu_B < 0$$

quando nell'altra direzione.

Nel caso di 2 campioni indipendenti, i gradi di libertà del **t** sono uguali a  $(n_A - 1) + (n_B - 1)$ , che possono anche essere scritti come  $(n_A + n_B - 2)$  oppure  $(N - 2)$ .

In valore del **t** è ottenuto mediante

$$t_{(n_A+n_B-2)} = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{s_p^2 \cdot \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

dove

- $\bar{x}_A$  e  $\bar{x}_B$  sono le medie rispettivamente del campione **A** e del campione **B**,
- $\mu_A$  e  $\mu_B$  sono le rispettive medie attese, espresse nell'ipotesi nulla,
- $n_A$  e  $n_B$  sono il numero di osservazioni nei campioni **A** e **B**,
- $s_p^2$  è la varianza associata (*pooled*) dei due gruppi a confronto,
- $N = n_A + n_B$

**La varianza associata o varianza pooled ( $s_p^2$ )** è data dal rapporto tra la somma delle due devianze e la somma dei rispettivi gradi di libertà

$$s_p^2 = \frac{\sum_{i=1}^{n_A} (x_{Ai} - \bar{x}_A)^2 + \sum_{i=1}^{n_B} (x_{Bi} - \bar{x}_B)^2}{n_A - 1 + n_B - 1}$$

dove

- $x_{Ai}$  e  $\bar{x}_A$  sono nell'ordine i dati e la media del gruppo **A**,
- $x_{Bi}$  e  $\bar{x}_B$  sono rispettivamente i dati e la media del gruppo **B**,
- $n_A$  e  $n_B$  sono il numero di osservazioni nei campioni **A** e **B**.

La varianza pooled è **una varianza media ponderata, calcolata sempre a partire dalle due devianze e dai loro gdl**, che attribuisce una importanza proporzionalmente maggiore al gruppo che ha un numero maggiore di dati.

Nella ricerca ambientale ed ecologica in genere, ricorre con relativa frequenza il caso in cui si debbono **confrontare i risultati di due ricercatori diversi** (indicati con **A** e **B**), ognuno dei quali ha pubblicato solo i tre valori fondamentali dei suoi dati:

- la media campionaria  $\bar{x}$ ,
- una misura della variabilità, quali varianza ( $s^2$ ) o deviazione standard ( $s$ ) o errore standard ( $es$ ),
- numero di dati raccolti ( $n$ ).

Per verificare se tra le due medie campionarie esiste differenza, è possibile utilizzare il test **t con la solita formula**

$$t_{(n_A+n_B-2)} = \frac{(\bar{x}_A - \bar{x}_B)}{\sqrt{s_p^2 \cdot \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

dopo aver calcolato la **varianza pooled** ( $s_p^2$ ).

Come risulta dalla formula precedente,

$$s_p^2 = \frac{\sum_{i=1}^{n_A} (x_{Ai} - \bar{x}_A)^2 + \sum_{i=1}^{n_B} (x_{Bi} - \bar{x}_B)^2}{n_A - 1 + n_B - 1}$$

occorre **ritornare alla devianza**

$$\sum_{i=1}^n (x_i - \bar{x})^2$$

di ognuno dei due gruppi.

A questo scopo è utile ricordare che

- se si dispone della varianza  $s^2$  oppure delle deviazione standard ( $s$ ) si ottiene la devianza attraverso la relazione

$$\sum_{i=1}^n (x_i - \bar{x})^2 = s^2 \cdot (n - 1)$$

- mentre se si dispone dell'errore standard ( $es$ )

si ottiene la devianza attraverso la relazione

$$\sum_{i=1}^n (x_i - \bar{x})^2 = (es \cdot \sqrt{n})^2 \cdot (n-1)$$

ESEMPIO. Si intende verificare due gruppi del Cladocero *Daphnia magna* (del quale sono stati misurati 5 individui per il campione 1 e 7 individui per il campione 2, come riportato nella tabella sottostante) dopo 20 giorni dalla schiusa delle uova hanno raggiunto dimensioni medie significativamente differenti.

$X_A$	$X_B$
4,290	3,120
3,900	3,112
3,783	3,120
3,900	3,847
4,095	3,081
---	3,042
---	3,742

Risposta

E' un **test bilaterale**, in cui

$$H_0: \mu_A = \mu_B \quad e \quad H_1: \mu_A \neq \mu_B$$

Dopo aver calcolato

- la media del gruppo  $X_A$

$$\bar{x}_A = \frac{4,29 + 3,9 + 3,783 + 3,9 + 4,095}{5} = \frac{19,968}{5} = 3,994$$

- la media del gruppo  $X_B$

$$\bar{x}_B = \frac{3,12 + 3,112 + 3,12 + 3,847 + 3,081 + 3,042 + 3,742}{7} = \frac{23,064}{7} = 3,295$$

- la devianza del gruppo A ( $SQ_A$ )

$$SQ_A = (4,29 - 3,994)^2 + (3,9 - 3,994)^2 + (3,783 - 3,994)^2 + (3,9 - 3,994)^2 + (4,095 - 3,994)^2$$

$$SQ_A = 0,0876 + 0,0088 + 0,0441 + 0,0088 + 0,0102 = 0,1595$$

- la devianza del gruppo B ( $SQ_B$ )

$$SQ_B = (3,12 - 3,295)^2 + (3,112 - 3,295)^2 + (3,12 - 3,295)^2 + (3,847 - 3,295)^2 + \\ + (3,081 - 3,295)^2 + (3,042 - 3,295)^2 + (3,642 - 3,295)^2 \\ SQ_B = 0,0306 + 0,0335 + 0,0306 + 0,3047 + 0,0458 + 0,0590 + 0,1204 = 0,6246$$

si stima

- la varianza associata ( $s_p^2$ )

$$s_p^2 = \frac{0,1595 + 0,6246}{4 + 6} = \frac{0,7841}{10} = 0,0784$$

Con  $n_A = 5$  e  $n_B = 7$

si calcola il valore di **t** con **10** gdl

$$t_{(4+6)} = \frac{3,994 - 3,295}{\sqrt{0,0784 \cdot \left(\frac{1}{5} + \frac{1}{7}\right)}} = \frac{0,669}{\sqrt{0,0784 \cdot 0,3429}} = \frac{0,669}{0,164} = 4,26$$

che risulta uguale a 4,26.

Poiché il valore di **t** con **10 gdl** per un **test bilaterale** (vedi tabella relativa alla fine del capitolo)

- alla probabilità  $\alpha = 0,05$  è uguale a **2,228**
- alla probabilità  $\alpha = 0,01$  è uguale a **3,169**
- alla probabilità  $\alpha = 0,001$  è uguale a **4,587**

si rifiuta l'ipotesi nulla, con probabilità inferiore a 0.01 di commettere un errore di I Tipo (rifiutare l'ipotesi nulla quando essa è vera). Pertanto, si deve concludere che la differenza tra le medie dei due campioni risulta molto significativa.

Per ognuno dei due campioni, i dati possono essere riassunti in una tabella come la seguente

Campione	A	B
Dimensione (n)	5	7
Media ( $\bar{X}$ )	3,994	3,295
Varianza ( $s^2$ )	0,0399	0,1041
Deviazione standard (s)	0,1997	0,3226
Errore standard (es)	0,0893	0,1220

In realtà, oltre al numero di dati raccolti e alla media, è sufficiente riportare una sola coppia di valori delle ultime 3 (varianza, deviazione standard, errore standard), potendo da una qualsiasi di esse ricavare le altre due, quando appunto si conoscano le dimensioni del campione (**n**) oppure i suoi gdl (**n-1**).

Per valutare la significatività della differenza tra le due medie, sempre con un **test bilaterale** in cui

$$H_0: \mu_A = \mu_B \quad e \quad H_1: \mu_A \neq \mu_B$$

prima di tutto è necessario ricavare la devianza (SQ) di ognuno dei due campioni:

- dalla varianza  $s^2$

$$SQ_A = 0,0399 \cdot 4 = 0,1596$$

$$SQ_B = 0,1041 \cdot 6 = 0,6246$$

- dalla deviazione standard  $s$

$$SQ_A = 0,1997^2 \cdot 4 = 0,1595$$

$$SQ_B = 0,3225^2 \cdot 6 = 0,6240$$

- dall'errore standard  $es$

$$SQ_A = (0,0893 \cdot \sqrt{5})^2 \cdot 4 = 0,1595$$

$$SQ_B = (0,1220 \cdot \sqrt{7})^2 \cdot 6 = 0,6248$$

si perviene sempre alle stesse stime delle due devianze ( $SQ_A = 0,1595$  e  $SQ_B = 0,6246$ ) a meno delle approssimazioni utilizzate nei calcoli.

Da esse si ricava la varianza associata

$$s_p^2 = \frac{0,1595 + 0,6246}{4 + 6} = \frac{0,7841}{10} = 0,0784$$

e quindi il valore del  $t$

$$t_{(4+6)} = \frac{3,994 - 3,295}{\sqrt{0,0784 \cdot \left(\frac{1}{5} + \frac{1}{7}\right)}} = \frac{0,669}{\sqrt{0,0784 \cdot 0,3429}} = \frac{0,669}{0,164} = 4,26$$

che ovviamente è uguale a quello calcolato in precedenza, utilizzando i singoli dati.

## 5.7 TEST F, TEST DI LEVENE E TEST DI BARTLETT PER IPOTESI BILATERALI E UNILATERALI SULL'UGUAGLIANZA DI DUE VARIANZE

Il **t di Student** è un test di statistica parametrica. Affinché possa essere ritenuto valido, come nel caso di un campione, devono essere rispettate le condizioni essenziali che

- **i dati (o gli scarti rispetto alla media) siano distribuiti normalmente,**
- **le osservazioni siano raccolte in modo indipendente.**

Con **due campioni indipendenti**, per calcolare la  $s^2$  pooled si ha l'ulteriore condizione essenziale, più importante delle precedenti, perché rispetto ad essa **il test t è meno robusto**, di

- **omoschedasticità o omoscedasticità, cioè che le due varianze siano statisticamente uguali.**

I termini *homoscedasticity* e *heteroscedasticity*, citati pure come *homoskedasticity* e *eteroskedasticity*, secondo **Walker M. H.** (vedi, del 1929: *Studies in History of Statistical Methods*. Williams and Wilkins, Baltimore, Maryland, 229 pp.) furono introdotti da K. Pearson nel 1905.

Calcolare **una varianza comune**, la  $s^2$  **pooled**, richiede che i due gruppi abbiano **varianze statisticamente simili**.

L'ipotesi di raccolta indipendente dei dati dipende dalla programmazione dell'esperimento. L'ipotesi di normalità dei dati o degli errori (gli scarti dei dati dalla loro media) può essere violata senza gravi effetti sulla potenza del test, a meno di un grave asimmetria; ma

**l'eguaglianza delle varianze dei due campioni indipendenti dovrebbe essere sempre dimostrata.**

Nella prassi statistica, fino a poco anni fa il test **t di Student** su due campioni veniva sempre accettato, se non era più che evidente la loro non omoschedasticità. Ora si è maggiormente severi con questa condizione e si imputa al ricercatore l'obbligo di dimostrare che le due varianze sono simili. Infatti la varianza associata ( $s^2_p$ ) è una quantità fondamentale per il calcolo del **t** e ha significato solamente se è rappresentativa delle varianze di ogni gruppo.

Per l'applicazione del test t, la omoschedasticità tra due gruppi (**A** e **B**) è verificata con un **test bilaterale**, dove l'ipotesi nulla  $H_0$  e l'ipotesi alternativa  $H_1$  sono

$$H_0: s_A^2 = s_B^2 \quad H_1: s_A^2 \neq s_B^2$$

I **test parametrici** proposti in letteratura per verificare l'omoschedasticità bilaterale o unilaterale sono sostanzialmente due:

A - il **test F** o del rapporto tra le due varianze,

B - il **test di Levene**

C - il **test di Bartlett**.

A) Il **test F bilaterale**, il primo ad essere proposto e tuttora il più diffuso, è fondato sul **rapporto tra la varianza campionaria ( $s^2$ ) maggiore e la varianza campionaria minore:**

$$F_{(n_{\max}-1); (n_{\min}-1)} = \frac{s_{\max}^2}{s_{\min}^2}$$

dove

- $s_{\max}^2$  è la varianza maggiore,
- $s_{\min}^2$  è la varianza minore,
- $n_{\max}$  è il numero di dati nel gruppo con varianza maggiore,
- $n_{\min}$  è il numero di dati nel gruppo con varianza minore.

Fondato sull'ipotesi che le due varianze siano uguali (cioè che l'ipotesi nulla  $H_0$  sia vera), il risultato del rapporto tra esse dovrebbe essere uguale a 1. Ovviamente è ammessa una certa tolleranza, poiché la stima delle due varianze non è mai esatta. Solamente quando sono calcolate sulle due popolazioni a confronto le varianze sono quelle reali; si ottengono le due  $s^2$  reali e per valutare se sono differenti è sufficiente il semplice confronto, non sussistendo la necessità di ricorrere all'inferenza statistica.

Nella ricerca ambientale, spesso le varianze sono stimate su campioni di dimensioni molto piccole, formati da poche unità di osservazioni. Di conseguenza, il rapporto tra le due varianze è una stima campionaria, che potrebbe variare

- **da uno a infinito** oppure
- **da uno a zero**.

Per non utilizzare entrambe queste misure, che darebbero una informazione ridondante, è stata scelta la distribuzione dei valori che è più sensibile alle variazioni: quella da uno a infinito.

Il valore ottenuto dal rapporto deve essere confrontato con una tabella di valori critici; in essa è riportata la probabilità di trovare per caso rapporti uguali o maggiori di quello calcolato, nella condizione che l'ipotesi nulla sia vera.

**I valori critici della distribuzione F** (che è spiegata nel capitolo dell'analisi della varianza e le cui tabelle sono riportate alla fine di quel capitolo) dipendono dai **gradi di libertà**

- del **numeratore**, riportati nella prima riga della tabella, e
- del **denominatore**, riportati nella prima colonna.

Solo se **si dimostra che l'ipotesi nulla è vera** e pertanto che i due gruppi hanno varianze statisticamente uguali, è possibile usare il **test t di Student per 2 campioni indipendenti**.

ESEMPIO. Verificare la omogeneità delle due varianze dell'esercizio precedente. Nella tabella sono riportati i dati essenziali per il test **t** e per la verifica della omoschedasticità.

Campione	A	B
Dimensione (n)	5	7
Media ( $\bar{X}$ )	3,994	3,295
Varianza ( $s^2$ )	0,0399	0,1041

Risposta

La varianza del gruppo A, calcolata su 5 dati, è risultata uguale a 0,0399.

La varianza del gruppo B, calcolata su 7 dati, è risultata uguale a 0,1041

Per verificare l'ipotesi nulla

$$H_0: \sigma_A^2 = \sigma_B^2$$

con ipotesi alternativa bilaterale

$$H_1: \sigma_A^2 \neq \sigma_B^2$$

si deve applicare il test F

$$F_{(n_{\max}-1); (n_{\min}-1)} = \frac{s_{\max}^2}{s_{\min}^2}$$

Con i dati dell'esempio,

$$F_{(6,4)} = \frac{0,1041}{0,0399} = 2,61$$

risulta un valore di **F**, con gdl **6 e 4**, uguale a **2,61**.

Per ottenere la probabilità  $\alpha$  di trovare per caso questa risposta o un rapporto ancora maggiore, aspettandoci **1**, si deve ricorrere alle tabelle sinottiche dei valori **F di Fisher-Snedecor**, riportate nel capitolo 8 dove è presentata l'analisi della varianza.

Nella tabella per la probabilità  $\alpha = 0.05$ , all'incrocio tra

- **gdl 6 nella prima riga o del numeratore e**
- **gdl 4 nella prima colonna o del denominatore,**

si trova il valore critico **6,16**.

Il valore calcolato con i dati dei due campioni è molto minore di quello critico, riportato nella tabella.

Di conseguenza, si accetta l'ipotesi nulla: si può affermare che le due varianze sono statisticamente simili.

Il **test F** può essere utilizzato anche per verificare, con un **test unilaterale**, se una varianza è significativamente minore oppure maggiore di un'altra. Quando si confrontano due reagenti o due strumenti con prove ripetute sullo stesso campione, è migliore quello che ha una varianza minore. In genetica, gli individui che presentano variabilità minore probabilmente hanno un patrimonio genetico più simile oppure vivono in condizioni ambientali più omogenee. Nelle misure d'inquinamento, una variabilità maggiore porta più facilmente a superare i limiti di legge, se i valori medi sono simili.

In un **test unilaterale sulle varianze**,

- se si vuole verificare che la **varianza del gruppo A sia minore di quella del gruppo B**, si formula l'ipotesi nulla

$$H_0: s_A^2 \geq s_B^2$$

contro l'ipotesi alternativa

$$H_1: s_A^2 < s_B^2$$

- mentre se si pensa che la **varianza del campione A sia maggiore di quella del campione B**

l'ipotesi nulla è

$$H_0: \mathbf{s}_A^2 \leq \mathbf{s}_B^2$$

e l'ipotesi alternativa

$$H_1: \mathbf{s}_A^2 > \mathbf{s}_B^2$$

Per la procedura, il test F unilaterale si distingue da quello bilaterale perché

**la varianza che si ipotizza maggiore (in  $H_1$ ) deve sempre essere posta al numeratore.**

Il motivo è facilmente comprensibile:

- per la significatività, si utilizza la stessa distribuzione F di Fisher,
- in cui il valore è sempre superiore a 1.

Ovviamente, se i dati sperimentali dessero risultati opposti a quanto ipotizzato in  $H_1$ , il test diventa inutile:

- se  $s_A^2$  risultasse maggiore di  $s_B^2$
- non sarà mai possibile dimostrare  $H_1$  cioè  $\mathbf{s}_A^2 < \mathbf{s}_B^2$
- e il test è inapplicabile o privo di senso

Il test ha significato solamente se  $s_A^2$  risulta minore di  $s_B^2$  e serve per verificare se la differenza tra esse è trascurabile oppure troppo grande per poter essere ritenuta casuale.

E' **importante** ricordare che nell'uso dell'inferenza per l'omoschedasticità, quindi per concludere che due varianze sono simili e quindi applicare il test t di Student sulle medie, occorre molta cautela. I **test statistici sono generalmente impostati per non rifiutare l'ipotesi nulla**, a meno di evidenze del suo contrario: **si rifiuta l'ipotesi nulla, solo quando si dimostra che la probabilità di trovare per caso differenze grandi come quella riscontrata o maggiori di essa è molto piccola**. Se il ricercatore dispone di pochi dati, le differenze tra i due gruppi devono essere molto grandi, perché il test risulti significativo: **un test con pochi dati è poco potente, cioè ha probabilità molto basse di rifiutare l'ipotesi nulla, anche quando è noto che essa è falsa**.

Ne consegue che, **quando non si rifiuta l'ipotesi nulla  $H_0$ , non si può affermare che essa è vera**; ma solamente che non si è in grado, per la scarsità delle informazioni raccolte, di sostenere che essa è falsa.

In termini operativi, è possibile uscire da questo uso illogico dei test per l'omoschedasticità, con **una valutazione più completa e dettagliata della probabilità a stimata**.

Se, con pochi dati, la probabilità  $\alpha$  è di poco superiore a **0.05** (ad esempio 0.10 oppure 0.15) non si può ovviamente rifiutare l'ipotesi nulla. Si parla quindi di **significatività tendenziale**: se fosse stato possibile disporre di un numero maggiore di dati, con elevata probabilità si sarebbe potuto rifiutare l'ipotesi nulla.

In pratica, secondo vari autori di testi di statistica, si dovrebbe accettare l'ipotesi nulla solo quando la probabilità  $\alpha$  calcolata è alta, superiore al 30 per cento; con un numero di dati maggiore, questa probabilità può essere abbassata al 20 per cento. Purtroppo, per definire un campione grande o piccolo e per decidere a quale livello di probabilità  $\alpha$  si possa affermare che l'ipotesi nulla è vera, non esistono regole precise. La scelta è fondata sul buon senso statistico, che può derivare solo dall'esperienza.

Con la disponibilità di tabelle che forniscono i valori critici solo per quattro o cinque valori di  $\alpha$ , da **0.05 a 0.001**, è difficile fare queste valutazioni più complete. Recentemente, con l'uso di programmi informatici, è possibile ottenere dai computer, che dispongono di memoria molto ampia, una stima più precisa, meno approssimata, della probabilità: quali  $\alpha = 0.42$  oppure  $\alpha = 0.085$  oppure  $\alpha = 0.64$ . Solo con il primo ( $\alpha = 0.42$ ) ed ancor più con il terzo risultato ( $\alpha = 0.64$ ) è possibile affermare che le due varianze sono sostanzialmente simili; con il secondo ( $\alpha = 0.085$ ) si dovrebbe parlare di tendenziale significatività della differenza tra le due varianze.

Nel caso dell'esempio, è possibile **ragionare sul valore di F** ottenuto: rispetto ad un valore critico di 6,16 per la probabilità  $\alpha = 0.05$ , è stato calcolato un valore molto minore ( $F = 2,61$ ); seppure non stimata con precisione, facilmente la probabilità  $\alpha$  ad esso associata è alta (forse intorno al 50% o 0.5). Di conseguenza, si può affermare, senza eccessivi timori di essere contestati, che le due varianze sono tendenzialmente simili.

B) Il **test di Levene** è un **metodo alternativo**; può essere usato anche per integrare l'analisi condotta con il test F, quando si voglia una valutazione più approfondita sull'omogeneità di due varianze.

Ritenuto da vari statistici **più robusto del test F** rispetto alla non normalità della distribuzione, deve la sua diffusione soprattutto al suo inserimento in alcuni pacchetti statistici. Per la sua applicazione è necessario disporre dei dati originari, in quanto utilizza gli scarti di ogni valore campionario dalla media del suo gruppo.

Esso mette a confronto gli scarti dei due gruppi, cioè le differenze ( $d_i$ ) rispetto alle medie di gruppi,

$$d_i = x_i - \bar{x}$$

dopo

- averle elevate al quadrato

$$d_i = (x_i - \bar{x})^2$$

- oppure prese in valore assoluto

$$d_i = |x_i - \bar{x}|$$

per **eliminare i segni negativi**.

**I due metodi forniscono risultati differenti.**

Il primo metodo è suggerito da alcuni testi e viene presentato come la proposta originaria di Levene (vedi: *The statistical SLEUTH: A Course in Methods of Data Analysis* di Fred L. Ramsey e Daniel W. Schafer, Duxbury Press, 1997, a pag. 99 per 2 campioni e a pag. 133 per k campioni).

Il secondo è quello ora più utilizzato, in particolare nei programmi informatici; ha test analoghi in statistica non parametrica, presentati nei capitoli relativi.

Per confrontare la varianza di due gruppi (A e B),

con ipotesi nulla

$$H_0: \sigma_A^2 = \sigma_B^2$$

ed ipotesi alternativa bilaterale

$$H_1: \sigma_A^2 \neq \sigma_B^2$$

la proposta di **Levene consiste nell'applicare alla due serie di scarti** (al quadrato o in valore assoluto) **il test t di Student**, nell'assunzione che, **se i loro valori medi risultano significativamente diversi, le due varianze dei dati originali sono diverse.**

Se, utilizzando **gli scarti dalla media**, si rifiuta l'ipotesi nulla

$$H_0: \mu_A = \mu_B$$

per accettare l'ipotesi alternativa

$$H_1: \mu_A \neq \mu_B$$

implicitamente deriva che

sui **dati originali** si rifiuta l'ipotesi nulla

$$H_0: \sigma_A^2 = \sigma_B^2$$

per accettare l'ipotesi alternativa

$$H_1: \sigma_A^2 \neq \sigma_B^2$$

ESEMPIO. Utilizzando le due serie di dati (A e B), sulle quali in precedenza è stato applicato il test t di Student per verificare la significatività della differenza tra le due medie e il test F per la omoschedasticità, verificare con il test di Levene se hanno varianze significativamente diverse.

$X_A$	$X_B$
4,290	3,120
3,900	3,112
3,783	3,120
3,900	3,847
4,095	3,081
---	3,042
---	3,742

Risposta

Per verificare sui dati l'ipotesi nulla

$$H_0: \sigma_A^2 = \sigma_B^2$$

con ipotesi alternativa bilaterale

$$H_1: \sigma_A^2 \neq \sigma_B^2$$

è possibile utilizzare

- sia gli scarti al quadrato

$$d_i = (x_i - \bar{x})^2$$

- sia gli scarti in valore assoluto

$$d_i = |x_i - \bar{x}|$$

come riportati nella tabella successiva, ricordando che nei dati originari

$$\bar{x}_A = 3,994 \quad \text{e} \quad \bar{x}_B = 3,295$$

Campione A			Campione B		
$X_{Ai}$	$(x_{Ai} - \bar{x}_A)^2$	$ x_{Ai} - \bar{x}_A $	$X_{Bi}$	$(x_{Bi} - \bar{x}_B)^2$	$ x_{Bi} - \bar{x}_B $
4,290	0,0876	0,296	3,120	0,0306	0,175
3,900	0,0088	0,094	3,112	0,0335	0,183
3,783	0,0445	0,211	3,120	0,0306	0,175
3,900	0,0088	0,094	3,847	0,3047	0,552
4,095	0,0102	0,101	3,081	0,0458	0,214
---	---	---	3,042	0,0640	0,253
---	---	---	3,742	0,1998	0,447

Applicando il test alle **differenze al quadrato**  $(x_i - \bar{x})^2$ , si devono stimare

- le dimensioni dei due campioni

$$n_A = 5 \quad \text{e} \quad n_B = 7$$

- le medie

$$\bar{x}_A = \frac{0,0876 + 0,0088 + 0,0445 + 0,0088 + 0,0102}{5} = \frac{0,1599}{5} = 0,03198$$

$$\bar{x}_B = \frac{0,0306 + 0,0335 + 0,0306 + 0,03047 + 0,0458 + 0,0640 + 0,1998}{7} = \frac{0,709}{7} = 0,10129$$

- le devianze (ottenute in modo più rapido con la formula abbreviata)

$$SQ_A = (0,0876^2 + 0,0088^2 + 0,0445^2 + 0,0088^2 + 0,0102^2) - \frac{0,1599^2}{5}$$

$$SQ_A = (0,00767 + 0,00008 + 0,00198 + 0,00008 + 0,0001) - 0,00486$$

$$SQ_A = 0,00911 - 0,00486 = 0,00425$$

$$SQ_B = (0,0306^2 + 0,0335^2 + 0,0306^2 + 0,3047^2 + 0,0458^2 + 0,064^2 + 0,1998^2) - \frac{0,709^2}{7}$$

$$SQ_B = (0,00094 + 0,00112 + 0,00094 + 0,09284 + 0,0021 + 0,0041 + 0,03992) - 0,07181$$

$$SQ_B = 0,14196 - 0,07181 = 0,07015$$

Da esse si calcolano dapprima la varianza associata

$$s_p^2 = \frac{0,00425 + 0,07015}{4 + 6} = \frac{0,0744}{10} = 0,00744$$

ed infine il valore di  $t$

$$t_{(10)} = \frac{0,03198 - 0,10129}{\sqrt{0,00744 \cdot \left(\frac{1}{5} + \frac{1}{7}\right)}} = \frac{-0,06931}{\sqrt{0,00744 \cdot 0,34286}} = \frac{-0,06931}{0,0505} = -1,3724$$

che risulta uguale a  $-1,3724$  con 10 gradi di libertà.

Nella tabella dei valori critici, con 10 gdl e per un test bilaterale,

- alla probabilità  $\alpha = 0.05$  corrisponde un valore di  $t$  uguale a **2,228**
- alla probabilità  $\alpha = 0.10$  corrisponde un valore di  $t$  uguale a **1,812**
- alla probabilità  $\alpha = 0.20$  corrisponde un valore di  $t$  uguale a **1,372**
- alla probabilità  $\alpha = 0.40$  corrisponde un valore di  $t$  uguale a **0,879**

Il valore di  $t$  calcolato corrisponde ad una probabilità  $\alpha$  del 20%

Applicando il test alle **differenze in valore assoluto**  $|x_i - \bar{x}|$ , si devono stimare

- le dimensioni dei due campioni

$$n_A = 5 \quad \text{e} \quad n_B = 7$$

- le medie

$$\bar{x}_A = \frac{0,296 + 0,094 + 0,211 + 0,094 + 0,101}{5} = \frac{0,796}{5} = 0,1592$$

$$\bar{x}_B = \frac{0,175 + 0,183 + 0,175 + 0,552 + 0,214 + 0,253 + 0,447}{7} = \frac{1,999}{7} = 0,2856$$

- le devianze

$$SQ_A = (0,296^2 + 0,094^2 + 0,211^2 + 0,094^2 + 0,101^2) - \frac{0,796^2}{5}$$

$$SQ_A = (0,0876 + 0,0088 + 0,0445 + 0,0088 + 0,0102) - 0,1267$$

$$SQ_A = 0,1599 - 0,1267 = 0,0332$$

$$SQ_B = (0,175^2 + 0,183^2 + 0,175^2 + 0,552^2 + 0,214^2 + 0,253^2 + 0,447^2) - \frac{1,999^2}{7}$$

$$SQ_B = (0,0306 + 0,0335 + 0,0306 + 0,3047 + 0,0458 + 0,064 + 0,1998) - 0,5709$$

$$SQ_B = 0,709 - 0,5709 = 0,1381$$

Da esse si calcolano dapprima la varianza associata

$$s_p^2 = \frac{0,0332 + 0,1381}{4 + 6} = \frac{0,1713}{10} = 0,01713$$

ed infine il valore di  $t$

$$t_{(10)} = \frac{0,1592 - 0,2856}{\sqrt{0,01713 \cdot \left(\frac{1}{5} + \frac{1}{7}\right)}} = \frac{-0,1264}{\sqrt{0,01713 \cdot 0,34286}} = \frac{-0,1264}{0,0706} = -1,65$$

che risulta uguale a  $-1,65$  con 10 gradi di libertà.

Riprendendo i valori critici appena riportati,  $t = -1,65$  in modulo corrisponde ad una probabilità compresa tra  $\alpha = 0.10$  e  $\alpha = 0.20$ . Con un numero così ridotto di dati, con questa probabilità non è certamente possibile rifiutare l'ipotesi nulla, con sufficiente sicurezza.

Anche il **test di Levene** può servire per verificare l'ipotesi **unilaterale**, cioè

- l'ipotesi nulla

$$H_0: \mathbf{s}_A^2 \geq \mathbf{s}_B^2$$

contro l'ipotesi alternativa

$$H_1: \mathbf{s}_A^2 < \mathbf{s}_B^2$$

- oppure l'ipotesi nulla

$$H_0: \mathbf{s}_A^2 \leq \mathbf{s}_B^2$$

contro l'ipotesi alternativa

$$H_1: s_A^2 > s_B^2$$

**E' sufficiente applicare il test t unilaterale sulle medie dei residui**

C) Il **test di Bartlett** ricorre alla distribuzione  $\chi^2$ ; per questo motivo è classificato tra i test non parametrici. Il suo uso più generale è per più campioni. Per questo motivo, anche se a volte è utilizzato per il confronto tra le varianze di due campioni indipendenti, la sua illustrazione è stata collocata tra i metodi per k campioni, nel capitolo dell'analisi della varianza ad un criterio (vedi indice).

Per verificare l'ipotesi di omoschedasticità tra due campioni indipendenti, si utilizza la tavola sinottica del  $\chi^2_{(1)}$  con **1 grado di libertà**.

## 5.8 SIGNIFICATIVITA' E INTERVALLO FIDUCIALE DI UNA DIFFERENZA

**Se le varianze risultano statisticamente differenti, si può utilizzare il test t solo dopo adeguata trasformazione dei dati originali** (che sarà discussa alla fine dei capitoli sull'analisi della varianza).

**Con la trasformazione dei dati, le medie devono rimanere differenti mentre le varianze devono diventare uguali. Se le varianze dei due gruppi a confronto restano significativamente differenti, si deve ricorrere a un test di statistica non parametrica per 2 campioni indipendenti.**

Solo se le varianze dei due gruppi risultano statisticamente omogenee, eventualmente dopo trasformazione, sono corretti sia l'uso test t sia il calcolo dell'**intervallo fiduciale della differenza tra le 2 medie campionarie**  $(\bar{x}_1 - \bar{x}_2)$ .

Il metodo per il calcolo dell'intervallo fiduciale della differenza tra le due medie è simile al test **t di Student** per due campioni dipendenti, ricordando che **i gdl di t sono  $(n_1-1) + (n_2-1)$  (scritti anche come  $n_1 + n_2 - 2$ )** e soprattutto che **l'errore standard della differenza ( $es_d$ )** è

$$es_d = \sqrt{s_p^2 \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

L'errore standard della differenza tra due campioni indipendenti può essere calcolata anche attraverso

$$es_d = s_p \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}$$

**Per calcolare l'intervallo fiduciale della differenza tra due medie campionarie con varianze statisticamente uguali** si applica la formula

$$\mathbf{m}_1 - \mathbf{m}_2 = (\bar{x}_1 - \bar{x}_2) \pm t_{\left(\frac{\alpha}{2}\right); (n_1+n_2-2)} \cdot s_p \cdot \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Quando le due varianze non sono omogenee, si può calcolare l'**errore standard della differenza** ( $es_d$ ) come

$$es_d = \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$$

ed il valore **d**

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

è approssimativamente una **deviata normale standardizzata**, quando  $n_1$  e  $n_2$  sono abbastanza grandi. E' possibile calcolare anche l'intervallo fiduciale

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \cdot es_d$$

Quando  $n_1$  e  $n_2$  sono piccoli si deve ricorrere ad altre soluzioni, diverse sia dal test **z** che dal test **t**, come il **metodo proposto da Welch oppure quello di Behrens**.

L'ultima formula riportata, che verifica la significatività di una differenza,

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

nella parte che considera l'errore standard evidenzia che **la varianza di una differenza tra due serie di dati è uguale alla somma delle due varianze campionarie**.

Anche nel caso di **due campioni dipendenti**, il calcolo della varianza attuato sulla colonna delle differenze risponde allo stesso principio: le variazioni d'errore (o residuo) presenti in ognuna delle due colonne dei dati possono essere nella stessa direzione, avere lo stesso segno rispetto alla media, e quindi annullarsi nella colonna delle differenze; possono altresì essere di segno opposto e quindi sommarsi. Il risultato complessivo è che **nella colonna delle differenze le variazioni d'errore diventano più ampie rispetto alle due colonne di dati originali; pertanto, la varianza delle differenze diviene la somma della varianza presente nelle due colonne**.

ESEMPIO 1. Si vuole saggiare se la concentrazione algale influisce positivamente sulla crescita del Cladocero *Daphnia magna*. In laboratorio sono stati allevati 40 individui dello stesso ceppo: con attribuzione casuale, successivamente 20 individui sono stati posti in una soluzione con concentrazione algale di 120.000 cellule per ml e gli altri 20 in una soluzione con concentrazione algale di 24.000 celle per ml.

Dopo 20 giorni, è stata misurata la lunghezza (in mm) dei 20 individui di ognuno dei due gruppi e si sono ottenuti i valori riportati nella tabella sottostante:

conc. Algale 120.000/ml ( $X_1$ )	conc. Algale 24.000/ml ( $X_2$ )
4,290	3,120
3,900	3,112
3,783	3,120
3,900	2,847
4,095	3,081
4,056	3,042
4,173	3,042
4,095	3,198
4,095	3,081
4,056	2,964
3,939	3,120
3,978	2,964
4,017	3,003
4,251	3,081
4,017	3,042
3,900	2,925
4,095	3,198
4,173	3,120
3,978	2,964
4,095	3,003

Più dettagliatamente, si vuole sapere

1- se gli animali cresciuti nella soluzione con concentrazione algale maggiore (gruppo  $X_1$ ) hanno raggiunto dimensioni significativamente superiori a quelli cresciuti nella soluzione con concentrazione algale minore (gruppo  $X_2$ );

2- quale è il **vero valore della differenza** ( $\delta$ ) nella crescita tra le due differenti situazioni di cibo

a) alla probabilità  $\alpha = 0.05$

b) alla probabilità  $\alpha = 0.01$

Risposte.

1 - Saggiare se esistono differenze significative nelle dimensioni degli animali allevati nelle due differenti situazioni significa verificare l'ipotesi nulla

$$H_0: \mu_1 = \mu_2$$

con ipotesi alternativa unilaterale

$$H_1: \mu_1 > \mu_2$$

Per ognuno dei due gruppi ( $X_1$  e  $X_2$ ), dai dati campionari si devono calcolare

- il numero di osservazioni ( $n$ ),
- la media ( $\bar{x}$ ),
- la devianza (SQ) e la varianza ( $s^2$ )

Campione	$X_1$	$X_2$
$n$	20	20
Media ( $\bar{x}$ )	4,0443	3,04335
Devianza (SQ)	0,30075	0,15326
Varianza ( $s^2$ )	0,015828	0,008066

Prima di procedere all'applicazione del test  $t$ , si deve controllare se le due varianze sono statisticamente uguali.

E' possibile utilizzare il rapporto fra la varianza maggiore (che dal confronto risulta essere quella del gruppo  $X_1$ ) e quella minore (del gruppo  $X_2$ ):

$$F_{(19,19)} = \frac{0,015828}{0,008066} = 1,962$$

Si confronta il risultato (1,962) con il valore critico di  $F_{19,19}$  per il livello del 5% (2,16); poiché il valore calcolato è minore di quello tabulato, si può assumere che le due varianze campionarie sono statisticamente uguali (anche se tale conclusione potrebbe essere contestata, data la differenza ridotta con il valore critico prefissato).

E' quindi corretto calcolare la varianza pooled

$$s_p^2 = \frac{0,30075 + 0,15326}{20 - 1 + 20 - 1} = \frac{0,45401}{38} = 0,01194$$

e da essa l'errore standard della differenza ( $es_d$ ) fra le medie:

$$es_d = \sqrt{0,01194 \cdot \left( \frac{1}{20} + \frac{1}{20} \right)} = 0,034554$$

Il valore del  $t$  con 38 gdl è dato da

$$t_{20+20-2} = \frac{4,0443 - 3,04355}{0,034554} = 28,96$$

e risulta uguale a 28,96.

Si tratta di un test ad una coda perché interessa valutare se la maggior concentrazione algale produce anche una maggiore crescita delle *Daphnie*. Il valore critico del t di Student associato alla probabilità 1% con 38 gradi di libertà è 2,429, nettamente inferiore al 29,157 calcolato; si conclude quindi che la maggior concentrazione algale influisce in modo altamente significativo sulla maggior crescita delle *Daphnie*.

**2a - L'intervallo fiduciale alla probabilità  $\alpha = 0.05$  della differenza fra le due medie** dato da

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0,05; (n_1+n_2-2)} \cdot es_d$$

con i dati campionari diviene

$$(4,0443 - 3,04335) \pm 1,686 \cdot 0,034554$$

Da essa si calcolano i due limiti dell'intervallo

$$I_1 = 0,94269 \quad I_2 = 1,059208$$

**2b - L'intervallo fiduciale alla probabilità  $\alpha = 0.01$  della differenza fra le due medie** dato da

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0,01; (n_1+n_2-2)} \cdot es_d$$

con i dati del campione è

$$(4,0443 - 3,04335) \pm 2,429 \cdot 0,034554$$

I due limiti dell'intervallo ( $I_1$  e  $I_2$ ) risultano

$$I_1 = 0,917 \quad I_2 = 1,086$$

Si può osservare come, sia alla probabilità 0.05 che a quella 0.01, l'intervallo fiduciale calcolato non comprenda lo 0, espresso come differenza attesa ( $\delta$ ) nell'ipotesi nulla.

La differenza reale ( $\delta$ ) è positiva e significativa; di conseguenza, si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa espressa.

ESEMPIO 2. E' stato misurato il valore del pH in due gruppi di laghi appenninici: 12 laghi hanno un bacino imbrifero ricco di rocce carbonatiche affioranti ( $X_1$ ) e 13 laghi sono collocati in un bacino imbrifero senza rocce carbonatiche affioranti ( $X_2$ ).

I valori misurati sono riportati nella tabella sottostante:

$X_1$	$X_2$
7,94	7,30
8,03	7,26
8,18	6,82
8,03	7,08
8,19	7,13
8,01	7,37
8,16	7,42
8,16	7,16
8,18	6,89
8,29	6,96
7,94	7,13
8,29	7,08
---	7,17

Si vuole conoscere:

- 1 - se la differenza media del pH dei 2 gruppi di laghi collocati in bacini imbriferi con diversa presenza di rocce carbonatiche è statisticamente significativa;
- 2 - quale è la reale differenza media del pH tra le due situazioni
  - a) alla probabilità  $\alpha = 0.05$
  - b) alla probabilità  $\alpha = 0.01$ .

Risposte.

1 - In mancanza di conoscenze aprioristiche certe sul valore del pH nelle due differenti situazioni ambientali, il test è bilaterale; sono due campioni indipendenti e le ipotesi possono essere espresse mediante

$$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 \neq \mu_2$$

Dalle rilevazioni campionarie raccolte, si calcolano i dati necessari alla stima del valore del  $t$  per 2 campioni indipendenti

Campioni	X <sub>1</sub>	X <sub>2</sub>
n	12	13
Media ( $\bar{x}$ )	8,117	7,136
Devianza (SQ)	0,16656	0,37690
Varianza ( $s^2$ )	0,015	0,0314

Prima di procedere all'applicazione del test si deve controllare se le due varianze possono essere considerate statisticamente simili, mediante il rapporto fra la varianza maggiore (che risulta essere quella appartenente al gruppo X<sub>2</sub>) e quella minore (del gruppo X<sub>1</sub>):

$$F_{(12,11)} = \frac{0,0314}{0,015} = 2,093$$

Si confronta il risultato con il valore critico di  $F_{(12,11)}$  per il livello di probabilità 0.05 che è 2,79; poiché il valore calcolato (2,093) è minore di quello tabulato, le due varianze possono essere giudicate statisticamente uguali.

Di conseguenza, è possibile confrontare le due medie con il test t.

Dapprima si calcola la varianza pooled :

$$s_p^2 = \frac{0,16656 + 0,37690}{12 - 1 + 13 - 1} = \frac{0,54346}{23} = 0,02362$$

e da essa l'errore standard della differenza fra le due medie ( $es_d$ ):

$$es_d = \sqrt{0,02362 \cdot \left( \frac{1}{12} + \frac{1}{13} \right)} = 0,06152$$

Applicando alla formula del test t per due campioni indipendenti

$$t_{n_1+n_2-2} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mathbf{m}_1 - \mathbf{m}_2)}{\sqrt{s_p^2 \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

i dati precedentemente calcolati, si stima il valore del t con 23 gdl

$$t_{12+13-2} = \frac{8,117 - 7,136}{0,06152} = 15,946$$

che risulta uguale a 15,946.

Il valore critico del t di Student associato alla probabilità  $\alpha = 0.01$  con 23 gradi di libertà è 2,807 e risulta nettamente inferiore al valore calcolato (15,946).

I due gruppi di laghi hanno un pH medio statisticamente molto diverso.

**2a - L'intervallo fiduciale della differenza fra le due medie associato alla probabilità  $\alpha = 0.05$**  può essere calcolato mediante la formula:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0,05; (n_1+n_2-2)} \cdot es_d = 0,981 \pm 2,069 \cdot 0,06152$$

Applicata ai dati dei due campioni a confronto, diventa

$$(8,117 - 7,136) \pm 2,069 \cdot 0,06152$$

e permette di stimare i due limiti dell'intervallo

$$I_1 = 0,851 \quad I_2 = 1,111$$

**2b - L'intervallo fiduciale della differenza fra le due medie associato alla probabilità  $\alpha = 0.01$**  è:

$$(8,117 - 7,136) \pm 2,807 \cdot 0,06152$$

e i limiti sono

$$I_1 = 0,804 \quad I_2 = 1,158$$

## 5.9 POTENZA A PRIORI E A POSTERIORI DEL TEST $t$ , CON UN CAMPIONE E CON DUE CAMPIONI DIPENDENTI O INDIPENDENTI

Quando si programma un esperimento, la prima domanda è quasi sempre: “**Quanti dati devo raccogliere?**” Effettuato un test, se il risultato non è significativo è importante rispondere al quesito: “**Con il campione raccolto, che probabilità avevo che il test risultasse significativo?**”

La prima domanda, chiamata anche **potenza a priori**, e la seconda, **potenza a posteriori**, hanno già avuto una prima risposta nel capitolo IV, con l'uso della distribuzione  $z$ .

Ricordando i concetti principali,

- il numero ( $n$ ) di dati necessari a rendere significativo un test e
- la misura della sua potenza ( $1-b$ )

possono essere stimate sulla base delle relazioni che esistono tra 5 quantità:

- la **probabilità  $a$**  alla quale il test  $t$  di Student dovrebbe risultare significativo, dopo aver raccolto i dati; per essa deve essere specificata pure la **direzione dell'ipotesi  $H_1$** , se **unilaterale** oppure **bilaterale**;
- la **probabilità  $b$**  (sempre **unilaterale**), cioè il rischio che si intende correre che il test non risulti significativo; questo concetto risulta più chiaro utilizzando quello, ad esso alternativo, di **potenza ( $1-b$ )**, cioè della probabilità di rifiutare (correttamente) l'ipotesi nulla, quando essa è falsa;
- la **differenza  $d$**  che **a priori** si intende dimostrare significativa (con  $d$  uguale alla differenza tra  $m$  e  $m_0$  quando si applica il test ad un campione oppure alla differenza tra  $m_1$  e  $m_2$  nel caso di due campioni indipendenti); la **differenza campionaria  $d$**  per un test **a posteriori** (con  $d$  uguale a  $\bar{x}_1 - \bar{x}_2$  nel caso di due campioni indipendenti);
- la **varianza  $s^2$**  del fenomeno, misurata su uno studio pilota o un campione preliminare quando la **varianza reale  $s^2$  è ignota**;
- la **dimensione  $n$**  del campione.

L'uso di  $s^2$  al posto di  $s^2$  impone il ricorso alla distribuzione  $t$  in sostituzione della distribuzione  $z$ . Quando  $n$  è sufficientemente **grande**, la distribuzione normale  $z$  e la distribuzione  $t$  **praticamente coincidono**, fornendo risposte molto simili; le differenze sono ritenute importanti, quando il campione ha meno di 30 osservazioni e quindi i gdl del  $t$  sono una quantità limitata.

A differenza della distribuzione normale  $z$ , che è unica, la distribuzione  **$t$  di Student** è formata da una intera famiglia di distribuzioni, una per ogni grado di libertà.

Per stimare le dimensioni ( $n$ ) del campione, occorre quindi iniziare da un valore di  $t$  che dipende da  $n$ , non ancora calcolato; di conseguenza, per ottenere  $n$  si deve utilizzare un metodo iterativo, che stima  $n$  con approssimazioni successive, sempre più vicine al valore reale (in pratica, sono sufficienti due o al massimo tre calcoli).

Una impostazione didattica chiara e semplice dei concetti e dei metodi relativi alla potenza di un test richiede solo tre formule, una per ognuna delle **tre quantità fondamentali** che servono al ricercatore per verificare l'ipotesi nulla sulle medie attraverso **il test t di Student**.

Esse forniscono un aiuto importante

- sia nella programmazione di un esperimento, cioè nella fase iniziale di ogni ricerca empirica chiamata disegno sperimentale,
- sia nella discussione del risultato, che dovrebbe avvenire sempre, dopo la raccolta dei dati e l'applicazione del test.

Queste tre quantità, delle cinque elencate in precedenza, sono:

- **n**, le dimensioni richieste (o minime) del campione, necessarie per rifiutare l'ipotesi nulla,
- **d**, la differenza utile (o minima) che si vuole dimostrare significativa con i dati raccolti,
- **1-b**, la potenza del test.

Le **altre due** quantità fondamentali sono meno condizionate dal ricercatore:

- **s<sup>2</sup>**, la varianza campionaria, poiché dipende dalla natura dei dati ed è tipica di ogni fenomeno biologico e naturale,
- **a**, prefissata dal ricercatore ma in modo standard, quale la convenzione sui valore soglia (0.05 – 0.01 – 0.001) dei limiti di significatività, oppure sulla base di fattori esterni all'esperimento, quale il rischio che egli corre se la conclusione si rivelasse errata (vedi capitolo IV, sui criteri per scelta del valore di **a**).

Il calcolo delle tre quantità fondamentali (**n**, **d**, **1-b**) utilizza gli stessi concetti, ma ricorre a formule leggermente differenti, se sono riferite al caso di:

- un campione oppure due campioni dipendenti, in quanto utilizza solo la colonna delle differenze,
- due campioni indipendenti, in cui si utilizzano entrambe le serie di rilevazioni.

Affinché, con un campione o con due campioni dipendenti, la differenza tra la **media campionaria e la media attesa oppure la media delle differenze e quella attesa** risultino significative, cioè per rifiutare l'ipotesi nulla

$$H_0: \mu = \mu_0 \quad \text{nel caso di un campione,}$$

$$H_0: \delta = 0 \quad \text{nel caso di due campioni dipendenti}$$

- il numero minimo di osservazioni **n** (numero di dati in un campione e numero di coppie in due campioni dipendenti) deve essere

$$n \geq (t_{a,n} + t_{b,n})^2 \times \frac{s^2}{d^2}$$

- la differenza minima che può risultare significativa **d** è

$$d \geq \sqrt{\frac{s^2}{n}} \cdot (t_{a,n} + t_{b,n})$$

- la potenza **1-b** del test, stimata conoscendo **b**, è

$$t_{b,n} \leq \frac{d}{\sqrt{\frac{s^2}{n}}} - t_{a,n}$$

ESEMPIO 1. Ad un tecnico di laboratorio è stato chiesto di valutare se la concentrazione di una sostanza utilizzata per conservare un alimento (in grammi di sostanza secca per Kg di alimento) differisce di oltre 1 gr. dalla quantità stabilita, ritenuta ottimale: può avere conseguenze negative sia se in difetto sia se in eccesso.

Dopo alcune misure campionarie (12,5; 13,1; 14,7; 12,8;...) per avere una misura della variabilità, egli ha stimato la varianza dei suoi campioni ottenendo  $s^2 = 3,24$ .

Quanti misure deve effettuare per dimostrare che la quantità media di conservante si discosta di almeno un 1 gr. dalla quantità ritenuta ottimale alla probabilità del 95% ed abbia una probabilità del 90% di ottenere il risultato richiesto?

Risposta

Dalla presentazione del problema, si ricava:

-  $\alpha = 0.05$  in un **test bilaterale**; di esso occorre conoscere il valore di **t**; ma poiché nella tabella dei valori critici sono richiesti anche i gdl, in prima approssimazione si può assumere che servano circa 30 dati ( $n = 29$ ); quindi

$$t_{(0.025; 29)} = 2,045$$

-  $\beta = 0.10$  in un **test unilaterale**; il valore di **t**, sempre per il numero di dati preventivato in prima approssimazione ( $n = 29$ ), è

$$t_{(0.10; 29)} = 1,3114$$

-  $s^2 = 3,24$

-  $d = 1,0$

Da essi con

$$n \cong (t_{a,n} + t_{b,n})^2 \times \frac{s^2}{d^2}$$

si calcola un primo valore

$$n \cong (2,045 + 1,3114)^2 \times \frac{3,24}{1,0^2} = 11,27 \times 3,24 = 36,5$$

di **n** uguale a 36,5.

Poiché il valore calcolato (36,5) si discosta da quello iniziale (30), si devono individuare i valori di  $t$  corrispondenti alla nuova stima; tuttavia, quando la tabella non riporta i valori di  $t$  per 36 gdl, ma solo per 35 o 40 gdl, si può utilizzare quello di  $n = 35$ :

- sia per  $a = 0.05$        $t_{(0.025; 35)} = 2,030$

- sia per  $b = 0.010$        $t_{(0.10; 35)} = 1,3062$

Con essi si ottiene

$$n \cong (2,030 + 1,3062)^2 \times \frac{3,24}{1,0^2} = 11,13 \times 3,24 = 36,06$$

una seconda stima di  $n$ ,

che risulta uguale a 36,06; praticamente è identica alla stima precedente, poiché il risultato può essere solo un intero. Arrotondando in eccesso, servono **almeno 37** dati.

ESEMPIO 2. Se il tecnico, a causa del tempo richiesto e dei fondi a disposizione, valutasse che può effettuare solo 20 misure, quale differenza minima potrà dimostrare significativa, ovviamente a parità degli altri 3 parametri? (a, direzione dell'ipotesi  $H_1$ , b)

Risposta

Poiché

-  $n = 19$

- per  $a = 0.05$  **bilaterale** il valore di  $t_{(0.025; 19)}$  è **2,093**

- per  $b = 0.10$  **unilaterale** il valore di  $t_{(0.10; 19)}$  è **1,3277**

con

$$d \geq \sqrt{\frac{s^2}{n}} \cdot (t_{a,n} + t_{b,n})$$

si calcola una differenza minima

$$d \geq \sqrt{\frac{3,24}{20}} \cdot (2,093 + 1,3277) = 0,402 \cdot 3,42 = 1,38$$

uguale a 1,38.

Non potrà verificare, cioè non potrà dimostrare significativa, una differenza  $d$  inferiore a 1,38.

ESEMPIO 3. Raccolti i 20 dati, che probabilità avrebbe avuto il tecnico di evidenziare una differenza pari a 1 gr. rispetto al valore atteso?

Risposta

Per stimare la potenza ( $1-b$ ) del test, si deve calcolare il valore  $t_b$  e poi da esso la probabilità  $b$ .

Con

- $d = 1$
  - e gli altri parametri uguali all'esempio 2
- mediante

$$t_{b,n} \leq \frac{d}{\sqrt{\frac{s^2}{n}}} - t_{a,n}$$

si calcola

$$t_{b,19} \leq \frac{1}{\sqrt{\frac{3,24}{20}}} - 2,095 = \frac{1}{0,403} - 2,095 = 2,48 - 2,095 = 0,385$$

un valore del  $t$  uguale a **0,385** con **19** gdl.

Purtroppo le tabelle sinottiche riportano solamente alcuni valori per le probabilità fondamentali.

Da quella **unilaterale**, per  $n = 19$ , il valore critico per  $b = 0.25$  (anche per  $b$  si utilizza la tabella  $a$ ) è **0,6876**.

Di conseguenza, la probabilità  $b$  è **nettamente maggiore di 0.25**. Poiché **la potenza del test è  $1-b$** , la probabilità di rifiutare una differenza  $d$  pari a **1** gr. è **ampiamente inferiore a 0.75**.

Nel caso di **due campioni indipendenti**, affinché la differenza  $d$  tra le due medie  $m_1$  e  $m_2$  risulti significativa,

- il campione, per **ognuno dei due gruppi**, deve avere dimensioni minime  $n$

$$n \geq (t_{a,n} + t_{b,n})^2 \times \frac{2s_p^2}{d^2}$$

- la differenza minima  $d$  che può risultare significativa è

$$d \geq \sqrt{\frac{2s_p^2}{n}} \cdot (t_{a,n} + t_{b,n})$$

- la potenza  $1-b$  del test, stimata conoscendo  $b$ , è

$$t_{b,n} \leq \frac{d}{\sqrt{\frac{2s_p^2}{n}}} - t_{a,n}$$

Rispetto al caso precedente, riguardante un campione o due campioni dipendenti, con **due campioni indipendenti** nella scelta dei valori  $t_a$  e  $t_b$  i gradi di libertà(n) non sono  $n-1$ , ma **2n-2**, poiché il numero di dati complessivo utilizzato nei due gruppi a confronto è uguale a **2n**.

ESEMPIO 4. Si vuole verificare se una sostanza tossica, somministrata con il cibo a cavie, determina un accrescimento più limitato, rispetto alla situazione normale. Da un esperimento pilota su due gruppi di animali (sperimentale e controllo) è stata stimata la crescita (in millimetri) dopo 20 giorni e la varianza comune  $s_p^2$  è risultata uguale 0,54.

Per programmare un test in cui una differenza di 0,4 millimetri tra le dimensioni medie dei due gruppi risulti significativa alla probabilità 0.05 e la probabilità di scoprire tale differenza sia pari a 0.9, quanti individui servono?

Risposta

Dalla presentazione del problema, si ricava:

-  $\alpha = 0.05$  in un **test unilaterale**; di esso occorre conoscere il valore di  $t$ ; ma poiché nella tabella dei valori critici sono richiesti anche i gdl, in prima approssimazione si può assumere che servano circa **30** dati per gruppo ( $n = 58$ ); quindi

$$t_{(0.05; 58)} = 1,6716$$

-  $\beta = 0.10$  in un **test unilaterale**; il valore di  $t$ , sempre per il numero di dati preventivato in prima approssimazione ( $n = 58$ ), è

$$t_{(0.10; 58)} = 1,2963$$

-  $s_p^2 = 0,54$

-  $d = 0,4$ .

Da essi con

$$n \geq (t_{\alpha, n} + t_{\beta, n})^2 \times \frac{2s_p^2}{d^2}$$

si ottiene una **dimensione minima di ognuno dei due gruppi**

$$n \geq (1,6716 + 1,2963)^2 \cdot \frac{2 \cdot 0,54}{0,4^2} = 2,9679^2 \cdot \frac{1,08}{0,16} = 8,808 \cdot 6,75 = 59,45$$

pari a 60 individui.

La differenza tra i 30 preventivati e i 60 stimati è grande; di conseguenza, si deve procedere ad una seconda stima, usando i valori di  $t$  corrispondenti a  $v = 118$ . Poiché nella tabella il valore più vicino è quello per  $v = 120$ , è possibile servirsi di questo valore (quando i gdl sono così alti, la differenza tra i valore di  $t$  è minima).

In una distribuzione ad una coda,

- per a  $t_{(0.05; 120)} = 1,6577$

- per b  $t_{(0.10; 120)} = 1,2886$

la seconda stima

$$n \geq (1,6577 + 1,2886)^2 \cdot \frac{2 \cdot 0,54}{0,4^2} = 2,9463^2 \cdot \frac{1,08}{0,16} = 8,681 \cdot 6,75 = 58,60$$

calcola  $n = 58,60$ .

Poiché è poco diverso dal precedente, questo ne calcola 59 rispetto ai 60 della prima stima, anche senza passare alla terza si può concludere che ognuno dei due gruppi debba essere formato da almeno 60 individui, per un totale di 120.

ESEMPIO 5. Se il ricercatore dispone solo di 80 cavie, formando due gruppi di 40 individui quale è la differenza minima  $d$  che potrà dimostrare significativa, alle stesse condizioni del test precedente?

Risposta

Poiché

-  $n = 78$

- per a = **0.05 unilaterale** il valore di  $t_{(0.025; 78)}$  è **1,6646**

- per b = **0.10 unilaterale** il valore di  $t_{(0.10; 78)}$  è **1,2925**

con

$$d \geq \sqrt{\frac{2s_p^2}{n}} \cdot (t_{a,n} + t_{b,n})$$

si calcola una differenza minima

$$d \geq \sqrt{\frac{2 \cdot 0,54}{40}} \cdot (1,6646 + 1,2925) = \sqrt{0,027} \cdot 2,9571 = 0,486$$

$\delta$  uguale a 0,486.

Non potrà verificare, cioè non potrà risultare significativa, una differenza  $d$  inferiore a 0,486.

ESEMPIO 6. Effettuando un esperimento con 40 dati per gruppo, che probabilità avrebbe avuto di evidenziare una differenza pari a 0,4 tra le due medie?

Risposta

Per stimare la potenza (1-b) del test, si deve calcolare il valore  $t_b$ , e poi da esso la probabilità b.

Con

- $d = 0,4$
  - e gli altri parametri uguali all'esempio 5
- mediante

$$t_{b,n} \leq \frac{d}{\sqrt{\frac{2s_p^2}{n}}} - t_{a,n}$$

stima

$$t_{b,78} \leq \frac{0,4}{\sqrt{\frac{2 \cdot 0,54}{40}}} - 1,6646 = \frac{0,4}{0,1643} - 1,6646 = 2,4346 - 1,6646 = 0,77$$

un valore del  $t$  uguale a **0,77** con **78** gdl.

Purtroppo le tabelle sinottiche spesso riportano solamente alcuni valori.

Da quella **unilaterale**, per  $n = 78$ ,

- il valore critico per  $b = 0.25$  (nella tabella è indicato  $\alpha$ , ma le due distribuzioni sono identiche) è **0,6776** e
- il valore critico per  $b = 0.10$  è **1,2925**.

Di conseguenza, la probabilità b è **leggermente minore di 0.25**. Poiché la **potenza del test è 1-b**, la probabilità di rifiutare una differenza d pari a **0,4 è di poco superiore a 0.75**.

L'uso della distribuzione t di Student per calcolare la probabilità di b (o la potenza **1-b**) fornisce una risposta approssimata, poiché vincolata alla probabilità riportate nelle tabelle sinottiche: se il valore calcolato non coincide esattamente con uno dei valori riportati, ci si deve limitare ad indicare un intervallo entro il quale la probabilità è compresa, ad esempio, tra 0,2 e 0,1 oppure tra 0,025 e 0,05 come evidenzia l'intestazione della tabella sottostante, che riporta le probabilità di norma utilizzate nelle tavole sinottiche dei test di statistica.

Gradi di Libertà	?								
	0.500	0.400	0.200	0.100	0.050	0.025	0.010	0.005	0.001

Per una risposta più precisa, nel caso di **due o più campioni** è possibile ricorrere ai **metodi grafici**. Diffusi da tempo, ora queste tecniche appaiono superate dai programmi informatici, con i quali la stima della probabilità (a e b) è fornita alla seconda cifra decimale, a causa dell'ampia memoria dei computer.

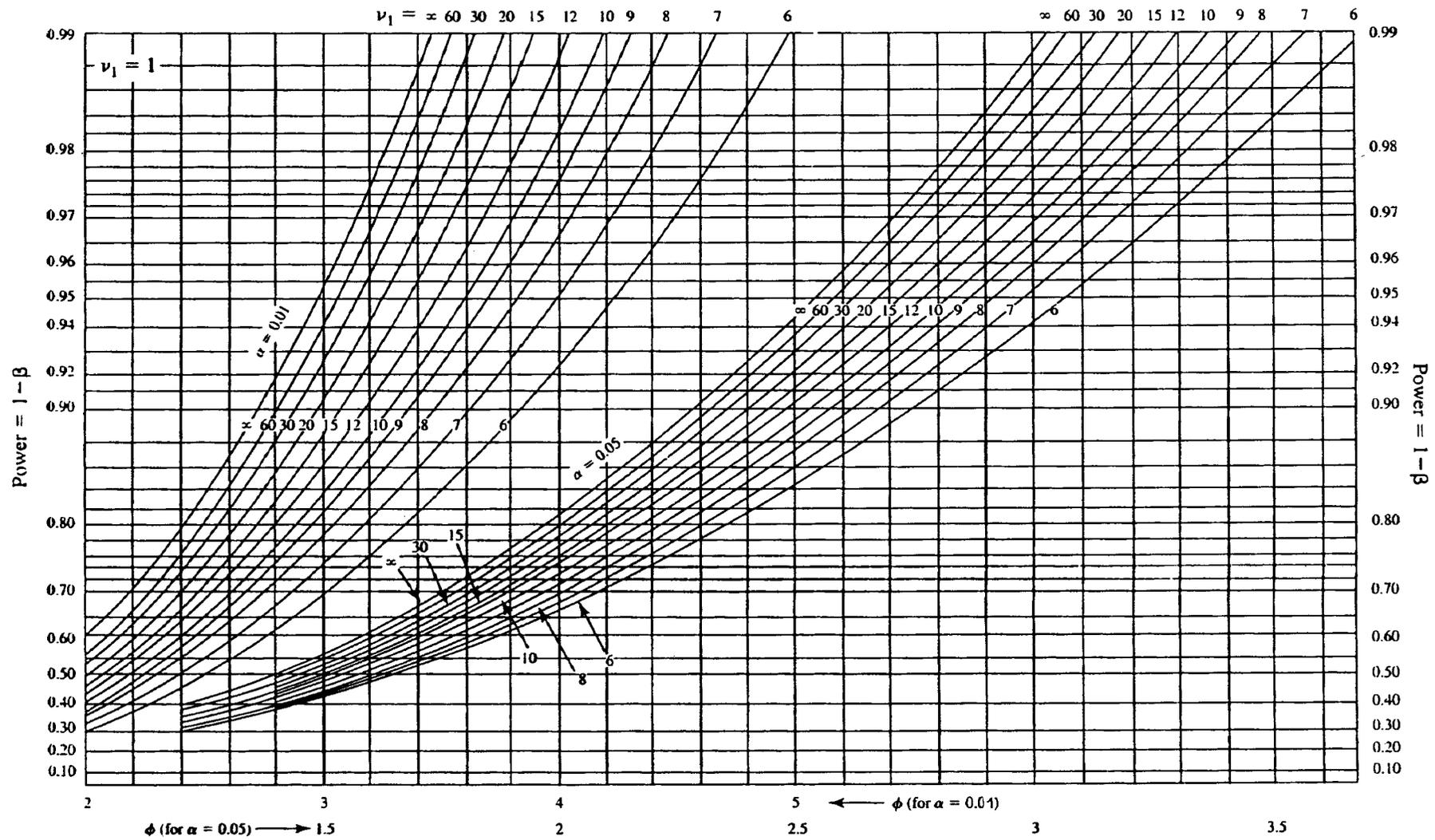
Per entrare in un grafico di curve di probabilità **1-b**, come riportato nella pagina successiva, servono 4 informazioni:

- il numero di **gruppi a confronto**, in questo caso sono 2: in una ANOVA corrispondono  $n = 1$ , i gdl della varianza tra trattamenti;
- la probabilità **a** per la quale si programma la significatività del test; questi grafici riportano solo le curve per  $a = 0.01$  e  $a = 0.05$ ;
- il valore di  $f$  (*phi*) ottenuto con la formula seguente, in accordo con la simbologia consueta

$$f = \sqrt{\frac{nd^2}{4s_p^2}}$$

- il numero **n** di dati in ogni gruppo, che devono essere distribuiti in modo bilanciato ( $n_1 = n_2$ ).

(Maggiori informazioni sulla costruzione di questo grafico sono riportate nel paragrafo sulla potenza in un test ANOVA).



Complessa nella costruzione, questa procedura è semplice nella sua utilizzazione:

1) Il grafico riportato può essere utilizzato solo

- in un **test bilaterale**
- per la probabilità  $\alpha = 0.01$  (ultima riga inferiore) e per la probabilità  $\alpha = 0.05$  (prima riga sotto la figura, appena sopra la precedente).

2) Una volta scelta la probabilità  $\alpha$  e stimato il valore di  $f$  ( $\phi$ ) che

- per la probabilità  $\alpha = 0.01$  varia approssimativamente da **1,5 a 3,5** per una potenza che varia da **0.30 a 0.99** in rapporto a  $n$ ,
- per la probabilità  $\alpha = 0.05$  varia da **2 a 5** sempre per una potenza **1-b** che varia da **0.30 a 0.99** in rapporto a  $n$ ,

3) dall'asse dell'ascissa si sale perpendicolarmente dal valore di  $f$  lungo le rette tracciate (parallele all'ordinata) fino ad incontrare la curva dei gdl  $n$ , dove  $n = 2(n-1)$ ;

4) da quel punto, spostandosi lateralmente lungo le rette parallele all'ascissa si giunge all'ordinata, dove (riportato sia a destra sia a sinistra) si legge il valore della potenza, la probabilità **1- b**.

In conclusione, per risalire da  $f$  alla potenza **1-b** servono altri due valori:  $\alpha$ ,  $n_2$ ;

- il primo solo per le due probabilità ricordate,
- il secondo per gli 11 gdl riportati (6, 7, 8, 9, 10, 12, 15, 20, 30, 60,  $\infty$ ); per altri gdl, si deve interpolare tra le due curve che li comprendono.

ESEMPIO 7. Stimare il valore di  $b$  sia

- a) con il **test t** (bilaterale) che
- b) con l'**uso del grafico**,

in un esperimento programmato per valutare,

- alla probabilità  $\alpha = 0.05$ ,
- se due sostanze tossiche hanno effetti che differiscono di una quantità  $d = 1$
- utilizzando 15 osservazioni per gruppo:  $n_1 = n_2 = 15$ .

Un esperimento pilota ha fornito una

- stima della varianza associata  $s_p^2 = 0,52$ .

Risposta

A) Con

- gdl  $n = 2(n-1) = 28$

- un valore di **t** alla probabilità **a = 0.05**, per un test **bilaterale** uguale a **2,048**

$$t_{0,025,28} = 2,048$$

si stima un valore di **t** per la probabilità **b** (sempre **unilaterale**)

$$t_{b,28} = \frac{1}{\sqrt{\frac{2 \cdot 0,52}{15}}} - 2,048 = \frac{1}{0,2633} - 2,048 = 3,798 - 2,048 = 1,75$$

uguale a 1,75

Poiché il valore critico di **t unilaterale** con **28** gdl

- alla probabilità 0.05 è uguale a 1,7011
- alla probabilità 0.025 è uguale a 2,0484

la probabilità **b** stimata è **compresa tra 0.05 e 0.025**; di conseguenza **la potenza 1-b è compresa tra 0,95 e 0,975**.

B) Utilizzando il grafico, si ottiene un valore di **f (phi)**

$$f = \sqrt{\frac{15 \cdot 1^2}{4 \cdot 0,52}} = \sqrt{\frac{15}{2,08}} = \sqrt{7,21} = 2,685$$

uguale a 2,685

Nel grafico delle curve di potenza,

- nella riga **a = 0.05** (ultima riga), il valore di 2,685 può essere collocato approssimativamente vicino alla seconda riga verticale dopo 2.5;
- salendo perpendicolarmente, si incontra la curva per **n = 30** in un punto che, proiettato sull'asse delle ordinate, cade tra **0.95 e 0.96**.

La potenza di questo test è appunto compresa tra 0.95 e 0.96.

La precisione della potenza stimata dipende dalle caratteristiche del grafico, che spesso è più dettagliato di quello qui utilizzato, ai soli fini di spiegare i concetti e le procedure.

I metodi descritti per 2 campioni sono utili **a priori**, **nella fase di programmazione dell'esperimento**, per avere un'idea di

- quanti dati **n** raccogliere,
- quale sia la differenza **d** che è possibile evidenziare come significativa,
- la potenza **1-b** del test,

se esso verrà effettuato seguendo le idee espresse nella programmazione, dopo aver misurato la varianza associata  $s_p^2$  con un esperimento pilota.

**A posteriori, dopo la raccolta dei dati**, spesso attuata senza la programmazione appena indicata, può avvenire che **il test non risulti significativo**.

Per comprenderne le cause, per valutare se la significatività non dimostrata molto probabilmente nella realtà non esiste oppure è dovuta ad un errore nella realizzazione dell'esperimento, diventa importante il calcolo della potenza, poiché  $b$  fornisce la probabilità di commettere l'errore di II Tipo.

Nel caso di **2 campioni indipendenti**, avviene spesso che essi siano formati da un numero differente di osservazioni ( $n_1 \neq n_2$ ); per utilizzare le formule precedenti,  $n$  è calcolato attraverso la media armonica.

Dalla formula generale per  $k$  gruppi

$$n = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2} + \dots + \frac{1}{n_k}}$$

nel caso di 2 soli campioni si ricava la formula abbreviata

$$n = \frac{2n_1n_2}{n_1 + n_2}$$

Il valore del  $f$  è ottenuto da

$$f = \sqrt{\frac{nd^2 - 2s_p^2}{4s_p^2}}$$

dove

- $n$ , il numero di dati in ognuno dei 2 gruppi, è stimato con la formula precedente quando  $n_1 \neq n_2$
- $d$  è la differenza campionaria tra le due medie:  $d = \bar{x}_1 - \bar{x}_2$
- $s_p^2$  è la varianza associata dei 2 campioni di cui si verifica la significatività della differenza tra le medie,
- la potenza  $1 - b$  è stimata dal valore  $f$  con  $n = n_1 + n_2 - 2$

ESEMPIO 8. Per verificare la differenza tra il livello medio d'inquinamento di due corpi idrici, sono state effettuate 8 prelievi ( $n_1 = 8$ ) nel primo e 6 ( $n_2 = 6$ ) nel secondo, ottenendo

- $\bar{x}_1 = 7,52$  e  $\bar{x}_2 = 6,95$
- $s_p^2 = 0,32$

Stimare

- A) la significatività della differenza tra le due medie
- B) la potenza del test effettuato.

Risposta

A) E' un **test bilaterale**, con

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

nel quale

$$t_{n_1+n_2-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = t_{12} = \frac{7,52 - 6,95}{\sqrt{0,32 \cdot \left(\frac{1}{8} + \frac{1}{6}\right)}} = \frac{0,57}{\sqrt{0,32 \cdot 0,292}} = \frac{0,57}{0,306} = 1,863$$

il **t** con 12 gdl risulta uguale a **1,863**.

Poiché, come riportato nella tavola sinottica, il valore di **t bilaterale** con  $n = 12$

- alla probabilità  $\alpha = 0.20$  è uguale a **1,356**
- alla probabilità  $\alpha = 0.10$  è uguale a **1,782**
- alla probabilità  $\alpha = 0.05$  è uguale a **2,8179**

si deve concludere che non è possibile rifiutare l'ipotesi nulla, per il livello di significatività del 5%: la probabilità di trovare una differenza uguale o superiore a quella misurata si trova tra il 10 e il 5 per cento (più vicina al 10%).

B) La potenza del test, la probabilità di rifiutare (correttamente) l'ipotesi nulla quando è falsa, trattandosi di

- un **test bilaterale**
- per un valore critico di  $\alpha = 0.05$
- oppure  $\alpha = 0.01$

può essere stimata mediante il valore di

- $f$  con  $n = n_1 + n_2 - 2$  e l'uso del **grafico** presentato.

Con **due campioni non bilanciati**,

- dapprima si stima **n**

$$n = \frac{2 \cdot 8 \cdot 6}{8 + 6} = \frac{96}{14} = 6,86$$

che risulta uguale a 6,86

(osservare ancora una volta come campioni non bilanciati facciano perdere informazione, essendo **n** minore della media aritmetica, che si otterrebbe con due gruppi formati da 7 repliche)

- successivamente la differenza campionaria tra le due medie:  $d = \bar{x}_1 - \bar{x}_2 = 0,57$  e  $s_p^2 = 0,32$

si stima  $f$

$$f = \sqrt{\frac{6,86 \cdot 0,57^2 - 2 \cdot 0,32}{4 \cdot 0,32}} = \sqrt{\frac{2,23 - 0,64}{1,28}} = \sqrt{1,24} = 1,11$$

che risulta uguale a **1,11**.

Nel grafico, alla probabilità  $\alpha = 0,05$  il valore di  $f = 1,11$  si trova a sinistra del valore minimo 1,5 riportato. Presumibilmente incontra la curva con 12 gdl con un valore di potenza intorno a 0.40.

Per  $\alpha = 0,01$  è nettamente inferiore al valore 2 riportato come minimo e quindi la potenza ha un valore intorno a 0.30 o minore.

In conclusione, anche se  $\alpha$  è risultata minore del 10%, vi era

- una probabilità  $\beta$  intorno al 60% che il test non risultasse significativo per  $\alpha = 0,05$
- una probabilità  $\beta$  intorno al 70% che il test non risultasse significativo per  $\alpha = 0,01$

ESEMPIO 9. L'aumento del numero di dati, a parità degli altri parametri, accresce notevolmente la potenza del test. Rifare i calcoli dell'esempio 8, con 15 dati per gruppo ( $n = n_1 = n_2 = 15$ ).

Risposta

A) Il test  $t$

$$t_{28} = \frac{7,52 - 6,95}{\sqrt{0,32 \cdot \left(\frac{1}{15} + \frac{1}{15}\right)}} = \frac{0,57}{\sqrt{0,32 \cdot 0,133}} = \frac{0,57}{0,206} = 2,767$$

sarebbe risultato uguale a **1,863** con 28 gdl.

Poiché, nella tavola sinottica, il valore di **t bilaterale** con  $n = 28$

- alla probabilità  $\alpha = 0,05$  è uguale a **2,368**
- alla probabilità  $\alpha = 0,025$  è uguale a **2,763**
- alla probabilità  $\alpha = 0,01$  è uguale a **3,047**

si sarebbe rifiutata l'ipotesi nulla alla probabilità del 5 per cento: la probabilità di trovare una differenza uguale o superiore a quella misurata si sarebbe trovata tra il 2,5 e l'1 per cento (appena inferiore al 2,5%).

B) Per stimare la potenza del test,  
si calcola f

$$f = \sqrt{\frac{15 \cdot 0,57^2 - 2 \cdot 0,32}{4 \cdot 0,32}} = \sqrt{\frac{4,47 - 0,64}{1,28}} = \sqrt{3,30} = 1,82$$

che risulta uguale a **1,82**.

Nel grafico, alla probabilità  $\alpha = 0,05$  il valore di  $f = 1,82$  si trova a destra del valore minimo 1,5 riportato. Approssimativamente, incontra la curva con  $n = 28$  (nel grafico  $n = 30$ ) per un valore di **potenza (1-b)** intorno a **0.70**.

Vi era una probabilità del 30% di commettere l'errore  $\beta$ , cioè di non rifiutare l'ipotesi nulla alla probabilità  $\alpha = 0,05$ .

#### 5.10. DIMENSIONE DEL CAMPIONE E PRECISIONE NELLA STIMA SIA DI UNA MEDIA SIA DELLA DIFFERENZA TRA DUE MEDIE

Nella ricerca sul campo ed in laboratorio, è frequente il caso in cui un tecnico o un ricercatore debbano fornire un **valore medio attendibile**, con una differenza piccola e non superiore ad una quantità certa, rispetto alla media reale.

**A posteriori**, dopo aver raccolto i dati, l'**intervallo fiduciale** permette di valutare entro quali limiti si trova la media reale alla probabilità desiderata.

**A priori**, prima della raccolta dei dati, sovente occorre essere in grado di conoscere **quanti dati è necessario raccogliere**, quante misure effettuare, affinché **la media campionaria  $\bar{x}$  abbia la precisione desiderata**, cioè non si discosti dalla media reale  $m$  o della popolazione di una quantità superiore a un valore  $d$  prefissato.

Nel caso della **media un campione** o **della media delle differenze tra due campioni dipendenti**, il numero  $n$  di osservazioni necessario per fornire una stima con la precisione desiderata è dato da

$$n = \frac{s^2 \cdot t_{\alpha/2, n}^2}{d^2}$$

dove

-  $s^2$  è la varianza del campione (stimata su un sondaggio preliminare),

- il valore di **t** deve essere preso per un **test bilaterale**, alla probabilità **a** prefissata,
- con gdl **n = n-1**,
- **d** è la differenza massima accettabile, scelta a priori o individuata come la metà dell'intervallo di confidenza.

Anche in questo caso, per la relazione esistente tra **n** e i gradi di libertà **n** del **t** prescelto (**n = n-1**), si **deve procedere in modo iterativo**.

ESEMPIO 1. Ad un tecnico di laboratorio è stato chiesto di indicare, con precisione, la quantità media di un conservante (in gr.) presente per Kg di peso di un alimento. L'errore massimo accettato, lo scarto **d** rispetto alla media reale, non deve superare 1 gr. alla probabilità del 99 per cento. Dopo l'analisi di un campione pilota (12,4; 11,9 12,7 ...), che gli ha permesso di stimare la varianza ( $s^2 = 4,12$ ), quante misure deve effettuare per fornire la media con la precisione richiesta?

Risposta

La formula da utilizzare

$$n = \frac{s^2 \cdot t_{a/2,n}^2}{d^2}$$

indica le informazioni indispensabili:

- la **varianza di un campione**, che in un sondaggio preliminare è risultata  $s^2 = 4,12$
- la **differenza massima** ( $d = |\bar{x} - \mu| = 1$ ) che si vuole accettare;
- la **probabilità a** con la quale si accetta di commettere l'errore; nell'esempio, **a = 0.01 per un test bilaterale**;
- il **numero di gdl n** collegato al **valore di t**, per la probabilità prefissata; come **indicazione preliminare**, si può sempre partire da una **stima soggettiva di n** che varia **da 20 a 40**.

Sulla base delle ultime due serie di indicazioni(a,n) relative all'esempio, assumendo **n = 30** il valore di **t**

- alla probabilità **a = 0.01 bilaterale** e
- per **n = 29**

$$t_{0,005,29} = 3,038$$

è uguale a **3,038**.

Da queste quantità, si deriva una prima stima di **n**

$$n = \frac{4,12 \cdot 3,038^2}{1^2} = \frac{38,03}{1} = 38,03$$

che risulta uguale a **38,03**

differente da  $n = 30$  scelto in via preliminare.

Poiché il valore di  $t$  per  $\alpha = 0.01$  bilaterale e  $n = 38$  non è riportato nella tabella, in modo cautelativo si può prendere quello per  $n = 35$  (che ha un valore maggiore di quello per  $n = 40$ )

$$t_{0.005,35} = 2,996$$

Da esso si stima un secondo valore di  $n$

$$n = \frac{4,12 \cdot 2,996^2}{1^2} = \frac{36,98}{1} = 36,98$$

che indica in **37** il numero minimo di misurazioni da effettuare.

Poiché la seconda stima di  $n$  è molto vicina alla prima discostandosi di una quantità inferiore all'unità, si può concludere che **servono almeno 37 misurazioni**.

Il calcolo dell'intervallo fiduciale permette di condurre una **verifica della stima effettuata**.

Alla probabilità  $\alpha = 0.01$ , la media reale  $m$  è compresa entro un intervallo **I**

$$I = 0 \pm t_{\alpha/2,n} \cdot \sqrt{\frac{s^2}{n}}$$

Con

- $t_{0.005,35} = 2,996$  (poiché la tabella non riporta il valore esatto per  $n = 36$ )
- $s^2 = 4,12$  e  $n = 37$

si ottiene un intervallo fiduciale **I** intorno alla media reale  $m$

$$I = 0 \pm 2,996 \cdot \sqrt{\frac{4,12}{37}} = \pm 2,996 \cdot 0,334 = \pm 1,000664$$

uguale a **1,000664** (leggermente differente da **1,0** a causa delle approssimazioni introdotte nel calcolo).

Nel caso della **la differenza tra due medie** ( $m_A - m_B$ ), per una stima campionaria ( $x_A - x_B$ ) che abbia la precisione desiderata, comunque con uno scarto **d** non superiore al limite prefissato alla probabilità  $\alpha$ , la dimensione del campione o numero **n** di **dati per ognuno dei due campioni** è data da

$$n = \frac{2s_p^2 \cdot t_{\alpha/2,n}^2}{d^2}$$

dove

- $s_p^2$  è la varianza associata, misurata in un esperimento pilota con due campioni indipendenti,
- il valore di  $t$  deve essere preso per un **test bilaterale**, alla probabilità  $\alpha$  prefissata,
- con gdl  $n = 2(n-1)$  trattandosi di due campioni indipendenti,
- $d$  è la differenza massima accettabile, scelta a priori.

Per la dipendenza reciproca tra  $n$  e  $n$ , la stima di  $n$  avviene in modo iterativo; ma quasi sempre è sufficiente la seconda stima, presa con un arrotondamento in eccesso.

ESEMPIO 2. Per misurare l'effetto di una sostanza tossica sull'accrescimento normale di un gruppo di cavie, si intende programmare un esperimento con due gruppi: al primo è somministrata la sostanza tossica, mentre il secondo serve come controllo. Un esperimento precedente, non significativo, ha permesso di stimare la varianza associata  $s_p^2 = 0,5193$  dei due gruppi.

Che dimensioni devono avere i due gruppi, affinché una differenza  $d$  massima uguale a **0,4** risulti significativa alla probabilità del 5%?

Risposta

Assumendo, in prima approssimazione, che servano **30** dati per gruppo,

- si deve scegliere il valore di  $t$  alla probabilità  $\alpha = 0,05$  **bilaterale**, per  $n = 58$ ;
- non essendo riporta nella tabella, è possibile utilizzare il valore per  $n = 60$ , che risulta uguale a **2,000**

$$t_{0,025, 60} = 2,000$$

- con  $s_p^2 = 0,5193$  e  $d = 0,4$

Con la formula proposta in precedenza, la prima stima di  $n$

$$n = \frac{2 \cdot 0,5193 \cdot 2,000^2}{0,4^2} = \frac{4,1544}{0,16} = 25,96$$

indica **26** osservazioni per gruppo.

Poiché si discosta dalla stima preliminare (30),

dopo aver scelto il valore di  $t$  per  $n = 50$  che risulta uguale a **2,008**

$$t_{0,025, 50} = 2,008$$

si perviene alla seconda stima di  $n$

$$n = \frac{2 \cdot 0,5193 \cdot 2,008^2}{0,4^2} = \frac{4,1877}{0,16} = 26,17$$

uguale a 26,17.

Con un arrotondamento in eccesso, servono almeno 27 cavie per gruppo, 54 in totale.

Ovviamente, anche in questo caso, il calcolo dell'**intervallo fiduciale** permette di condurre una **verifica della stima effettuata**.

Alla probabilità  $\alpha = 0,05$ , la differenza reale tra le due medie ( $m_A - m_B$ ), è compresa entro un intervallo **I** intorno alla differenza campionaria  $D = \bar{x}_A - \bar{x}_B$ .

Con

$$I = 0 \pm t_{\alpha/2, n} \cdot \sqrt{s_p^2 \cdot \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}$$

e

- con  $s_p^2 = 0,5193$  e  $n = 27$
- $t_{0,025, 50} = 2,008$  ( $n = 50$  poiché nella tabella non è riportato il valore esatto per  $n = 52$ )

si ottiene un intervallo fiduciale **I**

$$I = 0 \pm 2,008 \cdot \sqrt{0,5193 \cdot \left( \frac{1}{27} + \frac{1}{27} \right)} = \pm 2,008 \cdot \sqrt{0,5193 \cdot 0,1428} = \pm 2,008 \cdot 0,196 = \pm 0,3936$$

uguale a  $\pm 0,3936$  (leggermente differente da **0,4** per le approssimazioni introdotte nel calcolo).

### 5.11. IL BILANCIAMENTO DI 2 CAMPIONI INDIPENDENTI: VANTAGGI E COSTI

Determinato il numero totale di osservazioni che servono per condurre, con la potenza o la precisione prestabilite, un esperimento con 2 campioni indipendenti, il problema successivo che spesso si pone è il loro **bilanciamento**. Si tratta di sapere se, per utilizzare al meglio i dati, sia **più conveniente formare due gruppi uguali, bilanciati, oppure due gruppi con un numero differente di dati**.

Per esempio, avendo stimato che per un confronto tra un trattamento e il controllo servono complessivamente **40** dati, è più utile fare **2** campioni con

- **20** osservazioni per ogni gruppo oppure
- utilizzarne **30** per il trattamento e solo **10** come controllo?

Vari ricercatori ritengono più importante il risultato del trattamento, l'effetto della sostanza analizzata rispetto a quello del controllo. Pensano che sia più vantaggioso costruire campioni non bilanciati, per avere più misure per il trattamento e meno per il controllo. Credono che, nell'economia complessiva

dell'esperimento, convenga misurare con precisione maggiore l'effetto del trattamento, ignoto, che non quello del controllo, spesso già noto.

In realtà il test  $t$  di Student per 2 campioni indipendenti è fondato sullo scarto tra le due medie campionarie e la varianza pooled è stimata a partire dalle due devianze; quindi i due gruppi hanno la stessa importanza per il risultato finale.

La formula del  $t$  per 2 campioni indipendenti

$$t_{N-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

fornisce gli elementi per rispondere al quesito sulla convenienza del bilanciamento.

Si può osservare che l'errore standard della differenza ( $es_d$ )

$$es_d = \frac{1}{\sqrt{s_p^2 \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

a parità di varianza pooled  $s_p^2$  raggiunge il valore minimo quando i due campioni ( $n_1 = n_2$ ) sono bilanciati. Allontanandosi dal bilanciamento, l'errore standard diventa maggiore e quindi il valore di  $t$ , a parità di tutte le altre condizioni, diventa minore e con ciò meno significativo.

**Allontanarsi da una distribuzione bilanciata dei dati, mettere più dati in un gruppo togliendoli all'altro, comporta sempre una perdita d'informazione dell'esperimento.**

E' semplice calcolare che, se si dispone in totale di 40 dati,

- con 20 dati per gruppo si moltiplica la varianza pooled per 0,1 ( $1/20 + 1/20$ ),
- con 30 dati in un gruppo e 10 nell'altro si moltiplica la varianza pooled per 0,133 ( $1/30 + 1/10$ ).

Nel secondo caso, con un errore standard maggiore, il valore di  $t$  è inferiore e quindi la differenza tra le due medie più difficilmente risulterà significativa.

Un test per due campioni indipendenti raggiunge il **massimo di potenza e robustezza quando i due gruppi sono bilanciati.** E' quindi la scelta più conveniente, in particolare quando il numero di dati a disposizione è piccolo.

Ma **non sempre è possibile o conveniente formare due campioni con lo stesso numero di dati.** A volte **i costi morali od economici** del gruppo sottoposto a trattamento e del controllo **non sono uguali.**

Supponiamo di avere un gruppo di ammalati di tumore e di dover sperimentare un farmaco che potrebbe portare alla guarigione:

- ad un primo gruppo di ammalati viene somministrato il farmaco,
- al gruppo di controllo, sempre composto da ammalati nelle stesse condizioni del primo, occorre somministrare il placebo, una sostanza apparentemente uguale al farmaco ma senza il principio attivo che, presumibilmente, aumenta la probabilità di guarigione.

Considerazioni etiche vietano l'assegnazione casuale dei pazienti a trattamenti differenziati in esperimenti clinici, quando a priori si abbiano forti ragioni per credere che uno dei due trattamenti sia indubbiamente migliore dell'altro. E' quindi moralmente doveroso ridurre il più possibile il gruppo al quale somministrare il placebo, aumentando di conseguenza il gruppo sottoposto a trattamento, per **non ridurre la potenza e la robustezza del test.**

Ma di quanto aumentare il secondo gruppo, se una distribuzione non bilanciata, a parità del numero totale di dati, comporta sempre una perdita complessiva d'informazione del test, rispetto alla distribuzione uniforme?

Con il rapporto

$$n_2 = \frac{\hat{n} \cdot n_1}{2n_1 - \hat{n}}$$

dove

- $\hat{n}$  è il numero di osservazioni stimato come necessario in ognuno dei due campioni,
- $n_1$  è il numero di osservazioni assegnate al primo gruppo,
- $n_2$  è il numero di osservazioni che devono essere assegnate secondo gruppo,

è possibile stimare quante osservazioni ( $n_2$ ) deve avere il secondo gruppo al diminuire delle osservazioni ( $n_1$ ) del primo, una volta che sia stata stimata la loro dimensione minima ( $\hat{n}$ ), per mantenere costante la potenza del test.

ESEMPIO 1. Si supponga che, per sperimentare un farmaco antitumorale, sia stato calcolato che servono due gruppi di persone, composti ognuno da almeno 15 individui ( $\hat{n} = 15$ ).

Per l'alto costo morale insito nella somministrazione del placebo ad ammalati, si vuole ridurre il numero di persone di questo gruppo. Mettendo nel gruppo trattato con placebo 10 individui ( $n_1 = 10$ ), quanti dovranno essere gli individui trattati con il farmaco ( $n_2$ )?

Risposta.

Applicando la formula

$$n_2 = \frac{\hat{n} \cdot n_1}{2n_1 - \hat{n}}$$

con

-  $\hat{n} = 15$  e  $n_1 = 10$ ,

il valore di  $n_2$

$$n_2 = \frac{15 \cdot 10}{2(10) - 15} = \frac{150}{5} = 30$$

risulta uguale a 30 individui.

Si osservi che **le dimensioni complessive dell'esperimento aumentano**: in totale si richiedono non più **30** (15 + 15) persone ma **40** (10 + 30).

Il vantaggio consiste solo nella **diminuzione del gruppo che ha costi morali alti**. Ma è applicabile solo quando, come spesso succede, **il numero di persone ammalate che vogliono farsi curare è più alto del numero minimo stimato come necessario** per l'esperimento. Nessuno obietta al fatto che il numero di persone che vengono seguite nel decorso della malattia, alla quali distribuire gratuitamente il farmaco e da visitare periodicamente, sia alto. L'unico problema può essere rappresentato dai costi dell'operazione, per l'aumento del numero totale di ammalati.

Il numero di persone al quale somministrare il placebo può **diminuire ulteriormente**.

Ad esempio, abbassandolo a 8 persone ( $n_1 = 8$ ),

$$n_2 = \frac{15 \cdot 8}{2 \cdot 8 - 15} = \frac{120}{1} = 120$$

il numero di persone ai quali somministrare il farmaco sale a 120 ( $n_2 = 120$ ).

Tuttavia, come si può dimostrare con facilità, **non è conveniente scendere oltre una certa soglia**.

Ad esempio, con 7 persone nel primo gruppo ( $n_1 = 7$ )

$$n_2 = \frac{15 \cdot 7}{2 \cdot 7 - 15} = \frac{105}{-1} = ?$$

**$n_2$  diverrebbe negativo.**

E' una **situazione da evitare**, anche utilizzando un secondo campione molto più grande di quello stimato in precedenza: **differenze importanti nelle dimensioni dei due campioni determinano**

**alterazioni profonde, non stimate, della robustezza e della potenza del test**, con conseguenze rilevanti

- sia sulla probabilità a
- sia della probabilità b.

Anche nella ricerca ambientale sovente si pongono problemi etici. Quando occorre condurre esperimenti su animali per valutare effetti di sostanze tossiche, qualcuno deve essere sacrificato.

Soprattutto occorre **attenzione ai problemi economici**, al costo dell'esperimento, per ottenere il risultato massimo dal finanziamento assegnato. Le dimensioni minime dell'esperimento rappresentano una prima risposta, in quanto permettono di evitare analisi con campioni troppo piccoli (quindi inutili ai fini del risultato) oppure con campioni molto più grandi di quanto necessario (quindi con spreco dell'investimento in risorse, persone e tempo).

Una seconda serie di problemi deriva dal fatto che, in varie situazioni, anche **i costi economici sono molto diversi tra le osservazioni del trattamento e quelle del controllo**. Le cavie sottoposte all'effetto di una sostanza tossica hanno probabilità maggiori di morire o comunque di non essere più utilizzabili per un secondo esperimento, rispetto al controllo. Confrontando livelli d'inquinamento di due laghi, uno potrebbe distare maggiormente dalla sede del ricercatore e quindi richiedere spese maggiori per il campionamento; analisi di laboratorio condotte con metodologie diverse che richiedano tempi differenti quasi sempre hanno costi diversi; il farmaco e il principio attivo costano più del placebo.

La risposta ottimale a questa serie di problemi è leggermente più complessa della scelta qualitativa legata ai problemi etici; ma è possibile **determinare il rapporto numerico ottimale tra i due gruppi, per ridurre i costi complessivi dell'esperimento**.

Ad esempio, se ogni prova sperimentale con il trattamento costa 10.000 lire e una con il controllo 2.000 lire, nell'esperimento bilanciato di 15 osservazioni per gruppo

il costo complessivo sarebbe

$$15 \times 10.000 + 15 \times 2.000 = 180.000$$

di 180.000 lire .

L'esperimento con 10 osservazioni nel controllo e 30 nel trattamento (che ha la stessa potenza)

avrebbe un costo di

$$10 \times 10.000 + 30 \times 2.000 = 160.000$$

di 160.000 lire.

Infine l'esperimento con 8 osservazioni per il trattamento e 120 per il controllo vedrebbe aumentare i costi a

$$8 \times 10.000 + 120 \times 2.000 = 320.000$$

320.000 lire.

Ovviamente, con rapporti diversi tra i costi delle due rilevazioni la risposta sarebbe stata differente.

Per trovare la soluzione ottimale, evitando metodi più complessi, la procedura può essere semplice:

- dopo aver calcolato le dimensioni minime necessarie  $\hat{n}$ , **per ognuno dei due gruppi**,
- si stimano le coppie di osservazioni ( $n_1$  e  $n_2$ ) che hanno la stessa potenza di quella bilanciata,
- senza superare il valore soglia, rappresentato da una differenza  $2n_1 - \hat{n}$  **negativa**.
- Delle possibili pianificazioni accettabili, si **stimano i costi**, scegliendo quella **più conveniente**,
- ricordando che, dal punto di vista statistico, **è sempre preferibile utilizzare campioni con dimensioni non troppo differenti**.

## 5.12 POTENZA A PRIORI E A POSTERIORI DEL TEST F PER L'OMOSCHEDASTICITA'

Anche quando si ricorre all'uso del test **F** cioè al rapporto tra

la varianza maggiore ( $s_{\max}^2$ ) e la varianza minore ( $s_{\min}^2$ )

$$F = \frac{s_{\max}^2}{s_{\min}^2}$$

per verificare l'ipotesi di omoschedasticità tra due gruppi (A e B),

con

$$H_0: \sigma_A^2 = \sigma_B^2$$

ed ipotesi alternativa **bilaterale**

$$H_1: \sigma_A^2 \neq \sigma_B^2$$

è possibile chiedersi:

**“Quanti dati servono affinché l'ipotesi nulla possa essere respinta alla probabilità  $\alpha$ , con una probabilità  $\beta$  di commettere un errore di II Tipo?”**

Secondo quanto riportato nel testo di M. M. **Desu** e D. **Raghavarao** del 1990 (*Sample Size Methodology*, Academic Press, Boston, Massachusetts, 135 pp.) e ripreso da Jerrold H. **Zar** nel testo del 1999 *Biostatistical Analysis* (fourth edition, Prentice Hall, Upper Saddle River, New Jersey), è possibile utilizzare l'approssimazione alla distribuzione normale, per cui **in ognuno dei due campioni** il numero minimo **n** di dati è

$$n \geq \left[ \frac{z_a + z_b}{\ln\left(\frac{s_{\max}^2}{s_{\min}^2}\right)} \right]^2 + 2$$

dove

- $z_a$  = il valore di  $z$  alla probabilità  $a$  in una distribuzione **bilaterale**,
- $z_b$  = il valore di  $z$  alla probabilità  $b$  in una distribuzione **unilaterale**,
- **ln** = logaritmo naturale o neperiano,
- $s_{\max}^2$  = valore della varianza maggiore stimata in uno studio pilota,
- $s_{\min}^2$  = valore della varianza minore stimata nello stesso studio pilota.

Volendo utilizzare il logaritmo a base 10 (**log**) al posto del logaritmo naturale (**ln**) l'equazione diventa

$$n \geq \left[ \frac{z_a + z_b}{2,30259 \log\left(\frac{s_{\max}^2}{s_{\min}^2}\right)} \right]^2 + 2$$

Nel caso (più raro nella ricerca ambientale) in cui si voglia, con un test **unilaterale**, dimostrare che la varianza di un gruppo di osservazioni è maggiore di quella dell'altro gruppo, per cui l'ipotesi alternativa  $H_1$  è

$$s_A^2 > s_B^2 \quad \text{oppure} \quad s_A^2 < s_B^2$$

nelle formule precedenti

- la varianza ipotizzata come maggiore, che deve risultare effettivamente tale, va posta al numeratore,
- $z_a$  = il valore di  $z$  alla probabilità  $a$  in una distribuzione **unilaterale**.

Come già rilevato in precedenza nell'analisi della omoschedasticità con un test unilaterale, **questo test ha significato solo se il rapporto tra le due varianze a confronto risulta maggiore di 1.**

Il valore  $n$  stimato rappresenta il numero di osservazioni necessario **in ognuno dei due campioni, assunti come uguali.**

In alcune condizioni, per il diverso costo delle osservazioni nei due gruppi, si può pensare che i gdl di un gruppo ( $n_A$ ) siano  $m$  volte maggiori dei gdl dell'altro gruppo ( $n_B$ ).

Secondo i due testi citati,

$$\mathbf{m} = \frac{n_A - 1}{n_B - 1} = \frac{\mathbf{n}_A}{\mathbf{n}_B}$$

Avendo ricavato  $\mathbf{n}$  dalle formule precedenti e prefissato  $\mathbf{m}$  sulla base dei costi, si ricava prima  $\mathbf{n}_B$

$$\mathbf{n}_B = \frac{(m+1) \cdot (n-2)}{2m} + 2$$

e da esso  $\mathbf{n}_A$

$$\mathbf{n}_A = m \cdot (\mathbf{n}_B - 1) + 1$$

La stima di  $\mathbf{n}$ , il numero di dati necessari a rendere significativo il test, è chiamata **potenza a priori**.

Spesso, effettuato un test, soprattutto quando non risulti significativo è utile chiedersi quale era la probabilità che esso lo risultasse: è la **potenza a posteriori** o, più semplicemente, **la potenza del test**.

Se due campioni hanno lo **stesso numero di dati** ( $n = \mathbf{n}_A = \mathbf{n}_B$ ), il valore di  $\mathbf{z}_b$  (sempre in una **distribuzione unilaterale, per stimare la potenza sia a priori che a posteriori**)

- usando il logaritmo naturale (**ln**) è

$$\mathbf{z}_b = \sqrt{n-2} \cdot \ln \left( \frac{s_{\max}^2}{s_{\min}^2} \right) - \mathbf{z}_a$$

- usando il logaritmo a base 10 (**log**) è

$$\mathbf{z}_b = \sqrt{n-2} \cdot 2,30259 \log \left( \frac{s_{\max}^2}{s_{\min}^2} \right) - \mathbf{z}_a$$

dove

-  $\mathbf{z}_a$  deve essere preso in una distribuzione bilaterale oppure unilaterale, in rapporto a quanto esplicitato nell'ipotesi  $\mathbf{H}_1$ .

Se due campioni hanno un **numero differente di dati** ( $\mathbf{n}_A \neq \mathbf{n}_B$ ), si deve introdurre un **termine di correzione** alla dimensione  $\mathbf{n}$ : il nuovo valore risulterà inferiore alla medie aritmetica tra  $\mathbf{n}_A$  e  $\mathbf{n}_B$

Dopo aver stimato  $\mathbf{m}$  con

$$\mathbf{m} = \frac{n_{\max} - 1}{n_{\min} - 1}$$

dove

-  $n_{\max}$  è il numero di dati del gruppo con **varianza ( $s^2$ ) maggiore** (non del gruppo con più dati)

- $n_{\min}$  è il numero di dati del gruppo con **varianza minore** (non del gruppo più piccolo) nelle formule precedenti al posto di

$$\sqrt{n-2}$$

si introduce

$$\sqrt{\frac{2m(n_B - 2)}{m + 1}}$$

dove

- $n_B$  = numero di dati della varianza al denominatore

Di conseguenza, il valore di  $z_b$  (sempre in una **distribuzione unilaterale**)

- usando il logaritmo naturale (**ln**) è

$$z_b = \sqrt{\frac{2m \cdot (n_B - 2)}{m + 1}} \cdot \ln\left(\frac{s_{\max}^2}{s_{\min}^2}\right) - z_a$$

- usando il logaritmo a base 10 (**log**) è

$$z_b = \sqrt{\frac{2m \cdot (n_B - 2)}{m + 1}} \cdot 2,30259 \log\left(\frac{s_{\max}^2}{s_{\min}^2}\right) - z_a$$

dove

- $z_a$  deve essere preso in una distribuzione bilaterale oppure unilaterale, in rapporto a quanto esplicitato nell'ipotesi **H<sub>1</sub>**

#### ESEMPIO

Per applicare il test t di Student su due campioni indipendenti, si richiede l'omoschedasticità della varianza. In uno studio preliminare, sui campione A e B si sono ottenuti i seguenti risultati

Campioni	X <sub>A</sub>	X <sub>B</sub>
N	12	13
Varianza (s <sup>2</sup> )	0,015	0,0314

con i quali

$$F_{(12,11)} = \frac{0,0314}{0,015} = 2,093$$

non era stata rifiutata l'ipotesi nulla, poiché il loro rapporto è uguale a 2,093 mentre il valore critico di  $F_{(12,11)}$  per il livello di probabilità  $\alpha = 0.05$  è 2,79.

A) Quanti dati sarebbero stati necessari per dimostrare che le due varianze sono significativamente differenti

- alla probabilità  $\alpha = 0.05$
- con un rischio d'errore di II tipo ( $\beta$ ) = 0.10 (o una potenza  $1-\beta = 0.90$ )?

B) Quale è la potenza del test effettuato?

Risposte

A) E' un test bilaterale, in cui

- per  $\alpha = 0.05$  in una distribuzione **bilaterale** il valore di  $z$  è 1,96
- per  $\beta = 0.10$  in una distribuzione **unilaterale** il valore di  $z$  è 1,28

con

$$n \geq \left[ \frac{z_a + z_b}{\ln\left(\frac{s_{\max}^2}{s_{\min}^2}\right)} \right]^2 + 2$$

si ottiene

$$n \geq \left( \frac{1,96 + 1,28}{\ln\left(\frac{0,0314}{0,015}\right)} \right)^2 + 2 = \left( \frac{3,24}{0,7386} \right)^2 + 2 = 4,3867^2 + 2 = 19,24 + 2 = 21,24$$

un numero minimo di osservazioni pari a 22 per ognuno dei due gruppi.

Anche con questo semplice esempio si dimostra quanto fosse errato concludere che le due varianze erano simili, perché non si poteva rifiutare l'ipotesi nulla alla probabilità del 5%: sarebbe stato sufficiente raddoppiare il numero di osservazioni per arrivare, con una probabilità del 90 per cento, alla conclusione opposta.

B) Con due campioni sbilanciati, in cui il numero di dati della varianza al numeratore è 13 e quelli della varianza al denominatore è 12,

$$m = \frac{13-1}{12-1} = 1,091$$

il valore di  $z_b$

$$z_b = \sqrt{\frac{2m \cdot (n_B - 2)}{m + 1}} \cdot \ln\left(\frac{s_{\max}^2}{s_{\min}^2}\right) - z_a$$

con

- $m = 1,091$
- il valore di  $z$  alla probabilità  $\alpha = 0.05$  in una distribuzione **bilaterale** uguale a 1,96
- $n_B = 12$

risulta

$$z_b = \sqrt{\frac{2 \cdot 1,091 \cdot (12 - 2)}{1,091 + 1}} \cdot \ln\left(\frac{0,0314}{0,015}\right) - 1,96$$

$$z_b = \sqrt{\frac{2,182 \cdot 10}{2,091}} \cdot \ln 2,093 - 1,96 = \sqrt{10,43} \cdot 0,7386 - 1,96 = 2,385 - 1,96 = 0,425$$

uguale a **0,425**.

In una distribuzione unilaterale, ad esso corrisponde un valore di  $b$  collocato tra 0,337 e 0,334. Approssimativamente, la probabilità che il test non risultasse significativo era del 33,5% e la potenza del test era pari a 0,665.

### 5.13. CORREZIONE PER IL CAMPIONAMENTO IN UNA POPOLAZIONE FINITA E IL CONCETTO DI SUPERPOPOLAZIONE

Nel modello classico dell'inferenza statistica, **la popolazione è composta da un numero infinito di dati**; per quanto grande, **il campione da essa estratto ne rappresenta una frazione prossima allo zero**.

In vari settori della ricerca applicata, succede che il numero di individui che formano la popolazione, oggetto dell'indagine, in realtà sia composto solo da alcune decine di individui, al massimo da poche centinaia. In biologia e in medicina, è il caso dei portatori di malattie rare: poche decine di persone in una regione, difficilmente rintracciabili nella loro totalità. Nella ricerca ambientale, è il caso dei laghi di una provincia, delle discariche di una regione, dei pozzi d'acqua potabile gestiti da un acquedotto. Per ridurre i costi e i tempi, per superare la difficoltà oggettiva di rintracciare tutti i soggetti utili all'analisi, può essere conveniente limitare i prelievi ad un campione, ovviamente con la possibilità di estendere i risultati a tutto l'universo, mediante l'inferenza.

In termini tecnici, **quando il campione di  $n$  osservazioni rappresenta una quota non trascurabile della popolazione di  $N$  individui, per convenzione consolidata maggiore del 5%, si parla di campionamento in una popolazione finita** (*sampling a finite population*).

Per quanto attiene la **media**, è intuitivo che

**quanto più  $n$  si avvicina ad  $N$ , tanto più  $\bar{x}$  si avvicina a  $m$ .**

Nel caso in cui si analizzino tutti gli individui che compongono la popolazione, si afferma che

$$\bar{x} = m \quad \text{se} \quad n = N$$

Di conseguenza, l'errore standard (es)

della media campionaria  $\bar{x}$

$$es = \frac{s}{\sqrt{n}}$$

deve essere ridotto, mediante

la **correzione per la popolazione finita** (*finite population correction*),

Nelle tre formule analoghe, più frequentemente citate nei testi

l'errore standard diviene

$$es = \sqrt{\frac{s^2}{n}} \cdot \sqrt{1 - \frac{n}{N}} = \sqrt{\left(\frac{s^2}{n}\right) \cdot \left(1 - \frac{n}{N}\right)} = \sqrt{\left(\frac{s^2}{n}\right) \cdot \left(\frac{N-n}{N}\right)}$$

dove

- $\frac{n}{N}$  è la frazione campionata
- $\sqrt{1 - \frac{n}{N}}$  oppure  $\sqrt{\frac{N-n}{N}}$  è la correzione per la popolazione finita.

Anche dalle formule, è semplice dedurre che

- quando  $n$  è uguale a  $N$ , **l'errore standard  $es$  diventa 0**; cioè
- **la media campionaria  $\bar{x}$  coincide con la media reale  $m$** ; di conseguenza, **la sua dispersione, misurata appunto dall'errore standard, è nulla.**

Pure nel caso di una popolazione finita, è possibile stimare **quanti dati sono necessari, per ottenere la media con la precisione prefissata**. Secondo il metodo proposto da **W. G. Cochran** nel 1977, (vedi: *Sampling Techniques*, 3<sup>rd</sup> ed. John Wiley, New York, 428 pp. pagg. 77-78)

- dapprima si stima  $n'$  come se il calcolo fosse riferito ad una popolazione infinita,

$$n' = \frac{s^2 \cdot t_{\alpha/2n}^2}{d^2}$$

(vedi paragrafo 10 ed esempio 1 di questo capitolo)

- successivamente si passa alla sua stima corretta  $n$   
con il rapporto

$$n = \frac{n'}{1 + \frac{n'}{N}}$$

dove  $N$  è il numero di individui che formano la popolazione.

ESEMPIO 1. In molte zone, durante gli ultimi decenni la qualità delle acque sotterranee è peggiorata, per la diffusione sempre maggiore di fonti d'inquinamento, quali insediamenti civili e industriali, allevamenti zootecnici, pratiche agricole intensive, scarico di reflui. La concentrazione in nitrati spesso supera il limite di legge imposto per le acque potabili, fissato dal DPR 236/88 in 50 mg/l; quasi sempre supera la soglia d'attenzione.

Una analisi dettagliata, condotta su tutti i 36 pozzi che alimentano l'acquedotto di una città, ha dato un valore medio pari a 47,8 mg/l. Dopo un'azione di risanamento, basata sul controllo degli allevamenti zootecnici e lo scarico di reflui, a distanza di tempo con una indagine rapida sono stati prelevati campioni da 12 pozzi, sui 34 attivi: la loro media è risultata pari a 47,1 mg/l con una varianza uguale a 2,54.

A) Si può dimostrare che la quantità media di nitrati è diminuita?

B) Quale è l'intervallo fiduciale della media della seconda rilevazione alla probabilità del 95%?

Risposta

A) La quantità media di nitrati (mg/l) calcolata con un prelievo in tutti i 36 pozzi è la media reale, riferita alla prima rilevazione:  $m = 47,8$

La seconda analisi condotta su 12 pozzi dei 34 attivi ha fornito

- una media campionaria  $\bar{x} = 47,1$  con varianza  $s^2 = 2,54$
- che rappresenta solo una stima della media reale  $m_0$  alla seconda rilevazione.

Per verificare se la quantità media di nitrati è diminuita, quindi per scegliere tra l'ipotesi nulla

$$H_0: m_0 = m \quad \text{oppure} \quad H_0: m_0 = 47,8$$

e l'ipotesi alternativa unilaterale

$$H_1: m_0 < m \quad \text{oppure} \quad H_1: m_0 < 47,8$$

si dovrebbe applicare il test  $t$  di Student per un campione

$$t_{(n-1)} = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}}$$

se esso fosse stato estratto da una popolazione teoricamente infinita.

Ma, trattandosi di un campione di 12 unità estratto da una popolazione di soli 34 pozzi, si deve apportare la correzione per una popolazione finita.

Pertanto, il test diventa

$$t_{(n-1)} = \frac{\bar{x} - m}{\sqrt{\frac{s^2}{n} \cdot \sqrt{1 - \frac{n}{N}}}}$$

Con i dati dell'esempio,

- se la popolazione fosse stata infinita

$$t_{11} = \frac{47,1 - 47,8}{\sqrt{\frac{2,54}{12}}} = \frac{-0,7}{0,46} = -1,522$$

si sarebbe ottenuto un valore di  $t = -1,522$  con 11 gdl

- con un popolazione composta in totale di 34 individui

$$t_{11} = \frac{47,1 - 47,8}{\sqrt{\frac{2,54}{12} \cdot \sqrt{1 - \frac{12}{34}}}} = \frac{-0,7}{0,46 \cdot 0,804} = \frac{-0,7}{0,37} = -1,892$$

si ottiene un valore di  $t = -1,892$  con 11 gdl.

Poiché il valore di  $t$  con **11** gdl, per un **test unilaterale**

- alla probabilità  $\alpha = 0.05$  è uguale a **1,7959**
- alla probabilità  $\alpha = 0.01$  è uguale a **2,7181**

con la correzione per una popolazione finita, il test è significativo ad una probabilità  $\alpha < 0.05$ .

B) Per stimare l'intervallo fiduciale entro il quale, alla probabilità prefissata, si trova la media reale  $m$  della seconda rilevazione, si deve utilizzare

$$m = \bar{x} \pm t_{\alpha/2, n} \cdot \sqrt{\frac{s^2}{n} \cdot \sqrt{1 - \frac{n}{N}}}$$

Con i dati dell'esempio e con la solita simbologia,

poiché il valore di  $t$  per un test bilaterale, alla probabilità **0.05** e con **11** gdl, è uguale a **2,201** si stima che la media reale  $m$  si troverà

$$m = 47,1 \pm 2,201 \cdot \sqrt{\frac{2,54}{12}} \cdot \sqrt{1 - \frac{12}{34}}$$
$$\mathbf{n} = 47,1 \pm 2,201 \cdot 0,46 \cdot 0,804 = \pm 0,814$$

entro un intervallo di più o meno 0,814 intorno alla media campionaria di 47,1.

Poiché la media precedente, uguale a 47,8 è compresa entro questo intervallo, tra le due medie non esiste una differenza significativa, se si conduce un test bilaterale.

Se anche la seconda analisi fosse stata condotta su tutti i 34 pozzi che al momento formavano la popolazione, avremmo ottenuto direttamente il suo valore  $m$ . Per verificare una diminuzione della quantità media di nitrati rispetto alla prima rilevazione, sarebbe sufficiente il solo confronto tra le due medie, poiché il loro errore standard sarebbe uguale a 0.

ESEMPIO 2. Se il ricercatore avesse voluto dimostrare significativa una diminuzione di **0,5** mg/l alla probabilità  $\alpha = \mathbf{0.01}$  quanti pozzi avrebbe dovuto campionare dei **34** che formano la popolazione?

Risposta

Dalla formula generale

$$n' = \frac{s^2 \cdot t_{\alpha/2, n}^2}{d^2}$$

in cui

-  $d = 0,5$  e  $s^2 = 2,54$

- poiché  $t$  per un **test unilaterale** alla probabilità  $\alpha = \mathbf{0.01}$  con  $n$  scelto, in prima approssimazione, uguale a **30** è uguale a **2,4573**

si ottiene

$$n' = \frac{2,54^2 \cdot 2,4573^2}{0,5^2} = \frac{6,4516 \cdot 6,0383}{0,25} = \frac{38,96}{0,25} = 155,84$$

una prima stima  $\mathbf{n'} = 155,84$  cioè 156 dati.

Essendo molto diverso dai **31** preventivati con  $n = \mathbf{30}$ ,

- dopo aver scelto il nuovo valore di **t** che per un **test unilaterale** alla probabilità **a = 0.01** con **n = 150** (poiché le stime diventano approssimate e i valori di **t** tra loro sono molto simili) è uguale a **2,3515**

si effettua una seconda stima

$$n' = \frac{2,54^2 \cdot 2,3515^2}{0,5^2} = \frac{6,4516 \cdot 5,5296}{0,25} = \frac{35,67}{0,25} = 142,68$$

che conduce a **n' = 142,68** cioè **143** dati.

Poiché l'ultimo valore stimato (143) è differente dal precedente (156), si può effettuare un terzo tentativo, che ha significato solamente per **n = 140**; infatti se scegliessimo lo stesso valore di **t** (2,3515) precedente per **n = 150** otterremmo la stessa stima di **n'**.

Con un valore di **t**, per un **test unilaterale** alla probabilità **a = 0.01** con **n = 140** uguale a **2,3533** si perviene alla terza stima

$$n' = \frac{2,54^2 \cdot 2,3533^2}{0,5^2} = \frac{6,4516 \cdot 5,5380}{0,25} = \frac{35,78}{0,25} = 143,12$$

dove

- **n' = 143,12** cioè 144 dati, confermando la seconda stima.

Questo valore è chiaramente assurdo, date le dimensioni complessive della popolazione (**N = 34**); comunque sia il suo valore, **n'** deve essere corretto, essendo riferito ad una popolazione finita.

Con

$$n = \frac{n'}{1 + \frac{n'}{N}}$$

si ottiene

$$n = \frac{144}{1 + \frac{144}{34}} = \frac{144}{5,24} = 27,48$$

una stima di **n** uguale a **27,48** arrotondato a **29** dati

Se la domanda fosse stata: “Quanti dati servono per misurare la media della seconda rilevazione con un errore massimo di 0,5mg/l?” la procedura sarebbe stata identica a quella appena illustrata, però

utilizzando un valore di  $t$  bilaterale. Con il numero di dati stimati, alle stesse condizioni l'intervallo fiduciale sarebbe risultato alla differenza stimata.

L'approccio presentato è quello classico.

I concetti e i calcoli illustrati portano alla conclusione che, quando l'analisi è estesa all'universo dei dati, cadono

- sia il concetto di "significatività statistica"
- sia quello di distribuzione di probabilità.

**L'inferenza diviene teoricamente superflua, poiché i valori e le differenze riscontrate sui dati dell'universo non hanno bisogno d'inferenza**, essendo quelli reali o della popolazione: sono significativi per definizione, per quanto essi siano piccoli.

Tuttavia, vari statistici enunciano concetti differenti. A loro parere, ai fini dell'analisi statistica spesso risulta utile ed opportuno **considerare i dati della popolazione come frutto di un campionamento casuale semplice di una superpopolazione**. Il gruppo può cambiare continuamente nel tempo, come la comunità di un paese o gli animali che vivono in un territorio, per nuove nascite, decessi e migrazioni. I pozzi che formano l'universo di quelli utilizzati in certo momento dall'acquedotto non sono sempre gli stessi, potendo essere ciclicamente sostituiti, con la disattivazione di alcuni e l'attivazione di altri.

**Gli individui che formano una popolazione finita possono essere considerati come un campione casuale di una superpopolazione di numerosità ignota**. Sulla base di tale assunzione, questi statistici ritengono che tutti gli strumenti inferenziali ritrovano il loro significato corretto, anche in una popolazione finita. La precedente correzione per campioni estratti da popolazioni finite sarebbe superflua o addirittura errata.

**L'idea di superpopolazione è un artificio** che ad alcuni non appare in contrasto con considerazioni statistiche sostanziali; è adottato con frequenza, tanto che **quando si utilizzano i dati di tutta la popolazione da molti è comunemente accettato il ricorso a test statistici con le formule presentate nella prima parte del capitolo, senza necessità di ricorrere alle correzioni proposte in questo paragrafo**.

I testi classici consigliano la correzione, come presentata in questo paragrafo.

#### **5.14. TEST PER LA DIFFERENZA TRA DUE COEFFICIENTI DI VARIAZIONE CON DISTRIBUZIONE $z$ OPPURE DISTRIBUZIONE $t$ DI STUDENT**

Come in vari settori della statistica applicata, spesso nella ricerca ambientale sullo stesso prelievo vengono effettuate varie misure. Ad esempio, nella composizione dell'aria si misurano le ppm (parti

per milione) di Azoto, Ossigeno, Argon, Biossido di Carbonio, Neon, Elio, Metano, Krypton, Idrogeno; nell'analisi dell'inquinamento atmosferico, si rilevano le particelle totali sospese, i composti dello Zolfo, i composti dell'Azoto, quelli inorganici del Carbonio, quelli organici volatili,...

Oltre a confronti tra i valori medi e tra le varianze, trattandosi di fenomeni che hanno dimensioni diverse ( ad esempio, a livello del mare e lontano da sorgenti inquinanti, la composizione dell'aria secca è 780.900 ppm per l'Azoto, 9.300 ppm per l'Argon, 18 ppm per il Neon, 0,5 ppm per l'Idrogeno,...) può essere utili il confronto tra due coefficienti di variazione, appunto per valutare la variabilità in rapporto alla dimensione media del fenomeno.

Ricordando (dal cap. I) che il coefficiente di variazione di un campione di dati (CV) è

$$CV = \left( \frac{s}{\bar{X}} \right) \cdot 100$$

dove

s = deviazione standard del campione

$\bar{X}$  = media del campione

e che, quando il numero di osservazioni è limitato, è proposta una correzione di una quantità  $1/4N$ , dove N è il numero di osservazioni del campione, per cui

**il coefficiente di variazione corretto CV'** diventa

$$CV' = CV \left( 1 + \frac{1}{4N} \right)$$

è possibile valutare se il tra due coefficienti di variazione (indicati in  $V_A$  e  $V_B$ ), ovviamente campionari, esiste una differenza significativa.

Il test , come nel confronto tra due medie, può essere **bilaterale** oppure **unilaterale**, è può utilizzare la **distribuzione z** oppure la **distribuzione t di Student**, in rapporto alle dimensioni dei due campioni.

Se i dati sono distribuiti in modo normale (senza trasformazione), secondo G. E. Miller (vedi articolo del 1991 *Asymptotic test statistics for coefficients of variation*, pubblicato su *Communic. Statist. – Theor. Meth.* 20: 2251 – 2262), si può utilizzare il valore di **z**

$$z = \frac{V_A - V_B}{\sqrt{\left( \frac{V_p^2}{n_A - 1} + \frac{V_p^2}{n_B - 1} \right) \cdot (0,5 + V_p^2)}}$$

dove

-  $V_p^2$  è il quadrato del coefficiente di variazione pooled  $V_p$  stimato come

$$V_p = \frac{(n_A - 1) \cdot V_A + (n_B - 1) \cdot V_B}{(n_A - 1) + (n_B - 1)}$$

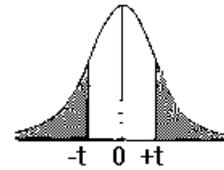
Accettato generalmente per grandi campioni, il test da alcuni è esteso anche al confronto tra campioni di dimensioni ridotte, ricorrendo al valore di  $t$  con gdl  $N - 2$

$$t_{(N-2)} = \frac{V_A - V_B}{\sqrt{\left( \frac{V_p^2}{n_A - 1} + \frac{V_p^2}{n_B - 1} \right) \cdot (0,5 + V_p^2)}}$$

dove

-  $N$  è il numero totale di dati dei due campioni.

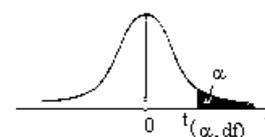
**Valori critici della distribuzione t di Student per un test bilaterale**



Gradi Di Libertà	a								
	0.500	0.400	0.200	0.100	0.050	0.025	0.010	0.005	0.001
1	1.000	1.376	3.078	6.314	12.706	25.452	63.657		
2	.816	1.061	1.886	2.920	4.303	6.205	9.925	14.089	31.598
3	.765	0.978	1.638	2.353	3.182	4.176	5.841	7.453	12.941
4	.741	.941	1.533	2.132	2.776	3.495	4.604	5.598	8.610
5	.727	.920	1.476	2.015	2.571	3.163	4.032	4.773	6.859
6	.718	.906	1.440	1.943	2.447	2.969	3.707	4.317	5.959
7	.711	.896	1.415	1.895	2.365	2.841	3.499	4.029	5.405
8	.706	.889	1.397	1.860	2.306	2.752	3.355	3.832	5.041
9	.703	.883	1.383	1.833	2.262	2.685	3.250	3.690	4.781
10	.700	.879	1.372	1.812	2.228	2.634	3.169	3.581	4.587
11	.697	.876	1.363	1.796	2.201	2.593	3.106	3.497	4.437
12	.695	.873	1.356	1.782	2.179	2.560	3.055	3.428	4.318
13	.694	.870	1.350	1.771	2.160	2.533	3.012	3.372	4.221
14	.692	.868	1.345	1.761	2.145	2.510	2.977	3.326	4.140
15	.691	.866	1.341	1.753	2.131	2.490	2.947	3.286	4.073
16	.690	.865	1.337	1.746	2.120	2.473	2.921	3.252	4.015
17	.689	.863	1.333	1.740	2.110	2.458	2.898	3.222	3.965
18	.688	.862	.330	1.734	2.101	2.445	2.878	3.197	3.922
19	.688	.861	1.328	1.729	2.093	2.433	2.861	3.174	3.883
20	.687	.860	1.325	1.725	2.086	2.423	2.845	3.153	3.850
21	.686	.859	1.323	1.721	2.080	2.414	2.831	3.135	3.819
22	.686	.858	1.321	1.717	2.074	2.406	2.819	3.119	3.792
23	.685	.858	1.319	1.714	2.069	2.398	2.807	3.104	3.767
24	.685	.857	1.318	1.711	2.064	2.391	2.797	3.090	3.745
25	.684	.856	1.316	1.708	2.060	2.385	2.787	3.078	3.725
26	.684	.856	1.315	1.706	2.056	2.379	2.779	3.067	3.707
27	.684	.855	1.314	1.703	2.052	2.373	2.771	3.056	3.690
28	.683	.855	1.313	1.701	2.048	2.368	2.763	3.047	3.674
29	.683	.854	1.311	1.699	2.045	2.364	2.756	3.038	3.659
30	.683	.854	1.310	1.697	2.042	2.360	2.750	5.030	3.646
35	.682	.852	1.306	1.690	2.030	2.342	2.724	2.996	3.591
40	.681	.851	1.303	1.684	2.021	2.329	2.704	2.971	3.551
45	.680	.850	1.301	1.680	2.014	2.319	2.690	2.952	3.520
50	.680	.849	1.299	1.676	2.008	2.310	2.678	2.937	3.496
55	.679	.849	1.297	1.673	2.004	2.304	2.669	2.925	3.476
60	.679	.848	1.296	1.671	2.000	2.299	2.660	2.915	3.460
70	.678	.847	1.294	1.667	1.994	2.290	2.648	2.899	3.435
80	.678	.847	1.293	1.665	1.989	2.284	2.638	2.887	3.416
90	.678	.846	1.291	1.662	1.986	2.279	2.631	2.878	3.402
100	.677	.846	1.290	1.661	1.982	2.276	2.625	2.871	3.390
120	.677	.845	1.289	1.658	1.980	2.270	2.617	2.860	3.373
∞	.6745	.841	1.28161	1.6448	1.9600	2.2414	2.5758	2.8070	3.2905

Valori critici della distribuzione t di Student per un test unilaterale

(prima parte)

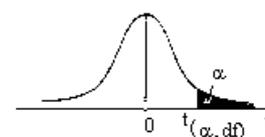


Gradi Di Libertà	Aree della coda superiore					
	0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5177	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.6800	1.3006	1.6794	2.0141	2.4121	2.6896
46	0.6799	1.3002	1.6787	2.0129	2.4102	2.6870
47	0.6797	1.2998	1.6779	2.0117	2.4083	2.6846
48	0.6796	1.2994	1.6772	2.0106	2.4066	2.6822
49	0.6795	1.2991	1.6766	2.0096	2.4049	2.6800
50	0.6794	1.2987	1.6759	2.0086	2.4033	2.6778
51	0.6793	1.2984	1.6753	2.0076	2.4017	2.6757
52	0.6792	1.2980	1.6747	2.0066	2.4002	2.6737
53	0.6791	1.2977	1.6741	2.0057	2.3988	2.6718
54	0.6791	1.2974	1.6736	2.0049	2.3974	2.6700
55	0.6790	1.2971	1.6730	2.0040	2.3961	2.6682
56	0.6789	1.2969	1.6725	2.0032	2.3948	2.6665
57	0.6788	1.2966	1.6720	2.0025	2.3936	2.6649
58	0.6787	1.2963	1.6716	2.0017	2.3924	2.6633
59	0.6787	1.2961	1.6711	2.0010	2.3912	2.6618
60	0.6786	1.2958	1.6706	2.0003	2.3901	2.6603

(segue)

**Valori critici della distribuzione t di Student per un test unilaterale**

(seconda parte)



Gradi di libertà	Aree della coda superiore					
	0.25	0.10	0.05	0.025	0.01	0.005
61	0.6785	1.2956	1.6702	1.9996	2.3890	2.6589
62	0.6785	1.2954	1.6698	1.9990	2.3880	2.6575
63	0.6784	1.2951	1.6694	1.9983	2.3870	2.6561
64	0.6783	1.2949	1.6690	1.9977	2.3860	2.6549
65	0.6783	1.2947	1.6686	1.9971	2.3851	2.6536
66	0.6782	1.2945	1.6683	1.9966	2.3842	2.6524
67	0.6782	1.2943	1.6679	1.9960	2.3833	2.6512
68	0.6781	1.2941	1.6676	1.9955	2.3824	2.6501
69	0.6781	1.2939	1.6672	1.9949	2.3816	2.6490
70	0.6780	1.2938	1.6669	1.9944	2.3808	2.6479
71	0.6780	1.2936	1.6666	1.9939	2.3800	2.6469
72	0.6779	1.2934	1.6663	1.9935	2.3793	2.6459
73	0.6779	1.2933	1.6660	1.9930	2.3785	2.6449
74	0.6778	1.4931	1.6657	1.9925	2.3778	2.6439
75	0.6778	1.2929	1.6654	1.9921	2.3771	2.6430
76	0.6777	1.2928	1.6652	1.9917	2.3764	2.6421
77	0.6777	1.2926	1.6649	1.9913	2.3758	2.6412
78	0.6776	1.2925	1.6646	1.9908	2.3751	2.6403
79	0.6776	1.2924	1.6644	1.9905	2.3745	2.6395
80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387
81	0.6775	1.2921	1.6639	1.9897	2.3733	2.6379
82	0.6775	1.2920	1.6636	1.9893	2.3727	2.6371
83	0.6775	1.2918	1.6634	1.9890	2.3721	2.6364
84	0.6774	1.2917	1.6632	1.9886	2.3716	2.6356
85	0.6774	1.2916	1.6630	1.9883	2.3710	2.6349
86	0.6774	1.2915	1.6628	1.9879	2.3705	2.6342
87	0.6773	1.2914	1.6626	1.9876	2.3700	2.6335
88	0.6773	1.2912	1.6624	1.9873	2.3695	2.6329
89	0.6773	1.2911	1.6622	1.9870	2.3690	2.6322
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316
91	0.6772	1.2909	1.6618	1.9864	2.3680	2.6309
92	0.6772	1.2908	1.6616	1.9861	2.3676	2.6303
93	0.6771	1.2907	1.6614	1.9858	2.3671	2.6297
94	0.6771	1.2906	1.6612	1.9855	2.3667	2.6291
95	0.6771	1.2905	1.6611	1.9853	2.3662	2.6286
96	0.6771	1.2904	1.6609	1.9850	2.3658	2.6280
97	0.6770	1.2903	1.6607	1.9847	2.3654	2.6275
98	0.6770	1.2902	1.6606	1.9845	2.3650	2.6269
99	0.6770	1.2902	1.6604	1.9842	2.3646	2.6264
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
110	0.6767	1.2893	1.6588	1.9818	2.3607	2.6213
120	0.6765	1.2886	1.6577	1.9799	2.3578	2.6174
130	0.6764	1.2881	1.6567	1.9784	2.3554	2.6142
140	0.6762	1.2876	1.6558	1.9771	2.3533	2.6114
150	0.6761	1.2872	1.6551	1.9759	2.3515	2.6090
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

**Confronto dei valori critici della distribuzione t tra un test bilaterale e un test unilaterale**

d.f.	Area nelle due code				
	0.10	0.05	0.02	0.01	0.001
	Area in una coda				
	0.05	0.025	0.01	0.005	0.0005
<b>1</b>	6.314	12.706	31.821	63.657	636.619
<b>2</b>	2.920	4.303	6.965	9.925	31.598
<b>3</b>	2.353	3.182	4.541	5.841	12941
<b>4</b>	2.132	2.776	3.747	4.604	8.610
<b>5</b>	2.015	2.571	3.365	4.032	6.859
<b>6</b>	1.943	2.447	3.143	3.707	5.959
<b>7</b>	1.895	2.365	2.998	3.499	5.405
<b>8</b>	1.860	2.306	2.896	3.355	5.041
<b>9</b>	1.833	2.262	2.821	3.250	4.781
<b>10</b>	1.812	2.228	2.764	3.169	4.587
<b>11</b>	1.796	2.201	2.718	3.106	4.437
<b>12</b>	1.782	2.179	2.681	3.055	4.318
<b>13</b>	1.771	2.160	2.650	3.01	4.221
<b>14</b>	1.761	2.145	2.624	2.977	4.140
<b>15</b>	1.753	2.131	2.602	2.947	4.073
<b>16</b>	1.746	2.120	2.583	2.921	4.015
<b>17</b>	1.740	2.110	2.567	2.898	3.965
<b>18</b>	1.734	2.101	2.552	2.878	3.922
<b>19</b>	1.729	2.093	2.539	2.861	3.883
<b>20</b>	1.725	2.086	2.528	2.845	3.850
<b>21</b>	1.721	2.080	2.518	2.831	3.819
<b>22</b>	1.717	2.074	2.508	2.819	3.792
<b>23</b>	1.714	2.069	2.500	2.807	3.767
<b>24</b>	1.711	2.064	2.492	2.797	3.745
<b>25</b>	1.708	2.060	2.485	2.787	3.725
<b>26</b>	1.706	2.056	2.479	2.779	3.707
<b>27</b>	1.703	2.052	2.473	2.771	3.690
<b>28</b>	1.701	2.048	2.467	2.763	3.674
<b>29</b>	1.699	2.045	2.462	2.756	3.659
<b>30</b>	1.697	2.042	2.457	2.750	3.646
<b>40</b>	1.684	2.021	.423	2.704	3.551
<b>60</b>	1.671	2.000	2.390	2.660	3.460
<b>120</b>	1.658	1.980	2.358	2.617	3.373
<b>∞</b>	1.645	1.960	2.326	2.576	3.291

## CAPITOLO VI

### METODI NON PARAMETRICI PER UN CAMPIONE

#### 6.1. CARATTERISTICHE DEI TEST NON PARAMETRICI

Il test t di Student per uno o per due campioni presentato nel capitolo precedente, il test F di Fisher per l'analisi della varianza, la correlazione e la regressione lineare semplice che saranno illustrati nei prossimi capitoli, la regressione multipla e la statistica multivariata che rappresentano lo sviluppo di tali tecniche applicate contemporaneamente a molte variabili sono i **metodi di inferenza classici o di statistica parametrica**.

Prima della applicazione di ognuno di questi test, è fondamentale che siano sempre verificati e soddisfatti alcuni assunti che riguardano la popolazione d'origine, dalla quale si presume che i dati campionari siano stati estratti. Nel caso in cui almeno uno dei presupposti non sia rispettato, neppure dopo appropriati tentativi di trasformazione dei dati, possono ragionevolmente sorgere dubbi sulla validità delle inferenze raggiunte. **Qualunque risultato statistico può essere messo in dubbio, quando non è certo che siano state rispettate compiutamente le condizioni di validità del test applicato.**

Il primo assunto da rispettare è **l'indipendenza dei gruppi campionari**: i campioni sottoposti ai differenti trattamenti dovrebbero essere generati per estrazione casuale da una popolazione, in cui ogni soggetto abbia la stessa probabilità di essere incluso in un gruppo qualsiasi. In questo modo, i fattori aleatori o non controllati, quelli che nel test t di Student formano l'errore standard e che nell'analisi della varianza formeranno la varianza d'errore o residuo, dovrebbero risultare casualmente distribuiti e non generare distorsioni od errori sistematici. E' una condizione che spesso è soddisfatta con facilità e che dipende quasi completamente dalla programmazione dell'esperimento. Per esempio, per verificare l'effetto di due tossici con il test t di Student, animali maschi e femmine, giovani ed anziani, grassi e magri devono essere distribuiti casualmente o in modo bilanciato nei due gruppi a confronto, se esiste il sospetto che il sesso, l'età ed il peso possano dare risultati differenti, rispetto all'effetto medio dei due tossici.

Il secondo assunto, distintivo della statistica parametrica, riguarda la **normalità delle distribuzioni**. Da essa deriva la relazione tra popolazione dei dati e medie dei campioni, secondo il **teorema del limite centrale**: **se da una popolazione con media  $\mu$  e varianza  $\sigma^2$ , i cui dati abbiano una forma di distribuzione non normale, si estraggono casualmente campioni di dimensione n, le loro medie si distribuiranno normalmente con media generale m ed errore standard  $s/\sqrt{n}$ .**

**La non-normalità della distribuzione delle medie è un indice serio di un'estrazione non casuale.**

La grande importanza pratica del **teorema del limite centrale** deriva dal fatto che **gruppi di dati** ( $x_{ij}$ ), estratti da una popolazione distribuita in modo differente dalla normale, hanno medie ( $\bar{x}_i$ ) che tendono a distribuirsi normalmente.

**La distribuzione normale è la forma limite della distribuzione delle medie campionarie  $\bar{x}_i$  per  $n$  che tende all'infinito.** Tuttavia, si può avere una buona approssimazione alla normale della distribuzione delle medie  $\bar{x}_i$  anche quando  $n$  è piccolo e la distribuzione dei dati ( $x_{ij}$ ) è molto distante dalla normale.

E' possibile comprendere il teorema del limite centrale in modo intuitivo, pensando come esempio al lancio dei dadi. Con un solo dado, i 6 numeri avranno la stessa probabilità e la distribuzione delle frequenze dei numeri ottenuti con i lanci ha una distribuzione rettangolare. Con due dadi, è possibile ottenere somme da 2 a 12 e tra esse quelle centrali sono più frequenti; all'aumentare del numero di dadi, **la distribuzione delle somme o delle medie (la legge è valida per entrambe poiché contengono la medesima informazione)** è sempre meglio approssimata ad una distribuzione normale.

Il terzo assunto riguarda la **omoschedasticità o omogeneità delle varianze**: se sono formati per estrazione casuale dalla medesima popolazione, come espresso nell'ipotesi nulla  $H_0$ , i vari gruppi devono avere varianze eguali. **Nella statistica parametrica, è possibile verificare se esistono differenze significative tra medie campionarie, solamente quando i gruppi a confronto hanno la stessa varianza.**

Quando la distribuzione di base è nota, ma non necessariamente normale, si possono calcolare probabilità esatte, come già mostrato con la distribuzione binomiale o con il metodo esatto di Fisher, fondato sulla distribuzione ipergeometrica. **Quando la forma della distribuzione dei dati è ignota, servono test che possano essere applicati con qualunque forma di distribuzione.** E' una situazione che nella ricerca sperimentale si realizza con frequenza e che richiede l'uso di **test indipendenti dalla forma della distribuzione**, come sono molti di quelli non parametrici.

**In questi anni, l'importanza della statistica non parametrica è fortemente aumentata.** Nelle riviste internazionali, è avvenuta una rapida evoluzione nelle scelte degli esperti di statistica. Fino a poco tempo fa, i test parametrici erano quasi sempre richiesti, quando non fosse dimostrato che la distribuzione doveva essere considerata, con elevata probabilità, differente dalla normale; ora spesso sono accettati solamente se è possibile dimostrare che la distribuzione è normale o approssimativamente tale.

**Si è rovesciato l'onere della prova, per accettare la validità di un test parametrico.**

Spesso nella ricerca sperimentale è possibile disporre solo di pochi dati, che sono assolutamente insufficienti per dimostrare la normalità della distribuzione; in particolare, quando il fenomeno studiato è nuovo e non è possibile citare dati di altre esperienze.

Nelle edizioni più recenti, vari testi importanti di statistica applicata consigliano di ricorrere alle tecniche non parametriche **quando gli assunti teorici relativi alle condizioni di validità della distribuzione normale non sono dimostrati.**

**In condizioni di incertezza sull'esistenza delle condizioni richieste da un test parametrico,** come quasi sempre succede quando si dispone di pochi dati, una soluzione sempre più diffusa suggerisce una duplice strategia:

1 - **utilizzare un test appropriato di statistica parametrica,**

2 - **convalidare tali risultati mediante l'applicazione di un equivalente test non parametrico.**

Se le probabilità stimate con i due differenti metodi risultano simili, sono confermate la robustezza del test parametrico e la sua sostanziale validità anche in quel caso. In questo caso, il test non parametrico può servire per confermare i risultati ottenuti con quello parametrico e come misura preventiva contro eventuali obiezioni sulla normalità ed omoschedasticità dei dati. Se **le probabilità** (non il valore del test) risultassero sensibilmente differenti, dovrebbe essere considerato come più attendibile il test non parametrico e sarebbe conveniente riportare nella pubblicazione solo esso, in quanto fondato su condizioni meno rigorose e caratterizzato da inferenze più generali.

In alcuni testi, questi concetti sono stati sintetizzati in un consiglio ai ricercatori: **quando sei incerto se utilizzare un test parametrico oppure uno non parametrico, usali entrambi.**

**I metodi non parametrici sono meno potenti, per cui è più difficile rifiutare l'ipotesi nulla; ma quando l'ipotesi nulla è rifiutata, generalmente le conclusioni non possono essere sospettate d'invalidità.**

**I test non parametrici presentano vantaggi e svantaggi.**

I test non parametrici sovente si fondano su una tecnica statistica semplice. Con poche eccezioni, **richiedono calcoli elementari**, spesso fondati sul calcolo combinatorio, che possono essere fatti in modo rapido, anche mentalmente, senza alcun supporto tecnico sofisticato. Per tale caratteristica è comprensibile la definizione, data anni fa da Tukey, di **“metodi rapidi e sporchi”**, in contrapposizione al tempo richiesto dai calcoli, soprattutto all'eleganza logica e alla pulizia matematica formale dei metodi parametrici.

Quando per la verifica delle ipotesi non è possibile o non è conveniente applicare i metodi classici, si può ricorrere a **test di statistica non parametrica**, detti anche **metodi indipendenti dalla forma della distribuzione** (*distribution-free*).

Per la maggior parte, questi metodi **sono fondati sulle statistiche di rango o d'ordine; non utilizzano la media, ma la mediana come misura della tendenza centrale; vengono applicati indifferentemente sia alle variabili casuali discrete che a quelle continue.**

**Quando le scale sono qualitative od ordinali e i campioni non sono di grandi dimensioni, non esistono alternative accettabili all'uso di test non parametrici.**

I metodi non parametrici presentano **diversi vantaggi**:

- forniscono **risposte rapide** con **calcoli elementari**, quando i campioni sono piccoli,
- richiedono **ipotesi meno rigorose**, in numero minore, più facilmente verificate nella realtà,
- sono **più estesamente applicabili** e quindi portano a **conclusioni più generali** e sono più **difficilmente confutabili**;
- alcune volte **permettono anche analisi differenti, non possibili con i metodi classici**, poiché non esistono test parametrici equivalenti, come nel caso del test delle successioni (presentato nel paragrafo successivo);
- **in certe condizioni, hanno addirittura una potenza maggiore**, in particolare quando gli assunti di validità del test parametrico sulla forma della distribuzione non sono rispettati rigorosamente ed il test classico non è robusto: è il caso del test di casualizzazione rispetto al test t di Student.

Impiegati vantaggiosamente in una varietà di situazioni, i test non parametrici presentano anche **alcuni svantaggi**.

Per scale d'intervalli o di rapporti, quando le condizioni di validità per i metodi classici sono rispettate in modo rigoroso,

- sovente **sfruttano in modo meno completo l'informazione contenuta nei dati**; quindi **hanno una potenza minore**, in particolare quando riducono l'informazione da scale d'intervalli o di rapporti a scale di rango o a risposte binarie.

Per campioni di grandi dimensioni i metodi non parametrici, soprattutto se fondati sul calcolo combinatorio,

- **a volte richiedono metodologie più lunghe, manualmente impossibili**, che pretendono l'uso del calcolatore. L'attuale divulgazione di alcuni di questi metodi, come sarà illustrato negli ultimi capitoli, è in parte dovuta anche alla diffusione dell'informatica.

Per molti test è complesso valutare la significatività delle ipotesi,

- poiché è **difficile disporre delle tavole dei valori critici**, pubblicati solo in testi per specialisti, quando non sia possibile disporre di campioni abbastanza grandi da permettere l'uso della distribuzione normale.

**I metodi non parametrici sono adatti a problemi relativamente semplici**, come il confronto tra due o più medie o tra due o più varianze, sempre relativamente ad un solo fattore. **Con strutture di dati complesse, in cui si vogliono considerare contemporaneamente più fattori e covariate, non esistono ancora alternative al modello parametrico. Una soluzione elegante è spesso la trasformazione dei dati nel loro rango: anche con poche osservazioni, la distribuzione diventa**

**approssimativamente normale e vengono ricostruite le condizioni di validità per l'uso dei test di statistica classica.**

Nella ricerca ambientale, si rivela sempre più utile la conoscenza della statistica non parametrica, almeno dei test che più frequentemente sono citati nella letteratura specifica. Esiste un'ampia varietà di situazioni in cui possono essere applicati con rilevante profitto. **Sotto l'aspetto didattico**, per la sua semplicità d'impostazione, **la statistica non parametrica si dimostra particolarmente utile nell'apprendimento dei processi logici, in riferimento alla formulazione delle ipotesi, alla stima delle probabilità mediante il test e all'inferenza sui parametri a confronto.**

I test di statistica classica formano una struttura logica unica, che ricorre ai medesimi presupposti ed elabora, in modo organico e con complessità crescente, una quantità di informazioni sempre maggiore, dal test *t* all'analisi della varianza, dalla regressione lineare semplice, all'analisi della covarianza, alla regressione multipla e alla statistica multivariata.

La statistica non parametrica invece è cresciuta per semplice accumulo di una serie ormai innumerevole di test, ognuno proposto per risolvere un problema specifico o poche situazioni particolari, anche se molti di essi si rifanno agli stessi principi elementari, come il calcolo dei ranghi o delle precedenze.

In questa frammentarietà d'elementi comuni e diversità d'approcci, diventa difficile ed ampiamente soggettiva una organizzazione logica e didattica delle varie centinaia di test non parametrici che è possibile rintracciare in letteratura. Nei testi è frequentemente risolta non sull'analogia dei metodi, ma sulla base del numero di campioni a confronto e delle ipotesi da verificare.

Nella presentazione dei metodi più utili alla ricerca e alla gestione ambientale, i test non parametrici sono sovente classificati in 3 gruppi:

- 1 - test per 1 campione e per 2 campioni dipendenti o indipendenti,
- 2 - test per *k* campioni dipendenti o indipendenti,
- 3 - test per l'associazione, la valutazione delle tendenze, la correlazione e la regressione.

In queste dispense sono stati suddivisi in

- 1 - test per un campione;
- 2 - test per due campioni dipendenti o indipendenti,
- 3 - test per più campioni dipendenti o indipendenti,
- 4 - misure di tendenza e di associazione,
- 5 - test per correlazione, concordanza e regressione lineare.

Ad essi è stato aggiunto un paragrafo (nel capitolo XV) sull'uso del bootstrap e del jackknife, le tecniche più recenti quando non sia possibile ricorrere alla statistica parametrica.

Tra i vari argomenti fino ad ora discussi, **il test  $\chi^2$  e il test G, il metodo esatto di Fisher e le differenti applicazioni del test di Kolmogorov-Smirnov devono essere classificati tra i test non**

**parametrici.** Sono stati trattati separatamente e prima della discussione generale sui metodi, perché utili a presentare in modo semplice la procedura dell'inferenza; inoltre essi sono considerati fondamentali in qualsiasi corso anche elementare di statistica, a causa delle loro numerose applicazioni nella ricerca sperimentale, sia in natura sia in laboratorio.

I test presentati nel terzo capitolo sul chi quadrato e il test G, con esclusione di quelli che si rifanno alla distribuzione z, sono parte integrante ed essenziale della statistica non parametrica. In vari casi, essi forniscono anche le procedure inferenziali, i valori critici e la distribuzione delle probabilità di altri test non parametrici; è il caso del test della mediana, che dopo aver diviso i 2 o più gruppi a confronto in due classi, ricorre al test  $\chi^2$  per la stima della significatività.

## **6.2. IL TEST DELLE SUCCESSIONI PER UN CAMPIONE**

Quando si dispone di un solo campione, i quesiti di inferenza statistica che ricorrono con maggior frequenza riguardano la verifica di un **accordo della distribuzione osservata con una distribuzione teorica od attesa.** E' quanto già discusso nel capitolo III, in cui **la distribuzione attesa può essere stimata sulla base di qualsiasi legge matematica, statistica o biologica.**

Un secondo gruppo importante di inferenze applicate ad un campione riguarda la verifica della tendenza centrale in una distribuzione simmetrica, sia intorno allo zero che a qualsiasi altro valore. Come già presentato nei capitoli precedenti di statistica parametrica, sono utilizzati il test **z** o il test **t**, in rapporto al fatto che si conosca la varianza della popolazione o che si debba ricorrere a quella campionaria come sua stima migliore.

Quando la distribuzione non è normale o il tipo di scala è ordinale, si può ricorrere ai test presentati nei capitoli successivi. Prima di essi è tuttavia importante discutere il test delle successioni, in quanto non ha l'equivalente in statistica parametrica ed è utile in varie circostanze che si ripetono con relativa frequenza nella ricerca ambientale ed ecologica.

**Il test delle successioni per un campione o per risposte alternative (*Run test with two attributes*), già illustrato da J. V. Bradley nel suo testo di Statistica non parametrica del 1968 (*Distribution-free Statistical Test*, edito da Englewood Cliffs, NJ: Prentice Hall), è utilizzato nel caso di osservazioni raccolte in una successione temporale o in successione geografica, a partire da un punto origine. Permette di saggiare se, in riferimento all'ordine, alla sequenza, successione (**run**) o serie (termini sinonimi), i dati del campione sono casuali.**

E' un quesito che nella ricerca sperimentale si pone spesso, sia in laboratorio che in natura: verificare se i risultati positivi e negativi, i valori alti e bassi di una serie temporale di osservazioni, due categorie di eventi alternativi si succedono nel tempo o nello spazio con casualità, oppure tendono ad aggregarsi o ad alternarsi con regolarità prevedibile, come possono essere fenomeni ciclici.

**L'importanza del test per l'analisi delle successioni deriva anche dall'essere privo di alternative: non esistono test analoghi nella statistica parametrica.**

Il test può essere applicato per dati binari, tipici di un processo binomiale. Può essere esteso a dati continui, misurati con scale d'intervalli o di rapporti solo dopo trasformazione in risposte binarie, che può avvenire mediante il confronto con un valore stimato o prefissato, come la mediana della distribuzione o un qualsiasi valore soglia.

Per chiarire più compiutamente i concetti relativi alla sua applicazione, può essere utile un esempio.

Si supponga di lanciare una moneta 15 volte e che si ottenga 8 volte testa (T) e 7 volte croce (C), con la seguente **serie temporale**

T C T C T C T C T C T C T C T

**E' evidente la sua non casualità, che si riferisce non al numero di T e di C, la cui probabilità può essere calcolata con il test binomiale, ma al regolare alternarsi degli eventi.** Parimenti non casuale sarebbe stata una serie, identica come numero di dati alternativi T (8) e C (7), ma in sequenza differente,

T T T T T T T T C C C C C C C

che concentra nella prima parte tutti gli eventi di un tipo e nella seconda quelli dell'altro tipo.

La **verifica degli effetti sistematici o periodici** è fondata sul **conteggio delle successioni**, definite come **il numero di simboli identici preceduti o seguiti da simboli differenti o da nessun simbolo.**

Nel primo caso dell'esempio, il numero delle successioni è

$$\frac{T}{1} \frac{C}{2} \frac{T}{3} \frac{C}{4} \frac{T}{5} \frac{C}{6} \frac{T}{7} \frac{C}{8} \frac{T}{9} \frac{C}{10} \frac{T}{11} \frac{C}{12} \frac{T}{13} \frac{C}{14} \frac{T}{15}$$

**15**, come il numero di osservazioni che ne rappresenta il valore massimo; nel secondo caso è

$$\frac{T T T T T T T T}{1} \quad \frac{C C C C C C C}{2}$$

solamente **2**, come il numero di eventi alternativi e che rappresenta il numero minimo.

E' intuitivo che la successione dei 15 lanci della moneta non può essere ritenuta casuale, in nessuno dei due casi teorici presentati. E' evidente una legge dell'alternanza nel primo caso e di

concentrazione nel secondo, che permettono di indovinare con facilità il risultato di eventuali lanci successivi.

Essi rappresentano i due casi estremi di tutti i modi in cui è possibile disporre in ordine differente gli elementi dei due gruppi.

In un gruppo di  $N$  oggetti di cui  $n_1$  di tipo 1 e  $n_2$  di tipo 2 si hanno

$$\frac{N!}{n_1!n_2!}$$

possibili differenti ordini differenti.

Nell'esempio precedente con  $N = 15$ ,  $n_1 = 8$  e  $n_2 = 7$

sono

$$\frac{15!}{8!7!} = 6435$$

6435 ordini differenti.

Ognuno di essi è caratterizzato da un numero specifico di successioni, che in totale hanno una distribuzione approssimativamente o asintoticamente normale, per campioni sufficientemente grandi.

Un campione può ragionevolmente essere ritenuto casuale solamente quando il numero delle successioni non è né troppo grande né troppo piccolo, in rapporto al numero di eventi dei due tipi alternativi. **Per essere casuale, il numero di successioni deve tendere ad una frequenza media ( $m_r$ ), che dipende dal numero dei due eventi alternativi e può essere calcolata con la formula**

$$m_r = \frac{2 \times n_1 \times n_2}{N} + 1$$

dove

$m_r$  è la media aritmetica attesa delle successioni,

$n_1$  è il numero di eventi di un tipo,

$n_2$  è il numero di eventi dell'altro tipo,

$N$  è il numero totale di dati od osservazioni ( $N = n_1 + n_2$ ).

Applicata sempre allo stesso esempio teorico, con  $N = 15$ ,  $n_1 = 8$  e  $n_2 = 7$ , la media stimata o attesa ( $m_r$ ) del numero di successioni (runs), nell'ipotesi che  $H_0$  (distribuzione casuale di T e C) sia vera

$$\mu_r = \frac{2 \cdot 8 \cdot 7}{15} + 1 = \frac{112}{15} + 1 = 8,4\bar{6}$$

è uguale a  $8,4\bar{6}$ .

In questo modo, il quesito sulla casualità delle successioni è trasformato nel problema di **verificare se il numero di successioni contato nella serie sperimentale di dati (15 o 2) è significativamente differente dal valore medio atteso (8,46̄)**.

**Nella condizione che l'ipotesi nulla  $H_0$  sia vera** (totale casualità degli eventi nella loro successione temporale), **la differenza tra il numero di successioni osservate ed il numero atteso segue una distribuzione approssimativamente normale, quando le dimensioni dei due campioni sono grandi**. La probabilità relativa può essere calcolata mediante

$$z = \frac{R - m_r}{S_r}$$

dove

**R** è il numero di successioni (Runs) osservate,

**$m_r$**  è la media attesa di successioni, nella condizione che l'ipotesi nulla  $H_0$  sia vera,

**$S_r$**  è la deviazione standard della media e può essere calcolata da

$$S_r = \sqrt{\frac{2 \cdot n_1 \cdot n_2 \cdot (2 \cdot n_1 \cdot n_2 - N)}{N^2 \cdot (N - 1)}}$$

sulla base del numero di dati  **$n_1$**  e  **$n_2$**  dei due eventi alternativi e del numero totale **N** di osservazioni.

**Nell'uso di questo test, l'ipotesi che ricorre con frequenza maggiore riguarda un numero troppo ridotto di successioni.** Nella ricerca etologica può essere, in animali a struttura sociale di tipo gerarchico, la modalità d'accesso al cibo d'individui appartenenti a due livelli sociali diversi o la precedenza degli anziani rispetto ai giovani. Nella ricerca ambientale, è il caso della successione di depositi geologici in una sezione, della quale si intenda verificare la non casualità dei differenti strati risalenti a due tipologie diverse.

**TAVOLA DEI VALORI CRITICI NEL TEST DELLE SUCCESSIONI  
 ALLA PROBABILITA' 0.05 PER TEST A DUE CODE**

La tabella superiore riporta i valori minimi e quella inferiore i valori massimi significativi.

n <sub>1</sub>	n <sub>2</sub>																			
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
2											2	2	2	2	2	2	2	2	2	
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6	
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6	
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7	
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8	
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9	
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9	
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10	
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10	
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11	
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12	
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12	
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13	
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13	
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13	
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14	

n <sub>1</sub>	n <sub>2</sub>																			
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
4				9	9															
5			9	10	10	11	11													
6			9	10	11	12	12	13	13	13	13									
7				11	12	13	13	14	14	14	14	15	15	15						
8				11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17	
9					13	14	14	15	16	16	16	17	17	18	18	18	18	18	18	
10					13	14	15	16	16	17	17	18	18	18	19	19	19	20	20	
11					13	14	15	16	17	17	18	19	19	19	20	20	20	21	21	
12					13	14	16	16	17	18	19	19	20	20	21	21	21	22	22	
13						15	16	17	18	19	19	20	20	21	21	22	22	23	23	
14						15	16	17	18	19	20	20	21	22	22	23	23	23	24	
15						15	16	18	18	19	20	21	22	22	23	23	24	24	25	
16							17	18	19	20	21	21	22	23	23	24	25	25	25	
17							17	18	19	20	21	22	23	23	24	25	25	26	26	
18							17	18	19	20	21	22	23	24	25	25	26	26	27	
19							17	18	20	21	22	23	23	24	25	26	26	27	27	
20							17	18	20	21	22	23	24	25	25	26	27	27	28	

**TAVOLA DEI VALORI CRITICI NEL TEST DELLE SUCCESSIONI  
ALLA PROBABILITA' 0.05 e 0.01 PER TEST A UNA CODA**

Le tabelle riportano i valori minimi significativi.

E' significativo ogni numero di successioni minore od uguale a quello riportato nella tabella.

**$\alpha = 0.05$**

$n_1$	$n_2$																			
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
4			2																	
5		2	2	3																
6		2	3	3	3															
7		2	3	3	4	4														
8	2	2	3	3	4	4	5													
9	2	2	3	4	4	5	5	6												
10	2	3	3	4	5	5	6	6	6											
11	2	3	3	4	5	5	6	6	7	7										
12	2	3	4	4	5	6	6	7	7	8	8									
13	2	3	4	4	5	6	6	7	8	8	9	9								
14	2	3	4	5	5	6	7	7	8	8	9	9	10							
15	2	3	4	5	6	6	7	8	8	9	9	10	10	11						
16	2	3	4	5	6	6	7	8	8	9	10	10	11	11	11					
17	2	3	4	5	6	7	7	8	9	9	10	10	11	11	12	12				
18	2	3	4	5	6	7	8	8	9	10	10	11	11	12	12	13	13			
19	2	3	4	5	6	7	8	8	9	10	10	11	12	12	13	13	14	14		
20	2	3	4	5	6	7	8	9	9	10	11	11	12	12	13	13	14	14	15	

**$\alpha = 0.01$**

$n_1$	$n_2$																			
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
5				2																
6			2	2	2															
7			2	2	3	3														
8			2	2	3	3	4													
9		2	2	3	3	4	4	4												
10		2	2	3	3	4	4	5	5											
11		2	2	3	4	4	5	5	5	6										
12		2	3	3	4	4	5	5	6	6	7									
13		2	3	3	4	5	5	6	6	6	7	7								
14		2	3	3	4	5	5	6	6	7	7	8	8							
15		2	3	4	4	5	5	6	7	7	8	8	8	9						
16		2	3	4	4	5	6	6	7	7	8	8	9	9	10					
17		2	3	4	5	5	6	7	7	8	8	9	9	10	10	10				
18		2	3	4	5	5	6	7	7	8	8	9	9	10	10	11	11			
19	2	2	3	4	5	6	6	7	8	8	9	9	10	10	11	11	12	12		
20	2	2	3	4	5	6	6	7	8	8	9	10	10	11	11	11	12	12	13	

**Per piccoli campioni ( $n_1$  e  $n_2 < 20$ )**, la distribuzione delle probabilità è distorta rispetto a quella normale. In tali condizioni, la significatività deve essere fornita da tabelle che riportano le **frequenze critiche minime e massime**. Nelle due pagine precedenti sono state riportate quattro tabelle: le prime due per test bilaterali, le altre due per test unilaterali.

**Le prime due tabelle** forniscono i valori critici del numero di successioni minimo e massimo alla probabilità  $\alpha = 0.05$ ; le due tabelle sono simmetriche, per cui è indifferente l'identificazione in  $n_1$  o  $n_2$  del numero di osservazioni di uno qualsiasi dei due gruppi.

**Nella tabella superiore, è riportato il valore minimo significativo: permette di rifiutare l'ipotesi nulla, con qualsiasi valore uguale o minore di quello riportato.** Qualunque valore osservato ( $R$ ) che sia uguale o minore a quello riportato ha una probabilità  $\alpha = 0.025$  o minore di verificarsi, nell'ipotesi che  $H_0$  sia vera.

**Nella tabella inferiore, è riportato il valore massimo significativo: permette di rifiutare l'ipotesi nulla, con qualsiasi valore uguale o superiore a quello riportato.** Qualunque valore osservato ( $R$ ) che sia uguale o maggiore di quello riportato corrisponde a una probabilità  $\alpha$  uguale o minore di 0.025 di essere casuale.

Con i dati dell'esempio ( $n_1 = 8$  e  $n_2 = 7$ ), il valore minimo, riportato nella tabella superiore, è 4; il valore massimo, riportato nella tabella inferiore, è 13. In un test bilaterale, sarebbero quindi significativi i valori uguali od inferiori a 4 e uguali o superiori a 13: i due valori (2 e 15) del numero di successioni osservate nell'esempio con il lancio delle monete permettono di rifiutare l'ipotesi nulla.

Quando il numero di successioni è compreso nell'intervallo fra la frequenza minima e quella massima riportate della tabella, con esclusione dei valori riportati, non si è nelle condizioni di rifiutare l'ipotesi nulla  $H_0$ : la sequenza dei due eventi può essere ritenuta casuale, alla probabilità prefissata (in questo caso  $\alpha = 0.05$ ).

**Le altre due tabelle** forniscono i valori critici per test ad una coda, al fine di **verificare se il numero di successioni osservato sia minore di quello casuale**, atteso nella condizione che  $H_0$  sia vera ( $H_1: R < m_r$ ).

La tabella superiore riporta i valori critici alla probabilità  $\alpha = 0.05$  e

la tabella inferiore i valori critici alla probabilità  $\alpha = 0.01$ .

A differenza delle due precedenti, **queste due tabelle non sono simmetriche:**

**$n_1$  è il numero di osservazioni del campione maggiore ed**

**$n_2$  identifica il numero di osservazioni del campione minore.**

**Per grandi campioni ( $n_1$  o  $n_2 > 20$ ) non eccessivamente sbilanciati, la **distribuzione delle successioni tende ad essere asintoticamente normale**. I valori critici sono pertanto forniti dalla **tabella della distribuzione normale standardizzata**.**

I valori critici alla probabilità  $\alpha = 0.05$  sono rispettivamente

- 1,96 per un test bilaterale ( $\alpha = 0.025$  nelle due code della distribuzione) e
- 1,645 per un test unilaterale ( $\alpha = 0.05$  in una sola coda della distribuzione).

Alla probabilità  $\alpha = 0.01$  sono rispettivamente

- 2,58 per un test bilaterale ( $\alpha = 0.005$  nelle due code della distribuzione) e
- 2,33 per un test unilaterale ( $\alpha = 0.01$  in una sola coda della distribuzione).

ESEMPIO 1. In laboratorio di analisi delle sostanze inquinanti oleose, si ha il timore che lo strumento di analisi non sia corretto, ma che, sporcandosi con una concentrazione alta, influenzi quella successiva. Sono stati misurati in successione 16 campioni d'acqua contenenti una sostanza oleosa e sono stati ottenuti i seguenti valori di concentrazione, espressi in mg per litro:

25	36	27	45	18	76	89	73	57	44	21	32	85	67	78	85
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Sulla base dei dati raccolti, si può affermare alla probabilità  $\alpha = 0.05$  che lo strumento non sia corretto?

Risposta.

Occorre dapprima classificare i valori in due categorie: bassi (-) e alti (+). E' una trasformazione possibile mediante il confronto con la mediana (uguale a 51 nella serie dei dati raccolti), per cui i due gruppi avranno un numero uguale di osservazioni basse e alte (8- e 8+).

La sequenza delle risposte, tradotte in valori bassi (-) e alti (+), diventa

-	-	-	-	-	+	+	+	+	-	-	-	+	+	+	+
1				2				3				4			

e risulta formata da 4 successioni.

Sulla base dell'ipotesi formulata, è un test ad una coda: si tratta infatti di verificare se esiste trascinamento dell'informazione e quindi se il numero di successioni sia significativamente inferiore al valore atteso, nell'ipotesi di totale casualità degli eventi come dovrebbe succedere con uno strumento corretto.

Secondo l'ipotesi nulla

$$H_0: R = m_r$$

il numero medio di successioni osservate ( $R$ ) non si discosta in modo significativo da quello atteso, mentre l'ipotesi alternativa unilaterale

$$H_1: R < m_r$$

che essa è significativamente minore.

Con  $n_1 = 8$  e  $n_2 = 8$ , tale media attesa  $\mu_r$

$$\mu_r = \frac{2 \cdot 8 \cdot 8}{16} + 1 = 9$$

risulta uguale a 9.

Per verificare l'ipotesi sulla casualità della sequenza delle analisi, occorre quindi risolvere il problema statistico di conoscere la probabilità di trovare 4 successioni o un valore inferiore, quando la media attesa è 9.

La tavola dei valori critici inferiori per un test ad una coda, per  $n_1 = 8$  e  $n_2 = 8$  alla probabilità  $\alpha = 0.05$  riporta la frequenza di 5, che è maggiore del valore di  $R$  (4) osservato.

E' un valore significativo alla probabilità  $\alpha = 0.05$ .

Di conseguenza, si rifiuta l'ipotesi nulla di una casualità del numero di successioni osservate e si accetta l'ipotesi alternativa: lo strumento non è preciso, ma risente del valore di concentrazione dell'ultimo campione.

ESEMPIO 2. Si vuole verificare se in un gruppo d'animali l'accesso al cibo avvenga in modo casuale oppure sia possibile ipotizzare un ordine differente: un alternarsi quasi regolare tra maschi e femmine secondo lo stato sociale delle coppie oppure a gruppi dello stesso sesso.

Un gruppo di 30 ( $N$ ) animali in cattività con un'organizzazione sociale a struttura gerarchica, composto da 17 ( $n_1$ ) femmine e 13 ( $n_2$ ) maschi adulti, deve percorrere uno stretto corridoio per accedere al cibo. L'ordine osservato d'accesso al cibo, classificato per sesso, è stato:

MFFMFFFMMMFMFMFFFFFMMMFMMMFFFF

L'ordine può essere definito casuale?

Risposta.

Dopo aver contato il numero di successioni ( $R$ )

(già definite come **il numero di simboli identici preceduti o seguiti da simboli differenti o da nessun simbolo**)

M	FF	M	FFF	MMM	F	M	F	M	FFFFF	MMM	F	MMM	FFFF
1	2	3	4	5	6	7	8	9	10	11	12	13	14

e che risulta uguale a 14,  
aver definito  $N$  (30),  $n_1$  (17) e  $n_2$  (13),  
si deve calcolare la media attesa, nella condizione che  $H_0$  sia vera.  
Dalla formula generale

$$m_r = \frac{2 \times n_1 \times n_2}{N} + 1$$

si stima una media attesa

$$m_r = \frac{2 \cdot 17 \cdot 13}{30} + 1 = 15,73$$

uguale a 15,73.

L'ipotesi nulla è

$$H_0: R = m_r$$

ossia che il numero medio di successioni osservate ( $R = 14$ ) non si discosta in modo significativo da quello atteso (15,73),  
mentre l'ipotesi alternativa bilaterale

$$H_1: R \neq m_r$$

è che esso differisca in modo non casuale.

Il numero osservato di successioni ( $R = 14$ ) è inferiore alla media attesa ( $m_r = 15,73$ ). Vi è quindi una tendenza a spostarsi in gruppi dello stesso sesso. Si tratta di valutare se tale differenza tra numero osservato e media attesa sia significativa.

Il test è bilaterale, in quanto a priori non si intendeva verificare una teoria precisa, ma solo evidenziare il comportamento degli individui esaminati.

Per 2 gruppi, di 17 e 13 dati rispettivamente, alla probabilità  $\alpha = 0.05$  nelle due tabelle sinottiche il valore critico inferiore è 10 ed il valore critico superiore è 22. Il numero di successioni osservato ( $R = 14$ ) è compreso in questo intervallo: si accetta l'ipotesi nulla. In riferimento al sesso, l'accesso al cibo è avvenuto in ordine casuale.

Il numero di osservazioni (17 e 13) è abbastanza alto e relativamente bilanciato, tale da permettere il ricorso alla distribuzione normale, al fine di stimare una probabilità più precisa di quella fornita dalla tabella attraverso i valori critici.

Dopo aver calcolato il numero di successioni osservato ( $\mathbf{R} = 14$ ) e la media attesa ( $\mathbf{m}_r = 15,73$ ) nell'ipotesi che  $H_0$  sia vera, attraverso la formula generale

$$s_r = \sqrt{\frac{2 \cdot n_1 \cdot n_2 \cdot (2 \cdot n_1 \cdot n_2 - N)}{N^2 \cdot (N - 1)}}$$

si stima la deviazione standard  
che, con i dati dell'esempio,

$$s_r = \sqrt{\frac{2 \cdot 17 \cdot 13 \cdot (2 \cdot 17 \cdot 13 - 30)}{30^2 \cdot (30 - 1)}} = \sqrt{\frac{442 \cdot 412}{900 \cdot 29}} = \sqrt{6,977} = 2,64$$

risulta uguale a 2,64.

Attraverso  $\mathbf{R}$  (14),  $\mathbf{m}_R$  (15,73) e  $\mathbf{s}_R$  (2,64)

si stima il valore di  $z$

$$z = \frac{\mathbf{R} - \mathbf{m}_r}{s_r}$$

che

$$z = \frac{14 - 15,73}{2,64} = 0,6553$$

fornisce un valore approssimato di 0,66.

Nella tavola della distribuzione normale standardizzata ad esso è associata una probabilità uguale a 0.5092 per un test bilaterale. E' una probabilità molto alta (50,92%), indicativa di una elevata casualità del numero di successioni osservato nell'esperimento.

**ESEMPIO 3.** Si ritiene importante verificare se aumenti e diminuzioni nel tasso medio d'inquinamento giornaliero avvengono casualmente ( $H_0$ ), oppure se esistono periodi di più giorni con una serie continua di valori con la stessa tendenza ( $H_1$ ).

Per un periodo di 50 giorni continuativi, è stata fatta una rilevazione media del tasso d'inquinamento atmosferico. Si è anche annotato costantemente se il valore era aumentato o diminuito, rispetto al giorno precedente.

Si è osservato che in 34 casi era aumentato (+) ed in 16 era diminuito (-), con la sequenza riportata nella tabella sottostante:

+++++	-	++	---	+++++	-	++	-	+++++	-	+++++	----	+++++	--	+++	---
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>

Quali indicazioni è possibile dedurre?

Risposta.

E' un test ad una coda: si chiede di verificare se le successioni osservate ( $\mathbf{R} = \mathbf{16}$ ) sono in numero significativamente minore dell'atteso, se fosse vera l'ipotesi di pura casualità dei valori d'inquinamento.

Il numero di osservazioni è sufficientemente alto ( $\mathbf{N} = 50$ ;  $\mathbf{n}_1 = 16$ ;  $\mathbf{n}_2 = 34$ ) da permettere il ricorso alla distribuzione normale standardizzata.

Secondo l'ipotesi nulla, il numero medio atteso ( $\mu_r$ ) è

$$\mu_r = \frac{2 \cdot 16 \cdot 34}{50} + 1 = 22,76$$

uguale a 22,76 e

la deviazione standard ( $\sigma_r$ ) è

$$s_r = \sqrt{\frac{2 \cdot 16 \cdot 34 \cdot (2 \cdot 16 \cdot 34 - 50)}{50^2 \cdot (50 - 1)}} = 3,036$$

uguale a 3,036.

La **probabilità di trovare per caso differenze uguali o superiori** a quella riscontrata tra numero di successioni osservate ( $\mathbf{R} = 16$ ) ed attese ( $\mu_r = 22,76$ ) è fornita dalla distribuzione normale standardizzata ( $\mathbf{z}$ )

$$z = \frac{16 - 22,76}{3,036} = -2,226$$

il cui valore è uguale a -2,226.

Nell'area di coda unilaterale, a tale valore (arrotondato a 2,23) è associata una probabilità  $\alpha$  uguale a 0.01287 o 1,287%.

La probabilità che il numero di successioni osservate sia casuale è molto piccola, nettamente inferiore a 0.05 scelta abitualmente come probabilità critica.

Si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa: i giorni di aumento e di diminuzione dei valori d'inquinamento atmosferico non si alternano casualmente, ma tendono a concentrarsi in serie temporali di più giorni.

### 6.3. TEST DEI SEGNI PER UN CAMPIONE

Il **test dei segni** (*the sign test*) per un campione è l'equivalente non parametrico del test t di Student per un campione, che ha ipotesi nulla

$$H_0: m = m_0$$

ed ipotesi alternativa sia bilaterale

$$H_1: \mu \neq \mu_0$$

che unilaterale

$$H_1: \mu < \mu_0 \quad \text{oppure} \quad H_1: \mu > \mu_0$$

Il **test dei segni** è il **test di posizione più semplice**, che al posto della media **come misura di tendenza centrale utilizza la mediana**, sia nella metodologia, sia nell'ipotesi nulla

$$H_0: me = me_0$$

che in quelle alternative bilaterali

$$H_1: me \neq me_0$$

o unilaterali

$$H_1: me < me_0 \quad \text{oppure} \quad H_1: me > me_0$$

La differenza fondamentale consiste nel fatto che al posto della distribuzione t di Student **si avvale della distribuzione binomiale, bene approssimata dalla distribuzione normale, nel caso di grandi campioni.**

Il **test dei segni rappresenta una delle procedure più antiche nella statistica inferenziale.** E' stato utilizzato già nei primi anni del 1700 da Arbuthnot, per verificare se a Londra il rapporto fra maschi e femmine alla nascita superava il valore di 0,5. In tempi più recenti, ma sempre nelle fasi iniziali della statistica moderna, è stato riproposto da Sir R. A. Fisher nel suo testo "*Statistical methods for research workers*" del 1925.

Nella ricerca sul campo ed in laboratorio, è frequente il caso in cui non tutti i dati hanno la stessa precisione o attendibilità. Nelle misure strumentali, quasi sempre si valutano correttamente quantità intorno alla media; ma sovente non si riesce a determinare valori troppo piccoli, che vengono indicati con minore di X, e/o valori molto grandi, fuori scala indicati con maggiore di Y. La serie dei dati riporta quantità intorno alla tendenza centrale, che possono essere ordinati ed altri agli estremi con molte sovrapposizioni.

Ad esempio, disponendo di un campione di N (12) osservazioni già ordinate in modo crescente,

$$<1 \quad <1 \quad 1 \quad 2 \quad 4 \quad 5 \quad 8 \quad 9 \quad 10 \quad 12 \quad 19 \quad >20$$

può sorgere il problema di valutare se la mediana (me) è significativamente minore di un valore di confronto, indicato nel caso in 15 ( $me_0$ ).

L'ipotesi nulla è

$$H_0: \mu_e = \mu_{e_0}$$

mentre l'ipotesi alternativa

$$H_1: \mu_e < \mu_{e_0}$$

è unilaterale.

**La procedura del test dei segni** per un campione è semplice:

- 1 - si confronta ogni punteggio con il valore di paragone (15), trasformando in segni negativi i punteggi inferiori ed in segni positivi quelli maggiori, ottenendo

- - - - - + +

- 2 - si contano i segni negativi (10) ed i segni positivi (2); la scala utilizzata dovrebbe essere continua e quindi **non dovrebbero esistere valori uguali a quello di confronto, che danno una differenza di 0 da esso; qualora esistessero, le differenze uguali a 0 devono essere ignorate, con una pari riduzione delle dimensioni N del campione;**

- 3 - se fosse vera l'ipotesi nulla, i segni negativi e quelli positivi dovrebbero essere approssimativamente uguali, con differenze imputabili alla casualità; si sceglie uno dei due valori, di solito quello minore (2): se è vera l'ipotesi nulla, dovrebbe non discostarsi troppo da  $N/2$ , corrispondente a 6 con i dati dell'esempio;

- 4 - con la distribuzione binomiale,

$$C_N^r \cdot p^r \cdot q^{N-r}$$

nella quale  $N = 12$      $r = 2$      $p = q = 1/2$

si stima la probabilità di trovare la distribuzione osservata e quelle più estreme nella stessa direzione (quindi per  $r$  che varia da 2 a 0); per evitare tanti calcoli si può ricorrere a tabelle che già forniscono le probabilità cumulate, per  $p = 1/2$ , con  $N$  e  $r$  che variano fino a 20 (riportata nella pagina successiva);

- 5 - applicando la distribuzione binomiale, si somma la probabilità relativa alla distribuzione osservata ( $r = 2$ ) con quelle più estreme nella stessa direzione; se insieme danno un valore inferiore alla probabilità  $\alpha$  prefissata (di solito 0.05), si può rifiutare l'ipotesi nulla in un test unilaterale.

**Nella tabella seguente è riportata la probabilità cumulata di r, per N che varia da 6 a 20.**

**PROBABILITA' CUMULATE DELLA DISTRIBUZIONE BINOMIALE**

$$C_N^r \cdot p^r \cdot q^{N-r}$$

N = numero di osservazioni

r = numero minore tra segni positivi e negativi

**N**

<b>r</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>
<b>0</b>	0.016	0.008	0.004	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<b>1</b>	0.109	0.062	0.035	0.020	0.011	0.006	0.003	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000
<b>2</b>	0.344	0.227	0.144	0.090	0.055	0.033	0.019	0.011	0.006	0.004	0.002	0.001	0.001	0.000	0.000
<b>3</b>	0.656	0.500	0.363	0.254	0.172	0.113	0.073	0.046	0.029	0.018	0.011	0.006	0.004	0.002	0.001
<b>4</b>	0.891	0.773	0.637	0.500	0.377	0.274	0.194	0.133	0.090	0.059	0.038	0.024	0.015	0.010	0.006
<b>5</b>	0.984	0.938	0.856	0.746	0.623	0.500	0.387	0.291	0.212	0.151	0.105	0.072	0.048	0.032	0.021
<b>6</b>	1.00	0.992	0.965	0.910	0.828	0.726	0.613	0.500	0.395	0.304	0.227	0.166	0.119	0.084	0.058
<b>7</b>		1.00	0.998	0.980	0.945	0.887	0.806	0.710	0.605	0.500	0.402	0.314	0.240	0.180	0.132
<b>8</b>			1.00	0.996	0.989	0.967	0.927	0.867	0.788	0.696	0.598	0.500	0.407	0.324	0.252
<b>9</b>				1.00	0.999	0.994	0.981	0.954	0.910	0.849	0.773	0.686	0.593	0.500	0.412
<b>10</b>					1.00	0.999	0.997	0.989	0.971	0.941	0.895	0.834	0.760	0.676	0.588
<b>11</b>						1.00	1.00	0.998	0.994	0.982	0.962	0.928	0.881	0.820	0.748
<b>12</b>							1.00	1.00	0.999	0.996	0.989	0.976	0.952	0.916	0.868
<b>13</b>								1.00	1.00	1.00	0.998	0.994	0.985	0.968	0.942
<b>14</b>									1.00	1.00	1.00	0.999	0.996	0.990	0.979
<b>15</b>										1.00	1.00	1.00	0.999	0.998	0.994
<b>16</b>											1.00	1.00	1.00	1.00	0.999
<b>17</b>												1.00	1.00	1.00	1.00
<b>18</b>													1.00	1.00	1.00
<b>19</b>														1.00	1.00
<b>20</b>															1.00

Con i dati dell'esempio,  $N = 12$  e  $r = 2$

la tabella riporta una probabilità uguale a 0.019, corrispondente a 1,9% quando espressa in percentuale. Essa significa che, se fosse vera l'ipotesi nulla, si ha una probabilità pari a 1,9% di trovare per caso una risposta uguale a quella trovata o ancor più estrema.

E' piccola, inferiore a 5%; di conseguenza, si rifiuta l'ipotesi nulla ed implicitamente si accetta quella alternativa.

6 - Per un test bilaterale, e quindi con ipotesi alternativa

$$H_1: \mu \neq \mu_0$$

poiché la distribuzione binomiale è simmetrica quando  $p = 1/2$  come atteso nell'ipotesi nulla, si deve moltiplicare la probabilità calcolata per 2: si rifiuta l'ipotesi nulla, quando questo ultimo valore è inferiore ad  $\alpha$ .

Con i dati dell'esempio, l'ipotesi bilaterale ha una probabilità pari a 3,8% ( $1,9 \times 2$ ); di conseguenza anche in questo caso si sarebbe rifiutata l'ipotesi nulla, ovviamente con una probabilità di errare pari a 3,8%.

**Per  $N > 12$  la distribuzione binomiale è già ritenuta sufficientemente grande per essere giudicata come approssimativamente normale; altri autori, più rigorosi, spostano questo limite intorno  $N > 20$  osservazioni.**

Per una distribuzione asintoticamente normale, si utilizza la distribuzione z

$$z = \frac{X - m}{s}$$

in cui

-  $X$  è il numero di segni positivi oppure negativi (di solito, in molti test viene consigliato di **scegliere il numero minore, per motivi pratici**, collegati alla tabella delle probabilità),

-  $m$  è il numero atteso del segno prescelto ed è uguale a  $N/2$  (con  $N$  = numero d'osservazioni),

mentre

-  $s$  è uguale a  $\sqrt{N \cdot p \cdot q} = \sqrt{N \cdot \frac{1}{4}} = \frac{1}{2} \sqrt{N}$

**Passando dalla distribuzione binomiale a quella normale, quindi da una misura discreta ad una continua, si deve apportare il termine di correzione per la continuità**, come illustrato in vari esercizi d'applicazione della distribuzione normale.

ESEMPIO. Da una serie di rilevazioni sulla quantità delle specie presenti in alcuni ambienti, sono stati derivati i seguenti 20 valori di biodiversità, già ordinati in modo crescente:

2,5 2,7 2,9 2,9 3,1 3,1 3,1 3,8 3,9 4,2 4,5 4,9 5,3 6,5 6,5 8,9 9,7 11,7 15,7 18,9

Si vuole valutare se la tendenza centrale di questa serie di rilevazioni è significativamente differente da 6,5 risultato il valore centrale dell'area in studi precedenti.

Risposta.

Per verificare l'ipotesi nulla

$$H_0: \mu = 6,5$$

con ipotesi alternativa bilaterale

$$H_1: \mu \neq 6,5$$

con il test dei segni, si calcolano le 20 differenze.

Poiché 2 risultano uguali a 0, restano  $N = 18$  osservazioni, delle quali solo 5 maggiori della mediana, per cui  $r = 5$ .

In un test unilaterale occorre calcolare le sei probabilità  $P_i$

$$P_i = C_{18}^r \cdot 0,5^r \cdot 0,5^{18-r}$$

con  $r$  che varia da 5 a 0.

La tabella delle probabilità cumulate in una distribuzione binomiale con  $p = q = 1/2$ , all'incrocio della colonna  $N = 18$  e della riga  $r = 5$  riporta 0,048.

Pertanto, in una serie di 18 misure, la probabilità di trovare per caso 5 valori positivi, o meno, è uguale a 4,8%. Poiché il test è bilaterale, si deve considerare anche la probabilità di avere 5 valori negativi.

In una distribuzione simmetrica come la binomiale con  $p = 1/2$ , la probabilità è uguale alla precedente; di conseguenza, si deve concludere che la probabilità di trovare scarti dall'atteso che siano uguali o superiori a quello trovato è uguale a 9,6%. Non è possibile rifiutare l'ipotesi nulla.

Per dimostrare come 18 osservazioni possano già essere considerate un grande campione ai fini pratici e per richiamare alla mente l'uso della distribuzione normale al posto della binomiale, con i dati dell'esempio si stima

$$z = \frac{|5 - 9| - 0,5}{\sqrt{18 \cdot \frac{1}{4}}} = \frac{3,5}{2,12} = 1,65$$

un valore di  $z$  uguale a 1,65.

In una coda della distribuzione normale corrisponde ad una probabilità di circa 0,047 **quasi identica a quella fornita (0,048) dalla distribuzione binomiale cumulata.**

In particolare quando si dispone di pochi dati, **nella scelta del test più adatto** insieme con il tipo di scala utilizzato e le caratteristiche della distribuzione dei dati **assume importanza rilevante anche la potenza-efficienza dei test a confronto.**

Come già ripetuto, l'ipotesi

$$H_0: \mu = \mu_0$$

in campo parametrico è verificata mediante il test t di Student. Pertanto, **l'efficienza asintotica relativa del test dei segni** deve essere confrontata con il test t di Student.

Poiché la potenza-efficienza di un test dipende dalla forma di distribuzione dei dati, in vari testi di statistica, dei quali si riportano solo le conclusioni, i confronti sono fatti nelle condizioni che i dati abbiano

- una distribuzione normale,
- una distribuzione rettangolare,
- una distribuzione esponenziale doppia.

La **potenza-efficienza relativa del test dei segni rispetto al test t di Student**

- con una **distribuzione normale** dei dati è uguale a circa **0,64 (2/p)**,
- con una **distribuzione rettangolare** dei dati è uguale a **0,33 (1/3)**,
- con una **distribuzione esponenziale doppia** è uguale a **2**.

Significa che, per avere la stessa probabilità di rifiutare l'ipotesi nulla, per ogni 100 dati ai quali sia stato applicato il test dei segni, il test t di Student richiede

- 64 dati nel caso di una distribuzione normale,
- 33 nel caso di una distribuzione rettangolare,
- 200 nel caso di una esponenziale doppia.

**Quando i dati hanno una distribuzione fortemente asimmetrica, il test dei segni si fa preferire al test t di Student non solo per il rispetto delle condizioni di validità, ma anche perché è più potente.**

#### **6.4. INTERVALLO DI CONFIDENZA DI UNA PROBABILITÀ O FREQUENZA RELATIVA, SECONDO IL METODO DI CLOPPER E PEARSON**

La probabilità media ( $\bar{p}$ ) di un evento, calcolata su un campione di dati raccolti in natura o con prove ripetute in laboratorio, è data dal rapporto tra i casi positivi (B) e il numero totale di osservazioni (N)

$$\bar{p} = B / N$$

La sua deviazione standard ( $S_p$ ) dipende da  $\bar{p}$  e da N, secondo la relazione

$$s_p = \sqrt{\frac{\bar{p} \cdot (1 - \bar{p})}{N}}$$

La varianza non è quindi misurata utilizzando la variabilità campionaria delle repliche, ma è una funzione **del valore medio** (come già illustrato in precedenza) sommando tutti i risultati.

Calcolata una probabilità o frequenza relativa dai dati di un campione, spesso si pone il problema di **conoscere la probabilità (p) vera o della popolazione**. E' una domanda alla quale la statistica frequentista risponde attraverso l'intervallo di confidenza.

Nel caso di **campioni molto piccoli (N da 1 a 10)**, uno dei metodi più semplici e rapidi per stimare l'intervallo di confidenza o intervallo fiduciale è **la procedura grafica, proposta da Clopper C. J. e Pearson E. S.** nel 1934 (nell'articolo *The use of confidence or fiducial limits illustrated in the case of binomial* comparso su *Biometrika* 26, pp.404-413). Di essa vengono riportati solo i risultati e la loro applicazione.

**Alcuni valori fondamentali (per  $\alpha = 0.01, 0.05, 0.10, 0.20$ ) di una distribuzione bilaterale sono riportati nelle tabelle della pagina successiva.**

L'intervallo di confidenza alla probabilità  $1 - \alpha$  della probabilità media  $\bar{p}$ , compreso tra il valore del limite inferiore  $p_I(\alpha/2)$  e il valore del limite superiore  $p_S(\alpha/2)$  può essere scritto come

$$Pp[ p_I(\alpha/2) < \bar{p} < p_S(\alpha/2) ] \cong 1 - \alpha$$

**In termini pratici, le risposte sono fornite direttamente dalle due tabelle seguenti.**

Ad esempio

1) assumendo di avere avuto una risposta positiva su 4 tentativi ( $N = 4$  e  $B = 1$ ), l'intervallo fiduciale del valore medio  $\bar{p}$  uguale a 0,25 è compreso tra 0,0063 e 0,8059 alla probabilità 5% e tra 0,0127 e 0,7514 alla probabilità 10%;

2) assumendo 4 risposte positive su 8 tentativi ( $N = 8$  e  $B = 4$ ), l'intervallo fiduciale del valore medio  $\bar{p}$  uguale a 0,50 è compreso tra 0,1570 e 0,8430 alla probabilità 5% e tra 0,1929 e 0,8071 alla probabilità 10%.

**Tabella dei limiti di confidenza di una probabilità  $P = B/N$  (0.00 £  $P$  £ 1.00 )  
calcolati con il metodo grafico di Clopper e Pearson (1934)**

N = dimensioni del campione (da 1 a 10); B = numero di successi (varia da 0 a N)

$\alpha$  = probabilità bilaterale;  $P_I(\alpha/2)$  = limite critico inferiore;  $P_S(\alpha/2)$  = limite critico superiore

B	$\alpha$	N = 1		N = 2		N = 3		N = 4		N = 5	
		$P_I(a/2)$	$P_S(a/2)$								
0	.010	.0000	.9950	.0000	.9293	.0000	.8290	.0000	.7341	.0000	.6534
	.020	.0000	.9900	.0000	.9000	.0000	.7846	.0000	.6838	.0000	.6019
	.050	.0000	.9750	.0000	.8419	.0000	.7076	.0000	.6024	.0000	.5218
	.100	.0000	.9500	.0000	.7764	.0000	.6316	.0000	.5271	.0000	.4507
	.200	.0000	.9000	.0000	.6838	.0000	.5358	.0000	.4377	.0000	.3690
1	.010	.0050	1.0000	.0025	.9975	.0017	.9586	.0013	.8891	.0010	.8149
	.020	.0100	1.0000	.0050	.9950	.0033	.9411	.0025	.8591	.0020	.7779
	.050	.0250	1.0000	.0126	.9874	.0084	.9057	.0063	.8059	.0051	.7164
	.100	.0500	1.0000	.0253	.9747	.0170	.8647	.0127	.7514	.0102	.6574
	.200	.1000	1.0000	.0513	.9487	.0345	.8042	.0260	.6795	.0209	.5839
2	.010			.0707	1.0000	.0414	.9983	.0294	.9706	.0229	.9172
	.020			.1000	1.0000	.0589	.9967	.0420	.9580	.0327	.8944
	.050			.1581	1.0000	.0943	.9916	.0676	.9324	.0527	.8534
	.100			.2236	1.0000	.1353	.9830	.0976	.9024	.0764	.8107
	.200			.3162	1.0000	.1958	.9655	.1426	.8574	.1122	.7534
3	.010					.1710	1.0000	.1109	.9987	.0828	.9771
	.020					.2154	1.0000	.1409	.9975	.1056	.9673
	.050					.2924	1.0000	.1941	.9937	.1466	.9473
	.100					.3684	1.0000	.2486	.9873	.1893	.9236
	.200					.4642	1.0000	.3205	.9740	.2466	.8878
4	.010							.2659	1.0000	.1851	.9990
	.020							.3162	1.0000	.2221	.9980
	.050							.3976	1.0000	.2836	.9949
	.100							.4729	1.0000	.3426	.9898
	.200							.5623	1.0000	.4161	.9791
5	.010									.3466	1.0000
	.020									.3981	1.0000
	.050									.4782	1.0000
	.100									.5493	1.0000
	.200									.6310	1.0000

CONTINUA NELLA PAGINA SEGUENTE

B	$\alpha$	N = 6		N = 7		N = 8		N = 9		N = 10	
		$p_1(a/2)$	$p_s(a/2)$								
0	.010	.0000	.5865	.0000	.5309	.0000	.4843	.0000	.4450	.0000	.4113
	.020	.0000	.5358	.0000	.4821	.0000	.4377	.0000	.4005	.0000	.3690
	.050	.0000	.4593	.0000	.4096	.0000	.3694	.0000	.3363	.0000	.3085
	.100	.0000	.3930	.0000	.3482	.0000	.3123	.0000	.2831	.0000	.2589
	.200	.0000	.3187	.0000	.2803	.0000	.2501	.0000	.2257	.0000	.2057
1	.010	.0008	.7460	.0007	.6849	.0006	.6315	.0006	.5850	.0005	.5443
	.020	.0017	.7057	.0014	.6434	.0013	.5899	.0011	.5440	.0010	.5044
	.050	.0042	.6412	.0036	.5787	.0032	.5265	.0028	.4825	.0025	.4450
	.100	.0085	.5818	.0073	.5207	.0064	.4707	.0057	.4291	.0051	.3942
	.200	.0174	.5103	.0149	.4526	.0131	.4062	.0116	.3684	.0105	.3369
2	.010	.0187	.8564	.0158	.7970	.0137	.7422	.0121	.6926	.0109	.6482
	.020	.0268	.8269	.0227	.7637	.0197	.7068	.0174	.6563	.0155	.6117
	.050	.0433	.7772	.0367	.7096	.0319	.6509	.0281	.6001	.0252	.5561
	.100	.0628	.7287	.0534	.6587	.0464	.5997	.0410	.5496	.0368	.5069
	.200	.0926	.6668	.0788	.5962	.0686	.5382	.0608	.4901	.0545	.4496
3	.010	.0663	.9337	.0553	.8823	.0475	.8303	.0416	.7809	.0370	.7351
	.020	.0847	.9153	.0708	.8577	.0608	.8018	.0534	.7500	.0475	.7029
	.050	.1181	.8819	.0990	.8159	.0852	.7551	.0749	.7007	.0667	.6525
	.100	.1532	.8468	.1288	.7747	.1111	.7108	.0978	.6551	.0873	.6066
	.200	.2009	.7991	.1696	.7214	.1469	.6554	.1295	.5994	.1158	.5517
4	.010	.1436	.9813	.1177	.9447	.0999	.9001	.0868	.8539	.0768	.8091
	.020	.1731	.9732	.1423	.9292	.1210	.8790	.1053	.8290	.0932	.7817
	.050	.2228	.9567	.1841	.9010	.1570	.8430	.1370	.7880	.1216	.7376
	.100	.2713	.9372	.2253	.8712	.1929	.8071	.1687	.7486	.1500	.6965
	.200	.3332	.9074	.2786	.8304	.2397	.7603	.2104	.6990	.1876	.6458
5	.010	.2540	.9992	.2030	.9842	.1697	.9525	.1461	.9132	.1283	.8717
	.020	.2943	.9983	.2363	.9773	.1982	.9392	.1710	.8947	.1504	.8496
	.050	.3588	.9958	.2904	.9633	.2449	.9148	.2120	.8630	.1871	.8129
	.100	.4182	.9915	.3413	.9466	.2892	.8889	.2514	.8313	.2224	.7776
	.200	.4897	.9826	.4038	.9212	.3446	.8531	.3010	.7896	.2673	.7327
6	.010	.4135	1.0000	.3151	.9993	.2578	.9863	.2191	.9584	.1909	.9232
	.020	.4642	1.0000	.3566	.9986	.2932	.9803	.2500	.9466	.2183	.9068
	.050	.5407	1.0000	.4213	.9964	.3491	.9681	.2993	.9251	.2624	.8784
	.100	.6070	1.0000	.4793	.9927	.4003	.9536	.3449	.9022	.3035	.8500
	.200	.6813	1.0000	.5474	.9851	.4618	.9314	.4006	.8705	.3542	.8124
7	.010			.4691	1.0000	.3685	.9994	.3074	.9879	.2649	.9630
	.020			.5179	1.0000	.4101	.9987	.3437	.9826	.2971	.9525
	.050			.5904	1.0000	.4735	.9968	.3999	.9719	.3475	.9333
	.100			.6518	1.0000	.5293	.9936	.4504	.9590	.3934	.9127
	.200			.7197	1.0000	.5938	.9869	.5099	.9392	.4483	.8842
8	.010					.5157	1.0000	.4150	.9994	.3518	.9891
	.020					.5623	1.0000	.4560	.9989	.3883	.9845
	.050					.6306	1.0000	.5175	.9972	.4439	.9748
	.100					.6877	1.0000	.5709	.9943	.4931	.9632
	.200					.7499	1.0000	.6316	.9884	.5504	.9455
9	.010							.5550	1.0000	.4557	.9995
	.020							.5995	1.0000	.4956	.9990
	.050							.6637	1.0000	.5550	.9975
	.100							.7169	1.0000	.6058	.9949
	.200							.7743	1.0000	.6631	.9895
10	.010									.5887	1.0000
	.020									.6310	1.0000
	.050									.6915	1.0000
	.100									.7411	1.0000
	.200									.7943	1.0000

**Con il metodo di Clopper e Pearson, l'intervallo fiduciale è simmetrico solo per i valori medi di  $\bar{p}$  uguali a 0,50.**

ESEMPIO. Per stimare la diffusione geografica di una specie, sono stati campionati cinque siti e la specie è stata trovata in due. Quale è la percentuale di presenza  $p$  reale di questa specie alla probabilità  $\alpha = 0.05$ , considerando l'universo dei siti?

Risposta.

Il campione è di dimensioni estremamente ridotte e quindi per l'intervallo di confidenza è utile ricorrere alla tabella costruita con il metodo grafico di Clopper-Pearson.

Per  $N$  uguale a 5 e  $B$  uguale a 2 (corrispondente ad una  $\bar{p}$  media di 0,4), alla probabilità 0.05 nella tabella viene riportato un intervallo di confidenza compreso tra 0,0527 e 0,8534.

Tradotto in percentuale, si può affermare che, con probabilità pari al 95%, la frequenza reale  $p$  di siti in cui è presente la specie studiata varia tra 5,27% e 85,34%. E' semplice osservare che, **rispetto al valore medio del campione (0,40), la distribuzione non è simmetrica.**

Per **grandi campioni**, (anche se non è mai definito chiaramente il confine tra campioni piccoli e grandi) si può **ricorrere alla distribuzione normale**, essendo la varianza definita dalla media (e quindi nota), come già presentato nel capitolo IV. La frequenza ( $p$ ) reale o della popolazione si trova, con probabilità  $1-\alpha$ , entro l'intervallo

$$\pi = \bar{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\bar{p} \cdot (1 - \bar{p})}{N}}$$

E' una distribuzione bilaterale; di conseguenza,

- alla probabilità complessiva  $\alpha = 0.05$  in ognuna delle due code si deve prendere  $\alpha = 0.025$  alla quale corrisponde un valore di  $z$  uguale a 1,96;
- alla probabilità complessiva  $\alpha = 0.01$  in ognuna delle due code si deve prendere  $\alpha = 0.005$  alla quale corrisponde un valore di  $z$  uguale a 2,58.

ESEMPIO. Nelle misure d'inquinamento dell'aria, in una città su 25 zone campionate il limite di legge è stato superato in 6 casi. Quale è la frequenza reale  $p$  di superamento dei limiti di legge, alla probabilità  $\alpha = 0.05$ ?

Risposta.

Con  $\bar{p} = 6/25 = 0,24$      $N = 25$      $z = 1,96$  alla probabilità  $\alpha = 0.05$  bilaterale, si ottiene un intervallo fiduciale della media reale

$$\pi = 0,24 \pm 1,96 \cdot \sqrt{\frac{0,24 \cdot 0,76}{25}} = 0,24 \pm 0,1674$$

che varia

da un limite inferiore uguale a 0,0726 e

un limite superiore uguale a 0,4074.

La frequenza reale di superamento dei limiti di legge, stimata con un campione di 25 osservazioni in una media pari a 24%, alla probabilità  $\alpha = 0.05$  varia tra 7,26% e 40,74%.

## 6.5. INTERVALLI DI CONFIDENZA NON PARAMETRICI E INTERVALLI DI TOLLERANZA

E' possibile costruire anche **intervalli di confidenza non parametrici**. Essi fanno riferimento non alla distribuzione normale (quando è nota la varianza  $s^2$  della popolazione) o alla distribuzione **t** di Student (quando si deve utilizzare la varianza campionaria  $s^2$ ), ma a distribuzioni non parametriche; oppure sono derivati da altre funzioni di distribuzione dei dati, diverse dalla normale.

Tra le più importanti funzioni di distribuzione dei dati è da ricordare la **disuguaglianza di Cebicev**, scritto anche come Tchebycheff (già discussa nel capitolo 2): è completamente generale ed è valida per qualsiasi forma di distribuzione. Essa consente di determinare intervalli di confidenza non parametrici, con coefficiente di confidenza non inferiore alla probabilità a **1-a** prefissata, in **assenza di qualsiasi informazione sulla forma di distribuzione della variabile casuale, eccetto l'esistenza della varianza**.

**Il limite fondamentale dell'intervallo di confidenza non parametrico deriva dal fatto che l'intervallo calcolato spesso è molto ampio**, troppo per avere applicazioni utili in varie situazioni. E' appunto l'effetto di non voler porre limiti alle sue applicazioni e richiedere che sia valido in tutte le situazioni.

Nel **caso di una probabilità**, è possibile rifarsi alla distribuzione bernouillana ed ottenere una stima dell'**intervallo di confidenza della probabilità p** mediante la relazione:

$$\pi = \bar{p} \pm \frac{1}{2\sqrt{Na}}$$

dove

$\bar{p}$  è la frequenza relativa media,

$N$  è il numero totale di casi analizzati,

$\alpha$  è la probabilità prefissata.

ESEMPIO (uguale al precedente, per un successivo **confronto tra intervallo fiduciale e intervallo di confidenza**). Nelle misure d'inquinamento dell'aria, in una città su 25 zone campionate il limite di legge è stato superato in 6 casi. Quale è la frequenza reale  $\pi$  di superamento dei limiti di legge, alla probabilità  $\alpha = 0.05$ ?

Risposta.

Con  $\bar{p} = 6/25 = 0,24$  e  $N = 25$  e  $\alpha = 0.05$

$$\pi = 0,24 \pm \frac{1}{2\sqrt{25 \cdot 0.05}} = 0,24 \pm \frac{1}{2,236} = 0,24 \pm 0,447$$

L'intervallo di confidenza non parametrico o intervallo di tolleranza al 5% come estremi fornisce i valori 0 e 0,687.

Per l'esatta comprensione del metodo, è importante osservare che **l'applicazione della formula produrrebbe un limite inferiore negativo** ( $0,24 - 0,447$ ), privo di senso e quindi **posto convenzionalmente uguale a 0**.

Una seconda osservazione è che il limite di confidenza, fondato sulla distribuzione normale, fornisce un intervallo nettamente minore (tra 7,26% e 40,74%) alla stessa probabilità  $\alpha$ : ma esso è valido solo per distribuzioni che siano normali, almeno in modo approssimato.

Quando è possibile costruire vari intervalli di confidenza sulla base di ipotesi diverse, si pone il problema di **scegliere quello "ottimale"**. Senza entrare nel dibattito tecnico, un criterio di preferenza sovente proposto è la lunghezza minima dell'intervallo, a parità di probabilità  $1-\alpha$  e a condizione che i dati campionari rispettino le condizioni di validità relative al test utilizzato.

Come già illustrato per il test t di Student, l'intervallo di confidenza permette la **verifica di ipotesi: non si può rifiutare l'ipotesi nulla  $H_0$  quando l'intervallo di confidenza include il valore di confronto**.

**Gli intervalli di tolleranza sono ottenuti dalla distribuzione dei valori**, ordinati spesso in rapporto alla mediana (come quartili e quantili, decili e centili), con le tecniche spiegate nel capitolo 1 sulla statistica descrittiva. Con gli intervalli di tolleranza è possibile individuare come variano i valori entro i quantili determinati, che di solito sono il 5%, il 10% o il 20%, in una sola coda od in entrambe, in rapporto al fatto che l'anomalia sia rappresentata da valori troppo alti, oppure troppo bassi, oppure troppo lontani dai valori centrali.

**Intervalli di confidenza parametrici e intervalli di tolleranza hanno usi analoghi, ma implicano concetti differenti:**

- un intervallo di confidenza è **un intervallo entro il quale si sostiene che sia contenuto un parametro (quasi sempre la media  $m$  o la varianza  $s^2$ , in questo caso la media  $p$  di una percentuale) fisso e ignoto, con fiducia prefissata;**
- un intervallo di tolleranza è **un intervallo del quale si sostiene che contiene una proporzione fissata dei valori assumibili da una variabile casuale, con probabilità prefissata.**

In sostanza, **l'intervallo di confidenza riguarda un'affermazione su un parametro costante ed ignoto, che sarà compreso o non compreso, mentre l'intervallo di tolleranza riguarda un'affermazione sulla variabile casuale (sulle singole osservazioni) nella sua interezza.**

## **6.6. IL TEST DEI SEGNI PER RANGHI DI WILCOXON**

Nel caso di una **variabile continua**, quindi ordinabile in ranghi senza sovrapposizioni (anche se alcune sono accettate) e che abbia **una distribuzione simmetrica**, l'ipotesi nulla sulla mediana

$$H_0: me = me_0$$

con ipotesi alternativa sia bilaterale che unilaterale può essere verificata anche mediante **il test, più potente di quello dei segni, proposto da Wilcoxon nel 1945** (con l'articolo *Individual comparison by ranking methods* pubblicato su *Biometrika* n. 1, pp. 80-83).

**Il test dei segni per ranghi di Wilcoxon (*the Wilcoxon signed rank test*), detto più semplicemente anche test T di Wilcoxon, nel caso di un campione permette di verificare se la tendenza centrale di una distribuzione si discosta in modo significativo da un qualsiasi valore prefissato di confronto. Analogo al test t di Student per il confronto tra una media campionaria ed una media attesa, come termini di confronto utilizza la mediana e viene utilizzato quando non è rispettata la normalità della distribuzione oppure i dati raccolti sono stati misurati con una scala di rango.**

La procedura del test dei segni di Wilcoxon **per un campione** può essere facilmente spiegata illustrando i vari passaggi di elaborazione dei dati, fino alle conclusioni.

Come esempio, si supponga di voler verificare se un terreno abbia una quantità unitaria di sostanze azotate pari a 300 oppure significativamente minore, come lascia supporre la qualità del prodotto; a questo scopo, su un campione di 13 lotti (indicati con lettere da A ad O) con caratteristiche geologiche e di coltivazione molto differenti, è stata misurata la quantità unitaria di sostanze azotate.

Sono state ottenuti i risultati seguenti:

A	B	C	D	E	F	G	H	I	L	M	N	O
235	230	180	250	280	330	440	430	260	225	240	235	215

I dati confermano l'ipotesi espressa?

Risposta.

Si tratta di un test ad una coda, con:

$$H_0: \text{mediana} = 300$$

$$H_1: \text{mediana} < 300$$

L'asimmetria dei dati, come indicano i due valori molto più alti (430 e 440), non permette di utilizzare il test t, ma richiede un test non parametrico.

Il metodo T di Wilcoxon è fondato su alcuni passaggi:

1 - Calcolare le differenze **d**, con relativo segno, tra i dati raccolti ed il valore dell'ipotesi nulla (eliminando le eventuali differenze, non presenti in questo esempio, che risultassero uguali a **zero**)

$$d = x - \text{mediana}$$

come riportato nella tabella sottostante

A	B	C	D	E	F	G	H	I	L	M	N	O
-65	-70	-120	-50	-20	+30	+140	+130	-40	-75	-60	-65	-85

2 - Calcolare i ranghi delle differenze **d**, considerate in valore assoluto (ordinare gli **N** valori assoluti dal minore al maggiore; se esistono valori che hanno lo stesso rango, assegnare ad ognuno di essi un punteggio dato dalla media dei loro ranghi).

Dai dati precedenti, si ottiene la seguente serie di ranghi:

A	B	C	D	E	F	G	H	I	L	M	N	O
6,5	8	11	4	1	2	13	12	3	9	5	6,5	10

3 - Attribuire ad ogni rango il segno della differenza, già calcolata al punto 1, per ottenere la stessa serie di ranghi della distribuzione precedente, ma con il relativo segno positivo oppure negativo:

A	B	C	D	E	F	G	H	I	L	M	N	O
-6,5	-8	-11	-4	-1	<b>+2</b>	<b>+13</b>	<b>+12</b>	-3	-9	-5	-6,5	-10

4 - Sommare i ranghi **R** dello stesso segno per calcolare **T** (ai fini del test è indifferente scegliere il valore minore o maggiore tra somma dei positivi e la somma dei negativi; abitualmente si sceglie il valore che in modulo è minore).

Con i dati dell'esempio, la scelta cade sulla somma dei ranghi con segno positivo (**R+**) appunto perché minore. (I dati sono già stati riportati in grassetto nella tabella precedente, per essere meglio evidenziati).

Da essi

$$\mathbf{T} = \sum \mathbf{R}_+ = 2 + 12 + 13 = 27$$

si ottiene un valore di **T** uguale a 27.

5 - Stimare il valore medio, al quale dovrebbe tendere la somma dei ranghi **T**, nella condizione che sia vera l'ipotesi nulla (ranghi positivi e negativi dovrebbero essere casualmente distribuiti e dare quindi la stessa somma, dipendente solo dal numero di dati).

La somma di **N** ranghi è

$$\mathbf{N(N+1)/2}$$

Quindi la media dei valori positivi o negativi  $\mu_T$ , che è la metà della somma di tutti i ranghi, dovrebbe essere

$$\mathbf{m}_T = \frac{N \cdot (N+1)}{4}$$

Applicata ai 13 dati dell'esempio, la media ( $\mu_T$ ) attesa nella condizione che sia vera l'ipotesi nulla, è

$$\mathbf{m}_T = \frac{13 \cdot 14}{4} = 45,5$$

uguale a 45,5.

6 - Se il valore espresso nell'ipotesi nulla (nell'esempio = 300) fosse la vera tendenza centrale della popolazione, la somma dei ranghi di segno positivo ( $T = 27$ ) non dovrebbe essere significativamente differente dalla media dei ranghi ( $\mu_T = 45,5$ ).

Nel **caso di piccoli campioni (N ≤ 20)**, la significatività è fornita dalla tavola che riporta il **valore critico inferiore** (vedere la tabella sottostante e quella, più dettagliata, riportata nella pagina successiva).

**Valori critici per il test dei ranghi con segno di Wilcoxon  
per test unilaterali e bilaterali, alle probabilità 0.05 e 0.01**  
(per campioni con N da 6 a 20)

	N	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
test a	a = 0.05	2	3	5	8	10	13	17	21	23	30	35	41	47	53	60
<b>1 coda</b>	a = 0.01	*	0	1	3	5	7	9	12	15	20	23	27	32	37	43
test a	a = 0.05	0	2	3	5	8	10	13	17	21	25	29	34	40	46	52
<b>2 code</b>	a = 0.01	*	*	0	1	3	5	7	9	12	15	19	23	27	32	37

Dove è riportato l'asterisco \*

il campione è troppo piccolo, per un test significativo al livello di probabilità  $\alpha$  stabilito.

Con i dati dell'esempio, per  $N = 13$  nella colonna  $a = 0.05$  **per un test unilaterale** il valore massimo di  $T$  è 21, al quale (nella tabella della pagina seguente) corrisponde una probabilità (calcolata in modo più preciso) di 0.0471.

Il valore  $T$  calcolato ( $T = 27$ ) con i dati dell'esempio è superiore a quello riportato nella tabella (21). Di conseguenza, nell'ipotesi che  $H_0$  sia vera, la probabilità di trovare un valore uguale o inferiore a 27 è superiore a 0.05.

Non si è in grado di rifiutare l'ipotesi nulla: la tendenza centrale dei dati raccolti non è significativamente minore di 300.

**Se il test fosse stato bilaterale**, quindi se la domanda fosse stata semplicemente se 300 poteva essere la tendenza centrale della nostra distribuzione, i valori di confronto per il  $T$  (sempre con  $N = 13$ ) sarebbero stati 17 per una probabilità  $a = 0.05$  e 9 per una probabilità  $a = 0.01$ .

**Tavola dei valori critici di T nel test di Wilcoxon per un campione  
e per due campioni dipendenti.**

Le probabilità sono riferite ad un test unilaterale. Per un test bilaterale occorre moltiplicare per 2 il valore di  $\alpha$ . Si può rifiutare l'ipotesi nulla alla probabilità  $\alpha$  se il valore di T calcolato sui dati è minore o uguale a quello riportato in grassetto alla colonna corrispondente.

Per i valori critici di T intorno al valore  $\alpha$  è riportata la probabilità esatta.

N	<b>T</b>	a = 0.05	<b>T</b>	a = 0.025	<b>T</b>	a = 0.01	<b>T</b>	a = 0.005
<b>5</b>	<b>0</b>	0.0313						
	1	0.0625						
<b>6</b>	<b>2</b>	0.0469	<b>0</b>	0.0156				
	3	0.0781	1	0.0313				
<b>7</b>	<b>3</b>	0.0391	<b>2</b>	0.0234	<b>0</b>	0.0078		
	4	0.0547	3	0.0391	1	0.0156		
<b>8</b>	<b>5</b>	0.0391	<b>3</b>	0.0195	<b>1</b>	0.0078	<b>0</b>	0.0039
	6	0.0547	4	0.0273	2	0.0117	1	0.0078
<b>9</b>	<b>8</b>	0.0488	<b>5</b>	0.0195	<b>3</b>	0.0098	<b>1</b>	0.0039
	9	0.0645	6	0.0273	4	0.0137	2	0.0059
<b>10</b>	<b>10</b>	0.0420	<b>8</b>	0.0244	<b>5</b>	0.0098	<b>3</b>	0.0049
	11	0.0527	9	0.0322	6	0.0137	4	0.0068
<b>11</b>	<b>13</b>	0.0415	<b>10</b>	0.0210	<b>7</b>	0.0093	<b>5</b>	0.0049
	14	0.0508	11	0.0269	8	0.0122	6	0.0068
<b>12</b>	<b>17</b>	0.0461	<b>13</b>	0.0212	<b>9</b>	0.0081	<b>7</b>	0.0046
	18	0.0549	14	0.0261	10	0.0105	8	0.0061
<b>13</b>	<b>21</b>	0.0471	<b>17</b>	0.0239	<b>12</b>	0.0085	<b>9</b>	0.0040
	22	0.0549	18	0.0287	13	0.0107	10	0.0052
<b>14</b>	<b>25</b>	0.0453	<b>21</b>	0.0247	<b>15</b>	0.0083	<b>12</b>	0.0043
	26	0.0520	22	0.0290	16	0.0101	13	0.0054
<b>15</b>	<b>30</b>	0.0473	<b>25</b>	0.0240	<b>19</b>	0.0090	<b>15</b>	0.0042
	31	0.0535	26	0.0277	20	0.0108	16	0.0051
<b>16</b>	<b>35</b>	0.0467	<b>29</b>	0.0222	<b>23</b>	0.0091	<b>19</b>	0.0046
	36	0.0523	30	0.0253	24	0.0107	20	0.0055
<b>17</b>	<b>41</b>	0.0492	<b>34</b>	0.0224	<b>27</b>	0.0087	<b>23</b>	0.0047
	42	0.0544	35	0.0253	28	0.0101	24	0.0055
<b>18</b>	<b>47</b>	0.0494	<b>40</b>	0.0241	<b>32</b>	0.0091	<b>27</b>	0.0045
	48	0.0542	41	0.0269	33	0.0104	28	0.0052
<b>19</b>	<b>53</b>	0.0478	<b>46</b>	0.0247	<b>37</b>	0.0090	<b>32</b>	0.0047
	54	0.0521	47	0.0273	38	0.0102	33	0.0054
<b>20</b>	<b>60</b>	0.0487	<b>52</b>	0.0242	<b>43</b>	0.0096	<b>37</b>	0.0047
	61	0.0527	53	0.0266	44	0.0107	38	0.0053

Nel **caso di grandi campioni ( $N > 20$ )**, sempre nella condizione che  $H_0$  sia vera, la somma dei ranghi dello stesso segno segue approssimativamente la distribuzione normale

$$Z \cong \frac{T - \mu_T}{\sigma_T}$$

dove

$\mu_T$  è calcolato con la formula precedente e

$\sigma_T$  è la deviazione standard di  $T$ , determinata solamente da  $N$  (il numero di dati)

secondo la relazione

$$s_T = \sqrt{\frac{N \cdot (N + 1) \cdot (2N + 1)}{24}}$$

La formula di approssimazione per grandi campioni fornisce buoni risultati già quando  $N$  è maggiore di 10 osservazioni. E' quindi possibile applicare le formule all'esercizio (con  $N = 13$ ), per calcolare prima la deviazione standard  $\sigma_T$

$$\sigma_T = \sqrt{\frac{13 \cdot 14 \cdot 27}{24}} = 14,31$$

e successivamente il valore di  $z$

$$Z = \frac{27 - 45,5}{14,31} = -1,29$$

Il valore di  $z = 1,29$  corrisponde ad una probabilità  $\alpha$  uguale a 0.0985 (o 9,85%) in una coda della distribuzione normale standardizzata.

Anche con questi calcoli, si deve concludere che i dati raccolti non permettono di rifiutare l'ipotesi nulla: il valore 300 può essere la tendenza centrale della popolazione dalla quale sono stati estratti i 13 valori campionari.

Il test dei segni per ranghi di Wilcoxon, come illustrato dalla metodologia, utilizza una quantità maggiore dell'informazione contenuta nei dati, rispetto al test dei segni.

Nei confronti del test dei segni, **la sua efficienza relativa asintotica**

- con una distribuzione normale dei dati è uguale a circa 1,50 (3/2);
- con una distribuzione rettangolare dei dati è uguale a 3,00 (3/1);
- con una distribuzione esponenziale doppia è uguale a 0,75 (3/4).

Significa che quando per un test con il T di Wilcoxon si hanno 100 dati, con il test dei segni servono 150 dati se la loro distribuzione è normale, 300 dati se la distribuzione è rettangolare e 75 se è esponenziale doppia.

**Il test T di Wilcoxon è più efficiente del test dei segni, eccetto quando la distribuzione dei dati è fortemente asimmetrica, come nel caso di una esponenziale doppia.**

Per la **scelta tra test parametrico e test non parametrico**, è importante il confronto tra test dei ranghi con segno di Wilcoxon e il test t di Student. L'**efficienza relativa asintotica del test T di Wilcoxon** rispetto al test t di Student

- con una distribuzione normale dei dati è uguale a circa 0,95 ( $3/\pi$ ),
- con una distribuzione rettangolare dei dati è uguale a 1,
- con una distribuzione esponenziale doppia è uguale a 1,50 ( $3/2$ ).

Significa che **il test T di Wilcoxon ha**

- **un grado di efficienza molto vicina a quella del test t di Student (seppure leggermente inferiore e pari a 0,95) quando la distribuzione dei dati è approssimativamente normale,**
- **un'efficienza uguale (1), quando la distribuzione è rettangolare,**
- **una sensibilmente maggiore quando ci si allontana dalla distribuzione normale (1,5).**

**Nella pratica della ricerca ambientale, in cui la distribuzione dei dati è spesso lontana dalla normalità, il test T di Wilcoxon è pertanto preferibile al test parametrico: assicura condizioni di validità più generali, senza perdere in potenza.**

Inoltre, il test T di Wilcoxon, pur richiedendo la simmetria della distribuzione dei dati, è molto **più robusto del test t di Student**: fornisce risultati attendibili, anche quando i dati si allontanano sensibilmente dalle condizioni teoriche di validità.

ESEMPIO 1. Con una serie di esperimenti è stato dimostrato che in condizioni ottimali di temperatura, in 15 giorni la popolazione di una specie di plancton aumenta in media del 45 per cento.

Per valutare l'effetto inibente delle basse temperature, sono stati misurati i tassi di crescita di 10 campioni, composti da un numero di individui molto variabile. La tabella sottostante riporta il tasso di crescita, in percentuale, di ogni campione:

A	B	C	D	E	F	G	H	I	L
22	28	30	15	48	37	50	24	29	36

Il loro tasso di crescita è significativamente minore del 45%?

Risposta.

Poiché i gruppi campionari sono di dimensioni nettamente diverse, i tassi di crescita misurati forniscono essenzialmente un'informazione di rango. Per confrontare tassi, rapporti o percentuali con test parametrici, è indispensabile che i dati siano riferiti a campioni di dimensioni simili; infatti campioni con dimensione diversa hanno un'attendibilità diversa, come ha dimostrato l'intervallo fiduciale di una media o di una percentuale.

Inoltre, nel caso di percentuali o rapporti a valori centrali differenti corrispondono varianze differenti; di conseguenza, l'uso di test parametrici richiederebbe la trasformazione dei dati (nel caso specifico servirebbe la trasformazione in arcoseno, che sarà trattata con le altre trasformazioni alla fine del capitolo X).

La domanda sulle percentuali di crescita richiede un test unilaterale; infatti, si vuole verificare se il valore centrale dei 10 dati sperimentali è inferiore al 45% stimato in condizioni ottimali.

Pertanto l'ipotesi nulla è

$$H_0: me = 45$$

e l'ipotesi alternativa unilaterale è

$$H_1: me < 45$$

Per rispondere a questo quesito, la procedura del test T di Wilcoxon prevede alcuni passaggi.

1 - Dapprima si calcolano le differenze rispetto a 45:

A	B	C	D	E	F	G	H	I	L
-23	-17	-15	-30	+3	-8	+5	-21	-15	-9

si ottengono 8 valori negativi e 2 positivi.

2 - Successivamente si trasformano nei ranghi, considerando i dati in valore assoluto,

A	B	C	D	E	F	G	H	I	L
9	7	5,5	10	1	3	2	8	5,5	4

e ad essi si attribuiscono i segni delle differenze già calcolate

A	B	C	D	E	F	G	H	I	L
-9	-7	-5,5	-10	+1	-3	+2	-8	-5,5	-4

3 - Si fanno le somme dei positivi e dei negativi; i dati con la somma minore sono quelli negativi, il cui valore di **T** è uguale a 3 (+1 e +2).

4 - Per la significatività, trattandosi di un campione piccolo, si ricorre alle tabelle dei valori critici. Con **N** = 10, la tabella dei valori critici nella colonna con  $\alpha = 0.05$  fornisce la probabilità relativa a **T** = 10, che risulta uguale a 0.0420. Il valore di **T** calcolato (uguale a 3) è nettamente minore; la probabilità che sia casuale è inferiore a 0.05.

Una lettura più dettagliata della tabella può evidenziare che per **N** = 10 nella colonna della probabilità  $\alpha = 0.01$  a **T** = 5 corrisponde una probabilità esatta  $\alpha = 0.0098$ . Il valore di **T** calcolato (uguale a 3) è minore di quello tabulato (uguale a 5) e quindi la probabilità che possa essere casuale è inferiore anche a 0.01.

In conclusione, si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa: con temperature più basse, i 10 campioni di plancton hanno avuto una crescita relativa la cui tendenza centrale è significativamente inferiore al tasso del 45%.

Secondo le indicazioni dei test di statistica applicata, con 10 osservazioni il campione è troppo piccolo per usare correttamente la distribuzione normale. Tuttavia, per dimostrare la robustezza del metodo e per ricordare il procedimento di calcolo della probabilità ad essa associata, è utile il suo uso anche ai dati di questo esempio.

Ricordando che per **N** = 10 e **T** = 3 nella condizione che l'ipotesi nulla sia vera, la media attesa  $\mu_T$

$$m_T = \frac{N \cdot (N + 1)}{4} = \frac{10 \cdot 11}{4} = 27,5$$

è uguale a 27,5

e la deviazione standard  $\sigma_T$  con la formula

$$s_T = \sqrt{\frac{N \cdot (N + 1) \cdot (2N + 1)}{24}}$$

risulta

$$s_T = \sqrt{\frac{10 \cdot 11 \cdot 21}{24}} = \sqrt{\frac{2310}{24}} = \sqrt{96,25} = 9,81$$

uguale a 9,81.

Da essi, con la distribuzione normale

$$z \cong \frac{T - m_i}{s_T}$$

si stima un valore di z

$$z = \frac{3 - 27,5}{9,81} = -2,497$$

uguale a 2,497 che può essere arrotondato in 2,50.

Nella tabella della distribuzione normale, ad un valore di 2,50 in una coda corrisponde una probabilità uguale a 0.00623 o 0,623%.

Per confrontare la probabilità stimata con z e quella riportata nella tabella per piccoli campioni, si può fare il calcolo per T = 10, che ha probabilità più alte, riportate nella tabella

$$z = \frac{10 - 27,5}{9,81} = -1,78$$

Il valore di z risulta uguale a 1,78 al quale corrisponde una probabilità uguale a 0.0378.

Si può osservare che è un valore vicino, ma inferiore, a quello riportato nella tabella (per N = 10 e T = 10) che risulta uguale a 0.0420.

**Il test delle successioni (seppure con ipotesi differenti) e il test T di Wilcoxon possono essere utilizzati anche nel caso di due campioni dipendenti.** Questi casi, che richiedono una metodologia simile a quella qui illustrata, saranno spiegati nei capitoli relativi.

## **6.7. CARATTERISTICHE DISTINTIVE DEI TEST SULLA BONTÀ DI ADATTAMENTO RISPETTO A QUELLI SU UN PARAMETRO**

Tra i test per un campione, devono essere inseriti anche quelli sulla **bontà di adattamento** (*goodness-of-fit test*). Essi servono per **verificare l'ipotesi che i dati campionari provengano da una variabile casuale di cui è nota la distribuzione di frequenza**, come già spiegato e discusso con il chi-quadrato, il test G e il test di Kolmogorov-Smirnov.

In questa serie di test per verificare l'accordo tra una distribuzione osservata ed una distribuzione attesa, rispetto ai concetti qui illustrati, **la differenza fondamentale è l'aspettativa del ricercatore, in rapporto all'ipotesi nulla.**

- Nei test inferenziali sulla media o sulla mediana, egli quasi sempre spera di rifiutare l'ipotesi nulla e quindi di dimostrare che la differenza riscontrata non è imputabile al caso, ma ad un fattore noto o da ricercare.
- Nei test sulla bontà dell'adattamento, egli si augura di non rifiutare l'ipotesi nulla e quindi di avere già individuato una curva che spieghi le caratteristiche della sua distribuzione osservata. Infatti, se rifiutasse l'ipotesi nulla egli dovrebbe concludere che la distribuzione teorica da lui indicata non è valida, senza che tale affermazione possa essere d'aiuto nell'individuare quale altra distribuzione teorica sia quella adatta.

Sulla valutazione della bontà di adattamento di una distribuzione, è didatticamente importante rileggere alcuni paragrafi di **un articolo del 1976** (Rivista di Statistica Applicata vol. 9 n. 4, pp. 239-255) **di Rodolfo Cristofaro**, uno dei maggiori statistici italiani degli ultimi decenni, sia per comprendere più esattamente quale valutazione dare ai risultati di questi test, sia per acquisire il linguaggio degli statistici.

“Il problema dell'adattamento di una distribuzione teorica ad un processo stocastico derivante dall'osservazione empirica di un fenomeno è stato generalmente affrontato nell'ambito della teoria dei test di ipotesi, contrapponendo all'ipotesi semplice

$$H_0: F(x) = F_0(x),$$

dove  $F_0(x)$  è una particolare funzione di ripartizione (continua o discreta),  
il suo complemento

$$H_1: F(x) \neq F_0(x)$$

Questo sistema d'ipotesi appare però poco adeguato nel valutare la bontà di adattamento di una distribuzione. Infatti, una distribuzione teorica può solo approssimare un processo stocastico derivante dall'osservazione di un fenomeno del mondo reale, così come qualunque modello teorico non riesce quasi mai a descrivere con assoluta esattezza una realtà empirica. Pertanto, se il numero  $N$  delle osservazioni disponibili è sufficientemente grande, qualunque test consistente scarterà l'ipotesi  $H_0$  anche se la distribuzione ipotizzata sotto  $H_0$  si approssima molto bene, pur se non in maniera perfetta, alla realtà. Ciò è stato sottolineato per la prima volta da J. Berkson nel 1938 (con l'articolo *Some difficulties of interpretation encountered in the application of chi-square test*. In *Journal of the American Statistical Association* vol 33, n. 256), con riferimento al test chi-quadro, tanto che egli giungeva a chiedersi se valga la pena di tormentarsi ad applicare un test quando è noto che con un campione abbastanza grande esso darà un risultato di significatività. Successivamente, J. Neyman nel 1949 (con l'articolo *Contribution to the theory of the test of the  $\chi^2$  test*. In *Proceed. First Berkeley Symposium on Mathem. Statist. and Probab.* Univ. of California

Press, 239), trattando l'argomento in maniera più tecnica, richiamava l'attenzione degli statistici sul fatto che qualunque test consistente basato sulla sola ipotesi nulla scarterà detta ipotesi con probabilità tendente a uno al crescere di  $N$  qualunque siano le reali differenze da  $H_0$  e quindi anche se esse sono piccolissime o trascurabili”.

Una strada che lo studioso dell'ambiente può percorrere, per uscire da questo apparente vicolo cieco, è una valutazione biologica o tecnica della differenza riscontrata come significativa. Anche in questo caso, può essere utile rileggere un altro passo dello stesso articolo, sempre in riferimento all'uso del test chi-quadro.

“ Un esempio può servire a chiarire i criteri di applicazione del test chi-quadro. A questo proposito si possono prendere in considerazione i dati di W.F.R. Weldon relativi all'adattamento di una distribuzione binomiale nel lancio di dadi, per un evento con probabilità di successo  $\pi = 1/3$ . Trattasi di dati, pubblicati ad esempio da R. A. Fisher (1925 e succ. ed.), che hanno formato oggetto di discussione già ai tempi di Weldon, tra lo stesso Weldon, K. Pearson e F. Galton, a causa dell'elevato numero di osservazioni che conduce a scartare l'ipotesi nulla oltre ogni usuale livello di significatività, malgrado il soddisfacente adattamento della distribuzione binomiale.

In particolare Weldon si lamentava in una lettera a F. Galton (riportata da E. S. Pearson nel 1965) che K. Pearson avesse rifiutato i suoi dati, sebbene essi non risultassero così “incredibili” come appariva dalle elaborazioni statistiche compiute dallo stesso K. Pearson. Naturalmente il giudizio di K. Pearson era ineccepibile, essendo molto verosimilmente i dati di Weldon imperfetti. Ma esistono in natura dati perfetti? In ogni caso Weldon aveva, a nostro avviso, validi motivi per lamentarsi, non avendo K. Pearson specificato l'entità dell'errore presente negli stessi dati”.

Nell'esempio di Weldon, erano stati fatti 26.306 lanci di due dadi apparentemente senza difetti, ottenendo nelle 11 classi (due dadi possono dare un totale che varia da 2 a 12) un chi-quadro uguale a 35,491 (ricordando che con d.f. 10 alla probabilità  $\alpha = 0.01$  il valore critico è uguale a 23,209 e alla probabilità  $\alpha = 0.001$  è uguale a 29,588). I due dadi, seppure senza apparenti anomalie, non erano quindi perfetti.

Per uscire da questo dilemma teorico, un modo pratico e semplice consiste nel valutare quale sia in percentuale lo scostamento tra valori osservati e i valori attesi nelle varie classi.

## **6.8. IL TEST DI LILLIEFORS E IL TEST DI CRAMÉR E VON MISES**

Quando si dispone di una serie di dati campionari, un caso che ricorre con relativa frequenza è la verifica del suo **accordo con la distribuzione normale, senza che  $m$  e  $s^2$  siano noti. E' una**

**informazione importante, sia per gli aspetti descrittivi, sia per la scelta del test. Si deve quindi ricorrere alla media e alla varianza campionaria, per verificare**

l'ipotesi nulla che la distribuzione X sia almeno approssimativamente Normale

$$H_0: X \sim N(\mu, \sigma^2)$$

contro l'ipotesi alternativa che la distribuzione X sia diversa dalla Normale

$$H_1: X \neq N(\mu, \sigma^2)$$

Tra i diversi metodi, uno dei più diffusi è il **test di Lilliefors**.

Il tempo richiesto dalla serie di operazioni è lungo. Di conseguenza, è opportuno utilizzare un programma informatico. Tuttavia, la procedura può essere presentata con semplicità mediante la descrizione dei passaggi fondamentali, utili più per comprenderne la logica che non per giungere a un'applicazione mediante calcoli manuali.

Anche in questo caso, la procedura è stata suddivisa in vari passaggi elementari:

1- Dato il campione osservato

$$x_1, x_2, \dots, x_N$$

con  $N$  osservazioni o repliche,

2- si stimano la sua media  $\bar{x}$  e la sua varianza  $s^2$ .

3 - Da esse si deriva il **campione standardizzato** ( $z_1, z_2, \dots, z_N$ ) attuando, per ogni dato campionario, la trasformazione (detta appunto standardizzazione dei dati)

$$z_i = \frac{x_i - \bar{x}}{s}, \quad i = 1, 2, \dots, N$$

ed ottenendo la ripartizione empirica  $F(z)$ .

4 - Dopo il raggruppamento in classi dei dati campionari trasformati, si stimano le frequenze teoriche relative, utilizzando la distribuzione normale (come spiegato nel capitolo 2) ed ottenendo la distribuzione Normale  $\Phi(z)$  corrispondente.

5 - Quindi si confronta la funzione di ripartizione empirica  $F(z)$  con la funzione di ripartizione della distribuzione Normale  $\Phi(z)$  della variabile casuale  $Z \sim N(0,1)$ , calcolando la differenza per ogni classe.

6 - Come indicatore della distanza (**D**) tra le due distribuzioni, si sceglie la differenza massima in valore assoluto:

$$D = \text{Diff. Massima } |F(z) - \Phi(z)|, \text{ per ogni } -\infty < z < +\infty$$

7 – Infine si stima il valore critico,

$$d\alpha / \sqrt{N}$$

dove

$d\alpha$  è il valore stimato di  $d$ , per la probabilità  $\alpha$ ,

da Lilliefors (1967) e più tardi migliorato da Dallal e Wilkinson (1986).

8 – Si rifiuta  $H_0$  quando la differenza massima  $D$  riscontrata è superiore al valore critico

$$D > d\alpha / \sqrt{N}$$

Per  $N > 30$  (al di sotto di queste dimensioni del campione con questo test è difficile valutare la normalità della distribuzione) Dallal e Wilkinson propongono i seguenti valori di  $d$  alle varie probabilità di  $\alpha$

$\alpha$	0.10	0.05	0.01	0.001
$d$	0,82	0,89	1,04	1,22

Il test di **Cramér e von Mises** rappresenta uno sviluppo della logica del **test di Kolmogorov-Smirnov**, applicata in modo specifico al caso di una variabile casuale continua, per verificare l'accordo tra una distribuzione campionaria ed una distribuzione attesa di qualsiasi forma.

Come in tutti i test per la bontà dell'adattamento, l'ipotesi riguarda tutti i parametri della distribuzione: quando il test risulta significativo, **la distribuzione osservata si differenzia da quella attesa per almeno un parametro**, senza alcuna informazione su quale esso sia.

Limitando anche in questo caso la spiegazione alla comprensione dei programmi informatici,

1 - dopo aver costruito le  $n$  classi della distribuzione osservata e della distribuzione attesa,

2 - si stima il valore di un indicatore  $w_n^2$  che è uguale a

$$w_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left[ F(x_i) - \frac{2i-1}{2n} \right]^2$$

dove

$F(x_i)$  è lo scarto tra osservato ed atteso nella classe  $i$ .

3 - Si rifiuta l'ipotesi nulla, quando  $w_n^2$  supera il valore critico  $C\alpha$  (riportato nella tabella seguente) stimato per il test.

Con  $n > 10$ , si possono usare i seguenti valori critici alla probabilità  $\alpha$  prefissata.

$\alpha$	0.10	0.05	0.01	0.001
C	0,347	0,461	0,743	1,168

Quando la distribuzione attesa è costruita sulla base di uno o più parametri (media, varianza, simmetria, curtosi) calcolati nella distribuzione osservata, il valore di  $w_n^2$  è inferiore. Sono stati stimati valori critici anche per queste analisi più specifiche.

Tra i test non parametrici per un campione, che in questo periodo assumono importanza crescente nella ricerca ambientale, può essere ricordato **il test dei segni per il trend** (*the sign test for trend*) proposto da **Cox e Stuart** nel 1955. Serve per verificare se la serie dei dati, raccolti in successione temporale o ordinati secondo la distanza da un'origine, che può essere anche geografica, hanno una tendenza monotonica alla crescita oppure alla diminuzione. **Il test è presentato in modo dettagliato nel capitolo specifico dedicato all'analisi delle tendenze**, insieme ad altri test come quello di Page e quello di Jonckheere, che in disegni sperimentali diversi servono per la verifica di ipotesi simili.

## CAPITOLO VII

### METODI NON PARAMETRICI PER DUE CAMPIONI

#### 7.1. TEST PER 2 CAMPIONI DIPENDENTI O PER DATI APPAIATI

Per confrontare l'effetto di due trattamenti in esperimenti di laboratorio o di due situazioni ambientali verificate sul campo, si possono usare i test statistici per due campioni. (Le caratteristiche distintive di due campioni dipendenti e due campioni indipendenti sono già state trattate ampiamente nel capitolo relativo al test  $t$  di Student).

In rapida sintesi, due gruppi a confronto possono essere formati in due modi diversi:

- 1) con gli stessi individui o coppie di individui scelti come simili e misurati in due differenti situazioni, per cui ogni dato ha il relativo controllo;
- 2) misurando gli individui di due gruppi diversi, per cui il confronto può essere solo complessivo.

**La soluzione migliore è quella di utilizzare gli stessi soggetti: si riduce la variabilità e le differenze tra i due gruppi sono imputabili più facilmente ai fattori in esame.**

Usare un soggetto come controllo di se stesso è un metodo preferibile all'appaiamento tra soggetti considerati uguali da colui che programma l'esperimento. Per una perfetta applicazione, quest'ultimo metodo infatti richiederebbe una esatta conoscenza, non sempre possibile e mai completa, dei fattori che determinano la variabilità individuale nelle risposte ai due trattamenti.

**Nessun appaiamento è più preciso di quello ottenuto per identità.**

**Con due campioni dipendenti, è possibile calcolare le differenze tra ogni coppia di dati.** Nella ricerca ambientale ed ecologica, queste differenze non sempre possono essere valutate con una scala d'intervalli o di rapporti; **in varie condizioni è possibile stimare solo il segno o la direzione della differenza.** E' tipico di analisi in realtà complesse poter indicare solo che la situazione è migliorata o peggiorata.

**Per inferenze sulla tendenza centrale di due campioni dipendenti, tra i test non parametrici più diffusi per questo corso ne sono stati riportati quattro:**

**il test di McNemar,**

**il test dei segni,**

**il test T di Wilcoxon,**

**il test di casualizzazione** detto anche test di permutazione.

Il **test di McNemar** può essere utilizzato quando le variabili sono espresse in una scala nominale binaria. I dati raccolti sono conteggi di risposte positive o negative, espresse in due tempi diversi (prima e dopo), riportati in una tabella 2x2.

Il **test dei segni** per due campioni dipendenti come metodologia è totalmente simile a quello per un campione (presentato nel capitolo precedente), poiché utilizza la serie delle differenze tra i due campioni dipendenti. Si ricorre al test dei segni quando per ogni coppia di dati, riferiti allo stesso individuo o caso, è possibile valutare solo il segno della differenza, cioè stabilire se la situazione è migliorata o peggiorata.

Si ricorre al **test T di Wilcoxon** quando si utilizza una misura più precisa della precedente; non solo per ogni coppia di dati è possibile valutare se la situazione è migliorata o peggiorata, ma è possibile stabilire una dimensione o ordine nelle differenze calcolate, cioè mettere in ordine gli individui secondo la quantità di miglioramento o peggioramento avuto.

Si ricorre al **test di casualizzazione** quando le risposte sono misurate in modo preciso, con una scala d'intervalli o di rapporti.

## 7.2. IL TEST DI McNEMAR

Il **test di Q. McNemar** (proposto nel 1947 con l'articolo *Note on the sampling error of the difference between correlated proportion on percentage*, pubblicato da *Psychometrika* 12, pp. 153-157) è utile per verificare se gli individui di un campione sottoposto a due diversi trattamenti forniscono risposte simili o differenti nei due tempi. Per la sua applicazione si deve disporre di **dati appaiati** e gli esiti devono essere misurati su una **scala nominale binaria**.

E' chiamato pure **test per la significatività dei cambiamenti**, poiché l'analisi della significatività utilizza solamente le risposte che hanno cambiato segno, passando da una situazione all'altra.

In biologia, può servire per valutare la condizione di benessere o malessere soggettivo prima e dopo un intervento, nella ricerca ambientale per analizzare i cambiamenti in una serie di situazioni prima e dopo l'azione di risanamento.

Il procedimento del test può essere più facilmente spiegato mediante un esempio.

Si supponga che ad un gruppo di persone, riuniti per un dibattito, sia stato chiesto se sono favorevoli (+) o contrari (-) all'energia nucleare, annotando la risposta di ognuno all'inizio della riunione. Si supponga sempre che, dopo la proiezione di filmati sull'argomento ed una discussione sui pericoli ed i vantaggi dei diversi modi di produrre energia elettrica, agli stessi individui sia stato chiesto di esprimere ancora il loro parere sulla convenienza delle centrali nucleari.

I dati sono riportati come nell'elenco sottostante, che per praticità contiene le risposte solo di 19 individui (indicati con lettere da A ad U), anche se il test di norma richiede campioni di dimensioni nettamente superiori.

Di ogni individuo è riportata la posizione prima e dopo il dibattito.

Individui	Prima	Dopo
A	+	-
B	+	+
C	+	-
D	-	-
E	-	-
F	+	+
G	-	+
H	-	-
I	+	-
L	+	+
M	+	-
N	+	-
O	+	-
P	-	-
Q	+	-
R	+	+
S	-	-
T	-	+
U	+	-

Si tratta di sapere se, alla fine della serata, vi è stato un cambiamento significativo nella convinzione degli intervistati.

**(Fondato su risposte binarie, il test è analogo a quello dei segni, anche se l'elaborazione dei dati è differente).**

1 - I dati devono essere riportati in una tabella più sintetica dell'elenco precedente, classificando le risposte in 4 gruppi, che sono le 4 combinazioni dei 2 segni positivi e dei 2 negativi.

Con i dati della tabella, i 19 individui del campione possono essere suddivisi in:

- A) 4 persone che prima erano favorevoli (+) e dopo si sono dichiarate ancora favorevoli (+),
- B) 8 persone che prima erano favorevoli (+) e dopo si sono dichiarate contrarie (-),
- C) 2 persone che prima erano contrarie (-) e dopo si sono dichiarate favorevoli (+),
- D) 5 persone che prima erano contrarie (-) e dopo si sono dichiarate ancora contrarie.

2 - I risultati sono presentati in una tabella 2x2, impostata come quella sottostante:

		DOPO	
		+	-
PRIMA	+	4	<b>8</b>
	-	<b>2</b>	5

**3 - Il test deve verificare se l'esperimento ha indotto significativi cambiamenti di parere nel campione interrogato nei due momenti differenti. Pertanto, si ignorano le persone che sono rimaste della stessa opinione: non forniscono alcuna informazione sull'effetto del dibattito.**

In modo più specifico, con i dati dell'esempio si prendono in considerazione solo

- le 8 persone che da favorevoli (+) sono divenute contrarie (-),
- le 2 persone che da contrarie (-) sono divenute favorevoli (+).

Come impostazione grafica, ad una lettura non attenta, la tabella può apparire identica a quelle di contingenza 2x2 utilizzate nel test  $\chi^2$ . La procedura d'analisi inferenziale è parzialmente simile e per la significatività del risultato si utilizza la stessa distribuzione dei valori critici; di conseguenza, alcuni confondono il test di McNemar con il test  $\chi^2$ .

**Ma il test  $\chi^2$  si applica a due campioni indipendenti: è utilizzato correttamente quando la tabella riporta la distribuzione di due gruppi diversi ed il calcolo dei valori utilizza tutti e quattro i dati.**

**Il test di McNemar si applica nel caso in cui righe e colonne riportano le risposte o gli esiti degli stessi individui; utilizza solamente i 2 gruppi che hanno cambiato segno ignorando gli altri 2; infine, risponde a un quesito differente.**

4 - **Se è vera l'ipotesi nulla** ( $H_0$ : il trattamento non determina un mutamento significativo), coloro che hanno cambiato la loro risposta dovrebbero aver scelto a caso. Di conseguenza, **il numero di coloro che sono passati dal segno positivo a quello negativo dovrebbe essere equivalente al numero di coloro che hanno cambiato nell'altra direzione**, dal negativo al positivo.

Sulla base di questi concetti, è semplice comprendere la metodologia.

Chiamando A, B, C e D i quattro gruppi

		PRIMA	
		+	-
DOPO	+	A	B
	-	C	D

come nella tabella riportata,

si calcola il numero atteso (E) nell'ipotesi che coloro che dovrebbero passare dal segno positivo (+) a quello negativo (-) debbano essere numericamente uguali a coloro che compiono il tragitto opposto.

$$E = \frac{B + C}{2}$$

**Il numero atteso di cambiamenti nelle due caselle è uguale alla media aritmetica dei due gruppi.**

5 - La significatività del cambiamento è stimata mediante il  $\chi^2$  **con 1 gdl** .

Rispetto alle dimensioni del campione, permangono le stesse condizioni di validità del  $\chi^2$ , riferite ai soli due gruppi utilizzati.

Per **grandi campioni ( $B + C > 100$ )** si può ricorrere alla formula

$$c_{(1)}^2 = \frac{(B - E)^2}{E} + \frac{(C - E)^2}{E}$$

Per **campioni di dimensioni medie ( $30 < B + C < 100$ )** è vantaggioso includere la correzione di Yates per la continuità

$$c_{(1)}^2 = \frac{(|B - E| - 0,5)^2}{E} + \frac{(|C - E| - 0,5)^2}{E}$$

dalla quale è utile derivare la formula abbreviata,

$$c_{(1)}^2 = \frac{(|B - C| - 1)^2}{B + C}$$

valida sia per campioni di dimensioni medie sia di grandi dimensioni, quindi sempre valida, poiché in questo ultimo caso l'effetto della correzione diviene trascurabile.

Per **campioni piccoli ( $B + C < 30$ )** si può ricorrere alla **distribuzione binomiale**, confrontando le frequenze dei due gruppi, che secondo l'ipotesi nulla dovrebbero essere uguali ( $p = q = 1/2$ ;  $B$  e  $C = 30 \times 1/2 = 15$ ).

ESEMPIO. Ad un gruppo di persone residenti in un centro storico, con un referendum nominativo è stato chiesto se erano favorevoli o contrari alla instaurazione dell'isola pedonale, con forti limitazioni al traffico di autoveicoli: 119 si sono dichiarati favorevoli e 100 contrari.

A distanza di alcuni mesi, su sollecitazione dell'amministrazione interessata al consenso dei cittadini, agli stessi individui è stata posta nuovamente la stessa domanda: 158 si sono dichiarati favorevoli e 61 contrari.

Una verifica individuale e nominativa dei voti prima e dopo l'evento fornisce la distribuzione riportata nella tabella seguente, dove con + si indicano i favorevoli e con - i contrari

		+	DOPO	-	Totale
PRIMA	+	84		35	119
	-	74		26	100
Totale		158		61	219

Si è avuto un mutamento significativo nell'opinione dei residenti interrogati?

Risposta.

Fra tutti i dati riportati nella tabella, per il test sono necessarie solo le due frequenze di coloro che hanno modificato la loro posizione, come riportato in grassetto nella tabella sottostante:

		DOPO	
		+	-
PRIMA	+	A    84	<b>B    35</b>
	-	<b>C    74</b>	D    26

Ricordando che, con la simbologia presentata,

$$B = 35; \quad C = 74; \quad E = \frac{35+74}{2} = 54,5;$$

si può ricorrere indifferentemente a una delle due formule presentate, tra loro matematicamente equivalenti:

1 -

$$c_{(1)}^2 = \frac{(|B-E|-0,5)^2}{E} + \frac{(|C-E|-0,5)^2}{E}$$

che con i dati diviene

$$c_{(1)}^2 = \frac{(|35-54,5|-0,5)^2}{54,5} + \frac{(|74-54,5|-0,5)^2}{54,5} = 13,25$$

e fornisce un valore di  $\chi^2$ , con un grado di libertà, uguale a 13,25.

2 -

$$c_{(1)}^2 = \frac{(|B-C|-1)^2}{B+C}$$

che con i dati diviene

$$c_{(1)}^2 = \frac{(|35-74|-1)^2}{35+74} = 13,25$$

e da un risultato identico al precedente.

Il valore calcolato (13,25) supera ampiamente il valore critico riportato nella tabella del  $\chi^2$  (3,84) per la probabilità  $\alpha = 0.05$  con 1 gdl.

Si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa ( $H_1$ ): il cambiamento è stato significativo.

Il numero di coloro che hanno cambiato opinione da favorevoli a contrari (35) è significativamente minore del numero di coloro che hanno cambiato nel senso opposto, da contrari a favorevoli (74) in merito al blocco del traffico nel centro storico.

### 7.3. IL TEST DEI SEGNI

**Quando non è possibile valutare la differenza esistente tra coppie di dati con una misura quantitativa, ma solo stabilire la direzione della differenza, si può utilizzare il test dei segni (*sign test*), per verificare se la tendenza centrale di un gruppo è uguale oppure significativamente diversa da quella della situazione precedente.**

Come nel test t per 2 campioni dipendenti, di cui è analogo, l'ipotesi è impostata sul valore della differenza: se  $\delta$  sia uguale a 0 oppure a un qualunque valore prefissato, con ipotesi alternative che possono essere bilaterali o unilaterali.

L'unico postulato (o condizione di validità) richiesto è che **il fenomeno** analizzato abbia una **distribuzione continua**, per cui è quasi sempre possibile stabilire il segno della differenza tra le misure appaiate dei due campioni dipendenti. Come nel test dei segni per un campione, le differenze uguali a zero non danno informazioni; quindi sono ignorate nell'elaborazione dei dati e dal conteggio delle dimensioni del campione.

Se è vera l'ipotesi nulla  $H_0$

$$H_0: \delta = 0$$

(il trattamento non determina effetti sistematici e le differenze tra le coppie di risposte pertanto sono casuali),

il numero di miglioramenti (+) dovrebbe essere uguale al numero di peggioramenti (-).

Il numero di segni + e quello di segni - (da cui il nome "test dei segni") dovrebbero equivalersi.

Indicando con  $X_1$  il segno positivo e con  $X_2$  il segno negativo, tale concetto è espresso dalla relazione

$$P(X_1 > X_2) = P(X_1 < X_2) = \frac{1}{2}$$

La procedura più diffusa richiede di contare sia i segni positivi che negativi e di utilizzare il numero minore.

Il problema di verificare la significatività della differenza assume quindi la seguente forma statistica:

- se è vera  $H_0$ , la frequenza del segno più raro tende al valore medio ( $N/2$ );

- se è vera  $H_1$ , la frequenza del segno più raro tende a 0.

Come nel test  $t$  per dati appaiati, **il test dei segni può essere sia unilaterale che bilaterale**, in dipendenza dell'ipotesi  $H_1$  sulla maggioranza di segni positivi o negativi oppure su una semplice loro differenza numerica.

Nel caso di **piccoli campioni**, la distribuzione delle probabilità è determinata mediante **la distribuzione binomiale**

$$P = C_N^x p^x q^{N-x}$$

dove:

$$p = q = \frac{1}{2};$$

$N$  = numero di coppie di dati o di segni;

$x$  = frequenza del segno più raro.

La distribuzione binomiale è una distribuzione discreta, che in questo caso fornisce la probabilità esatta di ottenere un dato numero di segni più e di segni meno, nell'ipotesi che essi abbiano la stessa probabilità (1/2) di comparire. Per l'inferenza, occorre sommare la probabilità di ottenere la risposta osservata con quelle delle eventuali risposte più estreme (il segno più raro diminuisce progressivamente di una unità dal valore osservato fino a 0).

Per **grandi campioni**, in pratica per un numero di osservazioni ( $N$ ) superiore a 20 o a 25 (il numero massimo riportato nelle tabelle per piccoli campioni nei vari testi non è sempre uguale) si ricorre alla distribuzione normale

$$Z = \frac{x - \mu_x}{\sigma_x}$$

dove

$x$  è il numero di segni con frequenza minore,

$\mu_x$  è la media attesa ed è uguale a

$$m_x = Np = \frac{N}{2}$$

$\sigma_x$  è la deviazione standard, calcolata come

$$s_x = \sqrt{N \cdot p \cdot q} = \frac{1}{2} \sqrt{N}$$

Trattandosi di conteggi e quindi di una variabile discreta, si introduce una correzione per la continuità: si aumenta di 0,5 il valore osservato (quindi si diminuisce di 0,5 la differenza tra il numero minore osservato e la media attesa)

$$z = \frac{(x + 0,5) - m_x}{s_x}$$

La **formula di calcolo abbreviata**, di uso abituale per semplificare i calcoli, diventa

$$z = \frac{2x + 1 - N}{\sqrt{N}}$$

ricordando che

$N$  è il numero di dati utili,

$x$  è la frequenza del segno più raro.

ESEMPIO 1. Dopo un anno di interventi per il risanamento ambientale di 12 laghi, si sono misurati alcuni parametri chimici e biologici che hanno permesso una valutazione complessiva: in 9 casi la situazione è migliorata, in 2 risulta peggiorata e in 1 caso non sembra mutata.

E' possibile sostenere che la situazione generale è migliorata in modo statisticamente significativo?

Risposta.

Si tratta di un test ad una coda ( $H_0: \delta = 0$ ;  $H_1: \delta < 0$ ), ricordando che l'**ipotesi è sulla mediana delle differenze**.

Escludendo il caso in cui non si è avuto un cambiamento (in quanto privo d'informazione), si tratta di sapere se **9+** e **2-** possono essere ritenuti una variazione casuale di **5,5+** e altrettanti -, oppure se la differenza deve essere ritenuta non casuale.

Per  $N$  uguale a **11**

con  $p$  e  $q$  uguali a **1/2**,

la probabilità di avere 2 volte il segno - ( $P_{(2)}$ ), deve essere cumulata con

la probabilità di averlo **1** volta ( $P_{(1)}$ ) sola e

quella di averlo **0** volte ( $P_{(0)}$ )

$$P_{(2)} = C_{11}^2 \cdot 0,5^2 \cdot 0,5^9$$

$$P_{(1)} = C_{11}^1 \cdot 0,5^1 \cdot 0,5^{10}$$

$$P_{(0)} = C_{11}^0 \cdot 0,5^0 \cdot 0,5^{11}$$

La somma delle 3 probabilità ( $P_{(2)} + P_{(1)} + P_{(0)}$ ) risulta uguale a 0.033.

**La tabella delle probabilità cumulate della distribuzione binomiale con  $p = q = \frac{1}{2}$  (già riportata nel caso di un campione) fornisce la risposta per  $N = 11$  e  $r = 2$ , evitando i calcoli.**

La probabilità complessiva di ottenere la risposta osservata e quelle più estreme nella stessa direzione risulta bassa, inferiore a 0.05.

Si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa: la situazione è migliorata in modo statisticamente significativo.

ESEMPIO 2. Per una ricerca sull'equilibrio tra due specie è stato calcolato il rapporto numerico tra la specie **A** e la specie **B** in 10 località. Dopo un anno, è stato ripetuto il campionamento. La tabella seguente riporta il rapporto tra le due specie nel primo e nel secondo campionamento, effettuati nelle stesse 10 località.

Località	A	B	C	D	E	F	G	H	I	L
Primo Campionamento	1,1	1,7	1,3	1,3	0,8	1,2	1,0	1,3	0,9	1,2
Secondo Campionamento	0,9	1,2	1,3	0,9	1,3	0,7	0,8	0,5	0,5	0,9

Si può sostenere che il rapporto tra le due specie sia variato?

Risposta.

**Tassi, percentuali e rapporti** rendono il valore indipendente dalle dimensioni del campione ed **agevolano il confronto tra i risultati ottenuti su campioni che possono avere dimensioni anche molto differenti**. Ma per i test parametrici questa caratteristica rappresenta un limite: **quando sono ottenuti su campioni di numerosità molto differente, tassi, percentuali e rapporti hanno un'attendibilità ed un campo di variazione sensibilmente diversi**.

In termini più semplici, per ricorrere ad un esempio,  $\frac{3}{4}$  e  $\frac{15}{20}$  pure fornendo lo stesso valore quando sono trasformati in tassi, percentuali o rapporti non possono essere direttamente confrontati, poiché la loro varianza è significativamente differente (vedi varianza delle percentuali).

Il confronto tra le due serie di valori fornisce quindi per ogni località un'informazione limitata al segno dell'incremento, per cui è appropriato l'uso del test dei segni.

Con i dati dell'esempio,

1 - il primo passo è il calcolo delle differenze, tradotte in segni

Località	A	B	C	D	E	F	G	H	I	L
Primo	1,1	1,7	1,3	1,3	0,8	1,2	1,0	1,3	0,9	1,2
Secondo	0,9	1,2	1,3	0,9	1,3	0,7	0,8	0,5	0,5	0,9
<b>Differenza</b>	+	+	=	+	-	+	+	+	+	+

I dieci campioni determinano 8 differenze positive, 1 negativa e una differenza nulla.

2 - L'ipotesi da verificare, riferita alla mediana, è bilaterale:

$$H_0: \delta = 0 ; \quad H_1: \delta \neq 0$$

Si trascura la differenza nulla, poiché ininfluente sulla possibile variazione complessiva; si deve stimare la probabilità totale di trovare 1 valore negativo  $P_{(1)}$  e 0 valori negativi  $P_{(0)}$  (che rappresenta l'unica risposta più estrema) su 9 dati.

3 - Il campione è di piccole dimensioni e quindi si deve ricorrere alla distribuzione binomiale:

$$P_{(1)} = C_9^1 \cdot 0,5^1 \cdot 0,5^8 = 9 \cdot 0,001953 = 0,017578$$

$$P_{(0)} = C_9^0 \cdot 0,5^0 \cdot 0,5^9 = 1 \cdot 0,001953 = 0,001953$$

con

$$P_{(1)} + P_{(0)} = 0,017578 + 0,001953 = 0,01953$$

che fornisce una probabilità complessiva in una coda della distribuzione ( $P_{(1)} + P_{(0)}$ ) uguale a 0.01953. Trattandosi di un test bilaterale, la probabilità calcolata deve essere raddoppiata e risulta uguale a 0.039062.

E' una probabilità inferiore a 0.05. Si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa.

Le due specie hanno modificato i loro rapporti numerici in modo significativo.

4 - La tabella delle probabilità cumulate, che evita i due calcoli precedenti, per  $N = 9$  e  $r = 1$  fornisce una probabilità uguale a 0,020 in una coda della distribuzione, che rappresenta un arrotondamento del 0,01953 (molto più rapido da ottenere).

ESEMPIO 3. Su 80 coppie di osservazioni, rilevate prima e dopo il trattamento, 5 non hanno mostrato differenze, 44 evidenziano un miglioramento e 31 un peggioramento.

Si può affermare che la situazione è migliorata in modo significativo?

Risposta.

E' un grande campione, per cui si ricorre alla distribuzione normale.

Si tratta di un test ad una coda, con

$x = 31$  (il segno meno frequente) e

$N = 75$  (44+31) da cui :

$$z = \frac{((2 \cdot 31) + 1) - 75}{\sqrt{75}} = -1,39$$

Ad un valore di  $z$  uguale a 1,39 in un test a una coda, nella tavola della distribuzione normale corrisponde una probabilità pari a 0.0823 o 8,23%. E' una probabilità superiore al valore soglia di 0.05 per cui non si è nelle condizioni di rifiutare l'ipotesi nulla.

#### **7.4. IL TEST DEI SEGNI PER RANGHI: TEST T DI WILCOXON**

Il test dei segni utilizza solo una parte ridotta delle informazioni contenute nei valori numerici; per esempio, non considera se i segni meno numerosi sono riferiti alle differenze maggiori o a quelle minori. Inoltre se una quota nettamente maggiore di segni è positiva oppure negativa, è logico considerare la distribuzione come asimmetrica intorno alla zero. E' quindi ragionevole considerare anche l'informazione data dalle dimensioni del valore assoluto delle differenze, senza utilizzarle compiutamente come nella statistica parametrica.

**Il test T di F. Wilcoxon si giova anche dell'informazione relativa al rango ed attribuisce un peso maggiore alle differenze più grandi.**

Principi e metodi sono identici a quelli del **test di Wilcoxon della somma dei ranghi con segno**, già applicati nel caso di un campione e derivati dall'articolo del 1945 *Individual comparisons by ranking methods*, pubblicato su *Biometrics Bulletin* 1, pp. 80-83.

Il test compare ormai nelle librerie informatiche a maggiore diffusione, per l'uso ampio che ne è fatto nella statistica applicata e la attendibilità dei risultati. La presentazione della metodologia e le tavole dei valori critici per piccoli campioni sono riportati anche dal testo di statistica non parametrica del 1975 di **E. L. Lehmann** dal titolo *Nonparametrics: Statistical Methods Based on Ranks* (San Francisco, Holden Day).

Anche in questo caso, il test è più facilmente spiegato ricorrendo ad un'applicazione.

Si supponga di voler verificare se esistono differenze significative nei livelli mediani d'inquinamento tra due giorni (chiamati convenzionalmente X e Y) con caratteristiche meteorologiche nettamente differenti, eliminando la variabilità presente tra ore. A questo scopo, dalle ore 6 alle ore 20 ogni 60 minuti nella stessa zona di una città è stato rilevato il tasso di inquinamento atmosferico.

I dati sono stati riportati nella tabella successiva (prime tre colonne):

Ora	Giorno X	Giorno Y	<b>diff.(X-Y)</b>	<b>Rango</b>
6	120	140	<b>- 20</b>	<b>- 3,5</b>
7	145	160	<b>- 15</b>	<b>- 2</b>
8	305	295	<b>+ 10</b>	<b>+1</b>
9	200	230	<b>- 30</b>	<b>- 7</b>
10	160	200	<b>- 40</b>	<b>-10</b>
11	135	185	<b>- 50</b>	<b>-12</b>
12	170	150	<b>+20</b>	<b>+3,5</b>
13	285	515	<b>-230</b>	<b>-14</b>
14	290	220	<b>+ 70</b>	<b>+13</b>
15	200	225	<b>- 25</b>	<b>-5</b>
16	150	180	<b>- 30</b>	<b>-7</b>
17	160	190	<b>- 30</b>	<b>-7</b>
18	115	115	<b>0</b>	<b>----</b>
19	105	140	<b>- 35</b>	<b>- 9</b>
20	105	150	<b>- 45</b>	<b>-11</b>

Per rispondere al problema, dopo aver definito che si tratta di un test bilaterale con ipotesi nulla

$$H_0 : \delta \text{ mediana} = 0$$

e ipotesi alternativa

$$H_1 : \delta \text{ mediana} \neq 0$$

la metodologia richiede vari passaggi.

1 - **Calcolare la colonna delle differenze con il loro segno:**  $d = X - Y$  (riportate in grassetto nella quarta colonna della tabella precedente, in cui sono presentati sia i dati originari sia le loro prime trasformazioni).

2 - **Eliminare dall'analisi le differenze nulle**; la numerosità del campione sarà proporzionalmente ridotta (nell'esempio si elimina l'osservazione delle ore 18: da 15, N diventa 14, come si osserva nella quinta colonna).

3 - **Trasformare le differenze, considerate in valore assoluto, nel loro rango**. Nel caso di due o più dati uguali, assegnare lo stesso valore, calcolato come media dei ranghi.

Nell'esempio sono presenti due differenze (X-Y) uguali in valore assoluto (+20 e -20), che occupano la 3<sup>a</sup> e la 4<sup>a</sup> posizione: per entrambe il rango è 3,5 dato da  $(3 + 4) / 2$ .

Vi sono altre tre differenze tra loro uguali (-30), che occupano la posizione 6<sup>a</sup>, 7<sup>a</sup> e 8<sup>a</sup>: ad ognuna è assegnato rango 7, dato da  $(6 + 7 + 8) / 3$ .

Anche se ampiamente dibattuto in vari testi di statistica per le sue implicazioni teoriche insieme con la riduzione del campione quando sono presenti differenze uguali a zero, l'attribuzione della media dei ranghi alle differenze identiche in valore assoluto (*ties*) ha effetti trascurabili sul successivo valore di T, se non ripetuto troppe volte.

4 - **Attribuire ad ogni rango il segno della differenza corrispondente** (si ottengono i valori in grassetto della quinta colonna, sempre della tabella precedente).

5 - **Sommare i ranghi con lo stesso segno**.

La somma dei ranghi positivi è

$$1 + 3,5 + 13 = \mathbf{17,5}$$

mentre la somma dei ranghi negativi è

$$-3,5 - 2 - 7 - 10 - 12 - 14 - 5 - 7 - 7 - 9 - 11 = 87,5$$

6 - **Scegliere il totale minore: è il valore di T**.

Nell'esempio, T è dato dai ranghi con segno positivo:

$$T = 1 + 3,5 + 13 = 17,5.$$

Per non commettere errori nei calcoli è utile ricordare, che la somma di N ranghi è uguale a

$$N(N+1)/2$$

Con i 14 dati dell'esempio (poiché una differenza è stata eliminata in quanto uguale a 0)

la somma dei valori positivi (17,5) e di quelli negativi (78,5)

$$17,5 + 78,5 = 105$$

deve essere uguale a

$$(14 \times 15) / 2 = 105$$

7 - Secondo l'ipotesi nulla  $H_0$ , la differenza tra le due serie di osservazioni appaiate dovrebbe essere uguale a zero. Di conseguenza, nella colonna delle differenze la somma dei ranghi con segno positivo e la somma dei ranghi con segno negativo dovrebbero essere uguali. Perciò il totale minore dovrebbe tendere ad un valore medio atteso  $\mu_T$  determinato da  $N$ , il numero di differenze o ranghi, secondo la relazione

$$m_T = \frac{N \cdot (N + 1)}{4}$$

Con i dati dell'esempio,

$$\mu_T = \frac{14 \cdot 15}{4} = 52,5$$

il valore atteso della somma minore tra segni positivi e negativi dovrebbe tendere a 52,5.

**8 - La significatività della differenza tra le due serie di dati appaiati è tradotta nella significatività della differenza tra  $T$  e  $m_T$ ; nell'esempio tra 17,5 e 52,5.**

Per **piccoli campioni (N ≤ 25)**, la significatività è fornita da tavole che riportano i valori critici di  $T$  in rapporto alla numerosità del campione.

A questo scopo è possibile utilizzare la tavola già presentata per un campione.

Altri testi riportano i valori critici in modo leggermente diverso, come riportato nella pagina successiva che fornisce i valori massimi accettabili di  $T$  ai vari livelli di significatività, secondo la dimensione  $N$  del campione delle differenze.

**La mediana della differenza è significativa alla probabilità  $\alpha$  prescelta, quando il valore di  $T$  è uguale o inferiore a quello riportato nella tabella.**

Per  $N = 14$  e  $\alpha = 0.05$  in un test bidirezionale, il valore critico è **21**.

Il valore di  $T$  (**17,5**) calcolato con i dati dell'esempio è minore: si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa della significatività della differenza.

Tra i due giorni esiste una differenza significativa nel valore mediano del tasso d'inquinamento.

Quando si programma un esperimento in cui utilizzare il test  $T$  di Wilcoxon è importante ricordare che **per rifiutare l'ipotesi nulla alla probabilità  $\alpha = 0.05$ ,**

- **in un test ad una coda, occorrono almeno 5 dati appaiati, la cui differenza sia sempre diversa da zero; si è in grado di rifiutare l'ipotesi nulla se le 5 differenze sono tutte dello stesso segno;**
- **in un test a due code, servono almeno 6 coppie di dati, con differenze che siano tutte dello stesso segno.**

## Valori critici del test T di Wilcoxon per 2 campioni dipendenti

	<b>Test a una coda</b>			
<b>P</b>	<b>0.05</b>	<b>0.025</b>	<b>0.01</b>	<b>0.001</b>
	<b>Test a due code</b>			
<b>P</b>	<b>0.1</b>	<b>0.05</b>	<b>0.02</b>	<b>0.002</b>
<b>N</b>				
<b>5</b>	0	--	--	--
<b>6</b>	2	0	--	--
<b>7</b>	3	2	0	--
<b>8</b>	5	3	1	--
<b>9</b>	8	5	3	--
<b>10</b>	10	8	5	0
<b>11</b>	13	10	7	1
<b>12</b>	17	13	9	2
<b>13</b>	21	17	12	4
<b>14</b>	25	21	15	6
<b>15</b>	30	25	19	8
<b>16</b>	35	29	23	11
<b>17</b>	41	34	27	14
<b>18</b>	47	40	32	18
<b>19</b>	53	46	37	21
<b>20</b>	60	52	43	26
<b>21</b>	67	58	49	30
<b>22</b>	75	65	55	35
<b>23</b>	83	73	62	40
<b>24</b>	91	81	69	45
<b>25</b>	100	89	76	51

Per **test a una coda**, per ottenere che la mediana delle differenze sia significativa

- alla probabilità  $\alpha = 0.01$  servono almeno 7 coppie di dati,
- alla probabilità  $\alpha = 0.001$  almeno 10, tutte dello stesso segno.

Per **grandi campioni** ( $N > 25$ ), il valore della somma dei ranghi (**T**) è distribuito in modo approssimativamente normale: la significatività della differenza può essere saggiata con la distribuzione **z**,

usando la relazione

$$z = \frac{T - \mu_T}{\sigma_T}$$

dove la media attesa  $\mu_T$  è data da

$$\mu_T = \frac{N \cdot (N + 1)}{4}$$

e la deviazione standard  $\sigma_T$  è data da

$$\sigma_T = \sqrt{\frac{N \cdot (N + 1) \cdot (2N + 1)}{24}}$$

**L'applicazione del test T di Wilcoxon richiede che i campioni siano estratti da popolazioni con una distribuzione simmetrica, ma non necessariamente normale.** Per il calcolo delle differenze tra dati appaiati, **le osservazioni devono essere misurate in un scala d'intervalli.**

La distribuzione normale, utilizzata per grandi campioni, fornisce **una buona approssimazione già quando N è uguale 14-15 coppie di dati o differenze tra essi.**

Nella tabella dei valori critici per piccoli campioni, per un test ad una coda è possibile osservare che, con  $N = 14$ , un valore di  $T = 25$  risulta significativo alla probabilità  $\alpha = 0.05$ .

Con la distribuzione normale, dove

$$\mu_T = \frac{14 \cdot 15}{4} = 52,5$$

e

$$\sigma_T = \sqrt{\frac{14 \cdot 15 \cdot 29}{24}} = 15,93$$

si ottiene un valore di z

$$z = \frac{25 - 52,5}{15,93} = -1,726$$

uguale a -1,726.

In una coda della distribuzione ad esso corrisponde una probabilità  $\alpha$  uguale a **0.0420**, vicino al valore fornito dalla tabella. Si conferma in modo empirico come la normalità della distribuzione dei ranghi sia approssimativamente raggiunta anche con meno di 20 dati.

### Ties

Quando i valori delle differenze sono distribuiti in un intervallo ristretto, si possono avere **alcuni punteggi uguali o valutazioni ex-aequo (ties)**, in particolare **in grandi campioni**.

La media attesa  $m_T$  non subisce variazioni; ma la varianza  $\sigma_T^2$  diviene più ridotta, rispetto ai casi in cui ogni rango è attribuito con precisione, come possibile quando la scala è continua.

A causa di queste **valutazioni ex-aequo (ties)** si impone una **correzione di  $S_T^2$** , che nel caso del test T di Wilcoxon consiste in una riduzione della varianza  $\sigma_T^2$  pari a

$$- \sum_{i=1}^g t_i \cdot (t_i - 1) \cdot (t_i + 1)$$

dove

$g$  è il numero di gruppi di ranghi identici,

$t$  è il numero di dati con lo stesso rango entro ogni gruppo.

ESEMPIO 1. Calcolare  $S_T^2$  con la correzione per **ties**, dalla seguente distribuzione di ranghi con segno

**1 2 4 4 -4 6 -8,5 8,5 8,5 -8,5 11 12 13,5 13,5 -15 16 17 19 19 19 -21 -22**

Senza correzione, il valore

$$\sigma_T^2 = \frac{N \cdot (N + 1) \cdot (2N + 1)}{24}$$

di  $\sigma_T^2$  con  $N = 22$

$$\sigma_T^2 = \frac{22 \cdot 23 \cdot 45}{24} = 948,75$$

risulta uguale a 948,75.

Ma sono presenti 4 gruppi (i ranghi 4 8,5 13,5 19 riportati in grassetto) con valori uguali

**1 2 **4 4** -4 6 -8,5 **8,5 8,5** -8,5 11 12 **13,5 13,5** -15 16 17 **19 19 19** -21 -22**

che possono essere riassunti in:

1 gruppo con 2 valori uguali (il rango 13,5);

2 gruppi con 3 valori uguali (i ranghi 4 e 19);

1 gruppo con 4 valori uguali (il rango 8,5).

Per una stima corretta, la varianza deve essere ridotta di una quantità

$$-\sum_{i=1}^g t_i \cdot (t_i - 1) \cdot (t_i + 1)$$

dove

$t$  è il numero di valori identici (2, 3, 3, 4),

$g$  è il numero di gruppi con valori identici.

Con i dati dell'esempio

$$(2 \cdot 1 \cdot 3) + (3 \cdot 2 \cdot 4) + (3 \cdot 2 \cdot 4) + (4 \cdot 3 \cdot 5) = 6 + 24 + 24 + 60 = 114$$

si ottiene un valore pari a 114;

di conseguenza, la varianza corretta è

$$s^2_T = 948,75 - 114 = \mathbf{834,75}$$

e la deviazione standard  $s_T$  corretta

$$\sigma_T = \sqrt{948,75 - 114} = \sqrt{834,75} = 28,89$$

diventa uguale a 28,89 mentre,

senza correzione, sarebbe stata

$$s_T = \sqrt{948,75} = 30,80$$

uguale a 30,80.

Con la diminuzione del valore di  $s_T$ , aumenta proporzionalmente il valore di  $z$ , che pertanto risulterà più facilmente significativo.

E' utile ricordare che se una differenza mediana risulta significativa senza la correzione per i valori identici, risulterà ancor più significativa con la correzione; pertanto, quando il valore del  $T$  risulta significativo, la correzione diviene superflua.

**La correzione della varianza per i ties determina una diminuzione della varianza**, che può essere importante quando la differenza tra 2 campioni dipendenti non risulta significativa, con una probabilità prossima al valore critico.

**ESEMPIO 2.** I composti poco solubili in acqua e persistenti come alcuni pesticidi clorurati (il DDT e i PCB) si accumulano negli organismi acquatici e per loro tramite si trasferiscono lungo le catene alimentari. Anche nelle catene alimentari terrestri gli inquinanti si concentrano ai livelli trofici più elevati.

E' stato catturato un gruppo di animali ed in essi è stata misurata la concentrazione di una sostanza tossica nel sangue. Dopo essere stati marcati sono stati liberati in un'area ritenuta inquinata.

A distanza di un mese, 13 sono stati ricatturati e per ognuno di essi è stata valutata nuovamente la presenza di sostanza tossica, rapportata ad un peso unitario ( $\mu\text{g}/\text{Kg}$ ).

Individui	Concentrazione alla 1 <sup>a</sup> misuraz.	Concentrazione alla 2 <sup>a</sup> misuraz.	Differenza	Rango con segno
A	13,2	18,5	+ 5,3	+ 12
B	12,4	15,2	+ 2,8	+ 8
C	13,7	14,6	+ 0,9	+ 3
D	12,1	13,1	+ 1,0	+ 4
E	10,8	14,2	+ 4,4	+ 11
F	12,1	12,1	0	-----
G	13,7	13,2	- 0,5	- 1
H	9,4	12,9	+ 3,5	+ 9
I	12,1	10,6	- 1,5	- 6
L	16,1	15,3	- 0,8	- 2
M	11,4	15,5	+ 4,1	+ 10
N	9,8	12,2	+ 2,4	+ 7
O	11,5	10,3	- 1,2	- 5

Si è realizzato un aumento significativo?

Risposta.

E' un test ad una coda:

$$H_0 : \text{mediana } 1^a = \text{mediana } 2^a, \quad H_1 : \text{mediana } 1^a < \text{mediana } 2^a;$$

Seguendo i dati nella tabella riportata, la metodologia richiede alcuni passaggi:

- 1 - si calcolano le 13 differenze con il segno, tra tutte le coppie di misure;
- 2 - ad esse si attribuisce il rango relativo, considerando la differenza in valore assoluto ed ignorando la differenza nulla (dell'individuo F);
- 3 - ai 12 ranghi rimasti si attribuisce il segno della differenza;
- 4 - i valori negativi determinano la somma minore, con  $T = 14$ ;
- 5 - nella tabella per un test ad una coda alla probabilità  $\alpha = 0.05$  per  $N = 12$ , il valore critico di  $T$  riportato è 17;
- 6 - Il valore di  $T$  calcolato (14) è inferiore a quello tabulato (17) alla probabilità  $\alpha$  prefissata (0.05).

Si rifiuta l'ipotesi nulla e quindi si accetta l'ipotesi alternativa: si è realizzato un aumento significativo nella concentrazione della sostanza tossica.

**Rispetto al test t di Student per due campioni dipendenti, la potenza di questo test non parametrico ha valori identici a quella per un campione.**

### 7.5. TEST DI CASUALIZZAZIONE PER 2 CAMPIONI DIPENDENTI

Tra tutti i test non parametrici utili al confronto sulla tendenza centrale della differenza tra due campioni dipendenti, il **test di casualizzazione (randomization test)** risulta il più utile per piccoli campioni, sia dal punto di vista pratico che didattico. Come quello dei segni, non ha un autore definito, essendo impostato sul calcolo combinatorio, una delle metodologie classiche della statistica non parametrica.

E' **chiamato anche test di permutazione (*permutation test*)**, anche se in realtà si fonda su disposizioni con replicazione, dal termine inglese che raggruppa sia permutazioni che disposizioni. E' riportato in alcuni testi di statistica non parametrica, tra cui quello di Sidney Siegel (*Statistica non parametrica per le Scienze del comportamento*), sia nella versione più recente che in quelle più antiche.

**Snedecor George W.** e **Cochran William G.** nel loro volume *Statistical Methods* (Sixth Edition, The Iowa University Press, Ames, Iowa, U.S.A., Seventh printing, 1974, cap. 5 a pag. 133) lo chiamano ***Fisher's randomization test***, attribuendolo appunto a **R. A. Fisher** in quanto descritto nel suo volume *The Design of Experiments* del 1960 (7<sup>th</sup> edition., p.44 Oliver and Boyd, Edinburgh).

Il test di casualizzazione può essere applicato solo con **scale di rapporti o d'intervallo; in queste condizioni** è più potente del test dei segni e di quello dei ranghi con segni di Wilcoxon; in molti casi è preferibile pure al test **t** di cui è analogo, in particolare in campioni con 5-10 coppie di dati. La sua potenza efficienza è del 100%. **Non richiede nessuna condizione sulla forma della distribuzione dei dati, né la normalità della distribuzione né l'omoschedasticità** (omogeneità della varianza).

A differenza del test T di Wilcoxon, **fornisce direttamente la probabilità complessiva, senza il ricorso a tavole dei valori critici.**

I limiti alla sua utilizzazione derivano solamente dalle difficoltà pratiche del calcolo manuale, quando le dimensioni del campione superano le 12-14 osservazioni.

**A differenza di quasi tutti gli altri test non parametrici, l'ipotesi è sulla media, non sulla mediana.**

Il metodo può essere spiegato con semplicità mediante una dimostrazione.

Come già nel test T di Wilcoxon, dal confronto tra due campioni dipendenti si supponga di aver ottenuto le 10 differenze seguenti

+ 10	+25	+7	+8	+2	+71	-5	+4	+15	-3
------	-----	----	----	----	-----	----	----	-----	----

Si intende verificare se la loro tendenza centrale è significativamente diversa da 0.

E' un test bidirezionale, con ipotesi nulla

$$H_0: d = 0$$

ed ipotesi alternativa

$$H_1: d \neq 0$$

I fondamenti logici e i passaggi metodologici del test possono essere riassunti in 6 punti:

1 - **Mantenere sempre fissi i valori assoluti delle differenze:** esse sono il risultato oggettivo dell'esperimento e misurano la variazione trovata tra le coppie di dati.

**Può cambiare solo il loro segno:** se l'ipotesi  $H_0$  è vera, **il segno di ogni differenza** avrebbe potuto essere indifferentemente + oppure -.

2 - **Calcolare il numero di possibili risposte,** nell'ipotesi che ogni differenza calcolata possa essere sia positiva che negativa. Con N dati, il numero di possibili risposte binarie è  $2^N$ .

Con i 10 valori dell'esempio, **le possibili risposte** sono  $2^{10} = 1024$ .

Esse **devono anche essere ugualmente probabili.**

3 - **Definire la zona di rifiuto,** secondo la direzionalità del test (a una coda o a due code) ed il livello di significatività  $\alpha$  prefissato.

Con 1024 possibili risposte, per una significatività  $\alpha = 0.05$  la zona di rifiuto esclude i 51 risultati ( $1024 \times 0,05 = 51,2$ ) possibili più estremi. Trattandosi di un test a due code, l'area di rifiuto è formata dalle 25 risposte più estreme in una coda e dalle altre 25 nell'altra coda (**nella zona di rifiuto, l'arrotondamento deve sempre essere per difetto**, al fine di non rifiutare l'ipotesi nulla con una probabilità  $\alpha$  anche solo leggermente superiore a quella prefissata).

4 - **Individuare i risultati più estremi,** collocati nella zona di rifiuto.

A tale scopo ordinare i valori ottenuti sulla base della loro somma, al variare progressivo dei segni. La risposta più estrema in una direzione è quella in cui tutti i valori hanno segno positivo; la più estrema nell'altra direzione è quella in cui tutti i valori hanno segno negativo.

Con i dati dell'esempio, le due risposte più estreme sono:

+10	+25	+7	+8	+2	+71	+5	+4	+15	+3	=	+150
-10	-25	-7	-8	-2	-71	-5	-4	-15	-3	=	-150

5 - Poiché la somma dei dati riportati nell'esempio è positiva, si può limitare la verifica ai soli 25 risultati positivi più estremi. Nell'elenco sottostante, per brevità sono riportate solo le 17 risposte più estreme, nella direzione dell'ipotesi  $\delta > 0$ .

Serie delle prime 17 risposte più estreme in una direzione											Rango		
1)	+10	+25	+7	+8	+2	+71	+5	+4	+15	+3	=	+150	1
2)	+10	+25	+7	+8	-2	+71	+5	+4	+15	+3	=	+146	2
3)	+10	+25	+7	+8	+2	+71	+5	+4	+15	-3	=	+144	3
4)	+10	+25	+7	+8	+2	+71	+5	-4	+15	+3	=	+142	4
5)	+10	+25	+7	+8	+2	+71	-5	+4	+15	+3	=	+140	5,5
6)	+10	+25	+7	+8	-2	+71	+5	+4	+15	-3	=	+140	5,5
7)	+10	+25	+7	+8	-2	+71	+5	-4	+15	+3	=	+138	7
8)	+10	+25	-7	+8	+2	+71	+5	+4	+15	+3	=	+136	9
9)	+10	+25	+7	+8	+2	+71	+5	-4	+15	-3	=	+136	9
10)	+10	+25	+7	+8	-2	+71	-5	+4	+15	+3	=	+136	9
11)	+10	+25	+7	-8	+2	+71	+5	+4	+15	+3	=	+134	11,5
<b>12)</b>	<b>+10</b>	<b>+25</b>	<b>+7</b>	<b>+8</b>	<b>+2</b>	<b>+71</b>	<b>-5</b>	<b>+4</b>	<b>+15</b>	<b>-3</b>	=	<b>+134</b>	<b>11,5</b>
13)	+10	+25	+7	+8	+2	+71	-5	-4	+15	+3	=	+132	13,5
14)	+10	+25	-7	+8	-2	+71	+5	+4	+15	+3	=	+132	13,5
15)	-10	+25	+7	+8	+2	+71	+5	+4	+15	+3	=	+130	16
16)	+10	+25	-7	+8	+2	+71	+5	+4	+15	-3	=	+130	16
17)	+10	+25	+7	+8	-2	+71	-5	+4	+15	-2	=	+130	16

- La prima risposta è fornita dalla serie di tutti i valori positivi. La sua somma fornisce il totale maggiore: con i dati dell'esempio è 150.

- La seconda risposta possibile nella stessa direzione è quella data dal cambiamento di segno nella differenza minore (2); il suo totale è 146 ed è il maggiore dopo il precedente.
- La terza possibile risposta è data dal cambiamento di segno della seconda differenza minore (3); il suo totale è 144.
- Tutte le risposte successive sono costruite nello stesso modo.

Secondo le diverse combinazioni delle differenze calcolate, alcuni totali sono uguali: tali risposte occuperanno lo stesso rango. E' il caso delle risposte 5 e 6, che danno un totale di 140, ottenuto rendendo negative insieme le differenze 2 e 3 oppure la differenza 5 da sola. Altro esempio di possibili risposte che occupano lo stesso rango, sono la 8, la 9 e la 10, riportate nell'elenco, che danno un totale di 136; esso è ottenuto rendendo negative la differenze 7, oppure 5 e 2 insieme, oppure 4 e 3 insieme.

**Verificare se la serie dei dati osservati nell'esperimento rientrano nella zona di rifiuto.**

6 - Dall'elenco parziale riportato, si evidenzia che **la risposta ottenuta nell'esperimento occupa la 12<sup>a</sup> posizione, tra le risposte positive più estreme.** La sua somma è uguale a quella fornita dalla 11<sup>a</sup> risposta e quindi **il rango della sua posizione è 11,5.**

**Ai fini dell'inferenza sulla significatività della serie delle differenze osservate, si osserva che essa rientra tra le 25 risposte collocate in un estremo della zona di rifiuto, per un test a due code.** Di conseguenza, **alla probabilità  $\alpha = 0.05$  si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa bilaterale:** esiste una differenza significativa tra **le medie** delle due serie di dati appaiati.

E' possibile una **valutazione più precisa della probabilità di trovare la risposta osservata o risposte più estreme.**

Con i dati dell'esempio, essa occupa il rango 11,5 su 1024 possibili risposte. Per un test ad una coda, nella condizione che l'ipotesi nulla sia vera ("i segni positivi e negativi sono attribuiti a caso alle differenze mantenute costanti in valore assoluto"), la probabilità  $\alpha$  che la serie osservata sia ottenuta casualmente è  $11,5 / 1024 = 0.0112$ .

Per un test a due code, tale probabilità deve essere moltiplicata per 2; quindi è  $\alpha = 0.0224$

Il limite maggiore di questo test dipende dal numero di dati.

**All'aumentare del numero di osservazioni, il test diventa rapidamente inapplicabile.** Come dimostrazione è sufficiente rilevare che con 20 sole differenze, il numero di possibili risposte è

$$2^{20} = 1.048.576$$

Anche limitando l'analisi al 5% delle risposte più estreme in una sola coda per un test bidirezionale, si tratta sempre di calcolare 26.214 serie di dati ( $1.048.576 \times 0.025 = 26.214,4$ )

**Nel caso di grandi campioni, se non è possibile utilizzare il test t di Student si può ricorrere al test T di Wilcoxon, con una perdita di potenza-efficienza ridotta, inferiore al 5%.**

ESEMPIO 1. Il contributo delle microalghe sia epifitiche che epibentiche alla produzione primaria degli ecosistemi estuariali può essere molto rilevante.

In 7 località è stata misurata la produzione primaria; le misure sono state ripetute nello stesso posto a distanza di un mese.

Località	A	B	C	D	E	F	G
1 <sup>a</sup> misurazione	22	23	26	19	17	23	24
2 <sup>a</sup> misurazione	24	28	25	27	26	29	27

Esiste una differenza significativa tra le due valutazioni?

Risposta.

E' un test bilaterale:

$$H_0: \delta = 0; \quad H_1: \delta \neq 0$$

Il problema fondamentale nella scelta del tipo di test non parametrico da utilizzare, in questo caso la scelta tra test dei segni, test T di Wilcoxon e test di casualizzazione per la verifica di differenze nella tendenza centrale dipende dal **valore reale della misura scelta**.

**Il test di casualizzazione può essere utilizzato solamente se la scala è di rapporti o d'intervalli; come le misure di peso o d'altezza;** se sono punteggi o stime, in cui l'informazione reale è di rango, occorre utilizzare il test T di Wilcoxon.

**Per utilizzare il test di casualizzazione, con i dati dell'esempio si devono calcolare le differenze tra le coppie di dati appaiati, conservando il loro segno**

Località	A	B	C	D	E	F	G
1 <sup>a</sup> misurazione	22	23	26	19	17	23	24
2 <sup>a</sup> misurazione	24	28	25	27	26	29	27
<b>Differenze</b>	<b>+2</b>	<b>+5</b>	<b>-1</b>	<b>+8</b>	<b>+9</b>	<b>+6</b>	<b>+3</b>

e, sulla base del loro numero ( $N = 7$ ), calcolare le possibili risposte

$$2^N, (2^7 = 128)$$

che risultano pari a 128.

Per un test bilaterale, alla probabilità  $\alpha = 0.05$  è compreso nella zona di rifiuto un numero di possibili risposte ( $128 \times 0.05 = 6,4$ ) pari a 6,4 che deve essere arrotondato all'unità per difetto (6): le 6 risposte saranno collocate 3 in una coda e altre 3 nell'altra coda della distribuzione.

Nella distribuzione delle differenze, si osserva che compare un solo valore negativo e che esso è riferito ad una differenza piccola. Per stimare le possibili risposte in una coda della distribuzione, è conveniente iniziare da quella in cui le differenze sono tutte positive e stimare da essa le successive, che danno totali sempre minori, come nella tabella successiva:

1)	+2	+5	+1	+8	+9	+6	+3	= 34
2)	+2	+5	-1	+8	+9	+6	+3	= 32
3)	-2	+5	+1	+8	+9	+6	+3	= 30
4)	+2	+5	+1	+8	+9	+6	-3	= 28
5)	-2	+5	-1	+8	+9	+6	+3	= 28
6)	+2	+5	-1	+8	+9	+6	-3	= 26

Nella tabella sono riportate le 6 risposte più estreme in una direzione.

Per un test bilaterale, alla probabilità  $\alpha = 0.05$  cadono nella zona di rifiuto le prime 3. La risposta fornita dall'esperimento è la 2<sup>a</sup>; di conseguenza, si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa.

E' possibile **stimare la probabilità in modo più preciso**, partendo dall'osservazione che la risposta ottenuta è la seconda ad iniziare da un estremo, sulle 128 teoricamente possibili se  $H_0$  fosse vera.

In un test ad una coda, **la probabilità di ottenerla per caso, insieme** (quindi **distribuzione cumulata**) **con risposte più estreme**, è data da  $2 / 128$  e a risulta uguale a 0.0156.

In un test a due code, occorre considerare anche le due possibili risposte più estreme nell'altra direzione; la probabilità  $\alpha$  diviene  $0.0156 \times 2 = 0.0312$  (oppure  $4 / 128 = 0.0312$ ).

ESEMPIO 2. Si supponga che nell'esempio 1, la risposta osservata come differenza tra le due misurazioni fosse stata

+2	+5	-1	+8	+9	+6	-3
----	----	----	----	----	----	----

coincidente con la 6<sup>a</sup> nell'elenco precedente.

Per un test bilaterale, alla probabilità  $\alpha = 0.05$  tale risposta sarebbe caduta nella zona di accettazione o non rifiuto dell'ipotesi nulla. Per un test unilaterale, alla stessa probabilità avrebbe permesso di rifiutare l'ipotesi nulla; ma non per un test bilaterale.

Infatti, la probabilità  $\alpha$ , calcolata in modo preciso, sarebbe stata ( $6 / 128$ ) uguale a 0.0469 per un test unilaterale e ( $12 / 128$ ) uguale a 0.09375 per un test bilaterale.

## 7.6. TEST PER 2 CAMPIONI INDIPENDENTI

La non-normalità di una distribuzione ha conseguenze rilevanti sulle probabilità calcolate mediante un test parametrico. Quando si dispone di piccoli campioni, è difficile, se non praticamente impossibile, valutare con attendibilità la forma della distribuzione. Spesso la scelta del tipo di approccio inferenziale, se parametrico o non parametrico, non è quindi fondata su elementi rigorosi, ma sulla intuizione o addirittura sulle preferenze personali.

Le opinioni sulla scelta tra test parametrici o corrispondenti test non parametrici divergono. Alcuni ritengono che sia più utile fare uso sempre dei metodi parametrici, quando non sia possibile dimostrare che i dati sono tratti da una popolazione distribuita non normalmente. Altri preferiscono i test non parametrici, in quanto forniscono risultati meno potenti, a volte di poco, ma nettamente più attendibili, soprattutto in caso d'incertezza sulla forma di distribuzione; inoltre essi hanno una perdita di potenza-efficienza trascurabile, in alcuni casi non superiore al 5%, quando sono scelti in modo appropriato rispetto alle caratteristiche dei dati.

Su questo dibattito non esistono risposte oggettive ed universali. E' quindi utile ripetere che in tutti i casi d'incertezza è bene utilizzarli entrambi, poiché solo il confronto tra un test parametrico ed il corrispondente non parametrico permette di ottenere le informazioni più utili sulle probabilità stimate.

**Alcuni test non parametrici per 2 campioni indipendenti sono già stati presentati nell'esposizione dei metodi di confronto tra 2 distribuzioni osservate:**

- il  $\chi^2$  e il test G in tabelle di contingenza 2x2 per grandi campioni,
- il metodo delle probabilità esatte di Fisher per piccoli campioni,
- il  $\chi^2$  in tabelle 2 x n,
- il test di Kolmogorov-Smirnov per valutare l'accordo tra due distribuzioni osservate.

Tra **i test per 2 campioni indipendenti più frequentemente utilizzati per inferenze sulla tendenza centrale**, di seguito illustrati, compaiono

- il test della mediana,
- il test di Wilcoxon-Mann-Whitney della somma dei ranghi,
- il test U di Mann-Whitney dell'ordine robusto dei ranghi insieme con il test S di Kendall,
- il test di permutazione o di casualizzazione.

Essi possono essere ritenuti equivalenti ai test per 2 campioni dipendenti già presentati:

- il test della mediana al test dei segni,
- il test T di Wilcoxon-Mann-Whitney, il test U di Mann-Whitney e il test S di Kendall al test T di Wilcoxon,
- il test di permutazione o casualizzazione al corrispondente test per 2 campioni dipendenti.

Come nel caso precedente, la scelta dipende soprattutto dal tipo di scala utilizzata.

Con due campioni indipendenti sono possibili anche **confronti tra altri parametri della distribuzione oltre alla tendenza centrale, come la variabilità e la forma**, per i quali in questo capitolo sono proposti alcuni test specifici:

- il test delle successioni per due campioni, per verificare se due distribuzioni indipendenti abbiano almeno un parametro significativamente differente;
- il test di Siegel-Tukey, il test di Freund-Ansari-Bradley, il test di Conover e il test dei ranghi equivalenti di Moses, per differenze nella dispersione o variabilità.

### **7.7. IL TEST DELLA MEDIANA O TEST DI MOOD**

Il **test della mediana** è utile **per verificare differenze nella tendenza centrale** tra due campioni indipendenti. In alcuni testi di statistica (informazione utile per capire dal titolo l'argomento in vari programmi informatici) è **chiamato anche test di Mood**, attribuendo appunto ad A. M. Mood il merito della sua divulgazione con il suo testo del 1950 *Introduction to the theory of Statistics* (pubblicato da *McGraw-Hill, New York*); in altri testi è chiamato **test di Brown-Mood, facendo derivare l'attuale presentazione** dall'articolo di G. W. Brown e A. M. Mood del 1951 *On median  $t$  for linear hypotheses* (pubblicato su *Proc. 2<sup>nd</sup>, Berkeley Symp.*, pp. 159-166).

**L'ipotesi nulla è che i due gruppi di osservazioni appartengano alla stessa popolazione o a due popolazioni con la stessa mediana:**

$$H_0 : Me_1 = Me_2$$

con ipotesi alternativa che può essere bilaterale

$$H_1 : Me_1 \neq Me_2$$

o unilaterale

$$H_1 : Me_1 > Me_2 \quad \text{oppure} \quad H_1 : Me_1 < Me_2$$

e dove  $Me_1$  e  $Me_2$  sono rispettivamente la mediana del gruppo 1 e la mediana del gruppo 2.

Il test della mediana è di solito consigliato quando i due gruppi hanno misure approssimate, rilevate con precisione differente in rapporto alle dimensioni del fenomeno, per cui le informazioni contenute hanno attendibilità e precisione differente, massime verso i valori centrali e minime agli estremi. Molti strumenti di misura sono tarati per valori non troppo diversi dalla norma: valutano con accuratezza quelli prossimi alla tendenza centrale, più frequenti, ma sono poco sensibili ai valori estremi nelle due direzioni. Le bilance tarate per misurare grammi non possono essere ugualmente sensibili ai milligrammi e agli ettogrammi; ad una rete per la cattura di animali di una certa specie sfuggono quelli di taglia troppo piccola o eccessivamente grande. Tutte queste misure sono essenzialmente da valutare come scale di rango; utilizzano una scala continua nei valori centrali, ma hanno un'ampia fascia di valori identici agli estremi.

Casi relativamente frequenti in cui applicare il test della mediana per due campioni indipendenti sono la verifica della tendenza centrale tra due serie di dati come quelli riportati

Gruppo A	≤9	≤9	12	15	18	22	26	27	32	38	51	88
Gruppo B	25	28	54	68	76	87	93	>99	>99	>99		

dove è evidente che **non è possibile determinare la somma e la media dei due gruppi, né tutte le statistiche da esse derivate**, per la presenza di alcuni valori non esattamente definiti.

La procedura del test della mediana per due campioni indipendenti è semplice e può essere schematizzata in 4 passaggi.

**1 - Disporre in un gruppo unico in ordine crescente tutti i dati, mantenendo l'indicazione del gruppo d'appartenenza.**

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
≤9	≤9	12	15	18	22	25	26	27	28	32	38	51	54	68	76	87	88	93	>99	>99	>99
A	A	A	A	A	A	B	A	A	B	A	A	A	B	B	B	B	A	B	B	B	B

L'ordine prende in considerazione **le grandezze algebriche: quando sono presenti valori negativi, i ranghi inferiori sono attribuiti ai valori negativi maggiori.**

2 - **Calcolare la mediana del gruppo unico.** Con 22 dati la mediana è tra l'11° (32) e il 12° (38) valore e corrisponde a 35.

Se è vera l'ipotesi nulla, i dati dei due gruppi sono casualmente mescolati. Rispetto alla mediana, i dati di ognuno dei due gruppi dovrebbero essere ripartiti equamente: metà con valore inferiore e metà con valore superiore.

Se è vera l'ipotesi alternativa, i due gruppi non sono mescolati: prima della mediana vi dovrebbe essere una presenza prevalente dei valori di un gruppo e dopo la mediana la presenza prevalente dei valori dell'altro gruppo.

3 - **Costruire una tabella 2x2 per sintetizzare la distribuzione dei valori dei due gruppi rispetto alla mediana:** riportare nella tabella quante misure di ognuno dei due gruppi si trovano prima della mediana e quante dopo. Se il numero di dati è dispari e la mediana coincide con un valore, porre nel primo quelli che non superano ( $\leq$ ) la mediana e nel secondo gruppo i valori che la superano ( $>$ ). Se la mediana cade su un gruppo di valori uguali, non dividere tale gruppo ma spostare la mediana.

Con i dati dell'esempio, la distribuzione in una tabella 2 x 2 risulta

		<b>M E D I A N A</b>		
		≤	>	
<b>Gruppo A</b>	<b>9</b>	<b>3</b>	<b>12</b>	
<b>Gruppo B</b>	<b>2</b>	<b>8</b>	<b>10</b>	
	<b>11</b>	<b>11</b>	<b>22</b>	

4 - **Calcolare la significatività della distribuzione** che con i dati dell'esempio, come nel metodo esatto di Fisher, segue la distribuzione ipergeometrica.

Per l'inferenza si utilizza

- **il test  $\chi^2$  per tabelle 2x2, nel caso di grandi campioni ( $N > 100$ );**
- **si apporta la correzione per la continuità di Yates, nel caso di campioni di dimensioni inferiori ( $100 > N > 30$ );**

- si ricorre al **metodo esatto di Fisher, nel caso di piccoli campioni** ( $N < 30$ ).

**Il test della mediana utilizza solo una parte dell'informazione contenuta nelle due serie di valori: si effettua un conteggio di quanti sono i valori grandi e quanti sono quelli piccoli. E' quindi il meno potente dei test per due campioni indipendenti** che sono qui presentati **e quindi per risultare significativo richiede un numero relativamente grande di dati**, come appunto il test chi quadrato.

ESEMPIO. Nel centro storico di una città, per l'analisi della qualità dell'aria sono state rilevate le quantità di solventi aromatici (Benzene, Toluene, Etilbenzene, Xileni in microgrammi/mc) presenti in un giorno festivo ed in un giorno feriale. Con i dati della tabella

FESTIVO	FERIALE
A	B
92	156
114	123
82	198
164	83
167	242
110	176
135	185
	217

verificare se nel giorno festivo la quantità è significativamente minore di quella presente nel giorno feriale.

Risposta.

E' un test ad una coda, con ipotesi nulla

$$H_0: Me_{\text{festivo}} = Me_{\text{feriale}}$$

ed ipotesi alternativa

$$H_1: Me_{\text{festivo}} < Me_{\text{feriale}}$$

Per costruire la tabella 2x 2, si ordinano i valori in ordine crescente e si individua la mediana: con 15 dati è il valore che occupa il rango 8 (corrispondente al valore 156)

82	83	92	110	114	123	135	<b>156</b>	164	167	176	185	198	217	242
A	B	A	A	A	B	A	<b>B</b>	A	A	B	B	B	B	B

Si contano le osservazioni di ogni gruppo che sono inferiori od uguali alla mediana e le osservazioni che sono superiori, costruendo una tabella 2x2 come la seguente.

	<b>M E D I A N A</b>		
	£	>	
<b>Giorno A</b>	<b>5</b>	<b>2</b>	7
<b>Giorno B</b>	<b>3</b>	<b>5</b>	8
	8	7	15

Poiché il numero di osservazioni è limitato, troppo ridotto per il test  $\chi^2$  anche apportando la correzione per la continuità, la significatività della distribuzione può essere verificata con il metodo esatto di Fisher.

A questo scopo, si calcolano sia la probabilità di ottenere la tabella osservata sia le risposte più estreme nella stessa direzione.

Sono facilmente identificabili: si varia verso 0 la frequenza osservata minore (2), mantenendo fissi i totali marginali.

La probabilità  $P_{(2)}$  di avere la distribuzione osservata

5	2
3	5

è uguale a  $\frac{7!8!8!7!}{5!2!3!5!15!} = 0,18275$ ;

mentre la probabilità  $P_{(1)}$  di avere la distribuzione più estrema

6	<b>1</b>
2	5

è uguale a  $\frac{7!8!8!7!}{6!1!2!6!15!} = 0,03046$

e la probabilità  $P_{(0)}$  di avere quella ancora più estrema

7	0
1	7

è uguale a  $\frac{7!8!8!7!}{7!0!1!7!15!} = 0,00124$

La probabilità totale risulta uguale a 0.21445 (0,18275 + 0,03046 + 0,00124), per un test ad una coda. E' molto elevata, per cui non si può rifiutare l'ipotesi nulla.

I dati riportati in questo ultimo esempio permettono analisi più potenti del test della mediana (come quelli di seguito riportati).

### 7.8. IL TEST DI WILCOXON-MANN-WHITNEY DELLA SOMMA DEI RANGHI

L'esempio precedente mostra come nel test della mediana un dato fornisca lo stesso contributo al calcolo della significatività, sia quando è prossimo alla mediana sia quando, pur essendo dalla stessa parte, se ne discosta sensibilmente, perché molto grande o molto piccolo. Da questa osservazione elementare, è semplice dedurre che il test della mediana utilizza solamente una parte ridotta dell'informazione contenuta nelle due serie di dati.

Tra vari **test che utilizzano appieno l'informazione del rango**, nel caso di due campioni indipendenti i più noti sono il test di **Wilcoxon-Mann-Whitney** (a volte abbreviato in **WMW test**) o **test della somma dei ranghi**, ed il **test U di Mann-Whitney**, detto anche **test dell'ordine robusto dei ranghi**. Entrambi servono per verificare ipotesi, che possono essere bilaterali od unilaterali, sulla tendenza centrale.

In entrambi i casi **l'ipotesi verte sulla mediana**.

**Il test proposto da F. Wilcoxon** (quindi in vari testi chiamato semplicemente **test di Wilcoxon per 2 campioni indipendenti**) nel 1945 **per due campioni indipendenti ma delle stesse dimensioni** (*Individual comparisons by ranking method*, in *Biometrics* 1, pp. 80-83), **generalizzato da H. B. Mann e D. R. Witney al caso di 2 campioni con un numero differente di repliche** nel 1947 (*On a test of whether one of two random variables is stochastically larger than the other*, in *Ann. Math. Statist.*, 18, pp. 50-60) **richiede che**

**le due popolazioni a confronto siano distribuite in modo continuo,**

**abbiano la stessa forma rispetto alla simmetria** (entrambe simmetriche o entrambe asimmetriche nello stesso modo) e che  
**i dati siano misurati con una scala almeno ordinale.**

Con un esempio, è possibile spiegare la metodologia in modo chiaro e semplice.

Si ripropone l'esercizio precedente, per un più evidente confronto tra le procedure ed i risultati.

A	B
92	156
114	123
82	198
164	83
167	242
110	176
135	185
	217

**1 - Combinare i dati dei due gruppi in una serie unica, disponendo i valori o punteggi in ordine crescente.** Per questo ordinamento in ranghi, quando esistono dati negativi si utilizzano i valori algebrici: si assegnano i ranghi minori ai valori negativi che sono maggiori in modulo.

Per ogni dato, **conservare l'informazione relativa al gruppo di appartenenza**, come nella tabella seguente:

82	83	92	110	114	123	135	156	164	167	176	185	198	217	242
<b>A</b>	<b>B</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>B</b>	<b>A</b>	<b>B</b>	<b>A</b>	<b>A</b>	<b>B</b>	<b>B</b>	<b>B</b>	<b>B</b>	<b>B</b>

**2 - Definire**

**$n_1$  la dimensione del gruppo minore e**

**$n_2$  la dimensione del gruppo maggiore.**

Se i 2 gruppi hanno la stessa dimensione, l'attribuzione di  $n_1$  e  $n_2$  può essere casuale, in quanto indifferente sui risultati.

Nell'esempio,  $n_1 = 7$  e  $n_2 = 8$ .

**Attribuire il rango ad ogni valore.**

82	83	92	110	114	123	135	156	164	167	176	185	198	217	242
A	B	A	A	A	B	A	B	A	A	B	B	B	B	B
<u>1</u>	2	<u>3</u>	<u>4</u>	<u>5</u>	6	<u>7</u>	8	<u>9</u>	<u>10</u>	11	12	13	14	15

3 - **Calcolare la somma dei ranghi (chiamato T) del gruppo con il numero di dati minore.**

Nell'esempio è il gruppo A: il numero suo di repliche è indicato con  $n_1$  e i suoi ranghi sono stati riportati nella tabella precedente in grassetto sottolineato.

La somma di questi ranghi

$$T = 1 + 3 + 4 + 5 + 7 + 9 + 10 = 39$$

è indicata con **T** e risulta uguale a 39.

4 - **Quando l'ipotesi nulla  $H_0$  è vera**, i valori del gruppo prescelto sono casualmente mescolati con quelli dell'altro gruppo. Di conseguenza, **il valore di T tende ad una media attesa  $m_T$** , che dipende dal numero di osservazioni  $n_1$  e  $n_2$  che compongono i 2 gruppi, secondo la relazione

$$\mu_T = \frac{n_1 \cdot (n_1 + n_2 + 1)}{2}$$

Con i dati dell'esempio,  $n_1 = 7$  e  $n_2 = 8$

$$\mu_T = \frac{7 \cdot (7 + 8 + 1)}{2} = 56$$

**la media attesa  $m_T$**  risulta uguale a 56.

Se è vera l'ipotesi alternativa  $H_1$ , in rapporto alla coda della distribuzione in cui è collocata la tendenza centrale del gruppo con meno dati (cioè se la sua media è minore o maggiore di quella dell'altro gruppo), il valore di **T** tende a un valore minimo, dato dalla somma degli  $n_1$  ranghi minori, oppure a un valore massimo, determinato dalla somma degli  $n_1$  ranghi che in questo caso saranno i maggiori.

5- **La significatività della differenza tra le mediane dei due gruppi può essere valutata confrontando il valore di T calcolato con il valore  $m_T$  atteso.**

Nel **caso di piccoli campioni ( $n_1$  e  $n_2 \leq 10$ )**, i valori critici alla **probabilità  $\alpha = 0.05$**  sono forniti dalla tabella per un **test unilaterale**. Per ogni coppia  $n_1$  ed  $n_2$  riporta la probabilità di trovare il valore di  $T$  osservato, o valori che si discostano maggiormente dalla media, nella coda inferiore ed in quella superiore della distribuzione.

Con i dati dell'esempio, per  $n_1 = 7$  e  $n_2 = 8$  la media attesa è 56 e il valore di  $T$  calcolato è stato 39. Si tratta quindi di verificare se è minore dell'atteso in modo significativo.

**Il valore critico fornito dalla tabella successiva per  $n_1 = 7$  e  $n_2 = 8$  è  $T = 41$ .**

Il valore osservato (39) è inferiore a quello critico riportato nella tabella (41): di conseguenza, la probabilità  $\alpha$  che possa essere ottenuto per caso è inferiore a 0.05.

Per un test ad una coda, come espresso nella domanda dell'esercizio, si rifiuta l'ipotesi nulla.

E' utile **osservare come questo test sia molto più potente di quello della mediana**: con gli stessi dati, il test della mediana stimava una probabilità pari a 0.21445, molto distante dal valore critico e che non permetteva di rifiutare l'ipotesi nulla.

**Tavola dei valori critici di T del test di Wilcoxon-Mann-Whitney  
per 2 campioni indipendenti  
alla probabilità  $\alpha = 0.05$  per test ad una coda**

$n_1$  = campione con il numero minore di osservazioni,

$n_2$  = campione con il numero maggiore di osservazioni,

T è significativo quando è uguale o minore del valore minore tabulato, se inferiore alla media attesa,

T è significativo quando è uguale o maggiore del valore maggiore tabulato, se superiore alla media  $m_T$

$n_1 = 3$		$n_1 = 4$		$n_1 = 5$		$n_1 = 6$	
$n_2$	T	$n_2$	T	$n_2$	T	$n_2$	T
3	6-15						
4	6-18	4	11-25				
5	7-20	5	12-28	5	19-36		
6	8-22	6	13-31	6	20-40	6	28-50
7	8-25	7	14-34	7	21-44	7	29-55
8	9-27	8	15-37	8	23-47	8	31-59
9	10-29	9	16-40	9	24-51	9	33-63
10	10-32	10	17-43	10	26-54	10	35-67

$n_1 = 7$		$n_1 = 8$		$n_1 = 9$		$n_1 = 10$	
$n_2$	T	$n_2$	T	$n_2$	T	$n_2$	T
7	39-66						
8	41-71	8	51-85				
9	43-76	9	54-90	9	66-105		
10	45-81	10	56-96	10	69-111	10	82-128

Nel caso di grandi campioni ( $n_1$  o  $n_2 > 10$ ) la statistica T segue una distribuzione approssimativamente normale. La significatività può pertanto essere saggiata mediante la distribuzione normale ridotta, con media uguale a 0 e varianza uguale a 1.

Anche quando in un gruppo si hanno più di 10 osservazioni, è conveniente applicare la correzione per la continuità, sommando  $\pm 0,5$  al valore di  $T$  in modo che lo scarto tra osservato ed atteso sia più piccolo.

La formula della stima di  $z$ , corretta per la continuità, diviene

$$z = \frac{(T \pm 0,5) - m_T}{s_T}$$

dove la media  $m_T$  è data da

$$\mu_T = \frac{n_1 \cdot (n_1 + n_2 + 1)}{2}$$

e la deviazione standard  $s_T$  da

$$\sigma_T = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}$$

Con i dati dell'esempio,

$$\mu_T = \frac{7 \cdot (7 + 8 + 1)}{2} = 56$$

dove  $m_T$  è uguale a 56

e la deviazione standard  $s_T$  è

$$\sigma_T = \sqrt{\frac{7 \cdot 8 \cdot (7 + 8 + 1)}{12}} = 8,64$$

uguale a 8,64

la significatività della differenza tra le due tendenze centrali può essere verificata attraverso il valore di  $z$

$$z = \frac{(39 + 0,5) - 56}{8,64} = -1,91$$

che risulta uguale a -1,91.

Il segno indica che il valore osservato è inferiore a quello atteso; al valore di 1,91 in una coda della distribuzione nella tabella della distribuzione normale corrisponde un'area di probabilità  $\alpha = 0.0281$ .

Si rifiuta l'ipotesi nulla.

Quando i punteggi non sono valutati con una scala continua, come postula il test, si possono avere diversi valori uguali od **osservazioni ex-aequo (ties)**. Nella trasformazione in ranghi, ad ognuna di queste osservazioni viene assegnata la media dei ranghi dei valori uguali. La media resta invariata, ma la deviazione standard  $s_T$  è minore; di conseguenza deve essere corretta e diventa

$$s_T = \sqrt{\frac{n_1 \cdot n_2}{N(N-1)} \cdot \left( \frac{N^3 - N}{12} - \sum_{j=1}^g \frac{t_j^3 - t_j}{12} \right)}$$

dove  $N = n_1 + n_2$  e la stima dei  $t$  è condotta come nel test  $T$  di Wilcoxon per 2 campioni dipendenti.

**La correzione per i ties diminuisce il valore della deviazione standard e quindi aumenta il valore di z.**

ESEMPIO. Quantità di precipitazione, intensità e durata delle piogge variano da luogo a luogo. L'irregolarità degli eventi atmosferici e una diversa configurazione sia del suolo che di aspetti ambientali (zone montuose, aree alberate od urbanizzate) possono influire sulla quantità di pioggia. Tuttavia è ritenuto approssimativamente corretto considerare valide per tutta l'area le misure effettuate in una stazione pluviometrica.

Per verificare se un'area montana (M) ha avuto una quantità di piogge significativamente superiore a quella di un'area collinare (C) limitrofa, si confrontano due serie mensili (in millimetri).

AREA	AREA
M	C
78	43
130	58
93	69
110	96
	72
	85

Risposta.

E' un test ad una coda, con le seguenti ipotesi sulle mediane

$$H_0: Me_M = Me_C \quad H_1: Me_M > Me_C$$

Per effettuare il test,

1 - Ordinare i valori, conservando l'informazione del gruppo d'appartenenza; successivamente riportare i ranghi relativi.

43	58	69	72	78	85	93	96	110	130
C	C	C	C	M	C	M	C	M	M
1	2	3	4	<b>5</b>	6	<b>7</b>	8	<b>9</b>	<b>10</b>

2 - Calcolare la somma dei ranghi del gruppo **M** (già riportati in grassetto), con  $n_1 = 4$  e  $n_2 = 6$

$$T = 5 + 7 + 9 + 10 = 31$$

3 - Se  $H_0$  fosse vera, il valore atteso di  $m_T$  dovrebbe essere

$$m_T = \frac{4(4 + 6 + 1)}{2} = \frac{44}{2} = 22$$

uguale a 22.

Il valore osservato è più alto di quello atteso; pertanto il confronto deve essere condotto con il valore critico superiore.

Nella tabella dei valori critici, per  $n_1 = 4$  e  $n_2 = 6$  il valore massimo riportato è 31.

Il valore di **T** calcolato è esattamente uguale (31).

Poiché il valore riportato nella tabella appartiene all'area di rifiuto, si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa.

La mediana delle quantità mensili di pioggia caduta nella zona montuosa è significativamente superiore alla mediana delle quantità mensili cadute in collina.

## 7.9. IL TEST U DI MANN-WHITNEY O DELL'ORDINE ROBUSTO DEI RANGHI

Il test U di Mann-Whitney o test dell'ordine robusto dei ranghi deriva dalle proposte di Mann e di Whitney di generalizzare il metodo di Wilcoxon: **non richiede alcuna ipotesi sulla simmetria dei due campioni**, può essere applicato quando essi hanno dimensioni diverse (come d'altronde il test di Wilcoxon nella versione presentata nel paragrafo precedente) e serve sempre per **verificare la significatività della differenza tra le mediane**.

L'**ipotesi nulla** (come nel test della mediana) è **che i due gruppi di osservazioni appartengano alla stessa popolazione o a due popolazioni con la stessa mediana**:

$$H_0 : Me_1 = Me_2$$

con ipotesi alternativa che può essere bilaterale

$$H_1 : Me_1 \neq Me_2$$

o unilaterale

$$H_1 : Me_1 > Me_2 \quad \text{oppure} \quad H_0 : Me_1 < Me_2$$

e dove

$Me_1$  e  $Me_2$  sono rispettivamente la mediana del gruppo 1 e la mediana del gruppo 2.

A motivo della sua più estesa applicabilità è preferibile al test di Wilcoxon-Mann-Whitney e molti programmi informatici recenti non accennano più al test precedente, sebbene esso mantenga ancora una relativa diffusione internazionale.

**La procedura del test U di Mann-Whitney è fondata sulle precedenze, che rappresentano l'altra metodologia più diffusa nei test non parametrici, alternativa ai ranghi.** Essa utilizza non la somma dei ranghi, ma la somma delle differenze tra i ranghi dei due gruppi; in molti testi, il metodo è spiegato in modo più semplice e comprensibile mediante il concetto, ad essa equivalente, delle precedenze.

Come già illustrato da Mann e Whitney nella loro prima presentazione (e come sarà in seguito dimostrato), è facile passare dal risultato di questo test a quello del test di Wilcoxon (come sarà spiegato nel paragrafo successivo).

E' didatticamente utile avvalersi ancora dell'esempio utilizzato in precedenza:

A	B
92	156
114	123
82	198
164	83
167	242
110	176
135	185
	217

Nel primo passaggio, la sequenza delle operazioni richieste è simile; le differenze iniziano a partire dal secondo passaggio:

**1 - Combinare i dati dei due gruppi in un insieme unico, disponendo i valori o punteggi in ordine crescente, secondo il valore algebrico.** Per ogni dato, conservare l'informazione relativa al gruppo di appartenenza.

82	83	92	110	114	123	135	156	164	167	176	185	198	217	242
A	B	A	A	A	B	A	B	A	A	B	B	B	B	B

2 - **Contare il numero di precedenze:** quante volte ogni dato di un gruppo è preceduto da dati dell'altro gruppo.

Per esempio, il valore 82, che appartiene al gruppo A non è preceduto da alcun valore di B; di conseguenza il suo numero di precedenze è 0;

i valori 92, 110 e 114 del gruppo A sono tutti tre preceduti da un valore di B (83); di conseguenza ognuno di questi tre valori come numero di precedenze ha 1.

**Come indicatore, chiamato U, è stato scelto il numero minore di precedenze.**

Con i dati dell'esempio, è corretto contare quante volte ogni dato di A è preceduto da dati di B. La somma di queste precedenze

82	83	92	110	114	123	135	156	164	167	176	185	198	217	242
A	B	A	A	A	B	A	B	A	A	B	B	B	B	B
<b>0</b>		<b>1</b>	<b>1</b>	<b>1</b>		<b>2</b>		<b>3</b>	<b>3</b>					

è il valore di U

$$U = 0 + 1 + 1 + 1 + 2 + 3 + 3 = 11$$

che risulta uguale a 11.

Si sarebbe anche potuto calcolare quante volte ogni valore del gruppo B è preceduto da valori di A

82	83	92	110	114	123	135	156	164	167	176	185	198	217	242
A	B	A	A	A	B	A	B	A	A	B	B	B	B	B
	<b>1</b>				<b>4</b>		<b>5</b>			<b>7</b>	<b>7</b>	<b>7</b>	<b>7</b>	<b>7</b>

ottenendo un valore

$$U' = 1 + 4 + 5 + 7 + 7 + 7 + 7 + 7 = 45$$

uguale a 45, maggiore del precedente.

**Il valore corretto dell'indice U è quello minore,**

mentre quello maggiore deve essere indicato con U'.

3 - Quando le differenze tra  $U$  e  $U'$  sono ridotte, non sempre è facile **trovare subito il valore corretto**. A questo scopo, è utile ricordare che  $U$  e  $U'$  sono legati dalla relazione

$$n_1 \cdot n_2 = U + U'$$

dove:

$n_1$  è il numero di dati del gruppo minore,

$n_2$  è il numero di dati del gruppo maggiore.

Con i dati dell'esempio, dove

$$n_1 = 7, \quad n_2 = 8, \quad U = 11, \quad U' = 45$$

$$7 \cdot 8 = 11 + 45$$

è possibile calcolare un primo valore di  $U$  e, mediante la relazione, stimare l'altro.

**Il valore da utilizzare, il vero  $U$ , è il valore minore,**

$$U = (n_1 \cdot n_2) - U'$$

che può essere verificato facilmente, per escludere eventuali errori di calcolo.

4 - **Nel caso in cui sia vera l'ipotesi  $H_1$** , quindi un campione abbia una mediana nettamente minore dell'altro, **il valore di  $U$  tenderà a 0**, poiché i dati del gruppo in esame, che deve fornire il totale minore, precederanno tutti i dati dell'altro gruppo e quindi ognuno di essi avrà 0 precedenze.

**Nel caso in cui sia vera l'ipotesi  $H_0$**  di uguaglianza od identità delle due tendenze centrali, i dati dei due gruppi saranno casualmente mescolati:  **$U$  tenderà ad un valore medio ( $\mu_U$ )**, dipendente dal numero di osservazioni presenti  $n_1$  e  $n_2$ , secondo la relazione

$$m_U = \frac{n_1 \cdot n_2}{2}$$

5 - **Per valutare la significatività del valore di  $U$  si seguono metodi diversi, in funzione delle dimensioni dei due campioni.**

**Tavola dei valori critici di U del test Mann-Whitney  
per 2 campioni indipendenti alla probabilità  $\alpha = 0.05$**

Valore critico per test a due code (2) nella parte superiore e per una coda (1) nella parte inferiore di ogni casella. U calcolato è significativo quando è uguale o minore del valore tabulato.

$n_1$  = campione con il numero minore di osservazioni.

$n_2$  = campione con il numero maggiore di osservazioni.

$n_1$	$n_2$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	2	-	-	-	-	-	-	0	0	0	0	1	1	1	1
	1	-	-	-	0	0	0	1	1	1	1	2	2	3	3
3	2		-	-	0	1	1	2	2	3	3	4	4	5	5
	1		0	0	1	2	2	3	4	4	5	5	6	7	7
4	2			0	1	2	3	4	4	5	6	7	8	9	10
	1			1	2	3	4	5	6	7	8	9	10	11	12
5	2				2	3	5	6	7	8	9	11	12	13	14
	1				4	5	6	8	9	11	12	13	15	16	18
6	2					5	6	8	10	11	13	14	16	17	19
	1					7	8	10	12	14	16	17	19	21	23
7	2						8	10	12	14	16	18	20	22	24
	1						11	13	15	17	19	21	24	26	28
8	2							13	15	17	19	22	24	26	29
	1							15	18	20	23	26	28	31	33
9	2								17	20	23	26	28	31	34
	1								21	24	27	30	33	36	39
10	2									23	26	29	33	36	39
	1									27	31	34	37	41	44
11	2										30	33	37	40	44
	1										34	38	42	46	50
12	2											37	41	45	49
	1											42	47	51	55
13	2												45	50	54
	1												51	56	61
14	2													55	59
	1													61	66
15	2														64
	1														72

## Tabella dei valori critici del test U di Mann-Whitney

(campioni di dimensioni diverse)

I valori della matrice triangolare superiore si riferiscono alla probabilità  $\alpha = 0.05$ .

I valori della matrice triangolare inferiore si riferiscono alla probabilità  $\alpha = 0.01$ .

### Test a 1 coda

		$n_2$																
		5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
$n_1$	5		5	6	8	9	11	12	13	15	16	18	19	20	22	23	25	
	6	2		8	10	12	14	16	17	19	21	23	25	26	28	30	32	
	7	3	4		13	15	17	19	21	24	26	28	30	33	35	37	39	
	8	4	6	7		18	20	23	26	28	31	33	36	39	41	44	47	
	9	5	7	9	11		24	27	30	33	36	39	42	45	48	51	54	
	10	6	8	11	13	16		31	34	37	41	44	48	51	55	58	62	
	11	7	9	12	15	18	22		38	42	46	50	54	57	61	65	69	
	12	8	11	14	17	21	24	28		47	51	55	60	64	68	72	77	
	13	9	12	16	20	23	27	31	35		56	61	65	70	75	80	84	
	14	10	13	17	22	26	30	34	38	43		66	71	77	82	87	92	
	15	11	15	19	24	28	33	37	42	47	51		77	83	88	94	100	
	16	12	16	21	26	31	36	41	46	51	56	61		89	95	101	107	
	17	13	18	23	28	33	38	44	49	55	60	66	71		102	109	115	
	18	14	19	24	30	36	41	47	53	59	65	70	76	82		116	123	
	19	15	20	26	32	38	44	50	56	63	69	75	82	88	94		130	
	20	16	22	28	34	40	47	53	60	67	73	80	87	93	100	107		

Valori delle diagonali (campioni di uguali dimensioni).

$n_1 = n_2$	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\alpha = 0.05$	4	7	11	15	21	27	34	42	51	61	72	83	96	109	123	138
$\alpha = 0.01$	1	3	6	9	14	19	25	31	39	47	56	66	77	88	101	114

## Tabella dei valori critici del test U di Mann-Whitney

(campioni di dimensioni diverse)

I valori della matrice triangolare superiore si riferiscono alla probabilità  $\alpha = 0.05$ .

I valori della matrice triangolare inferiore si riferiscono alla probabilità  $\alpha = 0.01$ .

### Test a 2 code

		$n_2$															
		5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$n_1$	5		3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
	6	1		6	8	10	11	13	14	16	17	19	21	22	24	25	27
	7	1	3		10	12	14	16	18	20	22	24	26	28	30	32	34
	8	2	4	6		15	17	19	22	24	26	29	31	34	36	38	41
	9	3	5	7	9		20	23	26	28	31	34	37	39	42	45	48
	10	4	6	9	11	13		26	29	33	36	39	42	45	48	52	55
	11	5	7	10	13	16	18		33	37	40	44	47	51	55	58	62
	12	6	9	12	15	18	21	24		41	45	49	53	57	61	65	69
	13	7	10	13	17	20	24	27	31		50	54	59	63	67	72	76
	14	7	11	15	18	22	26	30	34	38		59	64	69	74	78	83
	15	8	12	16	20	24	29	33	37	42	46		70	75	80	85	90
	16	9	13	18	22	27	31	36	41	45	50	55		81	86	92	98
	17	10	15	19	24	29	34	39	44	49	54	60	65		93	99	105
	18	11	16	21	26	31	37	42	47	53	58	64	70	75		106	112
	19	12	17	22	28	33	39	45	51	57	63	69	74	81	87		119
	20	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	

Valori delle diagonali (campioni di uguali dimensioni).

$n_1 = n_2$	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\alpha = 0.05$	2	5	8	13	17	23	30	37	45	55	64	75	87	99	113	127
$\alpha = 0.01$	0	2	4	7	11	16	21	27	34	42	51	60	70	81	93	105

Nel **caso di piccoli campioni ( $n_1$  e  $n_2 < 15$ )**, la tavola dei valori critici fornisce il valore di  $U$  significativo. Nella tabella sono riportati i valori critici alla probabilità  $\alpha$  uguale a 0.05

- per test a due code nella parte superiore di ogni casella,
- per test a una coda nella parte inferiore.

**E' significativo qualunque valore di  $U$  calcolato che sia uguale o inferiore a quello riportato nella tabella.**

Per  $n_1 = 7$  e  $n_2 = 8$  in un test ad una coda, il valore di  $U$  riportato nella tabella alla probabilità  $\alpha=0.05$  è 13. Con i dati dell'esempio, il valore  $U$  calcolato (11) risulta inferiore a quello tabulato (13). Pertanto alla probabilità 0.05 si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa: la mediana del primo gruppo è significativamente minore di quella del secondo gruppo.

**Ai fini dell'inferenza il confronto è tra le mediane**, anche se non è necessario calcolarle.

**Quando sono presenti valori identici (ties), si deve assegnare il rango medio** e stimare il numero di precedenze come media dei diversi ranghi possibili.

Nel caso di piccoli campioni non si dovrebbero avere valori identici, come richiesto dalla condizione che la scala sia continua. Valori identici non alterano la media, ma abbassano il valore della varianza, che non viene considerata quando si deve ricorrere alla tabella dei valori critici per piccoli campioni.

Un numero molto limitato di valori ex-aequo incide in modo trascurabile sulla stima della significatività; pertanto nella prassi viene abitualmente accettato il valore che non considera la correzione, che invece può essere utile nel caso di grandi campioni.

Nel **caso di grandi campioni ( $n_1$  o  $n_2 > 15$ )**, se è vera l'ipotesi nulla  $H_0$  la distribuzione di campionamento di  $U$  è bene approssimato dalla distribuzione normale, con media 0 e varianza unitaria:

$$z = \frac{U - m_U}{s_U}$$

dove  $U$  è lo stimatore osservato,

$\mu_U$  è il suo valore atteso nell'ipotesi  $H_0$

$$m_U = \frac{n_1 \cdot n_2}{2}$$

e la deviazione standard  $\sigma_U$  è

$$s_U = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}$$

Con i dati dell'esempio,  $U$  è risultato uguale a 11,  
mentre  $\mu_U$

$$\mu_U = \frac{7 \cdot 8}{2} = 28$$

è uguale a 28

e la deviazione standard  $\sigma_U$

$$s_U = \sqrt{\frac{7 \cdot 8 \cdot (7 + 8 + 1)}{12}} = 8,64$$

è uguale a 8,64.

La significatività della differenza tra le mediane dei due gruppi indipendenti può essere stimata mediante il valore di  $z$

$$z = \frac{11 - 28}{8,64} = -1,967$$

che risulta uguale a -1,967.

Nella distribuzione normale a  $z = 1,967$  in una coda della distribuzione corrisponde una probabilità leggermente inferiore a 0.0250.

Si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa di una minore presenza di solventi aromatici dispersi nell'aria della città, durante i giorni festivi.

**Il valore di  $z$  risulta sempre negativo e non assume un significato preciso, in quanto  $U$  è sempre inferiore alla media attesa  $m_U$ .**

Il test  $U$  ipotizza che la scala utilizzata sia continua e quindi che si abbiano poche osservazione ex-aequo. **In caso di ties**, ad ogni gruppo di punteggi uguali sarà assegnata la media del gruppo. La presenza di valori identici non altera la media ma modifica la varianza.

Nel caso di grandi campioni ed il ricorso alla distribuzione normale, la deviazione standard  $\sigma_U$  diventa

$$s_U = \sqrt{\frac{n_1 \cdot n_2}{N \cdot (N - 1)} \cdot \left( \frac{N^3 - N}{12} - \sum T_i \right)}$$

dove:

$N$  è uguale a  $n_1 + n_2$

$T_i$  è dato da:

$$T_i = \frac{t_i^3 - t_i}{12}$$

con  $t_i$  = numero di osservazioni ex-aequo dello stesso rango.

**La correzione riduce il valore della deviazione standard ed aumenta proporzionalmente il valore di z.** Se non si introduce la correzione, il valore di **z** risulta minore e quindi è più difficile raggiungere la significatività; si dice anche che, senza correzione, il test risulta più conservativo. Di norma anche molti valori simili determinano una correzione ridotta: una sola serie molto lunga di valori identici ha effetti maggiori di molti valori ripetuti solo due o tre volte.

ESEMPIO. Per verificare se un'area montana (M) ha avuto una quantità di piogge significativamente superiore a quella di un'area collinare (C) limitrofa si confrontano due serie mensili (in millimetri):

AREA	AREA
M	C
78	43
130	58
93	69
110	96
	72
	85

La quantità mediana di pioggia caduta nell'area montana è significativamente superiore a quella dell'area collinare?

Risposta.

E' un test ad una coda, con le seguenti ipotesi sulle mediane

$$H_0: Me_M = Me_C \quad H_1: Me_M > Me_C$$

Per applicare il test effettuare i seguenti passaggi operativi:

1 - Ordinare i valori, conservando l'informazione del gruppo d'appartenenza.

43	58	69	72	78	85	93	96	110	130
C	C	C	C	M	C	M	C	M	M

2 - Riportare le precedenze e sommarle, determinando U

43	58	69	72	78	85	93	96	110	130
C	C	C	C	M	C	M	C	M	M
<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>		<b>1</b>		<b>2</b>		

che risulta uguale a 3.

Per  $n_1 = 4$  e  $n_2 = 6$ , il valore critico riportato nella tabella per un test unilaterale alla probabilità 0.05 risulta uguale a 3. Il valore calcolato è uguale a quello riportato nella tabella: il test risulta significativo e si rifiuta l'ipotesi nulla.

#### 7.10. CENNI DEL TEST S DI KENDALL E RAPPORTI CON T ED U

##### POTENZA-EFFICIENZA DEI TRE TEST

Da vari studiosi di statistica, sono state proposte **varianti del test U di Mann-Whitney: tra esse il più diffuso attualmente è il test S di Kendall**. Per il test U è indifferente considerare la somma minore o maggiore delle precedenze. La loro somma è definita dal numero di osservazioni  $n_1$  e  $n_2$  dei due gruppi; nel caso in cui l'ipotesi nulla sia vera, tendono ad un valore medio ( $n_1 \times n_2 / 2$ ), mentre quando l'ipotesi nulla è falsa, il valore minore di U tende a 0 e l'altro tende simmetricamente al valore massimo ( $n_1 \times n_2$ ), che corrisponde al doppio della media.

Alcuni testi di statistica, forse per distaccarsi (apparentemente) dal metodo tradizionale, suggeriscono di utilizzare il valore maggiore e forniscono le tabelle relative; ma sono solo valori che invece di variare dalla media a zero, variano simmetricamente dalla media al valore massimo.

**La statistica S proposta da Kendall deriva dalla differenza tra le due possibili stime di U:**

$$S = U' - U$$

Quando l'ipotesi nulla  $H_0$  è vera,  $S = 0$ ;

quando l'ipotesi nulla  $H_0$  è falsa, tende al valore massimo per cui  $S = n_1 \times n_2$ .

**Anche il test T, fondato sulla somma dei ranghi, ed il test U, che utilizza la somma delle precedenze, sono tra loro matematicamente equivalenti.**

Pure per il test T è possibile utilizzare indifferentemente la somma dei ranghi che risulta minore ( $T_1$ ) oppure quella che risulta maggiore ( $T_2$ ): sono distribuite in modo simmetrico rispetto al valore medio e la loro somma è definita dal numero di osservazioni nei due campioni.

Come tra  $U$  e  $U'$  si ha la relazione

$$U + U' = n_1 \times n_2$$

tra  $T_1$  e  $T_2$  esiste la relazione

$$T_1 + T_2 = \frac{1}{2}(n_1 + n_2) \times (n_1 + n_2 + 1)$$

Poiché entrambe sono definite da  $n_1$  e  $n_2$ ,

dato  $T_1$  oppure  $T_2$  e conoscendo  $n_1$  e  $n_2$ , è semplice passare al valore di  $U$

$$U = T_1 - \frac{1}{2} n_1 (n_1 + 1)$$

o di  $U'$

$$U' = T_2 - \frac{1}{2} n_2 (n_2 + 1)$$

e da essi al valore di  $S$

$$S = U' - U$$

dove  $U'$  è il valore maggiore, secondo la simbologia già usata nel test relativo.

Queste relazioni sono già presenti nell'articolo di Mann e Whitney, per dimostrare che il loro metodo era solo un'elaborazione più semplice e generale del lavoro di Wilcoxon.

Alcuni autori forniscono un'interpretazione meno restrittiva sull'utilizzazione del test fondato sui ranghi, evidenziando che la condizione di validità riguarda la simmetria e non la varianza.

V. H. Kruskal in un'analisi del 1952 (*Use of ranks in one-criterion variance analysis*, pubblicato su *Amer. Statist. Ass.* 23, pp. 525-540) afferma che i test fondati sui ranghi sono poco sensibili al fatto di avere varianze diverse.

Pure con condizioni di validità differenti,

**i tre test (T, U, S) forniscono esattamente lo stesso livello di significatività.**

La scelta del test è spesso condizionata dalla disponibilità delle tabelle dei valori critici, dei programmi informatici, dalla familiarità o maggiore conoscenza di una procedura. E' quindi importante conoscere le relazioni esistenti tra i tre test, poiché in letteratura nella presentazione di ricerche quasi sempre viene utilizzato soltanto uno di essi.

**L'efficienza asintotica relativa del test T, del test U e del test S** può essere valutata **nei confronti del test t di Student per 2 campioni indipendenti**, il test parametrico a loro equivalente.

L'efficienza asintotica varia in funzione della distribuzione dei dati ed è uguale a quella già presentata per i test fondati sui ranghi nel caso di un campione e di 2 campioni dipendenti:

- con una distribuzione normale dei dati è uguale a circa 0,95 ( $3/\pi$ ),
- con una distribuzione rettangolare dei dati è uguale a 1,
- con una distribuzione esponenziale doppia è uguale a 1,50 ( $3/2$ ).

#### **7.11. TEST DI CASUALIZZAZIONE PER 2 CAMPIONI INDIPENDENTI**

**Il test di permutazione (*permutation test*) o di casualizzazione è il test non parametrico più potente, per il confronto tra le tendenze centrali di 2 campioni indipendenti.**

Rispetto al test **t** di Student e al test **F** di Fisher, presenta i vantaggi classici di molti test non parametrici:

- **non richiede la verifica di alcun postulato in merito alla popolazione d'origine dei dati campionari, in riferimento alla variabilità dei due gruppi a confronto e alla forma della distribuzione,**
- **fornisce direttamente le probabilità esatte, senza imporre il ricorso a tavole di distribuzione dei valori critici** (come già dimostrato per la distribuzione binomiale, il metodo esatto di Fisher e il test di casualizzazione per due campioni dipendenti).

A differenza degli altri test non parametrici che verificano la significatività della differenza tra 2 tendenze centrali, il test di casualizzazione **richiede non misure ordinali ma dati misurati con una scala d'intervalli o di rapporti**.

Sebbene non sia necessario il loro calcolo e si tratti di un test non parametrico, **il confronto e le ipotesi relative riguardano le medie** non le mediane.

Come sarà possibile comprendere facilmente dopo l'illustrazione della metodologia, **il limite fondamentale** del test di permutazione per 2 campioni indipendenti è la **possibilità pratica della sua applicazione manuale solo a campioni di dimensioni ridotte** (non oltre 8-10 dati per gruppo).

I calcolatori ora permettono di elevare facilmente tale limite, ma limitatamente a due, al massimo a tre decine di dati.

Sono gli stessi pregi e svantaggi dell'analogo test per 2 campioni dipendenti: **i metodi sono in parte differenti**, ma la logica è identica.

La procedura può essere illustrata con semplicità e chiarezza, mediante lo svolgimento di un esempio.

Si supponga di avere raccolto 2 gruppi di dati, con 4 osservazioni ( $n_1 = 4$ ) nel primo campione e 6 ( $n_2 = 6$ ) nel secondo, al fine di verificare l'ipotesi che la media del secondo gruppo è significativamente maggiore di quella del primo.

Gruppo 1	Gruppo 2
30	28
9	33
34	41
2	36
	15
	40

E' un test ad una coda con ipotesi nulla

$$H_0: m_1 = m_2$$

ed ipotesi alternativa

$$H_1: m_1 < m_2$$

**I fondamenti logici e i passaggi metodologici del test** possono essere riassunti in 5 punti:

1 - Il presupposto fondamentale è che **i punteggi osservati rappresentano, di fatto, il risultato dell'esperimento**; di conseguenza, essi **mantengono sempre costante il loro valore**.

Ma se l'estrazione dalla popolazione fosse fatta a caso, **ogni dato osservato potrebbe fare parte indifferente di un gruppo oppure dell'altro, con probabilità dipendente dal diverso numero di osservazioni**.

Quando  **$H_0$  è vera**, i punteggi grandi e quelli piccoli osservati tendono ad essere distribuiti in modo casuale, tra gruppo 1 e gruppo 2;

quando  **$H_0$  è falsa ed il test è ad una coda** (come nell'esempio), i punteggi minori (e simmetricamente quelli maggiori) tendono a concentrarsi in uno dei due gruppi, individuato a priori;

se  **$H_0$  fosse falsa e il test fosse stato a due code**, i valori minori (e simmetricamente quelli maggiori) tenderebbero a concentrarsi in uno dei due gruppi diversi, non individuabile a priori.

2 - **Calcolare il numero di possibili diverse risposte dell'esperimento che sarebbe possibile ottenere**, nella condizione che i dati possano cadere indifferente in uno dei due gruppi (come implicito nell'ipotesi nulla), mantenendo costanti le loro dimensioni.

Tale numero è **dato dalle combinazioni**

$$C_{n_1+n_2}^{n_1} \text{ equivalente a } C_{n_1+n_2}^{n_2}$$

Nell'esempio, con  **$n_1 = 4$**  e  **$n_2 = 6$**

$$C_{4+6}^4 = \frac{10!}{4! \cdot 6!} = 210$$

le possibili risposte sono 210.

3 - Con il valore di  $\alpha$  prefissato alla probabilità 0.05,

**la differenza tra le due medie risulta significativa se le due serie di dati osservati rientrano tra le risposte più estreme, collocate nella zona di rifiuto.**

Nell'esempio, tale numero è 10 (con arrotondamento all'unità inferiore per non superare il limite di probabilità prefissata), ottenuto moltiplicando quello delle possibili risposte per la probabilità  $\alpha$  ( $210 \times 0.05$ ).

**4 - Individuare i risultati più estremi che sono collocati nella zona di rifiuto.**

Con i dati dell'esempio, intuitivamente il risultato più estremo nella direzione scelta è quello determinato dall'insieme formato

- dai 4 punteggi minori nel gruppo 1 e
- dai 6 punteggi maggiori nel gruppo 2.

La verifica può essere effettuata mediante la differenza tra le somme dei due gruppi, che ha il valore negativo massimo (colonna Diff. nella tabella seguente).

I risultati successivi sono quelli che riducono la differenza tra i due gruppi, fino ad invertire eventualmente il segno della differenza.

L'ultimo risultato più estremo nell'altra direzione ha i 4 punteggi maggiori nel gruppo 1 e i 6 punteggi minori nel gruppo 2.

Nella tabella successiva, in ordine di rango (colonna R) sono riportati i 10 risultati più estremi nella coda prevista dall'ipotesi  $H_1$  e i 5 più estremi nell'altra coda della distribuzione (che sarebbero utili se l'ipotesi alternativa fosse bilaterale).

R.	GRUPPO 1				Som. 1	GRUPPO 2						Som. 2	Diff.
1	2	9	15	28	54	30	33	34	36	40	41	214	-160
2	2	9	15	30	56	28	33	34	36	40	41	212	-156
3	2	9	15	33	59	28	30	34	36	40	41	209	-150
4	2	9	15	34	60	28	30	33	36	40	41	208	-148
5	2	9	15	36	62	28	30	33	34	40	41	206	-144
6	2	9	15	40	66	28	30	33	34	36	41	202	-136
7	2	9	15	41	67	28	30	33	34	36	40	201	-134
8	2	9	30	33	74	15	28	34	36	40	41	194	-120
<b>9</b>	<b>2</b>	<b>9</b>	<b>30</b>	<b>34</b>	<b>75</b>	<b>15</b>	<b>28</b>	<b>33</b>	<b>36</b>	<b>40</b>	<b>41</b>	193	-118
10	2	9	30	36	77	15	28	33	34	40	41	191	-114
...	....	....	....	....	....	....	...	....	....	....	....	....	...
206	30	34	40	41	145	2	9	15	28	33	36	123	+22
207	30	36	40	41	147	2	9	15	28	33	34	121	+26
208	33	34	40	41	148	2	9	15	28	30	36	120	+28
209	33	36	40	41	150	2	9	15	28	30	34	118	+32
210	34	36	40	41	151	2	9	15	28	30	33	117	+34

**5. Verificare se la distribuzione dei risultati dell'esperimento è compresa nella zona di rifiuto.**

Con i dati dell'esempio, la distribuzione osservata coincide con il nono risultato più estremo (in grassetto, nella tabella precedente).

Poiché si tratta di un test ad una coda e la serie osservata dei dati cade nella zona di rifiuto, si rifiuta l'ipotesi nulla: la media del primo gruppo è significativamente minore di quella del secondo.

Se, con gli stessi dati, si fosse trattato di un test bilaterale, la zona di rifiuto sarebbe stata formata dai 5 risultati più estremi in ognuna delle due code. La combinazione dei dati osservata non sarebbe stata compresa in nessuna delle due zone di rifiuto; di conseguenza, si sarebbe dovuto concludere che le medie dei due gruppi non sono significativamente differenti.

**Nel caso di grandi campioni**, il test non è manualmente applicabile. Per esempio, con due gruppi di 10 dati ( $n_1$  e  $n_2 = 10$ ) il numero di combinazioni possibili è

$$C_{10+10}^{10} = \frac{20!}{10! \cdot 10!} = 184.756$$

uguale a 184.756.

Anche con il solo calcolo delle risposte più estreme, collocate nella zona di rifiuto, e limitato alla probabilità  $\alpha = 0.05$ , si tratta di scrivere e controllare 9.237 serie diverse di 20 dati.

Sono operazioni che possono essere eseguite in tempi accettabili solo con l'uso del calcolatore; ma anche in questo caso è possibile raggiungere rapidamente i limiti operativi.

Già con 2 gruppi di 20 dati,

$$C_{20+20}^{20} = \frac{40!}{20! \cdot 20!} = 15.163.120.000.000$$

si richiederebbe di scrivere 15.163.120.000.000 possibili distribuzioni, tra le quali rintracciare quella ottenuta dall'esperimento. Anche l'artificio di analizzare solo il 5% delle possibili risposte, limitando la verifica alla sola zona di rifiuto, richiede sempre di confrontare la serie dei dati sperimentali con 758.155.900.000 (ottenuto da  $15.163.120.000.000 \times 0.05$ ) serie diverse.

Nel **caso di grandi campioni**, si può ricorrere al **test t per 2 campioni indipendenti**, in quanto diviene possibile dimostrare l'esistenza delle condizioni di validità per un test parametrico, eventualmente mediante la trasformazione dei dati.

Se le condizioni di validità non fossero rispettate e si richiedesse un test non parametrico, si può utilizzare il **test U di Mann-Whitney**.

**ESEMPIO.** Per valutare la qualità delle acque da destinare sia ad uso potabile ed alimentare in genere, sia ad uso agricolo per irrigazione od abbeveraggio di bestiame, occorre quantificare la concentrazione di molti elementi o composti, come fluoruri, cloruri, cianuri, boro, ferro, manganese, mercurio, piombo, rame, zinco (quasi sempre misurati in mg/l). Quando sono superati i valori guida o limite, stabiliti per legge per ogni uso specifico, si definisce l'acqua non idonea od inquinata.

In due corpi idrici (individuati come A e B) sono stati estratti 4 campioni ed è stata misurata la presenza di un elemento o composto, per verificare se la località A abbia un'acqua con valori medi più bassi e quindi sia da preferire. I risultati dell'analisi chimica dell'elemento X è (in mg/l)

A	B
1,2	2,7
1,9	2,0
1,6	1,7
1,5	6,1

Si può affermare che la media dei quattro campioni della località A sia significativamente minore di quella dei quattro campioni della località B, per l'elemento o composto analizzato?

Risposta.

I dati campionari raccolti sono **misurati con una scala di rapporti**; tuttavia è da ritenere **non corretto l'uso di un test parametrico**, come il test t di Student per 2 campioni indipendenti, data la maggiore variabilità dei dati del gruppo B (da un minimo di 1,7 a un massimo di 6,1) rispetto a quelli di A (da 1,2 a 1,9).

**Con poche osservazioni**, anche se il test F per valutare la significatività del rapporto tra varianza maggiore e varianza minore non risultasse significativo, è **sempre discutibile affermare che è dimostrata la omoschedasticità delle due diverse distribuzioni**. Con campioni così piccoli, è **ugualmente impossibile dimostrare la normalità delle due distribuzioni**.

**Valori misurati con una scala di rapporti e distribuzione dei dati di forma ignota sono le condizioni necessarie e sufficienti per utilizzare in modo appropriato il test di casualizzazione.**

Dopo aver definito l'ipotesi nulla

$$H_0: m_A = m_B$$

e l'ipotesi alternativa unilaterale

$$H_1: m_A < m_B$$

per semplificare i confronti è utile

**riportare i dati di ogni gruppo in ordine crescente.**

LOCALITA' A				LOCALITA' B			
1,2	1,5	1,6	1,9	1,7	2,0	2,7	6,1

Nel presupposto che gli 8 valori osservati restino costanti, ma possano fare parte indifferentemente del gruppo A o di quello B, si deve **stimare il numero di combinazioni**: quante diverse distribuzioni

degli 8 dati è possibile formare, in modo che le serie dei dati del gruppo A o di quello B differiscano per almeno un elemento.

Con il calcolo combinatorio

$$C_{4+4}^4 = \frac{8!}{4!4!} = 70$$

risulta che il numero delle combinazioni di 8 elementi 4 a 4 è uguale a 70.

Dopo aver prefissato  $\alpha$  alla probabilità 0.05 si deve **calcolare quante sono le risposte che cadono nella zona di rifiuto.**

Con 70 possibili combinazioni alla probabilità  $\alpha = 0.05$

$$70 \times 0.05 = 3,5$$

sono 3, poiché il numero di risposte che cadono nella zona di rifiuto deve sempre essere arrotondato all'unità inferiore.

**E' lecito abbassare la probabilità, mentre è ritenuto errato alzarla:** la prima è una scelta conservativa, mentre la seconda permette di rifiutare più facilmente l'ipotesi nulla quando essa è vera determinando un errore di secondo tipo.

Nell'esempio, si tratta di un test ad una coda; di conseguenza, le 3 risposte cadono tutte in un estremo.

Nel passaggio successivo, occorre **individuare quali sono le risposte che cadono nella zona di rifiuto.** E' utile iniziare dalla risposta più estrema, nella direzione espressa dall'ipotesi alternativa  $H_1$ .

Trattandosi di un test ad una coda, in cui l'ipotesi alternativa è che la media del gruppo A sia inferiore a quella del gruppo B, la risposta più estrema nella direzione stabilita dalla domanda è quella in cui nel gruppo A si hanno i 4 valori minori ed in B i 4 valori maggiori.

Distribuzione teorica n. 1

GRUPPO A				Somma A	GRUPPO B				Somma B	SommaB - SommaA
1,2	1,5	1,6	1,7	6,0	1,9	2,0	2,7	6,1	12,7	6,7

Una verifica semplice è data dalla somma dei dati del gruppo A, che deve risultare la minore possibile, e dalla somma dei dati di B, che deve risultare la maggiore possibile. La loro differenza è massima.

La seconda possibile risposta più estrema nella stessa direzione è ottenuta spostando nel gruppo B il valore maggiore del gruppo A (1,9) e nel gruppo A il valore minore del gruppo B (1,7) della distribuzione precedente.

Distribuzione teorica n. 2

GRUPPO A				Somma A	GRUPPO B				Somma B	SommaB - SommaA
1,2	1,5	1,6	1,9	6,2	1,7	2,0	2,7	6,1	12,5	6,3

Come verifica si può osservare che la somma di A sarà maggiore della precedente, ma minore di tutte le altre possibili; mentre la somma di B sarà minore della precedente, ma maggiore di tutte le altre possibili.

La differenza tra le due somme risulterà minore della precedente, ma maggiore di tutte le altre possibili.

La terza possibile risposta è ottenuta spostando il secondo valore più alto del gruppo A in B (1,9) ed il valore minore del gruppo B in A (1,7).

Distribuzione teorica n. 3

GRUPPO A				Somma A	GRUPPO B				Somma B	SommaB - SommaA
1,2	1,5	1,6	2,0	6,3	1,7	1,9	2,7	6,1	12,4	6,1

Aumenta la somma di A, diminuisce quella di B e si riduce la loro differenza di una posizione, rispetto ad una classificazione ordinale.

**L'ultimo passaggio richiede il confronto tra la distribuzione osservata o sperimentale e le distribuzioni teoriche più estreme costruite, per verificare se esse la comprendono.** E' possibile osservare che la distribuzione sperimentale coincide con la distribuzione teorica n.2.

In conclusione, con i dati dell'esempio, la distribuzione osservata cade nella zona di rifiuto: si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa.

**E' possibile calcolare con precisione la probabilità di trovare per caso la distribuzione osservata, nella condizione che l'ipotesi nulla  $H_0$  sia vera.**

Le risposte possibili sono 70 e la nostra risposta sperimentale è la seconda nell'ordine delle distribuzioni teoriche possibili, considerando solo la coda definita dall'ipotesi unilaterale  $H_1$ . Pertanto, la probabilità di trovare per caso risposte uguali o più estreme di quella sperimentalmente osservata è

$$P = 2/70 = 0,0286 \text{ o } 2,86\%$$

Se il test fosse bilaterale, occorre considerare anche l'altra coda della distribuzione. La probabilità calcolata ( $P = 2,86\%$ ) deve essere moltiplicata per 2 ( $P = 5,72\%$ ).

## **7.12. IL TEST DELLE SUCCESSIONI PER DUE CAMPIONI O TEST DI WALD-WOLFOWITZ**

Il test della mediana, il test di Wilcoxon-Mann-Whitney, il test di Mann-Whitney per l'ordine robusto dei ranghi ed il test di casualizzazione servono per verificare se tra due gruppi di osservazioni indipendenti esistono differenze significative nella tendenza centrale (nella mediana per i primi 3, nella media per l'ultimo). Sono **test specifici per un solo parametro**, come altri proposti per il confronto tra la variabilità di due gruppi, tra cui il test di Moses che verrà discusso successivamente.

**Esistono anche test generalisti**, come quello di Kolmogorov-Smirnov già presentato. Tra essi è da classificare **il test delle successioni per 2 campioni indipendenti o test di Wald-Wolfowitz (Wald-Wolfowitz number of runs test)**.

Proposto nel 1940 da **A. Wald** e **J. Wolfowitz** (con l'articolo *On a test whether two samples are from the same population* sulla rivista *Ann. of Math. Statist.* n. 11 pp. 147-162), secondo alcuni autori di testi di statistica trova le sue origini già in un lavoro di K. Pearson del 1897 (*The chances of death and other studies in evolution*, London vol. I. cap. 2). Il test di Wald-Wolfowitz è **utile per saggiare simultaneamente qualunque differenza tra i 4 parametri di due distribuzioni campionarie: la tendenza centrale, la dispersione, la simmetria, la curtosi**.

**Nel test delle successioni di Wald-Wolfowitz l'ipotesi nulla è che due serie campionarie di dati siano statisticamente uguali per tutti i parametri, in quanto estratte in modo casuale dalla**

**stessa popolazione.** L'ipotesi alternativa è che tra le due distribuzioni a confronto esista una differenza significativa in almeno un parametro, senza individuarne uno in modo specifico.

**Il test è particolarmente utile quando si tratta di verificare l'ipotesi che due campioni possano essere stati estratti dalla medesima popolazione.** In ecologia e nella ricerca ambientale, è il caso di misure per qualsiasi variabile biologica o genetica tratta da animali o vegetali campionati in due diverse comunità o località: se appartengono alla stessa popolazione, con scambi genetici, i due gruppi di animali o vegetali a confronto non dovrebbero differire per alcun parametro, né per la somma dei loro effetti; se invece fossero ormai da considerare due popolazioni differenti, e quindi senza scambi genetici, dovrebbero differire nella media o nella varianza, nella simmetria o in quote significative e sistematiche di tali parametri.

**Trattandosi di un test generalista, per valutare la significatività della differenza in un solo parametro tra due serie di osservazioni campionarie, la potenza del test delle successioni per 2 campioni indipendenti è minore di quella dei test specifici.** Ma appunto perché considera anche fattori differenti da quelli analizzati da un test specifico, il confronto sulla potenza di un test generalista con un test specifico sul fattore considerato da quest'ultimo è ritenuto privo di senso da molti statistici.

**Nonostante le differenze nell'ipotesi, la metodologia è uguale a quella del test delle successioni in un solo campione per eventi alternativi, quando l'ipotesi è unilaterale.**

Come per i test precedenti, la metodologia ed i concetti di base possono essere presentati in modo semplice mediante un esempio.

Si supponga di voler verificare se esistono differenze significative nella distribuzione di due gruppi campionari (A e B sottoriportati) formati rispettivamente da 8 e da 9 osservazioni, per verificare se appartengono a due popolazioni differenti.

Gruppo A	8	12	14	25	36	37	39	65	
Gruppo B	28	30	31	34	48	49	60	67	69

**A differenza del test delle successioni per un solo campione, il test di Wald-Wolfowitz considera solo il caso di un numero di successioni inferiore all'atteso: se i due campioni a confronto appartengono a due popolazioni differenti, il numero di successioni tende ad essere minimo.**

Le operazioni richieste dalla metodologia possono essere riassunte in alcuni passaggi fondamentali.

1 - **Ordinare le due serie di dati campionari in ordine crescente**, secondo il loro valore algebrico, **conservando l'informazione del gruppo di appartenenza** (nel test per un campione, le risposte di tipo binario erano ordinate secondo il tempo d'osservazione o rilevazione).

8	12	14	25	28	30	31	34	36	37	39	48	49	60	65	67	69
A	A	A	A	B	B	B	B	A	A	A	B	B	B	A	B	B

2 – **Definire**

- $n_1$  la numerosità del campione con il numero maggiore di osservazioni,
- $n_2$  la numerosità del campione di dimensioni minori.

**Contare il numero di osservazioni dei due gruppi e il numero di successioni (R)**

8	12	14	25	28	30	31	34	36	37	39	48	49	60	65	67	69
A	A	A	A	B	B	B	B	A	A	A	B	B	B	A	B	B
1				2				3			4		5	6		

Con i dati dell'esempio,

per il gruppo A  $n_1 = 8$ ,

per il gruppo B  $n_2 = 9$ ,

numero di successioni  $R = 6$ .

E' importante ricordare **quale dovrebbe essere il numero di successioni, nel caso in cui le due distribuzioni differissero nella tendenza centrale, nella variabilità e/o nella forma** (simmetria o curtosi) della distribuzione; in altri termini, **se l'ipotesi nulla di appartenenza alla stessa popolazione fosse falsa**.

Quando due gruppi hanno due valori significativamente differenti della **tendenza centrale** ( $H_0$  falsa), nella sequenza ordinata dei dati i valori di un gruppo precedono quelli dell'altro gruppo. All'opposto, quando le due tendenze centrali sono uguali ( $H_0$  vera) i dati dei due gruppi sono casualmente mescolati.

Nel primo caso, **con  $H_0$  falsa, il numero delle successioni è ridotto**; nella situazione più estrema, quando la differenza tra le due tendenze centrali è notevole, il numero di successioni tende ad essere solamente 2.

Nell'altro caso, **quando  $H_0$  è vera, il numero di successioni tende ad un valore medio, dipendente da  $n_1$  e  $n_2$ .**

Quando i due gruppi hanno una **variabilità o dispersione** differente (con tendenza centrale uguale), i dati del gruppo con maggiore variabilità tendono ad essere più frequenti dell'altro nelle due code della distribuzione. All'opposto, se i due gruppi avessero la stessa variabilità i dati dei gruppi sarebbero casualmente mescolati.

Nel primo caso **con  $H_0$  falsa, si hanno poche successioni**; nel secondo caso **con  $H_0$  vera, il numero di successioni tende ad un valore medio, dipendente da  $n_1$  e  $n_2$ .**

Effetti simili provoca una **differenza nella simmetria e nella curtosi**: **quando l'ipotesi nulla  $H_0$  è falsa, si hanno poche successioni**; **quando l'ipotesi nulla  $H_0$  è vera, il numero di successioni tende ad un valore medio che dipende dal numero di osservazioni nei due gruppi.**

**3 - Nel caso in cui i due campioni appartengono alla stessa popolazione e quindi l'ipotesi nulla su tendenza centrale, variabilità, simmetria e curtosi è vera, il numero medio atteso di successioni  $m_R$  è uguale a**

$$m_R = \frac{2 \times n_1 \times n_2}{n_1 + n_2} + 1$$

Con i dati dell'esempio:

$$\mu_R = \frac{2 \cdot 8 \cdot 9}{8 + 9} + 1 = 9,47$$

il numero di successioni dovrebbe essere 9,47 (ovviamente arrotondato all'unità).

**4 - L'ipotesi nulla  $H_0$  viene verificata mediante la significatività della differenza tra il numero di successioni osservato  $R$  ed il numero atteso  $\mu_R$ .**

Nell'esempio, si deve verificare se  **$R = 6$**  è significativamente diverso da  **$m_R = 9,47$** .

**TAVOLA DEI VALORI CRITICI NEL TEST DELLE SUCCESSIONI  
PER DUE CAMPIONI INDIPENDENTI DI WALD-WOLFOWITZ  
ALLA PROBABILITA' 0.05**

Le tabelle riportano i valori minimi significativi.

E' significativo ogni numero R di successioni minore od uguale a quello riportato nella tabella.

Con  $n_1$  si indica il campione maggiore, con  $n_2$  il campione minore.

$$\alpha = 0.05$$

		$N_2$																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$n_1$																				
4				2																
5			2	2	3															
6			2	3	3	3														
7			2	3	3	4	4													
8		2	2	3	3	4	4	5												
9		2	2	3	4	4	5	5	6											
10		2	3	3	4	5	5	6	6	6										
11		2	3	3	4	5	5	6	6	7	7									
12		2	3	4	4	5	6	6	7	7	8	8								
13		2	3	4	4	5	6	6	7	8	8	9	9							
14		2	3	4	5	5	6	7	7	8	8	9	9	10						
15		2	3	4	5	6	6	7	8	8	9	9	10	10	11					
16		2	3	4	5	6	6	7	8	8	9	10	10	11	11	11				
17		2	3	4	5	6	7	7	8	9	9	10	10	11	11	12	12			
18		2	3	4	5	6	7	8	8	9	10	10	11	11	12	12	13	13		
19		2	3	4	5	6	7	8	8	9	10	10	11	12	12	13	13	14	14	
20		2	3	4	5	6	7	8	9	9	10	11	11	12	12	13	13	14	14	15

**TAVOLA DEI VALORI CRITICI NEL TEST DELLE SUCCESSIONI  
PER DUE CAMPIONI INDIPENDENTI DI WALD-WOLFOWITZ  
ALLA PROBABILITA' 0.01**

Le tabelle riportano i valori minimi significativi.

E' significativo ogni numero R di successioni minore od uguale a quello riportato nella tabella.

Con  $n_1$  si indica il campione maggiore, con  $n_2$  il campione minore.

$$\alpha = 0.01$$

		$N_2$																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$n_1$																				
4																				
5																				
6																				
7																				
8																				
9																				
10																				
11																				
12																				
13																				
14																				
15																				
16																				
17																				
18																				
19	2	2	3	4	5	6	6	7	8	8	9	9	10	10	11	11	12	12		
20	2	2	3	4	5	6	6	7	8	8	9	10	10	11	11	11	12	12	13	

Nel caso di **piccoli campioni**, si ricorre alla tabella dei valori critici, che forniscono il numero massimo di osservazioni nel caso di differenza significativa. Si utilizzano le stesse tabelle dei valori critici del test delle successioni per un campione, nel caso di ipotesi unilaterali. **Quando il numero di successioni calcolato è più alto di quello riportato nella tabella, non si può rifiutare l'ipotesi nulla.**

Con i dati dell'esempio, alla probabilità  $\alpha = 0.05$  il numero massimo riportato nella tabella per  $n_1 = 9$  e  $n_2 = 8$  ( $n_1$  è il numero di osservazioni del campione maggiore) è **5**, mentre il numero osservato con i dati sperimentali è **6**.

Non si può rifiutare l'ipotesi nulla; non si è potuto dimostrare che le due distribuzioni di dati siano differenti per qualche caratteristica.

Nel caso di **grandi campioni**,  $R$  è distribuito in modo approssimativamente normale con

$$z = \frac{R - m_R}{s_R}$$

dove  $\mu_R$

$$m_R = \frac{2 \times n_1 \times n_2}{n_1 + n_2} + 1$$

e  $\sigma_R$

$$s_R = \sqrt{\frac{2 \cdot n_1 \cdot n_2 \cdot (2 \cdot n_1 \cdot n_2 - n_1 - n_2)}{(n_1 + n_2)^2 \cdot (n_1 + n_2 - 1)}}$$

Il **campione dell'esempio** è di piccole dimensioni; ma può essere utilizzato per mostrare la successione delle operazioni da eseguire nel caso di grandi campioni.

Con i dati già riportati, dove  $R = 6$  e  $m_R = 9,47$

dopo aver calcolato  $s_R$

$$s_R = \sqrt{\frac{2 \cdot 8 \cdot 9 \cdot (2 \cdot 8 \cdot 9 - 8 - 9)}{(8 + 9)^2 \cdot (8 + 9 - 1)}} = 1,988$$

che risulta uguale a 1,988,

si stima il valore di  $z$

$$z = \frac{6 - 9,47}{1,988} = -1,74$$

che risulta uguale a -1,74.

Nella **tavola della distribuzione normale**, in una coda della distribuzione al valore di 1,74 corrisponde una probabilità uguale a 0.0409 che permetterebbe di rifiutare l'ipotesi nulla alla probabilità  $\alpha = 0.05$ .

Quando i dati dei due campioni a confronto sono pochi, con la distribuzione normale si sottostima la probabilità. E' un errore che può essere corretto, quando il campione non è troppo piccolo.

Per **un numero di osservazioni non grande**, si deve apportare **il termine di correzione per la continuità**, che riduce il valore della differenza di 0,5 lasciando inalterata la varianza; abbassa quindi anche la significatività del test.

Il valore di **z** diviene

$$z = \frac{|R - m_R| - 0,5}{S_R}$$

Con i dati dell'esempio

$$z = \frac{|6 - 9,47| - 0,5}{1,988} = \frac{2,97}{1,988} = 1,49$$

si ottiene un valore di **z** uguale a 1,49.

Ad esso, in una coda della distribuzione normale, corrisponde una probabilità uguale  $\alpha = 0.0681$ ; è un valore sufficientemente alto, che non permette di rifiutare l'ipotesi nulla alla probabilità 0.05.

**La correzione per la continuità**, da applicare nel caso di campioni non grandi, **riduce la significatività della differenza tra successioni osservate ed attese; quindi abbassa la probabilità di rifiutare l'ipotesi nulla**.

I **ties**, le osservazioni ex-aequo, non dovrebbero esistere nel test delle successioni per due campioni indipendenti; infatti, per applicare il test delle successioni, i punteggi dovrebbero essere misurati in una scala continua. La presenza di **ties** tra i due gruppi altera il conteggio delle successioni, ponendo situazioni di difficile soluzioni.

Il test delle successioni per due campioni indipendenti saggia l'ipotesi nulla  $H_0$  di estrazione casuale dei due campioni indipendenti dalla stessa popolazione, contro ipotesi alternative multiple di differenze significative in almeno uno dei parametri delle due distribuzioni.

**Il rifiuto dell'ipotesi nulla spesso richiede il ricorso ad altri test, per individuare la causa specifica o più importante della differenza riscontrata.**

ESEMPIO. Gli inquinanti organici ad effetto tossico formano una categoria estremamente ampia ed eterogenea; le molecole inquinanti o potenzialmente inquinanti sono centinaia di migliaia. Una classificazione elementare li suddivide in pesticidi, oli e idrocarburi; essa comprende anche la voce altri tossici organici, tra cui aromatici alogenati, aromatici volatili, diossine, cloroparaffine, clorofenoli, eteri difenilici alogenati, idrocarburi policiclici.

In due corpi idrici (A e B) è stata effettuata una serie di misure dell'inquinamento organico in condizioni simili.

GRUPPO A	34	12	36	31	43	16	15	10			
GRUPPO B	65	76	18	27	21	49	20	45	41	17	58

Si vuole verificare se l'origine o causa può essere comune.

Risposta.

Se tale ipotesi è vera, in termini statistici i parametri delle due distribuzioni campionarie non dovrebbero differire in modo significativo ( $H_0$  vera).

Se la causa è diversa ( $H_0$  falsa), i due gruppi di dati dovrebbero differire per almeno un parametro della distribuzione.

Per applicare il test di Wald-Wolfowitz per due campioni indipendenti, si devono ordinare i due gruppi in una serie unica, mantenendo l'informazione del gruppo d'appartenenza.

10	12	15	16	17	18	20	21	27	31	34	36	41	43	45	49	58	65	76
A	A	A	A	B	B	B	B	B	A	A	A	B	A	B	B	B	B	B

Successivamente, si contano il numero di dati per campione ed il numero di successioni tra i due campioni a confronto

10	12	15	16	17	18	20	21	27	31	34	36	41	43	45	49	58	65	76
A	A	A	A	B	B	B	B	B	A	A	A	B	A	B	B	B	B	B
<b>1</b>				<b>2</b>				<b>3</b>			<b>4</b>	<b>5</b>	<b>6</b>					

Il numero di dati del campione maggiore è  $n_1 = 11$ ;

il numero di dati del campione minore è  $n_2 = 8$ ;

il numero di successioni tra i due campioni indipendenti è  $R = 6$ .

Sono due campioni di piccole dimensioni; per la stima della probabilità si deve ricorrere alla tabella dei valori critici.

Alla probabilità  $\alpha = 0.05$  per  $n_1 = 11$  e  $n_2 = 8$ , il valore critico riportato dalla tabella è **6**.

Il valore di  $R$  calcolato è uguale a quello tabulato. Di conseguenza, alla probabilità  $\alpha \leq 0.05$  si può rifiutare l'ipotesi nulla: le due serie di dati sono estratte da due popolazioni diverse.

Con probabilità uguale o inferiore a 0.05 di commettere un errore, si può sostenere che l'origine dell'inquinamento è differente.

### **7.13. TEST DI SIEGEL-TUKEY PER L'UGUAGLIANZA DELLA VARIANZA**

#### **CENNI DEL TEST DI FREUND-ANSARI-BRADLEY E DEL TEST DI CONOVER**

Nella ricerca ambientale, spesso l'attenzione del ricercatore non è rivolta alla tendenza centrale, ma alla **dispersione o variabilità dei dati di due serie indipendenti d'osservazioni**.

Nel confronto tra le caratteristiche fisiche o comportamentali di due gruppi d'animali, è atteso che quelli geneticamente identici presentino una minore variabilità del gruppo geneticamente eterogeneo. La variabilità di un campione di animali o vegetali tratto da un ambiente in condizioni normali è spesso inferiore a quella di un campione tratto da un ambiente sotto stress, dove alcuni individui subiscono limiti alla crescita con intensità differente. Anche la differenza nella variabilità è un indice dell'appartenenza a popolazioni differenti.

Nelle analisi di laboratorio, quando si confrontano due reagenti per valutare la qualità del prodotto, la scelta deve essere basata non sulle differenze nel valore medio, che dipende dalla concentrazione della sostanza analizzata, ma sulle differenze della varianza, calcolata su misure ripetute dello stesso campione. Sarà migliore il reagente che fornisce una variabilità minore; sarà più preciso, perché meno influenzato nelle sue risposte da fattori di perturbazione.

Se le due variabili casuali  $X_1$  e  $X_2$  sono distribuite normalmente, la verifica sulla variabilità coincide con quella effettuata con il test  $F$  di Fisher-Snedecor.

In campo non parametrico, le proposte sono state numerose. Tra i test più diffusi, sono da ricordare

- il **test di Siegel-Tukey** del 1960 (proposto con l'articolo *A nonparametric sum ranks procedure for relative spread in unpaired samples* pubblicato su *J. Amer. Statist. Ass.*, vol. 55, pp. 429-445),
- il **test di Freund-Ansari-Bradley** divulgato con un articolo del 1957 (J. E. Freund, A. R. Ansari, *Two-way rank-sum test for variances*, pubblicato su *Technical Report*, n. 34, Department of Statistics, Virginia Polytechnic Institute) ed uno successivo del 1960 (A. R. Ansari, R. A. Bradley, *Rank-sum test for dispersion*, pubblicato su *Ann. Math. Statist.*, vol. 31, pp. 1174-1189),

- il **test di Conover** del 1980 (presentato da W. J. Conover nel volume *Practical Nonparametric Statistics*, 2<sup>nd</sup> edition New York, John Wiley & Sons).

Sono **utili nel caso di campioni piccoli, ma richiedono la conoscenza della tendenza centrale (la mediana)** dei due campioni o almeno della differenza tra essi.

Non è vincolato a questo limite, non richiedendo la conoscenza delle due mediane, il test di Moses, che sarà presentato successivamente; ma esso ha una potenza inferiore a questi tre, nel caso di campioni con poche osservazioni; quando le osservazioni sono poche unità, addirittura non può essere applicato.

Il **test di Siegel-Tukey** e il **test di Freund-Ansari-Bradley** si fondano su concetti simili. Tuttavia, mentre il primo per la significatività utilizza il test U di Mann-Whitney, del quale è relativamente facile avere la tabella dei valori critici, il test di Freund-Ansari-Bradley richiede tabelle specifiche non facilmente reperibili. Viene quindi presentato, in modo dettagliato ed operativo per calcoli manuali, il test di Siegel-Tukey mentre del test di Freund-Ansari-Bradley e del test di Conover vengono illustrati solo i concetti di base, utili per capire i risultati di eventuali programmi informatici.

Il **test di Siegel-Tukey** si fonda su concetti e procedure semplici. L'idea di base è che se due campioni, estratti dalla stessa popolazione, hanno varianze diverse, quello con la varianza maggiore avrà una dispersione maggiore dei dati. Se i due gruppi hanno tendenze centrali diverse, è necessario ricondurli allo stesso valore, attraverso l'eliminazione della differenza tra le mediane.

Disponendo di due serie di dati, già ordinate in modo crescente,

Gruppo A	176	201	225	230	232	269	276	291				
Gruppo B	44	102	124	142	157	172	194	218	225	232	264	271

di cui il gruppo A con 8 osservazioni ed il gruppo B con 12 osservazioni, si supponga di voler verificare se hanno variabilità simile oppure significativamente differente.

La procedura richiede vari passaggi.

1 – Dopo aver **ordinato i dati in modo crescente entro ogni gruppo** (già fatto nella tabella di presentazione dei dati),

**calcolare le mediane** rispettive.

Nel gruppo A, composto da 8 dati, essa si colloca tra il 4° valore (230) e il 5° (232) e quindi è uguale a 231;

Gruppo A	176	201	225	<b>230</b>	<b>232</b>	269	276	291
	1	2	3	<b>4</b>	<b>5</b>	6	7	8

nel gruppo B, composto da 12 valori, si colloca tra il 6°(172) e il 7° (194) e quindi, calcolata come media di questi due valori, risulta uguale a 183.

Gruppo B	44	102	124	142	157	<b>172</b>	<b>194</b>	218	225	232	264	271
	1	2	3	4	5	<b>6</b>	<b>7</b>	8	9	10	11	12

**Le due distribuzioni non hanno la stessa mediana;** tra esse esiste una differenza

$$231 - 183 = 48$$

uguale a 48.

2 - Occorre **riconducere le due serie di dati ad una situazione in cui abbiano la stessa mediana.**

In questo caso, il metodo più semplice è togliere 48 agli 8 valori del gruppo A.

Gruppo A	128	153	177	182	184	221	228	243
----------	-----	-----	-----	-----	-----	-----	-----	-----

Di conseguenza, la nuova serie di dati, con il solo gruppo A trasformato e quindi con mediana (183) uguale a quella del gruppo B, diviene

Gruppo A	128	153	177	182	184	221	228	243				
Gruppo B	44	102	124	142	157	172	194	218	225	232	264	271

Se **non hanno la stessa varianza ( $H_0$  falsa)**, quello con variabilità maggiore dovrebbe avere contemporaneamente sia valori minori sia valori maggiori in numero significativo; cioè una **quota maggiore di valori collocati agli estremi;**

mentre, se **hanno la stessa varianza ( $H_0$  vera)**, i due gruppi dovrebbero avere la stessa quota di valori grandi e piccoli.

3 – Per valutare la significatività di una diversa presenza di valori collocati ai due estremi della distribuzione, si **costruisce una distribuzione unica ordinata in modo crescente, mantenendo l'informazione del gruppo d'appartenenza.**

44	10	12	12	14	15	15	17	17	18	18	19	21	22	22	22	23	24	26	27
	2	4	8	2	3	7	2	7	2	4	4	8	1	5	8	2	3	4	1
B	B	B	A	B	A	B	B	A	A	A	B	B	A	B	A	B	A	B	B

Successivamente si attribuiscono i ranghi, dando il punteggio minore a quelli collocati agli estremi e maggiore a quelli collocati verso il centro, alternando il punto di partenza nei due estremi.

4 – In modo più dettagliato si attribuisce,

rango 1 al valore minore e rango 2 al valore maggiore (o ultimo),

rango 3 al penultimo valore e rango 4 al secondo valore,

rango 5 al terzo valore e rango 6 al terzultimo valore,

rango 7 al quartultimo valore e rango 8 al quarto,

fino ai ranghi massimi per i valori collocati al centro,

come nella tabella sottostante

44	10	12	12	14	15	15	17	17	18	18	19	21	22	22	22	23	24	26	27
	2	4	8	2	3	7	2	7	2	4	4	8	1	5	8	2	3	4	1
B	B	B	A	B	A	B	B	A	A	A	B	B	A	B	A	B	A	B	B
<b>1</b>	<b>4</b>	<b>5</b>	<b>8</b>	<b>9</b>	<b>12</b>	<b>13</b>	<b>16</b>	<b>17</b>	<b>20</b>	<b>19</b>	<b>18</b>	<b>15</b>	<b>14</b>	<b>11</b>	<b>10</b>	<b>7</b>	<b>6</b>	<b>3</b>	<b>2</b>

Su questa serie di ranghi, in cui i punteggi minori sono attribuiti ai valori collocati agli estremi e i punteggi maggiori ai valori centrali, si applica il **test U di Mann-Whitney**.

5 – A questo scopo, anche se esistono modalità differenti, più rapide ma meno intuitive, per calcolare le precedenze si ordinano i ranghi dal minore al maggiore, mantenendo l'informazione di gruppo

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
<b>B</b>	<b>B</b>	<b>B</b>	<b>B</b>	<b>B</b>	<b>A</b>	<b>B</b>	<b>B</b>	<b>A</b>	<b>B</b>	<b>A</b>	<b>A</b>								

Dalla lettura della tabella appare evidente che i ranghi minori sono occupati dal gruppo **B** e i maggiori dal gruppo **A**.

Per ottenere U, con questa serie di ranghi è semplice calcolare quante volte i dati del gruppo B sono preceduti da dati del gruppo A.

6 - Si ottiene un conteggio, riportato nella tabella sottostante per ogni valore di B,

B	B	B	B	B	A	B	A	B	A	B	A	B	A	B	B	A	B	A	A
0	0	0	0	0	--	1	--	2	--	3	--	4	--	5	5	--	6	--	--

che in totale determina

$$U = 0 + 0 + 0 + 0 + 0 + 1 + 2 + 3 + 4 + 5 + 5 + 6 = 26$$

un valore di U uguale a 26.

7 - Come verifica di aver calcolato il valore corretto di U, si può stimare U', determinato dal numero di precedenze di B rispetto ad A e

B	B	B	B	B	A	B	A	B	A	B	A	B	A	B	B	A	B	A	A
--	--	--	--	--	5	--	6	--	7	--	8	--	9	--	--	11	--	12	12

il cui totale risulta

$$U' = 5 + 6 + 7 + 8 + 9 + 11 + 12 + 12 = 70$$

uguale a 70,

ricordando che

$$U + U' = n_1 \times n_2 = 26 + 70 = 8 \times 12 = 96$$

Per **piccoli campioni** si utilizza la tabella dei valori critici della distribuzione U di Man-Whitney, che può essere sia a due code con

$$H_1 : S_A^2 \neq S_B^2$$

sia a una coda, quando sia prefissato quale gruppo dovrebbe avere variabilità maggiore.

Con i dati dell'esempio, alla probabilità  $\alpha = 0.05$  e con  $n_1 = 8$  e  $n_2 = 12$

per un test bilaterale, nella tabella il valore critico di U è uguale a 22; di conseguenza, non è possibile rifiutare l'ipotesi nulla, poiché il valore di U calcolato è maggiore (26).

Per **grandi campioni**, è preferibile **utilizzare il test di Moses**, che ha condizioni di validità meno rigide. Volendo utilizzare ugualmente il **test di Siegel-Tukey** proposto, per la significatività di U vedere l'approssimazione alla normalità del valore di U

$$z = \frac{U - \mu_U}{S_U}$$

illustrato dettagliatamente nel paragrafo relativo.

Il **test di Freund-Ansari-Bradley** permette di verificare la stessa ipotesi del test precedente e nella prima parte (fino al punto 3) ha una metodologia identica.

1 - Utilizzando gli stessi dati di partenza,

Gruppo A	176	201	225	230	232	269	276	291				
Gruppo B	44	102	124	142	157	172	194	218	225	232	264	271

richiede il calcolo delle due mediane (231 e 183, con differenza uguale a 48)

2 – e la sottrazione della differenza, in modo che i due gruppi abbiano la stessa mediana

Gruppo A	128	153	177	182	184	221	228	243				
Gruppo B	44	102	124	142	157	172	194	218	225	232	264	271

3 – fino alla costruzione di una sola serie ordinata, mantenendo l'informazione del gruppo di appartenenza

44	10	12	12	14	15	15	17	17	18	18	19	21	22	22	22	23	24	26	27
	2	4	8	2	3	7	2	7	2	4	4	8	1	5	8	2	3	4	1
B	B	B	A	B	A	B	B	A	A	A	B	B	A	B	A	B	A	B	B

4 – Come indicatore della diversa distribuzione dei dati di un gruppo agli estremi, con il **test di Freund-Ansari-Bradley** si attribuiscono ranghi minori ai due estremi e progressivamente maggiori avvicinandosi al centro :

- il rango 1 è assegnato sia al primo valore, sia all'ultimo,
- il rango 2 è assegnato sia al secondo valore, sia al penultimo,
- il rango 3 sia al terzo valore, sia al terzultimo,

ottenendo la seguente serie di ranghi

44	10	12	12	14	15	15	17	17	18	18	19	21	22	22	22	23	24	26	27
	2	4	8	2	3	7	2	7	2	4	4	8	1	5	8	2	3	4	1
B	B	B	A	B	A	B	B	A	A	A	B	B	A	B	A	B	A	B	B
1	2	3	4	5	6	7	8	9	10	10	9	8	7	6	5	4	3	2	1

5 - Come indicatore della differente variabilità dei dati dei due gruppi, si prende la somma dei loro ranghi,

$$\text{Somma dei Ranghi del Gruppo A} = 4 + 6 + 9 + 10 + 10 + 7 + 5 + 3 = 54$$

$$\text{Somma dei Ranghi del Gruppo B} = 1 + 2 + 3 + 5 + 7 + 8 + 9 + 8 + 6 + 4 + 2 + 1 = 56$$

di cui gli autori hanno stimato la tabella dei valori critici, in rapporto alla probabilità  $\alpha$  e alle dimensioni  $n_1$  e  $n_2$  dei due campioni a confronto, sia **per test unilaterali che bilaterali**.

Gli autori hanno stimato anche l'efficienza asintotica relativa del loro test rispetto al test parametrico F di Fisher-Snedecor (il rapporto tra la varianza maggiore e quella minore), che

- con una distribuzione normale dei dati è uguale a circa 0,61 ( $6/\pi^2$ ),

- con una distribuzione rettangolare dei dati è uguale a 0,60,

- con una distribuzione esponenziale doppia è uguale a 0,94.

Il test non parametrico ha sempre una potenza inferiore a quello parametrico.

Il **test di Conover** per l'uguaglianza delle varianze è fondato sulle devianze dei due gruppi, ottenute calcolando per ogni valore

$$(X_{ij} - \bar{X}_i)^2$$

il **quadrato dello scarto rispetto alla media** del suo gruppo.

Successivamente,

si sostituiscono le devianze ai valori e

si attribuiscono ad essi i ranghi come nel test di Wilcoxon- Mann- Whitney.

Di tale test si utilizza la tabella dei valori critici nel caso di piccoli campioni e l'approssimazione alla normale nel caso di campioni grandi.

Al posto della devianza, alcuni testi suggeriscono l'uso dello **scarto assoluto**

$$|X_{ij} - \bar{X}_i|$$

che non varia l'ordine dei ranghi.

**Il test di Conover appare meno potente del test di Siegel -Tukey.**

#### 7.14. IL TEST DEI RANGHI EQUIVALENTI DI MOSES PER LE DIFFERENZE NELLA DISPERSIONE O VARIABILITÀ

Tra i vari test proposti per misurare la dispersione o variabilità di due serie di osservazioni, il **test di Moses (Moses run like test)** proposto nel 1963 e nel 1964 (L. E. Moses *Rank test of dispersion* in *Ann. Math. Statist.* n. 34, pp. 973-983; *One sample limits of some two-sample rank test* in *J. Amer. Statist. Ass.* vol. 59, pp. 645-651) è **quello di più generale applicazione: non richiede la conoscenza delle mediane** delle due popolazioni; **ma solo l'uso di una scala d'intervallo o di rapporti**, poiché utilizza misure della devianza (è possibile utilizzare anche la varianza, ma si richiede un'operazione aggiuntiva).

**Le ipotesi da verificare riguardano la varianza dei due campioni.**

L'ipotesi nulla  $H_0$  prevede sempre che le varianze di due campioni A e B siano uguali

$$H_0 : \mathbf{s}_A^2 = \mathbf{s}_B^2$$

mentre l'ipotesi alternativa  $H_1$  dipende dalla domanda posta nel problema:

**può essere sia bilaterale**

$$H_1 : \mathbf{s}_A^2 \neq \mathbf{s}_B^2$$

**sia unilaterale**

$$H_1 : \mathbf{s}_A^2 > \mathbf{s}_B^2 \quad \text{oppure} \quad H_0 : \mathbf{s}_A^2 < \mathbf{s}_B^2$$

in una delle due forme previste.

Lo sviluppo di un esempio permette una illustrazione didattica più chiara e semplice di un elenco di passaggi logici ed operativi.

Si supponga di avere raccolto due serie di misure, chiamate gruppo A e gruppo B

GRUPPO A	8,4	7,6	9,2	8,2	9,7	6,4	5,6	8,1			
GRUPPO B	5,4	3,7	6,8	5,3	9,1	4,1	8,6	8,2	3,9	4,2	6,7
Posizione	1	2	3	4	5	6	7	8	9	10	11

per verificare se esiste una differenza significativa tra la variabilità o dispersione dei due gruppi.

La metodologia richiede alcuni passaggi per l'elaborazione dei dati sperimentali raccolti.

**1 - Per ognuno dei due gruppi a confronto, mediante estrazione casuale, formare sottogruppi o sotto-insiemi con almeno 2 osservazioni.**

Questi sottoinsiemi devono essere composti riunendo casualmente (appunto, per estrazione di numeri random) le osservazioni presenti entro ogni gruppo.

Le osservazioni che non formano un sottoinsieme completo devono essere eliminate.

**Ai fini dei confronti successivi, è utile che ogni gruppo abbia almeno 4-5 sottoinsiemi.**

Poiché il gruppo A è formato da soli 8 dati ed il gruppo B da 11 dati, è conveniente formare sottoinsiemi con due sole osservazioni.

Mediante estrazione casuale, con i dati dell'esempio **per il gruppo A** sono risultati abbinati i numeri che occupano le posizioni **1-5, 8-7, 6-4, 2-3**.

Con un numero pari di osservazioni e sottoinsiemi di 2 elementi, in questo gruppo non esistono numeri da scartare.

**Per il gruppo B** sono risultati abbinati i numeri che occupano le posizioni **6-1, 2-5, 11-9, 8-10, 4-7**.

Il numero che occupa la posizione 3 **viene escluso dalle analisi successive, in quanto parte di un sottoinsieme incompleto.**

GRUPPO	SOTTOINSIEMI DI 2 DATI	
Gruppo A	1)	8,4      9,7
	2)	8,1      5,6
	3)	6,4      8,2
	4)	7,6      9,2
Gruppo B	1)	4,1      5,4
	2)	3,7      9,1
	3)	6,7      3,9
	4)	8,2      4,2
	5)	5,3      8,6

**2 - Per ogni sottoinsieme calcolare la devianza (SQ)**

di cui si ricorda la formula generale

$$SQ = \sum (x - \bar{x})^2$$

e la formula abbreviata

$$SQ = \sum x^2 - \frac{(\sum x)^2}{n}$$

Con i dati dell'esempio, si ottengono 4 devianze per il gruppo A e 5 devianze per il gruppo B:

GRUPPO	SOTTOINSIEMI DI 2 DATI		DEVIANZE	
<b>Gruppo A</b>	1)	8,4	9,7	<b>0,845</b>
	2)	8,1	5,6	<b>3,125</b>
	3)	6,4	8,2	<b>1,620</b>
	4)	7,6	9,2	<b>1,280</b>
<b>Gruppo B</b>	1)	4,1	5,4	<b>0,770</b>
	2)	3,7	9,1	<b>0,080</b>
	3)	6,7	3,9	<b>3,920</b>
	4)	8,2	4,2	<b>8,000</b>
	5)	5,3	8,6	<b>5,445</b>

dei quali sono riportati i valori.

**3 - Alle devianze dei due gruppi applicare il test U di Wilcoxon-Mann-Whitney per due campioni indipendenti, secondo l'ipotesi  $H_1$  espressa.**

E' il test per due campioni indipendenti più adatto per verificare la significatività di differenze nelle tendenze centrali delle devianze calcolate.

Nell'esercizio, lo scopo è di verificare se questi indici di dispersione (le devianze) dei due gruppi hanno la medesima tendenza centrale o sono significativamente diversi. E' un test bilaterale.

Per applicare il test U di Mann-Whitney, fondato sulle precedenze, ordinare i valori dei due gruppi in ordine crescente come se fossero un insieme unico, mantenendo per ogni devianza l'informazione del gruppo di appartenenza.

DEVIANZE	0,080	0,770	0,845	1,280	1,620	3,125	3,920	5,445	8,000
GRUPPO	B	B	A	A	A	A	B	B	B

Con i dati dell'esempio, si hanno 9 valori di devianze nell'ordine riportato.

#### 4 - Contare il numero di precedenze di ogni gruppo

DEVIANZE	0,080	0,770	0,845	1,280	1,620	3,125	3,920	5,445	8,000
GRUPPO	B	B	A	A	A	A	B	B	B
<b>Precedenze del gruppo 1</b>			<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>			

Per il gruppo A le precedenze sono in totale 8,  
mentre per il gruppo B

DEVIANZE	0,080	0,770	0,845	1,280	1,620	3,125	3,920	5,445	8,000
GRUPPO	B	B	A	A	A	A	B	B	B
<b>Precedenze del gruppo 2</b>	<b>0</b>	<b>0</b>					<b>4</b>	<b>4</b>	<b>4</b>

le precedenze sono 12.

**Come valore di U si deve scegliere la somma minore.**

Nell'esercizio, il valore di U risulta uguale a 8.

5- **Stimare la significatività**, con la tabella dei valori critici nel caso di piccoli campioni e con la distribuzione normale nel caso di grandi campioni (vedere metodo del test U di Mann-Whitney).

I dati dell'esempio formano due campioni piccoli.

Il numero massimo di precedenze riportato nella tabella dei valori critici di U alla probabilità  $\alpha = 0.05$  per  $n_1 = 4$  e  $n_2 = 5$  in un test bilaterale è 1.

Il numero di precedenze U calcolato (8) è maggiore di quello tabulato (1): non si può rifiutare l'ipotesi nulla. I due gruppi a confronto hanno una variabilità dei dati non significativamente differente.

Moses ha studiato l'efficienza asintotica relativa del suo test rispetto a quella del test F di Fisher. Essa dipende da **k**, il numero di osservazioni utilizzate per costruire i sotto insiemi

Nel caso di dati distribuiti normalmente

<b>la potenza-efficienza è</b>	0,30 quando $k = 2$
	0,50 quando $k = 3$
	0,61 quando $k = 4$
per crescere lentamente fino a	0,88 quando $k = 15$
e	0,96 quando $k$ tende all'infinito.

Nel caso di dati distribuiti come una esponenziale doppia, la potenza varia da 0,59 a 0,96.

Rispetto **al test di Freund-Ansari-Bradley**, nel caso di una distribuzione normale, il test di Moses

- è meno potente quando i sottoinsiemi sono composti da 2 o 3 unità,
- ha la stessa potenza quando sono formati da 4 unità,
- ha una potenza di 1,44 con 20 dati.

Il test di Moses è meno potente dell'altro test non parametrico, quando i dati hanno una distribuzione molto lontana dalla normalità e asimmetrica, come l'esponenziale doppia.

**ESEMPIO.** Due gruppi del Cladocero *Daphnia magna*, geneticamente identici, sono stati allevati in condizioni ambientali differenti: il gruppo A in abbondanza di cibo e il gruppo B in carenza, per verificare se questi ultimi sono caratterizzati da una variabilità maggiore.

Alla fine del periodo di crescita, con il microscopio binoculare sono state misurate le lunghezze (in millimetri) di tutti gli individui dei due gruppi, come riportato nella tabella sottostante.

Gruppo	4,290	3,900	3,783	3,900	4,095	4,329	4,173	4,095	4,095	4,056
A	3,939	3,978	4,017	4,251	4,017	---	---	---	---	---

Gruppo	3,120	3,112	3,120	2,847	3,081	3,042	3,042	3,042	3,081	2,964
A	3,120	2,964	3,003	3,081	3,042	2,925	3,198	3,120	2,964	3,003

Risposta.

E' un test con ipotesi nulla

$$H_0: \mathbf{s}_A^2 = \mathbf{s}_B^2$$

ed ipotesi alternativa

$$H_1: \mathbf{s}_A^2 < \mathbf{s}_B^2$$

ad una coda;

infatti, si vuole verificare se la dispersione dei dati del gruppo B è significativamente maggiore di quella del gruppo A.

Con i dati dell'esempio, disponendo di 15 osservazioni nel gruppo A e 20 nel gruppo B, per non avere un numero troppo ridotto di sotto-insiemi si può scegliere di formare sottogruppi di 3 unità. Con due sole unità si avrebbe il vantaggio di utilizzare un numero maggiore di devianze, ma con lo svantaggio di una loro minore stabilità.

Con il gruppo A si formano 5 sotto-insiemi,

con il gruppo B 6 sottoinsiemi,

come riportato nella tabella successiva:

GRUPPO		SOTTOINSIEMI DI TRE DATI			DEVIANZE
Gruppo A	1)	3,900	4,290	4,329	<b>0,1120</b>
	2)	3,783	4,095	4,251	<b>0,1135</b>
	3)	3,978	4,017	3,900	<b>0,0073</b>
	4)	4,017	4,095	3,939	<b>0,0121</b>
	5)	4,056	4,173	4,095	<b>0,0070</b>
Gruppo B	1)	3,112	3,120	3,042	<b>0,0040</b>
	2)	3,120	3,081	3,081	<b>0,0015</b>
	3)	2,964	3,003	3,120	<b>0,0135</b>
	4)	3,042	3,042	3,120	<b>0,0041</b>
	5)	2,964	3,042	3,003	<b>0,0030</b>
	6)	2,847	3,081	2,925	<b>0,0288</b>

I rimanenti 2 valori del gruppo B (2,964 e 3,198) sono eliminati dalle analisi successive, perché insufficienti a formare un sotto-insieme completo.

Come riportato nella tabella precedente, per ogni tripletta di dati è stata calcolata la devianza SQ con la formula abbreviata

$$SQ = \sum x^2 - \frac{(\sum x)^2}{n}$$

Alle due serie di devianze,

Devianze A	0,1120	0,1135	0,0073	0,0121	0,0070
------------	--------	--------	--------	--------	--------

Devianze B	0,0040	0,0015	0,0135	0,0041	0,0030	0,0288
------------	--------	--------	--------	--------	--------	--------

è possibile **applicare il test U di Wilcoxon-Mann-Whitney**, per verificare se il gruppo B ha indici di dispersione maggiori di quelli calcolati per A.

Per applicare il test, si dispongono i valori dei due gruppi insieme in ordine crescente, mantenendo l'informazione del gruppo d'appartenenza.

Successivamente occorre fare la somma delle precedenze per calcolare U

.0015	.0030	.0040	.0041	.0070	.0073	.0121	.0135	.0288	.1120	.1135
B	B	B	B	A	A	A	B	B	A	A
<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>				<b>3</b>	<b>3</b>		

che risulta uguale a 6.

I due campioni a confronto sono di piccole dimensioni.

Per l'uso della tabella dei valori critici si definisce

**n<sub>1</sub>** il campione con il numero minore di dati e **n<sub>2</sub>** quello con il numero maggiore.

Con i dati dell'esempio, per **n<sub>1</sub> = 5** e **n<sub>2</sub> = 6** il valore critico alla **probabilità  $\alpha = 0.05$  per test unilaterali** (riportato nella tabella dei valori critici di U) è uguale a 5, mentre il valore calcolato è 6.

Non è possibile rifiutare l'ipotesi nulla.

Non è dimostrato che la varianza del gruppo B sia significativamente maggiore di quella del gruppo A.

## CAPITOLO VIII

### ANALISI DELLA VARIANZA (ANOVA I) A UN CRITERIO DI CLASSIFICAZIONE E CONFRONTI TRA PIU' MEDIE

Nella ricerca sperimentale è frequente **il confronto simultaneo tra le medie di più di due gruppi**, formati da soggetti sottoposti a trattamenti differenti o con dati raccolti in condizioni diverse. Al fine di evidenziare tutte le possibili differenze significative tra le medie, **non è corretto ricorrere al test t di Student per ripetere l'analisi tante volte quanti sono i possibili confronti** a coppie tra i singoli gruppi; è errato “dragare” i dati con il test **t**, per effettuare a posteriori (*comparisons post-hoc*) tanti confronti tra le medie dei gruppi quante sono le combinazioni degli  $n$  trattamenti 2 a 2, alla ricerca di qualunque differenza tra le medie, da giustificare in seguito.

Con il metodo del **t** di Student, **si utilizza solo una parte dei dati** e la probabilità  $\alpha$  prescelta per l'accettazione dell'ipotesi nulla, **la probabilità di commettere un errore di primo tipo** (rifiutare l'ipotesi nulla quando in realtà è vera) **è valida solamente per ogni singolo confronto. Se i confronti sono numerosi, la probabilità complessiva che almeno uno di essi si dimostri significativo solo per effetto del caso aumenta proporzionalmente.**

Se è vera l'ipotesi nulla  $H_0$ , la probabilità che nessun confronto risulti casualmente significativo è  $(1-\alpha)^n$  dove  $n$  è il numero di confronti effettuati.

Per esempio, se si effettuano **10** confronti tra le medie di gruppi estratti a caso dalla stessa popolazione e per ognuno di essi  $\alpha$  è uguale a **0.05**, la probabilità che nessun confronto risulti casualmente significativo diminuisce a **0.60** (corrispondente a  $0,95^{10}$ ); reciprocamente, la probabilità complessiva che almeno uno risulti significativo solo per effetto di fluttuazioni casuali diventa **0.40**.

Espresso in termini più formali, effettuando  $k$  confronti con il test **t** di Student ognuno alla probabilità  $\alpha$ , la probabilità complessiva  $\alpha'$  di commettere almeno un errore di I tipo (che il test rifiuti l'ipotesi nulla quando in realtà essa è vera) diventa

$$\alpha' = 1 - (1 - \alpha)^k$$

Nell'analisi della varianza, con apparente paradosso dei termini, il confronto è tra due o più medie. Essa permette il **confronto simultaneo tra esse, mantenendo invariata la probabilità  $\alpha$  complessiva prefissata.**

L'ipotesi  $H_0$  e l'ipotesi alternativa  $H_1$  assumono una formulazione più generale, rispetto al confronto tra due medie:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

$H_1$ : non tutte le  $m_i$  sono uguali  
(oppure, almeno una  $m_i$  è diversa)

L'ipotesi nulla  $H_0$  afferma che **le medie delle popolazioni dalle quali sono estratti casualmente i vari campioni sono tra loro tutte uguali** oppure che tutti i campioni a confronto sono stati estratti dalla medesima popolazione.

L'ipotesi alternativa  $H_1$  può essere solamente bilaterale: **le medie non sono tra loro tutte uguali**. In questo caso, possono realizzarsi varie situazioni, delle quali le due più estreme sono:

- le medie sono tutte differenti tra loro,
- una sola media è diversa dalle altre.

La metodologia sviluppata per verificare la significatività delle differenze tra le medie aritmetiche di vari gruppi, chiamata **analisi della varianza** e sintetizzata in **ANOVA** dall'acronimo dell'inglese **ANalysis Of VAriance**, utilizza **la distribuzione F**.

Fondata sul rapporto tra varianze, è stata così denominata in onore di Sir Ronald Aylmer **Fisher** (1890-1962), giudicato il più eminente statistico contemporaneo e ritenuto il padre della statistica moderna. La metodologia attuale è tuttavia dovuta a **Snedecor**, un suo allievo, che ne perfezionò il metodo e ne semplificò la forma, per cui la distribuzione **F** è ricordata anche come distribuzione di **Fisher-Snedecor**.

Nel 1925 Fisher, al quale tra gli argomenti già affrontati si devono la definizione dei gradi di libertà, gli indici di simmetria e curtosi, il metodo esatto per tabelle  $2 \times 2$ , completò il metodo di Student per il confronto tra due medie. La sua proposta **permette di scomporre e misurare l'incidenza delle diverse fonti di variazione sui valori osservati di due o più gruppi**. E' **la metodologia che sta alla base della statistica moderna**, che ha progressivamente elaborato i principi enunciati fino alle analisi più complesse, con le quali si possono tenere in considerazione contemporaneamente molti fattori sia indipendenti che correlati.

**La grande rivoluzione introdotta dall'analisi della varianza rispetto al test t consiste nel differente approccio alla programmazione dell'esperimento**. L'approccio del test t risente del vecchio assioma che **la natura risponde solo a domande semplici**. Per organizzare un esperimento, il materiale con il quale formare i gruppi a confronto doveva essere il più omogeneo possibile. Per esempio, per confrontare l'effetto di due tossici su un gruppo di cavie, gli animali dovevano essere dello stesso sesso, della stessa età, della stessa dimensione, ecc., se si riteneva che sesso, età, peso e qualunque altro carattere noto incidessero sulla risposta dell'esperimento. La differenza tra i due gruppi poteva risultare più facilmente significativa, in quanto l'errore standard risultava indubbiamente minore; ma le conclusioni erano ovviamente limitate al gruppo di animali con le caratteristiche prescelte, senza possibilità di estensione a cavie con caratteri diversi. Per rendere più generali le conclusioni, non rimaneva che ripetere l'esperimento, variando un carattere alla volta. Era

richiesto un forte aumento della quantità di materiale ed un allungamento dei tempi necessari all'esperimento; alla fine, rimaneva complesso trarre conclusioni generali.

**La grande novità introdotta dall'analisi della varianza**, come verrà evidenziato progressivamente con analisi sempre più complesse che considerano contemporaneamente un numero sempre più elevato di fattori e di loro interazioni, **è la scoperta dei vantaggi offerti all'analisi dall'uso di materiale molto diversificato**. Conoscendo le cause ed i diversi fattori, è possibile attribuire ad ognuno di essi la parte di effetto determinata e contemporaneamente ridurre la variabilità d'errore. **Le differenze tra le medie dei gruppi diventano molto più facilmente significative e le conclusioni possono essere immediatamente estese alle varie situazioni**.

Dall'introduzione dell'analisi della varianza è vantaggioso usare materiale non omogeneo per tutti i caratteri.

Nell'analisi della varianza, la fonte o causa delle variazioni dei dati viene chiamata **fattore sperimentale** o **trattamento**; essa può essere

- a più **livelli** quantitativi, come le dosi crescenti dello stesso farmaco, oppure
- a diverse **modalità** qualitative, come la somministrazione di farmaci differenti.

Ogni unità od osservazione del gruppo sperimentale viene chiamata **replicazione** o **replica**; per permettere di calcolare la media e la varianza, ovviamente ogni gruppo deve essere formato da almeno due repliche

### **8.1. ANALISI DELLA VARIANZA AD UN CRITERIO DI CLASSIFICAZIONE O A CAMPIONAMENTO COMPLETAMENTE RANDOMIZZATO**

Il modello più semplice di analisi della varianza, che può essere visto come un'estensione del test **t** di Student a più campioni indipendenti, è detto **ad un criterio di classificazione**: ogni dato è classificato solo sulla base del trattamento o del gruppo al quale appartiene. E' chiamato anche **modello completamente randomizzato** in quanto, soprattutto per analisi di laboratorio, prevede **un campionamento in cui gli n individui omogenei sono assegnati casualmente ai vari livelli del fattore**.

Quando si dispone di un gruppo di soggetti (ad esempio, cavie) da sottoporre a diversi trattamenti per confrontarne gli effetti, l'attribuzione di ogni individuo ad uno specifico trattamento deve avvenire per estrazione casuale da tutto il gruppo.

La metodologia di presentazione delle osservazioni, ormai codificata, prevede che i dati sperimentali raccolti siano riportati in modo ordinato secondo la tabella sottostante. Per l'analisi statistica, in questo modello **non è richiesto che i vari gruppi abbiano lo stesso numero ( $n_i$ ) di osservazioni o di repliche**.

		MODALITA' O LIVELLI DEI TRATTAMENTI				
		T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	...	T <sub>p</sub>
UNITÀ' SPERIMENTALI O REPLICAZIONI		X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>	...	X <sub>1p</sub>
		X <sub>21</sub>	X <sub>22</sub>	X <sub>23</sub>	...	X <sub>2p</sub>
		X <sub>31</sub>	X <sub>32</sub>	X <sub>33</sub>	...	X <sub>3p</sub>
		...	...	...	...	...
		X <sub>n<sub>1</sub>1</sub>	X <sub>n<sub>2</sub>2</sub>	X <sub>n<sub>3</sub>3</sub>	...	X <sub>n<sub>p</sub>p</sub>
Medie dei trattamenti		$\bar{X}_{.1}$	$\bar{X}_{.2}$	$\bar{X}_{.3}$	...	$\bar{X}_{.p}$
Media generale		$\bar{\bar{X}}$				

La **singola osservazione**  $X_{ij}$  viene riportata con 2 indici, relativi uno al trattamento o gruppo e l'altro alla posizione occupata entro il gruppo.

La **media di ogni gruppo o singolo trattamento**  $\bar{X}_i$  è riportata soprasssegnata da un tratto e con l'indice relativo al gruppo.

La **media generale**  $\bar{\bar{X}}$  di tutti i dati è indicata con un duplice tratto e senza indici.

**A partire da queste tre quantità, si stimano le devianze e le varianze utili all'analisi.**

**L'analisi della varianza è fondata sugli effetti additivi dei vari fattori considerati.** Nel modello più semplice, che considera un solo fattore a due o più livelli, ogni singola osservazione  $X_{ij}$  può essere scritta come

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

in quanto determinata

- dalla **media generale**  $\mu$ , che definisce la dimensione dell'esperimento,
- dal **fattore**  $\alpha_i$  **del trattamento** e
- da un **fattore casuale**  $\varepsilon_{ij}$ , detto **residuo** od **errore sperimentale**.

(E' importante ricordare che **errore** non è sinonimo di sbaglio, ma **indica l'effetto di uno o più fattori sconosciuti, comunque non valutati o non controllati nell'esperimento**).

In tale modello, l'**effetto** a **del trattamento** a sua volta è misurato come

$$\alpha_i = \mu_i - \mu$$

dove  $\mu_i$  è la media del trattamento e  $\mu$  la media generale.

Passando dall'enunciazione teorica ai dati sperimentali, si può scrivere che ogni singolo dato  $X_{ij}$  di uno specifico trattamento

$$X_{ij} = \bar{\bar{X}} + (\bar{X}_i - \bar{\bar{X}}) + \varepsilon_{ij}$$

è determinato

- dalla media generale  $\bar{\bar{X}}$ ,
- dall'effetto del trattamento  $(\bar{X}_i - \bar{\bar{X}})$  e
- da altri fattori non noti, simboleggiati da  $\varepsilon_{ij}$ .

Prima dell'applicazione di questo test parametrico, occorre verificare se ne esistono le condizioni.

**Le assunzioni di validità del test F dipendono dagli errori  $\varepsilon_{ij}$ , che**

- **devono essere tra loro indipendenti,**
- **devono essere distribuiti normalmente;** inoltre
- **le varianze dei vari gruppi devono essere omogenee.**

L'**indipendenza degli errori** comporta che la variazione casuale di ogni osservazione non sia influenzata da quella di un'altra: l'errore di una replica, il suo scarto rispetto alla media del gruppo di appartenenza, non deve essere influenzato né dal segno (quando si possono avere valori sia negativi che positivi) né dalle dimensioni del suo valore.

A questo fine, la randomizzazione deve essere fondata su elementi obiettivi (**effetto random**) e non lasciata all'arbitrio o all'intuito dello sperimentatore; ogni dato deve avere la stessa possibilità di essere influenzato dai fattori noti (**effetto trattamento**) e da quelli ignoti (**effetto ambiente** statistico). L'attribuzione (che sarà discussa nel capitolo sul campionamento) deve avvenire con modalità indipendenti dal ricercatore.

Gli **errori devono essere distribuiti normalmente intorno alla media**. Prima dell'applicazione del test deve essere attuato il controllo dell'asimmetria e della curtosi della distribuzione, per verificare che non si discosti eccessivamente dalla normale. Quando lo scostamento è significativo, sovente è possibile ricostruire le condizioni di validità attraverso la trasformazione dei dati (che saranno presentate successivamente).

L'**omogeneità della varianza**, per cui i diversi gruppi dei quali si confrontano le rispettive medie devono avere tutti la stessa varianza vera ( $\sigma^2$ ), è indispensabile per non determinare perdite nell'informazione sull'effetto dei trattamenti. Anche in questo caso, può essere necessario ricorrere alla trasformazione dei dati.

Dopo l'analisi dei dati per la verifica delle condizioni di validità, la **metodologia dell'analisi della varianza prevede il calcolo** delle seguenti quantità:

- la **devianza totale**, con i suoi **gdl**;
- la devianza **tra trattamenti** o **between**, con i suoi **gdl** e la **varianza** relativa;
- la devianza **entro trattamenti** o **within** od **errore**, con i suoi **gdl** e la **varianza** relativa.

Ai fini di una verifica dei risultati e delle successive loro elaborazioni, è utile ricordare che **la somma della devianza tra trattamenti e di quella entro trattamenti è uguale alla devianza totale; identica proprietà additiva hanno i rispettivi gradi di libertà.**

Devianze, gdl e varianze di un'analisi della varianza abitualmente vengono presentate come nella tabella seguente:

<b>Devianza Totale</b>	<b>gdl = n-1</b> (n = num. dati)	
<b>Devianza tra trattamenti</b>	<b>gdl = p-1</b> (p = num. gruppi)	<b>varianza tra</b> $S^2_{tra}$
<b>Devianza entro trattamenti</b>	<b>gdl = n-p</b>	<b>varianza entro</b> $S^2_{entro}$

(molti testi riportano la devianza totale e i suoi gdl alla fine, in quanto somma dei precedenti)

La **devianza totale** o **SQ totale** (Somma dei Quadrati degli scarti, in inglese **SS** da Sum of Squares) è calcolato da

$$SQ_{totale} = \sum_{j=1}^p \sum_{i=1}^{n_i} (X_{ij} - \bar{X})^2 = \sum_{j=1}^p \sum_{i=1}^{n_i} X_{ij}^2 - \frac{(\sum_{j=1}^p \sum_{i=1}^{n_i} X_{ij})^2}{n}$$

La prima è chiamata **formula euristica**, in quanto definisce il significato della devianza totale: la somma del quadrato degli scarti di ogni valore dalla media generale.

La seconda è la **formula abbreviata**, matematicamente equivalente alla prima, che rende più semplici e rapidi i calcoli necessari. Con essa, la devianza totale è ottenuta come differenza tra la somma dei quadrati di tutti i dati e il quadrato della somma di tutti i dati diviso il numero di dati.

La seconda formula ha il vantaggio di richiedere meno operazioni e di non utilizzare la media, che spesso è un valore approssimato; in queste condizioni, consente un calcolo più preciso della formula euristica.

La **devianza tra trattamenti** ( $SQ_{tra}$ ) o **between**

$$SQ_{tra} = \sum_{j=1}^p n_j (\bar{X}_j - \bar{\bar{X}})^2 = \sum_{j=1}^p \frac{\left(\sum_{i=1}^{n_j} X_{ij}\right)^2}{n_j} - \frac{\left(\sum_{j=1}^p \sum_{i=1}^{n_j} X_{ij}\right)^2}{n}$$

è per definizione (**formula euristica**) la somma dei quadrati degli scarti di ogni media di gruppo dalla media generale, moltiplicato il numero di dati del gruppo relativo.

La **formula abbreviata** utilizza le somme dei gruppi e la somma totale, determinando una maggiore precisione nei risultati.

La **devianza entro trattamenti** ( $SQ_{entro}$ ) o **within**, detta anche **errore**

$$SQ_{entro} = \sum_{j=1}^p \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2 = SQ_{totale} - SQ_{tra}$$

è la somma degli scarti al quadrato di ogni valore dalla media del suo gruppo.

Per la proprietà additiva delle devianze, può essere ottenuta sottraendo alla devianza totale la devianza tra trattamenti.

**I gradi di libertà sono determinati dal numero di somme richieste dal calcolo delle devianze relative**, nella formula euristica.

Per la **devianza totale**, dove la sommatoria è estesa a tutti gli **n** dati, i gdl sono **n-1**.

Per la **devianza tra trattamenti**, dove la sommatoria è estesa ai **p** gruppi, i gdl sono **p-1**.

Per la **devianza entro od errore**, la sommatoria è estesa a tutti i dati entro ogni gruppo. Per calcolare i gdl occorre quindi sottrarre 1 ai dati di ogni gruppo e quindi è determinata da **n-p**.

Per la proprietà additiva dei gdl, può essere scritta anche come **(n-1) - (p-1)**, semplificato in **n-p**.

Dividendo la devianza tra trattamenti e quella entro trattamenti per i rispettivi gradi di libertà, si ottengono la **varianza tra** e la **varianza entro** (la varianza totale è priva d'interesse ai fini di questo test).

**La varianza fra gruppi misura le differenze esistenti tra un gruppo e l'altro**, anche se il calcolo viene attuato rispetto alla media generale.

**La varianza entro gruppi misura la variabilità esistente attorno alla media aritmetica di ogni gruppo.**

**Se è vera l'ipotesi nulla**, i dati dei vari gruppi sono estratti casualmente dalla stessa popolazione. La varianza tra le medie dei trattamenti e la varianza entro ogni gruppo dipendono dalla variabilità esistente tra i dati: **varianza fra** ( $s_F^2$ ) e **varianza entro** ( $s_e^2$ ) sono due stime indipendenti della stessa varianza vera  $s^2$  e quindi **dovrebbero avere statisticamente lo stesso valore**.

Come indice dell'uguaglianza tra le due varianze, viene utilizzato

il test **F di Fisher**, fondato sul rapporto

**varianza-tra / varianza-entro**

indicato con la simbologia

$$F_{(p-1, n-p)} = \frac{S_F^2}{S_e^2}$$

Se è vera l'ipotesi nulla **H<sub>0</sub>**

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

il rapporto dovrebbe risultare **uguale ad 1**.

Se è vera l'ipotesi alternativa **H<sub>1</sub>**

$$H_1: \text{le } m_i \text{ non sono tutte uguali}$$

il rapporto dovrebbe risultare **superiore a 1**.

Con un numero infinito di trattamenti e di repliche, è sufficiente un rapporto superiore a 1 per rifiutare l'ipotesi nulla (come mostra la tabella dei valori critici di F); con un numero ridotto di dati, il rapporto può essere superiore a 1, per effetto delle variazioni casuali. I valori critici per i rispettivi gradi di libertà sono forniti dalla distribuzione F.

**Se il valore di F calcolato supera quello tabulato, alla probabilità prefissata, si rifiuta l'ipotesi nulla** e si accetta l'ipotesi alternativa: almeno una media è diversa dalle altre.

**Se il valore F calcolato è inferiore a quello riportato nella tabella, si accetta l'ipotesi nulla, o almeno non può essere rifiutato che le medie sono tutte uguali.**

#### **Valori critici della distribuzione F di Fisher-Snedecor**

I gradi di libertà del numeratore (o varianza maggiore) sono riportati in orizzontale (prima riga)

I gradi di libertà del denominatore (o varianza minore) sono riportati in verticale (prima colonna)

$$\alpha = 0.05$$

NUMERATORE

DEN.	1	2	3	4	5	6	7	8	12	24	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	243,9	249,1	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	2,91	2,74	2,54
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,69	2,51	2,30
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,53	2,35	2,13
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,42	2,24	2,01
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,34	2,15	1,92
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,28	2,08	1,84
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,83	1,61	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,75	1,52	1,00

**Valori critici della distribuzione F di Fisher-Snedecor**

I gradi di libertà del numeratore (o varianza maggiore) sono riportati in orizzontale (prima riga)

I gradi di libertà del denominatore (o varianza minore) sono riportati in verticale (prima colonna)

$$\alpha = 0.025$$

NUMERATORE

DEN.	1	2	3	4	5	6	7	8	12	24	∞
<b>1</b>	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7	976,7	997,2	1018
<b>2</b>	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,41	39,46	39,50
<b>3</b>	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,34	14,12	13,90
<b>4</b>	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,75	8,51	8,26
<b>5</b>	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,52	6,28	6,02
<b>6</b>	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,37	5,12	4,85
<b>7</b>	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,67	4,42	4,14
<b>8</b>	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,20	3,95	3,67
<b>9</b>	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	3,87	3,61	3,33
<b>10</b>	6,94	5,46	4,83	4,46	4,24	4,06	3,95	3,85	3,62	3,37	3,08
<b>12</b>	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,28	3,02	2,72
<b>14</b>	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,05	2,79	2,49
<b>16</b>	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	2,89	2,63	2,32
<b>18</b>	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,77	2,50	2,19
<b>20</b>	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,68	2,41	2,09
<b>30</b>	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,41	2,14	1,79
<b>40</b>	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,29	2,01	1,64
<b>60</b>	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,17	1,88	1,48
<b>120</b>	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,05	1,76	1,31
<b>∞</b>	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	1,94	1,64	1,00

**Valori critici della distribuzione F di Fisher-Snedecor**

I gradi di libertà del numeratore (o varianza maggiore) sono riportati in orizzontale (prima riga)

I gradi di libertà del denominatore (o varianza minore) sono riportati in verticale (prima colonna)

$$\alpha = 0.01$$

NUMERATORE

DEN.	1	2	3	4	5	6	7	8	12	24	∞
<b>1</b>	4052	5000	5403	5625	5764	5859	5928	5981	6106	6235	6366
<b>2</b>	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,41	99,46	99,50
<b>3</b>	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,05	26,60	26,13
<b>4</b>	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,37	13,93	13,46
<b>5</b>	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	9,89	9,47	9,02
<b>6</b>	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,72	7,31	6,88
<b>7</b>	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,47	6,07	5,65
<b>8</b>	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,67	5,28	4,86
<b>9</b>	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,11	4,73	4,31
<b>10</b>	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,71	4,33	3,91
<b>12</b>	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,16	3,78	3,36
<b>14</b>	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	3,80	3,43	3,00
<b>16</b>	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,55	3,18	2,75
<b>18</b>	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,37	3,00	2,57
<b>20</b>	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,23	2,86	2,42
<b>30</b>	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	2,84	2,47	2,01
<b>40</b>	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,66	2,29	1,80
<b>60</b>	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,50	2,12	1,60
<b>120</b>	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,34	1,95	1,38
<b>∞</b>	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,18	1,79	1,00

**Valori critici della distribuzione F di Fisher-Snedecor**

I gradi di libertà del numeratore (o varianza maggiore) sono riportati in orizzontale (prima riga)

I gradi di libertà del denominatore (o varianza minore) sono riportati in verticale (prima colonna)

**a = 0.005**

NUMERATORE

DEN.	1	2	3	4	5	6	7	8	12	24	¥
<b>1</b>	16211	20000	21615	22500	23056	23437	23715	23925	24426	24940	25465
<b>2</b>	198,5	199,0	199,2	199,2	199,3	199,3	199,4	199,4	199,4	199,5	199,5
<b>3</b>	55,55	49,80	47,47	46,19	45,39	44,84	44,43	44,13	43,39	42,62	41,83
<b>4</b>	31,33	26,28	24,26	23,15	22,46	21,97	21,62	21,35	20,70	20,03	19,32
<b>5</b>	22,78	18,31	16,53	15,56	14,94	14,51	14,20	13,96	13,38	12,78	12,14
<b>6</b>	18,63	14,54	12,92	12,03	11,46	11,07	10,79	10,57	10,03	9,47	8,88
<b>7</b>	16,24	12,40	10,88	10,05	9,52	9,16	8,89	8,68	8,18	7,65	7,08
<b>8</b>	14,69	11,04	9,60	8,81	8,30	7,95	7,69	7,50	7,01	6,50	5,95
<b>9</b>	13,61	10,11	8,72	7,96	7,47	7,13	6,88	6,69	6,23	5,73	5,19
<b>10</b>	12,83	9,43	8,08	7,34	6,87	6,54	6,30	6,12	5,66	5,17	4,64
<b>12</b>	11,75	8,51	7,23	6,52	6,07	5,76	5,52	5,35	4,91	4,43	3,90
<b>14</b>	11,06	7,92	6,68	6,00	5,56	5,26	5,03	4,86	4,43	3,96	3,44
<b>16</b>	10,58	7,51	6,30	5,64	5,21	4,91	4,69	4,52	4,10	3,64	3,11
<b>18</b>	10,22	7,21	6,03	5,37	4,96	4,66	4,44	4,28	3,86	3,40	2,87
<b>20</b>	9,94	6,99	5,82	5,17	4,76	4,47	4,26	4,009	3,68	3,22	2,69
<b>30</b>	9,18	6,35	5,24	4,62	4,23	3,95	3,74	3,58	3,18	2,73	2,18
<b>40</b>	8,83	6,07	4,98	4,37	3,99	3,71	3,51	3,35	2,95	2,50	1,93
<b>60</b>	8,49	5,79	4,73	4,14	3,76	3,49	3,29	3,13	2,74	2,29	1,69
<b>120</b>	8,18	5,54	4,50	3,92	3,55	3,28	3,09	2,93	2,54	2,09	1,43
<b>¥</b>	7,88	5,30	4,28	3,72	3,35	3,09	2,90	2,74	2,36	1,90	1,00

ESEMPIO. Per un controllo della qualità dell'aria, con rilevazioni in tre diverse zone di una città (denominate A, B e C) è stata misurata anche la quantità di ferro (in microgrammi/Nmc a 0°C e 1013 mbar) tra i metalli pesanti in sospensione.

FATTORE SPERIMENTALE		
A	B	C
2,71	1,75	2,22
2,06	2,19	2,38
2,84	2,09	2,56
2,97	2,75	2,60
2,55		2,72
2,78		

Esiste una differenza significativa tra le tre zone, per la quantità di ferro in sospensione?

Risposta.

L'ipotesi nulla  $H_0$  è che tra le medie dei tre campioni non esistano differenze significative

$$H_0: \mu_A = \mu_B = \mu_C$$

mentre l'ipotesi alternativa  $H_1$

$H_1$ : le  $\mu_i$  non sono tutte uguali.

Attraverso il test F è possibile stimare la probabilità di trovare per caso tra le medie scarti uguali o superiori a quelli sperimentalmente osservati, nell'ipotesi che  $H_0$  sia vera.

Come primo passo, dalle tre serie di dati occorre calcolare

- il totale di ogni colonna  $\sum X_j$
- il numero di osservazioni  $n_j$
- la media di ogni colonna  $\bar{X}_j$

Successivamente, da essi è necessario stimare

- la somma totale  $\sum X$
- il numero totale di osservazioni  $n$
- la media totale o generale  $\bar{X}$

come riportato nella tabella successiva:

	A	B	C		
$\sum X_j$	15,91	8,78	12,48	$\sum X$	37,17
$n_j$	6	4	5	<b>n</b>	15
$\bar{X}_j$	2,652	2,195	2,496	$\bar{X}$	2,478

A partire da queste quantità, si calcolano le devianze ed i gradi di libertà rispettivi.

La **devianza totale** può essere calcolata dalla somma del quadrato degli scarti di ognuna delle 15 osservazioni rispetto alla media totale, in accordo con la formula euristica

$$SQ_{\text{totale}} = \sum_{j=1}^p \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2$$

A	B	C
$(2,71 - 2,478)^2$	$(1,75 - 2,478)^2$	$(2,22 - 2,478)^2$
$(2,06 - 2,478)^2$	$(2,19 - 2,478)^2$	$(2,38 - 2,478)^2$
$(2,84 - 2,478)^2$	$(2,09 - 2,478)^2$	$(2,56 - 2,478)^2$
$(2,97 - 2,478)^2$	$(2,75 - 2,478)^2$	$(2,60 - 2,478)^2$
$(2,55 - 2,478)^2$		$(2,72 - 2,478)^2$
$(2,78 - 2,478)^2$		

Svolgendo i calcoli e sommando i risultati

A	B	C
0,053824	0,529984	0,066564
0,174724	0,082944	0,009604
0,131044	0,150544	0,006724
0,242064	0,073984	0,014884
0,005184		0,058564
0,091204		
<b>0,698040</b>	<b>0,837456</b>	<b>0,156340</b>

$$\text{Devianza totale} = 0,698040 + 0,837456 + 0,156340 = 1,691836$$

si ottiene una devianza totale uguale a 1,691836 con 14 gdl.

Questo metodo di procedere al calcolo della devianza totale è lungo e determina stime non precise, quando la media generale è approssimata. Pertanto, **per il calcolo manuale è sempre conveniente utilizzare la formula abbreviata**

$$SQ_{\text{totale}} = \sum_{j=1}^p \sum_{i=1}^{n_j} X_{ij}^2 - \frac{(\sum_{j=1}^p \sum_{i=1}^{n_j} X_{ij})^2}{n}$$

che, applicata ai dati dell'esempio,

	A	B	C	
	7,3441	3,0625	4,9284	
	4,2436	4,7961	5,6644	
	8,0656	4,3681	6,5536	
	8,8209	7,5625	6,7600	
	6,5025		7,3984	
	7,7284			
$\sum X^2$	<b>42,7051</b>	<b>19,7892</b>	<b>31,3048</b>	<b>93,7991</b>

dalle due diverse somme stima la

$$Devianza_{\text{totale}} = 93,7991 - \frac{(37,17)^2}{15} = 1,69184$$

devianza totale che risulta uguale a 1,69184.

La corrispondenza tra le due stime è una dimostrazione elementare ed intuitiva dell'equivalenza matematica delle due formule (la differenza tra 1,691836 e 1,69184 è dovuta agli arrotondamenti).

La **devianza tra trattamenti** o **between** misura la variabilità esistente tra la media aritmetica di ogni gruppo e la media aritmetica generale, ponderata per il numero di osservazioni presenti in ciascun gruppo. Se non esistesse la variabilità casuale ed il valore delle singole osservazioni fosse determinato solamente dal fattore specifico che le raggruppa, le repliche di ogni trattamento dovrebbero avere tutte lo stesso valore ed essere uguali alla media del gruppo, come evidenzia la formula euristica

$$SQ_{tra} = \sum_{j=1}^p n_j (\bar{X}_j - \bar{X})^2$$

**La devianza tra trattamenti o between è la somma degli scarti di ogni media di gruppo rispetto alla media generale, ponderata per il numero di repliche.**

Pertanto con la formula euristica il calcolo diventa:

$$\begin{aligned} \text{Devianza}_{tra} &= 6 \cdot (2,652 - 2,478)^2 + 4 \cdot (2,195 - 2,478)^2 + 5 \cdot (2,496 - 2,478)^2 = \\ &= 6 \cdot 0,030276 + 4 \cdot 0,080089 + 5 \cdot 0,000324 = \\ &= 0,181656 + 0,320356 + 0,00162 = 0,503632 \end{aligned}$$

e risulta uguale a 0,503632 con 2 gradi di libertà.

Anche in questo caso la formula abbreviata

$$SQ_{tra} = \sum_{j=1}^p \frac{\left( \sum_{i=1}^{n_j} X_{ij} \right)^2}{n_j} - \frac{\left( \sum_{j=1}^p \sum_{i=1}^{n_j} X_{ij} \right)^2}{n}$$

è più rapida e precisa, non richiedendo le approssimazioni determinate dalle medie;

$$SQ_{tra} = \frac{(15,91)^2}{6} + \frac{(8,78)^2}{4} + \frac{(12,48)^2}{5} - \frac{(37,17)^2}{15} = 92,610196 - 92,10726 = 0,502936$$

essa risulta uguale a 0,502936. Anche in questo caso le differenze sono minime (0,503632 e 0,502936), imputabili all'uso di un numero diverso di cifre decimali e alle differenti approssimazioni.

(Di solito sono sufficienti calcoli con due o tre cifre decimali; il numero più elevato qui utilizzato è motivato dalla necessità contingente di confrontare i risultati dei due metodi).

**La devianza entro trattamenti, within od errore**

$$SQ_{entro} = \sum_{j=1}^p \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2$$

misura la variazione tra il valore di ciascuna replica e la media aritmetica del suo gruppo.

Sommando queste differenze elevate al quadrato per ogni gruppo

A	B	C
$(2,71 - 2,652)^2$	$(1,75 - 2,195)^2$	$(2,22 - 2,496)^2$
$(2,06 - 2,652)^2$	$(2,19 - 2,195)^2$	$(2,38 - 2,496)^2$
$(2,84 - 2,652)^2$	$(2,09 - 2,195)^2$	$(2,56 - 2,496)^2$
$(2,97 - 2,652)^2$	$(2,75 - 2,195)^2$	$(2,60 - 2,496)^2$
$(2,55 - 2,652)^2$		$(2,72 - 2,496)^2$
$(2,78 - 2,652)^2$		

e sviluppando i calcoli si ottiene

	A	B	C
	0,003364	0,198025	0,076176
	0,350464	0,000025	0,013456
	0,035344	0,011025	0,004096
	0,101124	0,308025	0,010816
	0,010404		0,050176
	0,015376		
<b>Devianza<sub>entro</sub></b>	<b>0,516076</b>	<b>0,517100</b>	<b>0,154720</b>

$$\text{Devianza}_{\text{entro}} = 0,516076 + 0,517100 + 0,154720 = 1,187896$$

la devianza entro, che risulta uguale a 1,187896 con 12 gdl.

La devianza entro od errore può essere ottenuta molto più rapidamente per sottrazione della devianza tra dalla devianza totale, precedentemente calcolate:

$$\text{Devianza}_{\text{entro}} = \text{Devianza}_{\text{totale}} - \text{Devianza}_{\text{tra}} = 1,69184 - 0,502936 = 1,188904$$

Nello stesso modo, per la proprietà additiva, si possono calcolare i gdl:

$$\text{gdl}_{\text{entro}} = \text{gdl}_{\text{totale}} - \text{gdl}_{\text{tra}} = 14 - 2 = 12$$

Per una presentazione chiara e sintetica, **normalmente i valori calcolati sono riassunti in una tabella che riporta le tre devianze, i rispettivi gradi di libertà e le varianze utili al test:**

DEVIANZA		GDL	VARIANZA
<b>Totale</b>	1,69184	14	----
<b>Tra trattamenti (between)</b>	0,502936	<b>2</b>	<b>0,251468</b>
<b>Entro trattamenti (within)</b>	1,188904	<b>12</b>	<b>0,0990753</b>

Dividendo la devianza tra e la devianza entro per i rispettivi gradi di libertà, si ottengono la varianza tra e la varianza entro.

Dividendo la varianza tra per la varianza entro, si calcola il rapporto F, che deve essere riportato con i rispettivi gradi di libertà  $F_{(2,12)}$

$$F_{(2,12)} = \frac{0,251468}{0,0990753} = 2,538$$

Il valore critico di F con gdl 2 per il numeratore e 12 per il denominatore che è riportato nella tabella per la probabilità  $\alpha = 0.05$  è 3,89. Il valore calcolato (2,538) è inferiore a quello tabulato (3,89): la probabilità che l'ipotesi nulla sia vera è superiore al 5%. Di conseguenza, si accetta l'ipotesi nulla: i tre campioni sono stati estratti dalla stessa popolazione; non esiste una differenza significativa tra le 3 medie campionarie.

## 8.2. CONFRONTO TRA L'ANALISI DELLA VARIANZA CON DUE TRATTAMENTI E IL TEST t DI STUDENT PER 2 CAMPIONI INDIPENDENTI.

L'analisi della varianza può essere applicata anche a 2 soli trattamenti; per questo caso, è già stata presentata la metodologia del test **t** di Student. In realtà, test t e test F sono due modi solo apparentemente differenti per fare la stessa analisi: **il test t può essere visto come un caso speciale di analisi della varianza**, applicata solo a due gruppi; meglio ancora, l'analisi della varianza è l'estensione a più gruppi e a più fattori del test **t** di Student.

Nel caso di un solo fattore con due gruppi, tra t ed F esiste una relazione matematica precisa:

$$F_{(1,n)} = t_{(n)}^2$$

che ovviamente può anche essere scritta come

$$t_{(v)} = \sqrt{F_{(1,n)}}$$

dove n è il numero di gradi di libertà.

**Il valore di F con gradi di libertà 1 al numeratore e n al denominatore è uguale al quadrato di t con n gradi di libertà.**

**Le due distribuzioni dei valori critici per la stessa probabilità  $\alpha$  sono equivalenti**, come è possibile evidenziare dal semplice confronto tra le tabelle dei valori critici.

ESEMPIO. Due gruppi di 10 uova di *Daphnia magna*, estratte casualmente dallo stesso clone, sono state allevate in due vasche con diverse concentrazioni di cromo, per verificare se incidono significativamente sulla crescita.

Dopo un mese sono stati misurati gli individui sopravvissuti: 7 nel gruppo A e 8 nel gruppo B, con le dimensioni riportate:

A	2,7	2,8	2,9	2,5	2,6	2,7	2,8	---
B	2,2	2,1	2,2	2,3	2,1	2,2	2,3	2,6

Risposta.

L'ipotesi nulla è

$$H_0: \mu_A = \mu_B$$

e l'ipotesi alternativa  $H_1$  bilaterale è

$$H_1: \mu_A \neq \mu_B$$

Prima di procedere sia al test t che al test F, si deve **verificare se le due varianze sono omogenee**.

Quindi è preliminare al confronto tra le due medie il confronto tra le 2 varianze, per saggiare l'ipotesi nulla

$$H_0: \sigma_A^2 = \sigma_B^2$$

con l'ipotesi alternativa bilaterale

$$H_1: \sigma_A^2 \neq \sigma_B^2$$

A questo scopo si calcolano le 2 devianze ed i loro gradi di libertà, per stimare le varianze rispettive

	A	B
Devianza	0,10857	0,18000
Gdl	6	7
Varianza $s^2$	<b>0,018095</b>	<b>0,02571</b>

ed infine il rapporto F tra varianza maggiore (al numeratore) e varianza minore (al denominatore).

$$F_{(7,6)} = \frac{0,02571}{0,018095} = 1,42$$

Nella tabella dei valori critici, con 7 gradi di libertà per la varianza maggiore e 6 per la varianza minore, il valore critico alla probabilità  $\alpha = 0.05$  è uguale a 4,21. Il valore calcolato (1,42) è inferiore: di conseguenza, si accetta l'ipotesi nulla che le due varianze siano omogenee.

Di conseguenza, è corretto procedere al confronto tra le due medie.

**Per il test t di Student per 2 campioni indipendenti,**

si calcolano le due medie:

$$\text{media del gruppo A} = 2,714$$

$$\text{media del gruppo B} = 2,250$$

e la varianza mediata  $s_p^2$

$$s_p^2 = \frac{0,10825 + 0,18000}{6 + 7} = 0,022173$$

Da esse si stima il valore di t con gdl 13

$$t_{13} = \frac{2,714 - 2,250}{\sqrt{0,022173 \cdot \left(\frac{1}{7} + \frac{1}{8}\right)}} = 6,02$$

che risulta uguale a 6,02.

**Per l'analisi della varianza ad un criterio di classificazione,**

si devono calcolare la devianza totale, quella tra trattamenti e quella entro trattamenti, con i rispettivi gradi di libertà.

E' possibile una verifica dei calcoli effettuati, mediante la proprietà additiva delle devianze:

$$\text{Devianza totale} = \text{Devianza tra} + \text{Devianza entro}$$

	<b>Devianza</b>	<b>Gdl</b>	<b>Varianza</b>
<b>Totale</b>	1,093333	14	
<b>Tra</b>	0,804762	1	<b>0,804761</b>
<b>Entro</b>	0,288571	13	<b>0,022198</b>

Si calcolano la varianza tra e la varianza entro e da esse

si stima F con gdl 1 e 13

$$F_{(1,13)} = \frac{0,804761}{0,022198} = 36,25$$

che risulta uguale a 36,25.

E' semplice **verificare che le due risposte coincidono:**

$$t^2_{(13)} = F_{(1,13)}; \quad (6,02)^2 = 36,25$$

a meno delle approssimazioni dei calcoli effettuati.

Sulle tabelle dei valori critici del test t di Student e del test F di Fisher si controlla la probabilità, che per entrambe risulta ovviamente uguale e nettamente inferiore a 0.001.

Con entrambi i test si rifiuta l'ipotesi nulla alla stessa probabilità.

### **8.3. TEST PER L'OMOGENEITA' DELLA VARIANZA TRA PIU' CAMPIONI:**

#### **TEST DI HARTLEY, COCHRAN, BARTLETT, LEVENE**

**Il confronto tra medie con l'analisi della varianza richiede che i diversi gruppi abbiano varianze uguali.** Allontanarsi sensibilmente da questa condizione di validità influenza gravemente la varianza d'errore, quindi la significatività del test. Si utilizzerebbe una varianza d'errore media  $s^2$ , come stima della varianza vera  $\sigma^2$ , che risulterebbe troppo grande per alcuni trattamenti e troppo piccola se riferita ad altri.

Oltre alla verifica delle condizioni di validità per il confronto tra medie, spesso si ha anche **un interesse esplicito a un confronto tra le varianze.** Per esempio, gruppi di animali o piante geneticamente identici dovrebbero avere varianze significativamente minori di gruppi geneticamente eterogenei; gruppi cresciuti in condizioni ambientali molto differenti dovrebbero avere una varianza maggiore di gruppi allevati in condizioni simili; nelle analisi di laboratorio, uno strumento di misura più preciso od un reagente di qualità superiore dovrebbero fornire varianze minori rispetto a strumenti e reagenti di bassa qualità, in esperimenti ripetuti nelle stesse condizioni.

L'ipotesi di **omoscedasticità**, chiamata in alcuni testi in italiano anche **omoscedalità** od **omogeneità delle varianze**, nel caso di **più gruppi** richiede la verifica dell'ipotesi nulla

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_p^2$$

contro l'ipotesi alternativa

$$H_1 : \text{non tutte le } \sigma_i^2 \text{ sono uguali}$$

I metodi proposti sono numerosi; tra i più diffusi, di norma utilizzati anche nei programmi informatici standard per calcolatori, sono da ricordare

**A - il test  $F_{\max}$  di Hartley,**

**B - il test di Cochran,**

**C - il test di Bartlett,**

**D - il test di Levene.**

A) Il procedimento  **$F_{\max}$  di Hartley** è quello più semplice e rapido. Le difficoltà alla sua utilizzazione derivano solo dalla scarsa reperibilità di testi ad ampia diffusione con la tabella dei valori critici. Questa tabella (riportata nella pagina successiva) non è da confondere con quella di Fisher-Snedecor, riportata in tutti i testi.

Tale difficoltà è ora superata in molti programmi informatici recenti, che stampano il valore esatto della probabilità relativa.

Secondo il test di Hartley, **esiste una differenza significativa tra più varianze quando il rapporto tra la varianza maggiore  $s_{\max}^2$  e la varianza minore  $s_{\min}^2$**

$$F_{\max(p, n-1)} = \frac{s_{\max}^2}{s_{\min}^2}$$

**supera il valore critico** riportato nelle tabelle corrispondenti.

**Gli indici dei valori di  $F_{\max}$  considerano il numero p di gruppi a confronto simultaneo ed il numero di gradi di libertà n-1 di ogni gruppo.**

**Il test richiede che i gruppi abbiano tutti lo stesso numero n di osservazioni.**

E' un test semplice, ma non robusto: l'assunzione fondamentale è che i dati siano distribuiti normalmente. Se non è possibile supporre la normalità della distribuzione per ciascun gruppo, si dovrebbe ricorrere ad altri test, come quelli non parametrici.

**Valori critici per il test di Hartley  
sull'omogeneità della varianza tra k gruppi**

**$\alpha = 0.05$**

**numero k di varianze a confronto**

$Df_2$	2	3	4	5	6	7	8	9	10	11	12
2	39.0	87.5	142	202	266	333	403	475	550	626	704
3	15.4	27.8	39.2	60.7	62.0	72.9	83.5	93.9	104	114	124
4	9.60	15.5	20.6	26.2	29.5	33.6	37.5	41.1	44.6	48.0	51.4
5	7.15	10.3	13.7	16.3	18.7	20.8	22.9	24.7	26.5	28.2	29.9
6	5.82	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6	19.7	20.7
7	4.99	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3	15.1	15.8
8	4.43	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7	12.2	12.7
9	4.03	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91	10.3	10.7
10	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.01	9.34
12	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.25	7.48
15	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.77	5.93
20	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.49	4.59
30	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.36	3.39
60	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.33	2.36
$\infty$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

**$\alpha = 0.01$**

**numero k di varianze a confronto**

$Df_2$	2	3	4	5	6	7	8	9	10	11	12
2	199	448	729	1036	1362	1705	2063	2432	2813	3204	3605
3	47.5	85	120	151	184	216	249	281	310	337	361
4	23.2	37	49	59	69	79	89	97	106	113	120
5	14.9	22	28	33	38	42	46	50	54	57	60
6	11.1	15.5	19.1	22	25	27	30	32	34	36	37
7	8.89	12.1	14.5	16.5	18.4	20	22	23	24	26	27
8	7.50	9.9	11.7	13.2	14.5	15.8	16.9	17.9	18.9	19.8	21
9	6.54	8.5	9.9	11.1	12.1	13.1	13.9	14.7	15.3	16.0	16.6
10	5.85	7.4	8.6	9.6	10.4	11.1	11.8	12.4	12.9	13.4	13.9
12	4.91	6.1	6.9	7.6	8.2	8.7	9.1	9.5	9.9	10.2	10.6
15	4.07	4.9	5.5	6.0	6.4	6.7	7.1	7.3	7.5	7.8	8.0
20	3.32	3.8	4.3	4.6	4.9	5.1	5.3	5.5	5.6	5.8	5.9
30	2.63	3.0	3.3	3.4	3.6	3.7	3.8	3.9	4.0	4.1	4.2
60	1.96	2.2	2.3	2.4	2.4	2.5	2.5	2.6	2.6	2.7	2.7
$\infty$	1.00	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

**Infatti non esistono test parametrici adatti alla verifica della omogeneità della varianza, quando le distribuzioni dei dati si discostano dalla normalità.**

B ) Anche **il test proposto da Cochran** nel 1967 può essere applicato solo ad esperimenti bilanciati. E' metodologicamente semplice come il precedente e permette una verifica rapida dell'ipotesi nulla di omoscedasticità dei vari trattamenti.

E' fondato sul rapporto tra la varianza maggiore e la somma di tutte le altre varianze.

Si calcola il valore del rapporto  $R_{n,p}$

$$R_{n,p} = \frac{s_{\max}^2}{s_1^2 + s_2^2 + \dots + s_p^2}$$

dove

$s_{\max}^2$  la varianza campionaria maggiore,

$s_1^2, s_2^2, \dots, s_p^2$  sono le varianze dei **p** gruppi,

con un numero **n** di repliche uguali in ogni gruppo.

Anche in questo caso, i limiti derivano dall'esigenza di un numero uguale di osservazioni in tutti i gruppi e dalla ridotta diffusione delle tabelle specifiche. **Con un numero di osservazioni molto alto (infinito) il rapporto tende a 1/p.**

B) Più complessa è la metodologia per il **test di significatività approssimato di Bartlett**; basato su un principio di **J. Neyman e E. S. Pearson** (vedi, del 1931: *On the problem of k samples*. Bull. Acad. Polon. Sci. Lett. Ser. A, 3: 460-481), è stato presentato da **M. S. Bartlett nel 1937** in due articoli (vedi a: *Some examples of statistical methods of research in agriculture and applied biology*. Journal Royal Statist. Soc. Suppl. 4: 137-140; vedi b: *Properties of sufficiency and statistical tests*. Proc. Royal Statist. Soc. Ser. A, 160: 268-282).

Attualmente è il più diffuso e offre **due vantaggi** rispetto ai due test precedenti:

- 1) **i trattamenti a confronto possono contenere un numero di repliche differente;**
- 2) per verificare **la significatività** tra **p** gruppi utilizza la **distribuzione  $\chi^2_{(p-1)}$  con gradi di libertà p-1**, più facilmente reperibile delle distribuzioni specifiche precedenti di Hartley e Cochran.

Con **p** misure di varianze campionarie  $s^2$  che abbiano **gradi di libertà  $n_i$ , eventualmente tra loro diversi**, estratte casualmente da popolazioni distribuite in modo normale, il test approssimato di Bartlett segue **una distribuzione  $C^2_{(p-1)}$**  fondata sul rapporto

$$c_{(p-1)}^2 = \frac{M}{C}$$

dove

- **C** è il **fattore di correzione** proposto successivamente per utilizzare la distribuzione  $c_{(p-1)}^2$ .  
E' uguale a

$$C = 1 + \frac{1}{3 \cdot (p-1)} \cdot \left( \sum \left( \frac{1}{n_i} \right) - \frac{1}{\sum n_i} \right)$$

e risulta un valore prossimo ad 1.

- M è uguale a

$$M = \left( \sum n_i \cdot \ln \bar{s}^2 - \sum n_i \cdot \ln s_i^2 \right)$$

con  $\bar{s}^2$  = media ponderata delle varianze,

data da

$$\bar{s}^2 = \frac{\sum v_i \cdot s_i^2}{\sum v_i}$$

Per il calcolo di **M** (in alcuni testi è indicato con **B**), in diversi autori propongono l'uso del logaritmo a base 10, preferibile alla logaritmo naturale a base **e**;

quindi un altro modo per calcolare **M** è

$$M = 2.30259 \left[ (\log \bar{s}^2) \cdot \left( \sum n_i \right) - \sum n_i \log s_i^2 \right]$$

Questo test per l'omoschedasticità è molto potente, ma solo quando la distribuzione dei dati è normale.

Se la distribuzione dei dati è

- **platicurtica**, il valore della probabilità  $\alpha$  calcolata è più alto di quello reale; (il test è conservativo, meno potente: diventa più difficile rifiutare l'ipotesi nulla e quindi è più facile commettere un errore di II Tipo)
- **leptocurtica**, il valore della probabilità  $\alpha$  calcolata è più basso di quello reale, rovesciando i concetti e la conclusione precedenti.

### Valori critici $R_{(n,p)}$ di Cochran per il confronto simultaneo tra più varianze.

**n** = numero di osservazioni per gruppo, con campioni bilanciati.

**p** = numeri di gruppi o varianze a confronto simultaneo.

$$\alpha = 0.05$$

NUMERO **n** DI OSSERVAZIONI PER GRUPPO

<b>P</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>¥</b>
<b>2</b>	0,9985	0,9750	0,9392	0,9057	0,8772	0,8534	0,8332	0,8159	0,8010	0,5000
<b>3</b>	0,9669	0,8709	0,7977	0,7457	0,7071	0,6771	0,6530	0,6333	0,6167	0,3333
<b>4</b>	0,9065	0,7679	0,6841	0,6287	0,5895	0,5598	0,5365	0,5175	0,5017	0,2500
<b>5</b>	0,8412	0,6838	0,5981	0,5441	0,5065	0,4783	0,4564	0,4387	0,4241	0,2000
<b>6</b>	0,7808	0,6161	0,5321	0,4803	0,4447	0,4184	0,3980	0,3817	0,3682	0,1667
<b>7</b>	0,7271	0,5612	0,4800	0,4307	0,3974	0,3726	0,3535	0,3384	0,3259	0,1429
<b>8</b>	0,6798	0,5157	0,4377	0,3910	0,3595	0,3362	0,3185	0,3043	0,2926	0,1250
<b>9</b>	0,6385	0,4775	0,4027	0,3584	0,3286	0,3067	0,2901	0,2768	0,2659	0,1111
<b>10</b>	0,6020	0,4450	0,3733	0,3311	0,3029	0,2823	0,2666	0,2541	0,2439	0,1000

$$a = 0.01$$

NUMERO **n** DI OSSERVAZIONI PER GRUPPO

<b>P</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>¥</b>
<b>2</b>	0,9999	0,9950	0,9794	0,9586	0,9373	0,9172	0,8988	0,8823	0,8674	0,500
<b>3</b>	0,9933	0,9423	0,8831	0,8335	0,7933	0,7606	0,7335	0,7107	0,6912	0,3333
<b>4</b>	0,9676	0,8643	0,7814	0,7212	0,6761	0,6410	0,6129	0,5897	0,5702	0,2500
<b>5</b>	0,9279	0,7885	0,6957	0,6329	0,5875	0,5531	0,5259	0,5037	0,4854	0,2000
<b>6</b>	0,8828	0,7218	0,6258	0,5635	0,5195	0,4866	0,4608	0,4401	0,4229	0,1667
<b>7</b>	0,8376	0,6644	0,5685	0,5080	0,4659	0,4347	0,4105	0,3911	0,3751	0,1429
<b>8</b>	0,7945	0,6152	0,5209	0,4627	0,4226	0,3932	0,3704	0,3522	0,3373	0,1250
<b>9</b>	0,7544	0,5727	0,4810	0,4251	0,3870	0,3592	0,3378	0,3207	0,3067	0,1111
<b>10</b>	0,7175	0,5358	0,4469	0,3934	0,3572	0,3308	0,3106	0,2945	0,2813	0,1000

Il test può essere applicato su campioni non eccessivamente piccoli, per cui **si richiede che ogni varianza sia calcolata su un campione con almeno 5-6 osservazioni.**

Diversi autori sono molto critici sull'uso dei test per l'omogeneità della varianza. Infatti sono **fortemente alterati dalla non normalità della distribuzione e con pochi dati è impossibile**

verificare se le varie distribuzioni campionarie possano essere ritenute prossime alla normale; inoltre, le varianze campionarie  $s^2$  devono evidenziare differenze molto grandi per risultare significative, poiché i test sono impostati per non rifiutare l'ipotesi nulla in condizioni d'incertezza. Può quindi essere frequente il caso in cui varianze  $s^2$ , anche se rilevanti sotto l'aspetto ecologico od ambientale, non risultano significative ai test per l'omogeneità.

ESEMPIO. Per verificare l'esistenza di differenze nella qualità dell'aria, in 4 zone di una città si è misurata la quantità di solventi aromatici in sospensione.

Z O N E			
I	II	III	IV
190	138	173	198
210	149	164	207
205	128	185	232
208	136	179	184
206	152	188	193

Con le osservazioni riportate nella tabella, si intende verificare se le varianze dei quattro gruppi possono essere ritenute omogenee.

Risposta.

Si calcolano dapprima le 4 varianze,

V A R I A N Z E			
I	II	III	IV
63,20	96,81	92,70	335,70

ricordando che ognuna di esse ha 4 gdl.

L'ipotesi nulla è

$$H_0 : \sigma^2_I = \sigma^2_{II} = \sigma^2_{III} = \sigma^2_{IV}$$

mentre l'ipotesi alternativa  $H_1$  è che almeno una di esse sia diversa.

Con il **metodo di Hartley**, si calcola il rapporto tra la varianza maggiore (335,7) e la varianza minore (63,2): si ottiene un F con indici 4 (numero di gruppi) e 4 (numero di osservazioni entro ogni gruppo meno 1)

$$F_{(4,4)} = \frac{335,70}{63,20} = 5,30$$

che risulta uguale a 5,30.

Per la significatività si confronta il valore calcolato (5,3) con il valore tabulato alla probabilità prefissata per il medesimo numero di gruppi (4) e di gdl entro gruppi (4): per  $\alpha = 0.05$  risulta 20,6.

Con il **metodo di Cochran** si stima un rapporto R

$$R_{5,4} = \frac{335,70}{63,2 + 96,81 + 92,7 + 335,7} = 0,57$$

che per **n** uguale a 5 e **p** uguale a 4 risulta 0,57

Nelle tabelle, il valore critico alla probabilità  $\alpha = 0.01$  risulta uguale a 0,7212 e alla probabilità 0.05 uguale a 0,6287.

Il valore calcolato è inferiore, per cui non è dimostrata una differenza significativa tra le 4 varianze.

Con il **metodo di Bartlett**, si deve dapprima calcolare la varianza media  $\bar{s}^2$

$$\bar{s}^2 = \frac{\sum v_i \cdot s_i^2}{\sum v_i}$$

dividendo la somma delle 4 devianze per la somma dei gradi di libertà:

$$\bar{s}^2 = \frac{4 \cdot 63,2 + 4 \cdot 96,81 + 4 \cdot 92,7 + 4 \cdot 335,7}{16} = \frac{2353,64}{16} = 147,1$$

Successivamente si stimano

- sia il valore di M

$$M = \left( \sum n_i \cdot \ln \bar{s}^2 - \sum n_i \cdot \ln s_i^2 \right)$$

che, con i dati dell'esempio,

$$\begin{aligned} M &= 16 \cdot \ln 147,1 - (4 \cdot \ln 63,2 + 4 \cdot \ln 96,81 + 4 \cdot \ln 92,7 + 4 \cdot \ln 335,7) = \\ &= 16 \times 4,991 - (4 \times 4,146 + 4 \times 4,573 + 4 \times 4,529 + 4 \times 5,186) = \\ &= 79,856 - (16,584 + 18,292 + 18,116 + 20,744) = 79,856 - 73,736 = 6,12 \end{aligned}$$

risulta uguale a **6,12**,

- sia il valore di C

$$C = 1 + \frac{1}{3 \cdot (p-1)} \cdot \left( \sum \left( \frac{1}{n_i} \right) - \frac{1}{\sum n_i} \right)$$

che risulta

$$C = 1 + \frac{1}{3 \cdot 3} \cdot \left( \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} - \frac{1}{16} \right) = 1 + \frac{1}{9} \cdot \left( \frac{4}{4} - \frac{1}{16} \right) =$$
$$= 1 + 0,111 \cdot (1 - 0,0625) = 1 + (0,111 \times 0,9375) = 1 + 0,104 = 1,104$$

uguale a **1,104**.

Il valore del  $c_{(3)}^2$  (chi quadrato con 3 gradi di libertà)

$$c_{(3)}^2 = \frac{M}{C} = \frac{6,12}{1,104} = 5,54$$

è uguale a **5,54**.

Nella tabella dei valori critici alla probabilità  $\alpha = 0,05$  il valore tabulato è 7,81. Il valore calcolato è inferiore: non si può rifiutare l'ipotesi nulla.

Secondo il test, i 4 gruppi a confronto hanno varianze non significativamente diverse.

Per accettare le conclusioni raggiunte mediante l'analisi, restano le perplessità già evidenziate sulla potenza di questi test e quelle che derivano dal numero ridotto di osservazioni per campione, come nel caso dell'esempio utilizzato.

D) Il **test di Levene** è l'estensione a K gruppi del metodo già illustrato per due campioni indipendenti. Ritenuto da vari statistici più robusto rispetto alla non normalità della distribuzione, deve la sua recente diffusione nella pratica statistica soprattutto all'inserimento in alcuni pacchetti statistici. Per la sua applicazione è necessario disporre dei dati originari, in quanto utilizza gli scarti di ogni valore campionario dalla media del suo gruppo.

Esso mette a **confronto gli scarti dei k gruppi**, cioè le differenze ( $d_i$ ) rispetto alle medie di gruppi,

$$d_i = x_i - \bar{x}$$

dopo

- averle elevate al quadrato

$$d_i = (x_i - \bar{x})^2$$

- oppure prese in valore assoluto

$$d_i = |x_i - \bar{x}|$$

per **eliminare i segni negativi**.

Come già dimostrato, **i due metodi forniscono risultati differenti**. Il metodo che utilizza gli scarti in valore assoluto attualmente è il più diffuso; ha una potenza maggiore, cioè una capacità maggiore di risultare significativo.

Per confrontare la varianza di K gruppi (A , B, C),  
con ipotesi nulla

$$H_0: \sigma^2_A = \sigma^2_B = \sigma^2_C$$

ed ipotesi alternativa

$$H_1: \text{non tutte le } \sigma^2 \text{ sono uguali} \quad \text{oppure} \quad H_1: \text{almeno due } \sigma^2 \text{ sono diverse tra loro}$$

la proposta di **Levene consiste nell'applicare alla k serie di scarti** (al quadrato o in valore assoluto) **l'analisi della varianza ad 1 criterio**, nell'assunzione che, **se i valori medi degli scarti risultano significativamente diversi, le k varianze dei dati originali sono diverse**.

Con un linguaggio più tecnico, se, utilizzando **gli scarti dalla media**, si rifiuta l'ipotesi nulla

$$H_0: \mu_A = \mu_B = \mu_C$$

per accettare l'ipotesi alternativa

$$H_1: \text{non tutte le } m \text{ sono uguali} \quad \text{oppure} \quad H_1: \text{almeno due } m \text{ sono diverse tra loro}$$

implicitamente deriva che

sui **dati originali** si rifiuta l'ipotesi nulla

$$H_0: \sigma^2_A = \sigma^2_B = \sigma^2_C$$

per accettare l'ipotesi alternativa

$$H_1: \text{non tutte le } s^2 \text{ sono uguali} \quad \text{oppure} \quad H_1: \text{almeno due } s^2 \text{ sono diverse tra loro}$$

Come nell'analisi della varianza ad un criterio, i gruppi possono avere un numero differente di osservazioni.

ESEMPIO. Riprendendo gli stessi dati dei tre test precedenti sull'omoscedaticità di  $k$  gruppi (dopo aver tolto due valori per costruire campioni non bilanciati)

Z O N E			
I	II	III	IV
190	138	173	198
210	149	164	207
205	128	185	232
208	136	179	184
206	---	188	---

verificare se le varianze dei 4 gruppi sono tra loro statisticamente differenti.

Risposta

Dopo aver calcolato le medie dei 4 gruppi

$$\bar{x}_I = \frac{190 + 210 + 205 + 208 + 206}{5} = \frac{1019}{5} = 203,8$$

$$\bar{x}_{II} = \frac{138 + 149 + 128 + 136}{4} = \frac{551}{4} = 137,75$$

$$\bar{x}_{III} = \frac{173 + 164 + 185 + 179 + 188}{5} = \frac{889}{5} = 177,8$$

$$\bar{x}_{IV} = \frac{198 + 207 + 282 + 184}{4} = \frac{871}{4} = 217,75$$

si calcolano le  $d_i$ , **gli scarti al quadrato**

$$d_i = (x_i - \bar{x})^2$$

ottenendo la tabella seguente

I	II	III	IV
190,44	0,0625	23,04	390,0625
38,44	126,5625	190,44	115,5625
1,44	95,0625	51,84	203,0625
17,64	3,0625	1,44	1139,0625
4,84	---	104,04	---

La verifica su questi scarti dell'ipotesi nulla

$$H_0: \mu_I = \mu_{II} = \mu_{III} = \mu_{IV}$$

corrisponde a quella

$$H_0: \sigma^2_I = \sigma^2_{II} = \sigma^2_{III} = \sigma^2_{IV}$$

sui dati originari.

A questo scopo, dopo i calcoli preliminari,

$$\begin{array}{cccc} \sum x_{ij}^2 = 1.617.742,98 & \sum x_{ij} = 2696,5 & & \\ \sum x_I = 252,8 & \sum x_{II} = 224,75 & \sum x_{III} = 371,2 & \sum x_{IV} = 1847,75 \end{array}$$

si stimano le devianze

$$SQ_{TOT} = 1.617.742,98 - \frac{2696,5^2}{18} = 1.213.793,3$$

$$SQ_{TRA} = \left( \frac{252,8^2}{5} + \frac{224,75^2}{4} + \frac{371,2^2}{5} + \frac{1.867,75^2}{4} \right) - \frac{2.696,5^2}{18} = 502.561,94$$

$$SQ_{ENTRO} = 1.213.793,3 - 502.561,94 = 711.231,36$$

e si completa la tabella dell'ANOVA

	<b>Devianza</b>	<b>Gdl</b>	<b>Varianza</b>
<b>Totale</b>	1,213.793,30	17	---
<b>Tra</b>	502.561,94	3	167.520,65
<b>Entro</b>	711.231,36	14	50.802,22

che conduce alla stima di

$$F_{3,14} = \frac{167.520,65}{50.802,22} = 3,297$$

**F = 3,297** con gdl **3** e **14**.

Il valore critico alla probabilità  $\alpha = 0,05$  è uguale a **3,34**: le 4 varianze a confronto risultano significativamente differenti, con una probabilità  $\alpha$  leggermente inferiore al 5%

Utilizzando gli scarti **in valore assoluto**, che ovviamente presentano una variabilità minore, si ottiene un test più potente. Dalla tabella

Z O N E			
I	II	III	IV
13,8	0,25	4,8	19,75
6,2	11,25	13,8	10,75
1,2	9,75	7,2	14,25
4,2	1,75	1,2	33,75
2,2	---	10,2	---

e dai calcoli preliminari

$$\sum x_{ij}^2 = 2.696,5 \quad \sum x_{ij} = 166,3$$

$$\sum x_I = 27,6 \quad \sum x_{II} = 23,0 \quad \sum x_{III} = 37,2 \quad \sum x_{IV} = 78,5$$

si stimano le devianze

$$SQ_{TOT} = 2.696,5 - \frac{166,3^2}{18} = 2.696,5 - 1.536,43 = 1.160,07$$

$$SQ_{TRA} = \left( \frac{27,6^2}{5} + \frac{23,0^2}{4} + \frac{37,2^2}{5} + \frac{78,5^2}{4} \right) - \frac{166,3^2}{18} = 566,5$$

$$SQ_{ENTRO} = 1.160,07 - 566,5 = 593,57$$

e si completa la tabella dell'ANOVA

	Devianza	Gdl	Varianza
<b>Totale</b>	1.160,07	17	----
<b>Tra</b>	566,50	3	188,83
<b>Entro</b>	593,57	14	42,40

che conduce alla stima di

$$F_{3,14} = \frac{188,83}{42,4} = 4,453$$

**F = 4,453** con gdl **3 e 14**.

Il valore critico con gdl **3 e 14**

- alla probabilità  $\alpha = 0.05$  è uguale a **3,34**
- alla probabilità  $\alpha = 0.025$  è uguale a **4,24**
- alla probabilità  $\alpha = 0.01$  è uguale a **5,35**

Le 4 varianze a confronto risultano significativamente differenti, con una probabilità inferiore al 2,5%

I quattro test proposti sono validi quando **la distribuzione dei dati è normale**; con l'allontanamento da essa, a causa di asimmetria e/o curtosi, tutti subiscono variazioni nella potenza e nella loro attendibilità ma in modo diverso. Per ovviare a questi inconvenienti, spesso è raccomandato il ricorso ai test non parametrici, presentati nei capitoli relativi (vedi indice).

#### **8.4. I CONFRONTI TRA PIU' MEDIE**

Quando con l'analisi della varianza viene rifiutata l'ipotesi nulla, si può essere interessati ad **individuare tra quali medie esista una differenza significativa**. Come nel confronto tra una situazione normale e diversi ambienti inquinati oppure in misure di tossicità con sostanze differenti, **sovente l'attenzione è più rivolta ai confronti specifici tra alcune medie che non ad una valutazione complessiva**, per la quale è quasi ovvio il rifiuto dell'ipotesi nulla.

Con più confronti, aumenta la probabilità di commettere errori di I tipo; in particolare, è errato applicare tante volte il test t di Student, nella convinzione che se un test risulta significativo, sia dimostrato che l'ipotesi nulla (tutte le medie sono uguali) sia da rifiutare.

I confronti possono essere semplici, condotti su coppie di trattamenti, oppure più complessi, tra gruppi diversi di trattamenti. Un'altra distinzione che verrà discussa in seguito è tra i **confronti a priori o pianificati e i confronti a posteriori (detti anche post-hoc) o multipli**. Esistono confronti che sono già programmati nella fase iniziale dell'esperimento, prima della raccolta di dati in natura o dei risultati dell'esperimento in laboratorio; altri che, in carenza di idee sulle differenze possibili tra i diversi gruppi, dopo la significatività del confronto globale evidenziata da un test F, vengono utilizzati per la **ricerca di qualunque confronto significativo, come utile indicazione dell'esistenza di effetti o trattamenti diversi, le cui cause restano da approfondire con ulteriori analisi**.

Il problema fondamentale è che con **p** gruppi, e quindi **con p-1 gradi di libertà nell'ANOVA**, **si possono avere solamente altrettanti confronti a coppie, come scomposizione della devianza tra trattamenti, se si vuole mantenere costante la probabilità  $\alpha$  prescelta**, mentre il numero teorico di

confronti possibili è molto più alto, dato dalle combinazioni di  $p$  elementi 2 a 2. Con 5 gruppi, si hanno 4 gradi di libertà e sono ammessi solo 4 confronti a coppie; tuttavia i confronti possibili 2 a 2 tra coppie di trattamenti ( $C^2_5$ ) sono 10, senza considerare i confronti fra insiemi diversi degli stessi 5 gruppi.

#### 8.4.1 CONFRONTI A PRIORI O PIANIFICATI OD ORTOGONALI

I confronti a priori, chiamati in molti testi anche confronti pianificati od ortogonali (*planned comparisons, orthogonal comparisons*), vengono prestabiliti durante la fase di programmazione dell'esperimento. Con essi è possibile organizzare sia confronti parziali che un **confronto globale**, in modo da analizzare le differenze tra le medie dei gruppi.

Come prosecuzione ed approfondimento dell'analisi della varianza, è possibile la scomposizione della devianza tra trattamenti e dei gdl relativi. Questi metodi presentano alcuni vantaggi:

- **utilizzano tutti i dati,**
- **per la stima dell'errore impiegano la varianza d'errore,**
- **non abbassano il valore di  $\alpha$  per ognuno dei confronti possibili e quindi sono più potenti di quelli non pianificati.**

Trascurando i confronti con due soli gruppi, dove i gradi di libertà della devianza tra trattamenti è 1 solo ed il confronto può essere solo tra le loro 2 medie, nel caso di 3 o più trattamenti è possibile operare con diverse modalità.

Come primo esempio, è possibile citare il caso di 3 gruppi: un controllo ( C ) e due trattamenti ( $A_1$ ,  $A_2$ ). La devianza tra trattamenti ha 2 gdl; se il test F risulta significativo è logica la sua successiva scomposizione in un primo confronto tra il controllo ( C ) contro (*versus*, abbreviato in vs) i due trattamenti ( $A_1 + A_2$ ) ed un secondo confronto tra i due trattamenti ( $A_1$  vs  $A_2$ ).

In un secondo esempio, che considera 4 gruppi, con la sostanza A alla concentrazione 5% e 10% (gruppi  $A_1$  e  $A_2$ ) e con la sostanza B alla concentrazione 8% e 30% si ha un'analisi della varianza con 3 gdl nella devianza tra trattamenti. Essi possono essere scomposti in

- 1 - un primo gdl per il confronto di A ( $A_1 + A_2$ ) contro B ( $B_1 + B_2$ ),
- 2 - un secondo gdl per il confronto di A al 5% ( $A_1$ ) contro A al 10% ( $A_2$ ),
- 3 - un terzo gdl per il confronto di B a concentrazione 8% ( $B_1$ ) contro la concentrazione 30% ( $B_2$ ).

I casi da portare come esempio possono essere numerosi e sensibilmente più complessi; **le scelte sui confronti da effettuare dipendono dalla conoscenza dei fattori che si vogliono sperimentare.**

E' fondamentale comprendere che i confronti che si possono effettuare devono essere tra loro indipendenti od ortogonali; in termini più semplici, significa che ogni confronto non deve fornire informazioni sul risultato di tutti gli altri.

Nell'esempio precedente con i 4 gruppi ( $A_1$ ,  $A_2$  e  $B_1$ ,  $B_2$ ) la significatività del confronto di A contro B non fornisce alcuna informazione di una eventuale significatività né tra  $A_1$  e  $A_2$ , né tra  $B_1$  e  $B_2$ ;

nessuno dei tre confronti fornisce informazioni sulla significatività degli altri due. Se invece si facessero confronti non indipendenti, ad esempio  $A_1$  contro  $B_1$  o  $B_2$ , l'eventuale significatività di A contro B aumenta la probabilità che siano significativi anche gli altri due.

Per una loro corretta impostazione tecnica è utile ricorrere ai **coefficienti polinomiali**.

E' un metodo proposto originariamente anche per semplificare i calcoli ed abbreviare i tempi richiesti dalla scomposizione della devianza tra trattamenti, come verrà dimostrato; ora, con la diffusione dei computer, sono rimasti utili soprattutto per impostare correttamente i confronti da effettuare, limitatamente ai gdl disponibili.

**Si definiscono confronti ortogonali solo quelli in cui**

- **la somma dei coefficienti per riga e**
- **quella dei loro prodotti per colonna sono entrambi uguali a zero.**

**Questi confronti godono della proprietà di rappresentare una corretta scomposizione della devianza tra trattamenti, con 1 gdl per ogni confronto, senza presentare correlazioni tra i risultati.** In una analisi ortogonale, **i differenti effetti sono stimati in modo indipendente, senza mutue interferenze e quindi senza alterazione della probabilità a prefissata di trovare differenze significative.**

Per il caso precedentemente citato dei 3 gruppi, un controllo (C) e due trattamenti ( $A_1$  e  $A_2$ ), si può verificare nella tabella seguente con i **coefficienti polinomiali** come la somma dei valori per riga sia zero e come, nel contempo, sia uguale a zero anche la somma dei prodotti delle colonne:

**Confronti ortogonali tra 3 gruppi (controllo C e 2 trattamenti  $A_1, A_2$ )**

<b>Gruppi</b>	<b>C</b>	<b><math>A_1</math></b>	<b><math>A_2</math></b>	<b>Somma per riga</b>
C contro $A_1 + A_2$	<b>+ 1</b>	<b>- 1/2</b>	<b>- 1/2</b>	<b>0</b>
$A_1$ contro $A_2$	<b>0</b>	<b>+1</b>	<b>-1</b>	<b>0</b>
Prodotti per colonna	+1 x 0	-1/2 x +1	-1/2 x -1	
<b>Totale colonna</b>	<b>0</b>	<b>-1/2</b>	<b>+1/2</b>	<b>0</b>

Se i gruppi fossero 4, un controllo e tre trattamenti, per un confronto del controllo contro gli altri tre, si darebbe valore 1 al controllo e 1/3 ad ogni trattamento.

Per semplificare ulteriormente i calcoli, quasi sempre i pesi attribuiti ad ogni gruppo per un confronto specifico sono cambiati nei corrispondenti numeri interi, moltiplicando i dati della stessa riga per il denominatore, per cui la riga +1, -1/2, -1/2 diventa +2, -1, -1.

La tabella seguente, analoga alla precedente, riporta i coefficienti polinomiali con gli interi

**Confronti ortogonali tra 3 gruppi (controllo C e 2 trattamenti A<sub>1</sub>, A<sub>2</sub>)**

<b>Gruppi</b>	<b>C</b>	<b>A<sub>1</sub></b>	<b>A<sub>2</sub></b>	<b>Somma per riga</b>
C contro A <sub>1</sub> + A <sub>2</sub>	+ 2	- 1	- 1	0
A <sub>1</sub> contro A <sub>2</sub>	0	+1	-1	0
Prodotti per colonna	+2 x 0	-1 x +1	-1 x -1	
<b>Totale colonna</b>	<b>0</b>	<b>-1</b>	<b>+1</b>	<b>0</b>

Il confronto di C contro A<sub>1</sub> + A<sub>2</sub> permette di verificare se il controllo ha effetti diversi rispetto ai due trattamenti, uniti in un gruppo solo; il confronto di A<sub>1</sub> contro A<sub>2</sub> permette di evidenziare se i due trattamenti determinano effetti tra loro differenti.

Nel caso di 4 gruppi, A<sub>1</sub> e A<sub>2</sub> con la stessa sostanza a concentrazioni diverse, B<sub>1</sub> e B<sub>2</sub> con altra sostanza a concentrazioni differenti, i confronti ortogonali diventano 3 come descritti dalla tabella:

**Confronti ortogonali tra 4 gruppi (A<sub>1</sub>+ A<sub>2</sub> contro B<sub>1</sub>+ B<sub>2</sub>)**

<b>Gruppi</b>	<b>A<sub>1</sub></b>	<b>A<sub>2</sub></b>	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>Somma riga</b>
A <sub>1</sub> e A <sub>2</sub> contro B <sub>1</sub> e B <sub>2</sub>	+1	+1	-1	-1	0
A <sub>1</sub> contro A <sub>2</sub>	+1	-1	0	0	0
A <sub>1</sub> contro B <sub>2</sub>	0	0	+1	-1	0
Prodotti per colonna	+1x+1x0	+1x-1x0	-1x0x+1	-1x0x-1	
<b>Totale colonna</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

Nel caso di un controllo (C) e tre trattamenti per valutare l'effetto di tre farmaci, di cui A contenente una principio attivo naturale mentre B<sub>1</sub> e B<sub>2</sub> contenenti 2 differenti prodotti di sintesi, i confronti potrebbero essere come nella tabella successiva.

**Confronti ortogonali tra 4 gruppi (C, A, B<sub>1</sub>, B<sub>2</sub>)**

<b>Gruppi</b>	<b>C</b>	<b>A</b>	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>Totale riga</b>
C contro A + B <sub>1</sub> + B <sub>2</sub>	+1	-1/3	-1/3	-1/3	0
A contro B <sub>1</sub> + B <sub>2</sub>	0	+1	-1/2	-1/2	0
B <sub>1</sub> contro B <sub>2</sub>	0	0	+1	-1	0
Prodotti per colonna	+1 x 0 x 0	-1/3 x +1 x 0	-1/3x-1/2x+1	-1/3x-1/2x-1	
<b>Totale colonna</b>	<b>0</b>	<b>0</b>	<b>+1/6</b>	<b>-1/6</b>	<b>0</b>

Usando, come d'abitudine, gli interi, si ottiene

**Confronti ortogonali tra 4 gruppi (C, A, B<sub>1</sub>, B<sub>2</sub>)**

<b>Gruppi</b>	<b>C</b>	<b>A</b>	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>Somma riga</b>
C contro A + B <sub>1</sub> + B <sub>2</sub>	+3	-1	-1	-1	0
A contro B <sub>1</sub> + B <sub>2</sub>	0	+2	-1	-1	0
B <sub>1</sub> contro B <sub>2</sub>	0	0	+1	-1	0
Prodotti per colonna	+3 x 0 x 0	-1 x +2 x 0	-1 x -1 x +1	-1 x -1 x -1	
<b>Totale colonna</b>	<b>0</b>	<b>0</b>	<b>+1</b>	<b>-1</b>	<b>0</b>

**La legge dell'ortogonalità**, presentata per tutti i confronti, è **valida anche per ogni coppia di confronti**.

Con i **coefficienti polinomiali** è possibile mostrare quando un confronto è errato, non è ortogonale. Per esempio, dopo aver verificato se esiste una differenza significativa del controllo rispetto ai tre farmaci insieme, è errato confrontare ancora il controllo rispetto ai farmaci B<sub>1</sub> e B<sub>2</sub>, sia insieme che separatamente. E' logico pensare che **il primo risultato fornisca informazioni sul secondo**: se tra il controllo e i farmaci risulta una differenza significativa, è maggiormente probabile che risulti significativa anche la differenza tra il controllo e i due farmaci B<sub>1</sub> e B<sub>2</sub>.

I **coefficienti polinomiali di questi due confronti non indipendenti**, presentati nella tabella successiva, infatti risultano

$$(3 \times 2) + (-1 \times 0) + (-1 \times -1) + (-1 \times -1) = 6 + 0 + 1 + 1 = 8$$

diversi da 0.

Nella tabella precedente, risultano indipendenti

- sia il confronto riportato nella riga 1 con quello della riga 2

$$(+3 \times 0) + (-1 \times +2) + (-1 \times -1) + (-1 \times -1) = 0 - 2 + 1 + 1 = 0$$

- sia il confronto riportato nella riga 2 con quello della riga 3

$$(0 \times 0) + (+2 \times 0) + (-1 \times +1) + (-1 \times -1) = 0 + 0 - 1 + 1 = 0$$

- sia il confronto riportato nella riga 1 con quello della riga 3

$$(+3 \times 0) + (-1 \times 0) + (-1 \times +1) + (-1 \times -1) = 0 + 0 - 1 + 1 = 0$$

in quanto la somma dei prodotti dei loro coefficienti è sempre uguale a 0.

#### Confronti ortogonali tra 4 gruppi (C, A, B<sub>1</sub>, B<sub>2</sub>)

Gruppi	C	A	B <sub>1</sub>	B <sub>2</sub>	Somma riga
C contro A + B <sub>1</sub> + B <sub>2</sub>	+3	-1	-1	-1	0
C contro B <sub>1</sub> + B <sub>2</sub>	+2	0	-1	-1	0
Prodotti per colonna	+3 x +2	-1 x 0	-1 x -1	-1 x -1	
Totale colonna	<b>6</b>	<b>0</b>	<b>+1</b>	<b>+1</b>	<b>+8</b>

Nel caso di **più gruppi, con lo stesso numero di osservazioni e con varianze (s<sup>2</sup>) uguali** (questa ultima d'altronde è richiesta dall'analisi della varianza come condizione di validità), con i coefficienti polinomiali è semplice **scomporre la devianza tra trattamenti**.

Per ogni singolo test riportato nelle righe dei coefficienti polinomiali, si devono calcolare

- sia la media generale di tutti i gruppi implicati ( $\bar{\bar{x}}$ ),
- sia la media di ognuno dei due sottogruppi ( $\bar{x}_i$ ) a confronto.

Il valore della devianza, con 1 gdl, è ottenuto dalla formula di quella tra trattamenti già presentata

$$\sum_{k=1}^2 (\bar{x}_{ki} - \bar{\bar{x}})^2 n_i$$

dove n<sub>i</sub> è il numero di osservazioni entro ogni sottogruppo.

L'applicazione è illustrata in modo chiaro nell'esempio successivo.

**ESEMPIO.** In cinque zone di una città, con 6 misure in ognuna è stata rilevata la presenza di solventi aromatici (microgrammi/Nmc a 0° e 1013 mbar): 2 stazioni di rilevazione (A e B) sono state collocate in centro, ai due angoli estremi della piazza principale; la 3<sup>a</sup> stazione di rilevazione (C) ancora in

centro, ma in una piazza secondaria; la 4<sup>a</sup> (D) e la 5<sup>a</sup> (E) in due zone periferiche, ai due estremi della strada principale, che attraversa la città.

I risultati sono stati

<b>ZONE</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
Medie	208,2	199,8	141,0	123,3	119,1
$n_i$	6	6	6	6	6

L'analisi della varianza ha dimostrato che tra le 5 zone esiste complessivamente una differenza significativa e la varianza d'errore ( $s^2_e$ ), con 25 gdl, è risultata uguale a 146,5 (ovviamente, può essere ottenuta come somma delle devianze entro ognuno dei 5 gruppi, con gdl totali 5 x 5).

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>	<b>F</b>
<b>TOTALE</b>	47.301,628	29		
<b>Tra trattamenti o zone</b>	43.639,128	4	10.909,782	74,47
<b>Errore</b>	3.662,500	25	146,500	

Mediante i confronti ortogonali, verificare tra quali zone esiste una differenza significativa nel livello medio d'inquinamento.

Risposta.

Con 5 gruppi e 4 gdl della varianza tra trattamenti, sono possibili 4 confronti ortogonali. Sulla base di quanto noto sulla collocazione delle 5 stazioni di rilevamento, appare logico e senza alternative plausibili lo schema riportato nella tabella successiva:

#### **Confronti ortogonali con relativi punteggi polinomiali**

4 Confronti ortogonali tra le 5 zone	A	B	C	D	E
Centro (A+B+C) vs Periferia (D+E)	+2	+2	+2	-3	-3
Piazza princ. (A+B) vs Piazza sec. (C)	+1	+1	-2	0	0
Piazza principale: A vs B	+1	-1	0	0	0
Periferia: D vs E	0	0	0	+1	-1

Dopo aver verificato che sono tutti confronti tra loro ortogonali

come, ad esempio, il primo ed il secondo

$$(+2 \times +1) + (+2 \times +1) + (+2 \times -2) + (-3 \times 0) + (-3 \times 0) = 2 + 2 - 4 + 0 + 0 = 0$$

si calcolano le **4 devianze relative**:

1 – la prima (centro contro periferia) con

$$\bar{\bar{x}} = 208,2 + 199,8 + 141,0 + 123,3 + 119,1 = 791,4 / 5 = 158,28$$

$$\bar{x}_1 = 208,2 + 199,8 + 141,0 = 549,0 / 3 = 183,0 \quad \text{e} \quad n_1 = 18$$

$$\bar{x}_2 = 123,3 + 119,1 = 242,4 / 2 = 121,2 \quad \text{e} \quad n_2 = 12$$

che risulta

$$(183,0 - 158,28)^2 \cdot 18 + (121,2 - 158,28)^2 \cdot 12 = 10.999,4112 + 16.499,1168 = 27.498,528$$

uguale a 27.498,528;

2 – la seconda devianza (piazza principale contro piazza secondaria del centro) con

$$\bar{\bar{x}} = 208,2 + 199,8 + 141,0 = 549,0 / 3 = 183,0$$

$$\bar{x}_1 = 208,2 + 199,8 = 408,0 / 2 = 204,0 \quad \text{e} \quad n_1 = 12$$

$$\bar{x}_2 = 141,0 \quad \text{e} \quad n_2 = 6$$

che risulta

$$(204 - 183)^2 \cdot 12 + (141 - 183)^2 \cdot 6 = 5.292 + 10.584 = 15.876$$

uguale a 15.876;

3 – la terza devianza (la stazione A contro la B della piazza principale) con

$$\bar{\bar{x}} = 208,2 + 199,8 = 408,0 / 2 = 204,0$$

$$\bar{x}_1 = 208,2 \quad \text{e} \quad n_1 = 6$$

$$\bar{x}_2 = 199,8 \quad \text{e} \quad n_2 = 6$$

che risulta

$$(208,2 - 204)^2 \cdot 6 + (199,8 - 204)^2 \cdot 6 = 105,84 + 105,82 = 211,68$$

uguale a 211,68;

4 – la quarta devianza (la stazione D contro la E delle periferie) con

$$\bar{\bar{x}} = 123,3 + 119,1 = 242,4 / 2 = 121,2$$

$$\bar{x}_1 = 123,3 \quad \text{e} \quad n_1 = 6$$

$$\bar{x}_2 = 119,1 \quad \text{e} \quad n_2 = 6$$

che risulta

$$(123,3 - 121,2)^2 \cdot 6 + (119,1 - 121,2)^2 \cdot 6 = 26,46 + 26,46 = 52,92$$

uguale a 52,92.

(AVVERTENZA: Se i confronti sono ortogonali e non sono stati commessi errori di calcolo, la somma delle 4 devianze calcolate

$$27.498,528 + 15.876 + 211,68 + 52,92 = \mathbf{43.639,128}$$

risulta esattamente uguale alla devianza tra trattamenti, la cui media generale è 158,28

$$(208,2-158,28)^2 \cdot 6 + (199,8-158,28)^2 \cdot 6 + (141-158,28)^2 \cdot 6 + (208,2-158,28)^2 \cdot 6 + (208,2-158,28)^2 \cdot 6 \\ 14.952,0384 + 10.343,4624 + 1.791,5904 + 7.341,6024 + 9.210,4344 = \mathbf{43.639,1280}$$

come la somma dei 4 gdl).

Una volta calcolate le 4 devianze, si possono effettuare 4 test F, ognuno con gdl 1,25 se e solo se il test F della varianza tra trattamenti, con gdl 4, risulta significativo.

Con i dati dell'esempio, poiché il test F della varianza tra trattamenti ( $43.639,1280 / 4 = 10.909,782$ ) risulta

$$F_{4,25} = \frac{10.909,782}{146,5} = 74,47$$

uguale a 74,47

mentre i valori tabulati di  $F_{4,25}$  alla probabilità

$\alpha = 0.05$  risulta uguale a 2,76

$\alpha = 0.01$  risulta uguale a 4,18

si possono fare i 4 test F

$$1 - F_{(1,25)} = \frac{27.498,528}{146,5} = 187,7 \quad \text{altamente significativo}$$

$$2 - F_{(1,25)} = \frac{15.876}{146,5} = 108,37 \quad \text{altamente significativo}$$

$$3 - F_{(1,25)} = \frac{211,68}{146,5} = 1,44 \quad \text{non significativo}$$

$$4 - F_{(1,25)} = \frac{111}{146,5} = < 1 \quad \text{non significativo}$$

Di norma, i risultati sono presentati in una tabella riassuntiva generale, come la seguente:

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>	<b>F</b>
<b>TOTALE</b>	47.301,628	29		
<b>Tra trattamenti o zone</b>	43.639,128	4	10.909,782	74,47
<b>A + B +C vs D + E</b>	27.498,528	1	27.498,528	187,70
<b>A + B vs C</b>	15.876,000	1	15.876,000	108,37
<b>A vs B</b>	211,680	1	211,680	1,44
<b>D vs E</b>	52,920	1	52,920	<1
<b>Errore</b>	3.662,500	25	146,500	

con le probabilità relative, riportate in una ulteriore colonna di fianco ai valori di F, quando effettuate dal computer con i programmi a maggior diffusione internazionale.

Da questi risultati è possibile trarre, in modo esplicito, le conclusioni a carattere ambientale sui valori d'inquinamento rilevati nel campionamento delle 5 zone:

- 1 tra le cinque zone, le medie aritmetiche dell'inquinamento da solventi aromatici hanno una differenza altamente significativa;
- 2 tale significatività è imputabile soprattutto alla differenza tra le 3 stazioni collocate in centro e le 2 situate in zone periferiche;
- 3 è altamente significativa anche la differenza tra le 2 stazioni collocate nella piazza principale e la stazione collocata in una piazza secondaria;
- 4 non esiste una differenza significativa tra le due stazioni di rilevazione situate a i due estremi della stessa piazza centrale;
- 5 non esiste alcuna differenza tra i valori medi delle due zone periferiche.

Quando i gruppi hanno un numero diverso di osservazioni, il confronto tra le medie non risulta più omogeneo: ogni media avrebbe un intervallo fiduciale diverso, ma i gradi di libertà di ogni F restano identici. Se le differenze nelle dimensioni campionarie dei vari gruppi non sono troppo differenti (ma resta la difficoltà di decidere quando lo siano), fino a poco tempo fa alcuni testi accettavano ugualmente il confronto con il metodo appena illustrato.

#### 8.4.2 TEST PER CONFRONTI MULTIPLI O A POSTERIORI: BONFERRONI, SNK DI NEUMAN-KEULS, LSD DI FISHER, STEP-UP DI WELSCH, HSD DI TUKEY, SCHEFFE', DUNNETT, DUNCAN

Se, nel confronto tra le medie di  $p$  gruppi, con il test  $F$  è stata rifiutata l'ipotesi nulla

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_p$$

si pone il problema di verificare tra quali esista una differenza significativa.

A questo scopo, i confronti a priori propongono i metodi migliori, poiché hanno una probabilità maggiore di evidenziare le significatività, quando l'ipotesi nulla è falsa. Ma con la diffusione dei computer, i confronti multipli o a posteriori hanno acquisito un rilevante vantaggio pratico: da vari anni sono sempre riportati in molti programmi informatici insieme con il test  $F$  e possono essere applicati con facilità. Sono quindi usati con frequenza maggiore, anche se la preferenza ad essi accordata appare illogica, ai fini di una corretta analisi statistica, che deve sempre preferire i test più potenti. Solo recentemente, anche i confronti a priori sono stati inseriti nei programmi informatici a maggior diffusione e perciò resi di uso semplice.

Quando è possibile scegliere, **la preferenza è sempre da attribuire ai test pre-pianificati**; tuttavia, in vari casi, i test a posteriori sono necessari.

**I confronti multipli o a posteriori** (detti anche **post-hoc**, **confronti non prestabiliti** o **non pianificati** e più raramente *incidental comparisons*) sono necessari quando non è possibile programmare i confronti a priori, al momento del disegno sperimentale, per carenza d'informazione. Quando i trattamenti non possono essere classificati in gruppi (come nei confronti ortogonali), che spieghino più utilmente di altri la differenza complessiva, già evidenziata dall'analisi della varianza, rimane solo la possibilità di **effettuare tutti i possibili confronti tra medie, alla ricerca di quelle differenze significative che hanno determinato la significatività totale**. E' detta "procedura di dragaggio" e serve per **individuare differenze** da studiare successivamente in modo più approfondito, con analisi ecologiche, chimiche, ambientali, **alla ricerca delle cause possibili**. Le ipotesi di lavoro non sono dedotte dalla conoscenza delle leggi della natura, ma partono dalle differenze osservate, nella convinzione che, se esistono, devono pure avere una causa. Il rischio di tale approccio è di "inventarsi" comunque una spiegazione nell'ambito della disciplina, giustificata esternamente dalla significatività statistica.

Per effettuare i confronti multipli a posteriori sono stati proposti diversi metodi, che derivano essenzialmente dal test  $t$  di Student per ipotesi bilaterali e dagli intervalli fiduciali. Questi metodi hanno lo scopo di ridurre la probabilità di commettere un **errore di I tipo** (la probabilità  $\alpha$  di trovare una differenza significativa, quando in realtà essa non esiste). Infatti, il tasso d'errore di I tipo che è

possibile commettere per un confronto tra due medie (con termine tecnico, chiamato tasso d'errore *comparison-wise*) all'aumentare del numero di confronti determina un tasso d'errore di I tipo per tutto l'esperimento (chiamato tasso d'errore *experiment-wise*) notevolmente maggiore.

La differenza tra *comparisonwise* e *experimentwise* può essere spiegata con un esempio semplice.

Si supponga che uno studente debba sostenere un esame su un testo di 1000 pagine e che egli ne abbia studiate 950:

- per **ogni domanda** (*comparisonwise*) egli ha una probabilità di 0,95 che cada su un argomento che ha studiato e quindi una probabilità di 0,05 su un argomento che non ha studiato;
- ma se l'esame prevede 10 domande, la probabilità che **tutto l'esame** (*experimentwise*) verta sulle 950 pagine che ha studiato è  $(0,95)^{10} = 0,5987$ ; quindi, la probabilità che almeno una domanda cada nelle pagine tralasciate è molto maggiore di 0,05 e pari a  $0,4013 (1 - 0,95^{10})$ .

Il problema diventa quante pagine può tralasciare, per non aumentare troppo la probabilità di essere respinto?

Anche nella scelta del test più adatto ai confronti multipli, si deve dare la preferenza a **quello più potente**; come spesso desiderato dal ricercatore, a quello che permette di **rifiutare l'ipotesi nulla in un singolo confronto**. Tuttavia, occorre anche **mantenere costante la probabilità a prescelta per tutto l'esperimento**, cioè la **protezione contro l'errore di I tipo**.

Nella scelta del test a posteriori, quindi è necessario un compromesso tra

- **comparisonwise**, collegata alla potenza del singolo test e
- **experimentwise**, collegata al principio di cautela o protezione di tutta la serie di test.

La ricerca del difficile equilibrio tra le esigenze contrastanti della potenza e della protezione, per il quale non è ancora stato trovata una soluzione universalmente condivisa, ha determinato una molteplicità di proposte, in funzione delle varie situazioni sperimentali e dei diversi rischi che un ricercatore intende correre; di conseguenza, i confronti multipli sono tra gli argomenti ancora più dibattuti ed in maggiore evoluzione della statistica univariata.

Il primo metodo di protezione, quindi per non innalzare troppo la probabilità d'errore di I tipo dell'*experimentwise*, suggerito da tutti i testi moderni, è **la prassi di**

- **fare precedere ai confronti multipli un'ANOVA e di**
- **effettuare i confronti a posteriori solo quando con essa si è rifiutata l'ipotesi nulla.**

Con questo accorgimento, si vuole evitare che anche un solo confronto tra due medie risulti significativo, quando l'analisi della varianza su tutti i dati non ha permesso di rifiutare l'ipotesi nulla.

Questo contrasto tra conclusioni è possibile, poiché i due test utilizzano probabilità a non identiche.

Tra i confronti multipli citati in letteratura, sia per l'importanza dei concetti, sia per la loro diffusione nei testi e nei programmi informatici, sono da ricordare:

- 1 – il **principio e il test t di Bonferroni**,
- 2 – la **procedura LSD (Least Significant Difference) di Fisher**,
- 3 - il **test di Student-Newman-Keuls** spesso citato come **test SNK o test Q**,
- 4 - la **procedura step-up di Welsch**, fondato sul **MSR (Minimum Significant Range)**,
- 5 - il **test di Tukey o procedura HSD (Honestly Significant Difference)**,
- 6 - il **test di Scheffé con l'estensione di Gabriel**,
- 7 - il **test di Dunnett** per confronti di vari trattamenti con un controllo,
- 8 - il **campo di variazione multiplo di Duncan**.

Informazioni più dettagliate su questi metodi possono essere rintracciate negli articoli:

- Newman D. (1939) *The distribution of the range in samples from a normal population expressed in terms of an independent estimate of the standard deviation*. Biometrika, 31, 20.
- Fisher R. A. (1948) *The design of experiments*. Edinburgh, Oliver and Boyd.
- Keuls M (1952) *The use of the Studentized range in connection with an analysis variance*. Euphytica 1, 112.
- Scheffé H (1953) *A method for judging all contrasts in the analysis of variance*. Biometrika 40, 87.
- Duncan D. B. (1955) *Multiple range and multiple F tests*. Biometrics 11,1.
- Dunnett C.W. (1964) *New tables for multiple comparisons with a control*. Biometrics 20, 482.
- Gabriel K. R. (1964) *A procedure for testing the homogeneity of all sets of means in analysis of variance*. Biometrics 20: 459-477.
- Welsch R. E. (1977) *Stepwise multiple comparison procedures*, J. Amer. Stat. Ass. 72, 566-575.

Una presentazione ampia e relativamente recente è fornita dall'ultima edizione (1995, third edition) di *Biometry* di Sokal e Rohlf (W. H. Freeman and Company, New York) pp. 240-261.

1) L'italiano **Bonferroni** è stato il primo ad affrontare l'argomento in modo operativo. Affermò che, **per effettuare p volte il test t di Student mantenendo costante la probabilità totale  $\alpha_T$  (experimentwise), la probabilità  $\alpha$  di ogni confronto (comparisonwise) deve essere minore di  $\alpha_T/p$** .

Tale relazione, detta **diseguaglianza di Bonferroni**, può essere scritta come

$$\alpha < \alpha_T / p$$

Per esempio, quando con 3 confronti la probabilità totale  $\alpha_T$  di commettere un errore di I tipo non deve essere superiore a 0.05, la probabilità  $\alpha$  di ogni singolo confronto deve essere minore di **0.0166** (0.05/3); se i confronti fossero 4, la probabilità  $\alpha$  di ogni confronto non deve superare **0.0125** (0.05/4).

In realtà, come sarà più avanti approfondito, per una stima più accurata della probabilità **comparisonwise** sulla base di quella dell'**experimentwise**, è utile ricordare che

- la probabilità d'errore complessivo ( $\alpha_T$ ) è legata
- alla probabilità di errore di ogni confronto ( $\alpha$ ) e
- al numero di confronti da effettuare ( $p$ )

secondo la relazione esponenziale

$$\alpha_T = 1 - (1 - \alpha)^p$$

A causa di questa approssimazione nella stima della probabilità, la procedura proposta da Bonferroni è ritenuta sostanzialmente accettabile quando si effettuano pochi confronti, perché le differenze tra le due stime sono minime. **Quando i confronti superano il numero di 8-10, il valore di  $\alpha$  stimato per ognuno di essi diventa troppo piccolo; di conseguenza, il metodo è ritenuto eccessivamente cautelativo.**

Con il metodo di Bonferroni, per il confronto tra due medie si utilizza una **formula analoga al t di Student per 2 campioni indipendenti**,

$$t_{(\text{Bonferroni})}(\alpha_T, p, n) = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{s_e^2 \cdot \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

dove

$\alpha_T$  = la probabilità prefissata globale per tutti i confronti (0.05 o 0.01),

$p$  = il numero di confronti che si intendono eseguire

$n$  = sono i gdl della varianza d'errore  $s_e^2$  utilizzata.

Se il confronto è più generale, volendo valutare se la differenza tra due medie è maggiore di una quantità prefissata ( $m_A - m_B$ ), in modo analogo al test  $t$  di Student per 2 campioni indipendenti, la formula precedente è scritta come

$$t_{(\text{Bonferroni})}(\alpha_T, p, n) = \frac{(\bar{X}_A - \bar{X}_B) - (m_A - m_B)}{\sqrt{s_e^2 \cdot \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

Rispetto al test  $t$  di Student per due campioni indipendenti, questi metodi offrono i vantaggi di

- utilizzare la varianza d'errore  $s_e^2$  calcolata con l'ANOVA tra tutti i gruppi, al posto della varianza associata  $s_p^2$  dei due soli gruppi a confronto;
- usare i gradi di libertà della varianza d'errore  $s_e^2$  (n) per la scelta del valore di **t**, al posto di quelli derivati solo dal numero dei dati presenti nei due gruppi a confronto ( $n_A - 1 + n_B - 1$ ).

Nel **caso di 2 campioni con lo stesso numero n d'osservazioni** o repliche (detti **campioni bilanciati**), il valore del **t**<sub>(Bonferroni)</sub> è più rapidamente calcolato con la formula equivalente

$$t_{(\text{Bonferroni})}(\alpha_T, p, n) = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{2s_e^2}{n}}}$$

Si evidenzia una differenza significativa tra ogni coppia di medie alla probabilità totale  $\alpha_T$  prefissata, quando il valore calcolato supera il valore critico riportato nella tabella successiva.

E' possibile ricavare il valore critico del **t** anche da una tabella dettagliata dei valori **F** di Fisher, per la nota relazione

$$t_{(\alpha;n)} = \sqrt{F_{(\alpha;1,n)}}$$

Tuttavia per **t** ed **F** esiste un problema pratico: non sempre questi valori sono disponibili alla probabilità  $\alpha$  richiesta.

Per esempio, con 3 confronti alla probabilità complessiva  $\alpha_T = 0.05$  occorrerebbe disporre di una tabella che fornisce il valore di **t** o di **F** alla probabilità  $\alpha = 0.0167$ .

E' semplice ottenere i valori del **t** di Bonferroni solamente quando il numero di confronti è 5 oppure 10 o 20, poiché con  $\alpha_T$  uguale a 0.05 (experimentwise) la probabilità  $\alpha$  di ogni confronto (comparisonwise) diventa rispettivamente 0.01 oppure 0.005 o 0.001, tutti valori riportati con frequenza sulle tavole sinottiche.

Di conseguenza, sono fornite tabelle specifiche che, come indici, utilizzano

- il numero **p** di confronti (che con **k** medie sono **k(k-1) / 2**),
- i gradi di libertà della varianza d'errore (n).

### Valori critici del test t di Bonferroni

**p** = numero di confronti simultanei

**gdl** o **n** = gradi di libertà della varianza d'errore

$$a_T = 0.05$$

NUMERO **p** DI CONFRONTI SIMULTANEI

n	2	3	4	5	6	7	8	9	10
5	3,17	3,54	3,81	4,04	4,22	4,38	4,53	4,66	4,78
7	2,84	3,13	3,34	3,50	3,64	3,76	3,86	3,95	4,03
10	2,64	2,87	3,04	3,17	3,28	3,37	3,45	3,52	3,58
12	2,56	2,78	2,94	3,06	3,15	3,24	3,31	3,37	3,43
15	2,49	2,69	2,84	2,95	3,04	3,11	3,18	3,24	3,29
20	2,42	2,61	2,75	2,85	2,93	3,00	3,06	3,11	3,16
24	2,39	2,58	2,70	2,80	2,88	2,94	3,00	3,05	3,09
30	2,36	2,54	2,66	2,75	2,83	2,89	2,94	2,99	3,03
40	2,33	2,50	2,62	2,71	2,78	2,84	2,89	2,93	2,97
60	2,30	2,47	2,58	2,66	2,73	2,79	2,84	2,88	2,92
120	2,27	2,43	2,54	2,62	2,68	2,74	2,79	2,83	2,86
∞	2,24	2,39	2,50	2,58	2,64	2,69	2,74	2,77	2,81

$$a_T = 0.01$$

NUMERO **p** DI CONFRONTI SIMULTANEI

n	2	3	4	5	6	7	8	9	10
5	4,78	5,25	5,60	5,89	6,15	6,36	6,56	6,70	6,86
7	4,03	4,36	4,59	4,78	4,95	5,09	5,21	5,31	5,40
10	3,58	3,83	4,01	4,15	4,27	4,37	4,45	4,53	4,59
12	3,43	3,65	3,80	3,93	4,04	4,13	4,20	4,26	4,32
15	3,29	3,48	3,62	3,74	3,82	3,90	3,97	4,02	4,07
20	3,16	3,33	3,46	3,55	3,63	3,70	3,76	3,80	3,85
24	3,09	3,26	3,38	3,47	3,54	3,61	3,66	3,70	3,74
30	3,03	3,19	3,30	3,39	3,46	3,52	3,57	3,61	3,65
40	2,97	3,12	3,23	3,31	3,38	3,43	3,48	3,51	3,55
60	2,92	3,06	3,16	3,24	3,30	3,34	3,39	3,42	3,46
120	2,86	2,99	3,09	3,16	3,22	3,27	3,31	3,34	3,37
∞	2,81	2,93	3,02	3,09	3,16	3,19	3,24	3,26	3,29

**Quando i campioni non sono delle stesse dimensioni n**, più recentemente vari autori hanno aggiunto una **ulteriore cautela**.

Nella formula generale precedente, al posto della correzione per il bilanciamento

$$\frac{1}{n_1} + \frac{1}{n_2}$$

propongono di **utilizzare la formula per due campioni bilanciati**

$$t_{(\text{Bonferroni})} (a, p, n) = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{2s_e^2}{n}}}$$

con **n uguale al valore minore tra n<sub>1</sub> e n<sub>2</sub>**

E' ovvio che non si supera la probabilità experimentwise prefissata, ma si perde in potenza nella comparisonwise.

ESEMPIO. Con le stesse misure d'inquinamento (utilizzate nell'esempio precedente per i confronti a priori) rilevate in 5 zone, delle quali sono stati riportati le medie e il numero di osservazioni:

<b>ZONE</b>	A	B	C	D	E
<b>Medie</b>	208,2	199,8	141,0	123,3	119,1
<b>n<sub>1</sub></b>	6	5	6	6	7

verificare con il test **t** di Bonferroni tra quali medie esiste una differenza significativa.

Risposta.

Vari autori affermano che i campioni dovrebbero essere bilanciati. Altri sostengono che, con leggere differenze nel numero di osservazioni, è lecito l'uso di questo test, ricorrendo alla formula generale che considera il diverso numero di osservazioni per gruppo.

Con 5 medie, si hanno 10 differenze, che possono essere presentate in modo chiaro ed utile in una tabella con tutti i confronti:

Confronti	Medie	Differenze
A vs B	208,2 - 199,8	8,4
A vs C	208,2 - 141,4	66,8
A vs D	208,2 - 123,3	84,9
A vs E	208,8 - 119,1	89,7
B vs C	199,8 - 141,4	58,4
B vs D	199,8 - 123,3	76,5
B vs E	199,8 - 119,1	80,7
C vs D	141,4 - 123,3	18,1
C vs E	141,4 - 119,1	22,3
D vs E	123,3 - 119,1	4,2

**Le differenze sono da considerare in valore assoluto, in quanto i confronti multipli comportano solo test bilaterali.**

(RICORDARE: nell'analisi della varianza con i **5** gruppi è stata rifiutata l'ipotesi nulla e che la varianza d'errore  $s_e^2$  è risultata uguale a **146,5** con **25** gdl.

Per ogni confronto si calcola il valore del  $t_{(\text{Bonferroni})}$  e si confronta il risultato con i valori critici riportati nella tabella.

Per esempio, **A vs B** diventa

$$t_{(\text{Bonferroni})} = \frac{208,2 - 199,8}{\sqrt{146,5 \cdot \left(\frac{1}{6} + \frac{1}{5}\right)}} = \frac{8,4}{\sqrt{146,5 \cdot 0,367}} = \frac{8,4}{\sqrt{53,766}} = \frac{8,4}{7,33} = 1,14$$

Esso fornisce un valore di **t = 1,14** per un numero di confronti **p = 10** e **gdl = 25**.

Quando il numero esatto di gdl non è riportato nella tabella dei valori critici, per trovare il valore critico nella tabella si utilizzano **i gdl immediatamente inferiori** (24 nella tabella), **in quanto rappresenta la scelta più cautelativa**.

Per **p = 10** e alla probabilità complessiva **a = 0.05** il valore critico riportato è **3,09**.

Il valore calcolato (1,14) è inferiore: non si può rifiutare l'ipotesi nulla e quindi la media di A e quella di B non differiscono in modo significativo.

Il confronto **A vs D** diventa

$$t_{(\text{Bonferroni})} = \frac{208,2 - 123,3}{\sqrt{146,5 \cdot \left(\frac{1}{6} + \frac{1}{6}\right)}} = \frac{84,9}{\sqrt{146,5 \cdot 0,334}} = \frac{84,9}{\sqrt{48,931}} = \frac{84,9}{6,995} = 12,13$$

e stima un valore del **t = 12,13** da confrontare sempre con il valore critico di 3,09. Infatti sono invariati sia la probabilità  $\alpha$  totale, sia il numero **p** di confronti, sia i **gdl** della varianza d'errore.

Alla probabilità **a = 0.01** il valore critico, sempre per **p = 10** e **gdl = 24**, risulta uguale a 3,74.

Di conseguenza, la differenza di questo confronto (208,2 e 123,3) risulta significativa: le due medie (A vs D) differiscono tra loro con probabilità  $\alpha < 0.01$ .

Poiché le ultime due medie a confronto (A vs D) hanno lo stesso numero d'osservazioni (6), si può ottenere lo stesso risultato più rapidamente con

$$t_{(\text{Bonferroni})} = \frac{208,2 - 123,3}{\sqrt{\frac{2 \cdot 146,5}{6}}} = \frac{84,9}{48,83} = \frac{84,9}{6,998} = 12,13$$

2) In vari testi di lingua inglese, **l'idea di condurre tutti i possibili confronti tra coppie di medie è attribuita a Fisher**, per la sua presentazione nel volume *The design of experiments*, edito da Oliver and Boyd di Edinburg nel 1948. Chiamata in inglese **Least Significant Difference** e tradotta in italiano come **Differenza Minima Significativa** (meno spesso, in modo letterale, come **Differenza Meno Significativa**), è abbreviata in **LSD**.

Effettuando un test **t di Student** per ogni coppia di medie ( $\bar{X}_A$  e  $\bar{X}_B$ ) con lo **stesso numero di repliche**, sono significative tutte le differenze **D** (con  $D = \bar{X}_A - \bar{X}_B$ ) che in valore assoluto superano la quantità **LSD**

$$\text{LSD} = t_{(\alpha/2, n)} \times \sqrt{\frac{2s_e^2}{n}}$$

dove

$t_{\alpha/2}$  = percentile con probabilità  $\alpha/2$  della distribuzione t

n = gdl della varianza d'errore stimata con l'ANOVA.

Come già discusso per il **t** di Bonferroni, per confronti tra tutte le coppie diventano ancor più importanti i **rapporti tra il valore di probabilità del comparison e quello dell'experiment-wise**. Con questo metodo è consigliata l'applicazione del **principio di cautela** già ricordato, cioè di **effettuare i confronti a coppie solo quando l'analisi della varianza su tutti i gruppi è risultata significativa**.

Con la generazione di numeri casuali, S. G. Carmer e M. R. Swanson nel 1973 (con l'articolo *An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods*, pubblicato su JASA, n. 68, pp. 66-74) hanno dimostrato che questa precauzione (procedere ai confronti multipli solo dopo aver rifiutato l'ipotesi nulla con l'ANOVA) è **una cautela efficiente contro gli errori di I tipo**: permette di non superare la probabilità experimentwise prefissata.

ESEMPIO. Con gli stessi dati dell'esempio precedente

ZONE	A	B	C	D	E
Medie	208,2	199,8	141,0	123,3	119,1
$n_i$	6	5	6	6	7

in cui la varianza d'errore  $s_e^2$  è risultata uguale a **146,5** con **25** gdl,

- assumendo  $n = 5$  come dimensione comune a tutti i gruppi (**scelta cautelativa**) e che

- alla probabilità  $\alpha = 0.05$  il valore di  $t$  per 25 gdl è uguale a **2,06**

la Differenza Minima Significativa (**LSD**)

$$\mathbf{LSD} = 2,06 \cdot \sqrt{\frac{146,5}{5}} = \mathbf{11,15}$$

risulta uguale a 11,15.

E' significativa qualsiasi differenza tra le 5 medie che superi questa quantità.

Per evitare una scelta così prudentiale come quella di utilizzare il valore minore, quando i campioni non hanno lo stesso numero di osservazioni, alcuni testi suggeriscono il metodo della **interpolazione armonica**. Quando i  $p$  gruppi sono di dimensioni ( $n_i$ ) non troppo differenti, è possibile stimare un valore  $\hat{n}$  corretto, dato dal rapporto

$$\hat{n} = \frac{p}{\frac{1}{n_1} + \frac{1}{n_2} + \dots + \frac{1}{n_p}}$$

Con i dati dell'esempio,

$$\hat{n} = \frac{5}{\frac{1}{6} + \frac{1}{5} + \frac{1}{6} + \frac{1}{6} + \frac{1}{7}} = 5,93$$

$\hat{n}$  risulta uguale a 5,93.

Si osservi che il valore ottenuto è inferiore alla media aritmetica di 6.

**Con una distribuzione non bilanciata, la quantità d'informazione media è inferiore a quella di una distribuzione bilanciata, che ovviamente abbia lo stesso numero totale di osservazioni.**

Con 5,93 al posto del 5 precedente (ritenuti da molti troppo cautelativo), si ottiene un errore standard minore e quindi un valore di **LSD** minore. Il test diventa più potente, con maggiori probabilità di evidenziare differenze significative.

Con gli stessi dati dell'esempio precedente,

si ottiene

$$\mathbf{LSD} = 2,06 \cdot \sqrt{\frac{146,5}{5,93}} = \mathbf{10,24}$$

una Differenza Minima Significativa pari a 10,24.

Per fare questi confronti tra coppie di medie, è stato proposto di agire in modo organizzato; quindi

- **disporre le varie medie in ordine di rango** (non importa se dalla minore alla maggiore o viceversa)

- **iniziare il confronto tra le due più estreme**, che quasi sempre risulterà significativo se l'ANOVA è risultata significativa.

Il primo confronto deve avvenire tra le due medie poste agli estremi; il secondo confronto tra una delle due medie agli estremi e quella che fornisce la differenza maggiore tra tutte le altre, dopo la precedente. Se anche questa risulta significativa si procede al terzo confronto, nel quale si riduce ancora la differenza tra le due medie interessate.

Da questa metodologia (chiamata *step-up procedure* o *stepwise approach*), sempre per evitare errori di I tipo, è stata derivata una **terza cautela** (dopo l'analisi della varianza prima dei confronti multipli e l'uso del numero di dati del campione minore per stimare la differenza minima significativa). Essa consiste nel

- **non considerare come significativamente differenti due medie, quando sono comprese entro due già ritenute non significative.**

Con un valore di **LSD** pari a 10,24 (vedi ultimo calcolo) si deduce che la media del gruppo

- A (208,2) risulta non significativamente diversa da B (199,8)
- D (123,3) non significativamente diversa da E (119,1).

I risultati di questi confronti a coppie sono spesso **rappresentati in modo grafico**, con la **convenzione** di congiungere quelle medie che non sono significativamente diverse tra loro.

Con i dati dell'esempio, si ottiene

A    B    C    D    E

Un'**altra convenzione grafica**, diffusa e di facile comprensione, consiste nel riportare le medie

208,2	199,8	141,0	123,3	119,1
A	A	B	C	C

e nell'indicare con la stessa lettera quelle che tra loro non sono significativamente differenti.

Spesso viene usata una **terza convenzione grafica**, data dalla combinazione di queste due tecniche, quale

208,2	199,8	141,0	123,3	119,1
A A A A	B	C C C C		

In situazione più complesse quando si hanno molte medie che differiscono tra loro per quantità ridotte, si determinano linee o serie di lettere a livelli diversi,

quali

A    B    C    D    E

-----  
-----

Alla informazione precedente (la media di A non diversa da quella di B; la media di D non diversa da quella di E), in questo caso è aggiunta l'informazione di una differenza non significativa tra le medie C-D e tra le medie B-C.

3) Il **test di Student-Neuman-Keuls** o **test SNK**, citato anche come **test di Neuman-Keuls** o **test Q** (oppure **q studentizzato**, detto pure delle **inferenze simultanee** di Newman (1939) e Keuls (1952), utilizza una stima più precisa della probabilità  $\alpha$  per ogni singolo confronto. Bonferroni ha avuto il merito di affrontare il problema dei confronti multipli in modo corretto e generale; ma i valori critici calcolati si sono dimostrati troppo cautelativi, soprattutto all'aumentare del numero di confronti. Una stima più corretta è quella che alcuni autori attribuiscono a **Dunn-Sidak** (vedi: Ury H. K. 1976, *A comparisons of four procedures for multipkle comparisons among means (pairwise contrasts) for arbitrary sample size*. Technometrics 18: 89-97). Con  $p$  confronti, la probabilità  $\alpha$  da assegnare ad ognuno di essi è

$$\alpha = 1 - (1 - \alpha_T)^{1/p}$$

Per esempio, con  $\alpha_T$  uguale a 0.05 e

-  $p$  uguale a 5,

la probabilità  $\alpha$  di ogni confronto non è uguale a 0.01 (0.05/5) ma

$$\alpha = 1 - 0.95^{1/5} = 1 - 0.98979 = 0.01021$$

è uguale a **0.01021**,

con una differenza relativa, rispetto alla stima del Bonferroni del 2,1 per cento;

- con  $p$  uguale a 10, non è più uguale a 0.005 (0.05/10) ma

$$\alpha = 1 - 0.95^{1/10} = 1 - 0.99488 = 0.00512$$

è uguale a **0.00512**, con un aumento relativo del 2,4 per cento rispetto alla stima prudentiale o cautelativa del Bonferroni.

Il confronto tra coppie di medie (A e B) avviene con una formula simile a quella del test  $t$  di Bonferroni

$$Q_{(\alpha_T, p, n)} = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_e^2}{2} \cdot \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

dove, con la consueta simbologia,

$\alpha_T$  = probabilità totale o complessiva di commettere l'errore di I tipo (di norma, 0.05 o 0.01),

$p$  = numero di confronti che si intendono effettuare,

$n$  = gradi di libertà della varianza d'errore del test F tra tutti i gruppi,

$\bar{x}_A - \bar{x}_B =$  differenza tra le due medie a confronto,

$s_e^2 =$  varianza d'errore ottenuta con il test **F** tra tutti i gruppi ,

$n_A$  e  $n_B =$  numero di osservazioni nei due gruppi A e B a confronto.

(Il **2**, riportato al denominatore, **non rappresenta una differenza reale dal test precedente**: è solamente un valore inglobato in quello di **Q**, per una scelta dei due autori).

Con la **procedura stepwise** prima descritta, nella metodologia **SNK** il valore di **Q studentizzato** dipende da quello di **p**, il numero di medie a confronto. Se, come nell'esempio precedente, le medie sono 5, per il confronto della media maggiore con quella minore il valore di **p** è uguale a 5. Il secondo passo prevede il confronto tra la media maggiore e la seconda minore, mentre il valore di **p** diviene 4. Quando si confrontano la maggiore con quella più vicina, la seconda in ordine decrescente, il valore di **p** è uguale a 2. Questa stima è fatta contando i "passi": due medie vicine distano 2 passi; due medie con un'altra intermedia distano 3 passi; in un gruppo di 5 medie, la minore e la maggiore distano 5 passi.

Con 5 medie ordinate in modo crescente (vedi tabella sottostante, in cui 5<sup>a</sup> indica la media maggiore e 1<sup>a</sup> la minore), i confronti da effettuare sono 10, con l'ordine ed il relativo valore dell'indice **p** riportato nella tabella:

Ordine	Confronto	P
1	5 <sup>a</sup> vs 1 <sup>a</sup>	5
2	5 <sup>a</sup> vs 2 <sup>a</sup>	4
3	5 <sup>a</sup> vs 3 <sup>a</sup>	3
4	5 <sup>a</sup> vs 4 <sup>a</sup>	2
5	4 <sup>a</sup> vs 1 <sup>a</sup>	4
6	4 <sup>a</sup> vs 2 <sup>a</sup>	3
7	4 <sup>a</sup> vs 3 <sup>a</sup>	2
8	3 <sup>a</sup> vs 1 <sup>a</sup>	3
9	3 <sup>a</sup> vs 2 <sup>a</sup>	2
10	2 <sup>a</sup> vs 1 <sup>a</sup>	2

Quando le due medie a confronto hanno lo stesso numero **n** di osservazioni, la formula può essere semplificata in

$$Q_{(\alpha_T, p, v)} = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_e^2}{n}}}$$

Sempre come nel test t di Bonferroni, quando le due medie hanno un numero differente **n** di osservazioni, è possibile utilizzare la frequenza minore.

Ricorrendo ai confronti multipli come proposto da Fisher, quando i gruppi a confronto hanno lo stesso numero di dati o frequenze simili, vari autori hanno generalizzato tale procedura, proponendo la tecnica delle **inferenze simultanee**: si calcola una **MSD** (Minimum Significant Difference) data da

$$MSD = Q_{(\alpha_T, p, n)} \cdot \sqrt{\frac{s_e^2}{n}}$$

Sono significative tutte le differenze superiori a MSD

### Valori critici del Q per il test SNK e per il W di Tukey

**p** = numero di confronti simultanei per il test SNK

**p** = numero di medie a confronto per il T di Tukey

**v** = gradi di libertà della varianza d'errore

$$\alpha_T = 0.05$$

#### INDICE **p** DEL CONFRONTO

<b>n</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
<b>10</b>	3,151	3,877	4,327	4,654	4,912	5,124	5,305	5,461	5,599
<b>11</b>	3,113	3,820	4,256	4,574	4,823	5,028	5,202	5,353	5,487
<b>12</b>	3,082	3,773	4,199	4,508	4,751	4,950	5,119	5,265	5,395
<b>13</b>	3,055	3,735	4,151	4,453	4,690	4,885	5,049	5,192	5,318
<b>14</b>	3,033	3,702	4,111	4,407	4,639	4,829	4,990	5,131	5,254
<b>15</b>	3,014	3,674	4,076	4,367	4,595	4,782	4,940	5,077	5,198
<b>16</b>	2,998	3,649	4,046	4,333	4,557	4,741	4,897	5,031	5,150
<b>17</b>	2,984	3,628	4,020	4,303	4,524	4,705	4,858	4,991	5,108
<b>18</b>	2,971	3,609	3,997	4,277	4,495	4,673	4,824	4,956	5,071
<b>19</b>	2,960	3,593	3,977	4,253	4,469	4,645	4,794	4,924	5,038
<b>20</b>	2,950	3,578	3,958	4,232	4,445	4,620	4,768	4,896	5,008
<b>24</b>	2,919	3,532	3,901	4,166	4,373	4,541	4,684	4,807	4,915
<b>30</b>	2,888	3,486	3,845	4,102	4,302	4,464	4,602	4,720	4,824
<b>40</b>	2,858	3,442	3,791	4,039	4,232	4,389	4,521	4,635	4,735
<b>60</b>	2,829	3,399	3,737	3,977	4,163	4,314	4,441	4,550	4,646
<b>120</b>	2,800	3,356	3,685	3,917	4,096	4,241	4,363	4,468	4,560
<b>∞</b>	2,772	3,314	3,633	3,858	4,030	4,170	4,286	4,387	4,474

$$a_T = 0.01$$

INDICE  $p$  DEL CONFRONTO

n	2	3	4	5	6	7	8	9	10
10	4,482	5,270	5,769	6,136	6,428	6,669	6,875	7,055	7,213
11	4,392	5,146	5,621	5,970	6,247	6,476	6,672	6,842	6,992
12	4,320	5,046	5,502	5,836	6,101	6,321	6,507	6,670	6,814
13	4,260	4,964	5,404	5,727	5,981	6,192	6,372	6,528	6,667
14	4,210	4,895	5,322	5,634	5,881	6,085	6,258	6,409	6,543
15	4,168	4,836	5,252	5,556	5,796	5,994	6,162	6,309	6,439
16	4,131	4,786	5,192	5,489	5,722	5,915	6,079	6,222	6,349
17	4,099	4,742	5,140	5,430	5,659	5,847	6,007	6,147	6,270
18	4,071	4,703	5,094	5,379	5,603	5,788	5,944	6,081	6,201
19	4,046	4,670	5,054	5,334	5,554	5,735	5,889	6,022	6,141
20	4,024	4,639	5,018	5,294	5,510	5,688	5,839	5,970	6,087
24	3,956	4,546	4,907	5,168	5,374	5,542	5,685	5,809	5,919
30	3,889	4,455	4,799	5,048	5,242	5,401	5,536	5,653	5,756
40	3,825	4,367	4,696	4,931	5,114	5,265	5,392	5,502	5,559
60	3,762	4,282	4,595	4,818	4,991	5,133	5,253	5,356	5,447
120	3,702	4,200	4,497	4,709	4,872	5,005	5,118	5,214	5,299
∞	3,643	4,120	4,403	4,603	4,757	4,882	4,987	5,078	5,157

ESEMPIO. Con gli stessi dati utilizzati nell'esempio per il test  $t$  di Bonferroni

Zone	A	B	C	D	E
Medie	208,2	199,8	141,0	123,3	119,1
$n_i$	6	5	6	6	7

verificare con il test  $Q$  di Student-Neuman-Keuls tra quali medie esiste una differenza significativa.

Risposta.

Dapprima ordinare le medie in modo decrescente:

Ordine	Confronto	P	Medie	Differenza
1	5 <sup>a</sup> vs 1 <sup>a</sup>	5	208,2 – 119,1	89,1
2	5 <sup>a</sup> vs 2 <sup>a</sup>	4	208,2 – 123,3	84,9
3	5 <sup>a</sup> vs 3 <sup>a</sup>	3	208,2 – 141,0	67,2
4	5 <sup>a</sup> vs 4 <sup>a</sup>	2	208,2 – 199,8	8,4
5	4 <sup>a</sup> vs 1 <sup>a</sup>	4	199,8 – 119,1	80,7
6	4 <sup>a</sup> vs 2 <sup>a</sup>	3	199,8 – 123,3	76,5
7	4 <sup>a</sup> vs 3 <sup>a</sup>	2	199,8 – 141,0	58,8
8	3 <sup>a</sup> vs 1 <sup>a</sup>	3	141,0 – 119,1	21,9
9	3 <sup>a</sup> vs 2 <sup>a</sup>	2	141,0 – 123,3	17,7
10	2 <sup>a</sup> vs 1 <sup>a</sup>	2	123,3 – 119,1	4,2

Successivamente calcolare le differenze, secondo l'ordine richiesto dalla procedura.

Limitando i calcoli a due soli casi (sufficienti ad illustrare il metodo), si può saggiare la significatività del confronto 3 (la media 5<sup>a</sup> vs la media 3<sup>a</sup> con indice **p** uguale a 3),

che determina un valore di Q

$$Q_{(3,25)} = \frac{208,2 - 141,0}{\sqrt{\frac{146,5}{6}}} = \frac{67,2}{\sqrt{24,416}} = \frac{67,2}{4,94} = 13,60$$

uguale a 13,60.

Il valore critico alla probabilità  $\alpha_T = 0.05$  con  $p = 3$  con  $gdl = 24$  (il valore minore più vicino a quello dei dati) è 3,532 mentre alla probabilità complessiva  $\alpha_T = 0.01$  è uguale a 4,546.

Il valore calcolato è nettamente superiore ad entrambi: nel confronto tra le medie 208,2 e 141,0 la differenza risulta significativa, con probabilità  $\alpha < 0.01$ .

Nel confronto 8, la 3<sup>a</sup> media vs la 1<sup>a</sup> media, il valore di Q

$$Q_{(3,25)} = \frac{141,0 - 119,1}{\sqrt{\frac{146,5}{2} \left( \frac{1}{6} + \frac{1}{7} \right)}} = \frac{21,9}{\sqrt{73,25 \cdot 0,310}} = \frac{21,9}{\sqrt{22,7075}} = \frac{21,9}{4,77} = 4,59$$

è uguale a 4,59 e risulta ugualmente significativo.

4 ) In alcuni testi, al metodo SNK viene preferita la **procedura step-up di Welsch** R. E. che differisce essenzialmente per l'uso di una tabella differente di valori critici, derivati da una logica diversa di procedere ai confronti. **Rovesciando la procedura stepwise precedente** che partiva dagli estremi, dopo aver ordinato le medie secondo la grandezza,

- dapprima si verifica la significatività della differenze tra **medie adiacenti**;
- solamente **quando una differenza non risulta significativa**, si passa al confronto tra le due medie appena più distanti, che comprendono la precedente.

E' ovvio che rispetto all'approccio che parte dalle medie più distanti, questo metodo determina test più potenti e meno cautelativi, controllando le probabilità  $\alpha$  di errore di I tipo per la **comparisonwise**, ma aumentandola per l'**experimentwise**.

Focalizzando l'attenzione sulla comparisonwise,

- la probabilità  $\alpha_i$  di un errore di I tipo nel confronto tra due medie adiacenti di rango  $i$ , è

$$\alpha_i = \alpha(i/p)$$

dove

$\alpha$  = probabilità dell'experimentwise (0.05 o 0.01)

$p$  = numero di medie a confronto.

Con il metodo di **Welsch** è possibile ricorrere ai confronti multipli simultanei, prima accennati, calcolando il valore **MSR** (Minimum Significant Range)

$$\text{MSR} = Q_{(\alpha, p, n)} \cdot \sqrt{\frac{s_e^2}{n}}$$

dove

- $Q_{(\alpha, p, n)}$  è fornito da tabelle specifiche
- $p$  = numero di medie
- $n$  = gdl della varianza d'errore.

Nel confronto tra coppie di medie, sono significative tutte le differenze superiori al valore di MSR

Nel solito esempio, con la varianza d'errore  $s_e^2 = 146,5$  e  $n = 25$  gdl, con **5** gruppi che hanno tutti  $n = 6$  osservazioni,

$$\text{MSR} = Q_{(\alpha, 5, 25)} \cdot \sqrt{\frac{146,5}{6}}$$

Il **Q di Welsch è fornito da tabelle specifiche** (non riportate nel testo).

Anche in questo caso, vari autori approvano l'uso di **campioni con un numero diverso di dati**. Di conseguenza, si deve ricorrere alla **interpolazione armonica** per stimare un valore di  $\hat{n}$  comune, accettabile quando le differenze sono ridotte (con i dati dell'esempio, 5,93 al posto di 6).

5) Il **test di Tukey** (chiamato anche test di **Tukey-Kramer**, poiché a Kramer è dovuto il cambiamento della procedura originaria di Tukey) è uno dei più diffusi nei programmi informatici, per determinare quali dei confronti tra le **p** medie diano differenze significative. A volte il test è indicato con **W** (altre con **T**) e la procedura dei confronti multipli simultanei **LSD** (Least Significant Difference) qui è chiamata **HSD** (Honestly Significant Difference); più raramente, forse per rendere omogenea la terminologia, anche in questo caso altri testi usano l'acronimo **MSD** (Minimum Significant Difference).

Fondato sull'**intervallo (o campo) di variazione del Q studentizzato**, il **metodo delle inferenze simultanee** è innovativo in quanto permette di esaminare contemporaneamente **i confronti semplici tra tutte le coppie di trattamenti**, escludendo quindi i contrasti (sinonimo di confronti) complessi.

Il **campo di variazione studentizzato (Q)**, quando stimato per campioni di dimensioni **n**, è determinato da

$$Q = \frac{\bar{X}_{\max} - \bar{X}_{\min}}{\sqrt{\frac{s_e^2}{n}}}$$

Per confronti tra coppie di medie, il livello di significatività è costruito sul caso peggiore (appunto sulla differenza massima, data da  $\bar{X}_{\max} - \bar{X}_{\min}$ ); di conseguenza, molti autori di testi di statistica ritengono che fornisca **una probabilità experimentwise appropriata per il complesso dei confronti**.

*(Per le differenze tra coppie di medie semplici, attualmente è ritenuto il test **più onesto**, usando la stessa terminologia dell'autore).*

Con **p** medie, si hanno **p(p-1)/2** differenze, delle quali risultano statisticamente significative quelle che in valore assoluto sono maggiori dell'**intervallo di confidenza o campo di variazione critico W**

$$W = Q_{(a,p,n)} \cdot \sqrt{\frac{s_e^2}{n}}$$

dove:

- **a** è la probabilità complessiva prescelta,
- **p** il numero di gruppi,
- **n** sono i gradi di libertà della varianza d'errore  $s_e^2$ ,
- **n** è il numero d'osservazioni di ogni gruppo (in campioni bilanciati),
- **Q** è il valore fornito dalla tabella alla probabilità complessiva  $a_T$  per **p** gruppi e gdl **n** (della varianza d'errore).

Il metodo è corretto solo per esperimenti bilanciati; se i trattamenti hanno un numero di repliche diverso, per ogni confronto tra due generici gruppi A e B con  $n$  differenti si può stimare  $W$  mediante

$$W = Q_{(a,p,n)} \cdot \sqrt{\frac{s_e^2}{2} \cdot \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}$$

Il risultato è approssimato, se esteso a tutti i confronti.

Per calcolare un solo valore nonostante l'uso di gruppi con un numero diverso di osservazioni, resta solamente la scelta più cautelativa, quella del numero  $n$  di osservazioni minore tra tutti i gruppi a confronto.

Questa **procedura di Tukey-Kramer**, nota anche come **metodo dell'intervallo di variazione**, oltre al confronto tra due medie specifiche permette di applicare la tecnica dei confronti simultanei, calcolando **la differenza minima significativa** (in inglese, MSD = Minimum Significant Difference) tra tutte le coppie di medie.

Il metodo richiede di **calcolare tutte le differenze tra le  $p$  medie: sono significative alla probabilità complessiva prefissata a quelle che superano il valore HSD.**

Per la presentazione dei risultati, sovente si ricorre ad una tabella che evidenzia contemporaneamente tutte le differenze. Con 4 medie si ottengono 6 differenze; tra esse quelle superiori al valore **HSD** calcolato sono significative e possono essere evidenziate in modo grafico (con gli asterischi), secondo il livello di significatività.

		MEDIE			
		(1)	(2)	(3)	(4)
		$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$
MEDIE	DIFFERENZE				
(2)	$\bar{X}_2$	$\bar{X}_1 - \bar{X}_2$	---	---	---
(3)	$\bar{X}_3$	$\bar{X}_1 - \bar{X}_3$	$\bar{X}_2 - \bar{X}_3$	---	---
(4)	$\bar{X}_4$	$\bar{X}_1 - \bar{X}_4$	$\bar{X}_2 - \bar{X}_4$	$\bar{X}_3 - \bar{X}_4$	---

ESEMPIO. Stimare le differenze significative tra le 5 medie utilizzate negli esempi precedenti, alle probabilità  $\alpha = 0.05$  e  $\alpha = 0.01$ .

Risposta.

Dalla tabella di distribuzione dei valori critici del **Q studentizzato**, scegliere il valore di **Q** per:

- la probabilità  $\alpha = 0.05$  e  $\alpha = 0.01$ ;
- il numero di trattamenti o medie **p**, che con i dati dell'esempio è uguale a **5**;
- i gradi di libertà della varianza d'errore (**n**), che nell'esempio sono uguali a **25**; nella tabella dei valori critici scegliere il numero inferiore più vicino (24), in quanto più cautelativo di quello superiore (30); un altro metodo, più preciso ma leggermente più difficile, suggerisce di stimare il valore di **Q** mediante interpolazione armonica tra i gdl riportati in tabella (24 e 30).

Nella tabella dei valori critici (riportata in precedenza), il valore di **Q**

- alla probabilità  $\alpha = 0.05$  è uguale a **4,166**
- alla probabilità  $\alpha = 0.01$  è uguale a **5,168**.

Ricordando che la varianza d'errore  $s_e^2 = 146,5$

calcolare il campo di variazione critico o intervallo di confidenza per un generico contrasto semplice tra 2 medie, mediante il valore di **HSD**.

Alla probabilità  $\alpha = 0.05$

$$\mathbf{HSD} = 4,166 \cdot \sqrt{\frac{146,5}{5,93}} = 4,166 \times 4,962 = 20,67$$

**HSD** risulta uguale a 20,67

mentre alla probabilità  $\alpha = 0.01$

$$\mathbf{HSD} = 5,168 \cdot \sqrt{\frac{146,5}{5,93}} = 5,168 \times 4,962 = 25,64$$

**HSD** è uguale a 25,64.

Successivamente, è utile costruire la matrice triangolare delle differenze tra le 5 medie ed effettuare i confronti con le due **HSD** calcolate per le due probabilità. Dall'analisi risulta che

- con probabilità  $\alpha < 0.01$  sono significative le differenze superiori a **25,64**,
- con probabilità  $\alpha < 0.05$  sono significative quelle comprese tra **25,64** e **20,67**
- le differenze minori di **20,67** non sono significative, avendo probabilità  $\alpha > 0.05$ .

(Le prime possono essere contrassegnate da un doppio asterisco; le seconde da un solo asterisco).

		MEDIE				
		A	B	C	D	E
		208,2	199,8	141,0	123,3	119,1
MEDIE		DIFFERENZE				
B	199,8	8,4	---	---	---	---
C	141,0	<b>67,2**</b>	<b>58,8**</b>	---	---	---
D	123,3	<b>85,5**</b>	<b>76,5**</b>	17,7	---	---
E	119,1	<b>89,1**</b>	<b>80,7**</b>	<b>21,9*</b>	4,2	---

L'interpretazione della tabella porta alla conclusione che sono molto significative ( $\alpha < 0.01$ ) le differenze (in grassetto con due asterischi) tra la media C, la media D e la media E rispetto sia alla media A che alla B e significativa ( $0.01 < \alpha < 0.05$ ) la differenza tra la media C e la E.

Esse sono le cause della differenza complessiva tra le 5 medie, valutata in precedenza con il test F.

*(Questa tecnica era usata alcuni anni fa; ora i computer permettono di riportare il valore esatto di  $\alpha$  per ogni confronto e quindi di avere una visione più dettagliata, forse a discapito della sintesi).*

Per l'importanza che il test ha assunto tra i confronti multipli, a ulteriore chiarimento della metodologia è qui riproposta la presentazione di George E. P. **Box**, William G. **Hunter** e J. Stuart **Hunter** (nel testo: "*Statistics for Experimenters. An introduction to Design, Data Analysis and Model Building*", pubblicato nel 1978 da John Wiley & Sons, New York, p. 653).

**La procedura di Tukey** J. W. (vedi: (1949) *Comparing individual means in the analysis of variance*. Biometrics, 5, 99) **per il confronto simultaneo tra p medie** suggerisce di stimare il limite di confidenza tra tutte le differenze di  $p(p-1)/2$  coppie di medie ( $\bar{X}_i$  e  $\bar{X}_j$ ) con

$$(\bar{X}_i - \bar{X}_j) \pm \frac{Q_{\alpha/2, p, n}}{\sqrt{2}} \cdot \sqrt{s_e^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

dove

$Q_{p, n}$  è il valore di **q studentizzato** per il confronto tra **p** medie, con gdl **n** della varianza d'errore.

ESEMPIO. Individua quali differenze sono significative tra le **7** medie  $\bar{X}_i$  riportate nella tabella

Trattamenti	A	B	C	D	E	F	G
$\bar{X}_i$	53	52	57	55	55	60	50
$s_i^2$	9,2	8,7	8,8	9,8	10,2	8,3	8,0

ognuna con 4 dati (**n = 4**) e la varianza entro gruppo  $s_i^2$  uguale a quella indicata.

1 – Dapprima si stima una varianza comune a tutti i gruppi (o varianza d'errore)  $s^2 = 9,0$  che in questo caso, con campioni bilanciati, è uguale alla media delle varianze

$$s_e^2 = 3 (9,2 + 8,7 + 8,8 + 9,8 + 10,2 + 8,3 + 8,0) / 21 = 9,0$$

e ha gdl **n = 21**, pari a **p(n-1) = 7 (4-1)**.

Per **a = 0.05** e con

$$\frac{Q_{\alpha/2, p, n}}{\sqrt{2}} = 3,26$$

(tratto dalle tabelle relative)

alla probabilità del 95% si stima **un intervallo fiduciale o differenza minima significativa delle differenze tra le medie** pari a

$$\pm \frac{Q_{\alpha/2, p, n}}{\sqrt{2}} \cdot \sqrt{s_e^2 \cdot \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} = \pm 3,26 \cdot \sqrt{9,0 \cdot \left( \frac{1}{4} + \frac{1}{4} \right)} = \pm 6,91$$

**6,91 senza considerare il segno.**

Con le sette medie precedenti,

Trattamenti	A = 53	B = 52	C = 57	D = 55	E = 55	F = 60	G = 50
A = 53	---	---	---	---	---	---	---
B = 52	1	---	---	---	---	---	---
C = 57	-4	-5	---	---	---	---	---
D = 55	-2	-3	2	---	---	---	---
E = 55	-2	-3	2	0	---	---	---
F = 60	<b>-7*</b>	<b>-8*</b>	-3	-5	-5	---	---
G = 50	3	2	<b>7*</b>	5	5	<b>-10*</b>	---

si possono stimare 21 differenze (riportate in grassetto nella tabella), di cui risultano significative le quattro in grassetto e con l'asterisco.

6) **La procedura di Scheffé è il test più versatile** per i confronti multipli: offre il vantaggio di eseguire anche **confronti complessi oltre a quelli semplici**. Il numero totale di confronti è quindi superiore a quello permesso dal test di Tukey. Pertanto, sulla base del principio del Bonferroni, cambiano i rapporti tra **comparisonwise** e **experimentwise**, cioè tra la **potenza** o probabilità  $\alpha$  di ogni confronto e la **protezione** o probabilità  $\alpha_T$  complessiva.

Il test di Scheffé ha varie applicazioni, delle quali le più diffuse sono tre:

A - confronto tra due medie, sia semplici che complesse;

B – scomposizione della devianza di un gruppo di medie, per valutare il contributo relativo di ogni confronto;

C – matrice di tutte le possibili differenze, sia semplici che complesse.

A - Nel caso di confronti **sia semplici che complessi**, con il test di Scheffé risulta significativa la differenza **D**, se il valore di **L**, ottenuto dalle medie  $\bar{X}_i$  e dai contrasti  $c_i$ ,

con la relazione

$$L = \sum (c_i \cdot \bar{X}_i)$$

in valore assoluto è maggiore di **S**

$$S = \sqrt{(p-1) \cdot F_{(a;p-1,n)} \cdot \sum c_i^2 \cdot \left(\frac{s_e^2}{n}\right)}$$

dove

- **p** = numero di gruppi interessati al confronto,
- **F<sub>(a;p-1,n)</sub>** = valore dell'**F di Fisher** alla probabilità **a** prescelta e con gdl (**p-1; n**) uguali a quelli del test ANOVA ad un criterio di classificazione,
- **c<sub>i</sub>** = valore dei singoli contrasti implicati nel confronto.

La metodologia può essere spiegata in modo chiaro mediante l'illustrazione di un esempio, scomposto dettagliatamente nei suoi passaggi fondamentali.

Si supponga che, con le 5 medie riportate nella tabella sottostante e stimate su campioni con lo stesso numero di dati (6)

<b>ZONE</b>	A	B	C	D	E
<b>Medie</b>	208,2	199,8	141,0	123,3	119,1
<b>n<sub>i</sub></b>	6	6	6	6	6

si intenda valutare specificatamente la significatività della differenza tra il livello medio dell'inquinamento rilevato nelle due stazioni del centro (A e B) rispetto a quello delle due stazioni collocate in periferia (D e E), ricordando che la varianza d'errore  $s_e^2$  è risultata uguale a **146,5** con **25** gdl.

1 - Si verifica l'ipotesi nulla

$$H_0: \delta = 0$$

contro l'ipotesi alternativa bilaterale

$$H_1: \delta \neq 0$$

valutando la significatività della differenza **D** tra la media del centro rispetto a quella della periferia

$$D = \frac{\bar{X}_A + \bar{X}_B}{2} - \frac{\bar{X}_D + \bar{X}_E}{2} = \frac{208,2 + 199,8}{2} - \frac{123,3 + 119,1}{2} = 204 - 121,2 = 82,8$$

in cui **D** risulta uguale a 82,8.

2 - A questo scopo, si calcola il valore del contrasto **L**

$$L = \bar{X}_A + \bar{X}_B - \bar{X}_D - \bar{X}_E = 208,2 + 199,8 - 123,3 - 119,1 = 165,6$$

che risulta uguale a 165,6

e il cui valore  $\sum c_i^2$ , derivato dai coefficienti ortogonali, è

$$\sum c_i^2 = +1^2 + 1^2 - 1^2 - 1^2 = 4$$

uguale a 4.

3 – Alla probabilità  $\alpha = 0.05$ , con  $F_{0.025, 3, 25} = 3,69$  (tratto dalla tabella F di Fisher),  
 $p = 4$  e  $n = 6$

$$S = \sqrt{3 \cdot 3,68 \cdot 4 \cdot \frac{146,5}{6}} = 32,8$$

S risulta uguale a 32,8.

4 – Poiché il valore del contrasto **L** (165,6) è nettamente maggiore del valore **S** (32,8) calcolato, la differenza **D** (82,8) tra le due medie è significativa ad una probabilità  $\alpha$  nettamente inferiore a 0.05.

**Il test di Scheffé è il più adatto, quando il numero di gruppi e quello di osservazioni per gruppo sono molto piccoli e parte dei confronti non sono ortogonali.** Poiché permette di verificare tutti i confronti possibili, **se l'ANOVA è risultata significativa alla probabilità  $\alpha$ , almeno uno di tutti i possibili confronti dovrebbe risultare significativo alla stessa probabilità  $\alpha$  o a una probabilità inferiore.**

B - Il metodo di Scheffé, nell'esempio precedente applicato ad un singolo confronto tra due medie complesse, con le modifiche proposte da **Gabriel K. R.** (vedi: *A simple method of multiple comparisons of means*, pubblicato nel 1978 da *J. Amer. Stat. Assn.* Vol. 73, pp 724-729) è stato esteso a tutti i possibili confronti tra gruppi di medie e a tutte le possibili suddivisioni delle **p** medie.

I concetti di base di **questa diversa applicazione del test di Scheffé**, indicata con l'acronimo **SS-STP** (*Sum of Squares – Simultaneous Test Procedure*), possono essere derivati dall'ANOVA ad un criterio di classificazione.

Con l'ANOVA l'ipotesi nulla su **p** medie

$$H_0: \mu_1 = \mu_2 = \dots = \mu_p$$

è rifiutata quando

$$\frac{s_{tra}^2}{s_{entro}^2} \geq F_{\alpha, p-1, n-p}$$

Questo rapporto tra le varianze può essere scritto in modo differente: si divide la **Devianza tra** (indicata con **SS<sub>tra</sub>** da *Sum of Squares*) per i suoi gdl (**p-1**), per cui la formula diventa

$$\frac{SS_{tra}}{s_e^2 \cdot (p-1)} \geq F_{\alpha, p-1, n-p}$$

Da questa relazione si deriva facilmente

$$SS_{tra} \geq (p-1) \cdot s_e^2 \cdot F_{\alpha, p-1, n-p}$$

il valore di significatività.

Questa relazione può essere usata per **qualsiasi confronto non pianificato tra p medie, con n osservazioni per gruppo.**

L'applicazione ai dati della tabella sottostante, in cui ognuna delle 4 medie interessate al confronto (A, B, D, E) ha lo stesso numero di dati (6) e la varianza d'errore  $s_e^2$  è assunta uguale a **146,5**

ZONE	A	B	C	D	E
Medie	208,2	199,8	141,0	123,3	119,1
$n_i$	6	6	6	6	6
$\sum x_i$	1249,2	1198,8	846,0	739,8	714,6

**spiega come effettuare tutti i possibili confronti tra le 4 medie A, B, D, E, prescelte.**

Il metodo permette **sia confronti ortogonali, sia non ortogonali** tra queste 4 medie, quali

- A vs. B + D + E
- A + B vs. D + E
- A vs. D; ecc....

1 - A questo scopo, si calcola il valore della **Devianza tra (SS)**

$$\sum \frac{(\sum x_i)^2}{n_i} - \frac{(\sum x)^2}{n}$$

Considerando le 4 medie a confronto

$$SS = (1249,2)^2/6 + (1198,8)^2/6 + (739,8)^2/6 + (714,6)^2/6 - (3902,4)^2/24 =$$

$$260.083,44 + 239.520,24 + 91.217,34 + 85.108,86 - 638.432,64 = 37.497,24$$

SS risulta uguale a 37.497,24.

2 - Il **valore critico** di confronto con

- **p = 4**
- **$s_e^2 = 146,5$**
- **F** per  $\alpha = 0.05/2$ , e **gdl 3, 20 = 3,10**

è

$$(4-1) \times 146,5 \times 3,1 = 1.362,45$$

uguale a 1.362,45.

Ovviamente, poiché **SS** (37.497,24) è molto maggiore del **valore critico** (1.362,45), si rifiuta l'ipotesi nulla: le 4 medie sono tra loro significativamente diverse.

3 - Con i confronti ortogonali, è possibile **scomporre la devianza SS** prima calcolata. Diventa così possibile valutare il contributo di ogni contrasto a questo valore **SS** complessivo, fondato sulle differenze tra le 4 medie.

Si tratta di applicare la tecnica dei confronti ortogonali su queste 4 medie, come illustrata nei confronti a priori.

C - Il **metodo più generale e più versatile di utilizzazione del test di Scheffé** permette di costruire una **matrice di tutte le possibili differenze** tra tutte le medie considerate; in questo caso, le 5 medie riportate nella tabella successiva. Con la **procedura MSD**, risultano significative alla probabilità  $\alpha$  prefissata tutte le differenze che sono maggiori della quantità

$$\text{MSD} \geq \sqrt{(p-1) \cdot F_{\alpha; (p-1), n}} \cdot \sqrt{s_e^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

dove,

(oltre alla consueta simbologia relativa al confronto tra due medie)

**p** è il numero di medie semplici o di più gruppi che si intende confrontare,

$F_{\alpha; (p-1), n}$  è il valore di **F** alla probabilità  $\alpha_T$  prescelta e per i gradi di libertà **(p-1)** (dati dal numero di medie **p**) e **n** della varianza d'errore.

Con

- **p = 5** e **n = 6**
- **s<sub>e</sub><sup>2</sup> = 146,5**
- **F** per  $\alpha = 0.05$ , e **gdl 4, 25 = 2,76**

si ottiene un valore MSD

$$\text{MSD} = \sqrt{(5-1) \cdot 2,76} \cdot \sqrt{146,5 \cdot \left( \frac{1}{6} + \frac{1}{6} \right)} = \sqrt{11,04} \cdot \sqrt{48,633} = 23,21$$

uguale a 23,21

Con

- **F** per  $\alpha = 0.01$ , e **gdl 4, 25 = 4,18**

$$\text{MSD} = \sqrt{(5-1) \cdot 4,18} \cdot \sqrt{146,5 \cdot \left( \frac{1}{6} + \frac{1}{6} \right)} = \sqrt{16,72} \cdot \sqrt{48,633} = 28,51$$

si ottiene un MSD uguale a 28,51.

Con la stessa tecnica già illustrata in precedenza, nella tabella delle **p(p-1)/2 differenze** si evidenziano con

- un asterisco tutte le differenze che sono significative ad una probabilità  $\alpha$  compresa tra 0.05 e 0.01
- due asterischi tutte le differenze significative ad una probabilità  $\alpha$  minore di 0.01

		MEDIE				
		A	B	C	D	E
		208,2	199,8	141,0	123,3	119,1
MEDIE		DIFFERENZE				
B	199,8	8,4	---	---	---	---
C	141,0	<b>67,2**</b>	<b>58,8**</b>	---	---	---
D	123,3	<b>85,5**</b>	<b>76,5**</b>	17,7	---	---
E	119,1	<b>89,1**</b>	<b>80,7**</b>	21,9	4,2	---

E' importante osservare che il valore la differenza (21,9) tra la media **C** (141,0) e la media **E** (119,1) non è significativa alla probabilità  $\alpha = 0.05$ , mentre la era **con il test di Tukey**.

Più in generale, in questo esempio, **con Scheffé** la **differenza minima significativa** sulle 5 medie è

- **23,21** alla probabilità  $\alpha = 0.05$
- **28,51** alla probabilità  $\alpha = 0.01$

mentre **con il test di Tukey** era

- **20,67** alla probabilità  $\alpha = 0.05$
- **25,64** alla probabilità  $\alpha = 0.01$

Questo confronto tra Tukey e Scheffé e il confronto tra i risultati della terza applicazione del test di Scheffé rispetto alle prime due evidenziano la **relazione inversa tra versatilità e potenza di un test**. Tanto più il test è generale, tanto meno è potente. Lo stesso concetto è applicato quando i campioni hanno dimensioni diverse: la scelta prudentiale o cautelativa della frequenza **n** minore rende le conclusioni generali, cioè estese a tutti i possibili confronti dell'esperimento, ma diminuisce la potenza di ogni confronto.

Il metodo di Scheffé ha applicazioni più ampie per numero di confronti; ma **è più prudentiale di quello SNK e di quello di Tukey**.

**Se interessano solo i confronti singoli, è svantaggioso utilizzare un metodo che permette anche confronti complessi; se nel piano dell'esperimento l'interesse è rivolto solo ad alcuni confronti, è svantaggioso per la significatività utilizzare un metodo che li permette tutti.**

7) Il **test Q di Dunnett** deve essere applicato in un caso particolare: **il confronto di due o più trattamenti con un controllo**.

In numero di confronti da effettuare diminuisce: è uguale al numero di trattamenti, escludendo il controllo. Con 5 gruppi, fra cui un controllo e 4 trattamenti, non è più uguale a 10 ( $C_5^2$ ) ma a 4.

Sulla base per principio del Bonferroni, aumenta quindi la potenza di ogni confronto, poiché con una probabilità experimentwise  $\alpha_T = 0.05$  la probabilità a comparisonwise diventa uguale 0.0125 (0.05 /4).

**Ovviamente questa scelta riduce la versatilità del test: si verifica la significatività della differenza tra ogni trattamento ed il controllo, senza poter dire nulla sulle eventuali differenze tra i trattamenti**, cioè se uno è migliore o peggiore degli altri in modo significativo.

**Il test Q di Dunnett utilizza la stessa formula del test SNK**

(con la medesima simbologia)

$$Q_{(\alpha, p, n)} = \frac{\bar{X}_c - \bar{X}_i}{\sqrt{s_e^2 \cdot \left( \frac{1}{n_c} + \frac{1}{n_i} \right)}}$$

dove

- **c** = **gruppo di controllo o placebo**
- **i** = **trattamento in oggetto**, per il quale si effettua il confronto con il controllo.

A differenza degli altri test per confronti multipli, che **a parità del numero totale di osservazioni** raggiungono la potenza maggiore quando tutti i gruppi sono bilanciati, nel confronto tra vari trattamenti con un controllo **si ottiene una utilizzazione migliore dei dati quando il controllo è di dimensioni ragionevolmente maggiori di quella dei trattamenti**.

Infatti, il controllo entra in tutti i confronti ed un numero più alto di osservazioni in esso aumenta la potenza di ogni confronto, anche se determina una parziale penalizzazione per il mancato bilanciamento.

### Valori critici del Q per il test di Dunnett

**p** = numero di medie a confronto (compreso il controllo)

**n** = gradi di libertà della varianza d'errore

$$a_T = 0.05$$

#### NUMERO **p** DI MEDIE A CONFRONTO

n	2	3	4	5	6	7	8	9	10
10	2,23	2,57	2,76	2,89	2,99	3,07	3,14	3,19	3,24
11	2,20	2,53	2,72	2,84	2,94	3,02	3,08	3,14	3,19
12	2,18	2,50	2,68	2,81	2,90	2,98	3,04	3,09	3,14
13	2,16	2,48	2,65	2,78	2,87	2,94	3,00	3,06	3,10
14	2,14	2,46	2,63	2,75	2,84	2,91	2,97	3,02	3,07
15	2,13	2,44	2,61	2,73	2,82	2,89	2,95	3,00	3,04
16	2,12	2,42	2,59	2,71	2,80	2,87	2,92	2,97	3,02
17	2,11	2,41	2,58	2,69	2,78	2,85	2,90	2,95	3,00
18	2,10	2,40	2,56	2,68	2,76	2,83	2,89	2,94	2,98
19	2,09	2,39	2,55	2,66	2,75	2,81	2,87	2,92	2,96
20	2,09	2,38	2,54	2,65	2,73	2,80	2,86	2,90	2,95
24	2,06	2,35	2,51	2,61	2,70	2,76	2,81	2,86	2,90
30	2,04	2,32	2,47	2,58	2,66	2,72	2,77	2,82	2,86
40	2,02	2,29	2,44	2,54	2,62	2,68	2,73	2,77	2,81
60	2,00	2,27	2,41	2,51	2,58	2,64	2,69	2,73	2,77
120	1,98	2,24	2,38	2,47	2,55	2,60	2,65	2,69	2,73
∞	1,96	2,21	2,35	2,44	2,51	2,57	2,61	2,65	2,69

$$a_T = 0.01$$

#### NUMERO **p** DI MEDIE A CONFRONTO

n	2	3	4	5	6	7	8	9	10
10	3,17	3,53	3,74	3,88	3,99	4,08	4,16	4,22	4,28
11	3,11	3,45	3,65	3,79	3,89	3,98	4,05	4,11	4,16
12	3,05	3,39	3,58	3,71	3,81	3,89	3,96	4,02	4,07
13	3,01	3,33	3,52	3,65	3,74	3,82	3,89	3,94	3,99
14	2,98	3,29	3,47	3,59	3,69	3,76	3,83	3,88	3,93
15	2,95	3,25	3,43	3,55	3,64	3,71	3,78	3,83	3,88
16	2,92	3,22	3,39	3,51	3,60	3,67	3,73	3,78	3,83
17	2,90	3,19	3,36	3,47	3,56	3,63	3,69	3,74	3,79
18	2,88	3,17	3,33	3,44	3,53	3,60	3,66	3,71	3,75
19	2,86	3,15	3,31	3,42	3,50	3,57	3,63	3,68	3,72
20	2,85	3,13	3,29	3,40	3,48	3,55	3,60	3,65	3,69
24	2,80	3,07	3,22	3,32	3,40	3,47	3,52	3,57	3,61
30	2,75	3,01	3,15	3,25	3,33	3,39	3,44	3,49	3,52
40	2,70	2,95	3,09	3,19	3,26	3,32	3,37	3,41	3,44
60	2,66	2,90	3,03	3,12	3,19	3,25	3,29	3,33	3,37
120	2,62	2,85	2,97	3,06	3,12	3,18	3,22	3,26	3,29
∞	2,58	2,79	2,92	3,00	3,06	3,11	3,15	3,19	3,22

Quando si programma un esperimento sul quale deve essere applicato il test di Dunnett, è conveniente che

- $n_c$ , il numero di dati del controllo, sia più numeroso di
- $n_i$ , il numero dei dati di ogni trattamento,
- in funzione del numero di trattamenti  $p$

secondo la relazione

$$n_c = n_i \cdot \sqrt{p}$$

Ad esempio, in un esperimento con 7 dati in ognuno dei 5 gruppi (il controllo più 4 trattamenti e quindi 35 osservazioni in tutto), si ottiene la migliore utilizzazione complessiva

$$n_c = 7 \cdot \sqrt{5} = 15,65$$

quando

- 15 cavie sono dedicate al controllo e
- le rimanenti 20 sono suddivise tra i 4 trattamenti.

E' una indicazione approssimata, in quanto è semplice verificare che nella formula

$$Q_{(a,p,n)} = \frac{\bar{X}_c - \bar{X}_i}{\sqrt{s_e^2 \cdot \left( \frac{1}{n_c} + \frac{1}{n_i} \right)}}$$

si ottiene il valore massimo di Q (quindi il risultato più significativo)

quando (a parità di tutti gli altri parametri)

$$\frac{1}{n_c} + \frac{1}{n_i} = X_{\min}$$

la somma dei due rapporti ha il valore minimo,

ovviamente mantenendo costante il numero totale  $n$  di dati.

Una stima più precisa ed una verifica degli effetti di questa concentrazione delle osservazioni sul campione di controllo può essere ottenuta con un confronto dettagliato delle varie possibili distribuzioni del numero complessivo di cavie disponibili nei vari gruppi.

Con 35 osservazioni in totale,

- nel caso di campioni bilanciati e quindi  $n_c = 7$  e  $n_i = 7$  si avrebbe  $\frac{1}{7} + \frac{1}{7} = 0,2857$

- nel caso di  $n_c = 11$  e  $n_i = 6$  si avrebbe  $\frac{1}{11} + \frac{1}{6} = 0,0909 + 0,1667 = 0,2576$

- nel caso di  $n_c = 15$  e  $n_i = 5$  si avrebbe  $\frac{1}{15} + \frac{1}{5} = 0,0667 + 0,2000 = 0,2667$

- nel caso di  $n_c = 19$  e  $n_i = 4$  si avrebbe  $\frac{1}{19} + \frac{1}{4} = 0,0526 + 0,2500 = 0,3026$

Per ottenere la maggiore potenza del test, con 35 cavie e 5 gruppi, la scelta più vantaggiosa è collocare 11 cavie nel gruppo di controllo e 6 in ognuno degli altri 4 trattamenti.

ESEMPIO 1. Si è voluto esaminare l'effetto di 6 diverse sostanze tossiche sull'accrescimento somatico di una specie planctonica (misurati in mm dopo 20 giorni dalla schiusa delle uova), per verificare quali di esse riducano significativamente le dimensioni medie (test unilaterale) allo stato adulto.

Con i seguenti risultati ottenuti in laboratorio:

	CONTROLLO	SOSTANZE TOSSICHE					
		A	B	C	D	E	F
Media	3,25	2,80	2,18	2,96	2,24	2,39	2,67
Osservazioni	10	7	7	7	7	7	7

per un totale di 52 osservazioni, di cui 10 nel gruppo controllo.

L'analisi della varianza con  $F_{(7, 45)}$  ha permesso di rifiutare l'ipotesi nulla; la varianza d'errore  $s_e^2$  con 45 gdl è risultata uguale a 0,36. Verificare quali sostanze hanno un effetto significativo alla probabilità  $\alpha = 0.05$  e quali anche alla probabilità  $\alpha = 0.01$  in rapporto al controllo.

Risposta.

I confronti da effettuare sono 6. E' possibile stimare una differenza minima significativa (MDS) unica, poiché i trattamenti hanno tutti lo stesso numero d'osservazioni

$$Q_{(\alpha; p, n-p)} = \sqrt{s_e^2 \cdot \left(\frac{1}{n_c} + \frac{1}{n_i}\right)}$$

Con i dati dell'esempio ( $p = 6$  e  $gdl = 40$ ), nella tavola dei valori critici

- alla probabilità  $\alpha = 0.05$  il valore del **Q di Dunnett** è uguale a 2,62
- alla probabilità  $\alpha = 0.01$  è uguale a 3,26.

Pertanto,

- alla probabilità  $\alpha = 0.05$

il valore della MDS

$$2,62 \cdot \sqrt{0,36 \cdot \left(\frac{1}{10} + \frac{1}{7}\right)} = 2,62 \cdot \sqrt{0,36 \cdot 0,243} = 2,62 \cdot 0,296 = 0,775$$

è uguale a 0,775 e

- alla probabilità  $\alpha = 0.01$

$$3,26 \cdot \sqrt{0,36 \cdot \left(\frac{1}{10} + \frac{1}{7}\right)} = 3,26 \cdot \sqrt{0,36 \cdot 0,243} = 3,26 \cdot 0,296 = 0,965$$

MDS è uguale a 0,965.

Si calcolano le differenze dei 6 trattamenti rispetto al controllo e si verifica la loro significatività mediante il confronto con i due valori MDS stimati. Possono essere segnate con

- due asterischi le differenze maggiori del valore 0,965 e
- un asterisco le differenze comprese tra 0,965 e 0,775.

A	$3,25 - 2,80 = \mathbf{0,45}$
B	$3,25 - 2,18 = \mathbf{1,07^{**}}$
C	$3,25 - 2,96 = \mathbf{0,29}$
D	$3,25 - 2,24 = \mathbf{1,01^{**}}$
E	$3,25 - 2,39 = \mathbf{0,86^*}$
F	$3,25 - 2,67 = \mathbf{0,58}$

La tabella evidenzia che, delle 6 sostanze tossiche esaminate nell'esperimento, rispetto al controllo hanno un effetto molto significativo ( $\alpha < 0.01$ ) la B e la D, mentre ha un effetto significativo ( $\alpha < 0.05$ ) la E. Le sostanze A, C ed F non hanno ridotto la crescita in modo significativo rispetto al controllo ( $\alpha > 0.05$ ).

ESEMPIO 2. Questo secondo esempio è tratto dal testo di George E. P. **Box**, William G. **Hunter** e J. Stuart **Hunter** (nel testo: "*Statistics for Experimenters. An introduction to Design, Data Analysis and Model Building*", pubblicato nel 1978 da John Wiley & Sons, New York, p. 653) che individua

nel metodo di Tukey e in quello di Dunnett le due proposte fondamentali, per analisi da effettuare con calcoli manuali.

**La procedura di Dunnett** per il confronto tra **k** medie con la media di un campione standard o controllo richiede ovviamente il calcolo e l'analisi di **k-1** differenze.

Per ogni differenza ( $\bar{X}_i - \bar{X}_c$ ) tra la media di un generico trattamento **i** ( $\bar{X}_i$ ) e la media del controllo ( $\bar{X}_c$ ) si stima un intervallo fiduciale

$$(\bar{X}_i - \bar{X}_c) \pm t_{k,n,a/2} \cdot \sqrt{s_e^2 \left( \frac{1}{n_i} + \frac{1}{n_c} \right)}$$

in cui,

- al posto del valore di **q studentizzato**, viene utilizzato
- quello di **t** per **k** confronti, con gdl n e la probabilità **a/2**.

Con le 7 medie di prima in cui G sia il controllo

Trattamenti	A	B	C	D	E	F	G = Controllo
$n_i$	4	4	4	4	4	4	4
$\bar{X}_i$	53	52	57	55	55	60	50

alla probabilità del 95% dove  $t_{7, 21, 0,025} = 2,80$

si stima una differenza minima significativa

$$\pm t_{k,n,a/2} \cdot \sqrt{s_e^2 \left( \frac{1}{n_i} + \frac{1}{n_c} \right)} = \pm 2,80 \cdot \sqrt{9,0 \cdot \left( \frac{1}{4} + \frac{1}{4} \right)} = \pm 5,94$$

che risulta uguale a 5,94 (osservare che è minore del valore precedente, in quanto **stimato non per k(k-1)/2 confronti ma per k-1**).

Di conseguenza, tra le 6 differenze riportate nella tabella successiva

Trattamenti	A	B	C	D	E	F
$\bar{X}_i$	53	52	57	55	55	60
Differenze $\bar{X}_i - \bar{X}_c$	<b>3</b>	<b>2</b>	<b><u>7</u></b>	<b>5</b>	<b>5</b>	<b><u>10</u></b>

Sono significativamente diverse, dalla media del controllo, la media del trattamento C e quella del trattamento F.

8) Il **metodo di Duncan**, chiamato **test del campo di variazione multiplo**, può essere applicato nelle **stesse condizioni del test di Tukey, ma con una significatività *experiment wise* meno cautelativa e quindi una potenza per un singolo test o *comparisonwise* maggiore**. Di conseguenza, rende significative differenze tra coppie di singole medie che non risultano tali con il test di Tukey e il test SNK, facendosi per questo preferire da vari ricercatori, soprattutto negli ultimi anni; altri continuano a preferire Tukey, appunto perché più prudentiale e quindi “meno contestabile”, una volta dimostrata la significatività di una differenza tra due medie.

Il metodo è fondato sulla **procedura step-wise**, ma con un approccio innovativo.

**La procedura classica del campo di variazione studentizzato**, dopo aver disposto le medie a confronto in ordine di rango, prende in considerazione solo il numero di medie coinvolto in ogni singolo confronto. Per confrontare due medie adiacenti (come già spiegato nel presentare il metodo **LSD**, con  $p = 2$  e detta **comparazione delle medie due passi avanti**) utilizza il campo di variazione studentizzato su due medie; per valutare la differenza tra due medie tre passi avanti, stima il campo di variazione studentizzato su tre medie; più in generale, per confrontare le differenze tra medie  $p$  passi avanti, calcola il campo di variazione su  $p$  medie.

Quando stimato per campioni di dimensioni  $n$ , il **campo di variazione studentizzato (Q)** determinato da

$$Q = \frac{\bar{X}_{\max} - \bar{X}_{\min}}{\sqrt{\frac{s_e^2}{n}}}$$

risulta minore quando il numero  $p$  di medie a confronto è minore.

Ne deriva che la stessa differenza, in valore assoluto, tra due medie può risultare

- **significativa** se il confronto è effettuato **tra due medie adiacenti**,
- **non significativa** se effettuata **tra medie distanti tre o più passi avanti**.

Per un **principio di cautela**, la **procedura classica** dei confronti a posteriori contrasta questo ultimo effetto: **esclude che possa essere dichiarata significativa la differenza tra due medie, quando trovano all'interno di un'altra coppia non significativa**.

La **procedura proposta da Duncan**, sempre dopo aver disposto le medie in ordine di rango, rovescia la logica precedente.

Dapprima definisce il **livello di protezione per due medie separate da  $r$  passi**

come

$$(1 - \alpha)^{r-1}$$

Di conseguenza, la probabilità di commettere un errore di I tipo, cioè quella di rifiutare l'ipotesi nulla quando essa è vera, tra due medie campionarie separate da  $r$  passi è

$$1 - (1 - \alpha)^{r-1}$$

dove

- $\alpha$  = livello di significatività di un singolo confronto
- $r$  = numero di passi, tra le due medie a confronto.

Tra **due medie adiacenti**, in cui  $r = 2$ , il livello di protezione è uguale a  $\alpha$ : è facile verificare, con un semplice confronto tra le tabelle relative, che il valore del campo di variazione multiplo di Duncan è uguale a quello del Q studentizzato.

Ma per **due medie non adiacenti**, al crescere del numero ( $r$ ) di passi, il livello di protezione o probabilità experiment-wise si riduce progressivamente, rendendo il test di Duncan sempre più potente nei confronti dei test fondati sul valore del Q. Per una indicazione semplice di questo effetto, è sufficiente confrontare la tabella di **Duncan** con quella del **Q studentizzato** per il test SNK e il test W di Tukey: alla stessa probabilità  $\alpha$  e per i medesimi gdl, il valore di **Duncan** è minore di quello della tabella **Q**, in modo sempre più accentuato al crescere di  $r$ .

- Con le stesse 5 medie utilizzate in precedenza,

Zone	A	B	C	D	E
Medie	208,2	199,8	141,0	123,3	119,1
$n_i$	6	6	6	6	6

supponendo

- un numero di osservazioni o repliche costante in ogni gruppo:  $n = 6$ ,
- una varianza d'errore  $s_e^2$  uguale a **146,5**
- gdl  $v = 25$

è possibile verificare la significatività della differenza tra ogni coppia di medie.

Dopo aver ordinato le medie in ordine decrescente (o crescente, come altri preferiscono), in relazione al rango di ognuna di esse, si stima il numero di passi  $r$ , che in questo caso, con 5 medie, può variare da 2 a 5.

Punteggi per il test del campo di variazione multiplo di Duncan  
 $\alpha = 0.05$

V	$r =$ numero di passi ordinati tra le medie													
	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
2	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
3	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
4	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
5	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
6	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
7	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
8	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
9	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48
11	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.46	3.46	3.46	3.47	3.48
12	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48
13	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45	3.46	3.46	3.47	3.47
14	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47
15	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47
16	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47
17	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47
18	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47
19	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47
20	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.46	3.47
22	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47
24	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47
26	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47
28	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47
30	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47
60	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47
100	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47
$\infty$	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47

Punteggi per il test del campo di variazione multiplo di Duncan  
 $\alpha = 0.01$

n	r = numero di passi ordinati tra le medie													
	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	8.26	8.50	8.60	8.70	8.80	8.90	8.90	9.00	9.00	9.00	9.10	9.20	9.30	9.30
4	6.51	6.80	6.90	7.00	7.10	7.10	7.20	7.20	7.30	7.30	7.40	7.40	7.50	7.50
5	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.50	6.60	6.60	6.70	6.70	6.80
6	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.00	6.10	6.20	6.20	6.30	6.30
7	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.80	5.80	5.90	5.90	6.00	6.00
8	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.50	5.60	5.70	5.70	5.80	5.80
9	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.40	5.50	5.50	5.60	5.70	5.70
10	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55
11	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39
12	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.23	5.26
13	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15
14	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91	4.96	5.00	5.04	5.06	5.07
15	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00
16	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84	4.88	4.91	4.93	4.94
17	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89
18	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.85
19	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82
20	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	4.76	4.78	4.79
22	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.68	4.71	4.74	4.75
24	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72
26	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69
28	3.91	3.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67
30	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65
40	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.57	4.59
60	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53
100	3.71	3.86	3.93	4.06	4.11	4.17	4.21	4.25	4.29	4.35	4.38	4.42	4.45	4.48
$\infty$	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41

Ognuna delle 10 differenze (colonna 5) risulta significativa, se superiore alla **differenza minima significativa** (MDS), ottenuta con la solita formula

$$\text{MDS} = C_{(\alpha, r, n)} \cdot \sqrt{\frac{s_e^2}{n}}$$

dove

$C_{(\alpha, r, v)}$  = valore riportato nella **tabella di Duncan**.

Nell'esempio e nella tabella sottostante, il valore di C è stato scelto per

- la probabilità  $\alpha = 0.05$
- il numero **r** di passi relativi al confronto
- $v = 24$ , in assenza del valore per gdl 25, in quanto più cautelativo (determina infatti una MDS maggiore di quella che si otterrebbe scegliendo in valore per  $v = 26$ )

1	2	3	4	5	6	7
Ordine	Confronto	r passi	Medie	Differenza	Q	DUNCAN
1	5 <sup>a</sup> vs 1 <sup>a</sup>	5	208,2 – 119,1	89,1	4,166	3,22
2	5 <sup>a</sup> vs 2 <sup>a</sup>	4	208,2 – 123,3	84,9	3,901	3,15
3	5 <sup>a</sup> vs 3 <sup>a</sup>	3	208,2 – 141,0	67,2	3,532	3,07
4	5 <sup>a</sup> vs 4 <sup>a</sup>	2	208,2 – 199,8	8,4	2,919	2,92
5	4 <sup>a</sup> vs 1 <sup>a</sup>	4	199,8 – 119,1	80,7	3,901	3,15
6	4 <sup>a</sup> vs 2 <sup>a</sup>	3	199,8 – 123,3	76,5	3,532	3,07
7	4 <sup>a</sup> vs 3 <sup>a</sup>	2	199,8 – 141,0	58,8	2,919	2,92
8	3 <sup>a</sup> vs 1 <sup>a</sup>	3	141,0 – 119,1	21,9	3,532	3,07
9	3 <sup>a</sup> vs 2 <sup>a</sup>	2	141,0 – 123,3	17,7	2,919	2,92
10	2 <sup>a</sup> vs 1 <sup>a</sup>	2	123,3 – 119,1	4,2	2,919	2,92

Come tutti gli altri test per i confronti multipli, anche quello di Duncan presenta differenti gradi di combinazione tra **specificità e sensitività**. Quando dalla stima della differenza minima significativa per ogni singolo confronto si passa alla stima di una quantità valida per tutti i confronti, **si aumenta la versatilità**, ma **si diminuisce in potenza**: il valore della MDS cresce.

**I confronti multipli o a posteriori sono uno dei settori in maggiore evoluzione, nell'attuale ricerca statistica.** Di conseguenza, vengono proposti metodi nuovi e non esiste ancora unanimità sulle scelte più adeguate, in rapporto al confronto da effettuare e alle dimensioni del campione. Tuttavia, in molti testi sono enunciati alcuni principi fondamentali:

1- **è conveniente programmare l'esperimento e attuare confronti pianificati**, poiché è più probabile che le differenze che interessano risultino significative;

2 – con i confronti pianificati, il numero di test è limitato dall'ortogonalità e dalla conoscenza delle caratteristiche dei vari gruppi; quindi, **per scoprire differenze non attese, può essere necessario ricorrere ai confronti a posteriori;**

3 – **confrontare la differenza tra la media minore e quella maggiore** (individuate sulla base dei dati raccolti) **è un test a posteriori**: la probabilità va calcolata sul numero di confronti possibili;

4 – **tra tutti i test a posteriori forniti dai programmi informatici, secondo vari autori di testi di statistica applicata appare lecito utilizzare il test che risulta più significativo, poiché sono tutti cautelativi e non è ancora dimostrata la maggiore correttezza di uno rispetto agli altri.;**

5 – per confronti tra **tutte le singole medie** di  $p$  gruppi, è possibile scegliere tra test di Tukey e test di Duncan: il primo è più cautelativo, ma meno potente.

#### **8.5. STIMA DELLA DIMENSIONE $N$ DI $K$ GRUPPI CAMPIONARI PER L'ANOVA**

Al momento di programmare il confronto tra più medie campionarie, un problema fondamentale è sapere **quanti dati è necessario raccogliere**, ovviamente allo scopo di rendere il test significativo.

Le **dimensioni  $n$  di ognuno dei  $k$  campioni dipendono essenzialmente da 4 variabili**, che occorre conoscere o determinare al momento della programmazione:

1 – la **differenza minima  $d$  tra almeno 2 medie**, di cui si intende verificare la significatività; la scelta del valore dipende dalla conoscenza del fenomeno o da uno studio preliminare;  
quanto minore è  $d$  tanto maggiore deve essere la dimensione  $n$  di ogni campione;

2 – la **deviazione standard  $s$ , tratta dalla corrispondente varianza d'errore**; anche in questo caso deve essere nota attraverso dati riportati in letteratura, per l'esperienza del ricercatore oppure determinata da un esperimento pilota;  
quanto minore è  $s$  tanto minore può essere la dimensione  $n$  di ogni campione;

3 – la **probabilità  $\alpha$** , alla quale si vuole che la differenza  $d$  risulti significativa, **in un test bilaterale**; di norma è fissato uguale a 0.05 oppure a 0.01;

quanto minore è  $a$  tanto maggiore deve essere la dimensione  $n$  di ogni campione;

4 – **la potenza  $1 - b$  del test, la probabilità di rifiutare l'ipotesi nulla quando è falsa, tratta da una distribuzione per test unilaterali;** è prassi accettare una probabilità pari a 80% oppure 90%, corrispondente ad una probabilità di  $b$  uguale a 0.20 oppure 0.10;

tanto minore è  $b$ , tanto maggiore è la potenza richiesta al test e quindi tanto maggiore deve essere anche la dimensione  $n$  di ogni campione.

Nel caso di un'analisi della varianza in cui si confrontano le medie di  $k$  gruppi, ognuno con  $n$  dati, i gdl dell'errore standard sono quelli della varianza d'errore, quindi uguali a  $n = k \times (n-1)$ .

Poiché è sufficiente che sia significativa la differenza tra 2 delle  $k$  medie a confronto,

- per la probabilità  $a$  **si ricorre alla distribuzione  $t$  di Student per un test bilaterale,**
- per la probabilità  $b$  **alla stessa distribuzione  $t$  di Student, ma per un test unilaterale.**

Affinché il test sia sufficientemente potente,

$n$  deve essere maggiore od uguale a

$$n \geq 2 \left( \frac{s}{d} \right)^2 \cdot (t(a, n) + t(b, n))^2$$

La stessa relazione è più frequentemente scritta come

$$n \geq 2j^2 \cdot (t(a, n) + t(b, n))^2$$

(gli indici di  $t$ , entro parentesi, non sono stati riportati a pedice per renderli di lettura più facile).

E' da sottolineare che **la formula richiede di conoscere**

- **il rapporto  $s / d$** , spesso indicato con  $j$ ; è più facile da ottenere che non i singoli valori, in quanto simile ad un coefficiente di variazione;

per utilizzare un valore indicativo, quando non si hanno informazioni è utile ricordare che l'esperienza ha dimostrato che il valore

- $j \gg 0,2$  è piccolo (variabilità ridotta rispetto alla media);
- $j \gg 0,5$  è medio;
- $j \gg 0,7$  è grande (variabilità ampia rispetto al valore della media);

- **il valore di  $t$  alla probabilità  $b$**  deve essere preso dalla tabella dei valori critici e nello stesso modo con il quale viene scelto quello della probabilità  $a$  **per un test bilaterale. Per prassi**, la probabilità di  $\beta$  è circa 4-5 volte quella di  $\alpha$ ; di conseguenza
- quando si ha  $\alpha=0.01$  si sceglie un valore di  $\beta = 0.05$ ,
- quando si ha  $\alpha=0.05$  si sceglie un valore di  $\beta = 0.20$ .

Sarebbe possibile prendere anche un valore di  $\beta = 0.5$ , che corrisponde alla probabilità del 50% che il campione raccolto non risulti significativo alla probabilità  $a$  prefissata; in questo caso, il valore di  $t$  ha distribuzione simmetrica ed è uguale a **0**.

**Quando, come tabella dei valori critici, si dispone solo di una distribuzione bilaterale,** (vantaggiosa per trovare direttamente il valore di  $a$ ) **per trovare il valore di  $b$  si deve utilizzare la colonna 2b.**

**Il calcolo di  $n$  è ottenuto con un processo iterativo**, quando non è possibile ricorrere a metodi grafici.

Di seguito è riportato il processo di calcolo, in quanto utile a comprendere i fattori in gioco nella scelta delle dimensioni del campione; sono anche le informazioni richieste dai programmi informatici più diffusi.

Il valore di  $t$  dipende dal numero  $n$  di gdl, determinato sulla base del numero  $k$  di gruppi e soprattutto del numero  $n$  di osservazioni entro ogni gruppo:  $n = k \times (n-1)$ .

Il metodo iterativo richiede:

- a) una prima stima di  $n$ , considerando che ogni gruppo abbia almeno  $n = 5-6$  osservazioni; con 4 gruppi, il valore di  $n$  diventa uguale a 16 – 20 e sulla base di questi gdl si scelgono i due valori di  $t$  (quello alla probabilità  $a$  e quello alla probabilità  $b$ );
- b) se il calcolo determina un valore di  $n$  maggiore dei 5-6 preventivati (ad esempio 10), si stima un nuovo  $n$  (uguale a 36 poiché  $(10-1) \times 4 = 36$ ) e si scelgono dalla tabella sinottica due nuovi valori di  $t$ ;
- c) dopo il nuovo calcolo, spesso si può osservare che il terzo valore di  $n$  è vicino al secondo: si sceglie quello più cautelativo, arrotondato all'unità per eccesso. Se la differenza tra il terzo valore di  $n$  ed il secondo fosse ritenuta ancora importante, si effettua un nuovo calcolo dopo aver modificato i valori di  $t$  corrispondenti ai nuovi gdl; quasi sempre la quarta stima è molto simile alla terza e con essa termina il processo iterativo.

**ESEMPIO.** Mediante un'analisi della varianza con 4 gruppi (un controllo e tre trattamenti), si intende dimostrare la significatività di una differenza (tra il controllo ed uno dei tre trattamenti) uguale a 11.

Dai dati già raccolti, è noto che la varianza è uguale a 150 e quindi  $\sigma$  è uguale a 12,2 (arrotondato alla prima cifra decimale), mentre il rapporto  $j$  ( $s / d$ ) è uguale a 0,9.

Quanti dati  $n$  occorre raccogliere per ognuno dei 4 campioni, affinché il test ANOVA risulti significativo alla probabilità  $\alpha$  uguale a 0.05 e con una potenza  $(1 - \beta)$  uguale al 90 per cento?

Risposta.

Si utilizza la formula

$$n \geq 2j^2 \cdot (t(a, n) + t(b, n))^2$$

in cui, con i dati del problema, si ha che

$$j (s / d) = 0,9$$

$$a = 0.05 \quad e \quad b = 0.10$$

**Nel 1° tentativo**, si scelgono i valori dei **gdl** e i valori di **t** corrispondenti, solo sulla base del buon senso (l'esperienza):

$$\text{con } k = 4 \quad e \quad v = 20,$$

se si ipotizza a priori che sia sufficiente  $n = 6$ ,

$$\text{poiché } n = k \times (n-1)$$

si devono scegliere i due valori di **t** con 20 gdl.

Dalla tabella dei valori critici si ricava che

$$t \text{ di } a (0.05, 20) = 2,086 \text{ (in una distribuzione per test bilaterale),}$$

$$t \text{ di } b (0.10, 20) = 1,325 \text{ (in una distribuzione per test unilaterale, corrispondente alla colonna 0.20 se la distribuzione è bilaterale).}$$

Dai parametri fissati, con la formula sopra riportata

si ottiene un valore di **n**

$$n \cong 2 \times 0,9^2 \times (2,086 + 1,325)^2 = 2 \times 0,81 \times 11.635 = 18,85$$

uguale a 19, per arrotondamento all'unità superiore.

Si può osservare che il valore stimato (19) è molto maggiore di quello ipotizzato all'inizio, uguale a 6.

Di conseguenza, il valore di **t** utilizzato con 20 gdl è errato e troppo grande in quanto fondato su pochi gdl. Si deve quindi procedere ad una iterazione, con un secondo tentativo di calcolo fondato su un valore di **t** più preciso.

**Nel 2° tentativo,**

prendendo come riferimento delle dimensioni di ogni gruppo  $n = 19$ ,

il valore di  $n$  è  $4 \times 18 = 72$ .

Poiché poche tabelle riportano i valori esatti di  $t$  per questo numero di gradi di libertà, ma approssimativamente per decine, come scelta cautelativa si utilizza  $n$  uguale a 70, che fornisce un valore di  $t$  maggiore di quello con 80 gdl e quindi anche un  $n$  maggiore.

I nuovi valori di  $t$  sono:

per  $\alpha = 0.05$  in un test bilaterale,  $t(0.05, 70) = 1,994$

per  $\beta = 0.10$  in un test unilaterale,  $t(0.10, 70) = 1,294$

La nuova stima di  $n$

$$n \cong 2 \times 0,9^2 \times (1,994 + 1,294)^2 = 2 \times 0,81 \times 10,81 = 17,51$$

risulta uguale a 18 per arrotondamento all'unità superiore.

Poiché il nuovo valore (18) non differisce sensibilmente dal valore calcolato in precedenza (19), si può concludere che per ognuno dei 4 gruppi sono sufficienti 18 o 19 dati.

L'esempio mette in evidenza che **per poter utilizzare pochi dati**, quindi avere un risparmio in costo di materiale e di tempo richiesti dall'esperimento,

è vantaggioso rendere il valore di  $j$  ( $s / d$ ) il **minimo** possibile, agendo

sulla **differenza**, affinché **sia grande** e

sulla **varianza** affinché **sia piccola**.

## CAPITOLO VIII

### ANALISI DELLA VARIANZA (ANOVA I) A UN CRITERIO DI CLASSIFICAZIONE E CONFRONTI TRA PIU' MEDIE

Nella ricerca sperimentale è frequente **il confronto simultaneo tra le medie di più di due gruppi**, formati da soggetti sottoposti a trattamenti differenti o con dati raccolti in condizioni diverse. Al fine di evidenziare tutte le possibili differenze significative tra le medie, **non è corretto ricorrere al test t di Student per ripetere l'analisi tante volte quanti sono i possibili confronti** a coppie tra i singoli gruppi; è errato “dragare” i dati con il test **t**, per effettuare a posteriori (*comparisons post-hoc*) tanti confronti tra le medie dei gruppi quante sono le combinazioni degli  $n$  trattamenti 2 a 2, alla ricerca di qualunque differenza tra le medie, da giustificare in seguito.

Con il metodo del **t** di Student, **si utilizza solo una parte dei dati** e la probabilità  $\alpha$  prescelta per l'accettazione dell'ipotesi nulla, **la probabilità di commettere un errore di primo tipo** (rifiutare l'ipotesi nulla quando in realtà è vera) **è valida solamente per ogni singolo confronto**. **Se i confronti sono numerosi, la probabilità complessiva che almeno uno di essi si dimostri significativo solo per effetto del caso aumenta proporzionalmente.**

Se è vera l'ipotesi nulla  $H_0$ , la probabilità che nessun confronto risulti casualmente significativo è  $(1-\alpha)^n$  dove  $n$  è il numero di confronti effettuati.

Per esempio, se si effettuano **10** confronti tra le medie di gruppi estratti a caso dalla stessa popolazione e per ognuno di essi  $\alpha$  è uguale a **0.05**, la probabilità che nessun confronto risulti casualmente significativo diminuisce a **0.60** (corrispondente a  $0,95^{10}$ ); reciprocamente, la probabilità complessiva che almeno uno risulti significativo solo per effetto di fluttuazioni casuali diventa **0.40**.

Espresso in termini più formali, effettuando  $k$  confronti con il test **t** di Student ognuno alla probabilità  $\alpha$ , la probabilità complessiva  $\alpha'$  di commettere almeno un errore di I tipo (che il test rifiuti l'ipotesi nulla quando in realtà essa è vera) diventa

$$\alpha' = 1 - (1 - \alpha)^k$$

Nell'analisi della varianza, con apparente paradosso dei termini, il confronto è tra due o più medie. Essa permette il **confronto simultaneo tra esse, mantenendo invariata la probabilità  $\alpha$  complessiva prefissata.**

L'ipotesi  $H_0$  e l'ipotesi alternativa  $H_1$  assumono una formulazione più generale, rispetto al confronto tra due medie:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

$H_1$ : non tutte le  $m_i$  sono uguali  
(oppure, almeno una  $m_i$  è diversa)

L'ipotesi nulla  $H_0$  afferma che **le medie delle popolazioni dalle quali sono estratti casualmente i vari campioni sono tra loro tutte uguali** oppure che tutti i campioni a confronto sono stati estratti dalla medesima popolazione.

L'ipotesi alternativa  $H_1$  può essere solamente bilaterale: **le medie non sono tra loro tutte uguali**. In questo caso, possono realizzarsi varie situazioni, delle quali le due più estreme sono:

- le medie sono tutte differenti tra loro,
- una sola media è diversa dalle altre.

La metodologia sviluppata per verificare la significatività delle differenze tra le medie aritmetiche di vari gruppi, chiamata **analisi della varianza** e sintetizzata in **ANOVA** dall'acronimo dell'inglese **ANalysis Of VAriance**, utilizza **la distribuzione F**.

Fondata sul rapporto tra varianze, è stata così denominata in onore di Sir Ronald Aylmer **Fisher** (1890-1962), giudicato il più eminente statistico contemporaneo e ritenuto il padre della statistica moderna. La metodologia attuale è tuttavia dovuta a **Snedecor**, un suo allievo, che ne perfezionò il metodo e ne semplificò la forma, per cui la distribuzione **F** è ricordata anche come distribuzione di **Fisher-Snedecor**.

Nel 1925 Fisher, al quale tra gli argomenti già affrontati si devono la definizione dei gradi di libertà, gli indici di simmetria e curtosi, il metodo esatto per tabelle  $2 \times 2$ , completò il metodo di Student per il confronto tra due medie. La sua proposta **permette di scomporre e misurare l'incidenza delle diverse fonti di variazione sui valori osservati di due o più gruppi**. E' **la metodologia che sta alla base della statistica moderna**, che ha progressivamente elaborato i principi enunciati fino alle analisi più complesse, con le quali si possono tenere in considerazione contemporaneamente molti fattori sia indipendenti che correlati.

**La grande rivoluzione introdotta dall'analisi della varianza rispetto al test t consiste nel differente approccio alla programmazione dell'esperimento**. L'approccio del test t risente del vecchio assioma che **la natura risponde solo a domande semplici**. Per organizzare un esperimento, il materiale con il quale formare i gruppi a confronto doveva essere il più omogeneo possibile. Per esempio, per confrontare l'effetto di due tossici su un gruppo di cavie, gli animali dovevano essere dello stesso sesso, della stessa età, della stessa dimensione, ecc., se si riteneva che sesso, età, peso e qualunque altro carattere noto incidessero sulla risposta dell'esperimento. La differenza tra i due gruppi poteva risultare più facilmente significativa, in quanto l'errore standard risultava indubbiamente minore; ma le conclusioni erano ovviamente limitate al gruppo di animali con le caratteristiche prescelte, senza possibilità di estensione a cavie con caratteri diversi. Per rendere più generali le conclusioni, non rimaneva che ripetere l'esperimento, variando un carattere alla volta. Era

richiesto un forte aumento della quantità di materiale ed un allungamento dei tempi necessari all'esperimento; alla fine, rimaneva complesso trarre conclusioni generali.

**La grande novità introdotta dall'analisi della varianza**, come verrà evidenziato progressivamente con analisi sempre più complesse che considerano contemporaneamente un numero sempre più elevato di fattori e di loro interazioni, **è la scoperta dei vantaggi offerti all'analisi dall'uso di materiale molto diversificato**. Conoscendo le cause ed i diversi fattori, è possibile attribuire ad ognuno di essi la parte di effetto determinata e contemporaneamente ridurre la variabilità d'errore. **Le differenze tra le medie dei gruppi diventano molto più facilmente significative e le conclusioni possono essere immediatamente estese alle varie situazioni**.

Dall'introduzione dell'analisi della varianza è vantaggioso usare materiale non omogeneo per tutti i caratteri.

Nell'analisi della varianza, la fonte o causa delle variazioni dei dati viene chiamata **fattore sperimentale** o **trattamento**; essa può essere

- a più **livelli** quantitativi, come le dosi crescenti dello stesso farmaco, oppure
- a diverse **modalità** qualitative, come la somministrazione di farmaci differenti.

Ogni unità od osservazione del gruppo sperimentale viene chiamata **replicazione** o **replica**; per permettere di calcolare la media e la varianza, ovviamente ogni gruppo deve essere formato da almeno due repliche

### **8.1. ANALISI DELLA VARIANZA AD UN CRITERIO DI CLASSIFICAZIONE O A CAMPIONAMENTO COMPLETAMENTE RANDOMIZZATO**

Il modello più semplice di analisi della varianza, che può essere visto come un'estensione del test **t** di Student a più campioni indipendenti, è detto **ad un criterio di classificazione**: ogni dato è classificato solo sulla base del trattamento o del gruppo al quale appartiene. E' chiamato anche **modello completamente randomizzato** in quanto, soprattutto per analisi di laboratorio, prevede **un campionamento in cui gli n individui omogenei sono assegnati casualmente ai vari livelli del fattore**.

Quando si dispone di un gruppo di soggetti (ad esempio, cavie) da sottoporre a diversi trattamenti per confrontarne gli effetti, l'attribuzione di ogni individuo ad uno specifico trattamento deve avvenire per estrazione casuale da tutto il gruppo.

La metodologia di presentazione delle osservazioni, ormai codificata, prevede che i dati sperimentali raccolti siano riportati in modo ordinato secondo la tabella sottostante. Per l'analisi statistica, in questo modello **non è richiesto che i vari gruppi abbiano lo stesso numero ( $n_i$ ) di osservazioni o di repliche**.

		MODALITA' O LIVELLI DEI TRATTAMENTI				
		T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	...	T <sub>p</sub>
UNITÀ' SPERIMENTALI O REPLICAZIONI		X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>	...	X <sub>1p</sub>
		X <sub>21</sub>	X <sub>22</sub>	X <sub>23</sub>	...	X <sub>2p</sub>
		X <sub>31</sub>	X <sub>32</sub>	X <sub>33</sub>	...	X <sub>3p</sub>
		...	...	...	...	...
		X <sub>n<sub>1</sub>1</sub>	X <sub>n<sub>2</sub>2</sub>	X <sub>n<sub>3</sub>3</sub>	...	X <sub>n<sub>p</sub>p</sub>
Medie dei trattamenti		$\bar{X}_{.1}$	$\bar{X}_{.2}$	$\bar{X}_{.3}$	...	$\bar{X}_{.p}$
Media generale		$\bar{\bar{X}}$				

La **singola osservazione**  $X_{ij}$  viene riportata con 2 indici, relativi uno al trattamento o gruppo e l'altro alla posizione occupata entro il gruppo.

La **media di ogni gruppo o singolo trattamento**  $\bar{X}_i$  è riportata soprasssegnata da un tratto e con l'indice relativo al gruppo.

La **media generale**  $\bar{\bar{X}}$  di tutti i dati è indicata con un duplice tratto e senza indici.

**A partire da queste tre quantità, si stimano le devianze e le varianze utili all'analisi.**

**L'analisi della varianza è fondata sugli effetti additivi dei vari fattori considerati.** Nel modello più semplice, che considera un solo fattore a due o più livelli, ogni singola osservazione  $X_{ij}$  può essere scritta come

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

in quanto determinata

- dalla **media generale**  $\mu$ , che definisce la dimensione dell'esperimento,
- dal **fattore**  $\alpha_i$  **del trattamento** e
- da un **fattore casuale**  $\varepsilon_{ij}$ , detto **residuo** od **errore sperimentale**.

(E' importante ricordare che **errore** non è sinonimo di sbaglio, ma **indica l'effetto di uno o più fattori sconosciuti, comunque non valutati o non controllati nell'esperimento**).

In tale modello, l'**effetto** a **del trattamento** a sua volta è misurato come

$$\alpha_i = \mu_i - \mu$$

dove  $\mu_i$  è la media del trattamento e  $\mu$  la media generale.

Passando dall'enunciazione teorica ai dati sperimentali, si può scrivere che ogni singolo dato  $X_{ij}$  di uno specifico trattamento

$$X_{ij} = \bar{\bar{X}} + (\bar{X}_i - \bar{\bar{X}}) + \varepsilon_{ij}$$

è determinato

- dalla media generale  $\bar{\bar{X}}$ ,
- dall'effetto del trattamento  $(\bar{X}_i - \bar{\bar{X}})$  e
- da altri fattori non noti, simboleggiati da  $\varepsilon_{ij}$ .

Prima dell'applicazione di questo test parametrico, occorre verificare se ne esistono le condizioni.

**Le assunzioni di validità del test F dipendono dagli errori  $\varepsilon_{ij}$ , che**

- **devono essere tra loro indipendenti,**
- **devono essere distribuiti normalmente;** inoltre
- **le varianze dei vari gruppi devono essere omogenee.**

L'**indipendenza degli errori** comporta che la variazione casuale di ogni osservazione non sia influenzata da quella di un'altra: l'errore di una replica, il suo scarto rispetto alla media del gruppo di appartenenza, non deve essere influenzato né dal segno (quando si possono avere valori sia negativi che positivi) né dalle dimensioni del suo valore.

A questo fine, la randomizzazione deve essere fondata su elementi obiettivi (**effetto random**) e non lasciata all'arbitrio o all'intuito dello sperimentatore; ogni dato deve avere la stessa possibilità di essere influenzato dai fattori noti (**effetto trattamento**) e da quelli ignoti (**effetto ambiente** statistico). L'attribuzione (che sarà discussa nel capitolo sul campionamento) deve avvenire con modalità indipendenti dal ricercatore.

Gli **errori devono essere distribuiti normalmente intorno alla media**. Prima dell'applicazione del test deve essere attuato il controllo dell'asimmetria e della curtosi della distribuzione, per verificare che non si discosti eccessivamente dalla normale. Quando lo scostamento è significativo, sovente è possibile ricostruire le condizioni di validità attraverso la trasformazione dei dati (che saranno presentate successivamente).

L'**omogeneità della varianza**, per cui i diversi gruppi dei quali si confrontano le rispettive medie devono avere tutti la stessa varianza vera ( $\sigma^2$ ), è indispensabile per non determinare perdite nell'informazione sull'effetto dei trattamenti. Anche in questo caso, può essere necessario ricorrere alla trasformazione dei dati.

Dopo l'analisi dei dati per la verifica delle condizioni di validità, la **metodologia dell'analisi della varianza prevede il calcolo** delle seguenti quantità:

- la **devianza totale**, con i suoi **gdl**;
- la devianza **tra trattamenti** o **between**, con i suoi **gdl** e la **varianza** relativa;
- la devianza **entro trattamenti** o **within** od **errore**, con i suoi **gdl** e la **varianza** relativa.

Ai fini di una verifica dei risultati e delle successive loro elaborazioni, è utile ricordare che **la somma della devianza tra trattamenti e di quella entro trattamenti è uguale alla devianza totale; identica proprietà additiva hanno i rispettivi gradi di libertà.**

Devianze, gdl e varianze di un'analisi della varianza abitualmente vengono presentate come nella tabella seguente:

<b>Devianza Totale</b>	<b>gdl = n-1</b> (n = num. dati)	
<b>Devianza tra trattamenti</b>	<b>gdl = p-1</b> (p = num. gruppi)	<b>varianza tra</b> $S^2_{tra}$
<b>Devianza entro trattamenti</b>	<b>gdl = n-p</b>	<b>varianza entro</b> $S^2_{entro}$

(molti testi riportano la devianza totale e i suoi gdl alla fine, in quanto somma dei precedenti)

La **devianza totale** o **SQ totale** (Somma dei Quadrati degli scarti, in inglese **SS** da Sum of Squares) è calcolato da

$$SQ_{totale} = \sum_{j=1}^p \sum_{i=1}^{n_i} (X_{ij} - \bar{X})^2 = \sum_{j=1}^p \sum_{i=1}^{n_i} X_{ij}^2 - \frac{(\sum_{j=1}^p \sum_{i=1}^{n_i} X_{ij})^2}{n}$$

La prima è chiamata **formula euristica**, in quanto definisce il significato della devianza totale: la somma del quadrato degli scarti di ogni valore dalla media generale.

La seconda è la **formula abbreviata**, matematicamente equivalente alla prima, che rende più semplici e rapidi i calcoli necessari. Con essa, la devianza totale è ottenuta come differenza tra la somma dei quadrati di tutti i dati e il quadrato della somma di tutti i dati diviso il numero di dati.

La seconda formula ha il vantaggio di richiedere meno operazioni e di non utilizzare la media, che spesso è un valore approssimato; in queste condizioni, consente un calcolo più preciso della formula euristica.

La **devianza tra trattamenti** ( $SQ_{tra}$ ) o **between**

$$SQ_{tra} = \sum_{j=1}^p n_j (\bar{X}_j - \bar{\bar{X}})^2 = \sum_{j=1}^p \frac{\left(\sum_{i=1}^{n_j} X_{ij}\right)^2}{n_j} - \frac{\left(\sum_{j=1}^p \sum_{i=1}^{n_j} X_{ij}\right)^2}{n}$$

è per definizione (**formula euristica**) la somma dei quadrati degli scarti di ogni media di gruppo dalla media generale, moltiplicato il numero di dati del gruppo relativo.

La **formula abbreviata** utilizza le somme dei gruppi e la somma totale, determinando una maggiore precisione nei risultati.

La **devianza entro trattamenti** ( $SQ_{entro}$ ) o **within**, detta anche **errore**

$$SQ_{entro} = \sum_{j=1}^p \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2 = SQ_{totale} - SQ_{tra}$$

è la somma degli scarti al quadrato di ogni valore dalla media del suo gruppo.

Per la proprietà additiva delle devianze, può essere ottenuta sottraendo alla devianza totale la devianza tra trattamenti.

**I gradi di libertà sono determinati dal numero di somme richieste dal calcolo delle devianze relative**, nella formula euristica.

Per la **devianza totale**, dove la sommatoria è estesa a tutti gli **n** dati, i gdl sono **n-1**.

Per la **devianza tra trattamenti**, dove la sommatoria è estesa ai **p** gruppi, i gdl sono **p-1**.

Per la **devianza entro od errore**, la sommatoria è estesa a tutti i dati entro ogni gruppo. Per calcolare i gdl occorre quindi sottrarre 1 ai dati di ogni gruppo e quindi è determinata da **n-p**.

Per la proprietà additiva dei gdl, può essere scritta anche come **(n-1) - (p-1)**, semplificato in **n-p**.

Dividendo la devianza tra trattamenti e quella entro trattamenti per i rispettivi gradi di libertà, si ottengono la **varianza tra** e la **varianza entro** (la varianza totale è priva d'interesse ai fini di questo test).

**La varianza fra gruppi misura le differenze esistenti tra un gruppo e l'altro**, anche se il calcolo viene attuato rispetto alla media generale.

**La varianza entro gruppi misura la variabilità esistente attorno alla media aritmetica di ogni gruppo.**

**Se è vera l'ipotesi nulla**, i dati dei vari gruppi sono estratti casualmente dalla stessa popolazione. La varianza tra le medie dei trattamenti e la varianza entro ogni gruppo dipendono dalla variabilità esistente tra i dati: **varianza fra** ( $s_F^2$ ) e **varianza entro** ( $s_e^2$ ) sono due stime indipendenti della stessa varianza vera  $s^2$  e quindi **dovrebbero avere statisticamente lo stesso valore**.

Come indice dell'uguaglianza tra le due varianze, viene utilizzato

il test **F di Fisher**, fondato sul rapporto

**varianza-tra / varianza-entro**

indicato con la simbologia

$$F_{(p-1, n-p)} = \frac{S_F^2}{S_e^2}$$

Se è vera l'ipotesi nulla **H<sub>0</sub>**

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

il rapporto dovrebbe risultare **uguale ad 1**.

Se è vera l'ipotesi alternativa **H<sub>1</sub>**

$$H_1: \text{le } m_i \text{ non sono tutte uguali}$$

il rapporto dovrebbe risultare **superiore a 1**.

Con un numero infinito di trattamenti e di repliche, è sufficiente un rapporto superiore a 1 per rifiutare l'ipotesi nulla (come mostra la tabella dei valori critici di F); con un numero ridotto di dati, il rapporto può essere superiore a 1, per effetto delle variazioni casuali. I valori critici per i rispettivi gradi di libertà sono forniti dalla distribuzione F.

**Se il valore di F calcolato supera quello tabulato, alla probabilità prefissata, si rifiuta l'ipotesi nulla** e si accetta l'ipotesi alternativa: almeno una media è diversa dalle altre.

**Se il valore F calcolato è inferiore a quello riportato nella tabella, si accetta l'ipotesi nulla, o almeno non può essere rifiutato che le medie sono tutte uguali.**

#### **Valori critici della distribuzione F di Fisher-Snedecor**

I gradi di libertà del numeratore (o varianza maggiore) sono riportati in orizzontale (prima riga)

I gradi di libertà del denominatore (o varianza minore) sono riportati in verticale (prima colonna)

$$\alpha = 0.05$$

NUMERATORE

DEN.	1	2	3	4	5	6	7	8	12	24	∞
<b>1</b>	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	243,9	249,1	254,3
<b>2</b>	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,41	19,45	19,50
<b>3</b>	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,74	8,64	8,53
<b>4</b>	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	5,91	5,77	5,63
<b>5</b>	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,68	4,53	4,36
<b>6</b>	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,00	3,84	3,67
<b>7</b>	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,57	3,41	3,23
<b>8</b>	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,28	3,12	2,93
<b>9</b>	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,07	2,90	2,71
<b>10</b>	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	2,91	2,74	2,54
<b>12</b>	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,69	2,51	2,30
<b>14</b>	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,53	2,35	2,13
<b>16</b>	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,42	2,24	2,01
<b>18</b>	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,34	2,15	1,92
<b>20</b>	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,28	2,08	1,84
<b>30</b>	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,09	1,89	1,62
<b>40</b>	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,00	1,79	1,51
<b>60</b>	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	1,92	1,70	1,39
<b>120</b>	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,83	1,61	1,25
<b>∞</b>	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,75	1,52	1,00

**Valori critici della distribuzione F di Fisher-Snedecor**

I gradi di libertà del numeratore (o varianza maggiore) sono riportati in orizzontale (prima riga)

I gradi di libertà del denominatore (o varianza minore) sono riportati in verticale (prima colonna)

$$\alpha = 0.025$$

NUMERATORE

DEN.	1	2	3	4	5	6	7	8	12	24	∞
<b>1</b>	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7	976,7	997,2	1018
<b>2</b>	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,41	39,46	39,50
<b>3</b>	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,34	14,12	13,90
<b>4</b>	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,75	8,51	8,26
<b>5</b>	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,52	6,28	6,02
<b>6</b>	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,37	5,12	4,85
<b>7</b>	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,67	4,42	4,14
<b>8</b>	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,20	3,95	3,67
<b>9</b>	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	3,87	3,61	3,33
<b>10</b>	6,94	5,46	4,83	4,46	4,24	4,06	3,95	3,85	3,62	3,37	3,08
<b>12</b>	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,28	3,02	2,72
<b>14</b>	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,05	2,79	2,49
<b>16</b>	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	2,89	2,63	2,32
<b>18</b>	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,77	2,50	2,19
<b>20</b>	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,68	2,41	2,09
<b>30</b>	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,41	2,14	1,79
<b>40</b>	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,29	2,01	1,64
<b>60</b>	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,17	1,88	1,48
<b>120</b>	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,05	1,76	1,31
<b>∞</b>	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	1,94	1,64	1,00

**Valori critici della distribuzione F di Fisher-Snedecor**

I gradi di libertà del numeratore (o varianza maggiore) sono riportati in orizzontale (prima riga)

I gradi di libertà del denominatore (o varianza minore) sono riportati in verticale (prima colonna)

$$\alpha = 0.01$$

NUMERATORE

DEN.	1	2	3	4	5	6	7	8	12	24	∞
1	4052	5000	5403	5625	5764	5859	5928	5981	6106	6235	6366
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,41	99,46	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,05	26,60	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,37	13,93	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	9,89	9,47	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,72	7,31	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,47	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,67	5,28	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,11	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,71	4,33	3,91
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,16	3,78	3,36
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	3,80	3,43	3,00
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,55	3,18	2,75
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,37	3,00	2,57
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,23	2,86	2,42
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	2,84	2,47	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,66	2,29	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,50	2,12	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,34	1,95	1,38
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,18	1,79	1,00

**Valori critici della distribuzione F di Fisher-Snedecor**

I gradi di libertà del numeratore (o varianza maggiore) sono riportati in orizzontale (prima riga)

I gradi di libertà del denominatore (o varianza minore) sono riportati in verticale (prima colonna)

**a = 0.005**

NUMERATORE

DEN.	1	2	3	4	5	6	7	8	12	24	¥
1	16211	20000	21615	22500	23056	23437	23715	23925	24426	24940	25465
2	198,5	199,0	199,2	199,2	199,3	199,3	199,4	199,4	199,4	199,5	199,5
3	55,55	49,80	47,47	46,19	45,39	44,84	44,43	44,13	43,39	42,62	41,83
4	31,33	26,28	24,26	23,15	22,46	21,97	21,62	21,35	20,70	20,03	19,32
5	22,78	18,31	16,53	15,56	14,94	14,51	14,20	13,96	13,38	12,78	12,14
6	18,63	14,54	12,92	12,03	11,46	11,07	10,79	10,57	10,03	9,47	8,88
7	16,24	12,40	10,88	10,05	9,52	9,16	8,89	8,68	8,18	7,65	7,08
8	14,69	11,04	9,60	8,81	8,30	7,95	7,69	7,50	7,01	6,50	5,95
9	13,61	10,11	8,72	7,96	7,47	7,13	6,88	6,69	6,23	5,73	5,19
10	12,83	9,43	8,08	7,34	6,87	6,54	6,30	6,12	5,66	5,17	4,64
12	11,75	8,51	7,23	6,52	6,07	5,76	5,52	5,35	4,91	4,43	3,90
14	11,06	7,92	6,68	6,00	5,56	5,26	5,03	4,86	4,43	3,96	3,44
16	10,58	7,51	6,30	5,64	5,21	4,91	4,69	4,52	4,10	3,64	3,11
18	10,22	7,21	6,03	5,37	4,96	4,66	4,44	4,28	3,86	3,40	2,87
20	9,94	6,99	5,82	5,17	4,76	4,47	4,26	4,009	3,68	3,22	2,69
30	9,18	6,35	5,24	4,62	4,23	3,95	3,74	3,58	3,18	2,73	2,18
40	8,83	6,07	4,98	4,37	3,99	3,71	3,51	3,35	2,95	2,50	1,93
60	8,49	5,79	4,73	4,14	3,76	3,49	3,29	3,13	2,74	2,29	1,69
120	8,18	5,54	4,50	3,92	3,55	3,28	3,09	2,93	2,54	2,09	1,43
¥	7,88	5,30	4,28	3,72	3,35	3,09	2,90	2,74	2,36	1,90	1,00

ESEMPIO. Per un controllo della qualità dell'aria, con rilevazioni in tre diverse zone di una città (denominate A, B e C) è stata misurata anche la quantità di ferro (in microgrammi/Nmc a 0°C e 1013 mbar) tra i metalli pesanti in sospensione.

FATTORE SPERIMENTALE		
A	B	C
2,71	1,75	2,22
2,06	2,19	2,38
2,84	2,09	2,56
2,97	2,75	2,60
2,55		2,72
2,78		

Esiste una differenza significativa tra le tre zone, per la quantità di ferro in sospensione?

Risposta.

L'ipotesi nulla  $H_0$  è che tra le medie dei tre campioni non esistano differenze significative

$$H_0: \mu_A = \mu_B = \mu_C$$

mentre l'ipotesi alternativa  $H_1$

$H_1$ : le  $\mu_i$  non sono tutte uguali.

Attraverso il test F è possibile stimare la probabilità di trovare per caso tra le medie scarti uguali o superiori a quelli sperimentalmente osservati, nell'ipotesi che  $H_0$  sia vera.

Come primo passo, dalle tre serie di dati occorre calcolare

- il totale di ogni colonna  $\sum X_j$
- il numero di osservazioni  $n_j$
- la media di ogni colonna  $\bar{X}_j$

Successivamente, da essi è necessario stimare

- la somma totale  $\sum X$
- il numero totale di osservazioni  $n$
- la media totale o generale  $\bar{X}$

come riportato nella tabella successiva:

	A	B	C		
$\sum X_j$	15,91	8,78	12,48	$\sum X$	37,17
$n_j$	6	4	5	<b>n</b>	15
$\bar{X}_j$	2,652	2,195	2,496	$\bar{X}$	2,478

A partire da queste quantità, si calcolano le devianze ed i gradi di libertà rispettivi.

La **devianza totale** può essere calcolata dalla somma del quadrato degli scarti di ognuna delle 15 osservazioni rispetto alla media totale, in accordo con la formula euristica

$$SQ_{\text{totale}} = \sum_{j=1}^p \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2$$

A	B	C
$(2,71 - 2,478)^2$	$(1,75 - 2,478)^2$	$(2,22 - 2,478)^2$
$(2,06 - 2,478)^2$	$(2,19 - 2,478)^2$	$(2,38 - 2,478)^2$
$(2,84 - 2,478)^2$	$(2,09 - 2,478)^2$	$(2,56 - 2,478)^2$
$(2,97 - 2,478)^2$	$(2,75 - 2,478)^2$	$(2,60 - 2,478)^2$
$(2,55 - 2,478)^2$		$(2,72 - 2,478)^2$
$(2,78 - 2,478)^2$		

Svolgendo i calcoli e sommando i risultati

A	B	C
0,053824	0,529984	0,066564
0,174724	0,082944	0,009604
0,131044	0,150544	0,006724
0,242064	0,073984	0,014884
0,005184		0,058564
0,091204		
<b>0,698040</b>	<b>0,837456</b>	<b>0,156340</b>

$$\text{Devianza totale} = 0,698040 + 0,837456 + 0,156340 = 1,691836$$

si ottiene una devianza totale uguale a 1,691836 con 14 gdl.

Questo metodo di procedere al calcolo della devianza totale è lungo e determina stime non precise, quando la media generale è approssimata. Pertanto, **per il calcolo manuale è sempre conveniente utilizzare la formula abbreviata**

$$SQ_{\text{totale}} = \sum_{j=1}^p \sum_{i=1}^{n_j} X_{ij}^2 - \frac{(\sum_{j=1}^p \sum_{i=1}^{n_j} X_{ij})^2}{n}$$

che, applicata ai dati dell'esempio,

	A	B	C	
	7,3441	3,0625	4,9284	
	4,2436	4,7961	5,6644	
	8,0656	4,3681	6,5536	
	8,8209	7,5625	6,7600	
	6,5025		7,3984	
	7,7284			
$\sum X^2$	<b>42,7051</b>	<b>19,7892</b>	<b>31,3048</b>	<b>93,7991</b>

dalle due diverse somme stima la

$$Devianza_{\text{totale}} = 93,7991 - \frac{(37,17)^2}{15} = 1,69184$$

devianza totale che risulta uguale a 1,69184.

La corrispondenza tra le due stime è una dimostrazione elementare ed intuitiva dell'equivalenza matematica delle due formule (la differenza tra 1,691836 e 1,69184 è dovuta agli arrotondamenti).

La **devianza tra trattamenti** o **between** misura la variabilità esistente tra la media aritmetica di ogni gruppo e la media aritmetica generale, ponderata per il numero di osservazioni presenti in ciascun gruppo. Se non esistesse la variabilità casuale ed il valore delle singole osservazioni fosse determinato solamente dal fattore specifico che le raggruppa, le repliche di ogni trattamento dovrebbero avere tutte lo stesso valore ed essere uguali alla media del gruppo, come evidenzia la formula euristica

$$SQ_{tra} = \sum_{j=1}^p n_j (\bar{X}_j - \bar{X})^2$$

**La devianza tra trattamenti o between è la somma degli scarti di ogni media di gruppo rispetto alla media generale, ponderata per il numero di repliche.**

Pertanto con la formula euristica il calcolo diventa:

$$\begin{aligned} \text{Devianza}_{tra} &= 6 \cdot (2,652 - 2,478)^2 + 4 \cdot (2,195 - 2,478)^2 + 5 \cdot (2,496 - 2,478)^2 = \\ &= 6 \cdot 0,030276 + 4 \cdot 0,080089 + 5 \cdot 0,000324 = \\ &= 0,181656 + 0,320356 + 0,00162 = 0,503632 \end{aligned}$$

e risulta uguale a 0,503632 con 2 gradi di libertà.

Anche in questo caso la formula abbreviata

$$SQ_{tra} = \sum_{j=1}^p \frac{\left( \sum_{i=1}^{n_j} X_{ij} \right)^2}{n_j} - \frac{\left( \sum_{j=1}^p \sum_{i=1}^{n_j} X_{ij} \right)^2}{n}$$

è più rapida e precisa, non richiedendo le approssimazioni determinate dalle medie;

$$SQ_{tra} = \frac{(15,91)^2}{6} + \frac{(8,78)^2}{4} + \frac{(12,48)^2}{5} - \frac{(37,17)^2}{15} = 92,610196 - 92,10726 = 0,502936$$

essa risulta uguale a 0,502936. Anche in questo caso le differenze sono minime (0,503632 e 0,502936), imputabili all'uso di un numero diverso di cifre decimali e alle differenti approssimazioni.

(Di solito sono sufficienti calcoli con due o tre cifre decimali; il numero più elevato qui utilizzato è motivato dalla necessità contingente di confrontare i risultati dei due metodi).

**La devianza entro trattamenti, within od errore**

$$SQ_{entro} = \sum_{j=1}^p \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2$$

misura la variazione tra il valore di ciascuna replica e la media aritmetica del suo gruppo.

Sommando queste differenze elevate al quadrato per ogni gruppo

A	B	C
$(2,71 - 2,652)^2$	$(1,75 - 2,195)^2$	$(2,22 - 2,496)^2$
$(2,06 - 2,652)^2$	$(2,19 - 2,195)^2$	$(2,38 - 2,496)^2$
$(2,84 - 2,652)^2$	$(2,09 - 2,195)^2$	$(2,56 - 2,496)^2$
$(2,97 - 2,652)^2$	$(2,75 - 2,195)^2$	$(2,60 - 2,496)^2$
$(2,55 - 2,652)^2$		$(2,72 - 2,496)^2$
$(2,78 - 2,652)^2$		

e sviluppando i calcoli si ottiene

	A	B	C
	0,003364	0,198025	0,076176
	0,350464	0,000025	0,013456
	0,035344	0,011025	0,004096
	0,101124	0,308025	0,010816
	0,010404		0,050176
	0,015376		
<b>Devianza<sub>entro</sub></b>	<b>0,516076</b>	<b>0,517100</b>	<b>0,154720</b>

$$\text{Devianza}_{\text{entro}} = 0,516076 + 0,517100 + 0,154720 = 1,187896$$

la devianza entro, che risulta uguale a 1,187896 con 12 gdl.

La devianza entro od errore può essere ottenuta molto più rapidamente per sottrazione della devianza tra dalla devianza totale, precedentemente calcolate:

$$\text{Devianza}_{\text{entro}} = \text{Devianza}_{\text{totale}} - \text{Devianza}_{\text{tra}} = 1,69184 - 0,502936 = 1,188904$$

Nello stesso modo, per la proprietà additiva, si possono calcolare i gdl:

$$\text{gdl}_{\text{entro}} = \text{gdl}_{\text{totale}} - \text{gdl}_{\text{tra}} = 14 - 2 = 12$$

Per una presentazione chiara e sintetica, **normalmente i valori calcolati sono riassunti in una tabella che riporta le tre devianze, i rispettivi gradi di libertà e le varianze utili al test:**

DEVIANZA		GDL	VARIANZA
<b>Totale</b>	1,69184	14	----
<b>Tra trattamenti (between)</b>	0,502936	<b>2</b>	<b>0,251468</b>
<b>Entro trattamenti (within)</b>	1,188904	<b>12</b>	<b>0,0990753</b>

Dividendo la devianza tra e la devianza entro per i rispettivi gradi di libertà, si ottengono la varianza tra e la varianza entro.

Dividendo la varianza tra per la varianza entro, si calcola il rapporto F, che deve essere riportato con i rispettivi gradi di libertà  $F_{(2,12)}$

$$F_{(2,12)} = \frac{0,251468}{0,0990753} = 2,538$$

Il valore critico di F con gdl 2 per il numeratore e 12 per il denominatore che è riportato nella tabella per la probabilità  $\alpha = 0.05$  è 3,89. Il valore calcolato (2,538) è inferiore a quello tabulato (3,89): la probabilità che l'ipotesi nulla sia vera è superiore al 5%. Di conseguenza, si accetta l'ipotesi nulla: i tre campioni sono stati estratti dalla stessa popolazione; non esiste una differenza significativa tra le 3 medie campionarie.

## 8.2. CONFRONTO TRA L'ANALISI DELLA VARIANZA CON DUE TRATTAMENTI E IL TEST t DI STUDENT PER 2 CAMPIONI INDIPENDENTI.

L'analisi della varianza può essere applicata anche a 2 soli trattamenti; per questo caso, è già stata presentata la metodologia del test **t** di Student. In realtà, test t e test F sono due modi solo apparentemente differenti per fare la stessa analisi: **il test t può essere visto come un caso speciale di analisi della varianza**, applicata solo a due gruppi; meglio ancora, l'analisi della varianza è l'estensione a più gruppi e a più fattori del test **t** di Student.

Nel caso di un solo fattore con due gruppi, tra t ed F esiste una relazione matematica precisa:

$$F_{(1,n)} = t_{(n)}^2$$

che ovviamente può anche essere scritta come

$$t_{(v)} = \sqrt{F_{(1,n)}}$$

dove n è il numero di gradi di libertà.

**Il valore di F con gradi di libertà 1 al numeratore e n al denominatore è uguale al quadrato di t con n gradi di libertà.**

**Le due distribuzioni dei valori critici per la stessa probabilità  $\alpha$  sono equivalenti**, come è possibile evidenziare dal semplice confronto tra le tabelle dei valori critici.

ESEMPIO. Due gruppi di 10 uova di *Daphnia magna*, estratte casualmente dallo stesso clone, sono state allevate in due vasche con diverse concentrazioni di cromo, per verificare se incidono significativamente sulla crescita.

Dopo un mese sono stati misurati gli individui sopravvissuti: 7 nel gruppo A e 8 nel gruppo B, con le dimensioni riportate:

A	2,7	2,8	2,9	2,5	2,6	2,7	2,8	---
B	2,2	2,1	2,2	2,3	2,1	2,2	2,3	2,6

Risposta.

L'ipotesi nulla è

$$H_0: \mu_A = \mu_B$$

e l'ipotesi alternativa  $H_1$  bilaterale è

$$H_1: \mu_A \neq \mu_B$$

Prima di procedere sia al test t che al test F, si deve **verificare se le due varianze sono omogenee**.

Quindi è preliminare al confronto tra le due medie il confronto tra le 2 varianze, per saggiare l'ipotesi nulla

$$H_0: \sigma_A^2 = \sigma_B^2$$

con l'ipotesi alternativa bilaterale

$$H_1: \sigma_A^2 \neq \sigma_B^2$$

A questo scopo si calcolano le 2 devianze ed i loro gradi di libertà, per stimare le varianze rispettive

	A	B
Devianza	0,10857	0,18000
Gdl	6	7
Varianza $s^2$	<b>0,018095</b>	<b>0,02571</b>

ed infine il rapporto F tra varianza maggiore (al numeratore) e varianza minore (al denominatore).

$$F_{(7,6)} = \frac{0,02571}{0,018095} = 1,42$$

Nella tabella dei valori critici, con 7 gradi di libertà per la varianza maggiore e 6 per la varianza minore, il valore critico alla probabilità  $\alpha = 0.05$  è uguale a 4,21. Il valore calcolato (1,42) è inferiore: di conseguenza, si accetta l'ipotesi nulla che le due varianze siano omogenee.

Di conseguenza, è corretto procedere al confronto tra le due medie.

**Per il test t di Student per 2 campioni indipendenti,**

si calcolano le due medie:

$$\text{media del gruppo A} = 2,714$$

$$\text{media del gruppo B} = 2,250$$

e la varianza mediata  $s_p^2$

$$s_p^2 = \frac{0,10825 + 0,18000}{6 + 7} = 0,022173$$

Da esse si stima il valore di t con gdl 13

$$t_{13} = \frac{2,714 - 2,250}{\sqrt{0,022173 \cdot \left(\frac{1}{7} + \frac{1}{8}\right)}} = 6,02$$

che risulta uguale a 6,02.

**Per l'analisi della varianza ad un criterio di classificazione,**

si devono calcolare la devianza totale, quella tra trattamenti e quella entro trattamenti, con i rispettivi gradi di libertà.

E' possibile una verifica dei calcoli effettuati, mediante la proprietà additiva delle devianze:

$$\text{Devianza totale} = \text{Devianza tra} + \text{Devianza entro}$$

	<b>Devianza</b>	<b>Gdl</b>	<b>Varianza</b>
<b>Totale</b>	1,093333	14	
<b>Tra</b>	0,804762	1	<b>0,804761</b>
<b>Entro</b>	0,288571	13	<b>0,022198</b>

Si calcolano la varianza tra e la varianza entro e da esse

si stima F con gdl 1 e 13

$$F_{(1,13)} = \frac{0,804761}{0,022198} = 36,25$$

che risulta uguale a 36,25.

E' semplice **verificare che le due risposte coincidono:**

$$t^2_{(13)} = F_{(1,13)}; \quad (6,02)^2 = 36,25$$

a meno delle approssimazioni dei calcoli effettuati.

Sulle tabelle dei valori critici del test t di Student e del test F di Fisher si controlla la probabilità, che per entrambe risulta ovviamente uguale e nettamente inferiore a 0.001.

Con entrambi i test si rifiuta l'ipotesi nulla alla stessa probabilità.

### **8.3. TEST PER L'OMOGENEITA' DELLA VARIANZA TRA PIU' CAMPIONI:**

#### **TEST DI HARTLEY, COCHRAN, BARTLETT, LEVENE**

**Il confronto tra medie con l'analisi della varianza richiede che i diversi gruppi abbiano varianze uguali.** Allontanarsi sensibilmente da questa condizione di validità influenza gravemente la varianza d'errore, quindi la significatività del test. Si utilizzerebbe una varianza d'errore media  $s^2$ , come stima della varianza vera  $\sigma^2$ , che risulterebbe troppo grande per alcuni trattamenti e troppo piccola se riferita ad altri.

Oltre alla verifica delle condizioni di validità per il confronto tra medie, spesso si ha anche **un interesse esplicito a un confronto tra le varianze.** Per esempio, gruppi di animali o piante geneticamente identici dovrebbero avere varianze significativamente minori di gruppi geneticamente eterogenei; gruppi cresciuti in condizioni ambientali molto differenti dovrebbero avere una varianza maggiore di gruppi allevati in condizioni simili; nelle analisi di laboratorio, uno strumento di misura più preciso od un reagente di qualità superiore dovrebbero fornire varianze minori rispetto a strumenti e reagenti di bassa qualità, in esperimenti ripetuti nelle stesse condizioni.

L'ipotesi di **omoscedasticità**, chiamata in alcuni testi in italiano anche **omoscedalità** od **omogeneità delle varianze**, nel caso di **più gruppi** richiede la verifica dell'ipotesi nulla

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_p^2$$

contro l'ipotesi alternativa

$$H_1 : \text{non tutte le } \sigma_i^2 \text{ sono uguali}$$

I metodi proposti sono numerosi; tra i più diffusi, di norma utilizzati anche nei programmi informatici standard per calcolatori, sono da ricordare

**A - il test  $F_{\max}$  di Hartley,**

**B - il test di Cochran,**

**C - il test di Bartlett,**

**D - il test di Levene.**

A) Il procedimento  **$F_{\max}$  di Hartley** è quello più semplice e rapido. Le difficoltà alla sua utilizzazione derivano solo dalla scarsa reperibilità di testi ad ampia diffusione con la tabella dei valori critici. Questa tabella (riportata nella pagina successiva) non è da confondere con quella di Fisher-Snedecor, riportata in tutti i testi.

Tale difficoltà è ora superata in molti programmi informatici recenti, che stampano il valore esatto della probabilità relativa.

Secondo il test di Hartley, **esiste una differenza significativa tra più varianze quando il rapporto tra la varianza maggiore  $s_{\max}^2$  e la varianza minore  $s_{\min}^2$**

$$F_{\max(p, n-1)} = \frac{s_{\max}^2}{s_{\min}^2}$$

**supera il valore critico** riportato nelle tabelle corrispondenti.

**Gli indici dei valori di  $F_{\max}$  considerano il numero p di gruppi a confronto simultaneo ed il numero di gradi di libertà n-1 di ogni gruppo.**

**Il test richiede che i gruppi abbiano tutti lo stesso numero n di osservazioni.**

E' un test semplice, ma non robusto: l'assunzione fondamentale è che i dati siano distribuiti normalmente. Se non è possibile supporre la normalità della distribuzione per ciascun gruppo, si dovrebbe ricorrere ad altri test, come quelli non parametrici.

**Valori critici per il test di Hartley  
sull'omogeneità della varianza tra k gruppi**

**$\alpha = 0.05$**

**numero k di varianze a confronto**

$Df_2$	2	3	4	5	6	7	8	9	10	11	12
2	39.0	87.5	142	202	266	333	403	475	550	626	704
3	15.4	27.8	39.2	60.7	62.0	72.9	83.5	93.9	104	114	124
4	9.60	15.5	20.6	26.2	29.5	33.6	37.5	41.1	44.6	48.0	51.4
5	7.15	10.3	13.7	16.3	18.7	20.8	22.9	24.7	26.5	28.2	29.9
6	5.82	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6	19.7	20.7
7	4.99	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3	15.1	15.8
8	4.43	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7	12.2	12.7
9	4.03	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91	10.3	10.7
10	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.01	9.34
12	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.25	7.48
15	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.77	5.93
20	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.49	4.59
30	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.36	3.39
60	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.33	2.36
$\infty$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

**$\alpha = 0.01$**

**numero k di varianze a confronto**

$Df_2$	2	3	4	5	6	7	8	9	10	11	12
2	199	448	729	1036	1362	1705	2063	2432	2813	3204	3605
3	47.5	85	120	151	184	216	249	281	310	337	361
4	23.2	37	49	59	69	79	89	97	106	113	120
5	14.9	22	28	33	38	42	46	50	54	57	60
6	11.1	15.5	19.1	22	25	27	30	32	34	36	37
7	8.89	12.1	14.5	16.5	18.4	20	22	23	24	26	27
8	7.50	9.9	11.7	13.2	14.5	15.8	16.9	17.9	18.9	19.8	21
9	6.54	8.5	9.9	11.1	12.1	13.1	13.9	14.7	15.3	16.0	16.6
10	5.85	7.4	8.6	9.6	10.4	11.1	11.8	12.4	12.9	13.4	13.9
12	4.91	6.1	6.9	7.6	8.2	8.7	9.1	9.5	9.9	10.2	10.6
15	4.07	4.9	5.5	6.0	6.4	6.7	7.1	7.3	7.5	7.8	8.0
20	3.32	3.8	4.3	4.6	4.9	5.1	5.3	5.5	5.6	5.8	5.9
30	2.63	3.0	3.3	3.4	3.6	3.7	3.8	3.9	4.0	4.1	4.2
60	1.96	2.2	2.3	2.4	2.4	2.5	2.5	2.6	2.6	2.7	2.7
$\infty$	1.00	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

**Infatti non esistono test parametrici adatti alla verifica della omogeneità della varianza, quando le distribuzioni dei dati si discostano dalla normalità.**

B ) Anche **il test proposto da Cochran** nel 1967 può essere applicato solo ad esperimenti bilanciati. E' metodologicamente semplice come il precedente e permette una verifica rapida dell'ipotesi nulla di omoscedasticità dei vari trattamenti.

E' fondato sul rapporto tra la varianza maggiore e la somma di tutte le altre varianze.

Si calcola il valore del rapporto  $R_{n,p}$

$$R_{n,p} = \frac{s_{\max}^2}{s_1^2 + s_2^2 + \dots + s_p^2}$$

dove

$s_{\max}^2$  la varianza campionaria maggiore,

$s_1^2, s_2^2, \dots, s_p^2$  sono le varianze dei **p** gruppi,

con un numero **n** di repliche uguali in ogni gruppo.

Anche in questo caso, i limiti derivano dall'esigenza di un numero uguale di osservazioni in tutti i gruppi e dalla ridotta diffusione delle tabelle specifiche. **Con un numero di osservazioni molto alto (infinito) il rapporto tende a 1/p.**

B) Più complessa è la metodologia per il **test di significatività approssimato di Bartlett**; basato su un principio di **J. Neyman e E. S. Pearson** (vedi, del 1931: *On the problem of k samples*. Bull. Acad. Polon. Sci. Lett. Ser. A, 3: 460-481), è stato presentato da **M. S. Bartlett nel 1937** in due articoli (vedi a: *Some examples of statistical methods of research in agriculture and applied biology*. Journal Royal Statist. Soc. Suppl. 4: 137-140; vedi b: *Properties of sufficiency and statistical tests*. Proc. Royal Statist. Soc. Ser. A, 160: 268-282).

Attualmente è il più diffuso e offre **due vantaggi** rispetto ai due test precedenti:

- 1) **i trattamenti a confronto possono contenere un numero di repliche differente;**
- 2) per verificare **la significatività** tra **p** gruppi utilizza la **distribuzione  $\chi^2_{(p-1)}$  con gradi di libertà p-1**, più facilmente reperibile delle distribuzioni specifiche precedenti di Hartley e Cochran.

Con **p** misure di varianze campionarie  $s^2$  che abbiano **gradi di libertà  $n_i$ , eventualmente tra loro diversi**, estratte casualmente da popolazioni distribuite in modo normale, il test approssimato di Bartlett segue **una distribuzione  $C^2_{(p-1)}$**  fondata sul rapporto

$$c_{(p-1)}^2 = \frac{M}{C}$$

dove

- **C** è il **fattore di correzione** proposto successivamente per utilizzare la distribuzione  $c_{(p-1)}^2$ .  
E' uguale a

$$C = 1 + \frac{1}{3 \cdot (p-1)} \cdot \left( \sum \left( \frac{1}{n_i} \right) - \frac{1}{\sum n_i} \right)$$

e risulta un valore prossimo ad 1.

- M è uguale a

$$M = \left( \sum n_i \cdot \ln \bar{s}^2 - \sum n_i \cdot \ln s_i^2 \right)$$

con  $\bar{s}^2$  = media ponderata delle varianze,

data da

$$\bar{s}^2 = \frac{\sum v_i \cdot s_i^2}{\sum v_i}$$

Per il calcolo di **M** (in alcuni testi è indicato con **B**), in diversi autori propongono l'uso del logaritmo a base 10, preferibile alla logaritmo naturale a base e;

quindi un altro modo per calcolare **M** è

$$M = 2.30259 \left[ (\log \bar{s}^2) \cdot \left( \sum n_i \right) - \sum n_i \log s_i^2 \right]$$

Questo test per l'omoschedasticità è molto potente, ma solo quando la distribuzione dei dati è normale.

Se la distribuzione dei dati è

- **platicurtica**, il valore della probabilità  $\alpha$  calcolata è più alto di quello reale; (il test è conservativo, meno potente: diventa più difficile rifiutare l'ipotesi nulla e quindi è più facile commettere un errore di II Tipo)
- **leptocurtica**, il valore della probabilità  $\alpha$  calcolata è più basso di quello reale, rovesciando i concetti e la conclusione precedenti.

### Valori critici $R_{(n,p)}$ di Cochran per il confronto simultaneo tra più varianze.

**n** = numero di osservazioni per gruppo, con campioni bilanciati.

**p** = numeri di gruppi o varianze a confronto simultaneo.

$$\alpha = 0.05$$

NUMERO **n** DI OSSERVAZIONI PER GRUPPO

<b>P</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>¥</b>
<b>2</b>	0,9985	0,9750	0,9392	0,9057	0,8772	0,8534	0,8332	0,8159	0,8010	0,5000
<b>3</b>	0,9669	0,8709	0,7977	0,7457	0,7071	0,6771	0,6530	0,6333	0,6167	0,3333
<b>4</b>	0,9065	0,7679	0,6841	0,6287	0,5895	0,5598	0,5365	0,5175	0,5017	0,2500
<b>5</b>	0,8412	0,6838	0,5981	0,5441	0,5065	0,4783	0,4564	0,4387	0,4241	0,2000
<b>6</b>	0,7808	0,6161	0,5321	0,4803	0,4447	0,4184	0,3980	0,3817	0,3682	0,1667
<b>7</b>	0,7271	0,5612	0,4800	0,4307	0,3974	0,3726	0,3535	0,3384	0,3259	0,1429
<b>8</b>	0,6798	0,5157	0,4377	0,3910	0,3595	0,3362	0,3185	0,3043	0,2926	0,1250
<b>9</b>	0,6385	0,4775	0,4027	0,3584	0,3286	0,3067	0,2901	0,2768	0,2659	0,1111
<b>10</b>	0,6020	0,4450	0,3733	0,3311	0,3029	0,2823	0,2666	0,2541	0,2439	0,1000

$$a = 0.01$$

NUMERO **n** DI OSSERVAZIONI PER GRUPPO

<b>P</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>¥</b>
<b>2</b>	0,9999	0,9950	0,9794	0,9586	0,9373	0,9172	0,8988	0,8823	0,8674	0,500
<b>3</b>	0,9933	0,9423	0,8831	0,8335	0,7933	0,7606	0,7335	0,7107	0,6912	0,3333
<b>4</b>	0,9676	0,8643	0,7814	0,7212	0,6761	0,6410	0,6129	0,5897	0,5702	0,2500
<b>5</b>	0,9279	0,7885	0,6957	0,6329	0,5875	0,5531	0,5259	0,5037	0,4854	0,2000
<b>6</b>	0,8828	0,7218	0,6258	0,5635	0,5195	0,4866	0,4608	0,4401	0,4229	0,1667
<b>7</b>	0,8376	0,6644	0,5685	0,5080	0,4659	0,4347	0,4105	0,3911	0,3751	0,1429
<b>8</b>	0,7945	0,6152	0,5209	0,4627	0,4226	0,3932	0,3704	0,3522	0,3373	0,1250
<b>9</b>	0,7544	0,5727	0,4810	0,4251	0,3870	0,3592	0,3378	0,3207	0,3067	0,1111
<b>10</b>	0,7175	0,5358	0,4469	0,3934	0,3572	0,3308	0,3106	0,2945	0,2813	0,1000

Il test può essere applicato su campioni non eccessivamente piccoli, per cui **si richiede che ogni varianza sia calcolata su un campione con almeno 5-6 osservazioni.**

Diversi autori sono molto critici sull'uso dei test per l'omogeneità della varianza. Infatti sono **fortemente alterati dalla non normalità della distribuzione e con pochi dati è impossibile**

**verificare se le varie distribuzioni campionarie possano essere ritenute prossime alla normale;** inoltre, **le varianze campionarie  $s^2$  devono evidenziare differenze molto grandi per risultare significative, poiché i test sono impostati per non rifiutare l'ipotesi nulla in condizioni d'incertezza.** Può quindi essere frequente il caso in cui varianze  $s^2$ , anche se rilevanti sotto l'aspetto ecologico od ambientale, non risultano significative ai test per l'omogeneità.

ESEMPIO. Per verificare l'esistenza di differenze nella qualità dell'aria, in 4 zone di una città si è misurata la quantità di solventi aromatici in sospensione.

Z O N E			
I	II	III	IV
190	138	173	198
210	149	164	207
205	128	185	232
208	136	179	184
206	152	188	193

Con le osservazioni riportate nella tabella, si intende verificare se le varianze dei quattro gruppi possono essere ritenute omogenee.

Risposta.

Si calcolano dapprima le 4 varianze,

V A R I A N Z E			
I	II	III	IV
63,20	96,81	92,70	335,70

ricordando che ognuna di esse ha 4 gdl.

L'ipotesi nulla è

$$H_0 : \sigma^2_I = \sigma^2_{II} = \sigma^2_{III} = \sigma^2_{IV}$$

mentre l'ipotesi alternativa  $H_1$  è che almeno una di esse sia diversa.

Con il **metodo di Hartley**, si calcola il rapporto tra la varianza maggiore (335,7) e la varianza minore (63,2): si ottiene un F con indici 4 (numero di gruppi) e 4 (numero di osservazioni entro ogni gruppo meno 1)

$$F_{(4,4)} = \frac{335,70}{63,20} = 5,30$$

che risulta uguale a 5,30.

Per la significatività si confronta il valore calcolato (5,3) con il valore tabulato alla probabilità prefissata per il medesimo numero di gruppi (4) e di gdl entro gruppi (4): per  $\alpha = 0.05$  risulta 20,6.

Con il **metodo di Cochran** si stima un rapporto R

$$R_{5,4} = \frac{335,70}{63,2 + 96,81 + 92,7 + 335,7} = 0,57$$

che per **n** uguale a 5 e **p** uguale a 4 risulta 0,57

Nelle tabelle, il valore critico alla probabilità  $\alpha = 0.01$  risulta uguale a 0,7212 e alla probabilità 0.05 uguale a 0,6287.

Il valore calcolato è inferiore, per cui non è dimostrata una differenza significativa tra le 4 varianze.

Con il **metodo di Bartlett**, si deve dapprima calcolare la varianza media  $\bar{s}^2$

$$\bar{s}^2 = \frac{\sum v_i \cdot s_i^2}{\sum v_i}$$

dividendo la somma delle 4 devianze per la somma dei gradi di libertà:

$$\bar{s}^2 = \frac{4 \cdot 63,2 + 4 \cdot 96,81 + 4 \cdot 92,7 + 4 \cdot 335,7}{16} = \frac{2353,64}{16} = 147,1$$

Successivamente si stimano

- sia il valore di M

$$M = (\sum n_i \cdot \ln \bar{s}^2 - \sum n_i \cdot \ln s_i^2)$$

che, con i dati dell'esempio,

$$\begin{aligned} M &= 16 \cdot \ln 147,1 - (4 \cdot \ln 63,2 + 4 \cdot \ln 96,81 + 4 \cdot \ln 92,7 + 4 \cdot \ln 335,7) = \\ &= 16 \times 4,991 - (4 \times 4,146 + 4 \times 4,573 + 4 \times 4,529 + 4 \times 5,186) = \\ &= 79,856 - (16,584 + 18,292 + 18,116 + 20,744) = 79,856 - 73,736 = 6,12 \end{aligned}$$

risulta uguale a **6,12**,

- sia il valore di C

$$C = 1 + \frac{1}{3 \cdot (p-1)} \cdot \left( \sum \left( \frac{1}{n_i} \right) - \frac{1}{\sum n_i} \right)$$

che risulta

$$C = 1 + \frac{1}{3 \cdot 3} \cdot \left( \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} - \frac{1}{16} \right) = 1 + \frac{1}{9} \cdot \left( \frac{4}{4} - \frac{1}{16} \right) = \\ = 1 + 0,111 \cdot (1 - 0,0625) = 1 + (0,111 \times 0,9375) = 1 + 0,104 = 1,104$$

uguale a **1,104**.

Il valore del  $c_{(3)}^2$  (chi quadrato con 3 gradi di libertà)

$$c_{(3)}^2 = \frac{M}{C} = \frac{6,12}{1,104} = 5,54$$

è uguale a **5,54**.

Nella tabella dei valori critici alla probabilità  $\alpha = 0,05$  il valore tabulato è 7,81. Il valore calcolato è inferiore: non si può rifiutare l'ipotesi nulla.

Secondo il test, i 4 gruppi a confronto hanno varianze non significativamente diverse.

Per accettare le conclusioni raggiunte mediante l'analisi, restano le perplessità già evidenziate sulla potenza di questi test e quelle che derivano dal numero ridotto di osservazioni per campione, come nel caso dell'esempio utilizzato.

D) Il **test di Levene** è l'estensione a K gruppi del metodo già illustrato per due campioni indipendenti. Ritenuto da vari statistici più robusto rispetto alla non normalità della distribuzione, deve la sua recente diffusione nella pratica statistica soprattutto all'inserimento in alcuni pacchetti statistici. Per la sua applicazione è necessario disporre dei dati originari, in quanto utilizza gli scarti di ogni valore campionario dalla media del suo gruppo.

Esso mette **a confronto gli scarti dei k gruppi**, cioè le differenze ( $d_i$ ) rispetto alle medie di gruppi,

$$d_i = x_i - \bar{x}$$

dopo

- averle elevate al quadrato

$$d_i = (x_i - \bar{x})^2$$

- oppure prese in valore assoluto

$$d_i = |x_i - \bar{x}|$$

per **eliminare i segni negativi**.

Come già dimostrato, **i due metodi forniscono risultati differenti**. Il metodo che utilizza gli scarti in valore assoluto attualmente è il più diffuso; ha una potenza maggiore, cioè una capacità maggiore di risultare significativo.

Per confrontare la varianza di K gruppi (A , B, C),  
con ipotesi nulla

$$H_0: \sigma^2_A = \sigma^2_B = \sigma^2_C$$

ed ipotesi alternativa

$$H_1: \text{non tutte le } \sigma^2 \text{ sono uguali} \quad \text{oppure} \quad H_1: \text{almeno due } \sigma^2 \text{ sono diverse tra loro}$$

la proposta di **Levene consiste nell'applicare alla k serie di scarti** (al quadrato o in valore assoluto) **l'analisi della varianza ad 1 criterio**, nell'assunzione che, **se i valori medi degli scarti risultano significativamente diversi, le k varianze dei dati originali sono diverse**.

Con un linguaggio più tecnico, se, utilizzando **gli scarti dalla media**, si rifiuta l'ipotesi nulla

$$H_0: \mu_A = \mu_B = \mu_C$$

per accettare l'ipotesi alternativa

$$H_1: \text{non tutte le } \mu \text{ sono uguali} \quad \text{oppure} \quad H_1: \text{almeno due } \mu \text{ sono diverse tra loro}$$

implicitamente deriva che

sui **dati originali** si rifiuta l'ipotesi nulla

$$H_0: \sigma^2_A = \sigma^2_B = \sigma^2_C$$

per accettare l'ipotesi alternativa

$$H_1: \text{non tutte le } s^2 \text{ sono uguali} \quad \text{oppure} \quad H_1: \text{almeno due } s^2 \text{ sono diverse tra loro}$$

Come nell'analisi della varianza ad un criterio, i gruppi possono avere un numero differente di osservazioni.

ESEMPIO. Riprendendo gli stessi dati dei tre test precedenti sull'omoscedaticità di  $k$  gruppi (dopo aver tolto due valori per costruire campioni non bilanciati)

Z O N E			
I	II	III	IV
190	138	173	198
210	149	164	207
205	128	185	232
208	136	179	184
206	---	188	---

verificare se le varianze dei 4 gruppi sono tra loro statisticamente differenti.

Risposta

Dopo aver calcolato le medie dei 4 gruppi

$$\bar{x}_I = \frac{190 + 210 + 205 + 208 + 206}{5} = \frac{1019}{5} = 203,8$$

$$\bar{x}_{II} = \frac{138 + 149 + 128 + 136}{4} = \frac{551}{4} = 137,75$$

$$\bar{x}_{III} = \frac{173 + 164 + 185 + 179 + 188}{5} = \frac{889}{5} = 177,8$$

$$\bar{x}_{IV} = \frac{198 + 207 + 232 + 184}{4} = \frac{821}{4} = 205,25$$

si calcolano le  $d_i$ , **gli scarti al quadrato**

$$d_i = (x_i - \bar{x})^2$$

ottenendo la tabella seguente

I	II	III	IV
190,44	0,0625	23,04	390,0625
38,44	126,5625	190,44	115,5625
1,44	95,0625	51,84	203,0625
17,64	3,0625	1,44	1139,0625
4,84	---	104,04	---

La verifica su questi scarti dell'ipotesi nulla

$$H_0: \mu_I = \mu_{II} = \mu_{III} = \mu_{IV}$$

corrisponde a quella

$$H_0: \sigma^2_I = \sigma^2_{II} = \sigma^2_{III} = \sigma^2_{IV}$$

sui dati originari.

A questo scopo, dopo i calcoli preliminari,

$$\begin{array}{cccc} \sum x_{ij}^2 = 1.617.742,98 & \sum x_{ij} = 2696,5 & & \\ \sum x_I = 252,8 & \sum x_{II} = 224,75 & \sum x_{III} = 371,2 & \sum x_{IV} = 1847,75 \end{array}$$

si stimano le devianze

$$SQ_{TOT} = 1.617.742,98 - \frac{2696,5^2}{18} = 1.213.793,3$$

$$SQ_{TRA} = \left( \frac{252,8^2}{5} + \frac{224,75^2}{4} + \frac{371,2^2}{5} + \frac{1.867,75^2}{4} \right) - \frac{2.696,5^2}{18} = 502.561,94$$

$$SQ_{ENTRO} = 1.213.793,3 - 502.561,94 = 711.231,36$$

e si completa la tabella dell'ANOVA

	<b>Devianza</b>	<b>Gdl</b>	<b>Varianza</b>
<b>Totale</b>	1,213.793,30	17	---
<b>Tra</b>	502.561,94	3	167.520,65
<b>Entro</b>	711.231,36	14	50.802,22

che conduce alla stima di

$$F_{3,14} = \frac{167.520,65}{50.802,22} = 3,297$$

**F = 3,297** con gdl **3** e **14**.

Il valore critico alla probabilità  $\alpha = 0,05$  è uguale a **3,34**: le 4 varianze a confronto risultano significativamente differenti, con una probabilità  $\alpha$  leggermente inferiore al 5%

Utilizzando gli scarti **in valore assoluto**, che ovviamente presentano una variabilità minore, si ottiene un test più potente. Dalla tabella

Z O N E			
I	II	III	IV
13,8	0,25	4,8	19,75
6,2	11,25	13,8	10,75
1,2	9,75	7,2	14,25
4,2	1,75	1,2	33,75
2,2	---	10,2	---

e dai calcoli preliminari

$$\sum x_{ij}^2 = 2.696,5 \quad \sum x_{ij} = 166,3$$

$$\sum x_I = 27,6 \quad \sum x_{II} = 23,0 \quad \sum x_{III} = 37,2 \quad \sum x_{IV} = 78,5$$

si stimano le devianze

$$SQ_{TOT} = 2.696,5 - \frac{166,3^2}{18} = 2.696,5 - 1.536,43 = 1.160,07$$

$$SQ_{TRA} = \left( \frac{27,6^2}{5} + \frac{23,0^2}{4} + \frac{37,2^2}{5} + \frac{78,5^2}{4} \right) - \frac{166,3^2}{18} = 566,5$$

$$SQ_{ENTRO} = 1.160,07 - 566,5 = 593,57$$

e si completa la tabella dell'ANOVA

	Devianza	Gdl	Varianza
<b>Totale</b>	1.160,07	17	----
<b>Tra</b>	566,50	3	188,83
<b>Entro</b>	593,57	14	42,40

che conduce alla stima di

$$F_{3,14} = \frac{188,83}{42,4} = 4,453$$

**F = 4,453** con gdl **3 e 14**.

Il valore critico con gdl **3 e 14**

- alla probabilità  $\alpha = 0.05$  è uguale a **3,34**
- alla probabilità  $\alpha = 0.025$  è uguale a **4,24**
- alla probabilità  $\alpha = 0.01$  è uguale a **5,35**

Le 4 varianze a confronto risultano significativamente differenti, con una probabilità inferiore al 2,5%

I quattro test proposti sono validi quando **la distribuzione dei dati è normale**; con l'allontanamento da essa, a causa di asimmetria e/o curtosi, tutti subiscono variazioni nella potenza e nella loro attendibilità ma in modo diverso. Per ovviare a questi inconvenienti, spesso è raccomandato il ricorso ai test non parametrici, presentati nei capitoli relativi (vedi indice).

#### **8.4. I CONFRONTI TRA PIU' MEDIE**

Quando con l'analisi della varianza viene rifiutata l'ipotesi nulla, si può essere interessati ad **individuare tra quali medie esista una differenza significativa**. Come nel confronto tra una situazione normale e diversi ambienti inquinati oppure in misure di tossicità con sostanze differenti, **sovente l'attenzione è più rivolta ai confronti specifici tra alcune medie che non ad una valutazione complessiva**, per la quale è quasi ovvio il rifiuto dell'ipotesi nulla.

Con più confronti, aumenta la probabilità di commettere errori di I tipo; in particolare, è errato applicare tante volte il test t di Student, nella convinzione che se un test risulta significativo, sia dimostrato che l'ipotesi nulla (tutte le medie sono uguali) sia da rifiutare.

I confronti possono essere semplici, condotti su coppie di trattamenti, oppure più complessi, tra gruppi diversi di trattamenti. Un'altra distinzione che verrà discussa in seguito è tra i **confronti a priori o pianificati e i confronti a posteriori (detti anche post-hoc) o multipli**. Esistono confronti che sono già programmati nella fase iniziale dell'esperimento, prima della raccolta di dati in natura o dei risultati dell'esperimento in laboratorio; altri che, in carenza di idee sulle differenze possibili tra i diversi gruppi, dopo la significatività del confronto globale evidenziata da un test F, vengono utilizzati per la **ricerca di qualunque confronto significativo, come utile indicazione dell'esistenza di effetti o trattamenti diversi, le cui cause restano da approfondire con ulteriori analisi**.

Il problema fondamentale è che con **p** gruppi, e quindi **con p-1 gradi di libertà nell'ANOVA**, **si possono avere solamente altrettanti confronti a coppie, come scomposizione della devianza tra trattamenti, se si vuole mantenere costante la probabilità  $\alpha$  prescelta**, mentre il numero teorico di

confronti possibili è molto più alto, dato dalle combinazioni di  $p$  elementi 2 a 2. Con 5 gruppi, si hanno 4 gradi di libertà e sono ammessi solo 4 confronti a coppie; tuttavia i confronti possibili 2 a 2 tra coppie di trattamenti ( $C^2_5$ ) sono 10, senza considerare i confronti fra insiemi diversi degli stessi 5 gruppi.

#### 8.4.1 CONFRONTI A PRIORI O PIANIFICATI OD ORTOGONALI

I confronti a priori, chiamati in molti testi anche confronti pianificati od ortogonali (*planned comparisons, orthogonal comparisons*), vengono prestabiliti durante la fase di programmazione dell'esperimento. Con essi è possibile organizzare sia confronti parziali che un **confronto globale**, in modo da analizzare le differenze tra le medie dei gruppi.

Come prosecuzione ed approfondimento dell'analisi della varianza, è possibile la scomposizione della devianza tra trattamenti e dei gdl relativi. Questi metodi presentano alcuni vantaggi:

- **utilizzano tutti i dati,**
- **per la stima dell'errore impiegano la varianza d'errore,**
- **non abbassano il valore di  $\alpha$  per ognuno dei confronti possibili e quindi sono più potenti di quelli non pianificati.**

Trascurando i confronti con due soli gruppi, dove i gradi di libertà della devianza tra trattamenti è 1 solo ed il confronto può essere solo tra le loro 2 medie, nel caso di 3 o più trattamenti è possibile operare con diverse modalità.

Come primo esempio, è possibile citare il caso di 3 gruppi: un controllo ( C ) e due trattamenti ( $A_1$ ,  $A_2$ ). La devianza tra trattamenti ha 2 gdl; se il test F risulta significativo è logica la sua successiva scomposizione in un primo confronto tra il controllo ( C ) contro (*versus*, abbreviato in vs) i due trattamenti ( $A_1 + A_2$ ) ed un secondo confronto tra i due trattamenti ( $A_1$  vs  $A_2$ ).

In un secondo esempio, che considera 4 gruppi, con la sostanza A alla concentrazione 5% e 10% (gruppi  $A_1$  e  $A_2$ ) e con la sostanza B alla concentrazione 8% e 30% si ha un'analisi della varianza con 3 gdl nella devianza tra trattamenti. Essi possono essere scomposti in

- 1 - un primo gdl per il confronto di A ( $A_1 + A_2$ ) contro B ( $B_1 + B_2$ ),
- 2 - un secondo gdl per il confronto di A al 5% ( $A_1$ ) contro A al 10% ( $A_2$ ),
- 3 - un terzo gdl per il confronto di B a concentrazione 8% ( $B_1$ ) contro la concentrazione 30% ( $B_2$ ).

I casi da portare come esempio possono essere numerosi e sensibilmente più complessi; **le scelte sui confronti da effettuare dipendono dalla conoscenza dei fattori che si vogliono sperimentare.**

E' fondamentale comprendere che i confronti che si possono effettuare devono essere tra loro indipendenti od ortogonali; in termini più semplici, significa che ogni confronto non deve fornire informazioni sul risultato di tutti gli altri.

Nell'esempio precedente con i 4 gruppi ( $A_1$ ,  $A_2$  e  $B_1$ ,  $B_2$ ) la significatività del confronto di A contro B non fornisce alcuna informazione di una eventuale significatività né tra  $A_1$  e  $A_2$ , né tra  $B_1$  e  $B_2$ ;

nessuno dei tre confronti fornisce informazioni sulla significatività degli altri due. Se invece si facessero confronti non indipendenti, ad esempio  $A_1$  contro  $B_1$  o  $B_2$ , l'eventuale significatività di A contro B aumenta la probabilità che siano significativi anche gli altri due.

Per una loro corretta impostazione tecnica è utile ricorrere ai **coefficienti polinomiali**.

E' un metodo proposto originariamente anche per semplificare i calcoli ed abbreviare i tempi richiesti dalla scomposizione della devianza tra trattamenti, come verrà dimostrato; ora, con la diffusione dei computer, sono rimasti utili soprattutto per impostare correttamente i confronti da effettuare, limitatamente ai gdl disponibili.

**Si definiscono confronti ortogonali solo quelli in cui**

- **la somma dei coefficienti per riga e**
- **quella dei loro prodotti per colonna sono entrambi uguali a zero.**

**Questi confronti godono della proprietà di rappresentare una corretta scomposizione della devianza tra trattamenti, con 1 gdl per ogni confronto, senza presentare correlazioni tra i risultati.** In una analisi ortogonale, **i differenti effetti sono stimati in modo indipendente, senza mutue interferenze e quindi senza alterazione della probabilità a prefissata di trovare differenze significative.**

Per il caso precedentemente citato dei 3 gruppi, un controllo (C) e due trattamenti ( $A_1$  e  $A_2$ ), si può verificare nella tabella seguente con i **coefficienti polinomiali** come la somma dei valori per riga sia zero e come, nel contempo, sia uguale a zero anche la somma dei prodotti delle colonne:

**Confronti ortogonali tra 3 gruppi (controllo C e 2 trattamenti  $A_1, A_2$ )**

<b>Gruppi</b>	<b>C</b>	<b><math>A_1</math></b>	<b><math>A_2</math></b>	<b>Somma per riga</b>
C contro $A_1 + A_2$	<b>+ 1</b>	<b>- 1/2</b>	<b>- 1/2</b>	<b>0</b>
$A_1$ contro $A_2$	<b>0</b>	<b>+1</b>	<b>-1</b>	<b>0</b>
Prodotti per colonna	+1 x 0	-1/2 x +1	-1/2 x -1	
<b>Totale colonna</b>	<b>0</b>	<b>-1/2</b>	<b>+1/2</b>	<b>0</b>

Se i gruppi fossero 4, un controllo e tre trattamenti, per un confronto del controllo contro gli altri tre, si darebbe valore 1 al controllo e 1/3 ad ogni trattamento.

Per semplificare ulteriormente i calcoli, quasi sempre i pesi attribuiti ad ogni gruppo per un confronto specifico sono cambiati nei corrispondenti numeri interi, moltiplicando i dati della stessa riga per il denominatore, per cui la riga +1, -1/2, -1/2 diventa +2, -1, -1.

La tabella seguente, analoga alla precedente, riporta i coefficienti polinomiali con gli interi

**Confronti ortogonali tra 3 gruppi (controllo C e 2 trattamenti A<sub>1</sub>, A<sub>2</sub>)**

<b>Gruppi</b>	<b>C</b>	<b>A<sub>1</sub></b>	<b>A<sub>2</sub></b>	<b>Somma per riga</b>
C contro A <sub>1</sub> + A <sub>2</sub>	+ 2	- 1	- 1	<b>0</b>
A <sub>1</sub> contro A <sub>2</sub>	<b>0</b>	+1	-1	<b>0</b>
Prodotti per colonna	+2 x 0	-1 x +1	-1 x -1	
<b>Totale colonna</b>	<b>0</b>	<b>-1</b>	<b>+1</b>	<b>0</b>

Il confronto di C contro A<sub>1</sub> + A<sub>2</sub> permette di verificare se il controllo ha effetti diversi rispetto ai due trattamenti, uniti in un gruppo solo; il confronto di A<sub>1</sub> contro A<sub>2</sub> permette di evidenziare se i due trattamenti determinano effetti tra loro differenti.

Nel caso di 4 gruppi, A<sub>1</sub> e A<sub>2</sub> con la stessa sostanza a concentrazioni diverse, B<sub>1</sub> e B<sub>2</sub> con altra sostanza a concentrazioni differenti, i confronti ortogonali diventano 3 come descritti dalla tabella:

**Confronti ortogonali tra 4 gruppi (A<sub>1</sub>+ A<sub>2</sub> contro B<sub>1</sub>+ B<sub>2</sub>)**

<b>Gruppi</b>	<b>A<sub>1</sub></b>	<b>A<sub>2</sub></b>	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>Somma riga</b>
A <sub>1</sub> e A <sub>2</sub> contro B <sub>1</sub> e B <sub>2</sub>	+1	+1	-1	-1	<b>0</b>
A <sub>1</sub> contro A <sub>2</sub>	+1	-1	<b>0</b>	<b>0</b>	<b>0</b>
A <sub>1</sub> contro B <sub>2</sub>	<b>0</b>	<b>0</b>	+1	-1	<b>0</b>
Prodotti per colonna	+1x+1x0	+1x-1x0	-1x0x+1	-1x0x-1	
<b>Totale colonna</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

Nel caso di un controllo (C) e tre trattamenti per valutare l'effetto di tre farmaci, di cui A contenente una principio attivo naturale mentre B<sub>1</sub> e B<sub>2</sub> contenenti 2 differenti prodotti di sintesi, i confronti potrebbero essere come nella tabella successiva.

**Confronti ortogonali tra 4 gruppi (C, A, B<sub>1</sub>, B<sub>2</sub>)**

<b>Gruppi</b>	<b>C</b>	<b>A</b>	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>Totale riga</b>
C contro A + B <sub>1</sub> + B <sub>2</sub>	+1	-1/3	-1/3	-1/3	0
A contro B <sub>1</sub> + B <sub>2</sub>	0	+1	-1/2	-1/2	0
B <sub>1</sub> contro B <sub>2</sub>	0	0	+1	-1	0
Prodotti per colonna	+1 x 0 x 0	-1/3 x +1 x 0	-1/3x-1/2x+1	-1/3x-1/2x-1	
<b>Totale colonna</b>	<b>0</b>	<b>0</b>	<b>+1/6</b>	<b>-1/6</b>	<b>0</b>

Usando, come d'abitudine, gli interi, si ottiene

**Confronti ortogonali tra 4 gruppi (C, A, B<sub>1</sub>, B<sub>2</sub>)**

<b>Gruppi</b>	<b>C</b>	<b>A</b>	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>Somma riga</b>
C contro A + B <sub>1</sub> + B <sub>2</sub>	+3	-1	-1	-1	0
A contro B <sub>1</sub> + B <sub>2</sub>	0	+2	-1	-1	0
B <sub>1</sub> contro B <sub>2</sub>	0	0	+1	-1	0
Prodotti per colonna	+3 x 0 x 0	-1 x +2 x 0	-1 x -1 x +1	-1 x -1 x -1	
<b>Totale colonna</b>	<b>0</b>	<b>0</b>	<b>+1</b>	<b>-1</b>	<b>0</b>

**La legge dell'ortogonalità**, presentata per tutti i confronti, è **valida anche per ogni coppia di confronti**.

Con i **coefficienti polinomiali** è possibile mostrare quando un confronto è errato, non è ortogonale. Per esempio, dopo aver verificato se esiste una differenza significativa del controllo rispetto ai tre farmaci insieme, è errato confrontare ancora il controllo rispetto ai farmaci B<sub>1</sub> e B<sub>2</sub>, sia insieme che separatamente. E' logico pensare che **il primo risultato fornisca informazioni sul secondo**: se tra il controllo e i farmaci risulta una differenza significativa, è maggiormente probabile che risulti significativa anche la differenza tra il controllo e i due farmaci B<sub>1</sub> e B<sub>2</sub>.

I **coefficienti polinomiali di questi due confronti non indipendenti**, presentati nella tabella successiva, infatti risultano

$$(3 \times 2) + (-1 \times 0) + (-1 \times -1) + (-1 \times -1) = 6 + 0 + 1 + 1 = 8$$

diversi da 0.

Nella tabella precedente, risultano indipendenti

- sia il confronto riportato nella riga 1 con quello della riga 2

$$(+3 \times 0) + (-1 \times +2) + (-1 \times -1) + (-1 \times -1) = 0 - 2 + 1 + 1 = 0$$

- sia il confronto riportato nella riga 2 con quello della riga 3

$$(0 \times 0) + (+2 \times 0) + (-1 \times +1) + (-1 \times -1) = 0 + 0 - 1 + 1 = 0$$

- sia il confronto riportato nella riga 1 con quello della riga 3

$$(+3 \times 0) + (-1 \times 0) + (-1 \times +1) + (-1 \times -1) = 0 + 0 - 1 + 1 = 0$$

in quanto la somma dei prodotti dei loro coefficienti è sempre uguale a 0.

#### Confronti ortogonali tra 4 gruppi (C, A, B<sub>1</sub>, B<sub>2</sub>)

Gruppi	C	A	B <sub>1</sub>	B <sub>2</sub>	Somma riga
C contro A + B <sub>1</sub> + B <sub>2</sub>	+3	-1	-1	-1	0
C contro B <sub>1</sub> + B <sub>2</sub>	+2	0	-1	-1	0
Prodotti per colonna	+3 x +2	-1 x 0	-1 x -1	-1 x -1	
Totale colonna	<b>6</b>	<b>0</b>	<b>+1</b>	<b>+1</b>	<b>+8</b>

Nel caso di **più gruppi, con lo stesso numero di osservazioni e con varianze (s<sup>2</sup>) uguali** (questa ultima d'altronde è richiesta dall'analisi della varianza come condizione di validità), con i coefficienti polinomiali è semplice **scomporre la devianza tra trattamenti**.

Per ogni singolo test riportato nelle righe dei coefficienti polinomiali, si devono calcolare

- sia la media generale di tutti i gruppi implicati ( $\bar{\bar{x}}$ ),
- sia la media di ognuno dei due sottogruppi ( $\bar{x}_i$ ) a confronto.

Il valore della devianza, con 1 gdl, è ottenuto dalla formula di quella tra trattamenti già presentata

$$\sum_{k=1}^2 (\bar{x}_{ki} - \bar{\bar{x}})^2 n_i$$

dove n<sub>i</sub> è il numero di osservazioni entro ogni sottogruppo.

L'applicazione è illustrata in modo chiaro nell'esempio successivo.

**ESEMPIO.** In cinque zone di una città, con 6 misure in ognuna è stata rilevata la presenza di solventi aromatici (microgrammi/Nmc a 0° e 1013 mbar): 2 stazioni di rilevazione (A e B) sono state collocate in centro, ai due angoli estremi della piazza principale; la 3<sup>a</sup> stazione di rilevazione (C) ancora in

centro, ma in una piazza secondaria; la 4<sup>a</sup> (D) e la 5<sup>a</sup> (E) in due zone periferiche, ai due estremi della strada principale, che attraversa la città.

I risultati sono stati

<b>ZONE</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
Medie	208,2	199,8	141,0	123,3	119,1
$n_i$	6	6	6	6	6

L'analisi della varianza ha dimostrato che tra le 5 zone esiste complessivamente una differenza significativa e la varianza d'errore ( $s^2_e$ ), con 25 gdl, è risultata uguale a 146,5 (ovviamente, può essere ottenuta come somma delle devianze entro ognuno dei 5 gruppi, con gdl totali 5 x 5).

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>	<b>F</b>
<b>TOTALE</b>	47.301,628	29		
<b>Tra trattamenti o zone</b>	43.639,128	4	10.909,782	74,47
<b>Errore</b>	3.662,500	25	146,500	

Mediante i confronti ortogonali, verificare tra quali zone esiste una differenza significativa nel livello medio d'inquinamento.

Risposta.

Con 5 gruppi e 4 gdl della varianza tra trattamenti, sono possibili 4 confronti ortogonali. Sulla base di quanto noto sulla collocazione delle 5 stazioni di rilevamento, appare logico e senza alternative plausibili lo schema riportato nella tabella successiva:

#### **Confronti ortogonali con relativi punteggi polinomiali**

4 Confronti ortogonali tra le 5 zone	A	B	C	D	E
Centro (A+B+C) vs Periferia (D+E)	+2	+2	+2	-3	-3
Piazza princ. (A+B) vs Piazza sec. (C)	+1	+1	-2	0	0
Piazza principale: A vs B	+1	-1	0	0	0
Periferia: D vs E	0	0	0	+1	-1

Dopo aver verificato che sono tutti confronti tra loro ortogonali

come, ad esempio, il primo ed il secondo

$$(+2 \times +1) + (+2 \times +1) + (+2 \times -2) + (-3 \times 0) + (-3 \times 0) = 2 + 2 - 4 + 0 + 0 = 0$$

si calcolano le **4 devianze relative**:

1 – la prima (centro contro periferia) con

$$\bar{\bar{x}} = 208,2 + 199,8 + 141,0 + 123,3 + 119,1 = 791,4 / 5 = 158,28$$

$$\bar{x}_1 = 208,2 + 199,8 + 141,0 = 549,0 / 3 = 183,0 \quad \text{e} \quad n_1 = 18$$

$$\bar{x}_2 = 123,3 + 119,1 = 242,4 / 2 = 121,2 \quad \text{e} \quad n_2 = 12$$

che risulta

$$(183,0 - 158,28)^2 \cdot 18 + (121,2 - 158,28)^2 \cdot 12 = 10.999,4112 + 16.499,1168 = 27.498,528$$

uguale a 27.498,528;

2 – la seconda devianza (piazza principale contro piazza secondaria del centro) con

$$\bar{\bar{x}} = 208,2 + 199,8 + 141,0 = 549,0 / 3 = 183,0$$

$$\bar{x}_1 = 208,2 + 199,8 = 408,0 / 2 = 204,0 \quad \text{e} \quad n_1 = 12$$

$$\bar{x}_2 = 141,0 \quad \text{e} \quad n_2 = 6$$

che risulta

$$(204 - 183)^2 \cdot 12 + (141 - 183)^2 \cdot 6 = 5.292 + 10.584 = 15.876$$

uguale a 15.876;

3 – la terza devianza (la stazione A contro la B della piazza principale) con

$$\bar{\bar{x}} = 208,2 + 199,8 = 408,0 / 2 = 204,0$$

$$\bar{x}_1 = 208,2 \quad \text{e} \quad n_1 = 6$$

$$\bar{x}_2 = 199,8 \quad \text{e} \quad n_2 = 6$$

che risulta

$$(208,2 - 204)^2 \cdot 6 + (199,8 - 204)^2 \cdot 6 = 105,84 + 105,82 = 211,68$$

uguale a 211,68;

4 – la quarta devianza (la stazione D contro la E delle periferie) con

$$\bar{\bar{x}} = 123,3 + 119,1 = 242,4 / 2 = 121,2$$

$$\bar{x}_1 = 123,3 \quad \text{e} \quad n_1 = 6$$

$$\bar{x}_2 = 119,1 \quad \text{e} \quad n_2 = 6$$

che risulta

$$(123,3 - 121,2)^2 \cdot 6 + (119,1 - 121,2)^2 \cdot 6 = 26,46 + 26,46 = 52,92$$

uguale a 52,92.

(AVVERTENZA: Se i confronti sono ortogonali e non sono stati commessi errori di calcolo, la somma delle 4 devianze calcolate

$$27.498,528 + 15.876 + 211,68 + 52,92 = \mathbf{43.639,128}$$

risulta esattamente uguale alla devianza tra trattamenti, la cui media generale è 158,28

$$(208,2-158,28)^2 \cdot 6 + (199,8-158,28)^2 \cdot 6 + (141-158,28)^2 \cdot 6 + (208,2-158,28)^2 \cdot 6 + (208,2-158,28)^2 \cdot 6 \\ 14.952,0384 + 10.343,4624 + 1.791,5904 + 7.341,6024 + 9.210,4344 = \mathbf{43.639,1280}$$

come la somma dei 4 gdl).

Una volta calcolate le 4 devianze, si possono effettuare 4 test F, ognuno con gdl 1,25 se e solo se il test F della varianza tra trattamenti, con gdl 4, risulta significativo.

Con i dati dell'esempio, poiché il test F della varianza tra trattamenti ( $43.639,1280 / 4 = 10.909,782$ ) risulta

$$F_{4,25} = \frac{10.909,782}{146,5} = 74,47$$

uguale a 74,47

mentre i valori tabulati di  $F_{4,25}$  alla probabilità

$\alpha = 0.05$  risulta uguale a 2,76

$\alpha = 0.01$  risulta uguale a 4,18

si possono fare i 4 test F

$$1 - F_{(1,25)} = \frac{27.498,528}{146,5} = 187,7 \quad \text{altamente significativo}$$

$$2 - F_{(1,25)} = \frac{15.876}{146,5} = 108,37 \quad \text{altamente significativo}$$

$$3 - F_{(1,25)} = \frac{211,68}{146,5} = 1,44 \quad \text{non significativo}$$

$$4 - F_{(1,25)} = \frac{111}{146,5} = < 1 \quad \text{non significativo}$$

Di norma, i risultati sono presentati in una tabella riassuntiva generale, come la seguente:

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>	<b>F</b>
<b>TOTALE</b>	47.301,628	29		
<b>Tra trattamenti o zone</b>	43.639,128	4	10.909,782	74,47
<b>A + B +C vs D + E</b>	27.498,528	1	27.498,528	187,70
<b>A + B vs C</b>	15.876,000	1	15.876,000	108,37
<b>A vs B</b>	211,680	1	211,680	1,44
<b>D vs E</b>	52,920	1	52,920	<1
<b>Errore</b>	3.662,500	25	146,500	

con le probabilità relative, riportate in una ulteriore colonna di fianco ai valori di F, quando effettuate dal computer con i programmi a maggior diffusione internazionale.

Da questi risultati è possibile trarre, in modo esplicito, le conclusioni a carattere ambientale sui valori d'inquinamento rilevati nel campionamento delle 5 zone:

- 1 tra le cinque zone, le medie aritmetiche dell'inquinamento da solventi aromatici hanno una differenza altamente significativa;
- 2 tale significatività è imputabile soprattutto alla differenza tra le 3 stazioni collocate in centro e le 2 situate in zone periferiche;
- 3 è altamente significativa anche la differenza tra le 2 stazioni collocate nella piazza principale e la stazione collocata in una piazza secondaria;
- 4 non esiste una differenza significativa tra le due stazioni di rilevazione situate a i due estremi della stessa piazza centrale;
- 5 non esiste alcuna differenza tra i valori medi delle due zone periferiche.

Quando i gruppi hanno un numero diverso di osservazioni, il confronto tra le medie non risulta più omogeneo: ogni media avrebbe un intervallo fiduciale diverso, ma i gradi di libertà di ogni F restano identici. Se le differenze nelle dimensioni campionarie dei vari gruppi non sono troppo differenti (ma resta la difficoltà di decidere quando lo siano), fino a poco tempo fa alcuni testi accettavano ugualmente il confronto con il metodo appena illustrato.

#### 8.4.2 TEST PER CONFRONTI MULTIPLI O A POSTERIORI: BONFERRONI, SNK DI NEUMAN-KEULS, LSD DI FISHER, STEP-UP DI WELSCH, HSD DI TUKEY, SCHEFFE', DUNNETT, DUNCAN

Se, nel confronto tra le medie di  $p$  gruppi, con il test  $F$  è stata rifiutata l'ipotesi nulla

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_p$$

si pone il problema di verificare tra quali esista una differenza significativa.

A questo scopo, i confronti a priori propongono i metodi migliori, poiché hanno una probabilità maggiore di evidenziare le significatività, quando l'ipotesi nulla è falsa. Ma con la diffusione dei computer, i confronti multipli o a posteriori hanno acquisito un rilevante vantaggio pratico: da vari anni sono sempre riportati in molti programmi informatici insieme con il test  $F$  e possono essere applicati con facilità. Sono quindi usati con frequenza maggiore, anche se la preferenza ad essi accordata appare illogica, ai fini di una corretta analisi statistica, che deve sempre preferire i test più potenti. Solo recentemente, anche i confronti a priori sono stati inseriti nei programmi informatici a maggior diffusione e perciò resi di uso semplice.

Quando è possibile scegliere, **la preferenza è sempre da attribuire ai test pre-pianificati**; tuttavia, in vari casi, i test a posteriori sono necessari.

**I confronti multipli o a posteriori** (detti anche **post-hoc**, **confronti non prestabiliti** o **non pianificati** e più raramente *incidental comparisons*) sono necessari quando non è possibile programmare i confronti a priori, al momento del disegno sperimentale, per carenza d'informazione. Quando i trattamenti non possono essere classificati in gruppi (come nei confronti ortogonali), che spieghino più utilmente di altri la differenza complessiva, già evidenziata dall'analisi della varianza, rimane solo la possibilità di **effettuare tutti i possibili confronti tra medie, alla ricerca di quelle differenze significative che hanno determinato la significatività totale**. E' detta "procedura di dragaggio" e serve per **individuare differenze** da studiare successivamente in modo più approfondito, con analisi ecologiche, chimiche, ambientali, **alla ricerca delle cause possibili**. Le ipotesi di lavoro non sono dedotte dalla conoscenza delle leggi della natura, ma partono dalle differenze osservate, nella convinzione che, se esistono, devono pure avere una causa. Il rischio di tale approccio è di "inventarsi" comunque una spiegazione nell'ambito della disciplina, giustificata esternamente dalla significatività statistica.

Per effettuare i confronti multipli a posteriori sono stati proposti diversi metodi, che derivano essenzialmente dal test  $t$  di Student per ipotesi bilaterali e dagli intervalli fiduciali. Questi metodi hanno lo scopo di ridurre la probabilità di commettere un **errore di I tipo** (la probabilità  $\alpha$  di trovare una differenza significativa, quando in realtà essa non esiste). Infatti, il tasso d'errore di I tipo che è

possibile commettere per un confronto tra due medie (con termine tecnico, chiamato tasso d'errore *comparison-wise*) all'aumentare del numero di confronti determina un tasso d'errore di I tipo per tutto l'esperimento (chiamato tasso d'errore *experiment-wise*) notevolmente maggiore.

La differenza tra *comparisonwise* e *experimentwise* può essere spiegata con un esempio semplice.

Si supponga che uno studente debba sostenere un esame su un testo di 1000 pagine e che egli ne abbia studiate 950:

- per **ogni domanda** (*comparisonwise*) egli ha una probabilità di 0,95 che cada su un argomento che ha studiato e quindi una probabilità di 0,05 su un argomento che non ha studiato;
- ma se l'esame prevede 10 domande, la probabilità che **tutto l'esame** (*experimentwise*) verta sulle 950 pagine che ha studiato è  $(0,95)^{10} = 0,5987$ ; quindi, la probabilità che almeno una domanda cada nelle pagine tralasciate è molto maggiore di 0,05 e pari a  $0,4013 (1 - 0,95^{10})$ .

Il problema diventa quante pagine può tralasciare, per non aumentare troppo la probabilità di essere respinto?

Anche nella scelta del test più adatto ai confronti multipli, si deve dare la preferenza a **quello più potente**; come spesso desiderato dal ricercatore, a quello che permette di **rifiutare l'ipotesi nulla in un singolo confronto**. Tuttavia, occorre anche **mantenere costante la probabilità a prescelta per tutto l'esperimento**, cioè la **protezione contro l'errore di I tipo**.

Nella scelta del test a posteriori, quindi è necessario un compromesso tra

- **comparisonwise**, collegata alla potenza del singolo test e
- **experimentwise**, collegata al principio di cautela o protezione di tutta la serie di test.

La ricerca del difficile equilibrio tra le esigenze contrastanti della potenza e della protezione, per il quale non è ancora stato trovata una soluzione universalmente condivisa, ha determinato una molteplicità di proposte, in funzione delle varie situazioni sperimentali e dei diversi rischi che un ricercatore intende correre; di conseguenza, i confronti multipli sono tra gli argomenti ancora più dibattuti ed in maggiore evoluzione della statistica univariata.

Il primo metodo di protezione, quindi per non innalzare troppo la probabilità d'errore di I tipo dell'*experimentwise*, suggerito da tutti i testi moderni, è **la prassi di**

- **fare precedere ai confronti multipli un'ANOVA e di**
- **effettuare i confronti a posteriori solo quando con essa si è rifiutata l'ipotesi nulla.**

Con questo accorgimento, si vuole evitare che anche un solo confronto tra due medie risulti significativo, quando l'analisi della varianza su tutti i dati non ha permesso di rifiutare l'ipotesi nulla.

Questo contrasto tra conclusioni è possibile, poiché i due test utilizzano probabilità a non identiche.

Tra i confronti multipli citati in letteratura, sia per l'importanza dei concetti, sia per la loro diffusione nei testi e nei programmi informatici, sono da ricordare:

- 1 – il **principio e il test t di Bonferroni**,
- 2 – la **procedura LSD (Least Significant Difference) di Fisher**,
- 3 - il **test di Student-Newman-Keuls** spesso citato come **test SNK o test Q**,
- 4 - la **procedura step-up di Welsch**, fondato sul **MSR (Minimum Significant Range)**,
- 5 - il **test di Tukey o procedura HSD (Honestly Significant Difference)**,
- 6 - il **test di Scheffé con l'estensione di Gabriel**,
- 7 - il **test di Dunnett** per confronti di vari trattamenti con un controllo,
- 8 - il **campo di variazione multiplo di Duncan**.

Informazioni più dettagliate su questi metodi possono essere rintracciate negli articoli:

- Newman D. (1939) *The distribution of the range in samples from a normal population expressed in terms of an independent estimate of the standard deviation*. Biometrika, 31, 20.
- Fisher R. A. (1948) *The design of experiments*. Edinburgh, Oliver and Boyd.
- Keuls M (1952) *The use of the Studentized range in connection with an analysis variance*. Euphytica 1, 112.
- Scheffé H (1953) *A method for judging all contrasts in the analysis of variance*. Biometrika 40, 87.
- Duncan D. B. (1955) *Multiple range and multiple F tests*. Biometrics 11,1.
- Dunnett C.W. (1964) *New tables for multiple comparisons with a control*. Biometrics 20, 482.
- Gabriel K. R. (1964) *A procedure for testing the homogeneity of all sets of means in analysis of variance*. Biometrics 20: 459-477.
- Welsch R. E. (1977) *Stepwise multiple comparison procedures*, J. Amer. Stat. Ass. 72, 566-575.

Una presentazione ampia e relativamente recente è fornita dall'ultima edizione (1995, third edition) di *Biometry* di Sokal e Rohlf (W. H. Freeman and Company, New York) pp. 240-261.

1) L'italiano **Bonferroni** è stato il primo ad affrontare l'argomento in modo operativo. Affermò che, **per effettuare p volte il test t di Student mantenendo costante la probabilità totale  $\alpha_T$  (experimentwise), la probabilità  $\alpha$  di ogni confronto (comparisonwise) deve essere minore di  $\alpha_T/p$ .**

Tale relazione, detta **diseguaglianza di Bonferroni**, può essere scritta come

$$\alpha < \alpha_T / p$$

Per esempio, quando con 3 confronti la probabilità totale  $\alpha_T$  di commettere un errore di I tipo non deve essere superiore a 0.05, la probabilità  $\alpha$  di ogni singolo confronto deve essere minore di **0.0166** (0.05/3); se i confronti fossero 4, la probabilità  $\alpha$  di ogni confronto non deve superare **0.0125** (0.05/4).

In realtà, come sarà più avanti approfondito, per una stima più accurata della probabilità **comparisonwise** sulla base di quella dell'**experimentwise**, è utile ricordare che

- la probabilità d'errore complessivo ( $\alpha_T$ ) è legata
- alla probabilità di errore di ogni confronto ( $\alpha$ ) e
- al numero di confronti da effettuare ( $p$ )

secondo la relazione esponenziale

$$\alpha_T = 1 - (1 - \alpha)^p$$

A causa di questa approssimazione nella stima della probabilità, la procedura proposta da Bonferroni è ritenuta sostanzialmente accettabile quando si effettuano pochi confronti, perché le differenze tra le due stime sono minime. **Quando i confronti superano il numero di 8-10, il valore di  $\alpha$  stimato per ognuno di essi diventa troppo piccolo; di conseguenza, il metodo è ritenuto eccessivamente cautelativo.**

Con il metodo di Bonferroni, per il confronto tra due medie si utilizza una **formula analoga al t di Student per 2 campioni indipendenti**,

$$t_{(\text{Bonferroni})}(\alpha_T, p, n) = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{s_e^2 \cdot \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

dove

$\alpha_T$  = la probabilità prefissata globale per tutti i confronti (0.05 o 0.01),

$p$  = il numero di confronti che si intendono eseguire

$n$  = sono i gdl della varianza d'errore  $s_e^2$  utilizzata.

Se il confronto è più generale, volendo valutare se la differenza tra due medie è maggiore di una quantità prefissata ( $m_A - m_B$ ), in modo analogo al test  $t$  di Student per 2 campioni indipendenti, la formula precedente è scritta come

$$t_{(\text{Bonferroni})}(\alpha_T, p, n) = \frac{(\bar{X}_A - \bar{X}_B) - (m_A - m_B)}{\sqrt{s_e^2 \cdot \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

Rispetto al test  $t$  di Student per due campioni indipendenti, questi metodi offrono i vantaggi di

- utilizzare la varianza d'errore  $s_e^2$  calcolata con l'ANOVA tra tutti i gruppi, al posto della varianza associata  $s_p^2$  dei due soli gruppi a confronto;
- usare i gradi di libertà della varianza d'errore  $s_e^2$  (n) per la scelta del valore di **t**, al posto di quelli derivati solo dal numero dei dati presenti nei due gruppi a confronto ( $n_A - 1 + n_B - 1$ ).

Nel **caso di 2 campioni con lo stesso numero n d'osservazioni** o repliche (detti **campioni bilanciati**), il valore del **t**<sub>(Bonferroni)</sub> è più rapidamente calcolato con la formula equivalente

$$t_{(\text{Bonferroni})}(\alpha_T, p, n) = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{2s_e^2}{n}}}$$

Si evidenzia una differenza significativa tra ogni coppia di medie alla probabilità totale  $\alpha_T$  prefissata, quando il valore calcolato supera il valore critico riportato nella tabella successiva.

E' possibile ricavare il valore critico del **t** anche da una tabella dettagliata dei valori **F** di Fisher, per la nota relazione

$$t_{(\alpha;n)} = \sqrt{F_{(\alpha;1,n)}}$$

Tuttavia per **t** ed **F** esiste un problema pratico: non sempre questi valori sono disponibili alla probabilità  $\alpha$  richiesta.

Per esempio, con 3 confronti alla probabilità complessiva  $\alpha_T = 0.05$  occorrerebbe disporre di una tabella che fornisce il valore di **t** o di **F** alla probabilità  $\alpha = 0.0167$ .

E' semplice ottenere i valori del **t** di Bonferroni solamente quando il numero di confronti è 5 oppure 10 o 20, poiché con  $\alpha_T$  uguale a 0.05 (experimentwise) la probabilità  $\alpha$  di ogni confronto (comparisonwise) diventa rispettivamente 0.01 oppure 0.005 o 0.001, tutti valori riportati con frequenza sulle tavole sinottiche.

Di conseguenza, sono fornite tabelle specifiche che, come indici, utilizzano

- il numero **p** di confronti (che con **k** medie sono **k(k-1) / 2**),
- i gradi di libertà della varianza d'errore (n).

### Valori critici del test t di Bonferroni

**p** = numero di confronti simultanei

**gdl** o **n** = gradi di libertà della varianza d'errore

$$a_T = 0.05$$

NUMERO **p** DI CONFRONTI SIMULTANEI

n	2	3	4	5	6	7	8	9	10
5	3,17	3,54	3,81	4,04	4,22	4,38	4,53	4,66	4,78
7	2,84	3,13	3,34	3,50	3,64	3,76	3,86	3,95	4,03
10	2,64	2,87	3,04	3,17	3,28	3,37	3,45	3,52	3,58
12	2,56	2,78	2,94	3,06	3,15	3,24	3,31	3,37	3,43
15	2,49	2,69	2,84	2,95	3,04	3,11	3,18	3,24	3,29
20	2,42	2,61	2,75	2,85	2,93	3,00	3,06	3,11	3,16
24	2,39	2,58	2,70	2,80	2,88	2,94	3,00	3,05	3,09
30	2,36	2,54	2,66	2,75	2,83	2,89	2,94	2,99	3,03
40	2,33	2,50	2,62	2,71	2,78	2,84	2,89	2,93	2,97
60	2,30	2,47	2,58	2,66	2,73	2,79	2,84	2,88	2,92
120	2,27	2,43	2,54	2,62	2,68	2,74	2,79	2,83	2,86
∞	2,24	2,39	2,50	2,58	2,64	2,69	2,74	2,77	2,81

$$a_T = 0.01$$

NUMERO **p** DI CONFRONTI SIMULTANEI

n	2	3	4	5	6	7	8	9	10
5	4,78	5,25	5,60	5,89	6,15	6,36	6,56	6,70	6,86
7	4,03	4,36	4,59	4,78	4,95	5,09	5,21	5,31	5,40
10	3,58	3,83	4,01	4,15	4,27	4,37	4,45	4,53	4,59
12	3,43	3,65	3,80	3,93	4,04	4,13	4,20	4,26	4,32
15	3,29	3,48	3,62	3,74	3,82	3,90	3,97	4,02	4,07
20	3,16	3,33	3,46	3,55	3,63	3,70	3,76	3,80	3,85
24	3,09	3,26	3,38	3,47	3,54	3,61	3,66	3,70	3,74
30	3,03	3,19	3,30	3,39	3,46	3,52	3,57	3,61	3,65
40	2,97	3,12	3,23	3,31	3,38	3,43	3,48	3,51	3,55
60	2,92	3,06	3,16	3,24	3,30	3,34	3,39	3,42	3,46
120	2,86	2,99	3,09	3,16	3,22	3,27	3,31	3,34	3,37
∞	2,81	2,93	3,02	3,09	3,16	3,19	3,24	3,26	3,29

**Quando i campioni non sono delle stesse dimensioni n**, più recentemente vari autori hanno aggiunto una **ulteriore cautela**.

Nella formula generale precedente, al posto della correzione per il bilanciamento

$$\frac{1}{n_1} + \frac{1}{n_2}$$

propongono di **utilizzare la formula per due campioni bilanciati**

$$t_{(\text{Bonferroni})} (a, p, n) = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{2s_e^2}{n}}}$$

con **n uguale al valore minore tra n<sub>1</sub> e n<sub>2</sub>**

E' ovvio che non si supera la probabilità experimentwise prefissata, ma si perde in potenza nella comparisonwise.

ESEMPIO. Con le stesse misure d'inquinamento (utilizzate nell'esempio precedente per i confronti a priori) rilevate in 5 zone, delle quali sono stati riportati le medie e il numero di osservazioni:

<b>ZONE</b>	A	B	C	D	E
<b>Medie</b>	208,2	199,8	141,0	123,3	119,1
<b>n<sub>1</sub></b>	6	5	6	6	7

verificare con il test **t** di Bonferroni tra quali medie esiste una differenza significativa.

Risposta.

Vari autori affermano che i campioni dovrebbero essere bilanciati. Altri sostengono che, con leggere differenze nel numero di osservazioni, è lecito l'uso di questo test, ricorrendo alla formula generale che considera il diverso numero di osservazioni per gruppo.

Con 5 medie, si hanno 10 differenze, che possono essere presentate in modo chiaro ed utile in una tabella con tutti i confronti:

Confronti	Medie	Differenze
A vs B	208,2 - 199,8	8,4
A vs C	208,2 - 141,4	66,8
A vs D	208,2 - 123,3	84,9
A vs E	208,8 - 119,1	89,7
B vs C	199,8 - 141,4	58,4
B vs D	199,8 - 123,3	76,5
B vs E	199,8 - 119,1	80,7
C vs D	141,4 - 123,3	18,1
C vs E	141,4 - 119,1	22,3
D vs E	123,3 - 119,1	4,2

**Le differenze sono da considerare in valore assoluto, in quanto i confronti multipli comportano solo test bilaterali.**

(RICORDARE: nell'analisi della varianza con i **5** gruppi è stata rifiutata l'ipotesi nulla e che la varianza d'errore  $s_e^2$  è risultata uguale a **146,5** con **25** gdl.

Per ogni confronto si calcola il valore del  $t_{(\text{Bonferroni})}$  e si confronta il risultato con i valori critici riportati nella tabella.

Per esempio, **A vs B** diventa

$$t_{(\text{Bonferroni})} = \frac{208,2 - 199,8}{\sqrt{146,5 \cdot \left(\frac{1}{6} + \frac{1}{5}\right)}} = \frac{8,4}{\sqrt{146,5 \cdot 0,367}} = \frac{8,4}{\sqrt{53,766}} = \frac{8,4}{7,33} = 1,14$$

Esso fornisce un valore di **t = 1,14** per un numero di confronti **p = 10** e **gdl = 25**.

Quando il numero esatto di gdl non è riportato nella tabella dei valori critici, per trovare il valore critico nella tabella si utilizzano **i gdl immediatamente inferiori** (24 nella tabella), **in quanto rappresenta la scelta più cautelativa**.

Per **p = 10** e alla probabilità complessiva **a = 0.05** il valore critico riportato è **3,09**.

Il valore calcolato (1,14) è inferiore: non si può rifiutare l'ipotesi nulla e quindi la media di A e quella di B non differiscono in modo significativo.

Il confronto **A vs D** diventa

$$t_{(\text{Bonferroni})} = \frac{208,2 - 123,3}{\sqrt{146,5 \cdot \left(\frac{1}{6} + \frac{1}{6}\right)}} = \frac{84,9}{\sqrt{146,5 \cdot 0,334}} = \frac{84,9}{\sqrt{48,931}} = \frac{84,9}{6,995} = 12,13$$

e stima un valore del **t = 12,13** da confrontare sempre con il valore critico di 3,09. Infatti sono invariati sia la probabilità  $\alpha$  totale, sia il numero **p** di confronti, sia i **gdl** della varianza d'errore.

Alla probabilità **a = 0.01** il valore critico, sempre per **p = 10** e **gdl = 24**, risulta uguale a 3,74.

Di conseguenza, la differenza di questo confronto (208,2 e 123,3) risulta significativa: le due medie (A vs D) differiscono tra loro con probabilità  $\alpha < 0.01$ .

Poiché le ultime due medie a confronto (A vs D) hanno lo stesso numero d'osservazioni (6), si può ottenere lo stesso risultato più rapidamente con

$$t_{(\text{Bonferroni})} = \frac{208,2 - 123,3}{\sqrt{\frac{2 \cdot 146,5}{6}}} = \frac{84,9}{48,83} = \frac{84,9}{6,998} = 12,13$$

2) In vari testi di lingua inglese, **l'idea di condurre tutti i possibili confronti tra coppie di medie è attribuita a Fisher**, per la sua presentazione nel volume *The design of experiments*, edito da Oliver and Boyd di Edinburgh nel 1948. Chiamata in inglese **Least Significant Difference** e tradotta in italiano come **Differenza Minima Significativa** (meno spesso, in modo letterale, come **Differenza Meno Significativa**), è abbreviata in **LSD**.

Effettuando un test **t di Student** per ogni coppia di medie ( $\bar{X}_A$  e  $\bar{X}_B$ ) con lo **stesso numero di repliche**, sono significative tutte le differenze **D** (con  $D = \bar{X}_A - \bar{X}_B$ ) che in valore assoluto superano la quantità **LSD**

$$\mathbf{LSD} = t_{(\alpha/2, n)} \times \sqrt{\frac{2s_e^2}{n}}$$

dove

$t_{\alpha/2}$  = percentile con probabilità  $\alpha/2$  della distribuzione **t**

$n$  = gdl della varianza d'errore stimata con l'ANOVA.

Come già discusso per il **t** di Bonferroni, per confronti tra tutte le coppie diventano ancor più importanti i **rapporti tra il valore di probabilità del comparison e quello dell'experiment-wise**. Con questo metodo è consigliata l'applicazione del **principio di cautela** già ricordato, cioè di **effettuare i confronti a coppie solo quando l'analisi della varianza su tutti i gruppi è risultata significativa**.

Con la generazione di numeri casuali, S. G. Carmer e M. R. Swanson nel 1973 (con l'articolo *An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods*, pubblicato su JASA, n. 68, pp. 66-74) hanno dimostrato che questa precauzione (procedere ai confronti multipli solo dopo aver rifiutato l'ipotesi nulla con l'ANOVA) è **una cautela efficiente contro gli errori di I tipo**: permette di non superare la probabilità experimentwise prefissata.

ESEMPIO. Con gli stessi dati dell'esempio precedente

<b>ZONE</b>	A	B	C	D	E
<b>Medie</b>	208,2	199,8	141,0	123,3	119,1
<b>n<sub>i</sub></b>	6	5	6	6	7

in cui la varianza d'errore  $s_e^2$  è risultata uguale a **146,5** con **25** gdl,

- assumendo  $n = 5$  come dimensione comune a tutti i gruppi (**scelta cautelativa**) e che

- alla probabilità  $\alpha = 0.05$  il valore di  $t$  per 25 gdl è uguale a **2,06**

la Differenza Minima Significativa (**LSD**)

$$\mathbf{LSD} = 2,06 \cdot \sqrt{\frac{146,5}{5}} = \mathbf{11,15}$$

risulta uguale a 11,15.

E' significativa qualsiasi differenza tra le 5 medie che superi questa quantità.

Per evitare una scelta così prudentiale come quella di utilizzare il valore minore, quando i campioni non hanno lo stesso numero di osservazioni, alcuni testi suggeriscono il metodo della **interpolazione armonica**. Quando i  $p$  gruppi sono di dimensioni ( $n_i$ ) non troppo differenti, è possibile stimare un valore  $\hat{n}$  corretto, dato dal rapporto

$$\hat{n} = \frac{p}{\frac{1}{n_1} + \frac{1}{n_2} + \dots + \frac{1}{n_p}}$$

Con i dati dell'esempio,

$$\hat{n} = \frac{5}{\frac{1}{6} + \frac{1}{5} + \frac{1}{6} + \frac{1}{6} + \frac{1}{7}} = 5,93$$

$\hat{n}$  risulta uguale a 5,93.

Si osservi che il valore ottenuto è inferiore alla media aritmetica di 6.

**Con una distribuzione non bilanciata, la quantità d'informazione media è inferiore a quella di una distribuzione bilanciata, che ovviamente abbia lo stesso numero totale di osservazioni.**

Con 5,93 al posto del 5 precedente (ritenuti da molti troppo cautelativo), si ottiene un errore standard minore e quindi un valore di **LSD** minore. Il test diventa più potente, con maggiori probabilità di evidenziare differenze significative.

Con gli stessi dati dell'esempio precedente,

si ottiene

$$\mathbf{LSD} = 2,06 \cdot \sqrt{\frac{146,5}{5,93}} = \mathbf{10,24}$$

una Differenza Minima Significativa pari a 10,24.

Per fare questi confronti tra coppie di medie, è stato proposto di agire in modo organizzato; quindi

- **disporre le varie medie in ordine di rango** (non importa se dalla minore alla maggiore o viceversa)

- **iniziare il confronto tra le due più estreme**, che quasi sempre risulterà significativo se l'ANOVA è risultata significativa.

Il primo confronto deve avvenire tra le due medie poste agli estremi; il secondo confronto tra una delle due medie agli estremi e quella che fornisce la differenza maggiore tra tutte le altre, dopo la precedente. Se anche questa risulta significativa si procede al terzo confronto, nel quale si riduce ancora la differenza tra le due medie interessate.

Da questa metodologia (chiamata *step-up procedure* o *stepwise approach*), sempre per evitare errori di I tipo, è stata derivata una **terza cautela** (dopo l'analisi della varianza prima dei confronti multipli e l'uso del numero di dati del campione minore per stimare la differenza minima significativa). Essa consiste nel

- **non considerare come significativamente differenti due medie, quando sono comprese entro due già ritenute non significative.**

Con un valore di **LSD** pari a 10,24 (vedi ultimo calcolo) si deduce che la media del gruppo

- A (208,2) risulta non significativamente diversa da B (199,8)
- D (123,3) non significativamente diversa da E (119,1).

I risultati di questi confronti a coppie sono spesso **rappresentati in modo grafico**, con la **convenzione** di congiungere quelle medie che non sono significativamente diverse tra loro.

Con i dati dell'esempio, si ottiene

A    B    C    D    E

Un'**altra convenzione grafica**, diffusa e di facile comprensione, consiste nel riportare le medie

208,2	199,8	141,0	123,3	119,1
A	A	B	C	C

e nell'indicare con la stessa lettera quelle che tra loro non sono significativamente differenti.

Spesso viene usata una **terza convenzione grafica**, data dalla combinazione di queste due tecniche, quale

208,2	199,8	141,0	123,3	119,1
A A A A	B	C C C C		

In situazione più complesse quando si hanno molte medie che differiscono tra loro per quantità ridotte, si determinano linee o serie di lettere a livelli diversi,

quali

A    B    C    D    E

-----  
-----

Alla informazione precedente (la media di A non diversa da quella di B; la media di D non diversa da quella di E), in questo caso è aggiunta l'informazione di una differenza non significativa tra le medie C-D e tra le medie B-C.

3) Il **test di Student-Neuman-Keuls** o **test SNK**, citato anche come **test di Neuman-Keuls** o **test Q** (oppure **q studentizzato**, detto pure delle **inferenze simultanee** di Newman (1939) e Keuls (1952), utilizza una stima più precisa della probabilità  $\alpha$  per ogni singolo confronto. Bonferroni ha avuto il merito di affrontare il problema dei confronti multipli in modo corretto e generale; ma i valori critici calcolati si sono dimostrati troppo cautelativi, soprattutto all'aumentare del numero di confronti. Una stima più corretta è quella che alcuni autori attribuiscono a **Dunn-Sidak** (vedi: Ury H. K. 1976, *A comparisons of four procedures for multipkle comparisons among means (pairwise contrasts) for arbitrary sample size*. Technometrics 18: 89-97). Con  $p$  confronti, la probabilità  $\alpha$  da assegnare ad ognuno di essi è

$$\alpha = 1 - (1 - \alpha_T)^{1/p}$$

Per esempio, con  $\alpha_T$  uguale a 0.05 e

-  $p$  uguale a 5,

la probabilità  $\alpha$  di ogni confronto non è uguale a 0.01 (0.05/5) ma

$$\alpha = 1 - 0.95^{1/5} = 1 - 0.98979 = 0.01021$$

è uguale a **0.01021**,

con una differenza relativa, rispetto alla stima del Bonferroni del 2,1 per cento;

- con  $p$  uguale a 10, non è più uguale a 0.005 (0.05/10) ma

$$\alpha = 1 - 0.95^{1/10} = 1 - 0.99488 = 0.00512$$

è uguale a **0.00512**, con un aumento relativo del 2,4 per cento rispetto alla stima prudentiale o cautelativa del Bonferroni.

Il confronto tra coppie di medie (A e B) avviene con una formula simile a quella del test  $t$  di Bonferroni

$$Q_{(\alpha_T, p, n)} = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_e^2}{2} \cdot \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

dove, con la consueta simbologia,

$\alpha_T$  = probabilità totale o complessiva di commettere l'errore di I tipo (di norma, 0.05 o 0.01),

$p$  = numero di confronti che si intendono effettuare,

$n$  = gradi di libertà della varianza d'errore del test F tra tutti i gruppi,

$\bar{x}_A - \bar{x}_B =$  differenza tra le due medie a confronto,

$s_e^2 =$  varianza d'errore ottenuta con il test **F** tra tutti i gruppi ,

$n_A$  e  $n_B =$  numero di osservazioni nei due gruppi A e B a confronto.

(Il **2**, riportato al denominatore, **non rappresenta una differenza reale dal test precedente**: è solamente un valore inglobato in quello di **Q**, per una scelta dei due autori).

Con la **procedura stepwise** prima descritta, nella metodologia **SNK** il valore di **Q studentizzato** dipende da quello di **p**, il numero di medie a confronto. Se, come nell'esempio precedente, le medie sono 5, per il confronto della media maggiore con quella minore il valore di **p** è uguale a 5. Il secondo passo prevede il confronto tra la media maggiore e la seconda minore, mentre il valore di **p** diviene 4. Quando si confrontano la maggiore con quella più vicina, la seconda in ordine decrescente, il valore di **p** è uguale a 2. Questa stima è fatta contando i "passi": due medie vicine distano 2 passi; due medie con un'altra intermedia distano 3 passi; in un gruppo di 5 medie, la minore e la maggiore distano 5 passi.

Con 5 medie ordinate in modo crescente (vedi tabella sottostante, in cui 5<sup>a</sup> indica la media maggiore e 1<sup>a</sup> la minore), i confronti da effettuare sono 10, con l'ordine ed il relativo valore dell'indice **p** riportato nella tabella:

Ordine	Confronto	P
1	5 <sup>a</sup> vs 1 <sup>a</sup>	5
2	5 <sup>a</sup> vs 2 <sup>a</sup>	4
3	5 <sup>a</sup> vs 3 <sup>a</sup>	3
4	5 <sup>a</sup> vs 4 <sup>a</sup>	2
5	4 <sup>a</sup> vs 1 <sup>a</sup>	4
6	4 <sup>a</sup> vs 2 <sup>a</sup>	3
7	4 <sup>a</sup> vs 3 <sup>a</sup>	2
8	3 <sup>a</sup> vs 1 <sup>a</sup>	3
9	3 <sup>a</sup> vs 2 <sup>a</sup>	2
10	2 <sup>a</sup> vs 1 <sup>a</sup>	2

Quando le due medie a confronto hanno lo stesso numero **n** di osservazioni, la formula può essere semplificata in

$$Q_{(\alpha_T, p, v)} = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_e^2}{n}}}$$

Sempre come nel test t di Bonferroni, quando le due medie hanno un numero differente **n** di osservazioni, è possibile utilizzare la frequenza minore.

Ricorrendo ai confronti multipli come proposto da Fisher, quando i gruppi a confronto hanno lo stesso numero di dati o frequenze simili, vari autori hanno generalizzato tale procedura, proponendo la tecnica delle **inferenze simultanee**: si calcola una **MSD** (Minimum Significant Difference) data da

$$MSD = Q_{(\alpha_T, p, n)} \cdot \sqrt{\frac{s_e^2}{n}}$$

Sono significative tutte le differenze superiori a MSD

### Valori critici del Q per il test SNK e per il W di Tukey

**p** = numero di confronti simultanei per il test SNK

**p** = numero di medie a confronto per il T di Tukey

**v** = gradi di libertà della varianza d'errore

$$\alpha_T = 0.05$$

#### INDICE **p** DEL CONFRONTO

<b>n</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
<b>10</b>	3,151	3,877	4,327	4,654	4,912	5,124	5,305	5,461	5,599
<b>11</b>	3,113	3,820	4,256	4,574	4,823	5,028	5,202	5,353	5,487
<b>12</b>	3,082	3,773	4,199	4,508	4,751	4,950	5,119	5,265	5,395
<b>13</b>	3,055	3,735	4,151	4,453	4,690	4,885	5,049	5,192	5,318
<b>14</b>	3,033	3,702	4,111	4,407	4,639	4,829	4,990	5,131	5,254
<b>15</b>	3,014	3,674	4,076	4,367	4,595	4,782	4,940	5,077	5,198
<b>16</b>	2,998	3,649	4,046	4,333	4,557	4,741	4,897	5,031	5,150
<b>17</b>	2,984	3,628	4,020	4,303	4,524	4,705	4,858	4,991	5,108
<b>18</b>	2,971	3,609	3,997	4,277	4,495	4,673	4,824	4,956	5,071
<b>19</b>	2,960	3,593	3,977	4,253	4,469	4,645	4,794	4,924	5,038
<b>20</b>	2,950	3,578	3,958	4,232	4,445	4,620	4,768	4,896	5,008
<b>24</b>	2,919	3,532	3,901	4,166	4,373	4,541	4,684	4,807	4,915
<b>30</b>	2,888	3,486	3,845	4,102	4,302	4,464	4,602	4,720	4,824
<b>40</b>	2,858	3,442	3,791	4,039	4,232	4,389	4,521	4,635	4,735
<b>60</b>	2,829	3,399	3,737	3,977	4,163	4,314	4,441	4,550	4,646
<b>120</b>	2,800	3,356	3,685	3,917	4,096	4,241	4,363	4,468	4,560
<b>∞</b>	2,772	3,314	3,633	3,858	4,030	4,170	4,286	4,387	4,474

$$a_T = 0.01$$

INDICE  $p$  DEL CONFRONTO

n	2	3	4	5	6	7	8	9	10
10	4,482	5,270	5,769	6,136	6,428	6,669	6,875	7,055	7,213
11	4,392	5,146	5,621	5,970	6,247	6,476	6,672	6,842	6,992
12	4,320	5,046	5,502	5,836	6,101	6,321	6,507	6,670	6,814
13	4,260	4,964	5,404	5,727	5,981	6,192	6,372	6,528	6,667
14	4,210	4,895	5,322	5,634	5,881	6,085	6,258	6,409	6,543
15	4,168	4,836	5,252	5,556	5,796	5,994	6,162	6,309	6,439
16	4,131	4,786	5,192	5,489	5,722	5,915	6,079	6,222	6,349
17	4,099	4,742	5,140	5,430	5,659	5,847	6,007	6,147	6,270
18	4,071	4,703	5,094	5,379	5,603	5,788	5,944	6,081	6,201
19	4,046	4,670	5,054	5,334	5,554	5,735	5,889	6,022	6,141
20	4,024	4,639	5,018	5,294	5,510	5,688	5,839	5,970	6,087
24	3,956	4,546	4,907	5,168	5,374	5,542	5,685	5,809	5,919
30	3,889	4,455	4,799	5,048	5,242	5,401	5,536	5,653	5,756
40	3,825	4,367	4,696	4,931	5,114	5,265	5,392	5,502	5,559
60	3,762	4,282	4,595	4,818	4,991	5,133	5,253	5,356	5,447
120	3,702	4,200	4,497	4,709	4,872	5,005	5,118	5,214	5,299
∞	3,643	4,120	4,403	4,603	4,757	4,882	4,987	5,078	5,157

ESEMPIO. Con gli stessi dati utilizzati nell'esempio per il test  $t$  di Bonferroni

Zone	A	B	C	D	E
Medie	208,2	199,8	141,0	123,3	119,1
$n_i$	6	5	6	6	7

verificare con il test  $Q$  di Student-Neuman-Keuls tra quali medie esiste una differenza significativa.

Risposta.

Dapprima ordinare le medie in modo decrescente:

Ordine	Confronto	P	Medie	Differenza
1	5 <sup>a</sup> vs 1 <sup>a</sup>	5	208,2 – 119,1	89,1
2	5 <sup>a</sup> vs 2 <sup>a</sup>	4	208,2 – 123,3	84,9
3	5 <sup>a</sup> vs 3 <sup>a</sup>	3	208,2 – 141,0	67,2
4	5 <sup>a</sup> vs 4 <sup>a</sup>	2	208,2 – 199,8	8,4
5	4 <sup>a</sup> vs 1 <sup>a</sup>	4	199,8 – 119,1	80,7
6	4 <sup>a</sup> vs 2 <sup>a</sup>	3	199,8 – 123,3	76,5
7	4 <sup>a</sup> vs 3 <sup>a</sup>	2	199,8 – 141,0	58,8
8	3 <sup>a</sup> vs 1 <sup>a</sup>	3	141,0 – 119,1	21,9
9	3 <sup>a</sup> vs 2 <sup>a</sup>	2	141,0 – 123,3	17,7
10	2 <sup>a</sup> vs 1 <sup>a</sup>	2	123,3 – 119,1	4,2

Successivamente calcolare le differenze, secondo l'ordine richiesto dalla procedura.

Limitando i calcoli a due soli casi (sufficienti ad illustrare il metodo), si può saggiare la significatività del confronto 3 (la media 5<sup>a</sup> vs la media 3<sup>a</sup> con indice **p** uguale a 3),

che determina un valore di Q

$$Q_{(3,25)} = \frac{208,2 - 141,0}{\sqrt{\frac{146,5}{6}}} = \frac{67,2}{\sqrt{24,416}} = \frac{67,2}{4,94} = 13,60$$

uguale a 13,60.

Il valore critico alla probabilità  $\alpha_T = 0.05$  con  $p = 3$  con  $gdl = 24$  (il valore minore più vicino a quello dei dati) è 3,532 mentre alla probabilità complessiva  $\alpha_T = 0.01$  è uguale a 4,546.

Il valore calcolato è nettamente superiore ad entrambi: nel confronto tra le medie 208,2 e 141,0 la differenza risulta significativa, con probabilità  $\alpha < 0.01$ .

Nel confronto 8, la 3<sup>a</sup> media vs la 1<sup>a</sup> media, il valore di Q

$$Q_{(3,25)} = \frac{141,0 - 119,1}{\sqrt{\frac{146,5}{2} \left( \frac{1}{6} + \frac{1}{7} \right)}} = \frac{21,9}{\sqrt{73,25 \cdot 0,310}} = \frac{21,9}{\sqrt{22,7075}} = \frac{21,9}{4,77} = 4,59$$

è uguale a 4,59 e risulta ugualmente significativo.

4 ) In alcuni testi, al metodo SNK viene preferita la **procedura step-up di Welsch** R. E. che differisce essenzialmente per l'uso di una tabella differente di valori critici, derivati da una logica diversa di procedere ai confronti. **Rovesciando la procedura stepwise precedente** che partiva dagli estremi, dopo aver ordinato le medie secondo la grandezza,

- dapprima si verifica la significatività della differenze tra **medie adiacenti**;
- solamente **quando una differenza non risulta significativa**, si passa al confronto tra le due medie appena più distanti, che comprendono la precedente.

E' ovvio che rispetto all'approccio che parte dalle medie più distanti, questo metodo determina test più potenti e meno cautelativi, controllando le probabilità  $\alpha$  di errore di I tipo per la **comparisonwise**, ma aumentandola per l'**experimentwise**.

Focalizzando l'attenzione sulla comparisonwise,

- la probabilità  $\alpha_i$  di un errore di I tipo nel confronto tra due medie adiacenti di rango  $i$ , è

$$\alpha_i = \alpha(i/p)$$

dove

$\alpha$  = probabilità dell'experimentwise (0.05 o 0.01)

$p$  = numero di medie a confronto.

Con il metodo di **Welsch** è possibile ricorrere ai confronti multipli simultanei, prima accennati, calcolando il valore **MSR** (Minimum Significant Range)

$$\text{MSR} = Q_{(\alpha, p, n)} \cdot \sqrt{\frac{s_e^2}{n}}$$

dove

- $Q_{(\alpha, p, n)}$  è fornito da tabelle specifiche
- $p$  = numero di medie
- $n$  = gdl della varianza d'errore.

Nel confronto tra coppie di medie, sono significative tutte le differenze superiori al valore di MSR

Nel solito esempio, con la varianza d'errore  $s_e^2 = 146,5$  e  $n = 25$  gdl, con **5** gruppi che hanno tutti  $n = 6$  osservazioni,

$$\text{MSR} = Q_{(\alpha, 5, 25)} \cdot \sqrt{\frac{146,5}{6}}$$

Il **Q di Welsch è fornito da tabelle specifiche** (non riportate nel testo).

Anche in questo caso, vari autori approvano l'uso di **campioni con un numero diverso di dati**. Di conseguenza, si deve ricorrere alla **interpolazione armonica** per stimare un valore di  $\hat{n}$  comune, accettabile quando le differenze sono ridotte (con i dati dell'esempio, 5,93 al posto di 6).

5) Il **test di Tukey** (chiamato anche test di **Tukey-Kramer**, poiché a Kramer è dovuto il cambiamento della procedura originaria di Tukey) è uno dei più diffusi nei programmi informatici, per determinare quali dei confronti tra le **p** medie diano differenze significative. A volte il test è indicato con **W** (altre con **T**) e la procedura dei confronti multipli simultanei **LSD** (Least Significant Difference) qui è chiamata **HSD** (Honestly Significant Difference); più raramente, forse per rendere omogenea la terminologia, anche in questo caso altri testi usano l'acronimo **MSD** (Minimum Significant Difference).

Fondato sull'**intervallo (o campo) di variazione del Q studentizzato**, il **metodo delle inferenze simultanee** è innovativo in quanto permette di esaminare contemporaneamente **i confronti semplici tra tutte le coppie di trattamenti**, escludendo quindi i contrasti (sinonimo di confronti) complessi.

Il **campo di variazione studentizzato (Q)**, quando stimato per campioni di dimensioni **n**, è determinato da

$$Q = \frac{\bar{X} \max - \bar{X} \min}{\sqrt{\frac{s_e^2}{n}}}$$

Per confronti tra coppie di medie, il livello di significatività è costruito sul caso peggiore (appunto sulla differenza massima, data da  $\bar{X} \max - \bar{X} \min$ ); di conseguenza, molti autori di testi di statistica ritengono che fornisca **una probabilità experimentwise appropriata per il complesso dei confronti**.

*(Per le differenze tra coppie di medie semplici, attualmente è ritenuto il test **più onesto**, usando la stessa terminologia dell'autore).*

Con **p** medie, si hanno **p(p-1)/2** differenze, delle quali risultano statisticamente significative quelle che in valore assoluto sono maggiori dell'**intervallo di confidenza o campo di variazione critico W**

$$W = Q_{(a,p,n)} \cdot \sqrt{\frac{s_e^2}{n}}$$

dove:

- **a** è la probabilità complessiva prescelta,
- **p** il numero di gruppi,
- **n** sono i gradi di libertà della varianza d'errore  $s_e^2$ ,
- **n** è il numero d'osservazioni di ogni gruppo (in campioni bilanciati),
- **Q** è il valore fornito dalla tabella alla probabilità complessiva  $a_T$  per **p** gruppi e gdl **n** (della varianza d'errore).

Il metodo è corretto solo per esperimenti bilanciati; se i trattamenti hanno un numero di repliche diverso, per ogni confronto tra due generici gruppi A e B con  $n$  differenti si può stimare  $W$  mediante

$$W = Q_{(a,p,n)} \cdot \sqrt{\frac{s_e^2}{2} \cdot \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}$$

Il risultato è approssimato, se esteso a tutti i confronti.

Per calcolare un solo valore nonostante l'uso di gruppi con un numero diverso di osservazioni, resta solamente la scelta più cautelativa, quella del numero  $n$  di osservazioni minore tra tutti i gruppi a confronto.

Questa **procedura di Tukey-Kramer**, nota anche come **metodo dell'intervallo di variazione**, oltre al confronto tra due medie specifiche permette di applicare la tecnica dei confronti simultanei, calcolando **la differenza minima significativa** (in inglese, MSD = Minimum Significant Difference) tra tutte le coppie di medie.

Il metodo richiede di **calcolare tutte le differenze tra le  $p$  medie: sono significative alla probabilità complessiva prefissata a quelle che superano il valore HSD.**

Per la presentazione dei risultati, sovente si ricorre ad una tabella che evidenzia contemporaneamente tutte le differenze. Con 4 medie si ottengono 6 differenze; tra esse quelle superiori al valore **HSD** calcolato sono significative e possono essere evidenziate in modo grafico (con gli asterischi), secondo il livello di significatività.

		MEDIE			
		(1)	(2)	(3)	(4)
		$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$
MEDIE		DIFFERENZE			
(2)	$\bar{X}_2$	$\bar{X}_1 - \bar{X}_2$	---	---	---
(3)	$\bar{X}_3$	$\bar{X}_1 - \bar{X}_3$	$\bar{X}_2 - \bar{X}_3$	---	---
(4)	$\bar{X}_4$	$\bar{X}_1 - \bar{X}_4$	$\bar{X}_2 - \bar{X}_4$	$\bar{X}_3 - \bar{X}_4$	---

ESEMPIO. Stimare le differenze significative tra le 5 medie utilizzate negli esempi precedenti, alle probabilità  $\alpha = 0.05$  e  $\alpha = 0.01$ .

Risposta.

Dalla tabella di distribuzione dei valori critici del **Q studentizzato**, scegliere il valore di **Q** per:

- la probabilità  $\alpha = 0.05$  e  $\alpha = 0.01$ ;
- il numero di trattamenti o medie **p**, che con i dati dell'esempio è uguale a **5**;
- i gradi di libertà della varianza d'errore (**n**), che nell'esempio sono uguali a **25**; nella tabella dei valori critici scegliere il numero inferiore più vicino (24), in quanto più cautelativo di quello superiore (30); un altro metodo, più preciso ma leggermente più difficile, suggerisce di stimare il valore di **Q** mediante interpolazione armonica tra i gdl riportati in tabella (24 e 30).

Nella tabella dei valori critici (riportata in precedenza), il valore di **Q**

- alla probabilità  $\alpha = 0.05$  è uguale a **4,166**
- alla probabilità  $\alpha = 0.01$  è uguale a **5,168**.

Ricordando che la varianza d'errore  $s_e^2 = 146,5$

calcolare il campo di variazione critico o intervallo di confidenza per un generico contrasto semplice tra 2 medie, mediante il valore di **HSD**.

Alla probabilità  $\alpha = 0.05$

$$\mathbf{HSD} = 4,166 \cdot \sqrt{\frac{146,5}{5,93}} = 4,166 \times 4,962 = 20,67$$

**HSD** risulta uguale a 20,67

mentre alla probabilità  $\alpha = 0.01$

$$\mathbf{HSD} = 5,168 \cdot \sqrt{\frac{146,5}{5,93}} = 5,168 \times 4,962 = 25,64$$

**HSD** è uguale a 25,64.

Successivamente, è utile costruire la matrice triangolare delle differenze tra le 5 medie ed effettuare i confronti con le due **HSD** calcolate per le due probabilità. Dall'analisi risulta che

- con probabilità  $\alpha < 0.01$  sono significative le differenze superiori a **25,64**,
- con probabilità  $\alpha < 0.05$  sono significative quelle comprese tra **25,64** e **20,67**
- le differenze minori di **20,67** non sono significative, avendo probabilità  $\alpha > 0.05$ .

(Le prime possono essere contrassegnate da un doppio asterisco; le seconde da un solo asterisco).

		MEDIE				
		A	B	C	D	E
		208,2	199,8	141,0	123,3	119,1
MEDIE		DIFFERENZE				
B	199,8	8,4	---	---	---	---
C	141,0	<b>67,2**</b>	<b>58,8**</b>	---	---	---
D	123,3	<b>85,5**</b>	<b>76,5**</b>	17,7	---	---
E	119,1	<b>89,1**</b>	<b>80,7**</b>	<b>21,9*</b>	4,2	---

L'interpretazione della tabella porta alla conclusione che sono molto significative ( $\alpha < 0.01$ ) le differenze (in grassetto con due asterischi) tra la media C, la media D e la media E rispetto sia alla media A che alla B e significativa ( $0.01 < \alpha < 0.05$ ) la differenza tra la media C e la E.

Esse sono le cause della differenza complessiva tra le 5 medie, valutata in precedenza con il test F.

*(Questa tecnica era usata alcuni anni fa; ora i computer permettono di riportare il valore esatto di  $\alpha$  per ogni confronto e quindi di avere una visione più dettagliata, forse a discapito della sintesi).*

Per l'importanza che il test ha assunto tra i confronti multipli, a ulteriore chiarimento della metodologia è qui riproposta la presentazione di George E. P. **Box**, William G. **Hunter** e J. Stuart **Hunter** (nel testo: "*Statistics for Experimenters. An introduction to Design, Data Analysis and Model Building*", pubblicato nel 1978 da John Wiley & Sons, New York, p. 653).

**La procedura di Tukey** J. W. (vedi: (1949) *Comparing individual means in the analysis of variance*. Biometrics, 5, 99) **per il confronto simultaneo tra p medie** suggerisce di stimare il limite di confidenza tra tutte le differenze di  $p(p-1)/2$  coppie di medie ( $\bar{X}_i$  e  $\bar{X}_j$ ) con

$$(\bar{X}_i - \bar{X}_j) \pm \frac{Q_{\alpha/2, p, n}}{\sqrt{2}} \cdot \sqrt{s_e^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

dove

$Q_{p, n}$  è il valore di **q studentizzato** per il confronto tra **p** medie, con gdl **n** della varianza d'errore.

ESEMPIO. Individua quali differenze sono significative tra le **7** medie  $\bar{X}_i$  riportate nella tabella

Trattamenti	A	B	C	D	E	F	G
$\bar{X}_i$	53	52	57	55	55	60	50
$s_i^2$	9,2	8,7	8,8	9,8	10,2	8,3	8,0

ognuna con 4 dati (**n = 4**) e la varianza entro gruppo  $s_i^2$  uguale a quella indicata.

1 – Dapprima si stima una varianza comune a tutti i gruppi (o varianza d'errore)  $s^2 = 9,0$  che in questo caso, con campioni bilanciati, è uguale alla media delle varianze

$$s_e^2 = 3 (9,2 + 8,7 + 8,8 + 9,8 + 10,2 + 8,3 + 8,0) / 21 = 9,0$$

e ha gdl **n = 21**, pari a **p(n-1) = 7 (4-1)**.

Per **a = 0.05** e con

$$\frac{Q_{\alpha/2, p, n}}{\sqrt{2}} = 3,26$$

(tratto dalle tabelle relative)

alla probabilità del 95% si stima **un intervallo fiduciale o differenza minima significativa delle differenze tra le medie** pari a

$$\pm \frac{Q_{\alpha/2, p, n}}{\sqrt{2}} \cdot \sqrt{s_e^2 \cdot \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} = \pm 3,26 \cdot \sqrt{9,0 \cdot \left( \frac{1}{4} + \frac{1}{4} \right)} = \pm 6,91$$

**6,91 senza considerare il segno.**

Con le sette medie precedenti,

Trattamenti	A = 53	B = 52	C = 57	D = 55	E = 55	F = 60	G = 50
A = 53	---	---	---	---	---	---	---
B = 52	1	---	---	---	---	---	---
C = 57	-4	-5	---	---	---	---	---
D = 55	-2	-3	2	---	---	---	---
E = 55	-2	-3	2	0	---	---	---
F = 60	<b>-7*</b>	<b>-8*</b>	-3	-5	-5	---	---
G = 50	3	2	<b>7*</b>	5	5	<b>-10*</b>	---

si possono stimare 21 differenze (riportate in grassetto nella tabella), di cui risultano significative le quattro in grassetto e con l'asterisco.

6) **La procedura di Scheffé è il test più versatile** per i confronti multipli: offre il vantaggio di eseguire anche **confronti complessi oltre a quelli semplici**. Il numero totale di confronti è quindi superiore a quello permesso dal test di Tukey. Pertanto, sulla base del principio del Bonferroni, cambiano i rapporti tra **comparisonwise e experimentwise**, cioè tra la **potenza** o probabilità  $\alpha$  di ogni confronto e la **protezione** o probabilità  $\alpha_T$  complessiva.

Il test di Scheffé ha varie applicazioni, delle quali le più diffuse sono tre:

A - confronto tra due medie, sia semplici che complesse;

B – scomposizione della devianza di un gruppo di medie, per valutare il contributo relativo di ogni confronto;

C – matrice di tutte le possibili differenze, sia semplici che complesse.

A - Nel caso di confronti **sia semplici che complessi**, con il test di Scheffé risulta significativa la differenza **D**, se il valore di **L**, ottenuto dalle medie  $\bar{X}_i$  e dai contrasti  $c_i$ ,

con la relazione

$$L = \sum (c_i \cdot \bar{X}_i)$$

in valore assoluto è maggiore di **S**

$$S = \sqrt{(p-1) \cdot F_{(a;p-1,n)} \cdot \sum c_i^2 \cdot \left(\frac{s_e^2}{n}\right)}$$

dove

- **p** = numero di gruppi interessati al confronto,
- **F<sub>(a;p-1,n)</sub>** = valore dell'**F di Fisher** alla probabilità **a** prescelta e con gdl (**p-1; n**) uguali a quelli del test ANOVA ad un criterio di classificazione,
- **c<sub>i</sub>** = valore dei singoli contrasti implicati nel confronto.

La metodologia può essere spiegata in modo chiaro mediante l'illustrazione di un esempio, scomposto dettagliatamente nei suoi passaggi fondamentali.

Si supponga che, con le 5 medie riportate nella tabella sottostante e stimate su campioni con lo stesso numero di dati (6)

<b>ZONE</b>	A	B	C	D	E
<b>Medie</b>	208,2	199,8	141,0	123,3	119,1
<b>n<sub>i</sub></b>	6	6	6	6	6

si intenda valutare specificatamente la significatività della differenza tra il livello medio dell'inquinamento rilevato nelle due stazioni del centro (A e B) rispetto a quello delle due stazioni collocate in periferia (D e E), ricordando che la varianza d'errore  $s_e^2$  è risultata uguale a **146,5** con **25** gdl.

1 - Si verifica l'ipotesi nulla

$$H_0: \delta = 0$$

contro l'ipotesi alternativa bilaterale

$$H_1: \delta \neq 0$$

valutando la significatività della differenza **D** tra la media del centro rispetto a quella della periferia

$$D = \frac{\bar{X}_A + \bar{X}_B}{2} - \frac{\bar{X}_D + \bar{X}_E}{2} = \frac{208,2 + 199,8}{2} - \frac{123,3 + 119,1}{2} = 204 - 121,2 = 82,8$$

in cui **D** risulta uguale a 82,8.

2 - A questo scopo, si calcola il valore del contrasto **L**

$$L = \bar{X}_A + \bar{X}_B - \bar{X}_D - \bar{X}_E = 208,2 + 199,8 - 123,3 - 119,1 = 165,6$$

che risulta uguale a 165,6

e il cui valore  $\sum c_i^2$ , derivato dai coefficienti ortogonali, è

$$\sum c_i^2 = +1^2 + 1^2 - 1^2 - 1^2 = 4$$

uguale a 4.

3 – Alla probabilità  $\alpha = 0.05$ , con  $F_{0.025, 3, 25} = 3,69$  (tratto dalla tabella F di Fisher),  
 $p = 4$  e  $n = 6$

$$S = \sqrt{3 \cdot 3,68 \cdot 4 \cdot \frac{146,5}{6}} = 32,8$$

S risulta uguale a 32,8.

4 – Poiché il valore del contrasto **L** (165,6) è nettamente maggiore del valore **S** (32,8) calcolato, la differenza **D** (82,8) tra le due medie è significativa ad una probabilità  $\alpha$  nettamente inferiore a 0.05.

**Il test di Scheffé è il più adatto, quando il numero di gruppi e quello di osservazioni per gruppo sono molto piccoli e parte dei confronti non sono ortogonali.** Poiché permette di verificare tutti i confronti possibili, **se l'ANOVA è risultata significativa alla probabilità  $\alpha$ , almeno uno di tutti i possibili confronti dovrebbe risultare significativo alla stessa probabilità  $\alpha$  o a una probabilità inferiore.**

B - Il metodo di Scheffé, nell'esempio precedente applicato ad un singolo confronto tra due medie complesse, con le modifiche proposte da **Gabriel K. R.** (vedi: *A simple method of multiple comparisons of means*, pubblicato nel 1978 da *J. Amer. Stat. Assn.* Vol. 73, pp 724-729) è stato esteso a tutti i possibili confronti tra gruppi di medie e a tutte le possibili suddivisioni delle **p** medie.

I concetti di base di **questa diversa applicazione del test di Scheffé**, indicata con l'acronimo **SS-STP** (*Sum of Squares – Simultaneous Test Procedure*), possono essere derivati dall'ANOVA ad un criterio di classificazione.

Con l'ANOVA l'ipotesi nulla su **p** medie

$$H_0: \mu_1 = \mu_2 = \dots = \mu_p$$

è rifiutata quando

$$\frac{s_{tra}^2}{s_{entro}^2} \geq F_{\alpha, p-1, n-p}$$

Questo rapporto tra le varianze può essere scritto in modo differente: si divide la **Devianza tra** (indicata con **SS<sub>tra</sub>** da *Sum of Squares*) per i suoi gdl (**p-1**), per cui la formula diventa

$$\frac{SS_{tra}}{s_e^2 \cdot (p-1)} \geq F_{\alpha, p-1, n-p}$$

Da questa relazione si deriva facilmente

$$SS_{tra} \geq (p-1) \cdot s_e^2 \cdot F_{\alpha, p-1, n-p}$$

il valore di significatività.

Questa relazione può essere usata per **qualsiasi confronto non pianificato tra p medie, con n osservazioni per gruppo.**

L'applicazione ai dati della tabella sottostante, in cui ognuna delle 4 medie interessate al confronto (A, B, D, E) ha lo stesso numero di dati (6) e la varianza d'errore  $s_e^2$  è assunta uguale a **146,5**

ZONE	A	B	C	D	E
Medie	208,2	199,8	141,0	123,3	119,1
$n_i$	6	6	6	6	6
$\sum x_i$	1249,2	1198,8	846,0	739,8	714,6

**spiega come effettuare tutti i possibili confronti tra le 4 medie A, B, D, E, prescelte.**

Il metodo permette **sia confronti ortogonali, sia non ortogonali** tra queste 4 medie, quali

- A vs. B + D + E
- A + B vs. D + E
- A vs. D; ecc....

1 - A questo scopo, si calcola il valore della **Devianza tra (SS)**

$$\sum \frac{(\sum x_i)^2}{n_i} - \frac{(\sum x)^2}{n}$$

Considerando le 4 medie a confronto

$$SS = (1249,2)^2/6 + (1198,8)^2/6 + (739,8)^2/6 + (714,6)^2/6 - (3902,4)^2/24 =$$

$$260.083,44 + 239.520,24 + 91.217,34 + 85.108,86 - 638.432,64 = 37.497,24$$

SS risulta uguale a 37.497,24.

2 - Il **valore critico** di confronto con

- **p = 4**
- **$s_e^2 = 146,5$**
- **F** per  $\alpha = 0.05/2$ , e **gdl 3, 20 = 3,10**

è

$$(4-1) \times 146,5 \times 3,1 = 1.362,45$$

uguale a 1.362,45.

Ovviamente, poiché **SS** (37.497,24) è molto maggiore del **valore critico** (1.362,45), si rifiuta l'ipotesi nulla: le 4 medie sono tra loro significativamente diverse.

3 - Con i confronti ortogonali, è possibile **scomporre la devianza SS** prima calcolata. Diventa così possibile valutare il contributo di ogni contrasto a questo valore **SS** complessivo, fondato sulle differenze tra le 4 medie.

Si tratta di applicare la tecnica dei confronti ortogonali su queste 4 medie, come illustrata nei confronti a priori.

C - Il **metodo più generale e più versatile di utilizzazione del test di Scheffé** permette di costruire una **matrice di tutte le possibili differenze** tra tutte le medie considerate; in questo caso, le 5 medie riportate nella tabella successiva. Con la **procedura MSD**, risultano significative alla probabilità  $\alpha$  prefissata tutte le differenze che sono maggiori della quantità

$$MSD \geq \sqrt{(p-1) \cdot F_{\alpha; (p-1), n}} \cdot \sqrt{s_e^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

dove,

(oltre alla consueta simbologia relativa al confronto tra due medie)

**p** è il numero di medie semplici o di più gruppi che si intende confrontare,

$F_{\alpha; (p-1), n}$  è il valore di **F** alla probabilità  $\alpha_T$  prescelta e per i gradi di libertà **(p-1)** (dati dal numero di medie **p**) e **n** della varianza d'errore.

Con

- **p = 5** e **n = 6**
- **s<sub>e</sub><sup>2</sup> = 146,5**
- **F** per  $\alpha = 0.05$ , e **gdl 4, 25 = 2,76**

si ottiene un valore MSD

$$MSD = \sqrt{(5-1) \cdot 2,76} \cdot \sqrt{146,5 \cdot \left( \frac{1}{6} + \frac{1}{6} \right)} = \sqrt{11,04} \cdot \sqrt{48,633} = 23,21$$

uguale a 23,21

Con

- **F** per  $\alpha = 0.01$ , e **gdl 4, 25 = 4,18**

$$MSD = \sqrt{(5-1) \cdot 4,18} \cdot \sqrt{146,5 \cdot \left( \frac{1}{6} + \frac{1}{6} \right)} = \sqrt{16,72} \cdot \sqrt{48,633} = 28,51$$

si ottiene un MSD uguale a 28,51.

Con la stessa tecnica già illustrata in precedenza, nella tabella delle **p(p-1)/2 differenze** si evidenziano con

- un asterisco tutte le differenze che sono significative ad una probabilità  $\alpha$  compresa tra 0.05 e 0.01
- due asterischi tutte le differenze significative ad una probabilità  $\alpha$  minore di 0.01

		MEDIE				
		A	B	C	D	E
		208,2	199,8	141,0	123,3	119,1
MEDIE		DIFFERENZE				
B	199,8	8,4	---	---	---	---
C	141,0	<b>67,2**</b>	<b>58,8**</b>	---	---	---
D	123,3	<b>85,5**</b>	<b>76,5**</b>	17,7	---	---
E	119,1	<b>89,1**</b>	<b>80,7**</b>	21,9	4,2	---

E' importante osservare che il valore la differenza (21,9) tra la media **C** (141,0) e la media **E** (119,1) non è significativa alla probabilità  $\alpha = 0.05$ , mentre la era **con il test di Tukey**.

Più in generale, in questo esempio, **con Scheffé** la **differenza minima significativa** sulle 5 medie è

- **23,21** alla probabilità  $\alpha = 0.05$
- **28,51** alla probabilità  $\alpha = 0.01$

mentre **con il test di Tukey** era

- **20,67** alla probabilità  $\alpha = 0.05$
- **25,64** alla probabilità  $\alpha = 0.01$

Questo confronto tra Tukey e Scheffé e il confronto tra i risultati della terza applicazione del test di Scheffé rispetto alle prime due evidenziano la **relazione inversa tra versatilità e potenza di un test**. Tanto più il test è generale, tanto meno è potente. Lo stesso concetto è applicato quando i campioni hanno dimensioni diverse: la scelta prudentiale o cautelativa della frequenza **n** minore rende le conclusioni generali, cioè estese a tutti i possibili confronti dell'esperimento, ma diminuisce la potenza di ogni confronto.

Il metodo di Scheffé ha applicazioni più ampie per numero di confronti; ma **è più prudentiale di quello SNK e di quello di Tukey**.

**Se interessano solo i confronti singoli, è svantaggioso utilizzare un metodo che permette anche confronti complessi; se nel piano dell'esperimento l'interesse è rivolto solo ad alcuni confronti, è svantaggioso per la significatività utilizzare un metodo che li permette tutti.**

7) Il **test Q di Dunnett** deve essere applicato in un caso particolare: **il confronto di due o più trattamenti con un controllo.**

In numero di confronti da effettuare diminuisce: è uguale al numero di trattamenti, escludendo il controllo. Con 5 gruppi, fra cui un controllo e 4 trattamenti, non è più uguale a 10 ( $C_5^2$ ) ma a 4.

Sulla base per principio del Bonferroni, aumenta quindi la potenza di ogni confronto, poiché con una probabilità experimentwise  $\alpha_T = 0.05$  la probabilità a comparisonwise diventa uguale 0.0125 (0.05 /4).

**Ovviamente questa scelta riduce la versatilità del test: si verifica la significatività della differenza tra ogni trattamento ed il controllo, senza poter dire nulla sulle eventuali differenze tra i trattamenti**, cioè se uno è migliore o peggiore degli altri in modo significativo.

**Il test Q di Dunnett utilizza la stessa formula del test SNK**

(con la medesima simbologia)

$$Q_{(\alpha, p, n)} = \frac{\bar{X}_c - \bar{X}_i}{\sqrt{s_e^2 \cdot \left( \frac{1}{n_c} + \frac{1}{n_i} \right)}}$$

dove

- **c = gruppo di controllo o placebo**
- **i = trattamento in oggetto**, per il quale si effettua il confronto con il controllo.

A differenza degli altri test per confronti multipli, che **a parità del numero totale di osservazioni** raggiungono la potenza maggiore quando tutti i gruppi sono bilanciati, nel confronto tra vari trattamenti con un controllo **si ottiene una utilizzazione migliore dei dati quando il controllo è di dimensioni ragionevolmente maggiori di quella dei trattamenti.**

Infatti, il controllo entra in tutti i confronti ed un numero più alto di osservazioni in esso aumenta la potenza di ogni confronto, anche se determina una parziale penalizzazione per il mancato bilanciamento.

### Valori critici del Q per il test di Dunnett

**p** = numero di medie a confronto (compreso il controllo)

**n** = gradi di libertà della varianza d'errore

$$a_T = 0.05$$

#### NUMERO **p** DI MEDIE A CONFRONTO

n	2	3	4	5	6	7	8	9	10
<b>10</b>	2,23	2,57	2,76	2,89	2,99	3,07	3,14	3,19	3,24
<b>11</b>	2,20	2,53	2,72	2,84	2,94	3,02	3,08	3,14	3,19
<b>12</b>	2,18	2,50	2,68	2,81	2,90	2,98	3,04	3,09	3,14
<b>13</b>	2,16	2,48	2,65	2,78	2,87	2,94	3,00	3,06	3,10
<b>14</b>	2,14	2,46	2,63	2,75	2,84	2,91	2,97	3,02	3,07
<b>15</b>	2,13	2,44	2,61	2,73	2,82	2,89	2,95	3,00	3,04
<b>16</b>	2,12	2,42	2,59	2,71	2,80	2,87	2,92	2,97	3,02
<b>17</b>	2,11	2,41	2,58	2,69	2,78	2,85	2,90	2,95	3,00
<b>18</b>	2,10	2,40	2,56	2,68	2,76	2,83	2,89	2,94	2,98
<b>19</b>	2,09	2,39	2,55	2,66	2,75	2,81	2,87	2,92	2,96
<b>20</b>	2,09	2,38	2,54	2,65	2,73	2,80	2,86	2,90	2,95
<b>24</b>	2,06	2,35	2,51	2,61	2,70	2,76	2,81	2,86	2,90
<b>30</b>	2,04	2,32	2,47	2,58	2,66	2,72	2,77	2,82	2,86
<b>40</b>	2,02	2,29	2,44	2,54	2,62	2,68	2,73	2,77	2,81
<b>60</b>	2,00	2,27	2,41	2,51	2,58	2,64	2,69	2,73	2,77
<b>120</b>	1,98	2,24	2,38	2,47	2,55	2,60	2,65	2,69	2,73
¥	1,96	2,21	2,35	2,44	2,51	2,57	2,61	2,65	2,69

$$a_T = 0.01$$

#### NUMERO **p** DI MEDIE A CONFRONTO

n	2	3	4	5	6	7	8	9	10
<b>10</b>	3,17	3,53	3,74	3,88	3,99	4,08	4,16	4,22	4,28
<b>11</b>	3,11	3,45	3,65	3,79	3,89	3,98	4,05	4,11	4,16
<b>12</b>	3,05	3,39	3,58	3,71	3,81	3,89	3,96	4,02	4,07
<b>13</b>	3,01	3,33	3,52	3,65	3,74	3,82	3,89	3,94	3,99
<b>14</b>	2,98	3,29	3,47	3,59	3,69	3,76	3,83	3,88	3,93
<b>15</b>	2,95	3,25	3,43	3,55	3,64	3,71	3,78	3,83	3,88
<b>16</b>	2,92	3,22	3,39	3,51	3,60	3,67	3,73	3,78	3,83
<b>17</b>	2,90	3,19	3,36	3,47	3,56	3,63	3,69	3,74	3,79
<b>18</b>	2,88	3,17	3,33	3,44	3,53	3,60	3,66	3,71	3,75
<b>19</b>	2,86	3,15	3,31	3,42	3,50	3,57	3,63	3,68	3,72
<b>20</b>	2,85	3,13	3,29	3,40	3,48	3,55	3,60	3,65	3,69
<b>24</b>	2,80	3,07	3,22	3,32	3,40	3,47	3,52	3,57	3,61
<b>30</b>	2,75	3,01	3,15	3,25	3,33	3,39	3,44	3,49	3,52
<b>40</b>	2,70	2,95	3,09	3,19	3,26	3,32	3,37	3,41	3,44
<b>60</b>	2,66	2,90	3,03	3,12	3,19	3,25	3,29	3,33	3,37
<b>120</b>	2,62	2,85	2,97	3,06	3,12	3,18	3,22	3,26	3,29
¥	2,58	2,79	2,92	3,00	3,06	3,11	3,15	3,19	3,22

Quando si programma un esperimento sul quale deve essere applicato il test di Dunnett, è conveniente che

- $n_c$ , il numero di dati del controllo, sia più numeroso di
- $n_i$ , il numero dei dati di ogni trattamento,
- in funzione del numero di trattamenti  $p$

secondo la relazione

$$n_c = n_i \cdot \sqrt{p}$$

Ad esempio, in un esperimento con 7 dati in ognuno dei 5 gruppi (il controllo più 4 trattamenti e quindi 35 osservazioni in tutto), si ottiene la migliore utilizzazione complessiva

$$n_c = 7 \cdot \sqrt{5} = 15,65$$

quando

- 15 cavie sono dedicate al controllo e
- le rimanenti 20 sono suddivise tra i 4 trattamenti.

E' una indicazione approssimata, in quanto è semplice verificare che nella formula

$$Q_{(a,p,n)} = \frac{\bar{X}_c - \bar{X}_i}{\sqrt{s_e^2 \cdot \left( \frac{1}{n_c} + \frac{1}{n_i} \right)}}$$

si ottiene il valore massimo di Q (quindi il risultato più significativo)

quando (a parità di tutti gli altri parametri)

$$\frac{1}{n_c} + \frac{1}{n_i} = X_{\min}$$

la somma dei due rapporti ha il valore minimo,

ovviamente mantenendo costante il numero totale  $n$  di dati.

Una stima più precisa ed una verifica degli effetti di questa concentrazione delle osservazioni sul campione di controllo può essere ottenuta con un confronto dettagliato delle varie possibili distribuzioni del numero complessivo di cavie disponibili nei vari gruppi.

Con 35 osservazioni in totale,

- nel caso di campioni bilanciati e quindi  $n_c = 7$  e  $n_i = 7$  si avrebbe  $\frac{1}{7} + \frac{1}{7} = 0,2857$

- nel caso di  $n_c = 11$  e  $n_i = 6$  si avrebbe  $\frac{1}{11} + \frac{1}{6} = 0,0909 + 0,1667 = 0,2576$

- nel caso di  $n_c = 15$  e  $n_i = 5$  si avrebbe  $\frac{1}{15} + \frac{1}{5} = 0,0667 + 0,2000 = 0,2667$

- nel caso di  $n_c = 19$  e  $n_i = 4$  si avrebbe  $\frac{1}{19} + \frac{1}{4} = 0,0526 + 0,2500 = 0,3026$

Per ottenere la maggiore potenza del test, con 35 cavie e 5 gruppi, la scelta più vantaggiosa è collocare 11 cavie nel gruppo di controllo e 6 in ognuno degli altri 4 trattamenti.

ESEMPIO 1. Si è voluto esaminare l'effetto di 6 diverse sostanze tossiche sull'accrescimento somatico di una specie planctonica (misurati in mm dopo 20 giorni dalla schiusa delle uova), per verificare quali di esse riducano significativamente le dimensioni medie (test unilaterale) allo stato adulto.

Con i seguenti risultati ottenuti in laboratorio:

	CONTROLLO	SOSTANZE TOSSICHE					
		A	B	C	D	E	F
Media	3,25	2,80	2,18	2,96	2,24	2,39	2,67
Osservazioni	10	7	7	7	7	7	7

per un totale di 52 osservazioni, di cui 10 nel gruppo controllo.

L'analisi della varianza con  $F_{(7, 45)}$  ha permesso di rifiutare l'ipotesi nulla; la varianza d'errore  $s_e^2$  con 45 gdl è risultata uguale a 0,36. Verificare quali sostanze hanno un effetto significativo alla probabilità  $\alpha = 0,05$  e quali anche alla probabilità  $\alpha = 0,01$  in rapporto al controllo.

Risposta.

I confronti da effettuare sono 6. E' possibile stimare una differenza minima significativa (MDS) unica, poiché i trattamenti hanno tutti lo stesso numero d'osservazioni

$$Q_{(\alpha; p, n-p)} = \sqrt{s_e^2 \cdot \left(\frac{1}{n_c} + \frac{1}{n_i}\right)}$$

Con i dati dell'esempio ( $p = 6$  e  $gdl = 40$ ), nella tavola dei valori critici

- alla probabilità  $\alpha = 0,05$  il valore del **Q di Dunnett** è uguale a 2,62
- alla probabilità  $\alpha = 0,01$  è uguale a 3,26.

Pertanto,

- alla probabilità  $\alpha = 0.05$

il valore della MDS

$$2,62 \cdot \sqrt{0,36 \cdot \left(\frac{1}{10} + \frac{1}{7}\right)} = 2,62 \cdot \sqrt{0,36 \cdot 0,243} = 2,62 \cdot 0,296 = 0,775$$

è uguale a 0,775 e

- alla probabilità  $\alpha = 0.01$

$$3,26 \cdot \sqrt{0,36 \cdot \left(\frac{1}{10} + \frac{1}{7}\right)} = 3,26 \cdot \sqrt{0,36 \cdot 0,243} = 3,26 \cdot 0,296 = 0,965$$

MDS è uguale a 0,965.

Si calcolano le differenze dei 6 trattamenti rispetto al controllo e si verifica la loro significatività mediante il confronto con i due valori MDS stimati. Possono essere segnate con

- due asterischi le differenze maggiori del valore 0,965 e
- un asterisco le differenze comprese tra 0,965 e 0,775.

A	$3,25 - 2,80 = \mathbf{0,45}$
B	$3,25 - 2,18 = \mathbf{1,07^{**}}$
C	$3,25 - 2,96 = \mathbf{0,29}$
D	$3,25 - 2,24 = \mathbf{1,01^{**}}$
E	$3,25 - 2,39 = \mathbf{0,86^*}$
F	$3,25 - 2,67 = \mathbf{0,58}$

La tabella evidenzia che, delle 6 sostanze tossiche esaminate nell'esperimento, rispetto al controllo hanno un effetto molto significativo ( $\alpha < 0.01$ ) la B e la D, mentre ha un effetto significativo ( $\alpha < 0.05$ ) la E. Le sostanze A, C ed F non hanno ridotto la crescita in modo significativo rispetto al controllo ( $\alpha > 0.05$ ).

ESEMPIO 2. Questo secondo esempio è tratto dal testo di George E. P. **Box**, William G. **Hunter** e J. Stuart **Hunter** (nel testo: "*Statistics for Experimenters. An introduction to Design, Data Analysis and Model Building*", pubblicato nel 1978 da John Wiley & Sons, New York, p. 653) che individua

nel metodo di Tukey e in quello di Dunnett le due proposte fondamentali, per analisi da effettuare con calcoli manuali.

**La procedura di Dunnett** per il confronto tra **k** medie con la media di un campione standard o controllo richiede ovviamente il calcolo e l'analisi di **k-1** differenze.

Per ogni differenza ( $\bar{X}_i - \bar{X}_c$ ) tra la media di un generico trattamento **i** ( $\bar{X}_i$ ) e la media del controllo ( $\bar{X}_c$ ) si stima un intervallo fiduciale

$$(\bar{X}_i - \bar{X}_c) \pm t_{k,n,a/2} \cdot \sqrt{s_e^2 \left( \frac{1}{n_i} + \frac{1}{n_c} \right)}$$

in cui,

- al posto del valore di **q studentizzato**, viene utilizzato
- quello di **t** per **k** confronti, con gdl n e la probabilità **a/2**.

Con le 7 medie di prima in cui G sia il controllo

Trattamenti	A	B	C	D	E	F	G = Controllo
$n_i$	4	4	4	4	4	4	4
$\bar{X}_i$	53	52	57	55	55	60	50

alla probabilità del 95% dove  $t_{7,21,0.025} = 2,80$

si stima una differenza minima significativa

$$\pm t_{k,n,a/2} \cdot \sqrt{s_e^2 \left( \frac{1}{n_i} + \frac{1}{n_c} \right)} = \pm 2,80 \cdot \sqrt{9,0 \cdot \left( \frac{1}{4} + \frac{1}{4} \right)} = \pm 5,94$$

che risulta uguale a 5,94 (osservare che è minore del valore precedente, in quanto **stimato non per k(k-1)/2 confronti ma per k-1**).

Di conseguenza, tra le 6 differenze riportate nella tabella successiva

Trattamenti	A	B	C	D	E	F
$\bar{X}_i$	53	52	57	55	55	60
Differenze $\bar{X}_i - \bar{X}_c$	<b>3</b>	<b>2</b>	<b><u>7</u></b>	<b>5</b>	<b>5</b>	<b><u>10</u></b>

Sono significativamente diverse, dalla media del controllo, la media del trattamento C e quella del trattamento F.

8) Il **metodo di Duncan**, chiamato **test del campo di variazione multiplo**, può essere applicato nelle **stesse condizioni del test di Tukey, ma con una significatività *experiment wise* meno cautelativa e quindi una potenza per un singolo test o *comparisonwise* maggiore**. Di conseguenza, rende significative differenze tra coppie di singole medie che non risultano tali con il test di Tukey e il test SNK, facendosi per questo preferire da vari ricercatori, soprattutto negli ultimi anni; altri continuano a preferire Tukey, appunto perché più prudentiale e quindi “meno contestabile”, una volta dimostrata la significatività di una differenza tra due medie.

Il metodo è fondato sulla **procedura step-wise**, ma con un approccio innovativo.

**La procedura classica del campo di variazione studentizzato**, dopo aver disposto le medie a confronto in ordine di rango, prende in considerazione solo il numero di medie coinvolto in ogni singolo confronto. Per confrontare due medie adiacenti (come già spiegato nel presentare il metodo **LSD**, con  $p = 2$  e detta **comparazione delle medie due passi avanti**) utilizza il campo di variazione studentizzato su due medie; per valutare la differenza tra due medie tre passi avanti, stima il campo di variazione studentizzato su tre medie; più in generale, per confrontare le differenze tra medie  $p$  passi avanti, calcola il campo di variazione su  $p$  medie.

Quando stimato per campioni di dimensioni  $n$ , il **campo di variazione studentizzato (Q)** determinato da

$$Q = \frac{\bar{X}_{\max} - \bar{X}_{\min}}{\sqrt{\frac{s_e^2}{n}}}$$

risulta minore quando il numero  $p$  di medie a confronto è minore.

Ne deriva che la stessa differenza, in valore assoluto, tra due medie può risultare

- **significativa** se il confronto è effettuato **tra due medie adiacenti**,
- **non significativa** se effettuata **tra medie distanti tre o più passi avanti**.

Per un **principio di cautela**, la **procedura classica** dei confronti a posteriori contrasta questo ultimo effetto: **esclude che possa essere dichiarata significativa la differenza tra due medie, quando trovano all'interno di un'altra coppia non significativa**.

La **procedura proposta da Duncan**, sempre dopo aver disposto le medie in ordine di rango, rovescia la logica precedente.

Dapprima definisce il **livello di protezione per due medie separate da  $r$  passi**

come

$$(1 - \alpha)^{r-1}$$

Di conseguenza, la probabilità di commettere un errore di I tipo, cioè quella di rifiutare l'ipotesi nulla quando essa è vera, tra due medie campionarie separate da  $r$  passi è

$$1 - (1 - \alpha)^{r-1}$$

dove

- $\alpha$  = livello di significatività di un singolo confronto
- $r$  = numero di passi, tra le due medie a confronto.

Tra **due medie adiacenti**, in cui  $r = 2$ , il livello di protezione è uguale a  $\alpha$ : è facile verificare, con un semplice confronto tra le tabelle relative, che il valore del campo di variazione multiplo di Duncan è uguale a quello del Q studentizzato.

Ma per **due medie non adiacenti**, al crescere del numero ( $r$ ) di passi, il livello di protezione o probabilità experiment-wise si riduce progressivamente, rendendo il test di Duncan sempre più potente nei confronti dei test fondati sul valore del Q. Per una indicazione semplice di questo effetto, è sufficiente confrontare la tabella di **Duncan** con quella del **Q studentizzato** per il test SNK e il test W di Tukey: alla stessa probabilità  $\alpha$  e per i medesimi gdl, il valore di **Duncan** è minore di quello della tabella **Q**, in modo sempre più accentuato al crescere di  $r$ .

- Con le stesse 5 medie utilizzate in precedenza,

Zone	A	B	C	D	E
Medie	208,2	199,8	141,0	123,3	119,1
$n_i$	6	6	6	6	6

supponendo

- un numero di osservazioni o repliche costante in ogni gruppo:  $n = 6$ ,
- una varianza d'errore  $s_e^2$  uguale a **146,5**
- gdl  $v = 25$

è possibile verificare la significatività della differenza tra ogni coppia di medie.

Dopo aver ordinato le medie in ordine decrescente (o crescente, come altri preferiscono), in relazione al rango di ognuna di esse, si stima il numero di passi  $r$ , che in questo caso, con 5 medie, può variare da 2 a 5.

Punteggi per il test del campo di variazione multiplo di Duncan  
 $\alpha = 0.05$

V	$r =$ numero di passi ordinati tra le medie													
	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
2	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
3	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
4	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
5	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
6	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
7	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
8	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
9	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48
11	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.46	3.46	3.46	3.47	3.48
12	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48
13	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45	3.46	3.46	3.47	3.47
14	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47
15	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47
16	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47
17	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47
18	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47
19	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47
20	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.46	3.47
22	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47
24	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47
26	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47
28	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47
30	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47
60	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47
100	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47
$\infty$	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47

Punteggi per il test del campo di variazione multiplo di Duncan  
 $\alpha = 0.01$

n	r = numero di passi ordinati tra le medie													
	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	8.26	8.50	8.60	8.70	8.80	8.90	8.90	9.00	9.00	9.00	9.10	9.20	9.30	9.30
4	6.51	6.80	6.90	7.00	7.10	7.10	7.20	7.20	7.30	7.30	7.40	7.40	7.50	7.50
5	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.50	6.60	6.60	6.70	6.70	6.80
6	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.00	6.10	6.20	6.20	6.30	6.30
7	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.80	5.80	5.90	5.90	6.00	6.00
8	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.50	5.60	5.70	5.70	5.80	5.80
9	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.40	5.50	5.50	5.60	5.70	5.70
10	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55
11	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39
12	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.23	5.26
13	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15
14	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91	4.96	5.00	5.04	5.06	5.07
15	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00
16	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84	4.88	4.91	4.93	4.94
17	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89
18	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.85
19	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82
20	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	4.76	4.78	4.79
22	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.68	4.71	4.74	4.75
24	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72
26	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69
28	3.91	3.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67
30	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65
40	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.57	4.59
60	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53
100	3.71	3.86	3.93	4.06	4.11	4.17	4.21	4.25	4.29	4.35	4.38	4.42	4.45	4.48
$\infty$	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41

Ognuna delle 10 differenze (colonna 5) risulta significativa, se superiore alla **differenza minima significativa** (MDS), ottenuta con la solita formula

$$\text{MDS} = C_{(\alpha, r, n)} \cdot \sqrt{\frac{s_e^2}{n}}$$

dove

$C_{(\alpha, r, v)}$  = valore riportato nella **tabella di Duncan**.

Nell'esempio e nella tabella sottostante, il valore di C è stato scelto per

- la probabilità  $\alpha = 0.05$
- il numero **r** di passi relativi al confronto
- $v = 24$ , in assenza del valore per gdl 25, in quanto più cautelativo (determina infatti una MDS maggiore di quella che si otterrebbe scegliendo in valore per  $v = 26$ )

1	2	3	4	5	6	7
Ordine	Confronto	r passi	Medie	Differenza	Q	DUNCAN
1	5 <sup>a</sup> vs 1 <sup>a</sup>	5	208,2 – 119,1	89,1	4,166	3,22
2	5 <sup>a</sup> vs 2 <sup>a</sup>	4	208,2 – 123,3	84,9	3,901	3,15
3	5 <sup>a</sup> vs 3 <sup>a</sup>	3	208,2 – 141,0	67,2	3,532	3,07
4	5 <sup>a</sup> vs 4 <sup>a</sup>	2	208,2 – 199,8	8,4	2,919	2,92
5	4 <sup>a</sup> vs 1 <sup>a</sup>	4	199,8 – 119,1	80,7	3,901	3,15
6	4 <sup>a</sup> vs 2 <sup>a</sup>	3	199,8 – 123,3	76,5	3,532	3,07
7	4 <sup>a</sup> vs 3 <sup>a</sup>	2	199,8 – 141,0	58,8	2,919	2,92
8	3 <sup>a</sup> vs 1 <sup>a</sup>	3	141,0 – 119,1	21,9	3,532	3,07
9	3 <sup>a</sup> vs 2 <sup>a</sup>	2	141,0 – 123,3	17,7	2,919	2,92
10	2 <sup>a</sup> vs 1 <sup>a</sup>	2	123,3 – 119,1	4,2	2,919	2,92

Come tutti gli altri test per i confronti multipli, anche quello di Duncan presenta differenti gradi di combinazione tra **specificità e sensitività**. Quando dalla stima della differenza minima significativa per ogni singolo confronto si passa alla stima di una quantità valida per tutti i confronti, **si aumenta la versatilità**, ma **si diminuisce in potenza**: il valore della MDS cresce.

**I confronti multipli o a posteriori sono uno dei settori in maggiore evoluzione, nell'attuale ricerca statistica.** Di conseguenza, vengono proposti metodi nuovi e non esiste ancora unanimità sulle scelte più adeguate, in rapporto al confronto da effettuare e alle dimensioni del campione. Tuttavia, in molti testi sono enunciati alcuni principi fondamentali:

1- **è conveniente programmare l'esperimento e attuare confronti pianificati**, poiché è più probabile che le differenze che interessano risultino significative;

2 – con i confronti pianificati, il numero di test è limitato dall'ortogonalità e dalla conoscenza delle caratteristiche dei vari gruppi; quindi, **per scoprire differenze non attese, può essere necessario ricorrere ai confronti a posteriori;**

3 – **confrontare la differenza tra la media minore e quella maggiore** (individuate sulla base dei dati raccolti) **è un test a posteriori**: la probabilità va calcolata sul numero di confronti possibili;

4 – **tra tutti i test a posteriori forniti dai programmi informatici, secondo vari autori di testi di statistica applicata appare lecito utilizzare il test che risulta più significativo, poiché sono tutti cautelativi e non è ancora dimostrata la maggiore correttezza di uno rispetto agli altri.;**

5 – per confronti tra **tutte le singole medie** di  $p$  gruppi, è possibile scegliere tra test di Tukey e test di Duncan: il primo è più cautelativo, ma meno potente.

#### **8.5. STIMA DELLA DIMENSIONE $N$ DI $K$ GRUPPI CAMPIONARI PER L'ANOVA**

Al momento di programmare il confronto tra più medie campionarie, un problema fondamentale è sapere **quanti dati è necessario raccogliere**, ovviamente allo scopo di rendere il test significativo.

Le **dimensioni  $n$  di ognuno dei  $k$  campioni dipendono essenzialmente da 4 variabili**, che occorre conoscere o determinare al momento della programmazione:

1 – la **differenza minima  $d$  tra almeno 2 medie**, di cui si intende verificare la significatività; la scelta del valore dipende dalla conoscenza del fenomeno o da uno studio preliminare;

quanto minore è  $d$  tanto maggiore deve essere la dimensione  $n$  di ogni campione;

2 – la **deviazione standard  $s$ , tratta dalla corrispondente varianza d'errore**; anche in questo caso deve essere nota attraverso dati riportati in letteratura, per l'esperienza del ricercatore oppure determinata da un esperimento pilota;

quanto minore è  $s$  tanto minore può essere la dimensione  $n$  di ogni campione;

3 – la **probabilità  $\alpha$** , alla quale si vuole che la differenza  $d$  risulti significativa, **in un test bilaterale**; di norma è fissato uguale a 0.05 oppure a 0.01;

quanto minore è  $a$  tanto maggiore deve essere la dimensione  $n$  di ogni campione;

4 – **la potenza  $1 - b$  del test, la probabilità di rifiutare l'ipotesi nulla quando è falsa, tratta da una distribuzione per test unilaterali;** è prassi accettare una probabilità pari a 80% oppure 90%, corrispondente ad una probabilità di  $b$  uguale a 0.20 oppure 0.10;

tanto minore è  $b$ , tanto maggiore è la potenza richiesta al test e quindi tanto maggiore deve essere anche la dimensione  $n$  di ogni campione.

Nel caso di un'analisi della varianza in cui si confrontano le medie di  $k$  gruppi, ognuno con  $n$  dati, i gdl dell'errore standard sono quelli della varianza d'errore, quindi uguali a  $n = k \times (n-1)$ .

Poiché è sufficiente che sia significativa la differenza tra 2 delle  $k$  medie a confronto,

- per la probabilità  $a$  **si ricorre alla distribuzione  $t$  di Student per un test bilaterale,**
- per la probabilità  $b$  **alla stessa distribuzione  $t$  di Student, ma per un test unilaterale.**

Affinché il test sia sufficientemente potente,

$n$  deve essere maggiore od uguale a

$$n \geq 2 \left( \frac{s}{d} \right)^2 \cdot (t(a, n) + t(b, n))^2$$

La stessa relazione è più frequentemente scritta come

$$n \geq 2j^2 \cdot (t(a, n) + t(b, n))^2$$

(gli indici di  $t$ , entro parentesi, non sono stati riportati a pedice per renderli di lettura più facile).

E' da sottolineare che **la formula richiede di conoscere**

- **il rapporto  $s / d$** , spesso indicato con  $j$ ; è più facile da ottenere che non i singoli valori, in quanto simile ad un coefficiente di variazione;

per utilizzare un valore indicativo, quando non si hanno informazioni è utile ricordare che l'esperienza ha dimostrato che il valore

- $j \gg 0,2$  è piccolo (variabilità ridotta rispetto alla media);
- $j \gg 0,5$  è medio;
- $j \gg 0,7$  è grande (variabilità ampia rispetto al valore della media);

- **il valore di  $t$  alla probabilità  $b$**  deve essere preso dalla tabella dei valori critici e nello stesso modo con il quale viene scelto quello della probabilità  $a$  **per un test bilaterale. Per prassi**, la probabilità di  $\beta$  è circa 4-5 volte quella di  $\alpha$ ; di conseguenza
- quando si ha  $\alpha=0.01$  si sceglie un valore di  $\beta = 0.05$ ,
- quando si ha  $\alpha=0.05$  si sceglie un valore di  $\beta = 0.20$ .

Sarebbe possibile prendere anche un valore di  $\beta = 0.5$ , che corrisponde alla probabilità del 50% che il campione raccolto non risulti significativo alla probabilità  $a$  prefissata; in questo caso, il valore di  $t$  ha distribuzione simmetrica ed è uguale a **0**.

**Quando, come tabella dei valori critici, si dispone solo di una distribuzione bilaterale, (vantaggiosa per trovare direttamente il valore di  $a$ ) per trovare il valore di  $b$  si deve utilizzare la colonna 2b.**

**Il calcolo di  $n$  è ottenuto con un processo iterativo**, quando non è possibile ricorrere a metodi grafici.

Di seguito è riportato il processo di calcolo, in quanto utile a comprendere i fattori in gioco nella scelta delle dimensioni del campione; sono anche le informazioni richieste dai programmi informatici più diffusi.

Il valore di  $t$  dipende dal numero  $n$  di gdl, determinato sulla base del numero  $k$  di gruppi e soprattutto del numero  $n$  di osservazioni entro ogni gruppo:  $n = k \times (n-1)$ .

Il metodo iterativo richiede:

- una prima stima di  $n$ , considerando che ogni gruppo abbia almeno  $n = 5-6$  osservazioni; con 4 gruppi, il valore di  $n$  diventa uguale a 16 – 20 e sulla base di questi gdl si scelgono i due valori di  $t$  (quello alla probabilità  $a$  e quello alla probabilità  $b$ );
- se il calcolo determina un valore di  $n$  maggiore dei 5-6 preventivati (ad esempio 10), si stima un nuovo  $n$  (uguale a 36 poiché  $(10-1) \times 4 = 36$ ) e si scelgono dalla tabella sinottica due nuovi valori di  $t$ ;
- dopo il nuovo calcolo, spesso si può osservare che il terzo valore di  $n$  è vicino al secondo: si sceglie quello più cautelativo, arrotondato all'unità per eccesso. Se la differenza tra il terzo valore di  $n$  ed il secondo fosse ritenuta ancora importante, si effettua un nuovo calcolo dopo aver modificato i valori di  $t$  corrispondenti ai nuovi gdl; quasi sempre la quarta stima è molto simile alla terza e con essa termina il processo iterativo.

**ESEMPIO.** Mediante un'analisi della varianza con 4 gruppi (un controllo e tre trattamenti), si intende dimostrare la significatività di una differenza (tra il controllo ed uno dei tre trattamenti) uguale a 11.

Dai dati già raccolti, è noto che la varianza è uguale a 150 e quindi  $\sigma$  è uguale a 12,2 (arrotondato alla prima cifra decimale), mentre il rapporto  $j$  ( $s / d$ ) è uguale a 0,9.

Quanti dati  $n$  occorre raccogliere per ognuno dei 4 campioni, affinché il test ANOVA risulti significativo alla probabilità  $\alpha$  uguale a 0.05 e con una potenza  $(1 - \beta)$  uguale al 90 per cento?

Risposta.

Si utilizza la formula

$$n \geq 2j^2 \cdot (t(a,n) + t(b,n))^2$$

in cui, con i dati del problema, si ha che

$$j (s / d) = 0,9$$

$$a = 0.05 \quad e \quad b = 0.10$$

**Nel 1° tentativo**, si scelgono i valori dei **gdl** e i valori di **t** corrispondenti, solo sulla base del buon senso (l'esperienza):

$$\text{con } k = 4 \quad e \quad v = 20,$$

se si ipotizza a priori che sia sufficiente  $n = 6$ ,

$$\text{poiché } n = k \times (n-1)$$

si devono scegliere i due valori di **t** con 20 gdl.

Dalla tabella dei valori critici si ricava che

$$t \text{ di } a (0.05, 20) = 2,086 \text{ (in una distribuzione per test bilaterale),}$$

$$t \text{ di } b (0.10, 20) = 1,325 \text{ (in una distribuzione per test unilaterale, corrispondente alla colonna 0.20 se la distribuzione è bilaterale).}$$

Dai parametri fissati, con la formula sopra riportata

si ottiene un valore di **n**

$$n \cong 2 \times 0,9^2 \times (2,086 + 1,325)^2 = 2 \times 0,81 \times 11.635 = 18,85$$

uguale a 19, per arrotondamento all'unità superiore.

Si può osservare che il valore stimato (19) è molto maggiore di quello ipotizzato all'inizio, uguale a 6.

Di conseguenza, il valore di **t** utilizzato con 20 gdl è errato e troppo grande in quanto fondato su pochi gdl. Si deve quindi procedere ad una iterazione, con un secondo tentativo di calcolo fondato su un valore di **t** più preciso.

**Nel 2° tentativo,**

prendendo come riferimento delle dimensioni di ogni gruppo  $n = 19$ ,

il valore di  $n$  è  $4 \times 18 = 72$ .

Poiché poche tabelle riportano i valori esatti di  $t$  per questo numero di gradi di libertà, ma approssimativamente per decine, come scelta cautelativa si utilizza  $n$  uguale a 70, che fornisce un valore di  $t$  maggiore di quello con 80 gdl e quindi anche un  $n$  maggiore.

I nuovi valori di  $t$  sono:

per  $\alpha = 0.05$  in un test bilaterale,  $t(0.05, 70) = 1,994$

per  $\beta = 0.10$  in un test unilaterale,  $t(0.10, 70) = 1,294$

La nuova stima di  $n$

$$n \cong 2 \times 0,9^2 \times (1,994 + 1,294)^2 = 2 \times 0,81 \times 10,81 = 17,51$$

risulta uguale a 18 per arrotondamento all'unità superiore.

Poiché il nuovo valore (18) non differisce sensibilmente dal valore calcolato in precedenza (19), si può concludere che per ognuno dei 4 gruppi sono sufficienti 18 o 19 dati.

L'esempio mette in evidenza che **per poter utilizzare pochi dati**, quindi avere un risparmio in costo di materiale e di tempo richiesti dall'esperimento,

è vantaggioso rendere il valore di  $j$  ( $s / d$ ) il **minimo** possibile, agendo

sulla **differenza**, affinché **sia grande** e

sulla **varianza** affinché **sia piccola**.

## CAPITOLO IX

### ANALISI DELLA VARIANZA A PIU' CRITERI DI CLASSIFICAZIONE

#### 9.1. ANALISI DELLA VARIANZA A DUE CRITERI DI CLASSIFICAZIONE O A BLOCCHI RANDOMIZZATI, CON UNA SOLA OSSERVAZIONE PER CASELLA

L'analisi della varianza ad un criterio di classificazione o completamente randomizzata è lo schema più semplice di confronto simultaneo tra più medie. Nella pratica sperimentale, **spesso rappresenta un'impostazione troppo elementare**; infatti, in modo implicito, assume che tutta la variabilità presente nei diversi gruppi a confronto sia determinata dai differenti livelli o dalle varie modalità del solo fattore in osservazione.

Sovente è utile, quando non necessario, **prendere in considerazione almeno due fattori di variabilità,**

- **sia per analizzare gli effetti di due o più cause contemporaneamente,**
- **sia per ridurre la varianza d'errore isolando gli effetti dovuti ad altre cause note.**

Nell'analisi ad un criterio di classificazione, si è ricorsi all'esempio di differenti livelli d'inquinamento dell'aria in aree diverse della stessa città. E' ovvio che, tra gli altri fattori, anche l'ora della rilevazione può determinare differenze sensibili: durante la giornata, il traffico, il riscaldamento domestico e l'attività di molte industrie possono variare significativamente il livello d'inquinamento dell'aria nella stessa zona. Diventa importante effettuare esperimenti in cui, oltre alla località, si considera anche l'ora del campionamento.

Nello stesso modo, quando si confronta l'effetto di farmaci o di tossici sulla crescita di cavie, può essere importante la loro classe d'età: cavie giovani, adulte od anziane possono dare risposte sensibilmente diverse, alla somministrazione della stessa sostanza, determinando una forte variabilità nei dati sperimentali.

**Questa doppia classificazione di ogni misura può avere una duplice finalità:**

- **analizzare separatamente quale sia il contributo del fattore principale e quale quello del secondo fattore;**
- **eliminare l'effetto del secondo fattore sulla varianza d'errore, quando l'interesse fosse indirizzato solo verso il primo ed il secondo fosse considerato esclusivamente come un elemento di forte perturbazione.**

In questo caso, il metodo permette di **ridurre sensibilmente la varianza d'errore e quindi di aumentare la probabilità di trovare differenze significative tra le medie del fattore ritenuto più importante.**

Le diverse modalità del primo fattore, di solito quello ritenuto più importante, sono chiamate **trattamenti**; le diverse modalità del secondo fattore sono chiamate **blocchi**.

L'analisi della varianza a due criteri di classificazione, è chiamata anche **analisi a blocchi randomizzati**. In particolare negli esperimenti di laboratorio, con l'analisi completamente randomizzata si richiedeva che gli individui, preventivamente riuniti in un gruppo unico, fossero distribuiti a caso tra i vari trattamenti. Nell'analisi a blocchi randomizzati si richiede che gli individui siano dapprima suddivisi in gruppi omogenei, detti blocchi, per il secondo fattore; successivamente gli individui di ogni blocco devono essere attribuiti in modo casuale (random) ai trattamenti (di norma, il metodo richiede l'estrazione di numeri casuali, dopo aver numerato separatamente gli individui dei differenti blocchi).

Per esempio, analizzando contemporaneamente l'effetto di tossici su cavie d'età diversa, si richiede dapprima una loro suddivisione per classi d'età, formando gruppi composti esclusivamente da individui giovani, da individui adulti oppure anziani; successivamente, separatamente ed indipendentemente per ogni classe d'età, si devono attribuire casualmente gli individui ai trattamenti.

E' ovvio che, **se i trattamenti sono n, ogni blocco deve comprendere almeno n individui**. Quelli eccedenti, non estratti nell'attribuzione casuale per blocco, sono esclusi dall'esperimento.

Nell'analisi della varianza a due criteri di classificazione, è utile alla comprensione del metodo statistico riportare le osservazioni in una tabella a doppia entrata, come quella sottostante. Per convenzione ampiamente diffusa, quasi sempre **i trattamenti sono riportati in colonna ed i blocchi nelle righe**:

		TRATTAMENTI					
BLOCCHI	1	2	3	...	p	medie	
1	$X_{11}$	$X_{12}$	$X_{13}$	...	$X_{1p}$	$\bar{X}_1$	
2	$X_{21}$	$X_{22}$	$X_{23}$	...	$X_{2p}$	$\bar{X}_2$	
...	...	...	...	...	...	...	
k	$X_{k1}$	$X_{k2}$	$X_{k3}$	...	$X_{kp}$	$\bar{X}_k$	
medie	$\bar{X}_{.1}$	$\bar{X}_{.2}$	$\bar{X}_{.3}$	...	$\bar{X}_{.p}$	$\bar{\bar{X}}$	

Per indicare valori e medie, si ricorre ad una simbologia a due indici, analoga a quella illustrata per l'analisi della varianza ad un criterio di classificazione, con la sola differenza di dover considerare anche le medie di riga o dei blocchi.

Nell'analisi della varianza a due criteri, si possono avere due situazioni differenti che richiedono metodi parzialmente diversi. **In ogni casella data dall'intersezione tra un trattamento ed un blocco si può avere una sola misura oppure almeno due misure.**

Quando si hanno due o più repliche per ogni casella, è plausibile supporre che queste misure siano tra loro più simili: l'errore o residuo sarà misurato mediante esse, sulla base dello scarto di ogni replica dalla media di casella. Tale approccio verrà analizzato nei prossimi paragrafi, presentando l'**analisi dell'interazione con gli esperimenti fattoriali.**

Il caso più semplice, analizzato in questo paragrafo, è quello di **una sola osservazione ad ogni intersezione tra riga (blocco) e colonna (trattamento).**

Come presentato nel caso di un solo criterio, l'**analisi della varianza utilizza il modello lineare od additivo.** Con due criteri di classificazione ed una sola osservazione, ogni dato può essere rappresentato con

$$X_{pk} = \mu + \alpha_p + \beta_k + R_{pk}$$

dove :

$\mu$  è la **media generale**,

$\alpha_p$  è l'**effetto del trattamento** che, in riferimento alla popolazione, può essere indicato come differenza della media  $\mu_p$  del p-esimo trattamento dalla media generale  $\mu$

$$\alpha_p = \mu_p - \mu$$

$\beta_k$  è l'**effetto del blocco**, indicato dalla differenza tra la media  $\mu_k$  del k-esimo trattamento e la media generale  $\mu$

$$\beta_k = \mu_k - \mu$$

$R_{pk}$  è la **quota residua**, che ingloba

- sia tutti gli altri fattori non considerati e gli effetti di campionamento, già presentati nell'errore dell'analisi ad un criterio e simboleggiati complessivamente da  $\epsilon_{pk}$ ,
- sia l'interazione tra i due fattori che, con una sola osservazione per casella, non può essere stimata.

In riferimento ai **dati campionari** raccolti con un esperimento di laboratorio oppure in campagna, ogni singolo valore  $X_{pk}$  è stimato come somma di 4 valori

$$X_{pk} = \bar{\bar{X}} + (\bar{X}_p - \bar{\bar{X}}) + (\bar{X}_k - \bar{\bar{X}}) + R_{pk}$$

che dipendono rispettivamente

- dalla **media generale**  $\bar{\bar{X}}$ ,
- dall'**effetto del fattore A**, che per ogni trattamento è stimato mediante la differenza  $\bar{X}_p - \bar{\bar{X}}$ ,
- dall'**effetto del fattore B**, che per ogni blocco è stimata dalla differenza  $\bar{X}_k - \bar{\bar{X}}$ ,

- da **tutti gli altri fattori non considerati e dall'interazione tra i due fattori A e B**, compresi entrambi nella quantità  $R_{pk}$ .

**L'analisi della varianza a due criteri di classificazione con una sola osservazione per casella permette di verificare, in modo simultaneo ed indipendente, la significatività delle differenze tra le medie dei trattamenti, o fattore A, e tra le medie dei blocchi, o fattore B.**

Per esempio, indicando con a, b, c, d quattro trattamenti e con 1, 2, 3, 4, 5 cinque blocchi, tale concetto viene espresso mediante

- una prima ipotesi nulla  $H_0$  di uguaglianza delle medie dei trattamenti o del fattore A,

$$\text{per il fattore A} \quad H_0: m_a = m_b = m_c = m_d$$

con ipotesi alternativa

$$H_1 : \text{non tutte le medie dei 4 trattamenti sono tra loro uguali}$$

- una seconda ipotesi nulla di uguaglianza delle medie dei blocchi o del fattore B

$$\text{per il fattore B} \quad H_0: m_1 = m_2 = m_3 = m_4 = m_5$$

con ipotesi alternativa

$$H_1 : \text{non tutte le medie dei 5 blocchi sono tra loro uguali.}$$

Per questi confronti tra medie, la metodologia richiede il calcolo delle seguenti quantità:

- 1 - la **devianza totale**, con **gdl**  $p \times k - 1 = n - 1$ ;
- 2 - la **devianza tra trattamenti**, con **gdl**  $p - 1$ , e la rispettiva varianza;
- 3 - la **devianza tra blocchi**, con **gdl**  $k - 1$ , e la rispettiva varianza;
- 4 - la **devianza d'errore**, con **gdl**  $(p - 1) \times (k - 1) = (n - 1) - [(p - 1) + (k - 1)]$  e la sua varianza.

Le 4 devianze e i relativi gradi di libertà godono della proprietà additiva:

$$\text{Devianza}_{\text{totale}} = \text{Devianza}_{\text{tra trattamenti}} + \text{Devianza}_{\text{tra blocchi}} + \text{Devianza}_{\text{errore}}$$

$$\text{gdl}_{\text{totale}} = \text{gdl}_{\text{tra trattamenti}} + \text{gdl}_{\text{tra blocchi}} + \text{gdl}_{\text{errore}}$$

Nella presentazione dei risultati, queste quantità vengono abitualmente riportate per convenzione in una tabella come la seguente:

<b>Devianza totale</b>	<b>gdl: n - 1= pk - 1</b>	<b>----</b>
<b>Devianza tra trattamenti</b>	<b>gdl: p - 1</b>	<b>varianza tra trattamenti</b>
<b>Devianza tra blocchi</b>	<b>gdl: k - 1</b>	<b>varianza tra blocchi</b>
<b>Devianza d'errore</b>	<b>gdl: (p - 1)(k - 1)</b>	<b>varianza d'errore</b>

(nei programmi informatici, sovente la devianza totale ed i suoi gdl sono riportati come ultimi, per **evidenziare graficamente che rappresentano la somma delle altre tre quantità**).

La **devianza totale** misura la variazione totale tra le osservazioni; con la formula euristica e quella abbreviata è calcolata come

$$\sum_{j=1}^p \sum_{i=1}^k (X_{ij} - \bar{X})^2 = \sum_{j=1}^p \sum_{i=1}^k X_{ij}^2 - \frac{(\sum_{j=1}^p \sum_{i=1}^k X_{ij})^2}{n}$$

La **devianza tra trattamenti** misura la variazione tra le medie dei trattamenti; con formula euristica e quella abbreviata è data da

$$\sum_{j=1}^p (\bar{X}_{.j} - \bar{X})^2 \cdot n_j = \sum_{j=1}^p \frac{(\sum_{i=1}^p X_{.j})^2}{n_j} - \frac{(\sum_{i=1}^k \sum_{j=1}^p X_{ij})^2}{n}$$

La **devianza tra blocchi** misura la variazione tra le medie dei blocchi; la formula euristica e quella abbreviata sono

$$\sum_{i=1}^k (\bar{X}_{.i} - \bar{X})^2 \cdot n_i = \sum_{i=1}^k \frac{(\sum_{j=1}^p X_{.i})^2}{n_i} - \frac{(\sum_{i=1}^k \sum_{j=1}^p X_{ij})^2}{n}$$

La **devianza d'errore**, detta anche **residuo**, misura **la variazione che rimane di ogni osservazione dopo aver tolto gli effetti dei fattori già considerati**. Il procedimento diretto, corrispondente alla formula euristica, è lungo; per una presentazione semplice e chiara verrà illustrato con i dati dell'esempio. Abitualmente, con formula equivalente a quella abbreviata, **il calcolo viene effettuato ricorrendo alla proprietà additiva**

- delle devianze

$$\text{Devianza}_{\text{errore}} = \text{Devianza}_{\text{totale}} - (\text{Devianza}_{\text{tra trattamenti}} + \text{Devianza}_{\text{tra blocchi}})$$

- e dei gradi di libertà relativi

$$\text{gdl Devianza}_{\text{errore}} = \text{gdl Devianza}_{\text{totale}} - (\text{gdl Devianza}_{\text{tra trattamenti}} + \text{gdl Devianza}_{\text{tra blocchi}})$$

La varianza tra trattamenti, la varianza tra blocchi e la varianza d'errore o residuo si ottengono dividendo le devianze omonime per i rispettivi gdl.

Il test F consiste nel **confrontare, mediante il rapporto, sia la varianza tra trattamenti sia quella tra blocchi separatamente con la varianza d'errore.**

- **Per la significatività delle differenze tra le medie dei trattamenti**, si calcola il valore di F

$$F_{(p-1), (p-1) \times (k-1)} = \frac{\text{varianza tra trattamenti}}{\text{varianza d'errore}}$$

e si confronta il risultato calcolato con il valore tabulato alla probabilità  $\alpha$  prefissata per i gradi di libertà relativi. Si rifiuta l'ipotesi nulla relativa al fattore A, se il valore calcolato è maggiore di quello riportato nella tabella.

- **Per la significatività delle differenze tra le medie dei blocchi** si calcola un secondo valore di F

$$F_{(k-1), (p-1) \times (k-1)} = \frac{\text{varianza tra blocchi}}{\text{varianza d'errore}}$$

e si confronta il risultato con il valore tabulato per i suoi gdl con le stesse modalità. Si rifiuta l'ipotesi nulla relativa al fattore B, se il valore calcolato è superiore a quello riportato nella tabella.

**Le due analisi sono totalmente indipendenti.**

ESEMPIO. Si vuole verificare se esiste una differenza significativa nella quantità di piombo in sospensione nell'aria di 5 zone di una città (A, B, C, D, E).

BLOCCHI (ORE)	TRATTAMENTI (ZONE)				
	A	B	C	D	E
I	28	25	30	22	26
II	34	32	37	31	30
III	22	21	24	20	19
IV	36	31	40	33	29

Poiché si suppone che esista una forte variabilità determinata dall'ora di campionamento, è stata fatta una rilevazione simultanea in ogni zona, con ripetizioni a distanza di 6 ore (alle ore 6, 12, 18 e 24), per un totale di 4 campioni per zona.

I dati raccolti sono riportati nella tabella successiva.

Esiste una differenza significativa nella presenza media di polveri di piombo in sospensione nell'aria delle 5 zone? Le differenze tra le ore sono significative?

Risposta.

Le ipotesi sono relative

- sia alle medie dei trattamenti o zone, con ipotesi nulla

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$$

ed ipotesi alternativa

$$H_1: \text{non tutte le } \mu \text{ delle zone sono uguali}$$

- sia alle medie dei blocchi o ore, con ipotesi nulla

$$H_0: \mu_I = \mu_{II} = \mu_{III} = \mu_{IV}$$

ed ipotesi alternativa

$$H_1: \text{non tutte le } \mu \text{ delle ore sono uguali}$$

Dapprima si devono calcolare i totali di colonna, di riga, il totale generale e le medie relative,

	TRATTAMENTI (ZONE)						
BLOCCHI (ORE)	A	B	C	D	E	Totale ore	medie
I	28	25	30	22	26	131	26,2
II	34	32	37	31	30	164	32,8
III	22	21	24	20	19	106	21,2
IV	36	31	40	33	29	169	33,8
<i>Totale zone</i>	120	109	131	106	104	570	
<b>medie</b>	<b>30,00</b>	<b>27,25</b>	<b>32,75</b>	<b>26,50</b>	<b>26,00</b>		<b>28,50</b>

utili ai calcoli successivi per le devianze.

Da essi si stimano:

1- la **devianza totale**, con 19 gdl, ottenuta come scarto al quadrato di ogni valore dalla media generale

$$(28-28,5)^2 + (34-28,5)^2 + (22-28,5)^2 + \dots + (29-28,5)^2 = 683,0$$

oppure con la formula abbreviata

$$(28^2 + 34^2 + 22^2 + 36^2 + 25^2 + 32^2 + \dots + 29^2) - \frac{570^2}{20} = 683,0$$

2 - la **devianza tra trattamenti o tra zone**, con 4 gdl, ottenuta come scarto quadratico di ognuna delle 5 medie di colonna dalla media generale, moltiplicato per il numero di dati della colonna

$$4 \cdot (30,00 - 28,5)^2 + 4 \cdot (27,25 - 28,5)^2 + \dots + 4 \cdot (26,00 - 28,5)^2 = 128,5$$

oppure con la formula abbreviata

$$\frac{120^2}{4} + \frac{109^2}{4} + \frac{131^2}{4} + \frac{106^2}{4} + \frac{104^2}{4} - \frac{570^2}{20} = 128,5$$

3 - la **devianza tra blocchi o tra ore**, con 3 gdl, ottenuta come scarto quadratico di ognuna delle 4 medie di riga dalla media generale, moltiplicato per il numero di dati su cui è calcolata la media

$$5 \cdot (26,2 - 28,5)^2 + 5 \cdot (32,8 - 28,5)^2 + \dots + 5 \cdot (33,8 - 28,5)^2 = 525,8$$

oppure con la formula abbreviata

$$\frac{131^2}{5} + \frac{164^2}{5} + \frac{106^2}{5} + \frac{169^2}{5} - \frac{570^2}{20} = 525,8$$

(La quantità  $\frac{(SX)^2}{n}$ , che nell'esercizio è  $\frac{570^2}{20}$  compare sia nel calcolo della devianza totale che in quello della devianza tra trattamenti e tra blocchi, è **chiamata termine di correzione**, in alcuni testi abbreviato in TC).

4 - la **devianza d'errore** e i suoi gdl possono essere calcolati in modo rapido per differenza

$$\text{devianza d'errore} = 683,0 - 128,5 - 525,8 = 28,7$$

$$\text{gdl della devianza d'errore} = 19 - 4 - 3 = 12$$

Per una presentazione sintetica dei dati raccolti, per una rapida verifica dei calcoli e per la successiva stima delle tre varianze necessarie ai due test F, con gli 8 valori stimati (4 devianze e relativi gdl) è utile costruire la tabella:

	DEVIANZA	GDL	VARIANZA
<b>Totale</b>	683,0	19	----
<b>Tra trattamenti (zone)</b>	128,5	4	<b>32,125</b>
<b>Tra blocchi (ore)</b>	525,8	3	<b>175,266</b>
<b>Errore</b>	28,7	12	<b>2,39</b>

La **significatività della differenza tra zone** è verificata con

$$F_{4,12} = \frac{32,125}{2,39} = 13,44$$

un test F che, con gdl 4 e 12, risulta uguale a 13,44,

mentre **la significatività delle differenze tra ore** è verificata con

$$F_{3,12} = \frac{175,266}{2,39} = 73,33$$

un test F che, con gdl 3 e 12, risulta uguale a 73,33.

I valori critici corrispondenti

- alla probabilità  $\alpha = 0.05$  per  $F_{4,12}$  è uguale a 3,26 e per  $F_{3,12}$  è 3,49
- alla probabilità  $\alpha = 0.01$  per  $F_{4,12}$  è uguale 5,41 e per  $F_{3,12}$  è 5,95.

Con probabilità  $\alpha$  inferiore a 0.01 si rifiuta l'ipotesi nulla, sia per le medie delle zone che per le medie delle ore. La differenza tra ore risulta statisticamente più significativa di quella tra zone.

**La devianza d'errore è stata calcolata per differenza**, sottraendo alla devianza totale quella tra trattamenti e quella tra blocchi. **Per comprenderne più esattamente il significato**, è proficuo vedere quanto del valore di ogni osservazione è imputabile agli effetti congiunti della media generale, del fattore A e del fattore B (considerati nelle devianze relative) e quanto ai rimanenti fattori raggruppati nel residuo o devianza d'errore.

Con i primi 3 fattori, per ogni valore  $X_{pk}$  osservato è possibile calcolare un valore  $\hat{X}_{pk}$  atteso, definito come

$$\hat{X}_{pk} = \bar{\bar{X}} + (\bar{X}_p - \bar{\bar{X}}) + (\bar{X}_k - \bar{\bar{X}})$$

Dopo semplificazione, risulta che può essere stimato mediante

$$\hat{X}_{pk} = \bar{X}_p + \bar{X}_k - \bar{\bar{X}}$$

la somma della media di riga e della media di colonna, alla quale viene sottratta la media generale.

Con i dati dell'esercizio, dopo aver calcolato le medie marginali e quella totale, è possibile stimare in ogni casella, all'intersezione tra ogni riga e ogni colonna, quale è il valore atteso qualora agissero solamente i tre effetti considerati.

La tabella sottostante riporta questi valori attesi  $\hat{X}_{pk}$ :

		TRATTAMENTI					
BLOCCHI	A	B	C	D	E	Medie	
I	27,70	24,95	30,45	24,20	23,70	26,20	
II	34,30	31,55	37,05	30,80	30,30	32,80	
III	22,70	19,95	25,45	19,20	18,70	21,20	
IV	35,30	32,55	38,05	31,80	31,30	33,80	
Medie	30,00	27,25	32,75	26,50	26,00	28,50	

Se si utilizzasse questa distribuzione per calcolare le 4 devianze, otterremmo valori identici a quelli dell'esempio precedente per la devianza totale, per quella tra trattamenti e per quella tra blocchi; ma la devianza d'errore o residuo risulterebbe uguale a 0.

La **devianza d'errore** calcolata precedentemente è

$$\text{Devianza d'errore} = \sum_{i=1}^p \sum_{j=1}^k (\hat{X}_{pk} - X_{pk})^2$$

data dalla **somma dei quadrati degli scarti tra questi valori stimati e quelli reali o osservati.**

## 9.2. CONFRONTO TRA ANALISI DELLA VARIANZA A DUE CRITERI

### E TEST t DI STUDENT PER 2 CAMPIONI DIPENDENTI

**Quando si confrontano le medie di due soli trattamenti applicati agli stessi soggetti**, quindi con valori riportati in una tabella con 2 trattamenti ed n blocchi, **l'analisi della varianza a due criteri di classificazione fornisce i medesimi risultati del test t di Student per 2 campioni dipendenti**, in riferimento al fattore principale.

Un esempio permette di confrontare in modo didatticamente semplice e quindi di comprendere sia le due metodologie sia i risultati relativi.

Durante una giornata lavorativa (Giorno I), in una stazione di rilevamento sono state misurate le quantità di inquinamento dell'aria in 4 ore differenti; durante il successivo giorno festivo (Giorno II), sono state ripetute le misure alle stesse ore.

La tabella riporta i valori rilevati nei due giorni alla stessa ora:

	<b>Giorno I</b>	<b>Giorno II</b>
<b>Ore 6</b>	150	120
<b>Ore 10</b>	172	151
<b>Ore 14</b>	193	165
<b>Ore 18</b>	175	150

Esiste una differenza significativa nel tasso medio d'inquinamento tra i due giorni?

Risposta.

Con i dati del quesito, è possibile **utilizzare il test t di Student per 2 campioni dipendenti** per verificare l'ipotesi nulla

$$H_0: \delta = 0$$

con l'ipotesi alternativa bilaterale

$$H_1: \delta \neq 0$$

Il metodo del test t di Student per 2 campioni dipendenti richiede che dapprima si calcoli la colonna delle differenze,

	Giorno I	Giorno II	<b>Differenze</b>
Ore 6	150	120	<b>30</b>
Ore 10	172	151	<b>21</b>
Ore 14	193	165	<b>28</b>
Ore 18	175	150	<b>25</b>

sulle quali applicare le operazioni richieste dalla formula

$$t_{(n-1)} = \frac{\bar{d} - d}{\frac{s_d}{\sqrt{n}}}$$

Per una corretta comprensione del metodo, è importante evidenziare due conseguenze del ricorso alle differenze:

- **i valori sono resi indipendenti dalle dimensioni dei dati originali**, che infatti vengono ignorati;
- **la varianza delle differenze**, che possono essere viste come somma delle variazioni casuali delle coppie di dati quando l'ipotesi nulla è vera, **è uguale alla somma delle varianze delle due serie di dati originali**.

Dopo aver calcolato la media delle differenze

$$\bar{d} = \frac{(30 + 21 + 28 + 25)}{4} = 26$$

si stima la devianza

$$(30-26)^2 + (21-26)^2 + (28-26)^2 + (25-26)^2 = 16 + 25 + 4 + 1 = 46$$

e da essa la varianza delle differenze

$$s_d^2 = \frac{46}{3} = 15,33$$

e la loro deviazione standard

$$s_d = \sqrt{15,33} = 3,916$$

Il valore di t con 3 gdl

$$t_{(3)} = \frac{26}{\frac{3,916}{\sqrt{4}}} = 13,28$$

risulta uguale a 13,28.

Il risultato permette di rifiutare l'ipotesi nulla con probabilità inferiore a 0.01 (il cui valore critico è 5,841).

**Con l'analisi della varianza a due criteri di classificazione**, le ipotesi considerano

- sia il confronto tra le medie dei due giorni con ipotesi nulla

$$H_0: \mu_I = \mu_{II}$$

ed ipotesi alternativa bilaterale

$$H_1: \mu_I \neq \mu_{II}$$

- sia quello tra le medie delle 4 rilevazioni orarie con ipotesi nulla

$$H_0: \mu_6 = \mu_{10} = \mu_{14} = \mu_{18}$$

ed ipotesi alternativa

$H_1$ : non tutte le 4  $\mu$  dei blocchi sono tra loro uguali.

Dalle due serie di dati appaiati si calcolano i totali e le medie, sia quelle parziali che quella generale.

	Giorno I	Giorno II	<b>Totali</b>	<b>Medie</b>
Ore 6	150	120	<b>270</b>	<b>135,0</b>
Ore 10	172	151	<b>323</b>	<b>161,5</b>
Ore 14	193	165	<b>358</b>	<b>179,0</b>
Ore 18	175	150	<b>325</b>	<b>162,5</b>
<b>Totali</b>	<b>690</b>	<b>586</b>	<b>1276</b>	----
<b>Medie</b>	<b>172,5</b>	<b>146,5</b>	----	<b>159,5</b>

La **devianza totale con 7 gdl**, calcolata come somma degli scarti al quadrato di ogni valore rispetto alla media totale,

$$(-9,5)^2 + (-39,5)^2 + (12,5)^2 + (-8,5)^2 + (33,5)^2 + (5,5)^2 + (15,5)^2 + (-9,5)^2 =$$

$$= 90,25 + 1560,25 + 156,25 + 72,25 + 1122,25 + 30,25 + 240,25 + 90,25 = 3362$$

risulta uguale a 3362.

La **devianza tra giorni ha 1 gdl** ed è

$$4(172,5 - 159,5)^2 + 4(146,5 - 159,5)^2 = 4 \times 169 + 4 \times 169 = 1352$$

uguale a 1352.

La **devianza tra ore ha 3 gdl** e risulta

$$2 \cdot (135 - 159,5)^2 + 2 \cdot (161,5 - 159,5)^2 + 2 \cdot (179 - 159,5)^2 + 2 \cdot (162,5 - 159,5)^2 =$$

$$= 2 \cdot 600,25 + 2 \cdot 4 + 2 \cdot 380,25 + 2 \cdot 9 = 1200,50 + 8 + 760,5 + 18 = 1987$$

uguale a 1987.

La **devianza d'errore con 3 gdl** [7 - (1+3)] ed ottenuta per differenza

$$3362 - 1352 - 1987 = 23$$

risulta uguale a 23.

Devianze e gdl possono essere riportati in tabella

	DEVIANZA	GDL	VARIANZA
Totale	3362	7	----
Tra giorni	1352	1	<b>1352</b>
Tra ore	1987	3	<b>662,333</b>
Errore	23	3	<b>7,666</b>

e da essi sono stimate le 3 varianze necessarie ai due test F.

Per la differenza tra giorni si calcola un test F

$$F_{(1,3)} = \frac{1352}{7,666} = 176,36$$

con gdl 1 e 3, che risulta uguale a 176,36.

Il test permette di rifiutare l'ipotesi nulla relativa ai due giorni, con probabilità  $\alpha$  inferiore a 0.01 (il cui valore critico è 34,12).

Per verificare la differenza tra ore si calcola un altro test F

$$F_{(3,3)} = \frac{662,33}{7,666} = 86,40$$

con gdl 3 e 3, che risulta uguale a 86,40.

Anche in questo caso si rifiuta l'ipotesi nulla, con probabilità  $\alpha$  inferiore a 0.01 (valore critico uguale a 29,46).

Il confronto effettuato serve soprattutto per dimostrare che con 2 campioni **test t e test F non solo permettono la stessa inferenza, ma forniscono anche risultati identici.**

**Il valore di F che permette il confronto tra le medie dei due giorni è identico al valore del t che verifica l'uguaglianza delle due medie giornaliere; ovviamente, coincidono anche i gdl che per il test F sono 1 e 3 e per il test t sono 3.**

$$t_{(3)}^2 = F_{(1,3)} : (13,28)^2 = 176,36$$

Corrispondono anche i valori critici per la stessa probabilità. Infatti, come dimostrazione elementare, è possibile verificare che per la probabilità  $\alpha = 0.01$ , i cui valori critici sono rispettivamente

- 5,841 per il test t e
- 34,12 per il test F

tra essi esiste la stessa relazione

$$t_{(3)}^2 = F_{(1,3)} : (5,841)^2 = 34,12$$

### 9.3. ANALISI DELLA VARIANZA A TRE O PIU' CRITERI

**I concetti, le finalità ed i metodi dell'analisi della varianza a due criteri possono essere facilmente estesi a tre o più criteri di classificazione.** Con **k fattori**, è possibile verificare **k ipotesi nulle**, con altrettante ipotesi alternative bilaterali; a questo scopo occorre calcolare le **k varianze**, per effettuare **k test F** mediante il loro rapporto con la varianza d'errore.

A dimostrazione ed esemplificazione didattica, per facilitare la comprensione dell'analisi della varianza a più criteri di classificazione, vengono di seguito presentati, analizzati e discussi i risultati di un esperimento di genetica ambientale applicata a un problema d'inquinamento, in cui sono stati vagliati contemporaneamente **4 fattori**.

Per ogni singola osservazione  $X_{ijkp}$ , il modello di riferimento diviene

$$X_{ijkp} = m + a_i + b_j + g_k + d_p + R_{ijkp}$$

dove

- **m** rappresenta la media generale,
- **a<sub>i</sub>, b<sub>j</sub>, g<sub>k</sub>, d<sub>p</sub>** rappresentano gli effetti indipendenti dei 4 fattori, nelle loro varie modalità,
- **R<sub>ijkp</sub>** rappresenta tutti gli altri fattori non considerati, sommati alla variazione casuale e agli errori di misura e campionamento, oltre a tutte le possibili interazioni tra i 4 fattori.

Per ognuno dei quattro fattori, si può verificare l'ipotesi nulla

$$H_0: \text{tutte le } m \text{ delle varie modalità dello stesso fattore sono uguali}$$

con l'ipotesi alternativa

$$H_1: \text{non tutte le } m \text{ delle varie modalità dello stesso fattore sono uguali}$$

Con **n** osservazioni in ogni modalità, **s** modalità per ogni fattore ed **N** osservazioni in totale ( $N = n \cdot s$ ), la devianza di ogni fattore può essere stimata con la formula euristica o la più rapida formula abbreviata di seguito riportate

$$\sum_{j=1}^s (\bar{X}_{.j} - \bar{X})^2 \cdot n_j = \sum_{i=1}^s \frac{\left( \sum_{j=1}^n X_{.j} \right)^2}{n_i} - \frac{\left( \sum_{i=1}^s \sum_{j=1}^n X_{ij} \right)^2}{N}$$

ESEMPIO. Si intende verificare se da parte di **4 essenze** (medica, plantago, trifoglio, tarassaco), cresciute in **2 località** ad alto inquinamento ( $T_1$  e  $T_2$ ) **più un controllo** a livelli normali ( C ), esistono differenze nell'assorbimento di **4 metalli** (cadmio, nichel, piombo e zinco), considerando anche i **5 sfalci** che vengono praticati durante un anno.

Nella pagina successiva sono riportati in tabella i valori unitari per ognuna delle 240 (4 x 3 x 4 x 5) misure campionarie effettuate.

Esistono differenze significative tra le medie di ognuno dei 4 fattori considerati?

Risposta.

Per rispondere alle 4 ipotesi nulle che si possono formulare sui 4 fattori, applicando le formule presentate si devono calcolare le devianze, con i rispettivi gdl e le varianze utili ai 4 test F.

I risultati dei calcoli sono riportati nella tabella sottostante, che segue lo schema applicato dai programmi informatici più diffusi.

	DEVIANZA	GDL	VARIANZA
<b>Totale</b>	<b>149765998.1</b>	<b>239</b>	----
<b>Tra metalli</b>	<b>44775228.1</b>	<b>3</b>	<b>14925076.0</b>
<b>Tra località</b>	<b>648842.9</b>	<b>2</b>	<b>324421.5</b>
<b>Tra essenze</b>	<b>1801355.4</b>	<b>3</b>	<b>600451.8</b>
<b>Tra sfalci</b>	<b>7719081.8</b>	<b>4</b>	<b>1929770.5</b>
<b>Errore</b>	<b>94821489.8</b>	<b>227</b>	<b>417715.8</b>

In realtà (come verrà spiegato successivamente nel paragrafo dedicato alla trasformazione dei dati) **l'analisi della varianza non può essere applicata direttamente ai valori riportati nella tabella e il risultato precedente è errato.** E' semplice osservare come **la distribuzione non** sia **normale** ed esistano **grandissime differenze tra le varianze**, data la variabilità presente tra metalli (lo zinco ha valori nettamente superiore agli altri metalli), tra erbe o essenze (come all'interno del nichel), tra sfalci (come per l'erba medica).

L'analisi è effettivamente corretta solamente dopo la **trasformazione di ogni dato (x) nel suo logaritmo naturale** o  $\log_e(x)$ .

METALLO	CONDIZIONE	ESSENZA	S F A L C I O				
			1	2	3	4	5
CADMIO	T1	MEDICA	3.00	2.40	3.00	4.20	2.00
		PLANTAGO	4.00	2.85	2.55	3.45	2.10
		TRIFOGLIO	3.00	2.32	1.65	2.85	1.50
		TARASSACO	3.00	1.95	4.50	2.70	2.00
	T2	MEDICA	3.00	1.95	4.50	2.70	2.00
		PLANTAGO	4.00	2.32	2.77	4.80	2.40
		TRIFOGLIO	3.00	2.85	3.07	3.60	1.25
		TARASSACO	2.00	2.40	2.67	4.50	2.00
	C	MEDICA	2.00	2.40	2.67	4.50	2.00
		PLANTAGO	3.00	2.32	2.40	3.45	1.70
		TRIFOGLIO	3.00	1.57	1.80	2.70	2.10
		TARASSACO	2.00	4.20	3.15	3.90	2.00
NICHEL	T1	MEDICA	20.00	9.00	4.50	18.00	8.00
		PLANTAGO	10.00	5.25	3.75	9.00	5.80
		TRIFOGLIO	10.00	4.50	3.75	9.00	4.50
		TARASSACO	10.00	10.50	15.00	30.00	11.00
	T2	MEDICA	10.00	10.50	15.00	30.00	11.00
		PLANTAGO	20.00	6.75	5.25	16.50	7.20
		TRIFOGLIO	10.00	4.50	1.75	9.00	5.50
		TARASSACO	20.00	13.50	6.00	24.00	11.00
	C	MEDICA	20.00	13.50	6.00	24.00	11.00
		PLANTAGO	20.00	6.00	3.00	6.00	4.40
		TRIFOGLIO	10.00	5.25	1.50	6.00	3.70
		TARASSACO	10.00	16.50	6.00	12.00	9.00
PIOMBO	T1	MEDICA	1.70	10.80	28.50	1.50	30.00
		PLANTAGO	7.00	1.50	15.00	2.40	21.00
		TRIFOGLIO	15.00	8.25	15.00	.90	20.00
		TARASSACO	3.00	5.70	60.00	3.60	40.00
	T2	MEDICA	3.00	5.70	60.00	3.60	40.00
		PLANTAGO	4.00	2.40	13.50	3.60	24.00
		TRIFOGLIO	3.00	3.60	17.50	2.40	16.00
		TARASSACO	2.00	9.00	30.00	1.50	30.00
	C	MEDICA	2.00	9.00	30.00	1.50	30.00
		PLANTAGO	4.00	2.85	22.50	1.20	22.00
		TRIFOGLIO	2.00	5.50	15.00	.90	10.50
		TARASSACO	2.00	11.00	28.50	2.70	30.00
ZINCO	T1	MEDICA	1500.00	330.00	360.00	760.00	280.00
		PLANTAGO	1650.00	325.00	360.00	530.00	310.00
		TRIFOGLIO	1750.00	385.00	200.00	805.00	210.00
		TARASSACO	1650.00	380.00	6900.00	440.00	280.00
	T2	MEDICA	1650.00	380.00	6900.00	440.00	280.00
		PLANTAGO	1550.00	590.00	630.00	935.00	240.00
		TRIFOGLIO	1550.00	520.00	1925.00	630.00	170.00
		TARASSACO	1550.00	800.00	3700.00	860.00	200.00
	C	MEDICA	1550.00	800.00	3700.00	860.00	200.00
		PLANTAGO	1400.00	215.00	260.00	415.00	180.00
		TRIFOGLIO	1300.00	200.00	180.00	435.00	180.00
		TARASSACO	1550.00	740.00	350.00	720.00	240.00

Dopo la trasformazione appena citata, la tabella dell'analisi della varianza risulta

	DEVIANZA	GDL	VARIANZA
<b>Totale</b>	<b>1213.70</b>	<b>239</b>	<b>----</b>
<b>Tra metalli</b>	<b>1049.12</b>	<b>3</b>	<b>349.71</b>
<b>Tra località</b>	<b>2.08</b>	<b>2</b>	<b>1.04</b>
<b>Tra essenze</b>	<b>10.44</b>	<b>3</b>	<b>3.48</b>
<b>Tra sfalci</b>	<b>7.90</b>	<b>4</b>	<b>1.98</b>
<b>Errore</b>	<b>144.15</b>	<b>227</b>	<b>.635</b>

- Per valutare la significatività della **differenza tra le medie dei 4 metalli** contenute nelle 4 essenze, si calcola un rapporto F con gdl 3 e 227

$$F_{(3,227)} = 349,71 / 0,635 = 550,7$$

che risulta uguale a 550,7.

- Per la differenza tra **le medie delle 3 località**,

si calcola un rapporto F con gdl 2 e 227

$$F_{(2,227)} = 1,04 / 0,635 = 1,64$$

che risulta uguale a 1,64.

- Per la differenza tra **le medie delle 4 essenze**,

si calcola un rapporto F con gdl 3 e 227

$$F_{(3,227)} = 3,48 / 0,635 = 5,48$$

che risulta uguale 5,48.

- Per la differenza tra **le medie dei 5 sfalci**,

si calcola un rapporto F con gdl 4 e 227

$$F_{(4,227)} = 1,98 / 0,635 = 3,12$$

che risulta uguale a 3,12.

Un modo empirico ma corretto di **valutare l'adeguatezza della trasformazione è dato dal rapporto F**, ottenuto per ogni fattore prima e dopo la trasformazione: è migliore la trasformazione che rende complessivamente più significativi i vari test F.

Nelle due tabelle precedenti si può osservare che

**- tra metalli**

nella prima tabella **F** (14925076 / 417715) risulta uguale a 35,7 contro 550,7 della seconda;

**- tra località**

nella prima tabella **F** (324421 / 417715) risulta uguale a 0,777 contro 1,64 della seconda;

**- tra essenze**

nella prima tabella **F** (600451 / 417715) risulta uguale a 1,437 contro 5,48 della seconda;

**- tra sfalci**

nella prima tabella **F** (1929770 / 417715) risulta uguale a 4,62 contro 3,12 della seconda;

E' evidente la maggiore capacità della trasformazione  $\log_e(x)$  di **ridurre soprattutto la varianza d'errore e quindi di rendere i test F complessivamente molto più significativi** (si ha solo una leggera riduzione di F per gli sfalci, contro un aumento grande di quello tra metalli e tra essenze).

I calcoli necessari all'esercizio proposto sono lunghi e la probabilità di commettere errori effettuandoli manualmente è alta. E' possibile trovare rapidamente la soluzione con probabilità molto inferiori di commettere errori, ricorrendo a programmi informatici. Essi forniscono tutte le informazioni necessarie all'analisi statistica, tra cui il valore di F e la probabilità relativa, permessa dall'ampia memoria dei calcolatori attuali; essi permettono di superare il limite determinato dal valore approssimato, che si ottiene con il ricorso alle tabelle dei valori critici.

Con questa maggiore precisione nelle stima delle probabilità, l'interpretazione può essere più sofisticata.

	DEVIANZA	GDL	VARIANZA	F	PROB.
Totale	<b>1213.70</b>	<b>239</b>	----	----	----
<b>Tra metalli</b>	<b>1049.12</b>	<b>3</b>	<b>349.71</b>	<b>550.7</b>	<b>.000</b>
<b>Tra località</b>	<b>2.08</b>	<b>2</b>	<b>1.04</b>	<b>1.64</b>	<b>.196</b>
<b>Tra essenze</b>	<b>10.44</b>	<b>3</b>	<b>3.48</b>	<b>5,48</b>	<b>.001</b>
<b>Tra sfalci</b>	<b>7.90</b>	<b>4</b>	<b>1.98</b>	<b>3,12</b>	<b>.016</b>
Errore	<b>144.15</b>	<b>227</b>	<b>.635</b>	----	----

Tra i 4 fattori considerati, due risultano altamente significativi (metalli ed essenze) con probabilità inferiori a 0.001, uno (sfalci) risulta significativo con probabilità leggermente inferiore al 2% e il quarto (località) non significativo, avendo una probabilità prossima al 20%.

Insieme con i risultati dell'analisi della varianza, i programmi informatici più sofisticati forniscono anche quelli dei confronti multipli, con vari test relativi ai fattori considerati. Compete al ricercatore

scegliere le sole risposte utili al problema. Con i dati dell'esempio, sulla base dell'analisi della varianza che ha permesso di rifiutare l'ipotesi nulla, è interessante verificare tra quali medie dei metalli, delle essenze e degli sfalci la differenza sia significativa: occorre effettuare confronti singoli, per i quali è appropriato il test di Tukey.

Nel caso dei **metalli**, ricordando che si devono effettuare tutti i possibili confronti tra 4 gruppi, nei programmi informatici viene riportato che,

- alla probabilità  $\alpha$  uguale a 0.05, con una varianza d'errore uguale a 0,635 e 60 repliche ( $n_i$ ) per metallo (dati dal problema),
- il valore del  $q$  studentizzato (Critical Value of Studentized Range) è uguale a 3,660 (fornito dalle tabelle per  $p = 4$  e  $gdl = 227$ )
- applicando la formula già riportata nella presentazione dei confronti multipli a posteriori,

$$T = q(\alpha, p, n_e) \cdot \sqrt{\frac{S_e^2}{n_i}}$$

si ottiene

$$T = 3,660 \cdot \sqrt{\frac{0,635}{60}} = 0,3765$$

- la Differenza Minima Significativa (Minimum Significant Difference), o valore del  $T$ , che risulta uguale a 0,3765.

I risultati sono riportati sinteticamente in modo grafico nella tabella

METALLO	Media	$n_i$	Gruppi di Tukey
Zinco	6,317	60	A
Nichel	2,134	60	B
Piombo	2,033	60	B
Cadmio	0,967	60	C

nella quale si evidenzia che alla probabilità  $\alpha = 0.05$

- 1) lo Zinco (gruppo A) ha una media significativamente diversa dagli altri 3 metalli,
- 2) il Nichel ed il Piombo (gruppo B) hanno medie tra loro uguali ma diverse da quelle degli altri 2,
- 3) il Cadmio (gruppo C) ha una media diversa dagli altri 3 metalli.

I confronti multipli a posteriori devono essere utilizzati solamente quando l'analisi della varianza permette di rifiutare l'ipotesi nulla alla probabilità prefissata, al fine di non evidenziare con essi differenze significative che non esistono. Tuttavia quasi sempre le conclusioni coincidono.

#### 9.4. QUADRATI LATINI E GRECO-LATINI

Analizzare contemporaneamente **2 fattori di variazione a p livelli** nel disegno sperimentale a blocchi randomizzati richiede  $p^2$  osservazioni. Con le stesse modalità di programmazione, **poiché ogni livello di un fattore deve incrociare tutti i livelli degli altri fattori, un esperimento con 3 fattori di variazione a p livelli richiede  $p^3$  osservazioni o repliche**. In un esperimento con 3 fattori, ognuno a 5 livelli, si richiedono 125 dati. All'aumentare dei fattori, si ha un rapido incremento delle misure che occorre raccogliere; poiché ognuna ha un costo e richiede tempo, sono stati sviluppati metodi che permettono di analizzare contemporaneamente più fattori con un numero minore di dati.

**Il disegno sperimentale a quadrati latini permette di analizzare contemporaneamente 3 fattori a p livelli con sole  $p^2$  osservazioni**: con 3 fattori a 5 livelli sono sufficienti 25 dati. A questo vantaggio, rappresentato da un risparmio di materiale e quindi di denaro e di tempo necessari all'esperimento, si associa lo svantaggio di una maggiore rigidità dell'esperimento stesso: **tutti e tre i fattori devono avere lo stesso numero di livelli**.

I quadrati latini furono applicati per la prima volta in esperimenti di agraria. La suddivisione in strisce tra loro perpendicolari e della stessa ampiezza di un appezzamento quadrato di terreno formava tante celle quadrate; da qui il nome, per la somiglianza del frazionamento dell'area all'accampamento romano.

**Per riportare in tabella i risultati di un esperimento a quadrati latini, due fattori vengono rappresentati nelle righe e nelle colonne, mentre il terzo, di solito il fattore principale, è rappresentato nelle celle formate dall'incrocio tra riga e colonna.**

**In esse, il terzo fattore è distribuito in modo casuale ma ordinato: deve comparire una volta sola sia in ogni riga che in ogni colonna.**

Per esempio, indicando nelle colonne i 4 livelli (I, II, III, IV) del primo fattore, nelle righe i 4 livelli del secondo fattore (1, 2, 3, 4) e con **A, B, C, D** i 4 livelli del trattamento principale o terzo fattore, la rappresentazione tabellare dell'esperimento a quadrati latini 4 x 4 diventa:

	COLONNE			
RIGHE	I	II	III	IV
1	D	B	C	A
2	C	D	A	B
3	B	A	D	C
4	A	C	B	D

In un disegno a due criteri di classificazione, la randomizzazione è attuata assegnando i trattamenti a caso all'interno di ciascun blocco; **in un quadrato latino, la randomizzazione è ottenuta permutando i trattamenti nello schema ordinato delle righe e delle colonne.** A questo scopo esistono tabelle di distribuzione casuale, da utilizzare nel caso di più esperimenti a quadrati latini con schemi differenti.

Il limite più grave a questo modo di programmazione dell'esperimento dipende dalla sua rigidità: se si vogliono analizzare 5 trattamenti serve un numero uguale di livelli anche negli altri 2 fattori da considerare nelle righe e nelle colonne. E' ovvio che **non tutti i fattori si prestano**; per esempio, non è possibile valutare l'effetto di 4 tossici su cavie di 4 età diversa (neonati, giovani, adulti, anziani) considerando anche il sesso, che ha due sole modalità.

In un disegno sperimentale a quadrati latini, **il modello additivo dell'analisi della varianza** richiede che la generica osservazione  $X_{ijk}$ , appartenente alla riga i-esima, alla colonna j-esima e al trattamento k-esimo, sia data da

$$X_{ijk} = \mu + a_i + b_j + g_k + R_{ijk}$$

dove:

$\mu$  è la media generale della popolazione, corrispondente a  $\bar{\bar{X}}$  con i dati campionari,

$\alpha_i$  è l'effetto medio della riga i-esima, sperimentalmente uguale a  $(\bar{X}_i - \bar{\bar{X}})$

$\beta_j$  è l'effetto medio della colonna j-esima, uguale a  $(\bar{X}_j - \bar{\bar{X}})$

$\gamma_k$  è l'effetto medio del trattamento k-esimo, uguale a  $(\bar{X}_k - \bar{\bar{X}})$

$R_{ijk}$  è la variabilità residua ed è uguale a  $(X_{ijk} - \bar{X}_i - \bar{X}_j - \bar{X}_k + 2\bar{\bar{X}})$

Il calcolo delle devianze è semplice: la devianza totale, la devianza tra righe e quella tra colonne sono calcolate con la stessa metodologia utilizzata nel disegno a blocchi randomizzati; la devianza tra

trattamenti viene calcolata rispetto alla somma e alla media ottenute dall'addizione dei valori collocati nelle celle contrassegnate con la stessa lettera. In simboli, ogni valore sperimentale  $X_{ijk}$  è dato da

$$X_{ijk} = \bar{\bar{X}} + (\bar{X}_i - \bar{\bar{X}}) + (\bar{X}_j - \bar{\bar{X}}) + (\bar{X}_k - \bar{\bar{X}}) + (X_{ijk} - \bar{X}_i - \bar{X}_j - \bar{X}_k + 2\bar{\bar{X}})$$

corrispondenti nel modello a

$$X_{ijk} = m + a + b + g + R_{ijk}$$

ESEMPIO. Si intende confrontare la produttività di 5 varietà (A, B, C, D, E) di sementi in rapporto al tipo di concime (1, 2, 3, 4, 5) e ad un diverso trattamento del terreno (I, II, III, IV, V). A questo scopo, si è diviso un appezzamento quadrato di terreno in 5 strisce di dimensioni uguali, nelle quali è stata fatta un'aratura di profondità differente; perpendicolarmente a queste, sono state tracciate altre 5 strisce, che sono state concimate in modo diverso.

Nei 25 quadrati, sono state seminate le 5 varietà di sementi, secondo lo schema a quadrati latini della tabella sottostante. In essa è riportata anche la produttività delle 5 varietà di sementi, nei 25 appezzamenti quadrati di terreno:

	TIPO DI TRATTAMENTO DEL TERRENO				
CONCIME	I	II	III	IV	V
1	A 42	C 47	B 55	D 51	E 44
2	E 45	B 54	C 52	A 44	D 50
3	C 41	A 46	D 57	E 47	B 48
4	B 56	D 52	E 49	C 50	A 43
5	D 47	E 49	A 45	B 54	C 46

Esistono differenze significative tra le medie delle 5 modalità, per ognuno dei 3 fattori considerati?

Risposta.

E' possibile formulare 3 ipotesi nulle, con le rispettive ipotesi alternative.

Per il tipo di **trattamento del terreno od aratura**:

$$H_0: \mu_I = \mu_{II} = \mu_{III} = \mu_{IV} = \mu_V \quad \text{con} \quad H_1: \text{non tutte le } \mu \text{ delle arature sono tra loro uguali.}$$

Per il tipo di **concime**:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 \quad \text{con} \quad H_1: \text{non tutte le } \mu \text{ dei concimi sono tra loro uguali.}$$

Per il tipo di **semente**:

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D = \mu_E \quad \text{con} \quad H_1: \text{non tutte le } \mu \text{ delle sementi sono tra loro uguali.}$$

Calcolate le somme e le medie sia per riga che per colonna dei primi due fattori, cioè del tipo di trattamento del terreno e del tipo di concime utilizzato,

TIPO DI TRATTAMENTO DEL TERRENO							
CONCIME	I	II	III	IV	V	Totale	Media
1	A 42	C 47	B 55	D 51	E 44	239	47,8
2	E 45	B 54	C 52	A 44	D 50	245	49,0
3	C 41	A 46	D 57	E 47	B 48	239	47,8
4	B 56	D 52	E 49	C 50	A 43	250	50,0
5	D 47	E 49	A 45	B 54	C 46	241	48,2
<b>Totale</b>	<b>231</b>	<b>248</b>	<b>258</b>	<b>246</b>	<b>231</b>	<b>1214</b>	
<b>Media</b>	<b>46,2</b>	<b>49,6</b>	<b>51,6</b>	<b>49,2</b>	<b>46,2</b>		<b>48,56</b>

si calcolano anche le somme e le medie del terzo fattore (riportate sempre a parte)

	A	B	C	D	E
<b>Totale</b>	<b>220</b>	<b>267</b>	<b>236</b>	<b>257</b>	<b>234</b>
<b>Media</b>	<b>44,0</b>	<b>53,4</b>	<b>47,2</b>	<b>51,4</b>	<b>46,8</b>

mediante l'addizione dei valori associati alla stessa lettera nelle varie caselle.

Con una delle due solite formule (euristica od abbreviata), si valutano le devianze tra i diversi livelli dello stesso fattore.

Con la formula euristica,

**la devianza tra arature con 4 gdl è**

$$231^2/5 + 248^2/5 + 258^2/5 + 246^2/5 + 231^2/5 - 1214^2/25 = 109,36$$

uguale a 109,36.

**La devianza tra concimi, con 4 gdl, è**

$$239^2/5 + 245^2/5 + 239^2/5 + 250^2/5 + 241^2/5 - 1214^2/25 = 17,76$$

uguale a 17,76.

**La devianza tra sementi, sempre con 4 gdl, è**

$$220^2/5 + 267^2/5 + 236^2/5 + 257^2/5 + 234^2/5 - 1214^2/25 = 286,16$$

uguale a 286,16.

I risultati delle devianze ed il calcolo delle varianze necessarie ai tre test F, sono riportati in tabella

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>
<b>Totale</b>	480,16	24	
<b>Tra sementi</b>	286,16	4	71,54
<b>Tra arature</b>	109,36	4	27,34
<b>Tra concimi</b>	17,76	4	4,44
<b>Errore</b>	66,88	12	5,57

Con le varianze, si possono calcolare 3 F, ognuno con gdl 4 e 12:

- per il **confronto tra le medie delle sementi**:  $F_{4,12} = \frac{71,54}{5,57} = 12,84$

- per il **confronto tra le medie delle arature**:  $F_{4,12} = \frac{27,34}{5,57} = 4,91$

- per il **confronto tra le medie dei concimi**, con i dati dell'esempio è inutile calcolare il rapporto F. Infatti, il valore di questa varianza risulta minore di quella d'errore; pertanto il rapporto sarebbe minore di 1.

Per gdl 4 e 12 la tabella dei valori critici alla probabilità  $\alpha = 0.05$  fornisce un valore uguale 3,26 e alla probabilità  $\alpha = 0.01$  un valore uguale a 5,41.

In conclusione:

- 1) risulta molto significativa la differenza tra sementi, il cui valore di F è superiore al valore critico della probabilità 0.01;
- 2) è significativa la differenza tra arature, con un valore di F compreso tra il valore critico della probabilità 0.01 e quello della probabilità 0.05;
- 3) non è assolutamente significativa, è anzi totalmente trascurabile, la differenza tra concimi con F minore di 1.

Il disegno sperimentale a quadrati latini è utile non solo nella ricerca di campagna per applicazioni sul terreno, ma anche in varie situazioni ricorrenti **nella ricerca di laboratorio**. E' il caso in cui si devono somministrare più farmaci agli stessi individui o più tossici alle stesse cavie e si vuole considerare contemporaneamente l'effetto di questi trattamenti insieme con le differenze tra gli individui e gli

effetti delle diverse successioni temporali o giorni in cui i trattamenti sono stati somministrati. Non sono poche le situazioni in cui si verifica un effetto biologico o psicologico che dipende dai trattamenti precedenti.

ESEMPIO. Si sperimentano gli effetti di 4 farmaci (A, B, C, D) somministrati a 4 individui (I, II, III, IV) in 4 giorni diversi (1 Lunedì; 2 Martedì; 3 Mercoledì; 4 Giovedì), con una successione di farmaci differente per ogni individuo.

Il lunedì è stato somministrato il farmaco A alla persona I, il farmaco D alla persona II, il farmaco B alla persona III e il farmaco C alla persona IV.

Il martedì sono stati scambiati farmaci e persone, somministrando alla persona I il farmaco C, alla persona II il farmaco B e così di seguito per i quattro giorni, in modo che tutti e quattro gli individui sperimentino i quattro farmaci e senza che nello stesso giorno il medesimo farmaco sia somministrato a due persone.

La tabella riporta il piano dell'esperimento ed i dati relativi ad ogni persona

PERSONE	GIORNI			
	1 Lunedì	2 Martedì	3 Mercoledì	4 Giovedì
I	A 48	C 35	D 40	B 51
II	D 37	B 50	C 33	A 45
III	B 42	D 64	A 53	C 39
IV	C 31	A 40	B 42	D 37

I risultati dell'analisi della varianza a quadrati latini 4x4 sono riportati nella tabella:

	DEVIANZA	GDL	VARIANZA	F
<b>Totale</b>	1098	15		
<b>Tra farmaci</b>	389	3	129,7	2,77
<b>Tra giorni</b>	125	3	41,7	< 1
<b>Tra persone</b>	303	3	101,0	2,16
<b>Errore</b>	281	6	46,8	

Per gdl 3 e 6, il valore critico alla probabilità  $\alpha = 0.05$  è 4,53.

Nessun confronto tra le 4 medie dei 3 fattori risulta significativo.

La causa potrebbe essere la limitata potenza del test, evidenziata dal numero ridotto di dati e dai gdl della varianza d'errore. Infatti con più dati, per esempio con gdl 3 per ogni fattore e gdl 60 per l'errore, il valore critico di **F** alla stessa probabilità  $\alpha = 0.05$  è 2,76 e le differenze tra farmaci sarebbero risultate significative.

**Il disegno sperimentale a quadrati latini ha limiti che dipendono dalle sue dimensioni. Il numero di dati non può essere troppo piccolo, né troppo grande.**

Il **limite inferiore o minimo** è imposto dai gradi di libertà della varianza d'errore.

Un quadrato latino **2 x 2** avrebbe **3 gdl** per la devianza totale, che sarebbero scomposti in

- **1** per il fattore principale,
- **1** per le colonne e
- **1** per le righe.

Non resterebbero gdl per la varianza d'errore.

In una tabella a quadrati latini **3 x 3** ( con 8 gdl per la devianza totale e 2 gdl per la devianza di ognuno dei 3 fattori) la varianza d'errore ha solamente 2 gdl; è possibile condurre l'analisi, ma i gdl sono pochi per rendere significative differenze tra medie che non siano molto grandi, poiché il valore critico alla probabilità  $\alpha = 0.05$  per un **F** con gdl 2 e 2 è uguale a 19,00.

**Il limite minimo utile di un quadrato latino è 4x4.**

**Il limite massimo o superiore** è determinato dalla complessità dell'esperimento; viene quindi abitualmente fissato per un quadrato latino che varia tra **10 x 10** e **12 x 12**. A questi livelli è complesso gestire esperimenti con tanti farmaci da somministrare ad altrettante persone, in un numero equivalente di giorni; molto raramente esistono problemi così complessi e la gestione dell'esperimento può incorrere in vari inconvenienti, che renderebbero molto difficile ottenere tutte le risposte.

I quadrati latini analizzano contemporaneamente 3 fattori, ognuno con p modalità o livelli. **Il disegno può essere esteso a 4 fattori, sempre con lo stesso numero p di modalità.**

**Sono i quadrati greco-latini**, la cui rappresentazione grafica è riportata nella tabella sottostante.

	COLONNE			
RIGHE	I	II	III	IV
1	<b>B g</b>	<b>A d</b>	<b>D a</b>	<b>C b</b>
2	<b>C d</b>	<b>D g</b>	<b>A b</b>	<b>B a</b>
3	<b>D b</b>	<b>C a</b>	<b>B d</b>	<b>A g</b>
4	<b>A a</b>	<b>B b</b>	<b>C g</b>	<b>D d</b>

Nelle righe e nelle colonne sono indicati le modalità di 2 fattori, il terzo fattore è indicato dalle lettere latine ed il quarto con le lettere greche. Ovviamente, tutti i 4 fattori hanno lo stesso numero di livelli (che nella tabella sono 4).

La costruzione della tabella dei quadrati greco-latini è più complessa della precedente a quadrati latini:

- 1) **non solo le lettere latine devono comparire una sola volta per riga e per colonna;**
- 2) **la stessa legge è valida anche per le lettere greche,**
- 3) **con la regola aggiuntiva che ogni lettera greca e ogni lettera latina devono essere associati una volta sola, ma ognuna di esse deve incontrare tutte quelle dell'altro tipo.**

Nella ricerca ambientale è il caso di considerare non solo aratura, concimi e tipo di semente, ma anche un quarto fattore che può essere il tipo di antiparassitari, sempre con lo stesso numero di modalità. Nella ricerca di laboratorio, per un esempio analogo al precedente ma usando cavie, oltre alla razza o specie, al giorno e al tossico somministrato, può essere considerata la classe d'età.

Per ogni singola osservazione  $X_{ijkp}$ , il modello di riferimento è uguale a quello per 4 fattori già presentato

$$X_{ijkp} = m + a_i + b_j + g_k + d_p + R_{ijkp}$$

dove

$m$  rappresenta la media generale,

$a_i$ ,  $b_j$ ,  $g_k$ ,  $d_p$  rappresentano gli effetti indipendenti dei 4 fattori, nelle loro varie modalità,

$R_{ijkp}$  rappresenta tutti gli altri fattori non considerati, sommando alla variazione casuale e agli errori di misura e di campionamento tutte le possibili interazioni tra i 4 fattori, che in questi disegni sperimentali, con una sola osservazione per casella, non possono essere valutate.

Le ipotesi nulle da verificare sono 4, relative alle medie di ognuno dei 4 fattori.

A tal fine, le devianze da calcolare sono aumentate di una unità. Dopo aver sommato tutti i dati contrassegnati dalla stessa lettera greca, con la solita formula deve essere calcolata la devianza relativa, che avrà gli stessi gdl degli altri tre fattori. Con i dati campionari, tali concetti sono espressi dalla relazione

$$X_{ijkp} = \bar{\bar{X}} + (\bar{X}_i - \bar{\bar{X}}) + (\bar{X}_j - \bar{\bar{X}}) + (\bar{X}_k - \bar{\bar{X}}) + (\bar{X}_p - \bar{\bar{X}}) + (X_{ijkp} - \bar{X}_i - \bar{X}_j - \bar{X}_k - \bar{X}_p + 3\bar{\bar{X}})$$

**Il grande vantaggio di questo disegno sperimentale a quadrati greco-latini è rappresentato da un rilevante risparmio di materiale**, superiore a quello dei quadrati latini. **Con 4 fattori a p livelli non sono più necessari  $p^4$  dati ma solamente  $p^2$** ; con 4 livelli per fattore non più 256 dati ma solamente 16; con 5 livelli per fattore, non più 625 osservazioni ( $5 \times 5 \times 5 \times 5$ ), ma solamente 25. Quando il costo od il tempo richiesto per ottenere un dato è alto, questo disegno sperimentale può rappresentare una soluzione possibile per effettuare la ricerca programmata. Tuttavia, occorre considerare che **la riduzione del numero di dati determina una riduzione notevole dei gdl nella varianza d'errore**.

Con lo schema riportato nella tabella precedente con 4 livelli per ognuno dei 4 fattori, si devono calcolare le seguenti quantità, che hanno i gdl di fianco riportati

	DEVIANZA	GDL	VARIANZA	F
TOTALE	XXX	<b>15</b>		
<b>Tra colonne</b>	XXX	<b>3</b>	XXX	XXX
<b>Tra righe</b>	XXX	<b>3</b>	XXX	XXX
<b>Tra lettere latine</b>	XXX	<b>3</b>	XXX	XXX
<b>Tra lettere greche</b>	XXX	<b>3</b>	XXX	XXX
Errore	XXX	<b>3</b>	XXX	

Pure con 16 dati di una tabella  $4 \times 4$ , i gdl per la varianza d'errore sono solamente 3 e i test F hanno gdl 3 e 3. Il valore critico di **F** alla probabilità  $\alpha = 0.05$  è 9,28 e alla probabilità  $\alpha = 0.01$  è 29,46. Per ricorrere ai quadrati greco-latini in tabelle  $4 \times 4$  è quindi necessario che la varianza d'errore sia molto piccola; è un risultato che può essere ottenuto quando i fattori considerati sono gli unici reali elementi di variabilità.

Un'altra soluzione può essere un aumento delle dimensioni dell'esperimento ad almeno  $5 \times 5$ , che ha 4 gdl per ogni fattore ed 8 gdl nell'errore, con valori critici di **F** uguali a 3,84 alla probabilità  $\alpha = 0.05$  e 7,01 alla probabilità  $\alpha = 0.01$ .

E' possibile utilizzare esperimenti di dimensioni ancora maggiori; ma con la grande difficoltà di gestire esperimenti così complessi con uno schema tanto rigido e trovare tante modalità per ognuno dei 4 fattori.

#### 9.5. DATI MANCANTI O ANOMALI IN DISEGNI A PIU' FATTORI

**Nei disegni a blocchi randomizzati e a tre o più fattori, negli esperimenti a quadrati latini e greco-latini, la mancanza anche di una sola osservazione pone vari problemi di elaborazione dei dati.** A differenza di quanto avviene nel disegno sperimentale ad un criterio di classificazione o completamente randomizzato, dove l'analisi della varianza non richiede l'eguaglianza del numero di repliche e la perdita di un dato riduce semplicemente il campione di una unità, per cui **non è necessario pensare alla sua sostituzione**, nei disegni a più criteri è vantaggioso avere un numero rigidamente prefissato e costante di osservazioni, cioè non avere **dati mancanti** (*missing data*, *missing values*), **per utilizzare le formule presentate.**

Nel corso di un esperimento può avvenire che

- **uno o più dati non possano essere rilevati**, come nel caso della morte di una cavia o dell'individuo sul quale il parametro avrebbe dovuto essere misurato,
- **dopo la rilevazione uno o più dati siano persi**, per cancellazione involontaria o smarrimento dell'annotazione, nei vari passaggi richiesti per l'elaborazione statistica,
- **lo strumento fornisca una o più misure molto approssimate**, per le quali non è tarato.

Per sostituire un dato mancante, è fondamentale conoscere la causa della sua non disponibilità, che ai fini statistici può avvenire essenzialmente per **2 motivi**:

- **per selezione contro certi valori**, che sono troppo grandi o troppo piccoli per essere misurati correttamente dallo strumento a disposizione,
- **per una causa accidentale.**

**Nella prima situazione** (impossibilità di quantificare esattamente uno o più valori), **il campione raccolto è viziato in modo irrimediabile.** Non è possibile sostituire in alcun modo le misure mancanti, ma solo indicarle in modo approssimato (superiore a X, inferiore a Y); per la loro analisi è utile ricorrere a test non parametrici. Quando un valore solo è estremamente basso, è accettabile la sua sostituzione con zero oppure la media fra zero e il valore minimo misurabile dallo strumento. Se questi valori sono pochi, rispetto al numero totale di osservazioni, questa procedura è ancora possibile, per

l'uso di test parametrici. Ma quando i valori uguali a zero, comunque molto bassi, sono numerosi, la distribuzione dei dati diventa fortemente asimmetrica; quindi per l'inferenza è corretto solo il ricorso a test fondati sul segno o sull'ordine, se la trasformazione dei dati non è in grado di ripristinare le condizioni di validità della statistica parametrica.

Un gruppo diverso di problemi deriva anche da un solo valore anomalo, quando eccezionalmente alto. La distribuzione dei dati diviene asimmetrica e la varianza d'errore molto grande, rendendo i test non più significativi. Il caso è discusso alla fine del paragrafo.

Nella seconda situazione, **quando uno o più valori sono stati perduti o non sono stati rilevati per cause accidentali, è possibile pervenire ad una loro stima.** La loro sostituzione è necessaria, per **mantenere i campioni bilanciati e quindi utilizzare le formule semplici presentate,** pure con correzioni successive.

I metodi, le formule e i dati di seguito riportati sono tratti, con alcune integrazioni, dal testo di George W. **Snedecor** e William G. **Cochran** *STATISTICAL METHODS* (Sixth edition, 1967, the Iowa State University Press, pp. 317-320).

In un **disegno a blocchi randomizzati in cui manca un valore** (*single missing observation in a two-way anova without replication*) come nella tabella sottostante (trattamento D, blocco I)

	TRATTAMENTI				
BLOCCHI	A	B	C	D	TOTALE
I	32,3	33,3	30,8	--X--	<b>96,4</b>
II	34,0	33,0	34,3	26,0	127,3
III	34,3	36,3	35,3	29,8	135,7
IV	35,0	36,8	32,3	28,0	132,1
V	36,5	34,5	35,8	28,8	135,6
TOTALE	172,1	173,9	168,5	<b>112,6</b>	<b>627,1</b>

dopo aver effettuato le somme (senza il dato mancante), si perviene ad una sua stima ( $\hat{X}$ ) mediante la formula proposta da F. **Yates** nel 1933,

$$\hat{X} = \frac{r \cdot R + c \cdot C - G}{(r-1) \cdot (c-1)}$$

dove

- $r$  = numero di righe o blocchi,
- $R$  = totale di riga o blocco senza il dato mancante,
- $c$  = numero di colonne o trattamenti,
- $C$  = totale di colonna o trattamento senza il dato mancante,
- $G$  = totale generale senza il dato mancante.

Per la stima del dato mancante, sono presi in considerazione non solo

- l'effetto **del blocco** ( $\bar{X}_i - \bar{\bar{X}}$ )
- l'effetto **del trattamento** ( $\bar{X}_j - \bar{\bar{X}}$ ) ai quali appartiene, oltre alla
- **media generale** ( $\bar{\bar{X}}$ ),

ma pure il loro **numero di osservazioni**; inoltre i calcoli sono effettuati sul **principio dei minimi quadrati**, cioè con lo scopo di ridurre al minimo l'errore.

Dai dati della tabella precedente, con

$$r = 5 \quad R = 96,4 \quad c = 4 \quad C = 112,6 \quad G = 627,1$$

si stima  $\hat{X}$ , il valore mancante,

$$\hat{X} = \frac{5 \cdot 96,4 + 4 \cdot 112,6 - 627,1}{(5-1) \cdot (4-1)} = 25,4$$

che risulta uguale a 25,4.

Con questo dato ( $\hat{X} = 25,4$ ) diventa possibile costruire la tabella completa e calcolare i nuovi totali

	TRATTAMENTI				
BLOCCHI	A	B	C	D	TOTALE
I	32,3	33,3	30,8	<b>25,4</b>	<b>121,8</b>
II	34,0	33,0	34,3	26,0	127,3
III	34,3	36,3	35,3	29,8	135,7
IV	35,0	36,8	32,3	28,0	132,1
V	36,5	34,5	35,8	28,8	135,6
TOTALE	172,1	173,9	168,5	<b>138,0</b>	<b>652,5</b>

ai quali **si applicano le formule note dell'analisi della varianza a due criteri di classificazione senza repliche**.

Ma l'inserimento di questo valore stimato dagli altri, quindi senza una sua reale informazione, ha indotto alcuni errori sistematici (**bias**), che devono essere corretti, per un'analisi più precisa.

Alle stime dei vari parametri, ottenute con i metodi classici e necessarie per i test di significatività e per il calcolo degli intervalli fiduciali, devono essere apportate le correzioni (in grassetto e con l'asterisco nella tabella successiva), che interessano tre elementi:

- i gradi di libertà,
- l'errore standard delle medie,
- la devianza e la varianza dei trattamenti e dei blocchi.

	DEVIANZA	GDL	VARIANZA
Totale	224,08	<b>18*</b>	----
Tra Trattamenti	171,36 ( <b>135,78*</b> )	3	57,12 ( <b>45,79*</b> )
Tra Blocchi	35,39 ( <b>29,73*</b> )	4	8,85 ( <b>7,34*</b> )
Errore	17,33	<b>11*</b>	1,58

1) La prima correzione riguarda **i gradi di libertà della devianza totale e dell'errore.**

Poiché il dato inserito è ricavato dagli altri **N-1** (19) e quindi non aggiunge un'informazione reale, i gdl della devianza totale sono **N-2** (18).

Non variano i gdl dei trattamenti e dei blocchi, mentre diminuisce di una unità ( $11 = 18 - 7$ ) il numero di gdl dell'errore, sia perché al dato inserito manca appunto la variabilità casuale o d'errore, sia per la proprietà additiva dei gdl.

2) La seconda correzione concerne il **confronto tra le medie di due trattamenti.**

Quando **nessuna delle due medie a confronto è calcolata sul gruppo con il dato mancante,**

l'errore standard (**e.s.**) rimane invariato

$$e.s. = \sqrt{\frac{2 \cdot s_e^2}{r}} = \sqrt{\frac{2 \cdot 1,58}{5}} = 0,795$$

Con

- $s_e^2$  = varianza d'errore stimata (1,58 nell'esempio)
- r = numero di blocchi (righe) o dati sui quali sono calcolate le medie dei trattamenti (5 nell'esempio)

per valutare la significatività della differenza, l'errore standard (e.s.) è 0,795.

Ma quando **nel confronto è compresa una media calcolata su un gruppo con il dato stimato**, la differenza minima significativa è maggiore di quella precedentemente stimata

$$\text{e.s.} = \sqrt{s_e^2 \cdot \left( \frac{2}{r} + \frac{c}{r \cdot (r-1) \cdot (c-1)} \right)} = \sqrt{1,58 \cdot \left( \frac{2}{5} + \frac{4}{5 \cdot 4 \cdot 3} \right)} = \sqrt{1,58 \cdot 0,4667} = 0,859$$

e il suo errore standard, con la simbologia precedente e i dati dell'esempio, risulta uguale a 0,859.

3) La terza variazione al metodo generale è **la correzione della devianza e/o della varianza sia tra trattamenti sia tra blocchi**, per verificare la significatività delle differenze tra le loro medie con il test F.

La varianza tra trattamenti e quella tra blocchi, come risultano quando calcolate con il dato inserito (57,12 e 8,85 nella ultima tabella) utilizzando le formule tradizionali, sono maggiori di quelle reali; quindi il test F per la significatività tra le medie dei trattamenti e quello tra le medie dei blocchi in realtà sono meno significativi di quanto risulta da queste stime, ottenute con i valori stimati.

Per eliminare questa distorsione (**bias**), è utile apportare una correzione che, per la **varianza tra trattamenti**, è data dal valore di **A**

$$\mathbf{A} = \frac{[R - (c-1) \cdot \hat{X}]^2}{c \cdot (c-1)^2}$$

Con i dati dell'esempio,

**A** risulta

$$\frac{[96,4 - (4-1) \cdot 25,4]^2}{4 \cdot (4-1)^2} = \frac{20,2^2}{36} = 11,33$$

uguale a 11,33.

Di conseguenza, per la significatività della varianza tra trattamenti con il test F non deve essere utilizzato il valore calcolato (**57,12**) con le formule generali; ma ad esso occorre sottrarre **A**, per ottenere il valore corretto (**45,79** = 57,12 - 11,33).

Il test F tra trattamenti con gdl 3 e 11 è

$$F_{3,11} = \frac{45,79}{1,58} = 28,98$$

In altri testi, **sempre per la correzione della varianza tra trattamenti**, è proposto l'uso non della somma ma della media **del blocco**, con la formula

$$A = \frac{[(\bar{X}_i - \hat{X}) \cdot (c-1)]^2}{c \cdot (c-1)^2}$$

dove

- $\bar{X}_i$  è la media del **blocco** senza il dato mancante ( $96,4 / 3 = 32,13$ )
- $\hat{X}$  è il valore stimato, in sostituzione di quello mancante (**25,4**)
- $c$  è il numero di colonne o **trattamenti** (**4**).

Ovviamente, essa fornisce risultati identici alla precedente,

$$A = \frac{[(32,133 - 25,4) \cdot (4-1)]^2}{4 \cdot (4-1)^2} = \frac{(6,733 \cdot 3)^2}{4 \cdot 9} = \frac{408,0}{36} = 11,33$$

come è possibile verificare dal confronto tra le formule e con i dati dell'esempio.

Per la significatività delle differenze tra le medie dei blocchi, **alla varianza tra blocchi deve essere sottratta una quantità B**

$$B = \frac{[(\bar{X}_j - \hat{X}) \cdot (r-1)]^2}{r \cdot (r-1)^2}$$

dove

- $\bar{X}_j$  è la media del **trattamento** al quale appartiene il dato mancante ( $112,6 / 4 = 28,15$ )
- $\hat{X}$  è il **valore stimato** per sostituire quello mancante (**25,4**)
- $r$  è il numero di righe o **blocchi** (**5**).

$$B = \frac{[(28,15 - 25,4) \cdot (5-1)]^2}{5 \cdot (5-1)^2} = \frac{(2,75 \cdot 4)^2}{5 \cdot 16} = \frac{121}{80} = 1,51$$

Pertanto, nella tabella delle devianze e delle varianze, la varianza tra blocchi risulta uguale a 7,34 (8,85 - 1,51) e il test F per la verifica delle differenze tra le medie dei blocchi con gdl 4 e 11 risulta

$$F_{4,11} = \frac{7,34}{1,58} = 4,645$$

uguale a 4,645.

Altri testi, quale *BIOMETRY* di Robert R. **Sokal** e F. James **Rohlf** (third edition, 1995, W. H. Freeman and Company, New York), propongono metodi per correggere le devianze, anche se poi esse perdono la proprietà additiva del modello classico.

Per la **devianza tra trattamenti** si deve stimare un correttore **A**

$$\mathbf{A} = \frac{(R + c \cdot C - G)^2}{c \cdot (c - 1) \cdot (r - 1)^2}$$

che con

$$r = 5 \quad R = 96,4 \quad c = 4 \quad C = 112,6 \quad G = 627,1$$

risulta

$$\mathbf{A} = \frac{(96,4 + 4 \cdot 112,6 - 627,1)^2}{4 \cdot (4 - 1) \cdot (5 - 1)^2} = \frac{(96,4 + 450,4 - 627,1)^2}{4 \cdot 3 \cdot 16} = \frac{6,448}{192} = 35,58$$

uguale a 35,58

La **devianza aggiustata (adjusted) tra trattamenti** è quindi uguale a **135,78** (171,36 – 35,58) e la sua varianza è uguale a **45,26** (135,78 / 3), non differente dalla stima precedente (45,79).

Per la **devianza tra blocchi** si deve stimare un correttore **B**

$$\mathbf{B} = \frac{(C + r \cdot R - G)^2}{r \cdot (r - 1) \cdot (c - 1)^2}$$

Con gli stessi valori usati per il calcolo di A, risulta

$$\mathbf{B} = \frac{(112,6 + 5 \cdot 96,4 - 627,1)^2}{5 \cdot (5 - 1) \cdot (4 - 1)^2} = \frac{(112,6 + 482,0 - 627,1)^2}{5 \cdot 4 \cdot 9} = \frac{1,056}{180} = 5,86$$

uguale a 5,86

La **devianza aggiustata (adjusted) tra blocchi** è quindi uguale a **29,73** (35,59 – 5,86); quindi la **varianza aggiustata** è uguale a **7,43** (29,73 / 4), non differente dal 7,34 precedente.

Un'altra situazione, riportata con relativa frequenza sui testi, in cui si pone il problema di stimare un valore mancante è il disegno sperimentale a **quadrati latini** (*latin square with one missing value*).

Il valore mancante ( $\hat{X}$ ) è calcolato con

$$\hat{X} = \frac{k \cdot (R + C + T) - 2G}{(k - 1) \cdot (k - 2)}$$

dove

- **k** è il numero di dati per fattore o dimensioni del quadrato latino,

- **R, C, T** sono i totali rispettivamente di Riga, di Colonna e del Trattamento (ovviamente, calcolati senza il dato mancante),
- **G** è il totale generale (sempre calcolato senza il dato mancante).

Nella tabella a quadrati latini 3 x 3, nella quale in colonna 1 e riga I manca un dato,

	COLONNE			
RIGHE	1	2	3	TOTALE
I	A --X--	B 885	C 940	<b>1825</b>
II	B 715	C 1087	A 766	2568
III	C 844	A 711	B 832	2387
TOTALE	<b>1559</b>	2683	2538	<b>6780</b>

dopo aver calcolato i totali di colonna, di riga e dei trattamenti

TRATTAMENTI	A	B	C
TOTALE	<b>1447</b>	2432	2871

e quindi con

$$k = 3, \quad R = 1825, \quad C = 1559, \quad T = 1447, \quad G = 6780$$

il valore stimato  $\hat{X}$  di questo dato mancante risulta

$$\hat{X} = \frac{3 \cdot (1825 + 1559 + 1477) - 2 \cdot 6780}{(3-1) \cdot (3-2)} = \frac{14583 - 13560}{2} = 511,5$$

uguale a 511,5 (arrotondato a 512, per avere un valore senza decimali, come gli altri dati).

Con l'inserimento del dato stimato ( $\hat{X} = 512$ ) e l'uso delle formule classiche, si perviene ai risultati della tabella seguente

	DEVIANZA	GDL	VARIANZA
Totale	210.460	<b>7</b>	----
Tra Righe	9.847	2	4.923,5
Tra Colonne	68.185	2	34.092,5
Tra Trattamenti	129.655	2	64.827,5 ( <b>40.407,5</b> )
Errore	2.773	<b>1</b>	2.773

(IMPORTANTE: anche se riportato su uno dei testi più classici, l'esempio è finalizzato solamente alla dimostrazione della stima del valore mancante e della correzione della varianza tra trattamenti. In realtà un solo gdl nella varianza d'errore è troppo piccolo, per effettuare un test sufficientemente potente.)

L'errore sistematico (*bias*), determinato dalla presenza del dato stimato, secondo lo scopo dell'analisi può essere calcolato per la varianza tra righe, la varianza tra colonne e quella tra trattamenti.

Limitando la dimostrazione ai soli trattamenti per illustrare i concetti, poiché ora le varianze corrette sono fornite dai programmi informatici,

il *bias* può essere stimato con

$$A = \frac{[G - R - C - (k-1) \cdot T]^2}{(k-1)^3 \cdot (k-2)^2}$$

utilizzando la stessa simbologia e gli stessi dati della formula per il calcolo del dato mancante.

Nell'esempio, **A** risulta

$$A = \frac{[6780 - 1825 - 1559 - (3-1) \cdot 1477]^2}{(3-1)^3 \cdot (3-2)^2} = \frac{442^2}{8 \cdot 1} = \mathbf{24.420}$$

uguale a 24.420; di conseguenza, la varianza tra trattamenti corretta risulta uguale a **40.407,5** (64.827,5 - 24.420).

Sempre in un quadrato latino di dimensione **k x k**, l'errore standard della differenza tra le medie di **due trattamenti, nei quali non sia compreso il dato mancante**, resta

$$e.s. = \sqrt{s_e^2 \cdot \frac{2}{k}}$$

Quando **una delle due medie a confronto** (ovviamente, in modo distinto per righe, colonne e trattamenti) è **stimata sul gruppo che comprende il dato mancante**,

l'errore standard della differenza diventa

$$\text{e.s.} = \sqrt{s_e^2 \cdot \left( \frac{2}{k} + \frac{1}{(k-1) \cdot (k-2)} \right)}$$

Con  $s_e^2 = 2773$  e  $k = 3$

per due medie calcolate su gruppi diversi da quello del dato sostituito

$$\text{e.s.} = \sqrt{2773 \cdot \frac{2}{3}} = \sqrt{1849} = 43,0$$

l'errore standard risulta uguale a 43,0.

L'**errore standard della differenza tra due medie delle quali una contenga il dato sostituito**, è

$$\text{e.s.} = \sqrt{2773 \cdot \left( \frac{2}{3} + \frac{1}{(3-1) \cdot (3-2)} \right)} = \sqrt{2773 \cdot \frac{7}{6}} = \sqrt{3239} = 56,9$$

uguale a 56,9.

Quando **i dati mancanti sono due o più**, la procedura per la loro sostituzione non è più complessa ma più lunga, perché iterativa.

Supponendo, per esclusiva comodità di calcolo, di avere un **disegno a blocchi randomizzati 3 x 3**,

	TRATTAMENTI			
BLOCCHI	A	B	C	TOTALE
I	6	15	$X_{3,1}$	<b>21</b>
II	5	$X_{2,2}$	15	<b>20</b>
III	4	8	12	24
TOTALE	15	<b>23</b>	<b>27</b>	<b>65</b>

con due valori mancanti ( $X_{2,2}$  e  $X_{3,1}$ ), il metodo richiede alcuni passaggi:

1 – la scelta del primo valore è indifferente, per cui è possibile iniziare a caso da  $X_{2,2}$ ; ad esso si deve attribuire un valore presunto, che non inciderà sul risultato finale, poiché quando è distante dal reale ha il solo effetto di aumentare il numero di iterazioni;

2 – ma poiché è conveniente un valore prossimo a quello che sarà stimato, dato che la media del trattamento al quale appartiene è **11,5** (23/2) e quella del blocco è **10,0** (20/2), in prima approssimazione può essere ragionevole scegliere un valore intermedio tra essi, come **10,5**;

3 – con l’inserimento di  $X_{2,2} = 10,5$  nella tabella, i totali diventano

	TRATTAMENTI			
BLOCCHI	A	B	C	TOTALE
I	6	15	$X_{3,1}$	<b>21</b>
II	5	<b>10,5</b>	15	30,5
III	4	8	12	24
TOTALE	15	33,5	<b>27</b>	<b>75,5</b>

4 – si calcola il secondo valore mancante ( $X_{3,1}$ )

con la solita formula

$$\hat{X} = \frac{r \cdot R + c \cdot C - G}{(r-1) \cdot (c-1)}$$

dove

$$r = 3 \quad R = 21 \quad c = 3 \quad C = 27 \quad G = 75,5$$

ottenendo

$$\hat{X}_{3,1} = \frac{3 \cdot 27 + 3 \cdot 21 - 75,5}{(3-1) \cdot (3-1)} = \frac{81 + 63 - 75,5}{4} = 17,125$$

un risultato uguale a 17,125;

5 - si inserisce  $X_{3,1} = 17,125$  nella tabella

	TRATTAMENTI			
BLOCCHI	A	B	C	TOTALE
I	6	15	<b>17,125</b>	38,125
II	5	$X_{2,2}$	15	<b>20</b>
III	4	8	12	24
TOTALE	15	<b>23</b>	44,125	<b>82,125</b>

e quindi con i totali

$$r = 3 \quad R = 20 \quad c = 3 \quad C = 23 \quad G = 82,125$$

si ritorna a calcolare nuovamente il valore di  $X_{2,2}$

$$\hat{X}_{3,1} = \frac{3 \cdot 20 + 3 \cdot 23 - 82,125}{(3-1) \cdot (3-1)} = \frac{60 + 69 - 82,125}{4} = 11,719$$

che diventa 11,719;

6 – sostituendo il precedente  $X_{2,2} = 10,5$  con l'attuale 11,719 si ottiene una tabella

	TRATTAMENTI			
BLOCCHI	A	B	C	TOTALE
I	6	15	$X_{3,1}$	<b>21</b>
II	5	<b>11,719</b>	15	31,719
III	4	8	12	24
TOTALE	15	34,719	<b>27</b>	<b>76,719</b>

e quindi i totali

$$r = 3 \quad R = 21 \quad c = 3 \quad C = 27 \quad G = 76,719$$

dai quali si perviene a un nuovo valore calcolato

per  $\hat{X}_{3,1}$

$$\hat{X}_{3,1} = \frac{3 \cdot 21 + 3 \cdot 27 - 76,719}{(3-1) \cdot (3-1)} = \frac{63 + 81 - 76,719}{4} = 16,82$$

che risulta uguale a 16,82;

7 – con questo ultimo valore di  $\hat{X}_{3,1}$  inserito nella tabella

	TRATTAMENTI			
BLOCCHI	A	B	C	TOTALE
I	6	15	<b>16,82</b>	37,82
II	5	$X_{2,2}$	15	<b>20</b>
III	4	8	12	24
TOTALE	15	<b>23</b>	43,82	<b>81,82</b>

e quindi con i totali

$$r = 3 \quad R = 20 \quad c = 3 \quad C = 23 \quad G = 81,82$$

il nuovo valore calcolato di  $X_{2,2}$

$$\hat{X}_{2,2} = \frac{3 \cdot 20 + 3 \cdot 23 - 81,82}{(3-1) \cdot (3-1)} = \frac{60 + 69 - 81,82}{4} = 11,795$$

diventa 11,795;

8 – questo ultimo valore ( $X_{2,2} = 11,795$ ) può essere ritenuto sostanzialmente non differente dalla stima precedente (11,719), data la evidente variabilità dei dati e il loro arrotondamento all'unità; di conseguenza, la tabella definitiva con i due dati calcolati diventa la seguente:

	TRATTAMENTI			
BLOCCHI	A	B	C	TOTALE
I	6	15	<b>16,8</b>	<b>37,8</b>
II	5	<b>11,7</b>	15	31,7
III	4	8	12	24
TOTALE	15	34,7	<b>43,8</b>	<b>93,5</b>

Se l'ultimo valore stimato fosse ritenuto ancora troppo distante dal precedente, si continua l'iterazione che determina valori tra loro sempre più vicini.

Per l'inferenza sulle medie dei trattamenti e dei blocchi, alla tabella definitiva è possibile **applicare l'analisi della varianza a due criteri di classificazione.**

Utilizzando, per maggiore semplicità di calcolo, le formula abbreviate,

- la **devianza totale**

$$(6^2 + 5^2 + 4^2 + 15^2 + 11,7^2 + 8^2 + 16,8^2 + 15^2 + 12^2) - (93,5^2 / 9) = 1.115,84 - 971,36 = 182,77$$

è uguale a 182,7

- la **devianza tra trattamenti**

$$(15^2 / 3) + (34,7^2 / 3) + (43,8^2 / 3) - (93,5^2 / 9) = 75 + 401,36 + 639,48 - 971,36 = 144,48$$

è uguale a 144,48

- la **devianza tra blocchi**

$$(37,8^2 / 3) + (31,7^2 / 3) + (24^2 / 3) - (93,5^2 / 9) = 476,28 + 334,96 + 192 - 971,36 = 31,88$$

è uguale a 31,88

- la **devianza d'errore**, ottenuta per differenza

$$182,77 - (144,48 + 31,88) = 6,41$$

risulta uguale a 6,41

Per i test F, **la tabella completa con i gdl corretti** diventa

	DEVIANZA	GDL	VARIANZA
Totale	182,77	<b>6</b>	----
Tra Trattamenti	144,48 ( <b>116,14</b> )	2	<b>58,07</b>
Tra Blocchi	31,88	2	
Errore	6,41	<b>2</b>	3,205

Ma, a causa dell'introduzione dei dati stimati, anche in questo caso **le devianze tra trattamenti e tra blocchi sono sovrastime del reale; quindi, per un'analisi più precisa, devono essere corrette.**

Supponendo che il confronto che interessa il ricercatore sia solo quello **tra trattamenti**, per correggere la sua devianza è necessario:

1 – calcolare la **devianza entro blocchi**, applicando ai blocchi **l'analisi della varianza ad un solo criterio di classificazione**, in cui il fattore noto siano i blocchi, con esclusione dei dati inseriti; mediante la formula abbreviata

$$(6 - 10,5)^2 + (15 - 10,5)^2 + (5 - 10)^2 + (15 - 10)^2 + (4 - 8)^2 + (8 - 8)^2 + (12 - 8)^2 = \\ 20,25 + 20,25 + 25 + 25 + 16 + 0 + 16 = 122,5$$

risulta uguale a 122,5;

2 – essa ha **4 gdl** (7 - 3) e comprende

- sia la **devianza tra trattamenti** (con 2 gdl)
- sia la **devianza d'errore** (con 2 gdl);

di conseguenza, la **stima corretta della devianza tra trattamenti** è ottenuta sottraendo, a questa ultima devianza entro blocchi (122,5), la devianza d'errore (6,41) calcolata in precedenza nel disegno a due criteri di classificazione; pertanto, la stima corretta della devianza tra trattamenti risulta

$$122,55 - 6,41 = 116,14$$

uguale a 116,14 e la sua varianza, avendo 2 gdl, è uguale a 58,07;

3 – il test F con gdl 2 e 2 per la **significatività delle differenze tra le medie dei trattamenti** è

$$F_{2,2} = \frac{58,07}{3,205} = 18,1$$

Ai **quadrati latini con due o più dati mancanti** sono applicati metodi e concetti simili.

Per **i dati anomali**, l'approccio è più complesso. Al momento della rilevazione strumentale, dell'annotazione manuale e della sua trascrizione in tabelle o fogli informatici per i calcoli e l'applicazione di test statistici, **una misura può essere riportata in modo sbagliato**. Ne consegue una varianza d'errore molto grande, che rende il test non significativo e ne pregiudica le condizioni di validità. Occorre quindi verificare attentamente i dati, almeno per comprendere se si tratta di un errore o di un valore effettivo, seppure anomalo o raro rispetto agli altri.

Un metodo per **individuare questi dati** e per valutare la probabilità che il dato appartenga allo stesso gruppo di valori raccolti è **fondato sul valore dei loro residui**.

Secondo il modello additivo, già presentato per i vari test di significatività sulle medie, in una **analisi della varianza ad un criterio**, con la consueta simbologia il singolo dato ( $X_{ij}$ ) è

determinato da

$$X_{ij} = \bar{\bar{X}} + (\bar{X}_i - \bar{\bar{X}}) + R_{ij}$$

Quindi, l'errore o residuo  $R_{ij}$  di questa osservazione è

$$R_{ij} = X_{ij} - \bar{X}_i$$

In un disegno a **blocchi randomizzati senza repliche**, la singola osservazione ( $X_{ij}$ ) è

data da

$$X_{ij} = \bar{\bar{X}} + (\bar{X}_i - \bar{\bar{X}}) + (\bar{X}_j - \bar{\bar{X}}) + R_{ij}$$

Quindi, il suo errore o residuo  $R_{ij}$  è stimato per differenza mediante

$$R_{ij} = X_{ij} - \bar{X}_i - \bar{X}_j + \bar{\bar{X}}$$

In un **quadrato latino senza repliche**, dove la singola osservazione ( $X_{ijk}$ ) è

data da

$$X_{ijk} = \bar{\bar{X}} + (\bar{X}_i - \bar{\bar{X}}) + (\bar{X}_j - \bar{\bar{X}}) + (\bar{X}_k - \bar{\bar{X}}) + R_{ijk}$$

l'errore o residuo  $R_{ijk}$  è calcolato per differenza con

$$R_{ijk} = X_{ijk} - \bar{X}_i - \bar{X}_j - \bar{X}_k + 2\bar{\bar{X}}$$

**Snedecor e Cochran** nel loro testo citato in precedenza hanno presentato un metodo, proposto da **Anscombe e Tukey** nel 1963, per valutare se, alla probabilità prefissata, il dato sospetto possa essere ritenuto anomalo nel contesto di tutti quelli raccolti.

Questo metodo è fondato sul valore del residuo **R**: per definire il dato come anomalo, **il residuo deve essere maggiore di una quantità minima**

$$|R| > C\sqrt{s_e^2}$$

dove

- $s_e^2$  è la varianza d'errore ottenuta con le formule classiche, comprendendo il dato in discussione
- $C$  è il risultato di una serie di calcoli, di seguito spiegati nei vari passaggi.

1 – Per il disegno sperimentale in oggetto, calcolare

$n$  = gdl della varianza d'errore

$N$  = numero totale di dati

- Nell'analisi della varianza **completamente randomizzata**, con gruppi bilanciati, in cui

$k$  = numero di gruppi    e     $n$  = numero di dati per gruppo

si ha

$$n = k \cdot (n-1) \quad \text{e} \quad N = k \cdot n$$

- Nell'analisi della varianza a **blocchi randomizzati**, con

$r$  = numero di righe    e     $c$  = numero di colonne

si ha

$$n = (r-1) \cdot (c-1) \quad \text{e} \quad N = r \cdot c$$

- Nell'analisi della varianza a **quadrati latini** di dimensioni  $k \times k$ , essi sono

$$n = (k-1) \cdot (k-2) \quad \text{e} \quad N = k^2$$

2 – Stimare  $\alpha$

$$\alpha = \frac{n \cdot P}{N}$$

dove P è la probabilità (es.: 0,025 oppure 0,005) alla quale si vuole rifiutare l'ipotesi che il dato in discussione appartenga alla stessa popolazione dalla quale sono stati estratti tutti gli altri

3 - Dal valore di  $\alpha$  risalire, con la tabella della distribuzione normale, a quello della **deviata normale z a una coda**,

4 - Calcolare

$$H = 1,40 + 0,85 \cdot z$$

5 - Il valore di C è

$$C = H \left( 1 - \frac{H^2 - 2}{4 \cdot n} \right) \cdot \sqrt{\frac{v}{N}}$$

ESEMPIO (per un disegno a blocchi randomizzati)

Si supponga che nella tabella a blocchi randomizzati, nella quale in precedenza è stato inserito il valore di 25,4 al posto del dato mancante, successivamente sia stato ritrovato il presunto valore originario e che esso corrisponda a 29,3.

La differenza da quello stimato può far sorgere il sospetto che il presunto valore originario sia anomalo. Come stimare **la probabilità che esso appartenga effettivamente alla stessa popolazione degli altri 19 dati raccolti?**

	TRATTAMENTI				
BLOCCHI	A	B	C	D	TOTALE
I	32,3	33,3	30,8	<b>29,3</b>	<b>125,7</b>
II	34,0	33,0	34,3	26,0	127,3
III	34,3	36,3	35,3	29,8	135,7
IV	35,0	36,8	32,3	28,0	132,1
V	36,5	34,5	35,8	28,8	135,6
TOTALE	172,1	173,9	168,5	<b>141,9</b>	<b>656,4</b>

Risposta

1 - Il primo passo consiste nel calcolare il residuo  $R_{ij}$  del dato in discussione.

A questo scopo, da tutti i 20 dati della tabella calcolare

- la media di colonna o trattamento  $\bar{X}_i = 31,425$  (125,7 / 4)
- la media di riga o blocco  $\bar{X}_j = 28,38$  (141,9 / 5)
- la media generale  $\bar{\bar{X}} = 32,82$  (656,4 / 20)

Attraverso essi, il residuo  $R_{ij}$  del valore in discussione (29,3) risulta

$$R_{ij} = 29,3 - 31,425 - 28,38 + 32,82 = 2,315$$

uguale a 2,315

2 - Sempre dai dati della tabella, si stimano

- $N = 20$
- la varianza d'errore  $s_e^2 = 2,19$  con  $n = 12$

3 - Per rifiutare alla probabilità unilaterale  $P = 0,025$  (perché il residuo più estremo può essere solo dalla stessa parte) l'ipotesi nulla che il dato appartenga alla stessa popolazione delle altre 19 osservazioni,

calcolare  $\alpha$

$$\alpha = \frac{12 \cdot 0,025}{20} = \frac{0,3}{20} = 0,015$$

che risulta uguale a 0,015.

4 - Ad  $\alpha = 0,015$  in una tabella normale unilaterale corrisponde un valore di  $z$  uguale a 2,17.

5 - Calcolare  $H$  che

$$H = 1,40 + (0,85) \cdot (2,17) = 3,24$$

risulta uguale a 3,24.

6 - Il valore di  $C$

$$C = 3,24 \cdot \left( 1 - \frac{3,24^2 - 2}{4 \cdot 12} \right) \cdot \sqrt{\frac{12}{20}} = 3,24 \cdot \left( 1 - \frac{8,50}{48} \right) \cdot \sqrt{0,6} = 2,07$$

è uguale a 2,07.

7 - Il residuo minimo significativo alla probabilità  $P$  unilaterale prefissata è

$$C \cdot \sqrt{s_e^2} = 2,07 \cdot \sqrt{2,19} = 3,06$$

uguale a 3,06.

8 – Poiché il residuo di 29,3 è uguale a 2,315 ed esso risulta minore del valore minimo 3,06 calcolato, non è possibile rifiutare l'ipotesi nulla alla probabilità  $\alpha = 0,05$ : il valore 29,3 non è significativamente differente dagli altri 19 dati raccolti.

Se la verifica delle procedure sperimentali (rilevazione e trascrizione delle misure rilevate) dimostra che il dato indicato come anomalo è effettivamente un errore e non è più possibile risalire al valore reale, è corretto **procedere alla sua sostituzione come se fosse un dato mancante**.

Quando la misura e la trascrizione sono state corrette, è possibile l'uso della statistica parametrica solo se il dato non risulta anomalo, con procedure quale quella indicata; in caso contrario, è bene ricorrere a test non parametrici, che per l'analisi della varianza ad uno e a due criteri di classificazione sono di uso corrente e ormai riportati su tutti i programmi informatici.

#### **9.6. EFFICIENZA RELATIVA DI DUE DISEGNI SPERIMENTALI**

Per una esatta conoscenza dell'effetto dei vari fattori, l'analisi della varianza a blocchi randomizzati e quella a più criteri di classificazione sono indubbiamente più vantaggiose di quella ad un solo criterio.

**Lo scopo di un disegno sperimentale a più criteri è anche quello di rendere più significativo il test F per i trattamenti**, mediante il controllo delle altre fonti di variazione e la riduzione della varianza d'errore. **Tuttavia, non sempre con i disegni a più criteri si ottiene un risultato più significativo, per le sole differenze tra le medie dei trattamenti.**

Rispetto all'analisi ad un criterio di classificazione, in una analisi della varianza a blocchi randomizzati o a quadrati latini **si ha sempre una riduzione sia della devianza d'errore che dei suoi gdl. Ma la varianza d'errore, determinata dal loro rapporto, non sempre diminuisce**; ovviamente, se la devianza d'errore diminuisce meno dei gdl che perde, la varianza aumenta.

**Occorre inoltre considerare che, al vantaggio derivante dall'eventuale abbassamento della varianza d'errore, si associa sempre lo svantaggio dovuto alla riduzione dei suoi gradi di libertà**; di conseguenza, il valore di F richiesto per dimostrare la significatività del test aumenta. Soprattutto quando i gradi di libertà sono pochi, lo svantaggio derivante dal calo dei gdl della varianza d'errore può essere grave.

Per esempio, il valore di F alla probabilità  $\alpha = 0.05$

- per gdl 4 e 12 è 3,26

mentre per gdl 4 e 2 è 19,25.

Di conseguenza, per il solo effetto della riduzione dei gdl, si richiede che la varianza d'errore diminuisca di circa 6 volte, per determinare un F ugualmente significativo.

**La convenienza ad impiegare un disegno complesso (a più fattori) o più semplice può essere valutata solo a posteriori**, mediante un confronto dei diversi risultati.

In un precedente esempio di questo capitolo (paragrafo 9.2), l'**analisi della varianza a blocchi randomizzati** ha fornito il seguente risultato:

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>
<b>Totale</b>	3362	7	----
<b>Tra giorni</b>	1352	1	1352
<b>Tra ore</b>	1987	3	662,333
<b>Errore</b>	23	3	7,666

Con i medesimi dati è sempre possibile condurre un'**analisi della varianza ad un solo criterio**, calcolando solamente la devianza tra giorni; di conseguenza, la devianza tra ore ed i suoi gdl devono essere cumulati con quelli d'errore, con il seguente risultato:

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>
<b>Totale</b>	3362	7	----
<b>Tra giorni</b>	1352	1	1352
<b>Errore</b>	2010	6	335

Nel caso riportato nella prima tabella, per verificare la differenza tra giorni si ottiene un test  $F_{1,3}$

$$F_{1,3} = \frac{1352}{7,666} = 176,36$$

uguale a 176,36 che, con gdl 1 e 3, risulta altamente significativo, poiché il valore critico alla probabilità 0.05 è uguale a 10,13 e alla probabilità 0.01 è uguale a 34,12.

Nel caso della seconda tabella, la differenza tra giorni è verificata con test  $F_{1,6}$

$$F_{1,6} = \frac{1352}{335} = 4,035$$

che risulta uguale a 4,035 e con gdl 1 e 6 non risulta significativo, poiché il suo valore critico alla probabilità 0.05 è uguale a 5,99.

La spiegazione del diverso risultato è nella differenza tra ore, che è molto alta, come dimostra la sua varianza rispetto a quella d'errore. **Anche se l'analisi intende saggiare solo le differenze tra giorni, è conveniente eliminare l'effetto dovuto alla variabilità tra ore, nonostante il dimezzamento dei gdl nell'errore.**

Tale convenienza ad utilizzare lo schema a 2 criteri rispetto a quello ad 1 solo criterio può essere misurata in modo oggettivo, traducendo i concetti e le valutazioni precedenti in una misura.

E' l'**efficienza relativa** (chiamata anche **efficacia relativa** e simboleggiato con **E.R.**) che misura l'efficacia di un disegno sperimentale rispetto ad un altro. Secondo la formula proposta di Fisher essa è data da

$$E. R. = \frac{(n_1 + 1) / (n_1 + 3) \cdot s_1^2}{(n_2 + 1) / (n_2 + 3) \cdot s_2^2}$$

che, in modo più conveniente per i calcoli, in molti testi è scritta come

$$E. R. = \frac{(n_1 + 1) \cdot (n_2 + 3) \cdot s_2^2}{(n_2 + 1) \cdot (n_1 + 3) \cdot s_1^2}$$

dove

$n_1$  è il numero di gdl della varianza d'errore nell'analisi con il 1° metodo,

$n_2$  è il numero di gdl della varianza d'errore nell'analisi con il 2° metodo,

$s_1^2$  è la varianza d'errore nell'analisi con il 1° metodo,

$s_2^2$  è la varianza d'errore nell'analisi con il 2° metodo.

Con i dati sopra riportati, l'efficienza relativa del disegno a due criteri rispetto a quello ad un solo criterio è

$$E. R. = \frac{(3 + 1) \cdot (6 + 3) \cdot 335}{(6 + 1) \cdot (3 + 3) \cdot 7,666} = \frac{12060}{391,97} = 30,77$$

uguale a 30,77.

E' una differenza molto grande, come evidenzia la differenza tra le due varianze d'errore.

Per effetto della grande variabilità tra ore, **l'esperimento a due criteri di classificazione risulta quasi 31 volte più efficace di quello ad un solo criterio. Il significato di tale rapporto è che per ottenere la stessa potenza del test a blocchi randomizzati, con il test a disegno completamente randomizzato occorrono quasi 31 volte più dati:** non le 8 osservazioni usate ma addirittura 248.

Sulla base degli stessi principi e con lo stesso metodo, è possibile calcolare l'efficienza di un quadrato latino sia rispetto al disegno a due criteri che a quello ad un solo criterio. In un esempio precedente sui quadrati latini, l'analisi della varianza aveva fornito i seguenti risultati:

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>
<b>Totale</b>	480,16	24	----
<b>Tra sementi</b>	286,16	4	71,54
<b>Tra concimi</b>	109,36	4	27,34
<b>Tra arature</b>	17,76	4	4,44
<b>Errore</b>	66,88	12	5,57

Con gli stessi dati, sarebbe stato possibile **ignorare la differenza tra arature**; la sua devianza e i suoi gdl sarebbero stati sommati a quelli d'errore e i risultati dell'analisi della varianza sarebbero diventati:

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>
<b>Totale</b>	480,16	24	----
<b>Tra sementi</b>	286,16	4	71,54
<b>Tra concimi</b>	109,36	4	27,34
<b>Errore</b>	84,64	16	5,29

Con questi nuovi dati sarebbe stato possibile stimare i due F, relativi alle differenze tra le medie delle sementi:

$$F_{4,16} = \frac{71,54}{5,29} = 13,52$$

e alla differenze tra le medie dei concimi

$$F_{4,16} = \frac{27,34}{5,29} = 5,17$$

che risultano più significative delle precedenti, sia a causa della riduzione della varianza d'errore sia per l'aumento dei gradi di libertà che abbassa il livello del valore critico.

Alla probabilità  $\alpha = 0.05$  passando da gdl 4 e 12 a gdl 4 e 16 il valore di F diminuisce da 3,26 a 3,01; alla probabilità  $\alpha = 0.01$  da 5,41 si riduce a 4,77.

**Dopo aver verificato con un quadrato latino che non esiste una differenza significativa tra arature, per dimostrare la significatività delle differenze tra sementi e tra concimi è pertanto più opportuno presentare i risultati secondo lo schema a blocchi randomizzati, anche se l'esperimento è stato effettivamente condotto secondo il precedente schema più complesso.**

L'ulteriore possibile riduzione allo schema con un solo criterio di classificazione comporta la scelta del fattore principale. Continuando sempre con il medesimo esempio, come fattore principale potrebbe essere scelto sia il tipo di semente che il concime utilizzato. Nel caso in cui si scegliesse di analizzare solamente le differenze tra i tipi di semente, la nuova analisi della varianza diventerebbe

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>
<b>Totale</b>	480,16	24	----
<b>Tra sementi</b>	286,16	4	71,54
<b>Errore</b>	194	20	9,7

mentre nel caso in cui si scegliesse le differenze tra concimi si avrebbe

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>
<b>Totale</b>	480,16	24	----
<b>Tra concimi</b>	109,36	4	27,34
<b>Errore</b>	370,80	20	18,54

Nel primo caso, il test F con gdl 4 e 20

$$F_{4,20} = \frac{71,54}{9,7} = 7,38$$

darebbe un valore di 7,38

e nel secondo caso

$$F_{4,20} = \frac{27,34}{18,54} = 1,47$$

sempre per gdl 4 e 20

si otterrebbe un valore di F uguale a 1,47.

**Le differenze tra sementi rimangono ancora significative, ma con un valore di F notevolmente minore; le differenze tra i concimi non risultano più significative, poiché nella varianza d'errore è stata cumulata anche la grande variabilità che esiste tra le medie delle varie specie di sementi a confronto.**

Per queste due analisi risulta più conveniente utilizzare il precedente disegno a due criteri, anche se l'interesse è rivolto solamente ad uno dei due fattori isolati.

**L'efficienza relativa di uno specifico esperimento a quadrati latini rispetto al corrispondente disegno a blocchi randomizzati** o ad un solo criterio di classificazione può essere stimata con la stessa formula usata precedentemente.

Per verificare la significatività delle differenze tra le medie delle sementi, l'efficienza relativa del disegno a quadrati latini rispetto a quello a due criteri, in cui non si considera l'effetto delle arature, è

$$E. R. = \frac{(12 + 1) \cdot (16 + 3) \cdot 5,29}{(16 + 1) \cdot (12 + 3) \cdot 5,57} = \frac{1306,63}{1420,35} = 0,92$$

uguale 0,92.

E' più conveniente fare l'analisi a due soli criteri di quella a tre criteri.

Sempre per saggiare la significatività delle differenze tra le medie delle sementi, l'efficienza relativa del disegno a quadrati latini rispetto a quello ad un solo criterio

$$E. R. = \frac{(12 + 1) \cdot (20 + 3) \cdot 9,7}{(20 + 1) \cdot (12 + 3) \cdot 5,57} = \frac{2900,3}{1754,55} = 1,65$$

è uguale a 1,65.

Per il fattore sementi, con i dati raccolti il disegno a quadrati latini è più efficiente di quello ad un solo criterio. Per ottenere un test con la stessa potenza, nel disegno sperimentale ad un solo criterio servono 1,65 volte i dati utilizzati nell'esperimento a quadrati latini.

## **9.7. POTENZA A PRIORI E A POSTERIORI NELL'ANOVA, CON I GRAFICI DI PEARSON E HARTLEY**

I concetti e i metodi per stimare la potenza di un test, già illustrati nei capitoli precedenti con la distribuzione **z** e la distribuzione **t**, prendono in considerazione un numero più alto di parametri nel **caso di k campioni**, senza per questo diventare molto più complessi.

Per stimare la **potenza di un test F nell'ANOVA**, i metodi proposti negli ultimi decenni sono numerosi. In letteratura è possibile trovare presentazioni ampie e dettagliate, come il volume di J. **Cohen** del 1988 (*Statistical Power Analysis for the Behavioral Sciences*, 2<sup>nd</sup> edition Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 567), che ne riassume il dibattito e i risultati. I

programmi informatici più sofisticati iniziano a presentare anche questa opzione, tra gli output offerti, e forniscono direttamente il valore **1-b** della potenza del test. Ma, sia per comprendere i concetti che stanno alla base di questi metodi, sia per individuare i parametri implicati nella stima della potenza di un test ANOVA, appare ancora utile rifarsi alle carte di E. S. **Pearson** e H. O. **Hartley** del 1951 (*Charts for the power function for analysis of variance tests, derived from the non-central F-distribution*. **Biometrika**, 38:112-130), che utilizzano la **distribuzione F non-centrale**. La presentazione successiva e le tabelle riportate sono tratte da questo articolo.

La distribuzione dei valori **F** per l'ANOVA è individuata da  $n_1$  e  $n_2$ , determinati dai gdl della **varianza tra** e della **varianza entro**, quando l'ipotesi nulla **H<sub>0</sub>** è vera. Ma quando **H<sub>0</sub>** è falsa, il rapporto tra le due varianze assume una forma diversa, detta **distribuzione F non centrale** (*noncentral F-distribution*), che si allontana dalla precedente in funzione della distanza **d**, esistente tra le due o più medie a confronto ( $d = m_{\max} - m_{\min}$ ). Di conseguenza, per identificare una distribuzione non centrale, ai due parametri  $n_1$  e  $n_2$  deve essere aggiunto un terzo, indicato con **f** e detto della **non-centralità** (*noncentrality parameter*), che dipende a sua volta dalla varianza tra trattamenti, dalla varianza d'errore e dai loro gdl.

La potenza (**1-b**) del test **F**, che è collegata alla probabilità **b** di commettere l'errore di **II Tipo** (accettare l'ipotesi nulla **H<sub>0</sub>** quando è falsa), dipende da queste caratteristiche di ogni **distribuzione F non centrale**.

I **grafici di E. O. PEARSON e H. O. HARTLEY**, riportati nelle pagine finali del capitolo, rappresentano una elaborazione delle tabelle scritte da P. C. Tang nel 1938 e sono di più facile lettura ed interpolazione. Sinteticamente le loro caratteristiche fondamentali sono:

- **I grafici sono 8**; servono per stimare la potenza di un test ANOVA con un numero di medie che varia da 2 ad un massimo di 9. Ognuno è identificato dai gdl della varianza tra, indicati con  $n_1$ , che variano da 1 a 8.
- **Ogni grafico riporta 2 famiglie** di curve di potenza, corrispondenti a  $\alpha = 0.05$  e  $\alpha = 0.01$ .
- **Entro ogni famiglia le curve sono 11**, relative ai gdl della varianza d'errore; sono indicate con  $n_2$  e sono state tracciate solo quelle che hanno gdl 6, 7, 8, 9, 10, 12, 15, 20, 30, 60,  $\infty$ . Per altri gdl si deve ricorrere all'interpolazione.
- **Dato un certo valore di f**, riportato in ascissa, ogni curva (individuata appunto da  $n_1$ ,  $\alpha$ ,  $n_2$ ) permette di stimare la potenza **1-b**, riportata in ordinata.
- **La scala di potenza 1-b è logaritmica per b**, in modo da espandere i valori di potenza nella regione più importante e di uso più frequente, tra **0.80** e **0.99**, facilitandone la lettura e l'interpolazione.

- **I valori di f**, riportati in ascissa, dipendono da **a** e hanno un campo di variazione limitato, in quanto determinato dalle potenze relative. Essi variano nei diversi grafici, in quanto influenzate sensibilmente anche dal numero di gruppi a confronto (**n<sub>1</sub>**)

Nell'**analisi della varianza ad un criterio**, il valore di f può essere ottenuto a partire:

- dalle **medie campionarie** quando **i campioni sono bilanciati**  
con

$$f = \sqrt{\frac{n \cdot \sum_{i=1}^p (\bar{x}_i - \bar{\bar{x}})^2}{p \cdot s_e^2}}$$

- dai **risultati dell'analisi della varianza**  
con

$$f = \sqrt{\frac{(p-1) \cdot s_{tra}^2}{p \cdot s_e^2}}$$

dove (per entrambe le formule)

- $\bar{x}_i$  e  $\bar{\bar{x}}$  sono rispettivamente la media di ogni gruppo **i** e la media generale,
- **n** è il numero di dati di ogni gruppo,
- **p** è il numero di gruppi,
- $s_e^2$  è la varianza d'errore
- $s_{tra}^2$  è la varianza tra trattamenti.

(osservare che in entrambe le formule **al numeratore** compare la **devianza tra trattamenti**; nelle formule successive, essa sarà valutata attraverso **la differenza massima** d tra le **p** medie a confronto).

Stimato f, per calcolare la potenza **1-b** (o derivare da essa la probabilità **b**) mediante le curve di probabilità, è necessario determinare:

- **n<sub>1</sub>** che indica **il grafico** da scegliere,
- **a** (uguale a **0.05** oppure a **0.01**, poiché questi grafici non permettono altre opzioni) che indica la riga (tra le due riportate sulle ascisse) nella quale **individuare il punto corrispondente al valore di f e il blocco di curve** entro il grafico,

- $n_2$  (i gdl della varianza d'errore) che sulle 2 **ordinate** esterne indicano **la curva** (approssimata a partire da 11) dalla quale risalire al valore della potenza **1-b**, partendo da quello di  $f$  riportato in **ascissa**.

#### ESEMPIO 1

In tre zone di una città, sono state effettuate 5 misurazioni del livello di benzopirene (ng/mc) presente nell'aria, con i seguenti valori medi

CAMPIONI	A	B	C
$n_i$	5	5	5
$\bar{X}_i$	2,652	2,195	2,496

L'analisi della varianza ad un criterio, della quale si riportano i risultati,

DEVIANZA		n	$s^2$	F	P
<b>TOTALE</b>	1,69	14	----	----	----
<b>Tra trattamenti</b>	0,54	<b>2</b>	0,27	2,81	<b>0,20</b>
<b>Errore</b>	1,15	<b>12</b>	0,096	----	----

non ha permesso di rifiutare l'ipotesi nulla. Infatti

- il valore critico di F, con gdl **2** per il numeratore e **12** per il denominatore, per la probabilità  $\alpha = 0.05$  è **3,89** mentre il valore calcolato è **2,81**;
- in modo più preciso, il valore calcolato (**2,81**) corrisponde ad una probabilità  $\alpha$  uguale a **0,20** per un test bilaterale.

Domande: sulla base di questi dati, stimare la potenza (**1-b**) del test, per la significatività ad una probabilità

a)  $\alpha = 0.05$

b)  $\alpha = 0.01$

Risposte

Con la formula che utilizza le medie in campioni bilanciati

$$\phi = \sqrt{\frac{n \cdot \sum_{i=1}^p (\bar{x}_i - \bar{\bar{x}})^2}{p \cdot s_e^2}}$$

dove

-  $\bar{x}_A = 2,652$ ;  $\bar{x}_B = 2,195$ ;  $\bar{x}_C = 2,496$  dalle quali è stato ricavato

-  $\bar{\bar{x}} = 2,448$  e con

-  $s_e^2 = 0,096$ ;  $n = 5$ ;  $p = 3$

si ottiene

$$\phi = \sqrt{\frac{5 \cdot [(2,652 - 2,448)^2 + (2,195 - 2,448)^2 + (2,496 - 2,448)^2]}{3 \cdot 0,096}}$$

$$\phi = \sqrt{\frac{5 \cdot (0,0416 + 0,0640 + 0,0023)}{3 \cdot 0,096}} = \sqrt{\frac{0,5395}{0,288}} = 1,37$$

un valore di  $f$  uguale a **1,37**.

Con la formula che ricorre ai risultati dell'ANOVA

$$f = \sqrt{\frac{(p-1) \cdot s_{tra}^2}{p \cdot s_e^2}}$$

dove

-  $s_{tra}^2 = 0,27$

-  $s_e^2 = 0,096$ ;  $p = 3$

si ottiene

$$f = \sqrt{\frac{(3-1) \cdot 0,27}{3 \cdot 0,096}} = \sqrt{\frac{0,54}{0,288}} = 1,37$$

un valore di  $f$  uguale a 1,37 (ovviamente identico al precedente).

Successivamente, mediante il grafico, si passa dal valore di  $f$  a quello di **1-b**, in modo indipendente per le due probabilità  $\alpha = 0,05$  e  $\alpha = 0,01$ .

a) Per  $\alpha = 0.05$ , con

- $n_1 = 2$  (secondo grafico),
- il valore di  $f$  uguale a **1,37** letto nella riga per  $\alpha = 0.05$  (riga superiore nell'ascissa e blocco di curve a sinistra), all'incrocio con
- $n_2 = 12$ , fornisce un valore di  $1-b$  uguale a **circa 0,45** (difficile da leggere con precisione per l'addensarsi del fascio di curve).

b) Per  $\alpha = 0.01$ , con

- $n_1 = 2$  (stesso grafico),
- il valore di  $f$  uguale a **1,37** letto nella riga per  $\alpha = 0.01$  (riga inferiore nell'ascissa e blocco di curve a destra)
- con  $n_2 = 12$  fornisce un valore di  $1-b$  uguale a **circa 0,20**.

In conclusione, l'analisi della potenza a posteriori del test con l'ANOVA permette di affermare che

- esisteva una probabilità  $b$  del 55% di non rifiutare (erroneamente) l'ipotesi nulla, scegliendo un livello di significatività  $\alpha = 0.05$ ;
- esisteva una probabilità  $b$  dell'80 % di non rifiutare (erroneamente) l'ipotesi nulla, scegliendo un livello di significatività  $\alpha = 0.01$ .

Come hanno evidenziato i concetti illustrati con i due esempi e come mostra la formula generale sotto-riportata,

$$\phi = \sqrt{\frac{n \cdot d^2}{2p \cdot s_e^2}}$$

dove

$\frac{n \cdot d^2}{2}$  corrisponde alla devianza tra trattamenti

la potenza  $1-b$  di un test, associata al valore di  $f$ ,

- diminuisce con la riduzione di  $\alpha$ , la probabilità di commettere l'errore di I Tipo,
- aumenta al crescere di  $n$ , il numero di repliche per campione,
- aumenta al crescere della differenza  $d$  esistente tra le medie,
- diminuisce al crescere del numero  $p$  di gruppi a confronto,
- diminuisce al crescere della varianza d'errore  $s_e^2$ .

Utilizzando i risultati precedenti come se fossero quelli di uno **studio pilota**, effettuato solo per programmare correttamente un successivo esperimento, quello che realmente interessa, è possibile

l'analisi della **potenza a priori**. E' la base del disegno sperimentale, insieme con il tipo di campionamento.

Per le relazioni che esistono tra i vari parametri, è possibile rispondere ad alcune domande. Per la programmazione di un esperimento al quale si pensa di applicare l'analisi della varianza ad un criterio, le più importanti sono 3:

- la **differenza minima** ( $d$ ) che sarà possibile evidenziare;
- il **numero minimo di dati** ( $n$ ) che è necessario raccogliere per ogni gruppo, o il numero totale ( $N$ ) di dati, ovviamente in campioni bilanciati;
- il **numero massimo di gruppi** ( $p$ ) che è possibile formare con il numero totale ( $N$ ) prefissato.

La **differenza minima significativa** ( $d$ ) può essere dedotta direttamente attraverso le relazioni matematiche

$$d = \sqrt{\frac{2ps_e^2 f^2}{n}}$$

dopo avere

- stimato  $p$ ,  $s_e^2$ ,  $n$  dall'esperimento pilota e
- identificato il valore richiesto di  $f$  nel grafico, sulla base dei 4 parametri  $n_1$ ,  $n_2$ ,  $a$ ,  $1-b$ .

Il **numero minimo di dati** ( $n$ ) che è necessario raccogliere e il **numero massimo di gruppi** ( $p$ ) che è possibile formare sono ricavabili direttamente attraverso il grafico.

I metodi, con tutti i loro passaggi logici, sono illustrati nel successivo esempio 3.

### ESEMPIO 3

Riprendendo i risultati discussi nell'esempio 1, si chiede: "Per ottenere la significatività alla probabilità  $a = 0.05$  e con una potenza  $1-b = 0.90$  (la potenza del test è a discrezione del ricercatore, ma spesso si sceglie  $1-b = 0.80$ , per la relazione standard di 1 a 4 della probabilità  $b$  rispetto ad  $a$ )

- a) quale era la **differenza minima** ( $d$ ) che era possibile evidenziare con il campione raccolto (quindi mantenendo invariati  $p$ ,  $n$ ,  $s_e^2$ ) ?
- b) quale dovrebbe essere il **numero minimo di dati** ( $N$ ) da raccogliere, ovviamente da distribuire in campioni bilanciati (predeterminando  $d$  e mantenendo invariati  $p$ ,  $s_e^2$ ) ?;
- c) quale è il **numero massimo di gruppi** ( $p$ ) che è possibile formare (predeterminando  $N$ ,  $d$  e mantenendo invariato  $s_e^2$ ) ?

Risposte

- a) Per stimare la **differenza minima**  $d$  che è possibile dimostrare significativa, con

- $n = 5$ ;  $p = 3$ ,  $s_e^2 = 0,096$  come nello studio pilota,
- dapprima si richiede che dal grafico sia ricavato il valore di  $f$  individuato da
- $n_1 = 2$ ;  $a = 0.05$ ;  $1-b = 0,90$ ;  $n_2 = 12$ ;
- che risulta uguale a circa **2,30**;
- successivamente si calcola  $d$ ,

$$d = \sqrt{\frac{2 \cdot 3 \cdot 0,096 \cdot 2,3^2}{5}} = \sqrt{\frac{3,047}{5}} = \sqrt{0,6094} = \mathbf{0,78}$$

che risulta uguale a **0,78**.

Nello studio pilota poteva risultare significativa, ai livelli  $a$  e  $b$  prescelti, una differenza superiore a **0,78** tra media maggiore e media minore. Infatti il test non è risultato significativo, poiché la differenza massima tra le tre medie era  $(2,652 - 2,195)$  uguale a **0,457**.

b) Per stimare il **numero minimo di dati (N)** che è necessario raccogliere, sempre con

$$a = 0.05; \quad 1-b = 0.90; \quad p = 3, \quad s_e^2 = 0,096$$

- dapprima si deve definire **la differenza minima  $d$**  che si intende verificare come significativa: è la scelta più **strettamente dipendente dalla discrezionalità del ricercatore** e influenza molto le dimensioni del campione che sarà necessario raccogliere; è bene scegliere il valore di  $d$  sulla base del suo **significato biologico**, come già evidenziato in paragrafi precedenti (nell'esempio, si assume come importante una differenza minima  $d = 0,5$ )
- successivamente si deve agire per tentativi, ipotizzando un numero  $n$  ragionevole per ogni campione (ad esempio,  $n = 8$ ; superiore ai 5 dell'esperimento pilota, non significativo)
- quindi stimare il valore di  $n_2$  che deriva da questa scelta ( $n_2 = p \cdot (n - 1) = 3 \cdot (8 - 1) = 21$ )
- e calcolare il valore di  $f$

$$f = \sqrt{\frac{n \cdot d^2}{2p \cdot s_e^2}}$$

che deriva dalle scelte appena effettuate

$$f = \sqrt{\frac{8 \cdot 0,5^2}{2 \cdot 3 \cdot 0,096}} = \sqrt{\frac{2,0}{0,576}} = 1,86$$

- Infine, utilizzando il grafico  $n_1 = 2$  (poiché  $p = 3$ )
- si stima la potenza  $1-b$  di  $f = 1,86$  per  $a = 0.05$  e  $n_2 = 21$ .

- Nel grafico, il valore di **1-b** risulta circa 0,80. E' troppo basso rispetto alla potenza richiesta.

Non resta che aumentare **n**, ipotizzando ad esempio **n = 12** per un secondo tentativo, mantenendo costanti tutti gli altri parametri

- Il nuovo valore di **f**

$$f = \sqrt{\frac{12 \cdot 0,5^2}{2 \cdot 3 \cdot 0,096}} = \sqrt{\frac{3,0}{0,576}} = \mathbf{2,28}$$

risulta uguale a **2,28** con  $n_2 = 33 = [3 \cdot (12 - 1)]$

- Sempre nel grafico  $n_1 = 2$  si stima la potenza **1-b** di  $f = 2,28$  per  $a = 0.05$  e  $n_2 = 33$ .
- Il valore di **1-b** risulta circa **0,92**.

Si avvicina a quanto ipotizzato, superandolo leggermente: quindi, si può concludere che servono almeno 11-12 dati per gruppo, in totale tra 33 e 36 distribuiti in modo bilanciato.

- c) Per **stimare il numero massimo p di gruppi** che si possono formare, è necessario ricordare che all'aumentare del numero **p** di gruppi la potenza del test diminuisce, anche mantenendo costante il numero **n** di dati in ogni gruppo (quindi aumentando N).

Supponendo di poter effettuare in totale 60 osservazioni (**N = 60**), per stimare il numero massimo **p** di gruppi che si possono formare, mantenendo invariati

- $d = 0,5$ ;  $a = 0.05$ ;  $1-b = 0.90$ ;  $s_e^2 = 0,096$

si procede per tentativi, in modo analogo a quanto fatto nella stima di **n**.

- Dapprima, essendo 60 superiore al numero 36 stimato in precedenza, si supponga di voler formare 4 gruppi (**p = 4**), per cui **n = 15**;
- Successivamente con

$$f = \sqrt{\frac{n \cdot d^2}{2p \cdot s_e^2}}$$

si stima il valore di **f**

$$f = \sqrt{\frac{15 \cdot 0,5^2}{2 \cdot 4 \cdot 0,096}} = \sqrt{\frac{3,75}{0,768}} = \mathbf{2,21}$$

che risulta uguale a **2,21**.

- Con il grafico  $n_1 = 3$  (poiché  $p = 4$ )
- si stima la potenza **1-b** di  $f = 2,21$  per  $a = 0.05$  e  $n_2 = 56 = (60 - 4)$ .
- Nel grafico, il valore di **1-b** risulta superiore al **0.96**.

E' molto alto, superiore alla potenza richiesta di 0.90. E' quindi possibile pensare di formare 5 gruppi, se l'estensione delle zone studiate è importante ai fini della ricerca.

Ipotizzando  $p = 5$ , con  $N$  costante (**60**) si ha  $n = 12$ . Mantenendo costanti tutti gli altri parametri

- il nuovo valore di  $f$

$$f = \sqrt{\frac{12 \cdot 0,5^2}{2 \cdot 5 \cdot 0,096}} = \sqrt{\frac{3,0}{0,96}} = 1,77$$

risulta uguale a **1,77**

- Successivamente, si stima la potenza **1-b** di  $f = 1,77$  per  $a = 0.05$  e  $n_2 = 55$  (**60 - 5**).
- Nel grafico  $n_1 = 4$  (poiché  $p = 5$ ) il valore di **1-b** risulta minore di **0,90**.

Poiché la potenza di quella seconda ipotesi è minore di quella prescelta, si può concludere che con 60 osservazioni il numero massimo di gruppi che si possono formare è 4, se la potenza del test non deve essere inferiore a 0.90.

Sempre nell'articolo di E. S. **Pearson** e H. O. **Hartley** del 1951 (*Charts for the power function for analysis of variance tests, derived from the non-central F-distribution. Biometrika*, 38:112-130), il metodo presentato è esteso a vari disegni dell'analisi della varianza:

- due criteri di classificazione, con **una** osservazione per cella,
- due criteri di classificazione, con  $n_{ij}$  osservazioni per cella,
- quadrati latini.

Nel caso di **due criteri di classificazione con una sola osservazione per cella**, i metodi illustrati non variano, ricordando che possono essere applicati indipendentemente per ognuno dei due fattori, in modo analogo al test F per la significatività. Di conseguenza, il valore di  $f$  è calcolato nello stesso modo

$$f = \sqrt{\frac{(p-1) \cdot s_{tra}^2}{p \cdot s_e^2}}$$

ricordando che

- $s_{tra}^2$  è la varianza del fattore che interessa, con  $p$  medie a confronto e  $n_1$  gdl (per la scelta del grafico)
- $s_e^2$  è la varianza d'errore e  $n_2$  sono i suoi gdl (per la scelta della curva).

Come per la significatività, la potenza del test per i due fattori sarà quasi sempre differente, essendo determinata dalla loro **varianza tra** (quindi dalla differenza tra le medie) e dal diverso numero di dati.

Nel caso di **due criteri di classificazione con n<sub>ij</sub> osservazione per cella**, occorre distinguere tra

- effetti principali e
- interazione.

Per gli effetti principali, il metodo è identico a quello dell'analisi a due criteri con una sola osservazione.

Per l'interazione, il cui la varianza relativa dipende dagli scarti tra le medie osservate e le medie attese sulla base delle medie di riga e di colonna,

il valore di f è

$$f = \sqrt{\frac{(DF \text{ interazione}) \cdot s_{\text{interazione}}^2}{(DF \text{ interazione} + 1) \cdot s_e^2}}$$

Con i **quadrati latini**, si ritorna al caso di due criteri con una sola osservazione, prendendo in considerazione per ognuno dei tre fattori la varianza e i gdl relativi.

Questo metodo è facilmente estensibile a disegni sperimentali più complessi, con tre o più fattori e con repliche, purché bilanciati.

Per la stima **a priori**

- delle dimensioni N del campione,
- della differenza minima d che è si vuole rendere significativa,
- del numero massimo di livelli (p) di uno specifico fattore,

i metodi sono uguali a quelli presentati, applicando gli accorgimenti appena descritti.

## **9.8. APPENDICE AL CAPITOLO :**

### **LETTURA DI TABULATI DELL'ANALISI DELLA VARIANZA EFFETTUATA CON UN PACCHETTO STATISTICO**

Gli scopi di un corso e di un testo di statistica applicata di solito sono 4:

- 1) fornire una **chiara visione dei principi** che sono alla base dell'analisi e dell'inferenza statistica;
- 2) guidare alla **scelta dei test più appropriati**, per i problemi specifici che si affrontano o per le domande alle quali si vuole rispondere con la ricerca ed i dati raccolti;
- 3) condurre all'**applicazione esatta delle formule dei test**, con una impostazione del calcolo metodologicamente appropriata;
- 4) indirizzare alla **corretta interpretazione dei risultati**.

Come negli esercizi già svolti, i risultati possono essere ottenuti con il calcolo manuale quando i dati sono pochi e la formula da applicare è relativamente semplice. Con molti dati o con calcoli che richiedono lunghe elaborazioni, la probabilità di errore è sempre alta e la possibilità di verifiche è limitata, ridotta alla pura ripetizione dei calcoli, eventualmente con altra formula. Per ottenere risultati senza errori di calcolo, non esiste alcuna alternativa pratica all'uso di pacchetti statistici, ormai consolidato dall'ampia diffusione dei personal computer e dai costi relativamente bassi.

Di norma, questi programmi cercano di fornire risposte a tutte le domande che è possibile porre o ai vari casi che è possibile esaminare con l'uso di un determinato test. Essi sono molto più numerosi di quanto abitualmente serve al ricercatore in un caso specifico; spesso i test offerti sono tra loro alternativi, come nel caso dei confronti multipli quando sono forniti i risultati del test di Tukey, di Scheffé, di Dunnett e vari altri analoghi, mentre il problema quasi sempre richiede l'uso specifico di uno solo di essi.

Per un impiego corretto e funzionale di un pacchetto statistico, è divenuto importante **imparare a scegliere le sole risposte utili**, tra la grande quantità di analisi e di informazioni fornite.

Come esercitazione all'uso corretto dell'analisi della varianza con un pacchetto statistico, sono stati svolti alcuni degli esempi già utilizzati. Per l'interpretazione, prima dei risultati è utile rileggere il problema con i dati relativi.

L'output fornito dal computer è stato inserito nel testo senza alcuna modifica sostanziale. Per istruire ricercatori ed operatori dell'ambiente alla loro lettura, è stato fornito un breve commento delle parti di output racchiuse nei riquadri numerati, che rappresentano le informazioni utili per rispondere in modo completo e dettagliato alle domande dell'esercizio, in accordo con gli argomenti fino ad ora sviluppati.

## **ESEMPIO sulle differenze d'inquinamento tra zone del capitolo 7.**

### **analisi della varianza ad un criterio di classificazione**

Analysis of Variance Procedure

**1)**

Class	Levels	Values
ZONE	3	1 2 3

Number of observations in data set = 15

**2)**

Dependent Variable: FERRO

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.50293667	0.25146833	2.54	0.1204
Error	12	1.18890333	0.09907528		
Corrected Total	14	1.69184000			

R-Square	C.V.	Root MSE	FERRO Mean
	0.297272	12.70227	0.314762
			2.47800000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
ZONE	2	0.50293667	0.25146833	2.54	0.1204

Duncan's Multiple Range Test for variable: FERRO

NOTE: This test controls the type I comparisonwise error rate,  
not the experimentwise error rate

Alpha= 0.05 df= 12 MSE= 0.099075

WARNING: Cell sizes are not equal.

Harmonic Mean of cell sizes= 4.864865

Number of Means	2	3
Critical Range	0.439	0.460

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	ZONE
A	2.652	6	1
A			
A	2.496	5	3
A			
A	2.195	4	2

3)

Tukey's Studentized Range (HSD) Test for variable: FERRO

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 12 MSE= 0.099075  
Critical Value of Studentized Range= 3.773  
Minimum Significant Difference= 0.5384  
WARNING: Cell sizes are not equal.  
Harmonic Mean of cell sizes= 4.864865

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	ZONE
A	2.652	6	1
A			
A	2.496	5	3
A			
A	2.195	4	2

Scheffe's test for variable: FERRO

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than REGWF for all pairwise comparisons

Alpha= 0.05 df= 12 MSE= 0.099075  
Critical Value of F= 3.88529  
Minimum Significant Difference= 0.5626  
WARNING: Cell sizes are not equal.  
Harmonic Mean of cell sizes= 4.864865

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	ZONE
A	2.652	6	1
A			
A	2.496	5	3
A			
A	2.195	4	2

1) E' un'analisi della varianza ad un criterio di classificazione (ZONE); sono presenti 3 livelli per un totale di 15 dati.

2) La devianza totale ha 14 d.f; quella per il confronto tra le medie 2 d.f. e quella d'errore 12 d. f. Il test  $F_{(2,12)}$  con d.f. 2 e 12 non risulta significativo (probabilità superiore al 12%). Dopo le informazioni che riguardano la variabilità dei dati e la media generale, è riportata la scomposizione delle devianze tra i vari fattori in esame. Poiché si tratta di un'analisi ad un solo criterio di

classificazione, l'informazione riportata è identica a quella già stampata nella parte superiore del riquadro.

Benché per il confronto tra le quantità medie di polvere di ferro campionate nelle tre zone il test F non risulti significativo nemmeno alla probabilità 0.05 (ma il tabulato riporta solo la probabilità esatta ( $Pr > 0.1204$ )) sono di seguito stampati i confronti tra le tre medie, con vari test. Il primo è il test di Duncan, che nei paragrafi precedenti non è stato spiegato. Ovviamente conviene utilizzare quelli noti, come il test di Tukey e il test di Scheffé riportati nel riquadro 3.

**3)** Le informazioni per i confronti tra le tre medie sono fornite dal test di Tukey e dal test di Scheffé. Il test di Tukey per confronti singoli, alla probabilità  $\alpha = 0.05$ , con  $s_e^2 = 0,099075$ , 12 d.f. ed un valore critico del Q studentizzato uguale a 3,773 fornisce una differenza minima significativa tra le tre medie uguale a 0,5324. Il tabulato ricorda pure che occorre fare attenzione (WARNING): i tre gruppi a confronto non hanno lo stesso numero di osservazioni, come richiesto dal test. Di conseguenza, è stata stimata la media armonica del numero di osservazioni presenti in ogni gruppo, che risulta uguale a 4,86. Di seguito, in ordine decrescente sono riportate le medie con il relativo numero di osservazioni e l'indicazione della ZONA di rilevazione. Poiché le differenze tra loro sono tutte inferiori al valore della differenza minima significativa, le tre medie sono state indicate con la stessa lettera, con una rappresentazione grafica dei gruppi molto semplice.

Segue il test di Scheffé che, alla stessa probabilità (0.05) (ovviamente per un valore identico della varianza d'errore e dei df) con un valore di F pari a 3,885 fornisce una differenza minima significativa uguale a 0,5626. Anche il test di Scheffé occorre fare attenzione, poiché i tre gruppi non hanno lo stesso numero d'osservazioni. La differenza minima significativa risulta un valore superiore a quella stimata con il test di Tukey, in quanto permette pure i confronti multipli. Anche per il test di Scheffé è riportata la rappresentazione grafica dei tre gruppi, con la media relativa in ordine decrescente, il numero di osservazioni di ogni gruppo e la zona di rilevazione.

## ESEMPIO su blocchi e trattamenti del cap. 8

### disegno a blocchi randomizzati o a due criteri di classificazione

Analysis of Variance Procedure

1)	Class	Levels	Values
	BLOCCHI	4	1 2 3 4
	TRATTAM	5	1 2 3 4 5

Number of observations in data set = 20

**2)**

Dependent Variable: VY

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	654.3000000	93.4714286	39.08	0.0001
Error	12	28.7000000	2.3916667		
Corrected Total	19	683.0000000			

R-Square      C.V.      Root MSE      VY Mean  
 0.957980      5.426321      1.546501      28.5000000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
BLOCCHI	3	525.8000000	175.2666667	73.28	0.0001
TRATTAM	4	128.5000000	32.1250000	13.43	0.0002

**3)**

Tukey's Studentized Range (HSD) Test for variable: VY

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05    df= 12    MSE= 2.391667  
 Critical Value of Studentized Range= 4.199  
 Minimum Significant Difference= 2.9038

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	BLOCCHI
A	33.800	5	4
A	32.800	5	2
B	26.200	5	1
C	21.200	5	3

Scheffe's test for variable: VY

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWF for all pairwise comparisons

Alpha= 0.05    df= 12    MSE= 2.391667  
 Critical Value of F= 3.49029  
 Minimum Significant Difference= 3.165

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	BLOCCHI
A	33.800	5	4
A	32.800	5	2
B	26.200	5	1
C	21.200	5	3

4)

Tukey's Studentized Range (HSD) Test for variable: VY

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 12 MSE= 2.391667  
 Critical Value of Studentized Range= 4.508  
 Minimum Significant Difference= 3.4857

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	TRATTAM
A	32.750	4	3
A			
B A	30.000	4	1
B			
B C	27.250	4	2
C			
C	26.500	4	4
C			
C	26.000	4	5

Scheffe's test for variable: VY

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than REGWF for all pairwise comparisons

Alpha= 0.05 df= 12 MSE= 2.391667  
 Critical Value of F= 3.25917  
 Minimum Significant Difference= 3.9484

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	TRATTAM
A	32.750	4	3
A			
B A	30.000	4	1
B			
B C	27.250	4	2
B			
B C	26.500	4	4
C			
C	26.000	4	5

1) E' un'analisi della varianza a 2 criteri di classificazione (chiamati BLOCCHI e TRATTAMENTI). I blocchi hanno 4 livelli e i trattamenti 5 livelli, per un totale di 20 osservazioni.

2) La devianza totale ha 19 d.f. dei quali 7 per confronti tra medie e 12 per l'errore o residuo. Il test F complessivo per il confronto tra medie risulta altamente significativo (probabilità inferiore a 0.0001). Dopo le informazioni sulla variabilità e la media generale è riportato il test F sia per le 4 medie dei blocchi (con 3 d.f.) sia per le 5 medie dei trattamenti (con 4 d.f.): sono entrambe altamente significative. Le differenze tra le medie dei blocchi risultano più significative (probabilità inferiore a 0.0001) di quelle tra le medie dei trattamenti (probabilità inferiore a 0.0002). Il valore di  $R^2$  (R square) pari a 0.9579 sta ad indicare che i due fattori noti (blocchi e trattamenti) assorbono il 95,79% della variabilità totale ( $654.3 / 683 = 0,9579$ ).

3) Per il confronto tra le 4 medie dei blocchi il test di Tuckey e quello di Scheffé danno i medesimi risultati, anche se la differenza minima significativa di Tuckey (2,9038), come più volte spiegato, è leggermente minore di quella di Scheffé (3,165). Tra le 4 medie risultano non significativamente diverse le due medie maggiori (33,8 e 32,8); rispetto a queste due sono significativamente differenti le altre due medie (26,2 e 21,2), che risultano significativamente differenti anche tra loro.

Come imposto dallo schema a due criteri di classificazione, le medie hanno lo stesso numero di osservazioni; di conseguenza, non è calcolata la loro media aritmetica.

4) Per le differenze tra le medie dei trattamenti, il test di Tukey e quello di Scheffé non danno gli stessi risultati. All'aumentare del numero di medie e quindi dei possibili confronti semplici e complessi, cresce il divario tra le due diverse stime della differenza minima significativa (con 5 medie Tukey e uguale 3,4858 e Scheffé a 3,9484). Sempre alla probabilità 0.05 secondo il test di Tukey la media del trattamento 3 non è significativamente diversa da quella del trattamento 1, mentre è differente dalle altre 3 medie; la media del gruppo 1 (la seconda in ordine decrescente) non è diversa da quella del gruppo 3 né da quella del trattamento 2, ma è differente da quelle dei trattamenti 4 e 5; la media del trattamento 2 non differisce significativamente da quella del trattamento 1, mentre diverge da quella degli altre 3 gruppi. Le medie dei gruppi 4 e 5 non divergono tra loro né dal gruppo 2, ma differiscono in modo significativo dalla media del gruppo 3 e da quella del gruppo 1.

Con il test di Scheffé si ha una differenza per la media del gruppo 4 che non divergerebbe in modo significativo non solo da quella del gruppo 5 e del gruppo 2, ma nemmeno da quella del gruppo 1.

La rappresentazione grafica riesce a fornire la stessa informazione in modo più chiaro e sintetico.

## ESEMPIO su aratura, concime e semente del cap. 8

### quadrati latini

Analysis of Variance Procedure

1)

Class	Levels	Values
ARATURA	5	1 2 3 4 5
CONCIME	5	1 2 3 4 5
SEMENTE	5	1 2 3 4 5

Number of observations in data set = 25

2)

Dependent Variable: VY

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	413.2800000	34.4400000	6.18	0.0018
Error	12	66.8800000	5.5733333		
Corrected Total	24	480.1600000			

R-Square	C.V.	Root MSE	VY Mean
0.860713	4.861596	2.360791	48.5600000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
ARATURA	4	17.7600000	4.4400000	0.80	0.5498
CONCIME	4	109.3600000	27.3400000	4.91	0.0141
SEMENTE	4	286.1600000	71.5400000	12.84	0.0003

3)

Tukey's Studentized Range (HSD) Test for variable: VY

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 12 MSE= 5.573333  
Critical Value of Studentized Range= 4.508  
Minimum Significant Difference= 4.7592

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	ARATURA
A	50.000	5	4
A	49.000	5	2
A	48.200	5	5
A	47.800	5	3
A	47.800	5	1

Scheffe's test for variable: VY

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than REGWF for all pairwise comparisons

Alpha= 0.05 df= 12 MSE= 5.573333  
Critical Value of F= 3.25917  
Minimum Significant Difference= 5.391

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	ARATURA
A	50.000	5	4
A	49.000	5	2
A	48.200	5	5
A	47.800	5	3
A	47.800	5	1

4)

Tukey's Studentized Range (HSD) Test for variable: VY

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 12 MSE= 5.573333  
Critical Value of Studentized Range= 4.508  
Minimum Significant Difference= 4.7592

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	CONCIME
A	51.600	5	3
A			
B A	49.600	5	2
B A			
B A	49.200	5	4
B			
B	46.200	5	1
B			
B	46.200	5	5

Scheffe's test for variable: VY

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than REGWF for all pairwise comparisons

Alpha= 0.05 df= 12 MSE= 5.573333  
Critical Value of F= 3.25917  
Minimum Significant Difference= 5.391

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	CONCIME
A	51.600	5	3
A			
B A	49.600	5	2
B A			
B A	49.200	5	4
B			
B	46.200	5	1
B			
B	46.200	5	5

5)

Tukey's Studentized Range (HSD) Test for variable: VY

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 12 MSE= 5.573333  
Critical Value of Studentized Range= 4.508  
Minimum Significant Difference= 4.7592

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	SEMEN
A	53.400	5	2
A			
B A	51.400	5	4
B			
B C	47.200	5	3
B C			
B C	46.800	5	5
C			
C	44.000	5	1

Scheffe's test for variable: VY

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than REGWF for all pairwise comparisons

Alpha= 0.05 df= 12 MSE= 5.573333  
Critical Value of F= 3.25917  
Minimum Significant Difference= 5.391

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	SEMEN
A	53.400	5	2
A			
B A	51.400	5	4
B			
B C	47.200	5	3
B C			
B C	46.800	5	5
C			
C	44.000	5	1

1) E' un'analisi della varianza a tre criteri di classificazione, chiamati ARATURA, CONCIME e SEMEN. Ha cinque livelli (1, 2, 3, 4, 5) per ogni criterio e 25 osservazioni in totale, come richiesto per un'analisi a quadrati latini.

2) La devianza dei fattori analizzati ha in totale 12 df, quella d'errore altri 12 per un totale di 24. Complessivamente, le fonti di variazione risultano altamente significative con un valore di F uguale a 6,18 per d. f. 12 al numeratore e 12 al denominatore, corrispondenti ad una probabilità di 0.0018. La media generale delle 25 osservazioni è 48,56 e la sua deviazione standard (la radice quadrata della varianza d'errore) è uguale a 2,36. Il coefficiente di variazione è pertanto limitato (4,86%) e i tre fattori considerati assorbono l'86% della variabilità totale ( $R^2 = 0.86$ ); la variabilità residua è così ridotta (14%), da lasciare intendere che oltre ai tre considerati non esistono altri fattori in grado di influire in modo sostanziale sui valori delle osservazioni. Fra i tre fattori considerati, le differenze tra le medie dei diversi tipi di aratura non risultano assolutamente significative ( $Pr = 0.5498$ ); sono invece molto significative le differenze tra le medie dei concimi ( $Pr = 0.0141$ ) e soprattutto risultano altamente significative quelle tra le medie delle sementi ( $Pr = 0.0003$ ).

3) Sono evidenziate in modo dettagliato le differenze tra le medie dei cinque tipi di aratura. Il loro F complessivo, riportato nel riquadro precedente, non risulta assolutamente significativo. Per la significatività è quindi da ritenere inutile questa analisi più dettagliata; tuttavia, essa viene quasi sempre riportata nei tabulati dei programmi informatici, per favorire una valutazione più precisa ed ugualmente importante delle differenze campionarie.

A ulteriore conferma della non significatività delle differenze tra queste cinque medie, il test di Tukey stima che alla probabilità 0.05 la differenza minima significativa è uguale a 4,7592 mentre la differenza tra la media con il valore massimo (la 4ª con 50.0) e quella con il valore minimo (la 1ª con 47.8) è uguale a 2,2.

La rappresentazione grafica esprime questi concetti riunendo le medie dei cinque tipi di aratura in un solo gruppo.

Con il test di Scheffè a maggior ragione non si evidenzia alcuna differenza significativa, essendo ancor più conservativo del test precedente.

4) Sono descritte in modo particolareggiato le differenze tra le medie dei cinque tipi di concime. Poiché la varianza d'errore ed il numero di medie a confronto sono identiche a quelle utilizzate per il confronto tra i 5 tipi di aratura, la differenza minima significativa è uguale a quella stimata nel riquadro precedente, per ognuno dei due test riportati (4,7592 per il test di Tukey e 5,391 per quello di Scheffè).

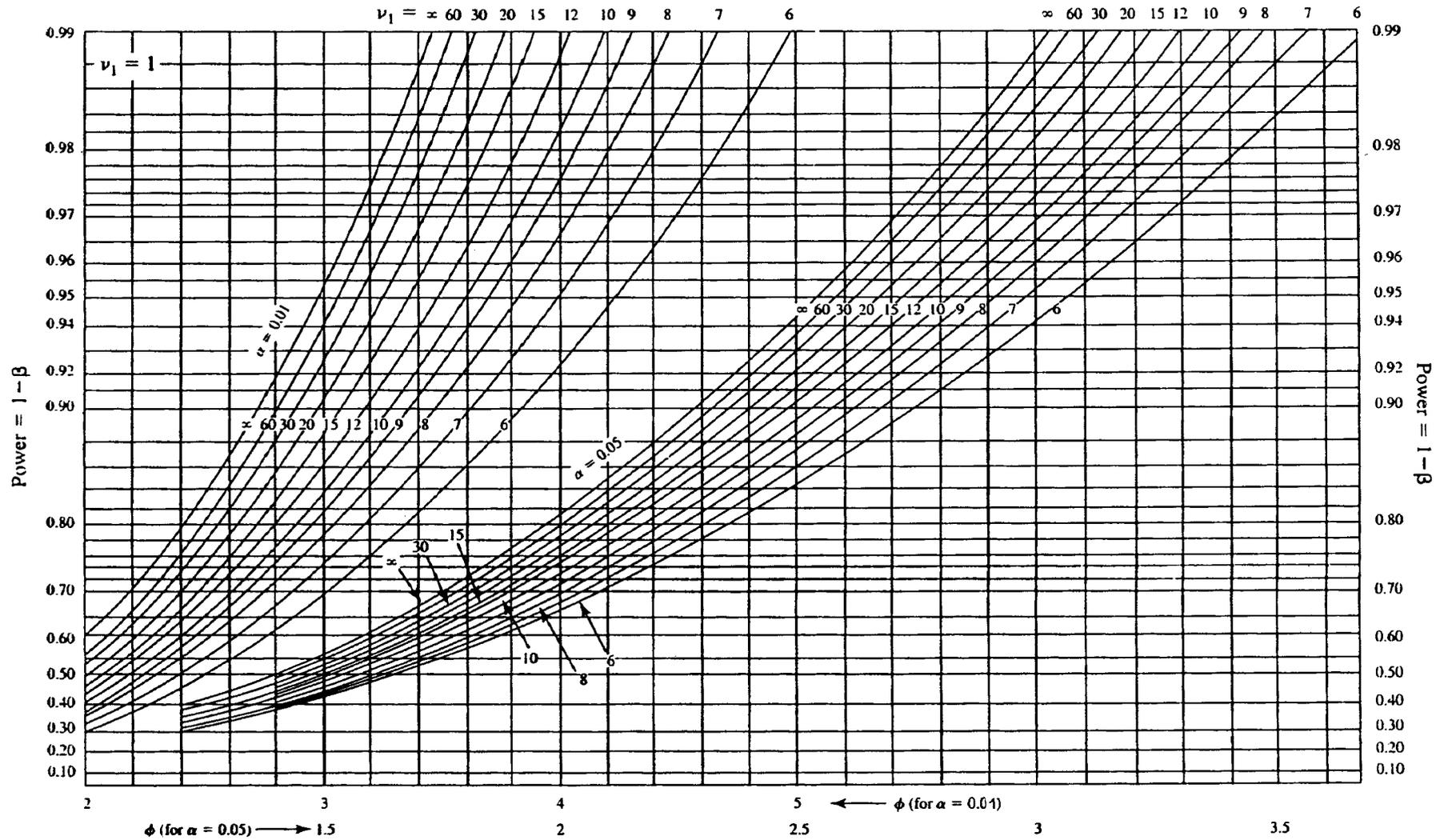
Con il test di Tukey, secondo la rappresentazione grafica la significatività complessiva è dovuta alla differenza del gruppo 3 (quello con la media maggiore) dai gruppi 1 e 5 (con le due medie minori).

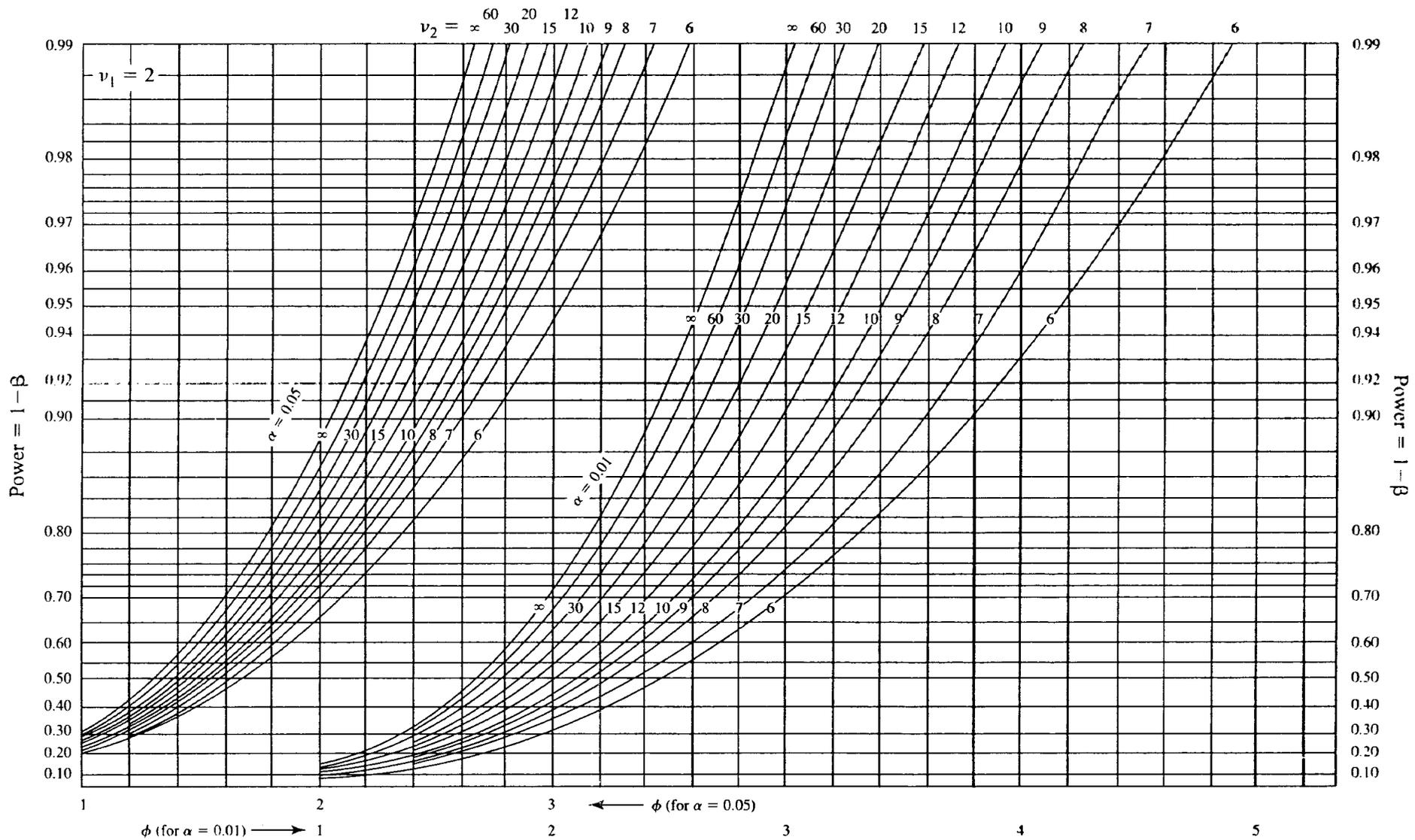
Il test di Scheffè arriva alle stesse conclusioni e pertanto fornisce la stessa risposta grafica.

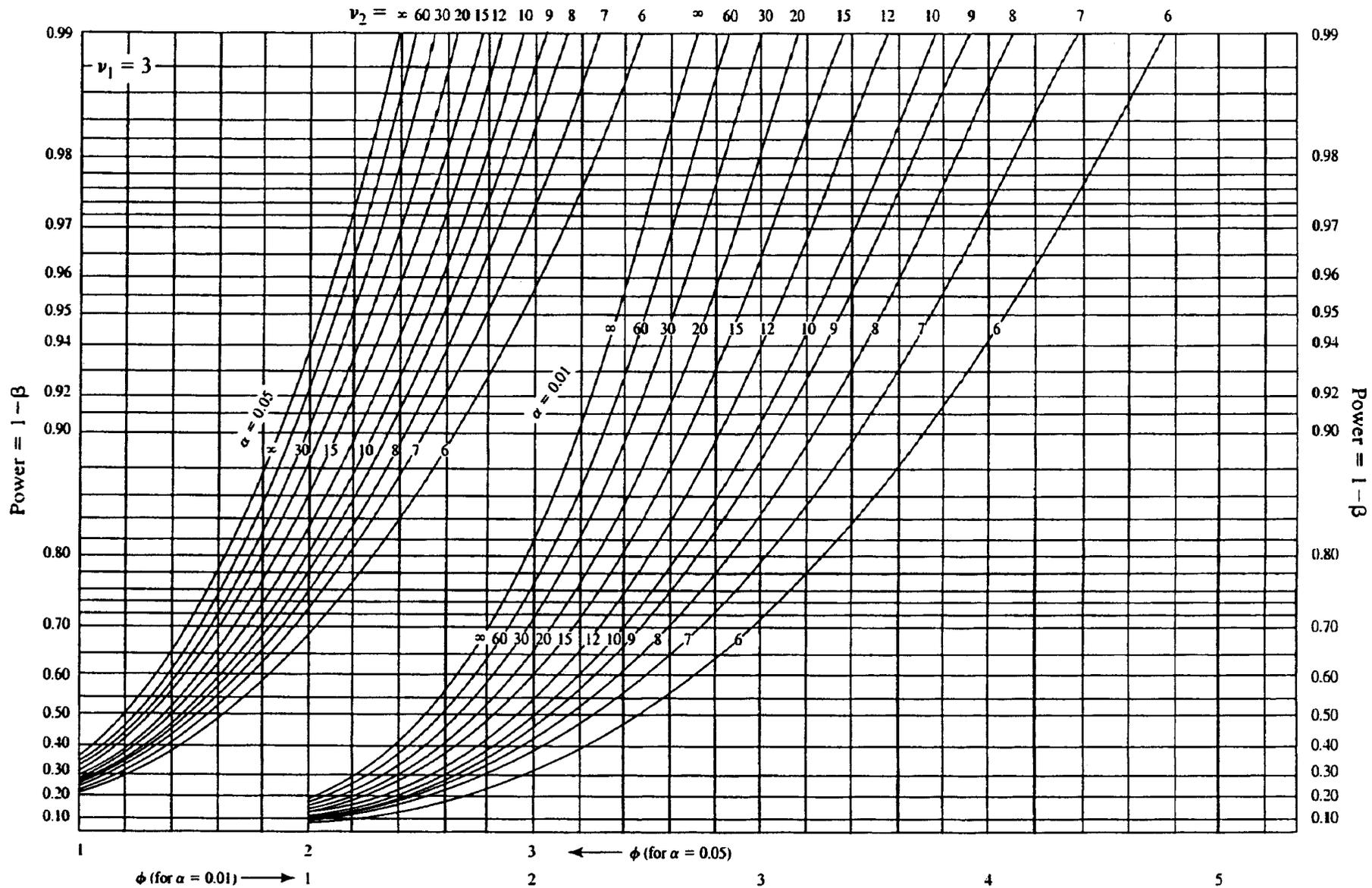
5) Nella resa dei cinque tipi di sementi a confronto, la differenza tra valore massimo e valore minimo (53,4 - 44,0) è maggiore di quella ottenuta nel confronto tra le medie dei cinque tipi di concime (51,6 - 46,2), riportate nel riquadro quattro. In questo disegno sperimentale, che utilizza la stessa varianza d'errore ed ha gli stessi gdl per ogni fattore, ciò spiega la maggiore significatività del test F tra le medie delle sementi rispetto a quello tra le medie dei concimi.

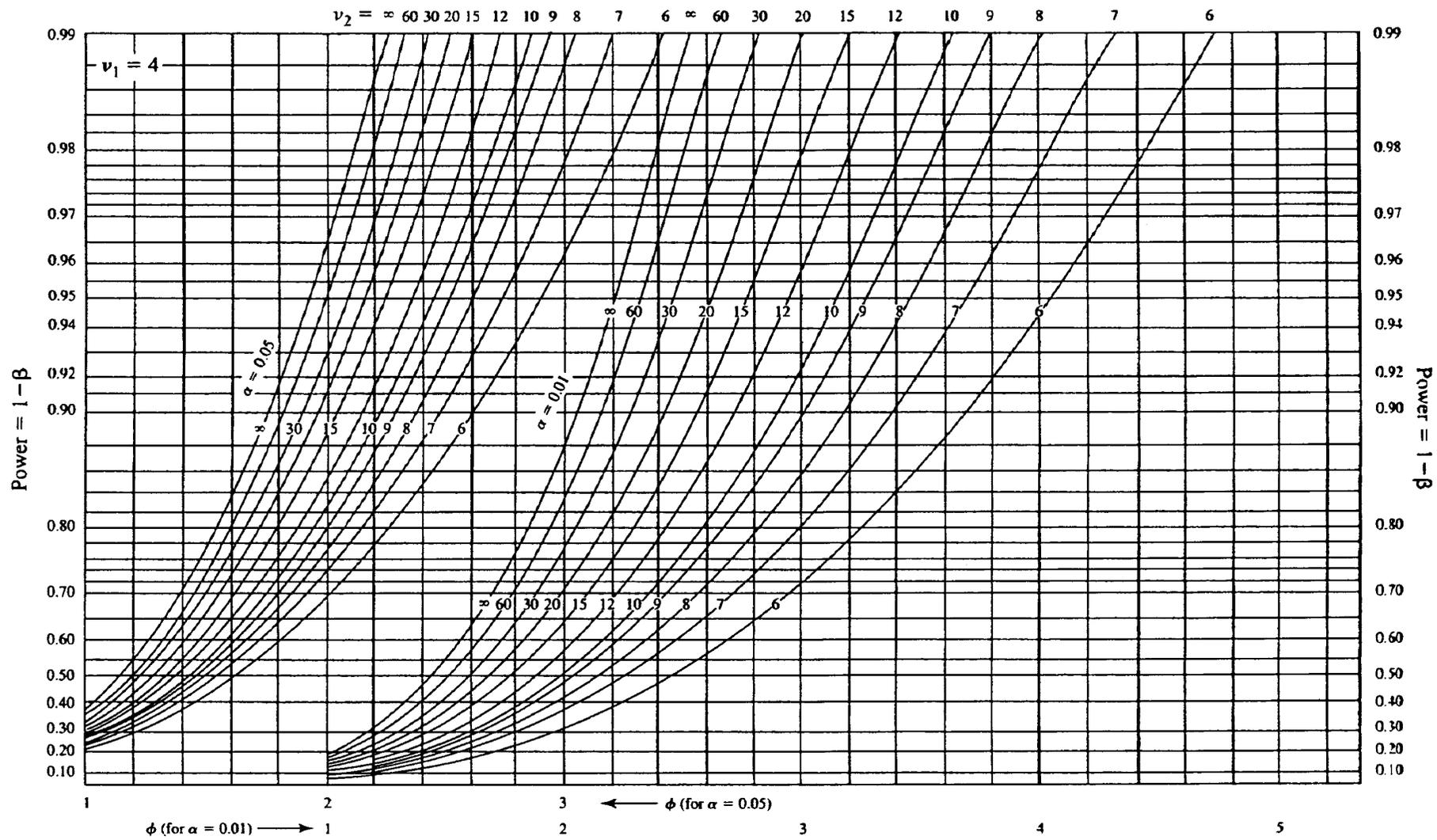
Secondo il test di Tukey, tra le cinque medie sono significative tutte le differenze superiori a 4,7592.

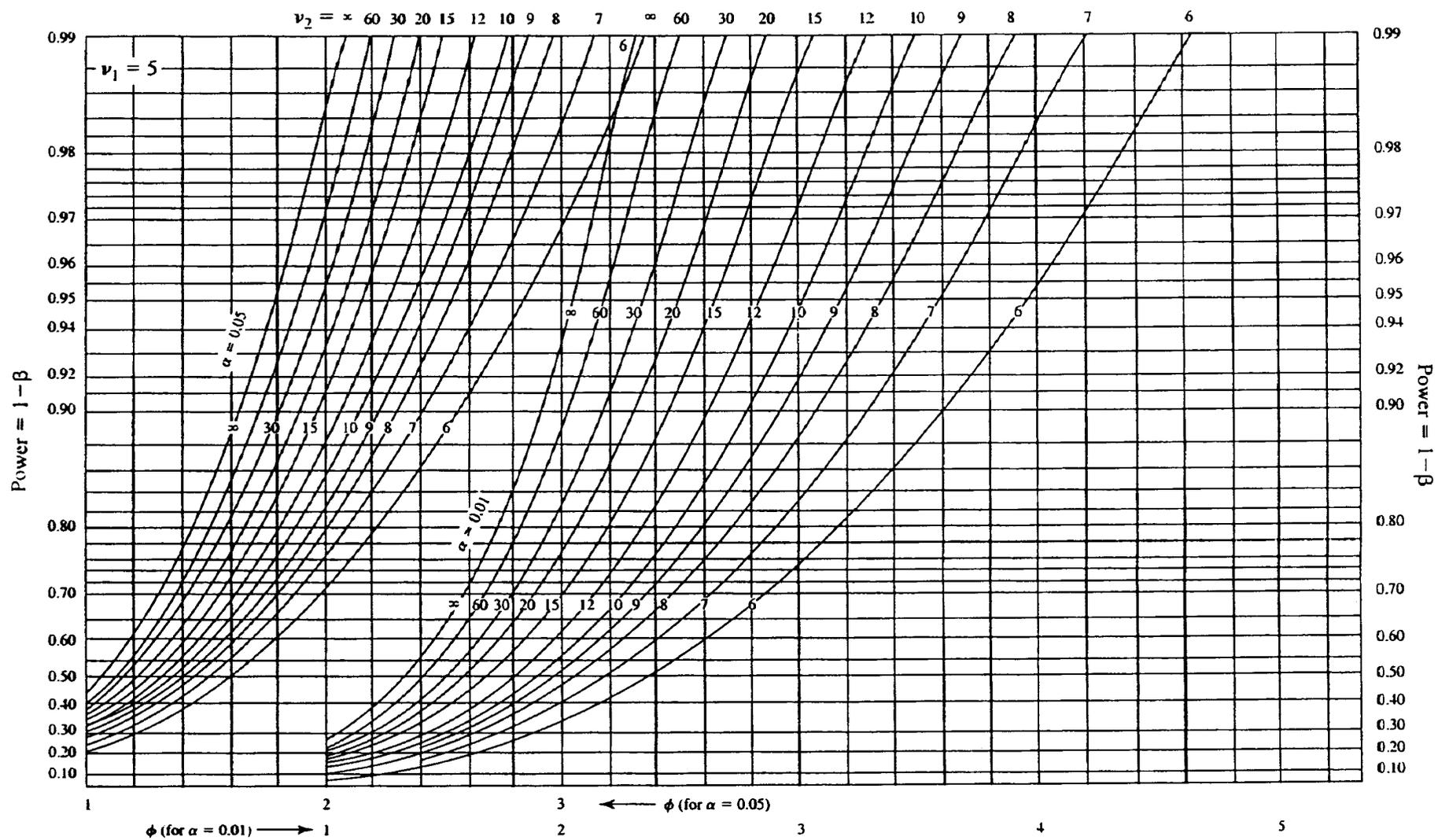
Secondo il test di Scheffè, per confronti sia semplici che complessi, sono significative le differenze superiori a 5,391.

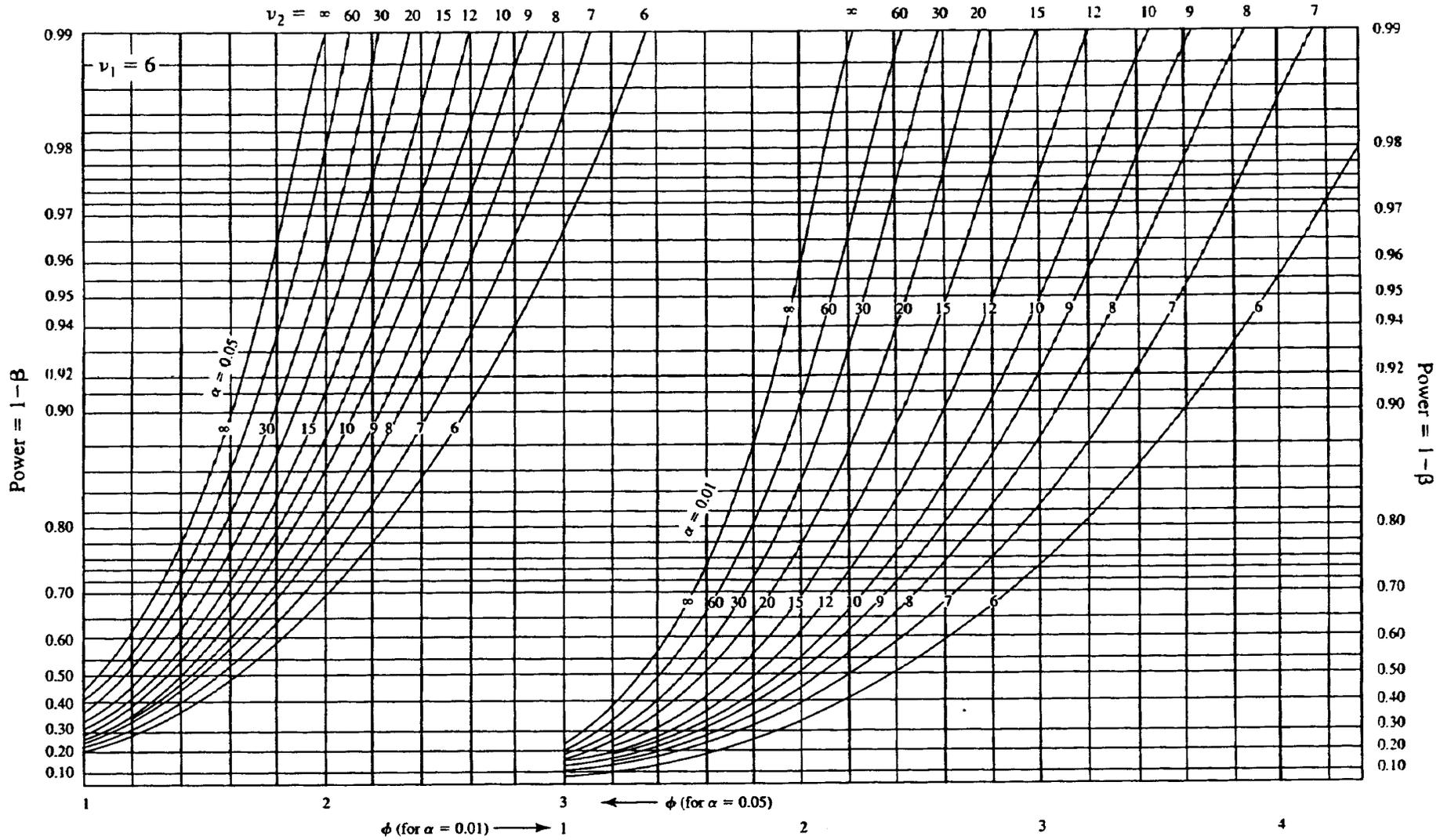


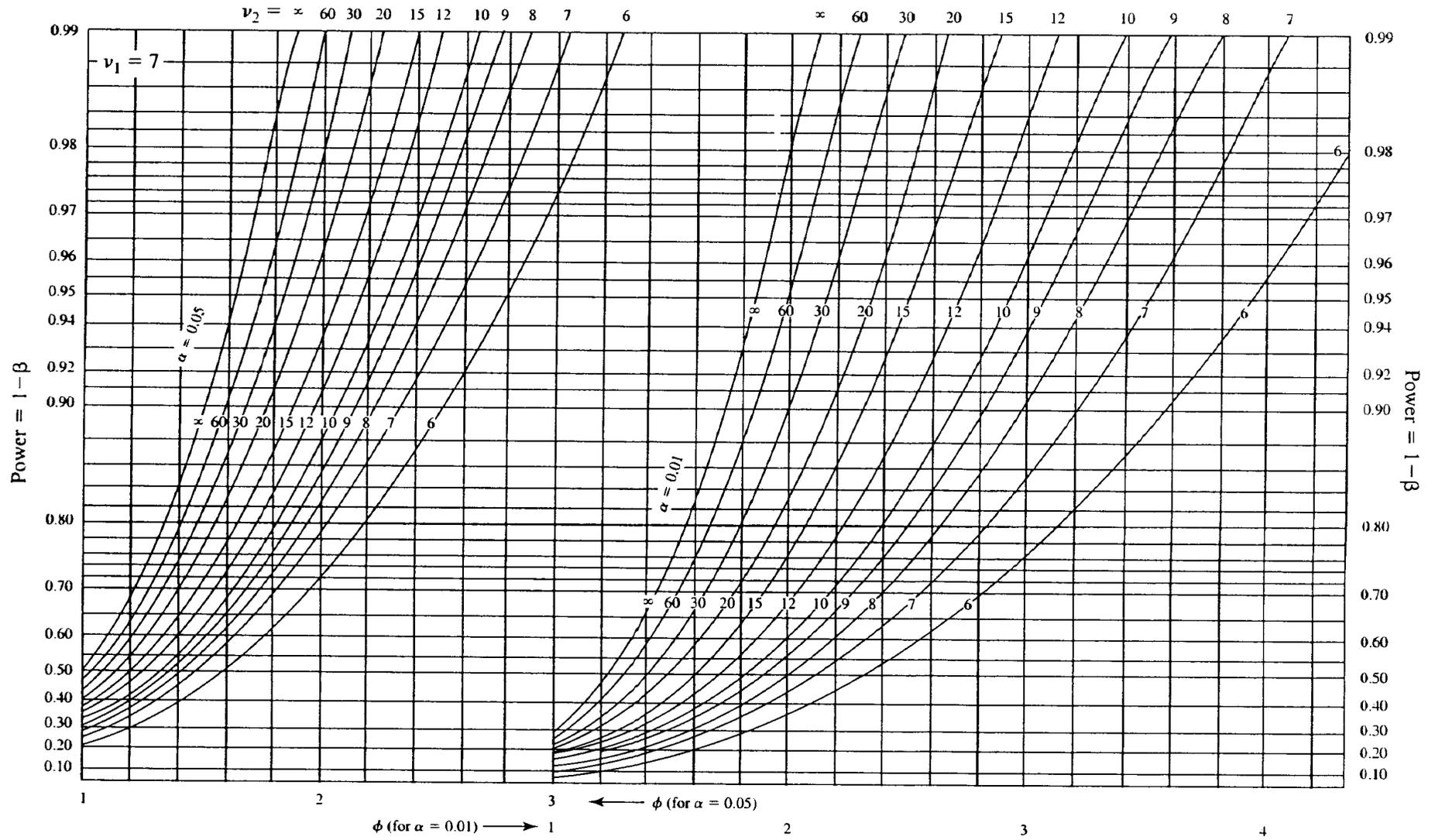


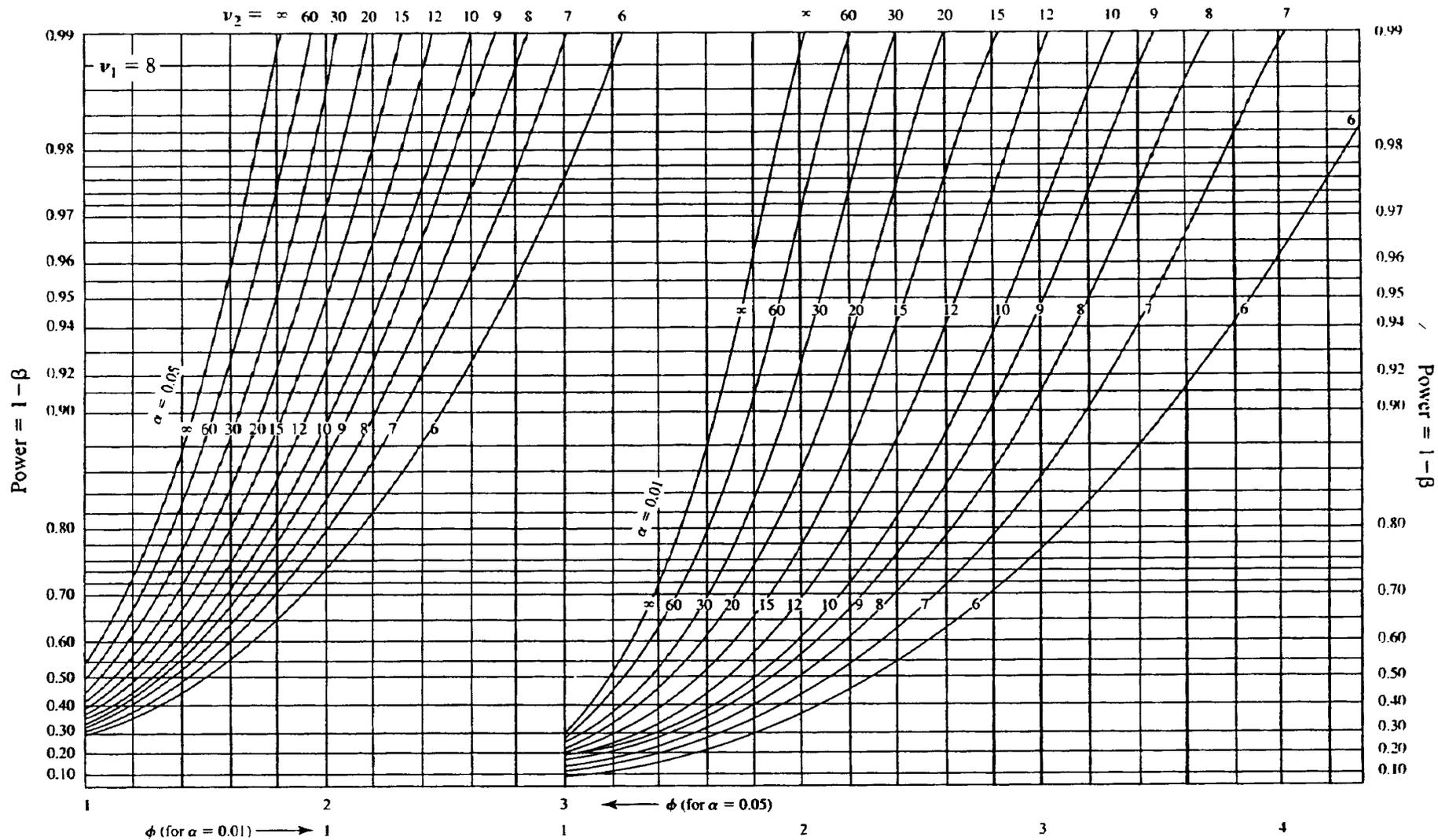












## CAPITOLO X

### ANALISI FATTORIALE, DISEGNI COMPLESSI TRASFORMAZIONI DEI DATI

#### 10.1. ANALISI FATTORIALE ED INTERAZIONE

Quando si studiano due o più fattori a vari livelli, non importa se qualitativi (come diversi tipi di farmaci) o quantitativi (quali la classe d'età), spesso l'interesse è rivolto ad analizzare non i singoli effetti, ma **le interazioni tra i fattori: se, come e quanto ogni livello o modalità di un fattore interagisce con quelli degli altri fattori, esaminati in tutte le combinazioni.** Nel caso in cui si debbano somministrare farmaci a persone di età o di sesso diversi, può essere di grande importanza sapere se lo stesso farmaco ha effetti differenti, potenziati o inibiti, nei giovani rispetto agli adulti o agli anziani, nei maschi rispetto alle femmine. Se non esiste interazione, il farmaco mediamente migliore sarà il più adatto per tutte le persone e potrà essere somministrato a tutti indifferentemente, in qualunque condizione; ma se esiste interazione con l'età o il sesso, occorre procedere ad una scelta appropriata del farmaco in rapporto ad essi.

Nella ricerca ambientale, quando si confrontano i livelli d'inquinamento in varie zone di una città tenendo in considerazione anche l'ora, è probabile che le aree ad inquinamento maggiore non siano sempre le stesse, ma che presentino variazioni in rapporto alla fase del giorno. Il traffico e l'attività delle fabbriche, che non hanno la stessa intensità a tutte le ore del giorno e sono distribuiti sul territorio in modo non omogeneo, possono determinare graduatorie d'inquinamento che si modificano nel corso della giornata. Di conseguenza, anche le politiche d'intervento potranno essere diverse, se l'interazione tra zone ed ore è significativa.

Gli esperimenti fino ad ora analizzati nei due precedenti capitoli sull'analisi della varianza permettono di evidenziare **gli effetti principali di ogni fattore**, ai suoi diversi livelli; ma esistono anche **effetti più complessi, determinati dalla loro azione congiunta, l'interazione.**

Quando si analizzano **due fattori** (a e b), come nel disegno a blocchi randomizzati **con repliche**, si parla di **interazione di primo ordine o di interazione a due fattori** ( $\alpha\beta$ ).

Con 3 fattori **con repliche**, si hanno

- gli **effetti principali** dei 3 fattori ( $\alpha$ ,  $\beta$ ,  $\gamma$ ),
- le 3 **interazioni di primo ordine** ( $\alpha\beta$ ,  $\alpha\gamma$ ,  $\beta\gamma$ ) causate dall'effetto dei fattori due a due ed infine
- una **interazione di secondo ordine** ( $\alpha\beta\gamma$ ), **determinata dall'effetto congiunto dei tre fattori.**

Quando si considerano contemporaneamente **molti fattori**, il numero delle loro combinazioni di primo ordine e di ordine superiore cresce in modo rapido; diventa difficile analizzare tutte le interazioni. Per

esempio, con **4 fattori** ( $\alpha, \beta, \gamma, \delta$ ) nel modello additivo dell'analisi della varianza con interazione, per ogni singolo valore  $X_{ijkp}$  è possibile stimare gli effetti di 17 quantità

$$X_{ijkp} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_p + \alpha\beta_{ij} + \alpha\gamma_{ik} + \alpha\delta_{ip} + \beta\gamma_{jk} + \beta\delta_{jp} + \gamma\delta_{kp} + \alpha\beta\gamma_{ijk} + \alpha\beta\delta_{ijp} + \\ + \alpha\gamma\delta_{ikp} + \beta\gamma\delta_{jkp} + \alpha\beta\gamma\delta_{ijkp} + \varepsilon_{ijkpr}$$

dove:

$m$  è la media generale;

$a_i, b_j, g_k, d_p$  sono i 4 **effetti principali** ( $C_4^1$ );

$ab_{ij}, ag_{ik}, ad_{ip}, bg_{jk}, bd_{jp}, gd_{kp}$  sono le 6 **interazioni di primo ordine o a due fattori** ( $C_4^2$ );

$abg_{ijk}, abd_{ijp}, agd_{ikp}, bgd_{jkp}$  sono le 4 **interazioni di secondo ordine o a tre fattori** ( $C_4^3$ );

$abgd_{ijkp}$  è l'**interazione di terzo ordine o a quattro fattori** ( $C_4^4$ );

$\varepsilon_{ijkpr}$  è la **varianza d'errore o residuo**, misurata attraverso le differenze delle repliche rispetto alla media di casella all'incrocio tra i 4 fattori.

Con l'analisi fattoriale della varianza, è possibile fornire una misura per ognuno di questi effetti e stimare la significatività sia dei fattori principali sia delle interazioni dei vari ordini.

In questo modello additivo, si hanno sia **interazioni positive** (quando la presenza di un fattore potenzia l'effetto dell'altro), sia **interazioni negative** (quando uno o più fattori si inibiscono in modo vicendevole). Non esiste interazione e si parla di **interazione nulla**, quando il risultato complessivo è determinato esclusivamente dalla somma dei singoli effetti principali.

## 10.2. INTERAZIONE TRA DUE FATTORI A PIU' LIVELLI

Per il rapido incremento della complessità dell'analisi all'aumentare del numero di fattori, lo studio dell'interazione è qui limitato ai casi più semplici. Come già avvenuto per vari altri argomenti e come sarà ripetuto anche in altri casi dei capitoli successivi, per analisi più complesse che superano le finalità di questo corso elementare si rinvia a trattazioni specifiche.

**Con due fattori** e con dati riportati in una tabella a due entrate, secondo le modalità già presentate nel caso di esperimenti a blocchi randomizzati, **è possibile analizzare l'interazione solo quando si dispone di più osservazioni in ognuna delle celle poste all'incrocio tra righe e colonne.**

Il modello additivo dell'ANOVA diventa

$$X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

dove:

$\mu$  è la **media generale** che, riferita ai dati dell'esperimento, è stimata nel modo migliore mediante la media campionaria di tutti i dati  $\bar{\bar{X}}$

$\alpha_i$  è l'**effetto del trattamento i** che, con i dati campionari, può essere calcolato come differenza della sua media dalla media generale:

$$\alpha_i = \bar{X}_i - \bar{\bar{X}}$$

$\beta_j$  è l'**effetto del blocco j** che, nell'esperimento, è determinato dalla differenza della sua media dalla media generale:

$$\beta_j = \bar{X}_j - \bar{\bar{X}}$$

$\alpha\beta_{ij}$  è l'**interazione tra l'effetto  $\alpha$  del trattamento i e l'effetto  $\beta$  del trattamento j** con dati sperimentali, può essere misurata come somma dei quadrati delle differenze tra la media osservata e quella attesa entro ogni casella

$$ab_{ij} = \sum_{i=1}^n \sum_{j=1}^p (\bar{X}_{ij} - \hat{X}_{ij})^2$$

dove

$\hat{X}_{ij}$ , la media attesa, può essere calcolata sommando alla media generale l'effetto-riga e l'effetto-colonna alla quale la casella appartiene:

$$\hat{X}_{ij} = \bar{\bar{X}} + (\bar{X}_i - \bar{\bar{X}}) + (\bar{X}_j - \bar{\bar{X}}) = \bar{X}_i + \bar{X}_j - \bar{\bar{X}}$$

$\varepsilon_{ijk}$  è la **differenza tra una singola osservazione e la sua media di casella**:

$$\varepsilon_{ijk} = X_{ijk} - \bar{X}_{ij}$$

La devianza totale, quella tra trattamenti e quella tra blocchi sono calcolate con le stesse modalità già spiegate nel capitolo precedente.

La **devianza d'interazione**  $SQ(\alpha\beta)$  secondo la formula euristica è la **somma dei quadrati degli scarti tra media osservata e media stimata in ogni casella**

$$SQ(ab) = \sum_{i=1}^n \sum_{j=1}^p (\bar{X}_{ij} - \hat{X}_{ij})^2 = \sum_{i=1}^n \sum_{j=1}^p (\bar{X}_{ij} - \bar{X}_i - \bar{X}_j + \bar{\bar{X}})^2$$

mentre la **devianza entro od errore**  $SQ(\varepsilon)$  è la **somma dei quadrati degli scarti di ogni replica dalla sua media di casella**

$$SQ(\varepsilon) = \sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^r (X_{ijk} - \bar{X}_{ij})^2$$

Di conseguenza, **nel modello ANOVA a due fattori con replicazioni, la devianza totale può essere suddivisa in 4 devianze:**

- devianza del fattore  $\alpha$ ;
- devianza del fattore  $\beta$ ;
- devianza dovuta all'effetto dell'interazione  $\alpha\beta$ ;
- devianza d'errore o residuo.

Esse ed i loro gdl godono della proprietà additiva

$$\text{Devianza}_{\text{totale}} = \text{Devianza}_{\alpha} + \text{Devianza}_{\beta} + \text{Devianza}_{\alpha\beta} + \text{Devianza}_{\text{errore}}$$

che permette di verificare i calcoli oppure di stimare una di esse a partire dalle altre quattro.

In questo modello, di norma i dati sono riportati in una tabella secondo lo schema generale sottostante.

In esso,

i trattamenti sono indicati in colonna,

i blocchi sono indicati nelle righe,

le repliche sono riportate nelle caselle, all'incrocio tra trattamenti e righe:

		TRATTAMENTI					
BLOCCHI	A	B	C	D	Totale di riga	Media di riga	
I	$X_{ij1} X_{ij2}$	Repliche	Repliche	Repliche	$S_{jr}$	$\bar{X}_j$	
II	Repliche	Repliche	Repliche	Repliche			
II	Repliche	Repliche	Repliche	Repliche			
Totale di colonna	$S_{ir}$				Totale $S_{ijr}$ generale		
Media di colonna	$\bar{X}_i$					Media $\bar{\bar{X}}$ generale	

e dove

$X_{ijk}$  è il valore della  $k$ -esima osservazione nel livello  $i$ -esimo del fattore A (trattamento) e nel livello  $j$ -esimo del fattore B (blocco),

$S_{ij}$  e  $\bar{X}_{ij}$  sono rispettivamente la somma e la media dei valori (repliche) della casella, collocata all'incrocio tra il trattamento  $i$ -esimo ed il blocco  $j$ -esimo,

ed inoltre, con la simbologia consueta,

$S_{ir}$  e  $\bar{X}_i$  sono la somma e la media dei valori per la colonna i del fattore A (trattamenti),

$S_{jr}$  e  $\bar{X}_j$  sono la somma e la media dei valori per la riga j del fattore B (blocchi),

$S_{ijr}$  e  $\bar{\bar{X}}$  sono la somma e la media generale.

Per valutare la significatività delle differenze tra i vari livelli del fattore A e del fattore B oltre a quella dell'interazione AB, il test F dell'analisi della varianza richiede il calcolo delle quantità riportate nella tabella sottostante, con i relativi gdl:

DEVIANZA	GDL	VARIANZA
<b>Totale</b>	<b>npr-1</b>	
Tra medie delle caselle o tra fattori	np-1	
<b>Tra trattamenti o del fattore A</b>	<b>n-1</b>	<b>Tra trattamenti</b>
<b>Tra blocchi o del fattore B</b>	<b>p-1</b>	<b>Tra blocchi</b>
<b>Interazione A × B</b>	<b>(n-1)×(p-1)</b>	<b>Interazione</b>
<b>Errore</b>	<b>(r-1)×(np)</b>	<b>Errore</b>

dove,

**n** è il numero di livelli del fattore A;

**p** è il numero di livelli del fattore B;

**r** è il numero di repliche entro ogni casella.

Come per la devianza totale,

**la devianza tra le medie di casella o devianza di tutti i fattori non è utile al calcolo dei test F;**

essa ha solo il duplice scopo di

- essere **didatticamente utile per comprendere più compiutamente il significato delle altre devianze** e

- soprattutto di permettere **calcoli più semplici per ottenere la devianza d'interazione**, mediante una semplice differenza con le devianze già note dei fattori A e B analizzati in modo separato.

**La devianza totale** SQ(T), definita come la somma dei quadrati degli scarti di ogni dato rispetto alla media generale, con la formula abbreviata può essere calcolata come differenza tra la somma dei quadrati di ogni dato ed il quadrato della somma generale diviso il numero totale di dati (chiamato anche termine di correzione):

$$SQ(T) = \sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^r (X_{ijk} - \bar{\bar{X}})^2 = \sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^r X_{ijk}^2 - \frac{(\sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^r X_{ijk})^2}{npr}$$

Il numero di gdl della devianza totale è uguale al numero totale di dati od osservazioni meno uno ( $npr - 1$ ).

**La devianza tra le medie delle caselle o devianza dei fattori**, indicata con  $SQ(\bar{X}_{ij})$ , per definizione è la somma dei quadrati degli scarti di ogni media di casella dalla media generale, moltiplicata per il numero di dati entro ogni casella; per comprendere il significato della formula, è utile ricordare che, se esistesse solamente la variabilità tra le caselle, tutte le repliche entro ogni casella avrebbero lo stesso valore. Con la formula abbreviata, questa devianza può essere calcolata per differenza tra la somma dei quadrati delle somme dei dati entro ogni casella diviso il numero di repliche e tra il quadrato della somma generale diviso il numero totale di dati (termine di correzione):

$$SQ(\bar{X}_{ij}) = \sum_{i=1}^n \sum_{j=1}^p r(\bar{X}_{ij} - \bar{\bar{X}})^2 = \sum_{i=1}^n \sum_{j=1}^p \frac{(\sum_{k=1}^r X_{ijk})^2}{r} - \frac{(\sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^r X_{ijk})^2}{npr}$$

**Il numero di gdl della devianza tra le medie di ogni casella è uguale al numero di caselle meno uno, corrispondente a  $(np-1)$ .**

**La devianza tra trattamenti o del fattore A**, indicata con  $SQ(A)$ , è la somma dei quadrati degli scarti tra la media di ogni trattamento e la media generale, moltiplicata per il numero di dati del trattamento; infatti se la variabilità fosse determinata solamente dagli effetti del trattamento, tutti i dati dello stesso trattamento dovrebbero essere tra loro uguali. Con la formula abbreviata, la devianza tra trattamenti può essere calcolata per differenza tra la somma dei quadrati della somma di ogni trattamento, diviso il numero di dati del trattamento, e tra il quadrato della somma generale diviso il numero totale di dati (termine di correzione):

$$SQ(A) = \sum_{i=1}^n pr(\bar{X}_i - \bar{\bar{X}})^2 = \sum_{i=1}^n \frac{(\sum_{j=1}^p \sum_{k=1}^r X_{ijk})^2}{pr} - \frac{(\sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^r X_{ijk})^2}{npr}$$

**La devianza tra blocchi o del fattore B**, indicata con  $SQ(B)$  è la somma dei quadrati degli scarti tra la media di ogni blocco e la media generale, moltiplicata per il numero di dati del blocco: se la variabilità tra le osservazioni fosse determinata solamente dall'effetto del blocco, tutti i dati dello stesso blocco dovrebbero avere lo stesso valore. Con la formula abbreviata, la devianza tra blocchi può essere calcolata per differenza tra la somma dei quadrati delle somme di ogni blocco, divisa per il numero di dati del blocco, e tra il quadrato della somma generale diviso il numero totale di dati (termine di correzione):

$$SQ(B) = \sum_{j=1}^p nr(\bar{X}_j - \bar{\bar{X}})^2 = \sum_{j=1}^p \frac{(\sum_{i=1}^n \sum_{k=1}^r X_{ijk})^2}{nr} - \frac{(\sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^r X_{ijk})^2}{npr}$$

**La devianza d'interazione tra i fattori A e B**, indicata con  $SQ(AB)$ , è la somma dei quadrati degli scarti di ogni media di casella rispetto al valore atteso dall'effetto trattamento e dall'effetto blocco ritenuti additivi

$$SQ(AB) = \sum_{i=1}^n \sum_{j=1}^p r(\bar{X}_{ij} - \bar{X}_i - \bar{X}_j + \bar{\bar{X}})^2$$

In modo molto più semplice e veloce, **la devianza d'interazione e i suoi gdl sono misurati per sottrazione dalla devianza tra le medie delle devianze sia del fattore A che del fattore B:**

$$SQ(AB) = SQ(\bar{X}_{ij}) - SQ(A) - SQ(B)$$

**La devianza d'errore o residuo**, indicata con  $SQ(e)$ , per definizione è la somma dei quadrati degli scarti di ogni valore rispetto alla media della sua casella:

$$SQ(e) = \sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^r (X_{ijk} - \bar{X}_{ij})^2$$

**Con r dati entro ogni casella, i suoi gdl sono**  $n \cdot p \cdot (r-1)$ .

Più rapidamente, **la devianza d'errore è ottenuta sottraendo dalla devianza totale la devianza tra le medie delle caselle:**

$$SQ(e) = SQ(T) - SQ(\bar{X}_{ij})$$

Dividendo la devianza tra trattamenti, quella tra blocchi, quella d'interazione e quella d'errore per i rispettivi gdl si ottengono le varianze corrispondenti:

$$S_A^2 = \frac{SQ(A)}{n-1}; \quad S_B^2 = \frac{SQ(B)}{p-1}; \quad S_{AB}^2 = \frac{SQ(AB)}{(n-1)(p-1)}; \quad S_e^2 = \frac{SQ(e)}{np(r-1)}$$

**Nel modello ANOVA a due fattori con repliche, si possono verificare 3 ipotesi distinte ed eseguire i tre test corrispondenti.**

1 - Si verifica l'ipotesi nulla di **nessuna differenza tra le medie del fattore A:**

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_n$$

contro l'ipotesi alternativa:

$$H_1: \text{non tutte le medie } \mu \text{ sono tra loro uguali}$$

mediante il test F, con gdl n-1 al numeratore e np(r-1) al denominatore,

$$F_{(n-1), np(r-1)} = \frac{S_A^2}{S_e^2}$$

dove l'ipotesi nulla viene respinta se il rapporto supera il valore critico alla significatività prescelta.

2 - Si verifica l'ipotesi nulla di **nessuna differenza tra le medie del fattore B**

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_p$$

contro l'ipotesi alternativa

$$H_1: \text{non tutte le } \mu \text{ sono tra loro uguali}$$

mediante il test F con gdl **p-1** al numeratore e **np(r-1)** al denominatore

$$F_{(p-1), np(r-1)} = \frac{S_B^2}{S_e^2}$$

dove l'ipotesi nulla viene respinta se il rapporto supera il valore critico alla significatività prescelta.

3 - Si verifica l'ipotesi nulla di **nessuna interazione tra i fattori A e B** ai vari livelli

$$H_0: AB_{ij} = 0 \text{ per ogni } i \text{ e } j$$

contro l'ipotesi alternativa

$$H_1: AB_{ij} \neq 0 \text{ per almeno un } ij$$

mediante il ricorso ad un test F con gdl

- **(n-1)(p-1)** al numeratore
- **np(r-1)** al denominatore

$$F_{(n-1)(p-1), np(r-1)} = \frac{S_{AB}^2}{S_e^2}$$

e si respinge l'ipotesi nulla se il rapporto supera il valore critico alla probabilità prefissata.

ESEMPIO. Si vuole verificare se insetti adulti della stessa specie che vivono in 4 località differenti (A, B, C, D) hanno differenze significative nelle loro dimensioni, considerando pure che dalla primavera all'autunno continuano ad aumentare.

	LOCALITA'			
	A	B	C	D
<b>Primavera</b>	<b>45</b>	<b>63</b>	<b>70</b>	<b>48</b>
	<b>50</b>	<b>57</b>	<b>65</b>	<b>52</b>
<b>Estate</b>	<b>57</b>	<b>77</b>	<b>74</b>	<b>60</b>
	<b>65</b>	<b>69</b>	<b>80</b>	<b>56</b>
<b>Autunno</b>	<b>70</b>	<b>82</b>	<b>88</b>	<b>70</b>
	<b>79</b>	<b>75</b>	<b>82</b>	<b>77</b>

Si vuole verificare se la crescita è diversa nelle 4 località, secondo le stagioni (per semplificare al massimo i calcoli, sono state riportate solamente 2 misure per località e stagione).

Risposta.

E' un'analisi a due fattori con repliche; si vuole verificare se nella crescita esiste interazione tra località e stagione. Dapprima si devono stimare i totali e le medie necessari per i calcoli successivi

	LOCALITA'				Totali	Medie
	A	B	C	D		
<b>Primavera</b>	<b>45</b>	<b>63</b>	<b>70</b>	<b>48</b>		
	<b>50</b>	<b>57</b>	<b>65</b>	<b>52</b>	450	56,250
<b>Totali</b>	(95)	(120)	(135)	(100)		
<b>Medie</b>	(47,5)	(60,0)	(67,5)	(50,0)		
<b>Estate</b>	<b>57</b>	<b>77</b>	<b>74</b>	<b>60</b>		
	<b>65</b>	<b>69</b>	<b>80</b>	<b>56</b>	538	67,250
<b>Totali</b>	(122)	(146)	(154)	(116)		
<b>Medie</b>	(61,0)	(73,0)	(77,0)	(58,0)		
<b>Autunno</b>	<b>70</b>	<b>82</b>	<b>88</b>	<b>70</b>		
	<b>79</b>	<b>75</b>	<b>82</b>	<b>77</b>	623	77,875
<b>Totali</b>	(149)	(157)	(170)	(147)		
<b>Medie</b>	(74,5)	(78,5)	(85,0)	(73,5)		
<b>Totali gen.</b>	366	423	459	363	1611	
<b>Medie gen.</b>	61,00	70,50	76,50	60,50		67,125

- La devianza totale , con 23 gdl (per tutti e 24 i dati) è ottenuta con

$$(45 - 67,125)^2 + (50 - 67,125)^2 + \dots + (70 - 67,125)^2 + (77 - 67,125)^2$$

e corrisponde a :

$$45^2 + 50^2 + \dots + 70^2 + 77^2 - \frac{1611^2}{24} = \mathbf{3.300}$$

- **La devianza tra le medie delle caselle o dei vari livelli dei due fattori, con 11 gdl** (le medie delle caselle all'incrocio località per stagione sono 12) è data da

$$2(47,5 - 67,125)^2 + 2(60,0 - 67,125)^2 + \dots + 2(73,5 - 67,125)^2$$

e corrisponde a (con i totali delle 12 caselle)

$$\frac{95^2}{2} + \frac{120^2}{2} \dots + \frac{147^2}{2} - \frac{1611^2}{24} = \mathbf{3.052}$$

- **La devianza tra trattamenti o del fattore A, con 3 gdl** è calcolata (per le 4 medie di colonna) da

$$\text{Devianza}_A = 6 \cdot (61,00 - 67,125)^2 + 6 \cdot (70,50 - 67,125)^2 + \dots + 6 \cdot (60,50 - 67,125)^2$$

ed è uguale a (con i totali dei 4 trattamenti):

$$\text{Devianza}_A = \frac{366^2}{6} + \frac{423^2}{6} + \frac{459^2}{6} + \frac{363^2}{6} - \frac{1611^2}{24} = \mathbf{1.084}$$

- **La devianza tra blocchi o del fattore B, con 2 gdl** è ottenuta mediante

$$\text{Devianza}_B = 8 \cdot (56,250 - 67,125)^2 + 8 \cdot (67,250 - 67,125)^2 + 8 \cdot (77,875 - 67,125)^2$$

e corrisponde a (con i totali dei 3 blocchi)

$$\text{Devianza}_B = \frac{450^2}{8} + \frac{538^2}{8} + \frac{623^2}{8} - \frac{1611^2}{24} = \mathbf{1.871}$$

- **La devianza d'interazione AB** viene stimata per differenza:

$$\text{Devianza}_{AB} = \text{Devianza}_{\text{tra medie}} - \text{Devianza}_A - \text{Devianza}_B = 3.052 - 1.084 - 1.871 = \mathbf{97}$$

e nello stesso modo vengono calcolati i rispettivi gdl

$$\text{gdl}_{AB} = \text{gdl}_{\text{tra medie}} - \text{gdl}_A - \text{gdl}_B = 11 - 3 - 2 = \mathbf{6}$$

- **La devianza d'errore o residuo** nei calcoli manuali viene quasi sempre stimata per differenza

$$\text{Devianza}_{\text{errore}} = \text{Devianza}_{\text{totale}} - \text{Devianza}_{\text{tra medie}} = 3.300 - 3.052 = \mathbf{248}$$

e nello stesso modo sono calcolati i suoi gdl

$$\text{gdl}_{\text{errore}} = \text{gdl}_{\text{totale}} - \text{gdl}_{\text{tra medie}} = 23 - 11 = \mathbf{12}$$

I gdl della devianza d'errore possono essere calcolati anche in modo diretto: **ognuna delle 12 caselle con 2 dati contribuisce alla devianza d'errore complessiva con 1 gdl, fornendo un totale di 12 gdl.**

E' utile disporre le devianze e i gdl in tabella, con il calcolo delle 4 varianze utili ai 3 test F:

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>
<b>Totale</b>	3.300	23	
<b>Tra medie</b>	3.052	11	
<b>Tra tratt. (A)</b>	1.084	3	361,33
<b>Tra blocchi (B)</b>	1.870	2	935,50
<b>Interazione (AB)</b>	97	6	16,16
<b>Errore</b>	248	12	20,66

Per le differenze tra trattamenti o effetto del fattore A si calcola un test F con gdl 3 e 12

$$F_{3,12} = \frac{361,33}{20,66} = 17,49$$

il cui valore critico alla probabilità  $\alpha = 0.05$  è uguale a 3,49.

Per le differenze tra blocchi o effetto del fattore B si calcola un test F con gdl 2 e 12

$$F_{2,12} = \frac{935,5}{20,66} = 45,28$$

il valore critico del quale alla probabilità  $\alpha = 0.05$  è 3,89.

Per l'effetto dell'interazione AB si calcola un test F con gdl 6 e 12

$$F_{6,12} = \frac{16,16}{20,66} = 0,78$$

che fornisce un rapporto inferiore ad 1 e

quindi non è significativo.

Con i risultati dell'esempio, si possono trarre le conclusioni relative ai tre test F:

- 1 - nelle 4 località, le dimensioni medie degli insetti sono significativamente differenti;
- 2 - tra primavera, estate ed autunno le dimensioni medie degli insetti variano significativamente;
- 3 - non esiste interazione: nelle 4 località le dimensioni medie degli insetti variano con intensità simile durante le stagioni.

Il disegno sperimentale dell'analisi della varianza a più criteri è la vera novità introdotta da Fisher; su di essa si fonda la statistica moderna. Con l'assunzione che la natura risponde solamente a domande

semplici, si era diffuso tra i ricercatori un metodo d'indagine che imponeva la variazione di un solo fattore sperimentale alla volta. Fisher dimostrò il vantaggio del disegno sperimentale di tipo fattoriale: **se si seguono contemporaneamente più fattori, si riesce ad evidenziare le loro interazioni e si ha una visione più corretta della complessità delle risposte. Si ottengono anche i vantaggi di poter utilizzare un numero minore di osservazioni e di ridurre sensibilmente la varianza d'errore.**

Gli esperimenti fattoriali per l'analisi dell'interazione tra 2 fattori a vari livelli sono condotti in laboratorio altrettanto spesso che nelle ricerche di campagna. Per chi non ha esperienza, l'approccio metodologico utilizzato in laboratorio appare diverso da quello impiegato in campagna, poiché i risultati spesso sono presentati in tante colonne, quante sono le combinazioni dei livelli dei due fattori a confronto, come se si trattasse di un'analisi ad un criterio di classificazione.

**Con il fattore A a tre livelli ( $a_1, a_2, a_3$ ) ed il fattore B a due modalità ( $b_1$  e  $b_2$ ) si devono formare 6 gruppi, date dalle loro combinazioni, ognuno con lo stesso numero d'osservazioni. Di norma i risultati sono riportati in una tabella che solo apparentemente è ad una sola entrata.**

$a_1b_1$	$a_1b_2$	$a_2b_1$	$a_2b_2$	$a_3b_1$	$a_3b_2$
$X_{111}$	$X_{121}$	$X_{211}$	---	---	---
$X_{112}$	$X_{122}$	$X_{212}$	---	---	$X_{322}$
$X_{113}$	$X_{123}$	$X_{213}$	---	---	$X_{323}$
---	---	---	---	---	---

E' sufficiente una corretta impostazione tabellare come quella sottostante

	$a_1$	$a_2$	$a_3$
$b_1$	Repliche	Repliche	Repliche
$b_2$	Repliche	Repliche	Repliche

per rendere evidente che si tratta di un esperimento a due fattori con repliche.

E' quindi un esperimento fattoriale, con il quale è possibile analizzare contemporaneamente se esiste differenza tra le varie modalità entro ognuno dei due fattori e se l'interazione tra loro è significativa.

ESEMPIO. Si vogliono verificare gli effetti di 3 mangimi industriali, contenenti ormoni sintetici, sulla crescita di animali: un aspetto fondamentale della ricerca è dimostrare se nei due sessi hanno un effetto di segno opposto.

A questo scopo, i 3 tipi di mangime ( $a_1, a_2, a_3$ ) sono stati somministrati

- a tre gruppi di femmine ( $b_1$ ), formati ognuno da 5 individui scelti per estrazione casuale da un grande gruppo di femmine, e

- a tre gruppi di maschi ( $b_2$ ), formati ognuno da 5 individui scelti per estrazione casuale da un grande gruppo di soli maschi.

Dopo un mese di dieta, è stato misurato l'accrescimento di ogni cavia.

Per ognuno dei gruppi, i risultati sono riportati nella tabella sottostante:

	$a_1b_1$	$a_1b_2$	$a_2b_1$	$a_2b_2$	$a_3b_1$	$a_3b_2$
	12	28	28	27	14	20
	18	32	21	29	17	17
	21	31	22	31	15	18
	25	34	25	30	18	21
	24	32	27	26	16	16
Medie	20,0	31,4	24,6	28,6	16,0	18,4

Esiste interazione tra farmaci e sesso?

Risposta.

Se non già fatto nella presentazione dei risultati, soprattutto per l'apprendimento dei metodi risulta utile impostare i dati secondo una tabella che evidenzi che si tratta di un'analisi a due fattori con repliche, non di un'ANOVA ad un solo criterio di classificazione:

	<b>a<sub>1</sub></b>	<b>a<sub>2</sub></b>	<b>a<sub>3</sub></b>
<b>B<sub>1</sub></b>	12	28	14
	18	21	17
	21	22	15
	25	25	18
	24	27	16
<b>B<sub>2</sub></b>	28	27	20
	32	29	17
	31	31	18
	34	30	21
	32	26	16

Nel passaggio successivo, da essa è utile ricavare una nuova tabella con tutte le medie da analizzare:

	<b>a<sub>1</sub></b>	<b>a<sub>2</sub></b>	<b>a<sub>3</sub></b>	<b>Medie</b>
<b>b<sub>1</sub></b>	<b>20,0</b>	<b>24,6</b>	<b>16,0</b>	20,200
<b>b<sub>2</sub></b>	<b>31,4</b>	<b>28,6</b>	<b>18,4</b>	26,133
<b>Medie</b>	25,7	26,6	17,2	23,166

Con questi dati riassuntivi, è più semplice formulare le tre ipotesi alle quali il test permette di rispondere.

1) La prima è relativa alle medie dei tre mangimi ( $a_1$ ,  $a_2$ ,  $a_3$ ) con ipotesi nulla

$$H_0: \mu_{a1} = \mu_{a2} = \mu_{a3}$$

ed ipotesi alternativa  $H_1$  che le  $\mu$  dei tre farmaci non sono tutte uguali.

2) La seconda ipotesi nulla è relativa alla media dei sessi ( $b_1$  e  $b_2$ )

$$H_0: \mu_{b1} = \mu_{b2}$$

con ipotesi alternativa bilaterale

$$H_1: \mu_{a1} \neq \mu_{a2}$$

3) La terza ipotesi riguarda la significatività dell'interazione tra il fattore A e il fattore B, con ipotesi nulla

$$H_0: \alpha\beta = 0$$

che afferma che non esiste interazione  
ed ipotesi alternativa

$$H_1: \alpha\beta \neq 0$$

che afferma che l'interazione è significativa, pure senza specificarne la direzione.

Per rispondere a questi quesiti, si deve utilizzare lo stesso schema dell'analisi a due criteri di classificazione con misure ripetute; pertanto si devono calcolare:

1 - **la devianza totale** che è fondata sugli scarti di ogni valore dalla media generale (23,1); con 30 dati, i gdl sono 29;

2 - **la devianza tra le medie di casella o delle combinazioni tra i due fattori**, che è fondata sugli scarti di ognuna delle 6 medie (20,0; 24,6; 16,0; 31,4; 28,6; 18,4) dalla media generale; i gdl sono 5;

3 - **la devianza dovuta al fattore A**, tra le medie dei 3 tipi di mangime (25,7; 26,6; 17,2) con 2 gdl;

4 - **la devianza dovuta al fattore B**, tra le medie dei 2 sessi (20,200; 26,133), con 1 gdl;

5 - **la devianza d'interazione AB**, mangime per sesso, ottenuta facilmente per sottrazione della devianza del fattore A più quella del fattore B dalla devianza tra le medie di casella; nello stesso modo si stimano i gdl; questa devianza stima **se le sei medie di casella sono determinate dalla somma degli effetti di riga e di colonna o se risentono in modo significativo dell'interazione;**

6 - **la devianza d'errore o residuo**, data dagli scarti di ognuna delle 30 osservazioni dalla loro media di gruppo; con 5 dati, ogni gruppo contribuisce con 4 gdl per un totale di 24 gdl (4 x 6); più rapidamente, la devianza d'errore ed i suoi gdl sono ottenuti per sottrazione della devianza tra le medie di casella dalla devianza totale.

Con le formule già presentate, si calcolano le devianze, i gdl e le varianze

	<b>DEVIANZA</b>	<b>GDL</b>	<b>VARIANZA</b>	<b>F</b>
Totale	1.128	29	---	---
Tra medie	917	5	---	---
<b>Fattore A</b>	<b>538</b>	<b>2</b>	<b>269</b>	30,74
<b>Fattore B</b>	<b>264</b>	<b>1</b>	<b>264</b>	30,17
<b>Interazione AB</b>	<b>115</b>	<b>2</b>	<b>57,5</b>	6,57
<b>Errore</b>	<b>210</b>	<b>24</b>	<b>8,75</b>	---

utili ai 3 test F.

Per il fattore A

$$F_{2,24} = \frac{269}{8,75} = 30,74$$

si ottiene un F con gdl 2 e 24 uguale a 30,74

mentre il valore critico alla probabilità  $\alpha = 0.05$  è uguale a 3,40.

Per il fattore B,

$$F_{1,24} = \frac{264}{8,75} = 30,17$$

si ottiene un F con gdl 1 e 24 uguale a 30,17

che deve essere confrontato con un valore critico alla probabilità  $\alpha = 0.05$  uguale a 4,26.

Per l'interazione A x B,

$$F_{2,24} = \frac{57,5}{8,75} = 6,57$$

il valore di F con gdl 2 e 24 è uguale a 6,57

mentre il suo valore critico alla probabilità  $\alpha = 0.05$  è uguale a 3,40.

Dal confronto dei tre valori di F con i rispettivi valori critici, si possono trarre

le **conclusioni relative alle tre ipotesi nulle:**

- 1 - i tre mangimi danno medie di accrescimento significativamente differenti;
- 2 - i due sessi hanno accrescimenti significativamente differenti;
- 3 - **l'interazione tra i due fattori (mangime e sesso) è significativo**, presumibilmente per la presenza di ormoni sintetici.

### 10.3. RAPPRESENTAZIONE GRAFICA PER L'INTERPRETAZIONE DELL'INTERAZIONE A DUE FATTORI

La semplice lettura della tabella che riporta le medie di casella, quelle di riga e quelle di colonna è utile per comprendere il significato dell'interazione. Quando il test F per l'interazione non risulta significativo, la media del livello del fattore A minore (o maggiore) è tale in tutti i livelli del fattore B; viceversa, quando il test F per l'interazione AxB risulta significativo, la media del livello del fattore A che risulta minore (o maggiore) è tale solamente per alcuni livelli del fattore B.

Nell'esempio precedente, dove l'interazione  $A \times B$  è risultata significativa, è possibile osservare che per il sesso femminile ( $b_1$ ) la crescita maggiore è determinata dal trattamento con l'ormone  $a_2$  (con media 24,6), mentre per il sesso maschile ( $b_2$ ) il valore medio maggiore è determinato dal trattamento con l'ormone  $a_1$  (con media 31,4).

		TRATTAMENTO			
SESSO		$a_1$	$a_2$	$a_3$	Medie
$b_1$		<b>20,0</b>	<b>24,6</b>	<b>16,0</b>	20,200
$b_2$		<b>31,4</b>	<b>28,6</b>	<b>18,4</b>	26,133
Medie		25,7	26,6	17,2	23,166

Se non esistesse interazione sesso x trattamento, ognuna delle medie dei 6 gruppi sperimentali risentirebbe solamente dell'effetto del fattore A (colonna) e del fattore B (riga) in modo additivo, secondo la relazione:

$$X_{a_i b_j} = \bar{X}_{a_i} + \bar{X}_{b_j} - \bar{\bar{X}}$$

Pertanto, se non esistesse interazione A x B, le medie dei 6 gruppi sperimentali avrebbero dovuto essere:

		TRATTAMENTO			
SESSO		$a_1$	$a_2$	$a_3$	Medie
$B_1$		<b>22,734</b>	<b>23,634</b>	<b>14,234</b>	20,200
$B_2$		<b>28,667</b>	<b>29,567</b>	<b>20,167</b>	26,133
Medie		25,7	26,6	17,2	23,166

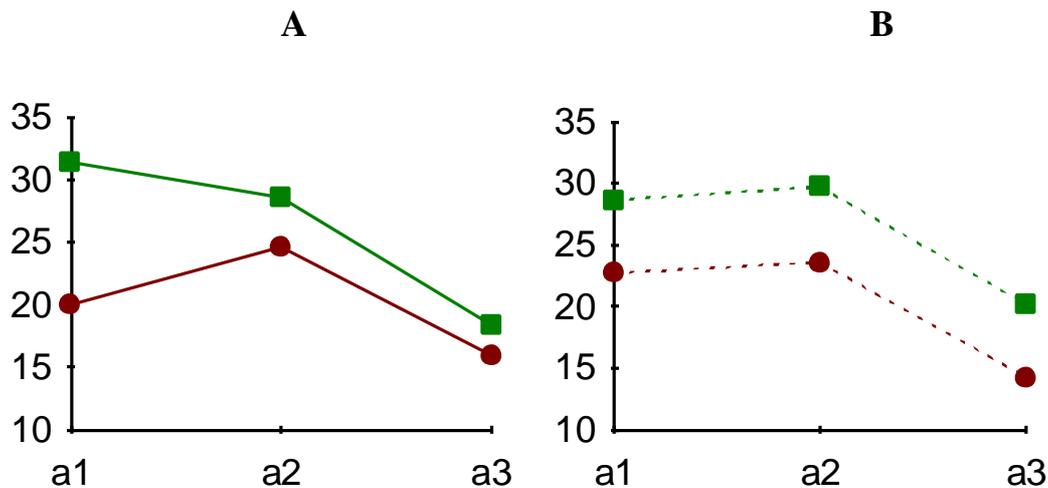
come quelli riportati nella tabella precedente.

Il trattamento che fornisce la media minore ( $a_3$  con 17,2) determina la media minore sia nel sesso  $b_1$  (14,234) che in quello  $b_2$  (20,167); simmetricamente, il trattamento che causa la media maggiore ( $a_2$  con 26,6) determina i valori medi più alti sia in  $b_1$  (23,634) che in  $b_2$  (29,567).

**La devianza d'interazione è la somma dei quadrati degli scarti tra questi valori medi stimati ed i valori medi osservati. Poiché i valori medi stimati nelle caselle sono calcolati dai totali marginali e da quello totale, i suoi gdl sono  $(n - 1) \times (p - 1)$ ; nell'esempio riportato,  $(3 - 1) \times (2 - 1) = 2$ , come già calcolato utilizzando la proprietà additiva dei gdl.**

Nell'analisi di un disegno fattoriale, in genere la varianza d'interazione ha numerosi gdl. Di conseguenza, è determinata da più differenze da quanto atteso e l'interpretazione non sempre è banale, ma richiede attenzione.

Un modo semplice per evidenziare l'esistenza dell'interazione ed i suoi effetti è la rappresentazione grafica dei valori medi osservati e del suo confronto con quelli stimati. Nelle due figure sottostanti sono riportati



**A** Grafico dei valori medi osservati

**B** Grafico dei valori medi stimati (in assenza di interazione)

- nella figura A i **valori medi osservati** per il sesso  $b_2$  (i quadrati) e per il sesso  $b_1$  (le palline),

- nella figura B sono riportati i **valori medi stimati in assenza d'interazione**.

Quando non esiste interazione, i segmenti che uniscono i punti risultano paralleli, come nella figura B, poiché le medie sono determinate solo dagli effetti additivi del fattore A e del fattore B. Nell'esempio con i valori medi stimati (esempio B), le linee risultano esattamente parallele, poiché mancano totalmente sia la variabilità casuale sia l'interazione.

Quando esiste interazione, come nel grafico con i valori medi osservati (esempio A), le linee si allontanano sensibilmente dal parallelismo: il trattamento che risulta migliore o peggiore per un sesso non ottiene lo stesso risultato con l'altro sesso, poiché si ha un ulteriore fattore di inibizione o di potenziamento per almeno una media.

#### 10.4. ANALISI DELLA VARIANZA A DUE FATTORI CON REPLICHE INEGUALI

Le procedure illustrate nei paragrafi precedenti per l'analisi della varianza nel caso di due fattori con repliche richiedono che **il loro numero (r) entro ogni cella sia sempre uguale**. La **potenza del test** è massima e **l'interpretazione dei risultati** è più semplice; ma nella pratica della ricerca non sempre è possibile rispettare uno schema così rigido.

Quando il numero di dati entro ogni casella **non è costante**, in un'analisi a due fattori è ancora possibile ricorrere a formule semplici, utilizzabili per calcoli manuali, se **il numero di repliche entro ogni casella è proporzionale**.

Come nella tabella sottostante (con 4 livelli per il fattore A e 3 livelli per il fattore B), nella quale i dati riportati ( $n_{ij}$ ) in ogni casella rappresentano il numero di osservazioni,

	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>	Totale
B <sub>1</sub>	3	6	9	6	<b>24</b>
B <sub>2</sub>	4	8	12	6	<b>32</b>
B <sub>3</sub>	2	4	6	2	<b>16</b>
Totale	<b>9</b>	<b>18</b>	<b>27</b>	<b>18</b>	<b>72</b>

si può parlare di **numero proporzionale di repliche** quando

$$n_{ij} = \frac{n_i \cdot n_j}{N}$$

dove

- $n_{ij}$  è il numero di dati nella cella all'incrocio della riga **i** e della colonna **j**,
- $n_i$  è il numero totale di dati nella riga **i**,
- $n_j$  è il numero totale di dati nella colonna **j**,
- $N$  è il numero complessivo di osservazioni riportate nella tabella.

Nell'esempio riportato, si può parlare di numero proporzionale di repliche poiché per tutte le celle è vera la relazione precedente; ad esempio,

- in quella posta all'incrocio tra  $A_3$  e  $B_2$  (12),

$$12 = \frac{32 \cdot 27}{72}$$

- in quella posta all'incrocio tra  $A_1$  e  $B_3$  (2),

$$2 = \frac{9 \cdot 16}{72}$$

Utilizzando le formule abbreviate, in uno schema con **repliche proporzionali**, in cui

- ogni riga  $i$  abbia  $q$  livelli
- ogni colonna  $j$  abbia  $p$  livelli
- ogni  $k$  cella abbia un numero variabile di repliche  $n_{ij}$ ,

si stima

- **la devianza totale**  $SQ(T)$  con  $gdl = N-1$  (nell'esempio,  $72-1 = 71$ )

$$SQ(T) = \sum_{i=1}^q \sum_{j=1}^p \sum_{k=1}^{n_{ij}} X_{ijk}^2 - \frac{(\sum_{i=1}^q \sum_{j=1}^p \sum_{k=1}^{n_{ij}} X_{ijk})^2}{N}$$

- **la devianza tra le celle**  $SQ(\bar{X}_{ij})$  con  $gdl = q \times p - 1$  (nell'esempio,  $4 \times 3 - 1 = 11$ )

$$SQ(\bar{X}_{ij}) = \sum_{i=1}^q \sum_{j=1}^p \frac{(\sum_{k=1}^{n_{ij}} X_{ijk})^2}{n_{ij}} - \frac{(\sum_{i=1}^q \sum_{j=1}^p \sum_{k=1}^{n_{ij}} X_{ijk})^2}{N}$$

- **la devianza del fattore A**  $SQ(A)$  con  $gdl = q-1$  (nell'esempio,  $4-1 = 3$ )

$$SQ(A) = \sum_{i=1}^q \frac{(\sum_{j=1}^p \sum_{k=1}^{n_{ij}} X_{ijk})^2}{\sum_{j=1}^p n_{ij}} - \frac{(\sum_{i=1}^q \sum_{j=1}^p \sum_{k=1}^{n_{ij}} X_{ijk})^2}{N}$$

- **la devianza del fattore B**  $SQ(B)$  con  $gdl = p-1$  (esempio,  $3-1 = 2$ )

$$SQ(B) = \sum_{j=1}^p \frac{(\sum_{i=1}^q \sum_{k=1}^{n_{ij}} X_{ijk})^2}{\sum_{i=1}^q n_{ij}} - \frac{(\sum_{i=1}^q \sum_{j=1}^p \sum_{k=1}^{n_{ij}} X_{ijk})^2}{N}$$

- **la devianza d'interazione**  $SQ(AB)$  con  $gdl = (q \cdot p - 1) - (q - 1) - (p - 1) = (q - 1) \times (p - 1)$  (nell'esempio,  $11 - 3 - 2 = 3 \cdot 2 = 6$ )

$$SQ(AB) = SQ(\bar{X}_{ij}) - SQ(A) - SQ(B)$$

- **la devianza d'errore**  $SQ(e)$  con  $gdl (N - 1) - (q \times p - 1)$ ; (nell'esempio,  $71 - 11 = 60$ )

$$SQ(e) = SQ(T) - SQ(\bar{X}_{ij})$$

Se le frequenze di osservazioni entro ogni cella non è proporzionale (l'uguaglianza del numero di repliche ne rappresenta solo un caso particolare), le formule sono più complesse e la loro applicazione diventa possibile solo con programmi informatici. E' quindi utile cercare di condurre i casi che se ne allontanano di poco allo schema del modello proporzionale,

- eliminando casualmente l'osservazione in eccesso entro le celle interessate,
- inserendo il dato mancante.

Come stimare un dato mancante nei casi più semplici è come correggere le varianze è già stato presentato nel capitolo precedente. In questo caso, G. P. **Shearer** nel 1973 (vedi l'articolo *Missing data in quantitative designs*. J. Royal Statist. Soc. Ser. C Appl. Statist. 22: 135-140), ripresa da Jerrold H. **Zar** nel testo *Biostatistical Analysis* del 1999 (fourth edition, Prentice Hall, New Jersey), propone di inserire il valore  $\hat{X}_{ijk}$  stimato con

$$\hat{X}_{ijk} = \frac{aA_i + bB_j - \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} X_{ijk}}{N + 1 - a - b}$$

dove

- $a$  = numero di livelli del fattore A (nell'esempio, è uguale a 4),
- $A_i$  = somma di tutti i valori nel livello  $i$  del fattore **A** (ovviamente senza il dato mancante),
- $b$  = numero di livelli del fattore B (nell'esempio, è uguale a 3),
- $B_j$  = somma di tutti i valori nel livello  $j$  del fattore **B** (ovviamente senza il dato mancante).

Se il numero di dati mancanti è superiore a 1, ma secondo vari autori non devono superare il 10% del numero totale di osservazioni, si deve usare un metodo iterativo, già illustrato nei suoi concetti nel capitolo precedente.

Nel caso di disegni sperimentali non proporzionali, la procedura è più complessa. In tutti questi casi, la soluzione reale è demandata ai programmi informatici.

E' importante ricordare che

- **eliminare uno o più dati riduce** di altrettanto i gdl totali e della devianza d'errore, senza incidere sulle altre;
- **aggiungere uno o più dati non modifica** i gdl totali e della devianza d'errore.

### 10.5. IL TEST T DI TUKEY PER IL CONFRONTO TRA LE MEDIE IN DISEGNI A DUE FATTORI CON REPLICHE

Quando almeno uno dei tre test F risulta significativo, è possibile chiedersi tra quali medie si abbia una differenza effettiva, volendo confrontare quelle di riga, quelle di colonna e/o quelle di casella.

Con un numero costante di repliche entro ogni casella e molte medie, i metodi più semplici sono quelli fondati sulla differenza minima significativa. Tra essi il metodo di Tukey è uno di quelli ai quali si ricorre con maggior frequenza.

Con la stessa tabella dei valori del Q studentizzato già riportata, per ognuno dei tre gruppi di medie sono da considerarsi significative le differenze che risultano uguali o maggiori del valore critico T determinato da

$$T = Q_{\alpha, p, np(r-1)} \cdot \sqrt{\frac{s_e^2}{k}}$$

dove

- il primo indice ( $\alpha$ ) è la probabilità prescelta,
- il secondo indice ( $p$ ) è il numero di medie a confronto simultaneo,
- il terzo indice ( $np(r-1)$ ) è il numero di gradi di libertà della varianza d'errore (con  $n$  uguale al numero di medie del secondo fattore ed  $r$  uguale al numero di repliche),
- $s_e^2$  è la varianza d'errore,
- $k$  è il numero di dati sui quali sono calcolate le medie a confronto.

Nel caso del confronto tra varie medie,

per il fattore A si utilizza la formula

$$T = Q_{\alpha, p, np(r-1)} \cdot \sqrt{\frac{s_e^2}{nr}}$$

per il fattore B la formula,

$$T = Q_{\alpha, n, np(r-1)} \cdot \sqrt{\frac{s_e^2}{pr}}$$

per le medie di casella la formula

$$T = Q_{\alpha, np, np(r-1)} \cdot \sqrt{\frac{S_e^2}{r}}$$

dove

- il secondo indice (p, n o np) sono il numero di medie a confronto simultaneo,
- i denominatori (nr, pr, r) sono il numero di dati sui quali sono stimate le medie a confronto.

Le tre formule di seguito riportate

$$\sqrt{\frac{S_e^2}{nr}}, \quad \sqrt{\frac{S_e^2}{pr}}, \quad \sqrt{\frac{S_e^2}{r}}$$

sono, nell'ordine, l'errore standard delle medie del fattore A, quello del fattore B e quello delle medie di casella.

ESEMPIO.

Con i dati dell'esercizio precedente (di cui vengono riportate tutte le medie),

		TRATTAMENTO			
SESSO		a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	Medie
b <sub>1</sub>		<b>20,0</b>	<b>24,6</b>	<b>16,0</b>	20,200
b <sub>2</sub>		<b>31,4</b>	<b>28,6</b>	<b>18,4</b>	26,133
Medie		25,7	26,6	17,2	23,166

nel quale con

$$S_e^2 = 8,75; \quad n = 2; \quad p = 3; \quad r = 5; \quad \alpha = 0.05$$

per il confronto tra le medie dei tre livelli del fattore A, si utilizza il valore di

$$Q_{3,24} = 3,53$$

per il confronto tra le medie dei due livelli del fattore B si utilizza

$$Q_{2,24} = 2,92$$

e per le 6 medie di casella, ottenute dalle combinazioni dei tre livelli del fattore A e dei due livelli del fattore B, si utilizza

$$Q_{6,24} = 4,37$$

Il confronto tra le 3 medie del fattore A ( $a_1 = 25,7$   $a_2 = 26,6$   $a_3 = 17,2$ )

	$a_1 = 25,7$	$a_2 = 26,6$	$a_3 = 17,2$
$a_2 = 26,6$	<b>0,9</b>	---	---
$a_3 = 17,2$	<b>8,5</b>	<b>9,4</b>	---

determina 3 differenze (in grassetto nella tabella); di esse sono significative alla probabilità  $\alpha = 0.05$  quelle che superano la quantità

$$T = 3,53 \cdot \sqrt{\frac{8,75}{10}} = \mathbf{3,30}$$

Risultano significative la differenza tra  $a_1$  e  $a_3$  (8,5) e quella tra  $a_2$  e  $a_3$  (9,4) mentre non è significativa la differenza tra le medie di  $a_1$  e  $a_2$  (0,9).

Il fattore B ha 2 medie ( $b_1 = 20,200$   $b_2 = 26,133$ ) e quindi una sola differenza (**5,933**). Il test F ha già dimostrato che essa è significativa; tuttavia, applicando ugualmente a scopo didattico il test T di Tukey

$$T = 2,92 \cdot \sqrt{\frac{8,75}{15}} = \mathbf{2,23}$$

si dimostra che la differenza (5,933) risulta significativa, perché maggiore del valore del T calcolato (2,23).

Il test T di Tukey dimostra la sua utilità soprattutto nel caso di confronti tra più medie: con N medie il numero di confronti è

$$\frac{N \cdot (N - 1)}{2}$$

Tra 6 gruppi dell'esempio, il numero di differenze è uguale a 15 ( $\frac{6 \cdot 5}{2}$ )

(tutte riportate in grassetto nella tabella sottostante):

		<b>a<sub>1</sub>b<sub>1</sub></b>	<b>a<sub>2</sub>b<sub>1</sub></b>	<b>a<sub>3</sub>b<sub>1</sub></b>	<b>a<sub>1</sub>b<sub>2</sub></b>	<b>a<sub>2</sub>b<sub>2</sub></b>	<b>a<sub>3</sub>b<sub>2</sub></b>
		20,0	24,6	16,0	31,4	28,6	18,4
<b>a<sub>2</sub>b<sub>1</sub></b>	24,6	<b>4,6</b>	---	---	---	---	---
<b>a<sub>3</sub>b<sub>1</sub></b>	16,0	<b>4,0</b>	<b>8,6*</b>	---	---	---	---
<b>a<sub>1</sub>b<sub>2</sub></b>	31,4	<b>11,4*</b>	<b>6,8*</b>	<b>15,4*</b>	---	---	---
<b>a<sub>2</sub>b<sub>2</sub></b>	28,6	<b>8,6*</b>	<b>4,0</b>	<b>12,6*</b>	<b>2,8</b>	---	---
<b>a<sub>3</sub>b<sub>2</sub></b>	18,4	<b>1,6</b>	<b>6,2*</b>	<b>2,4</b>	<b>13,0*</b>	<b>10,2*</b>	---

Tra esse risultano significative alla probabilità  $\alpha = 0.05$  quelle che superano la quantità

$$T = 4,37 \cdot \sqrt{\frac{8,75}{5}} = 5,78$$

T uguale a 5,78 (le 9 differenze con l'asterisco).

L'analisi può essere fatta contemporaneamente per più livelli di probabilità (oltre a 0.05 anche 0.01 e/o 0.001), riportando uno, due o tre asterischi vicino alle differenze che sono significative per la probabilità minore.

La rappresentazione tabellare, come quelle grafiche, permette di evitare la lunga descrizione necessaria ad elencare in modo dettagliato quali siano, e a che livello di probabilità, le differenze significative.

## 10.6. ESPERIMENTI FATTORIALI 2 X 2 E 2<sup>3</sup> CON I CONFRONTI ORTOGONALI

I metodi per calcolare le devianze ed i relativi gdl, nei disegni complessi a più fattori in cui si analizzano anche le interazioni, sono numerosi. In vari testi di statistica applicata dei decenni scorsi, è riportato anche il **metodo dei confronti ortogonali**; erano di grande utilità soprattutto in un periodo in cui i calcoli erano svolti manualmente, al massimo con l'aiuto delle calcolatrici da tavolo, per ridurre la probabilità di errori e abbreviare i tempi necessari. Ora, con i risultati ottenuti mediante programmi informatici, per i quali si richiede solo una corretta impostazione dell'input, diventa essenziale comprendere quali concetti sono alla base delle varie analisi, come si devono interpretare gli output e quale è l'esatto significato dei risultati ottenuti.

A questo fine, come generalizzazione dei metodi già presentati per due fattori con repliche, è utile considerare due casi semplici: un esperimento fattoriale 2 x 2 ed uno, relativamente più complesso, di 2 modalità per 3 fattori (2<sup>3</sup>).

In un **esperimento fattoriale 2 x 2** (cioè **2 modalità x 2 fattori = 4 trattamenti**), il fattore A ed il fattore B hanno due sole modalità: assente ( $a_0$  e  $b_0$ ) e presente ( $a_1$  e  $b_1$ ) oppure dose minima ( $a_1$  e  $b_1$ ) e dose massima ( $a_2$  e  $b_2$ ); di solito i dati sono riportati in una tabella come quella sottostante

		Fattore A	
		$a_0$	$a_1$
Fattore B			
$b_0$		<b><math>a_0b_0</math></b> repliche	<b><math>a_1b_0</math></b> repliche
$b_1$		<b><math>a_0b_1</math></b> repliche	<b><math>a_1b_1</math></b> repliche

Questo disegno sperimentale non deve essere confuso con quello del test  $\chi^2$  in tabelle 2 x 2, che serve per valutare l'associazione, mediante il confronto di frequenze o conteggi di variabili qualitative o nominali. In questo disegno 2 x 2, entro ogni casella sono riportate non le frequenze ma tante misure quante sono le repliche.

Dai confronti tra le 4 medie o i 4 totali (poiché è richiesto un numero di repliche sempre costante), si può stimare la devianza tra trattamenti con 3 gdl; successivamente, è possibile la sua scomposizione in 3 devianze, ognuna con 1 gdl.

ESEMPIO. Con un esperimento di tossicologia sulla crescita di semi, si vogliono valutare gli effetti indipendenti e quelli congiunti di una sostanza (presenza o assenza) e di due diversi livelli di temperatura (20 gradi e 25 gradi).

A questo fine sono stati coltivati 4 gruppi di semi, ognuno con 5 unità o repliche, posti in 4 situazioni differenti:

- il primo gruppo di semi a temperatura di 20° C e senza la sostanza tossica ( $a_1b_1$ );
- il secondo gruppo alla temperatura di 20° C e con la presenza della sostanza tossica ( $a_2b_1$ );
- il terzo alla temperatura di 25° C e senza la sostanza ( $a_1b_2$ );
- il quarto alla temperatura di 25° C e con la presenza della sostanza ( $a_2b_2$ ).

Dopo un mese, sono state misurate le crescite (in mm) di ogni seme:

Temperatura (B)	Sostanza (A)	
	Assente a <sub>1</sub>	Presente a <sub>2</sub>
b <sub>1</sub> (20° C)	10	8
	12	6
	8	8
	12	5
	9	7
b <sub>2</sub> (25 ° C)	15	7
	15	5
	12	7
	14	8
	13	10

Valutare se si è manifestato un effetto significativo dei due fattori e se tra loro esiste interazione.

Risposta. Come primo passaggio è utile calcolare le somme

GRUPPO	a <sub>1</sub> b <sub>1</sub>	a <sub>2</sub> b <sub>1</sub>	a <sub>1</sub> b <sub>2</sub>	a <sub>2</sub> b <sub>2</sub>	Totale
Somma	51	34	69	37	191

dei 4 gruppi e quella totale.

Con esse è possibile stimare:

- la devianza totale

$$(10^2 + 12^2 + 8^2 + 12^2 + 9^2 + 5^2 + 7^2 + 8^2 + 10^2) - 191^2/20 = 2017 - 1824 = 192,95$$

che risulta uguale a 192,5 ed ha 19 gdl,

- la devianza tra i 4 trattamenti

$$51^2/5 + 34^2/5 + 69^2/5 + 37^2/5 - 191^2/5 = 520,2 + 231,2 + 952,2 + 273,8 - 1824,05 = 153,35$$

che risulta uguale a 153,35 ed ha 3 gdl.

- la devianza d'errore, come somma dei quadrati delle differenze di ogni replica rispetto alla media del suo gruppo o, più rapidamente, per differenza,

$$192,95 - 153,35 = 39,6$$

che risulta uguale a 39,6 ed ha 16 (19-3) gdl.

La tabella di sintesi, che riporta anche il valore di F e la sua significatività,

	DEVIANZA	DF	VARIANZA	F	Pr
Totale	192,95	19	---	---	---
Tra Trattamenti	153,35	3	51,11	20,65	< 0.001
Errore	39,6	16	2,475	---	---

evidenzia che tra i 4 trattamenti esiste una differenza significativa.

La sua scomposizione in tre devianze, ognuna con 1 df, è possibile con una metodologia che non si discosta da quella già utilizzata, ma richiama l'approccio dei confronti ortogonali e permette di meglio comprendere sia l'effetto di ognuno dei due fattori che quello della loro interazione.

Tale regola di scomposizione della devianza tra trattamenti segue lo schema seguente

GRUPPO	a <sub>1</sub> b <sub>1</sub>	a <sub>2</sub> b <sub>1</sub>	a <sub>1</sub> b <sub>2</sub>	a <sub>2</sub> b <sub>2</sub>
Somma	51	34	69	37
Fattore A	-1	+1	-1	+1
Fattore B	-1	-1	+1	+1
Interazione AB	-1	+1	+1	-1

dove i gruppi con lo stesso segno devono essere sommati tra loro, per formare due gruppi per ogni confronto.

Per stimare **la devianza dovuta al fattore A**, tutti i dati devono essere riuniti in due soli gruppi: in uno (-1) quelli che sono indicati con a<sub>1</sub> (totale = 51+69 = 120) e nell'altro (+1) quelli con a<sub>2</sub> (totale = 34 + 37 = 71).

**Se fosse vera l'ipotesi nulla (cioè se il fattore A non ha un effetto significativo), le due somme dovrebbero essere uguali** (ricordando che tutti i gruppi hanno lo stesso numero di dati).

Con la formula abbreviata, tale devianza con 1 df è data da

$$\frac{120^2}{10} + \frac{71^2}{10} - \frac{191^2}{20} = 1440 + 504,1 - 1824,05 = 120,05$$

e risulta uguale a 120,05.

Per stimare **la devianza dovuta al fattore B**, devono essere riuniti in un gruppo (-1) quelli che sono indicati con  $b_1$  (totale = 51 + 34 = 85) e nell'altro (+1) quelli con  $b_2$  (totale = 69 + 37 = 106).

Una differenza tra le due somme che risulti minore (oppure maggiore) della precedente indica che l'effetto del fattore B è minore (oppure maggiore) di quello di A. Per una valutazione corretta della significatività di tali effetti, si calcola la devianza

$$\frac{85^2}{10} + \frac{106^2}{10} - \frac{191^2}{20} = 722,5 + 1129,6 - 1824,05 = 22,05$$

che risulta uguale a 22,05 ed ha 1 df.

La **devianza dell'interazione A x B** è ottenuta mediante il confronto tra la somma dei gruppi (-1) in cui A e B sono assenti o presenti in modo congiunto ( $a_1b_1$  e  $a_2b_2$  con totale = 51 + 37 = 88) rispetto a quella dei gruppi (+1) in cui A e B sono presenti separatamente ( $a_2b_1$  e  $a_1b_2$  con totale = 34 + 69 = 103).

Infatti, **se non esistesse interazione e quindi gli effetti di A e B fossero solo additivi, la somma in cui A e B compaiono separatamente dovrebbe essere uguale a quella in cui tali fattori non compaiono o sono presenti contemporaneamente** (sempre quando il numero di repliche è uguale in ogni gruppo).

**La differenza tra le due somme indica anche se l'effetto è di**

- **inibizione**, se la somma delle osservazioni in cui non compaiono oppure sono presenti in modo congiunto è minore,
- **potenziamento**, se questa somma è maggiore di quella dei valori in cui i due fattori sono presenti separatamente.

Con i dati dell'esempio, la devianza d'interazione è data da

$$\frac{88^2}{10} + \frac{103^2}{10} - \frac{191^2}{20} = 774,4 + 1060,9 - 1824,05 = 11,25$$

e risulta uguale a 11,25 con 1 df.

**I risultati delle 3 devianze e dei valori di F con la loro significatività** possono essere riportati in una tabella conclusiva, che evidenzia come la somma delle 3 devianze (A, B, AB) sia uguale a quella tra trattamenti e ne rappresenti la scomposizione.

L'effetto del fattore A, del fattore B e la loro interazione, utilizzando i dati dell'esempio sono verificati mediante un test F con df 1 e 16, il cui valore critico alla probabilità  $\alpha = 0.05$  è uguale a 4,49 e alla probabilità  $\alpha = 0.01$  è uguale a 8.86.

In conclusione sulla crescita dei semi, come risulta nella tabella seguente,

- la presenza della sostanza A inibisce la crescita in modo **altamente significativo** ( $< 0.001$ );
- la temperatura maggiore favorisce la crescita in modo **molto significativo** ( $< 0.01$ );
- l'alta temperatura potenzia l'effetto di riduzione del fattore A in modo **significativo** (con probabilità leggermente inferiore a 0.05)

	DEVIANZA	DF	VARIANZA	F	Pr
Totale	192,95	19	---	---	---
Tra Trattamenti	153,35	3	51,11	20,65	$< 0.001$
<b>A</b>	<b>120,05</b>	<b>1</b>	<b>120,05</b>	<b>48,50</b>	<b><math>&lt; 0.01</math></b>
<b>B</b>	<b>22,05</b>	<b>1</b>	<b>22,05</b>	<b>8,90</b>	<b><math>&lt; 0.01</math></b>
<b>AB</b>	<b>11,25</b>	<b>1</b>	<b>11,25</b>	<b>4,54</b>	<b><math>&lt; 0.05</math></b>
Errore	39,6	16	2,475	---	---

**Con 2 modalità per 3 fattori ( $2^3$ )**, è possibile utilizzare lo stesso schema esteso a 8 gruppi. Per un'analisi impostata sulla presenza-assenza di 3 fattori (A, B, C), si richiedono 8 gruppi bilanciati, che possono essere presentati come nell'ANOVA ad un criterio di classificazione

T R A T T A M E N T I							
$a_0b_0c_0$	$a_1b_0c_0$	$a_0b_1c_0$	$a_0b_0c_1$	$a_1b_1c_0$	$a_1b_0c_1$	$a_0b_1c_1$	$a_1b_1c_1$
replich e	replich e	Replich e	replich e	replich e	replich e	replich e	replich e

Il modello è

$$X_{ijk} = m + a_i + b_j + g_p + ab_{ij} + ag_{ip} + bg_{jp} + abg_{ijp} + e_{ijk}$$

dove

$a_i, b_j, g_p$  indicano gli **effetti principali**,

$ab_{ij}$ ,  $ag_{ip}$ ,  $bg_{jp}$  indicano le **interazioni di primo ordine od interazioni a due fattori**,

$abg_{ijp}$  indica l'**interazione di secondo ordine o interazione a tre fattori**,

$e_{ijk}$  indica l'errore o residuo di ogni dato o replica.

La devianza totale (che con k repliche ha gdl **(8 x k) - 1**),

la devianza tra trattamenti (con gdl **8-1**) e

la devianza d'errore (con gdl **8 x (k-1)**)

possono essere calcolati come nell'analisi della varianza ad un criterio.

La devianza tra trattamenti ( con 7 gdl) può essere scomposta in

- 3 effetti principali,
- 3 interazioni di primo ordine,
- 1 interazione di secondo ordine

mediante una scomposizione ortogonale, rappresentata in uno schema,

FATTORI E INTERAZIONI	T R A T T A M E N T I							
	$a_0b_0c_0$	$a_1b_0c_0$	$a_0b_1c_0$	$a_0b_0c_1$	$a_1b_1c_0$	$a_1b_0c_1$	$a_0b_1c_1$	$a_1b_1c_1$
	0	a	b	c	ab	ac	bc	abc
A	-1	+1	-1	-1	+1	+1	-1	+1
B	-1	-1	+1	-1	+1	-1	+1	+1
C	-1	-1	-1	+1	-1	+1	+1	+1
AB	+1	-1	-1	+1	+1	-1	-1	+1
AC	+1	-1	+1	-1	-1	+1	-1	+1
BC	+1	+1	-1	-1	-1	-1	+1	+1
ABC	-1	+1	+1	+1	-1	-1	-1	+1

dove +1 e -1 indicano i trattamenti che devono essere sommati tra loro, per formare ogni volta i due gruppi che devono essere confrontati.

La logica con cui si raggruppano i trattamenti è semplice:

- per ognuno dei 3 effetti principali (A, B, C) si sommano tra loro i trattamenti in cui il fattore è assente (-1) e quelli in cui è presente (+1);

- per le interazioni di primo ordine (AB, AC, BC) si sommano tra loro i trattamenti in cui è presente un solo fattore (-1) e quelli in cui o sono assenti entrambi o sono entrambi presenti (+1);
- per l'interazione di secondo ordine (ABC) si sommano tra loro i trattamenti in cui i tre fattori sono presenti singolarmente o tutti insieme (+1) e tra loro quelli in cui o sono tutti assenti o sono presenti a coppie (-1).

Per ognuno dei 7 confronti è possibile calcolare la devianza relativa, con 1 df, che è data da

$$\frac{(tot+)^2}{n+} + \frac{(tot-)^2}{n-} - \frac{TOT^2}{n}$$

dove

tot + è il totale del gruppo i cui trattamenti sono indicati con +

tot - è il totale del gruppo i cui trattamenti sono indicati con -

n+ è il numero complessivo delle repliche contenute nel gruppo +

n- è il numero complessivo delle repliche contenute nel gruppo -

TOT è la somma di tutti i dati

n è il numero complessivo di dati o repliche.

ESEMPIO. Per analizzare gli effetti della presenza-assenza di tre sostanze e delle loro interazioni sulla crescita di cavie, sono stati formati 8 gruppi, ognuno di 4 individui. Per un mese, ad ogni gruppo sono state somministrate con il cibo le 3 sostanze a due diversi livelli, secondo lo schema ed i risultati di seguito riportati

	TRATTAMENTI							
	<b>0</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>ab</b>	<b>ac</b>	<b>bc</b>	<b>abc</b>
Repliche	5,6	6,7	6,9	6,5	7,8	7,4	7,6	7,6
	6,2	6,6	7,0	7,1	7,5	7,2	7,9	7,9
	4,5	7,3	7,2	5,8	6,1	6,5	6,9	6,3
	5,2	5,9	6,8	6,0	7,6	7,3	8,1	7,0
<b>Totale</b>	<b>21,5</b>	<b>26,5</b>	<b>27,9</b>	<b>25,4</b>	<b>29,0</b>	<b>28,4</b>	<b>30,5</b>	<b>28,8</b>

in cui le lettere dei trattamenti indicano le sostanze aggiunte nel cibo somministrato e 0 è il controllo.

Calcolare la devianza tra gli 8 trattamenti e la sua scomposizione nei 7 confronti ortogonali, al fine di valutare

- gli effetti dei 3 fattori,

- le loro interazioni di primo ordine e
- quella di secondo ordine.

Risposta.

Ricordando che la somma totale dei 32 dati è uguale a 218,

la **devianza tra trattamenti** è data da

$$\begin{aligned} & (21,5^2 + 26,5^2 + 27,9^2 + 25,4^2 + 29,0^2 + 28,4^2 + 30,5^2 + 28,8^2)/4 - 218^2/32 = \\ & (462,25 + 702,25 + 778,41 + 645,16 + 841 + 806,56 + 930,25 + 829,44)/4 - 47524/32 = \\ & 1498,83 - 1485,125 = 13,705 \end{aligned}$$

e risulta uguale a 13,705 con 7 df.

**Per la sua scomposizione**, è utile costruire la tabella dei confronti ortogonali

	<b>0</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>ab</b>	<b>ac</b>	<b>bc</b>	<b>abc</b>	<b>totale</b>	<b>totale</b>	<b>diff.</b>
	21,5	26,5	27,9	25,4	29,0	28,4	30,5	28,8	+	-	
A	-	+	-	-	+	+	-	+	112,7	105,3	7,4
B	-	-	+	-	+	-	+	+	116,2	101,8	14,4
C	-	-	-	+	-	+	+	+	113,1	104,9	8,2
AB	+	-	-	+	+	-	-	+	104,7	113,3	-8,6
AC	+	-	+	-	-	+	-	+	106,6	111,4	-4,8
BC	+	+	-	-	-	-	+	+	107,3	110,7	-3,4
ABC	-	+	+	+	-	-	-	+	108,6	109,4	-0,8

con i totali dei due gruppi di trattamenti segnati con + e con -.

La **devianza del fattore A**, con 1 df, riportata come prima nella tabella è data

$$\frac{112,7^2}{16} + \frac{105,3^2}{16} - \frac{218^2}{32} = 793,83 + 693,01 - 1485,12 = 1,72$$

e così ogni fattore od interazione fino a quella di secondo ordine (ABC), riportata alla fine della tabella

Per una lettura più facile e una visione d'insieme, questi risultati sono riassunti in una tabella

	DEVIANZA	DF
<b>Tra trattamenti</b>	<b>13,705</b>	<b>7</b>
<b>A</b>	<b>1,72</b>	<b>1</b>
<b>B</b>	<b>6,48</b>	<b>1</b>
<b>C</b>	<b>2,11</b>	<b>1</b>
<b>AB</b>	<b>2,32</b>	<b>1</b>
<b>AC</b>	<b>0,72</b>	<b>1</b>
<b>BC</b>	<b>0,37</b>	<b>1</b>
<b>ABC</b>	<b>0,02</b>	<b>1</b>

per evidenziare che

- la loro somma corrisponde a quella tra trattamenti (a meno delle approssimazioni dei calcoli) e
- a differenze maggiori tra i due gruppi (+ e -) corrispondono ovviamente valori di devianza maggiori.

Per stimare la significatività occorre naturalmente procedere ai test F, mediante il rapporto di ognuna delle 7 varianze con la varianza d'errore.

### 10.7. ESPERIMENTI FATTORIALI CON P FATTORI A K LIVELLI

Con **p** fattori a **k** livelli, il modello additivo dei vari fattori e delle loro interazioni resta immutato, nei principi e nelle modalità di interpretazione, rispetto a quello presentato per esperimenti  $2^3$ ; ma diventa più complesso il calcolo delle devianze. Come dimostrazione, facilmente estensibile, si può utilizzare un esempio a tre fattori, di cui

- il primo (A) a 3 livelli,
- il secondo (B) a 4 livelli,
- il terzo (Sesso) a 2 livelli o modalità

con tre repliche per ognuno dei 24 (3 x 4 x 2) gruppi e quindi 72 dati.

Per **gli effetti principali**, il calcolo delle devianze è identico a quello già ripetutamente presentato.

Per **l'interazione tra due fattori**, si crea una tabella a due entrate di somme e si calcola la devianza relativa; per ottenere la devianza d'interazione e i suoi df, si sottraggono a questa ultima devianza quelle stimate per i singoli fattori.

Per **l'interazione a tre fattori**, si costruisce la tabella a tre entrate delle somme e si calcola la devianza relativa. Per ottenere la devianza d'interazione dei 3 fattori, al valore precedentemente calcolato si sottraggono sia le 3 devianze d'interazione a due fattori che le 3 degli effetti principali.

Per i df si segue lo stesso metodo.

Con i dati

	<b>a<sub>1</sub></b>		<b>a<sub>2</sub></b>		<b>a<sub>3</sub></b>	
	M	F	M	F	M	F
<b>b<sub>1</sub></b>	1,3	1,1	1,4	1,2	1,1	0,9
	1,5	0,9	1,7	1,5	1,2	0,8
	1,1	0,9	1,4	1,0	0,9	0,5
<b>b<sub>2</sub></b>	1,7	0,5	1,9	1,0	1,4	1,1
	1,6	0,8	2,1	1,7	1,5	1,2
	1,9	1,1	2,0	1,2	1,2	1,0
<b>b<sub>3</sub></b>	1,1	0,8	1,5	0,5	1,5	0,9
	1,1	0,7	1,7	0,6	1,2	0,7
	1,4	0,8	1,2	0,9	1,7	0,9
<b>b<sub>4</sub></b>	1,2	0,9	1,9	0,9	1,4	1,2
	1,5	1,3	2,3	1,1	1,2	1,3
	1,3	1,0	2,1	1,3	1,0	1,2

- la **devianza totale** è ottenuta come al solito a partire dagli scarti di ogni valore dalla media generale o con la formula abbreviata; il suo valore è 11,1329 ed ha 71 df;

- la **devianza del fattore A** è ottenuta dal confronto delle 3 medie di colonna: è uguale a 1,3086 ed ha 2 df;

- la **devianza del fattore B** è ottenuta dal confronto delle 4 medie di riga: è uguale a 1,2850 ed ha 3 df;

- la **devianza del fattore sesso** è ottenuta dal confronto tra le 2 medie relative: è uguale a 4,4006 ed ha 1 df.

Successivamente,

- per **la devianza d'interazione AB**, si deve dapprima calcolare la tabella delle somme dei fattori A e B, ignorando la distinzione tra sessi.

	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	Totali
b <sub>1</sub>	6,8	8,2	5,4	20,4
b <sub>2</sub>	7,6	9,9	7,4	24,9
b <sub>3</sub>	5,9	6,4	6,9	19,2
b <sub>4</sub>	7,2	9,6	7,3	24,1
Totali	27,5	34,1	27,0	88,6

Mediante le 12 somme (ognuna è la somma di 6 valori, 3 per ogni sesso come evidenziato nella tabella dei dati sperimentali) riportate nelle caselle all'incrocio dei 3 livelli di A e dei 4 livelli di B, si ottiene

- la **devianza tra le medie dei 12 trattamenti**

$$(6,8^2 + 7,6^2 + 5,9^2 + 7,2^2 + 8,2^2 + 9,9^2 + 6,4^2 + 9,6^2 + 5,4^2 + 7,4^2 + 6,9^2 + 7,3^2)/6 - 88,6^2/72 =$$

$$(46,24 + 57,76 + 34,81 + 51,84 + 67,24 + 98,01 + 40,96 + 92,16 + 29,16 + 54,76 + 47,61 + 53,29)/6 - 7849,96/72 = 673,79/6 - 7849,96/72 = 112,3066 - 109,0272 = 3,2794$$

che risulta uguale a 3,2794 con 11 df.

La **devianza d'interazione A x B** ed i suoi df sono stimati mediante la differenza

$$\text{Dev. tra i trattamenti AB} - \text{Dev. A} - \text{Dev. B} = \text{Dev. d'interazione A x B}$$

che, con i dati dell'esempio,

$$3,2794 - 1,3086 - 1,2850 = 0,6858$$

risulta uguale a 0,6858 con

$$11 - 2 - 3 = 6$$

6 df.

Come già dimostrato nei paragrafi precedenti con 2 fattori, questa devianza può essere ottenuta mediante la somma degli scarti al quadrato tra le 12 medie osservate e le medie attese, nell'ipotesi che non esista interazione. Poiché le medie attese sono stimate dai totali marginali, nella tabella sopra riportata di dimensioni 4 x 3 i df dell'interazione sono 3 x 2 = 6 come ottenuto mediante il metodo delle differenze.

Questo metodo può essere esteso alle 2 devianze d'interazione successive.

Per misurare **la devianza d'interazione A x sesso**, si devono dapprima calcolare le somme relative a questi due fattori, come nella tabella sottostante, ignorando l'effetto del fattore B

	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	Totali
M	<b>16,7</b>	<b>21,2</b>	<b>15,3</b>	53,2
F	<b>10,8</b>	<b>12,9</b>	<b>11,7</b>	35,4
Totali	27,5	34,1	27,0	88,6

La **devianza dei trattamenti** riportati nelle 6 caselle, in cui ogni valore è la somma di 12 osservazioni (4 trattamenti per 3 repliche), con la formula abbreviata è

$$\begin{aligned} & (16,7^2 + 10,8^2 + 21,2^2 + 12,9^2 + 15,3^2 + 11,7^2)/12 - 88,6^2/72 = \\ & (278,89 + 116,64 + 449,44 + 166,41 + 234,09 + 136,89)/12 - 7849,96/72 = \\ & 1382,36/12 - 7849,96/72 = 115,1967 - 109,0272 = 6,1695 \end{aligned}$$

e risulta uguale a 6,1695 con 5 df.

La **devianza d'interazione A x sesso** e i suoi gdl sono ottenuti rapidamente per differenza

$$\text{Dev. dei trattamenti A x sesso} - \text{Dev. A} - \text{Dev. sesso} = \text{Dev. d'interazione A x sesso}$$

Con i dati dell'esempio,

$$6,1695 - 1,3086 - 4,4006 = 0,4603$$

si ottiene una devianza d'interazione uguale a 0,4603 con

$$5 - 2 - 1 = 2$$

2 df.

Nello stesso modo, per **la devianza d'interazione B x sesso** si costruisce la tabella relativa, in cui i dati delle 8 caselle sono le somme di 9 dati (3 livelli del fattore A per 3 repliche)

	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	b <sub>4</sub>	Totali
M	<b>11,6</b>	<b>15,3</b>	<b>12,4</b>	<b>13,9</b>	53,2
F	<b>8,8</b>	<b>9,6</b>	<b>6,8</b>	<b>10,2</b>	35,4
Totali	20,4	24,9	19,2	24,1	88,6

La **devianza dovuta alle differenze tra le 8 medie di casella** con la formula abbreviata

$$\begin{aligned} & (11,6^2 + 8,8^2 + 15,3^2 + 9,6^2 + 12,4^2 + 6,8^2 + 13,9^2 + 10,2^2)/9 - 88,6^2/72 = \\ & (134,56 + 77,44 + 234,09 + 92,16 + 153,76 + 46,24 + 193,21 + 104,04)/9 - 7849,96/9 = \\ & 1035,5/9 - 7849,96/72 = 115,0556 - 109,0272 = 6,0284 \end{aligned}$$

risulta uguale a 6,0284 e i suoi df sono 7.

La **devianza d'interazione B x sesso**, ottenuta per differenza,

**Dev. dei trattamenti B x sesso - Dev. B - Dev. sesso = Dev. d'interazione B x sesso**

con i dati dell'esempio

$$6,0284 - 1,285 - 4,4006 = 0,3428$$

risulta uguale a 0,3428 e i suoi df

$$7 - 3 - 1 = 3$$

sono 3.

Per **l'interazione di secondo ordine o a 3 fattori** occorre costruire la tabella a 3 fattori

		MASCHI					FEMMINE		
		a <sub>1</sub>	a <sub>2</sub>	A <sub>3</sub>			a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>
b <sub>1</sub>	b <sub>1</sub>	3,9	4,5	3,2	b <sub>1</sub>	2,9	3,7	2,2	
	b <sub>2</sub>	5,2	6,0	4,1		b <sub>2</sub>	2,4	3,9	3,3
	b <sub>3</sub>	3,6	4,4	4,4		b <sub>3</sub>	2,3	2,0	2,5
	b <sub>4</sub>	4,0	6,3	3,6		b <sub>4</sub>	3,2	3,3	3,7

per la quale sono possibili varie soluzioni, tra cui quella sotto-riportata, in due passaggi :

1 - La **devianza che considera contemporaneamente le medie dei 3 fattori** (ognuno con 3 repliche) con la formula abbreviata

$$(3,9^2 + 5,2^2 + 3,6^2 + 4,0^2 + 4,5^2 + 6,0^2 + 4,4^2 + 6,3^2 + 3,2^2 + 4,1^2 + 4,4^2 + 3,6^2 + 2,9^2 + 2,4^2 + 2,3^2 + 3,2^2 + 3,7^2 + 3,9^2 + 2,0^2 + 3,3^2 + 2,2^2 + 3,3^2 + 2,5^2 + 3,7^2)/3 - 88,6^2/72 =$$

$$355,04/3 - 7849,96/72 = 118,3467 - 109,0272 = 9,3195$$

risulta uguale a 9,3195 ed ha 23 df.

2 - La **devianza d'interazione dei tre fattori** è ottenuta per sottrazione da questa ultima stima di tutte le precedenti già stimate

**Dev. tra medie di caselle - (Dev. A + Dev. B + Dev. sesso + Dev. A x B + Dev. A x sesso + Dev. B x sesso) = Dev. A x B x sesso**

Con i dati dell'esempio,

$$9,3195 - (1,3086 + 1,2850 + 4,4006 + 0,6858 + 0,4603 + 0,3428) = 0,8634$$

risulta uguale a 0,8634

e i suoi df

$$23 - (3 + 2 + 1 + 2 + 3 + 6) = 6$$

sono 6.

Per una visione generale ed una sintesi, i risultati ottenuti dai calcoli qui presentati sono di norma riportati in una tabella riassuntiva:

FATTORI e INTERAZIONI	DEVIANZA	DF	F	Pr > F
<b>TOTALE</b>	<b>11,1328</b>	<b>71</b>		
<b>TRA CASELLE A x B x sesso</b>	<b>9,3195</b>	<b>23</b>	<b>10,73</b>	<b>0.0001</b>
<b>FATTORE A</b>	<b>1,3086</b>	<b>2</b>	<b>17,32</b>	<b>0.0001</b>
<b>FATTORE B</b>	<b>1,2850</b>	<b>3</b>	<b>11,34</b>	<b>0.0001</b>
<b>FATTORE SESSO</b>	<b>4,4006</b>	<b>1</b>	<b>116,49</b>	<b>0.0001</b>
TABELLA A x B	3,2794	11		
<b>INTERAZIONE A x B</b>	<b>0,6858</b>	<b>6</b>	<b>3,03</b>	<b>0.0137</b>
TABELLA A x SESSO	6,1695	5		
<b>INTERAZIONE A x SESSO</b>	<b>0,4603</b>	<b>2</b>	<b>6,09</b>	<b>0.0044</b>
TABELLA B x SESSO	6,0284	7		
<b>INTERAZIONE B x SESSO</b>	<b>0,3428</b>	<b>3</b>	<b>3,02</b>	<b>0,0385</b>
<b>INTERAZIONE A x B x SESSO</b>	<b>0,8364</b>	<b>6</b>	<b>3,69</b>	<b>0.0043</b>
<b>ERRORE</b>	<b>1,8133</b>	<b>48</b>		

in cui i dati non in grassetto (quelli riferiti alle tabelle a due fattori) non sono riportate negli output dei programmi informatici, in quanto utili solo al calcolo manuale.

Gli unici dati riferiti a una tabella che vengono riportati sono quelli della tabella dei 3 fattori, poiché tutte le devianze successive sono una sua scomposizione.

**Per l'interpretazione delle interazioni è utile ricorrere alle tabelle relative in cui sono riportate le medie osservate e quelle attese nell'ipotesi di assenza d'interazione, oppure alla rappresentazione grafica, come già illustrato nei paragrafi precedenti.**

## 10.8. QUADRATI LATINI CON REPLICHE

Nella presentazione del disegno sperimentale a quadrati latini è stato ripetutamente evidenziato il suo limite operativo, imposto dall'esigenza che i tre fattori a confronto abbiano lo stesso numero di livelli.

Di conseguenza, si è indotti ad effettuare esperimenti di piccole dimensioni, con la forte

controindicazione che il numero di df dell'errore è molto piccolo: in un esperimento 3 x 3, i test F relativi alla significatività di ognuno dei tre fattori hanno df 2 e 2.

Spesso è **utile applicare forme di repliche, che aumentino i df dell'errore**: con df 2 e 2 alla probabilità  $\alpha = 0.05$  il valore critico è uguale a 18,51 e alla probabilità  $\alpha = 0.01$  è uguale a 99,00; già con df 2 e 4 i due valori critici si abbassano rispettivamente a 6,94 e 18,00 rendendo significative differenze che sono circa 1/3 e 1/5 delle prime.

Le modalità con cui si possono avere repliche di quadrati latini, soprattutto in esperimenti di laboratorio, sono numerose. A volte è possibile ripetere l'esperimento utilizzando le stesse cavie, altre volte ricorrendo a cavie appartenenti alla medesima nidiata, sempre distribuite nelle caselle in modo casuale. Negli esperimenti in natura, è possibile **ripetere l'esperimento** in altra località, **con una diversa distribuzione casuale dei fattori**.

In un esperimento con 2 differenti quadrati 3 x 3, estratti a caso dai 12 possibili determinati dalla diversa distribuzione del terzo fattore nelle 9 caselle, si ottengono 18 osservazioni. L'analisi separata dei 2 quadrati, ognuno con 9 osservazioni, facilmente porterebbe a conclusioni non significative per entrambi gli esperimenti.

L'analisi combinata consente

- **una stima della devianza d'errore e della significatività dei due fattori di disturbo con 4 df anziché 2;**
- **l'effetto del fattore ritenuto più importante è saggiato sul numero totale di osservazioni ottenute con i due esperimenti;**
- **infine è possibile l'analisi dell'interazione tra il fattore principale e le repliche, al fine di saggiare eventuali risultati diversi ottenuti con i due quadrati.**
- Se questa ultima interazione risulta significativa, è corretto analizzare separatamente i fattori nei due quadrati, con l'impostazione già presentata nei paragrafi precedenti; se non risulta significativa, è conveniente la loro analisi combinata.

In un disegno sperimentale con **q repliche di un quadrato latino di dimensioni n** è possibile calcolare le devianze sottoriportate, con i relativi df.

DEVIANZE	DF
TOTALE	$q (n^2 - 1)$
Tra quadrati	$q - 1$
Fattore X entro quadrati	$q (n - 1)$
Fattore Y entro quadrati	$q (n - 1)$
Trattamenti entro quadrati	$q (n - 1)$
<b>Trattamenti</b>	<b><math>n - 1</math></b>
<b>Interazione trattamenti x quadrati</b>	<b><math>(q - 1) (n - 1)</math></b>
Errore	$q (n - 1) (n - 2)$

La **devianza tra trattamenti entro quadrati ed i suoi df possono essere scomposti** in due parti:

- la **devianza tra trattamenti con df n-1**, con la quale si saggiano le differenze complessive tra le modalità dei trattamenti, senza considerare le differenze tra i due quadrati latini, ma valutando i dati come risultato di un unico esperimento;
- la devianza **d'interazione trattamenti x quadrati con df (q-1) (n-1)**, con la quale si verifica se il fattore in esame, quello ritenuto più importante dei tre sotto controllo, determina risultati significativamente diversi nei due quadrati.

ESEMPIO. Si intende saggiare la quantità di produzione unitaria di tre varietà dello stesso cereale (Z), considerando anche gli effetti di tre concimi (X) e di tre tipi di aratura (Y).

Per avere un numero di dati sufficientemente alto e per considerare anche l'effetto della località, sono stati impostati due esperimenti a quadrati latini: uno in pianura e l'altro in montagna, con i risultati che sono stati riportati nelle due tabelle.

		PIANURA		
		X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>
Y <sub>1</sub>	<b>B 18</b>	<b>A 12</b>	<b>C 17</b>	
Y <sub>2</sub>	<b>A 14</b>	<b>C 21</b>	<b>B 20</b>	
Y <sub>3</sub>	<b>C 18</b>	<b>B 16</b>	<b>A 12</b>	

		MONTAGNA		
		X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>
Y <sub>1</sub>	<b>C 23</b>	<b>A 16</b>	<b>B 22</b>	
Y <sub>2</sub>	<b>A 18</b>	<b>B 24</b>	<b>C 27</b>	
Y <sub>3</sub>	<b>B 20</b>	<b>C 20</b>	<b>A 15</b>	

Calcolare le devianze dei tre fattori e la devianza d'interazione delle varietà del cereale per località.

Risposta.

Separatamente per i due quadrati latini

		PIANURA			
		X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	Totale
Y <sub>1</sub>	B 18	A 12	C 17		<b>47</b>
Y <sub>2</sub>	A 14	C 21	B 20		<b>55</b>
Y <sub>3</sub>	C 18	B 16	A 12		<b>46</b>
Totale	<b>50</b>	<b>49</b>	<b>49</b>		<b>148</b>

		MONTAGNA			
		X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	Totale
Y <sub>1</sub>	C 23	A 16	B 22		<b>61</b>
Y <sub>2</sub>	A 18	B 24	C 27		<b>69</b>
Y <sub>3</sub>	B 20	C 20	A 15		<b>55</b>
Totale	<b>61</b>	<b>60</b>	<b>64</b>		<b>185</b>

e congiuntamente per le sementi si calcolano i totali

		SEMENTI			
		A	B	C	Totale
Pianura		38	54	56	148
Montagna		49	66	70	185
Totale		87	120	126	333

Con le solite modalità già illustrate per i quadrati latini, si possono ottenere le devianze e i df sia per l'esperimento in pianura che per quello in montagna.

		PIANURA		MONTAGNA	
		DEVIANZA	DF	DEVIANZA	DF
Tra dosi di X		0,22	2	2,89	2
Tra dosi di Y		16,22	2	32,89	2
Tra sementi		64,89	2	82,89	2
Errore		132,67	2	1,55	2
Totale		214,00	8	120,22	8

Mediante l'analisi congiunta dei due quadrati latini è possibile calcolare:

- la **devianza totale**, che è uguale alla somma degli scarti al quadrato di ognuno dei 18 valori dalla media generale, e può essere ottenuta come

$$\sum x_{ijk}^2 - \left( \sum x_{ijk} \right)^2 / n$$

(con n = 18 nel caso dell'esempio);

il valore calcolato è 410,28 con 17 df;

- la **devianza tra quadrati** (o tra località)

$$\frac{148^2}{9} + \frac{185^2}{9} - \frac{333^2}{18} = 2433,78 + 3802,78 - 6160,5 = 76,06$$

che risulta uguale a 76,06 con 1 df;

- la **devianza tra sostanza X entro località**, o entro quadrati,

$$\left(\frac{50^2}{3} + \frac{49^2}{3} + \frac{49^2}{3} - \frac{148^2}{9}\right) + \left(\frac{61^2}{3} + \frac{60^2}{3} + \frac{64^2}{3} - \frac{185^2}{9}\right)$$

che ovviamente coincide con la somma delle due devianze del fattore X calcolate entro ogni quadrato

$$0,22 + 2,89 = 3,11$$

e risulta uguale a 3,11 con 4 df (2 + 2);

- la **devianza tra sostanza Y entro località**

$$\left(\frac{47^2}{3} + \frac{55^2}{3} + \frac{46^2}{3} - \frac{148^2}{9}\right) + \left(\frac{61^2}{3} + \frac{69^2}{3} + \frac{55^2}{3} - \frac{185^2}{9}\right)$$

che coincide con la somma delle due devianze calcolate separatamente

$$16,22 + 32,89 = 49,11$$

e risulta uguale a 49,11 con 4 df;

- la **devianza tra sementi entro località**

$$\left(\frac{38^2}{3} + \frac{54^2}{3} + \frac{56^2}{3} - \frac{148^2}{9}\right) + \left(\frac{49^2}{3} + \frac{66^2}{3} + \frac{70^2}{3} - \frac{185^2}{9}\right)$$

che coincide con la somma delle due devianze relative

$$64,89 + 82,89 = 147,78$$

e risulta uguale a 147,78 con 4 df.

Questa ultima devianza calcolata entro quadrati, in particolare quando si considerano le sementi come il fattore più importante e gli altre due fattori come fonti di perturbazione da controllare, può essere suddivisa in due devianze:

- la **devianza tra sementi**

$$\frac{87^2}{6} + \frac{120^2}{6} + \frac{126^2}{6} - \frac{333^2}{18} = 1261,5 + 2400 + 2646 - 6160,5 = 147$$

che risulta uguale a 147 ed ha 2 df;

- la **devianza d'interazione sementi x località**

che è ottenuta per differenza

$$\text{dev. tra sementi entro località} - \text{dev. tra sementi} = \text{dev. d'interazione sementi x località}$$

$$147,78 - 147 = 0,78$$

e che, con i dati dell'esempio, risulta uguale a 0,78 ed ha 2 df.

**Questa scomposizione potrebbe essere fatta contemporaneamente per tutti e tre i fattori.**

Infine si calcola la devianza d'errore, che può essere facilmente ottenuta come somma delle due devianze d'errore calcolate in ogni quadrato:

con i dati dell'esempio

$$132,67 + 1,55 = 134,22$$

risulta uguale a 134,22 ed ha 4 df.

E' sempre utile presentare tutte le devianze in una tabella

FATTORI E INTERAZIONI	DEVIANZA	DF	VARIANZA
Totale	410,28	17	
Tra località	76,06	1	76,06
Fattore X entro località	3,11	4	0,78
Fattore Y entro località	49,11	4	12,28
Sementi entro località	147,78	4	36,94
<b>Sementi</b>	<b>147</b>	<b>2</b>	<b>73,5</b>
<b>Interazione sementi x località</b>	<b>0,78</b>	<b>2</b>	<b>0,39</b>
Residuo	134,22	4	33,55

che riassume i risultati di tutti i calcoli effettuati ed eventualmente la significatività di ogni test F.

### 10.9. ANALISI GERARCHICA (Nested ANOVA)

L'analisi della varianza a due o più fattori di variazione, come dimostrato con i blocchi randomizzati, con i quadrati latini e **k** fattori, prevede che **tutti i livelli di ogni fattore incontrino tutti i livelli degli altri fattori**. Ma non sempre è possibile programmare un esperimento e raccogliere i dati in questo modo. Soprattutto nella ricerca ambientale, in varie situazioni si devono suddividere le osservazioni in gruppi tra loro totalmente indipendenti, ciascuno ulteriormente suddiviso in sottogruppi che non incontrano mai i diversi livelli degli altri fattori. Un caso classico sono le stazioni di rilevazione lungo due fiumi, che non sono tra loro direttamente comparabili, anche se chiamate con nomi generici come alla sorgente, a medio percorso e alla foce.

E' il **disegno sperimentale gerarchico**, che **offre il vantaggio di non richiedere il bilanciamento completo delle osservazioni** e quindi di essere meno rigido; esso esige solo **che**

- **ogni gruppo principale sia diviso in almeno 2 sottogruppi** e che
- **i vari fattori previsti dal programma dell'esperimento debbano avere una successione prestabilita** (in gruppi, sottogruppi, sotto-sottogruppi, ecc. ...).

Nella letteratura inglese, questo tipo di analisi della varianza è detto **Nested ANOVA**; ossia analisi della varianza con **criteri di classificazione annidati l'uno dentro l'altro**.

Questo metodo è stato applicato inizialmente in agraria, in disegni sperimentali chiamati **split-plot** (dalla suddivisione dell'appezzamento) o disegni **split-unit** (dalla suddivisione dell'unità principale di analisi). Per verificare l'effetto di 2 fattori a vari livelli, si devono scegliere tanti appezzamenti di terreno quanti sono i livelli del fattore più importante e si procede ad una attribuzione casuale di ogni appezzamento ad una specifica modalità del fattore; successivamente, ogni appezzamento deve essere suddiviso in tanti lotti, tra loro uguali, quante sono le modalità del secondo fattore e si procede alla loro attribuzione casuale.

**Il confronto tra le modalità del secondo fattore**, analizzato entro la suddivisione principale, sovente è **soggetto ad una variabilità minore di quella che interessa il primo fattore**. La differenza di precisione è comunque accettata, quando ritenuta ridotta o ineliminabile. Nella ricerca ambientale, è il caso di confronti dei livelli d'inquinamento tra due o più città. Poiché ognuna di esse è geograficamente diversa da tutte le altre, entro ognuna possono essere effettuate rilevazioni della quantità d'inquinamento in zone funzionalmente simili, come un'area industriale, una zona residenziale ed un'altra commerciale; ma esse ovviamente sono diverse tra città e risentono in modo ineliminabile delle sue caratteristiche geografiche ed urbanistiche.

E' un disegno gerarchico, in cui le differenze tra aree (sottogruppi) devono essere analizzate entro la stessa unità di rilevazione, la loro città (gruppo).

Negli esperimenti di laboratorio, il disegno gerarchico ricorre sovente in incroci tra maschi e femmine per valutare l'effetto dei vari genotipi in una serie contemporanea d'incroci, poiché ogni femmina può essere ingravidata da un solo maschio, mentre ogni maschio può divenire contemporaneamente padre di più nidiate. I figli devono essere classificati dapprima secondo il padre (gruppo) e successivamente distinti per madre (sottogruppo).

Per un esempio più particolareggiato, si supponga di considerare 2 città (I e II, con  $k = 2$ ) e che nella prima siano state campionate 3 zone (A, B, C) e nella seconda 4 (D, E, F, G) per un totale di 7 ( $p = 7$ ):

CITTA'	I			II			
ZONE	A	B	C	D	E	F	G
Repliche	$X_{111}$ $X_{112}$ $X_{11r}$	$X_{121}$  $X_{12r}$	$X_{131}$   			$X_{231}$  $X_{23r}$	$X_{241}$ $X_{242}$  $X_{ijr}$
Medie delle zone	$\bar{x}_{11}$						$\bar{x}_{ij}$
Medie delle città	$\bar{x}_1$			$\bar{x}_i$			
Media generale	$\bar{\bar{x}}$						

si supponga inoltre, solo per semplificare la presentazione ma **non richiesto dalla metodologia, che in ognuna sia stato effettuato lo stesso numero di repliche** ( $n = 5$ ).

Per un'analisi gerarchica o nested ANOVA, è utile rappresentare i dati in una tabella come quella precedente, che evidenzia l'**annidamento** del fattore di I livello (la zona) entro quella di II livello (la città) e in cui

$X_{ijr}$  è una generica osservazione,

$\bar{x}_{ij}$  è la media del sottogruppo j-esimo entro il gruppo i-esimo,

$\bar{x}_i$  è la media del gruppo i-esimo,

$\bar{\bar{x}}$  è la media generale.

Con tale disegno sperimentale, rispetto a quello a blocchi randomizzati

- **le ipotesi da verificare restano immutate** sia in riferimento al fattore del primo livello

$$H_0: \mu_I = \mu_{II} \quad \text{con} \quad H_1: \mu_I \neq \mu_{II}$$

che al fattore del secondo livello

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D = \mu_E = \mu_F = \mu_G \quad \text{con} \quad H_1: \text{non tutte le } \mu \text{ delle zone uguali,}$$

- mentre **a partire dai sottogruppi cambiano le procedure di calcolo della devianza e del relativo numero di gdl.**

In modo più dettagliato:

- il calcolo della **devianza totale** e dei suoi df è identico a tutti gli altri casi già presentati:

con N osservazioni è

$$\sum_{i,j,r} (X_{ijr} - \bar{\bar{x}})^2 = \sum_{i,j,r} X_{ijr}^2 - (\sum_{i,j,r} X_{ijr})^2 / N$$

ed ha N-1 df,

- **la devianza tra gruppi**, con k gruppi e media del gruppo i-esimo  $\bar{x}_i$  di  $n_i$  osservazioni, è ancora uguale

$$\sum_i n_i (\bar{x}_i - \bar{\bar{x}})^2 = \sum_i (\sum_{j,r} X_{ijr})^2 / n_i - (\sum_{i,j,r} X_{ijr})^2 / N$$

ed i suoi df sono k-1,

- mentre **la devianza tra sottogruppi entro gruppi** è valutata come differenza tra la media di ogni sottogruppo ed il suo gruppo d'appartenenza:

con p sottogruppi entro il gruppo i-esimo e media  $\bar{x}_{ij}$  con  $n_{ij}$  osservazioni è

$$\sum_{i,j} n_{ij} (\bar{x}_{ij} - \bar{x}_i)^2 = \sum_{i,j} (\sum_r X_{ijr})^2 / n_{ij} - \sum_i (\sum_{j,r} X_{ijr})^2 / n_i$$

con df che sono

uguali al numero complessivo di sottogruppi (p) meno il numero (k) di gruppi (p-k).

La **devianza entro sottogruppi o errore** è la somma dei quadrati degli scarti di ogni osservazione r dalla media del sottogruppo (quello di ordine inferiore) al quale appartiene

$$\sum_{i,j,r} (X_{ijk} - \bar{x}_{ij})^2 = \sum_{i,j,r} X_{ijr}^2 - \sum_{i,j} (\sum_r X_{ijr})^2 / N$$

e i df sono

uguali al numero complessivo di osservazioni N meno il numero (p) di sottogruppi (N - p).

Nella stima delle devianze e nella costruzione della tabella riassuntiva, deve essere posta particolare attenzione al calcolo dei df, che permette di meglio comprendere i confronti effettuati. Con i dati riportati come esempio nella tabella precedente (N = 35; k = 2; P = 7), la sua distribuzione è

DEVIANZE	DF	
Totale	N - 1	34
Tra gruppi	k - 1	1
Tra sottogruppi (entro gruppi)	<b>p - k</b>	5
Errore	N - p	28

**Con 3 o più fattori programmati in un disegno sperimentale gerarchico**, i calcoli divengono più complessi ed il numero di dati aumenta in modo rilevante; di norma le stime sono effettuate con programmi informatici, che tuttavia per una comprensione corretta richiedono una impostazione chiara dei dati, anche nella loro rappresentazione tabellare.

L'esempio successivo spiega in modo dettagliato il passaggio a 3 fattori.

ESEMPIO. Con un **disegno sperimentale gerarchico non bilanciato**, si intendono confrontare i pesi di cavie neonate di 23 nidiate (generate da altrettante femmine), con un numero variabile di componenti da 3 a 6, per un totale di 106 individui. Si vogliono valutare gli effetti di 3 diversi tipi di alimentazione (fattore I, II, III), le differenze determinate dal genotipo di 9 maschi (indicati con numeri da 1 a 9) e quelle dovute al genotipo delle 23 madri (indicate con lettere maiuscole da A a Z).

Nidiata	Alimentaz.	Maschio	Femmina	n. di figli	Peso di ogni figlio
1	I	1	A	4	20, 22, 24, 22
2	I	1	B	5	22, 25, 21, 26, 22
3	I	2	C	6	27, 22, 24, 28, 21, 28
4	I	2	D	4	38, 36, 35, 38
5	I	2	E	5	31, 32, 35, 38, 32
6	II	3	F	4	29, 32, 30, 28
7	II	3	G	5	33, 39, 40, 31, 32
8	II	3	H	3	28, 27, 25
9	II	4	I	3	26, 27, 28
10	II	4	L	3	31, 32, 33
11	II	5	M	4	28, 32, 30, 26
12	II	5	N	3	33, 26, 43
13	II	6	O	5	27, 27, 26, 28, 30
14	II	6	P	5	35, 28, 28, 35, 30
15	II	6	Q	6	32, 33, 33, 33, 35, 31
16	II	6	R	5	33, 35, 31, 26, 27,
17	III	7	S	5	32, 31, 33, 32, 35
18	III	7	T	6	38, 37, 38, 37, 39, 35
19	III	8	U	5	31, 38, 38, 35, 34
20	III	8	V	5	38, 36, 33, 32, 33
21	III	9	X	6	32, 32, 36, 33, 34, 39
22	III	9	Y	6	31, 32, 32, 30, 33, 30
23	III	9	Z	3	37, 35, 40

Nella impostazione dell'esperimento, è **importante evidenziare che**

- **ogni tipo di alimentazione (gruppi) deve essere stato somministrato ad almeno 2 maschi (sottogruppi),**
- **ogni maschio (sottogruppi) deve avere almeno 2 femmine (sotto-sottogruppi) e**
- **ogni femmina deve avere almeno 2 figli (repliche).**

Se un tipo di alimentazione fosse stato somministrato ad un solo maschio, sarebbe impossibile valutare l'effetto dell'alimentazione in modo indipendente dal genotipo del maschio; nello stesso modo, se si

fossero riportate anche le femmine con un solo figlio, sarebbe divenuto impossibile stimare la **devianza dei figli rispetto alla madre** (identificata dalla media del gruppo dei suoi figli), **corrispondente alla devianza entro sottogruppi od errore.**

L'analisi della varianza, effettuato con un programma informatico standard, fornisce i seguenti risultati:

FONTE	DF	DEVIANZA	VARIANZA	F	Prob>F
Totale	<b>105</b>	2404,12	---	---	---
Nidiate	<b>22</b>	1880,06	85,46	<b>13,53</b>	<b>0.0001</b>
<b>Alimentazione</b>	<b>2</b>	673,03	336,51	<b>53,30</b>	<b>0.0001</b>
<b>Maschi</b>	<b>6</b>	419,95	69,99	<b>11,09</b>	<b>0.0001</b>
<b>Femmine</b>	<b>14</b>	787,08	56,22	<b>8,90</b>	<b>0.0001</b>
Errore	<b>83</b>	524,07	6,31	---	---

Per comprendere la tabella e le analisi, è utile seguire la distribuzione dei df e ricordare che:

1 - **la devianza totale** è determinata dagli scarti al quadrato di ognuna delle 106 misure di peso dei figli dalla media generale ha 105 df;

2 - **la devianza tra le 23 nidiate** è data dagli scarti della media di ognuna di esse dalla media generale; ha **22 df** e serve per verificare se esistono differenze tra esse; essa può essere scomposta nelle 3 devianze successive (2a, 2b, 2c), come è possibile verificare rapidamente con i df relativi;

2a) - **la devianza tra i 3 tipi di alimentazione** (fattore A o del I livello ed indicati con I, II, III) è data dagli scarti di ognuna delle sue medie dalla media generale ed ha **2 df**;

2b) - **la devianza tra i 9 maschi** (fattore B o del II livello ed indicati con numeri da 1 a 9) è calcolata entro ogni tipo di alimentazione, come scarto della media dei figli dello stesso maschio rispetto alla media del tipo di alimentazione:

il tipo di alimentazione I ha 2 maschi e quindi 1 df,

il tipo di alimentazione II ha 4 maschi e quindi 3 df,

il tipo di alimentazione III ha 3 maschi e quindi 2 df,

per un totale di **6 df** come risulta dalla tabella generale;

2c) - **la devianza tra le 23 femmine** (fattore C o del III livello ed indicate con le lettere da A a Z) è calcolata entro ognuno dei 9 maschi e quindi **i df risultano 14** (23 - 9); è possibile arrivare alla stessa stima sommando i gradi di libertà dei gruppi di femmine entro lo stesso maschio:

$$1 + 2 + 2 + 1 + 1 + 3 + 1 + 1 + 2 = 14$$

3 - **la devianza d'errore**, calcolata per differenza (devianza totale - devianza tra nidiate) o sulla base degli scarti di ogni osservazione dalla media della nidiate (corrispondente alla media della femmina).

**La varianza d'errore**, che è al denominatore nel rapporto F, è **la varianza entro nei sottogruppi di livello inferiore**: nell'esempio sull'inquinamento delle città, essa è determinata dalla differenza delle singole osservazioni dalla media dell'area entro la città; in quello sull'effetto dell'alimentazione e del genotipo sia paterno che materno, è lo scarto di peso di ogni figlio rispetto alla media riferita alla madre.

**Nell'analisi gerarchica è possibile calcolare due o più tipi di F**, secondo il numero di fattori considerati; in questi rapporti la varianza al denominatore è quella calcolata entro il livello inferiore.

Quando il test F per il confronto tra medie risulta significativo, è possibile **mettere in evidenza a quali differenze tra coppie** di medie è imputabile (vedi output alla fine del capitolo).

Per confronti tra le singole medie, il test più appropriato è quello di Tukey.

Con  $\alpha = 0.05$ ,  $df = 83$  (della varianza d'errore) e varianza d'errore uguale a 6,314 fornisce i seguenti valori della differenza minima significativa:

- per le medie dei 3 tipi di alimentazione, il valore critico studentizzato risulta 3,375;
- per le medie dei 9 maschi risultano significative le differenze che superano il valore critico 4,504;
- per le medie delle nidiate delle 23 femmine risultano significative le differenze che superano il valore critico 5,288.

La differenza tra i tre valori dipende dal diverso numero di confronti che è possibile effettuare nei 3 gruppi e dal numero di dati sui quali ognuna di esse è calcolata.

## 10.10. ASSUNZIONI DI VALIDITA' E TRASFORMAZIONI DEI DATI

I test di statistica parametrica già presentati (test t e test F in tutte le loro forme) oltre a quelli che verranno di seguito presentati (regressione, correlazione e covarianza) per essere validamente applicati richiedono che i dati rispettino **condizioni di carattere formale e di carattere sostanziale**. **Senza la dimostrazione che i dati sono in accordo con queste condizioni, qualunque conclusione raggiunta con un test parametrico può essere posta in discussione ed i risultati ottenuti non essere accettati.**

**Le condizioni di carattere formale** riguardano la completezza dei dati nei disegni sperimentali rigidi, come i blocchi randomizzati ed i quadrati latini:

- 1) **se mancano uno o più dati** occorre procedere alla loro integrazione.

Altre condizioni formali riguardano

- 2) la **presenza di dati uguali a zero o indefiniti**, che determinano l'impossibilità di calcolare la media (ad es.: almeno un dato uguale a 0, in caso del ricorso a logaritmi; almeno un dato uguale a infinito, quando si misurano tempi di risposta ad uno stimolo e la cavia non reagisce) oppure
- 3) una **diversa attendibilità dei risultati**, in quanto ottenuti su gruppi con un numero di osservazioni notevolmente diverso (ad es.: gruppo di percentuali o rapporti, calcolati contemporaneamente su campioni di poche unità e di alcune centinaia).

**Le condizioni di carattere sostanziale** sono:

- 1) gli **effetti del trattamento** devono essere **additivi**,
- 2) gli **errori e le osservazioni** devono essere **indipendenti**,
- 3) la **distribuzione** degli errori e quella delle osservazioni devono essere **normali**,
- 4) la **varianza** deve essere **omogenea**.

Gli **effetti di due o più trattamenti possono combinarsi tra loro per addizione o per moltiplicazione**. Si hanno effetti additivi quando un trattamento differisce da un altro di una quantità che non varia passando da un individuo all'altro; in caso contrario, si parla di effetti moltiplicativi. L'argomento è stato presentato nell'analisi dell'interazione, con la rappresentazione grafica e le differenze tra medie osservate e medie attese. In un trattamento a blocchi randomizzati senza repliche, occorre verificare se gli effetti sono indipendenti dalla media dei blocchi o viceversa. Quando si hanno effetti moltiplicativi, si può ritornare al modello additivo mediante la trasformazione logaritmica dei dati, utilizzando le proprietà matematiche dei logaritmi.

L'**indipendenza delle osservazioni** è realizzata quando una rilevazione non è influenzata da quella precedente o vicina. La dipendenza risulta più spesso da una correlazione nel tempo che nello spazio, per un trascinarsi dell'informazione; può succedere quando lo strumento di misura viene alterato o semplicemente influenzato dall'osservazione precedente oppure un individuo può essere più simile a quelli vicini e i campioni non sono stati raccolti in modo corretto.

Si ha **indipendenza degli errori** quando i termini che definiscono la varianza d'errore sono distribuiti in modo casuale. Sorgono forti sospetti sulla correttezza del campionamento, quando si evidenziano lunghe successioni di scarti positivi e di scarti negativi; oppure, scarti positivi e negativi tendono ad alternarsi con regolarità. Le modalità delle analisi sono già state presentate nel test delle successioni.

**La probabilità che una osservazione presenti un certo errore non deve dipendere né dal segno né dalla grandezza, ma essere assolutamente casuale.**

Per ottenere l'indipendenza delle osservazioni e degli errori, è necessario che nella sperimentazione il ricercatore si attenga ai seguenti principi:

- la randomizzazione deve essere fondata su elementi obiettivi, come l'estrazione di numeri casuali da una tabella, da un sacchetto, dal calcolatore; non deve mai essere lasciata all'arbitrio di un individuo (**effetto random**);
- ogni dato deve avere le stesse possibilità di essere influenzato da eguali circostanze di tempo e di luogo (**effetto ambiente**);
- tutti gli individui del campione devono avere le stesse possibilità di essere sottoposti a un trattamento qualunque (**effetto trattamento**).

I test parametrici sono validi se **la distribuzione dei dati è normale e quella degli errori è normale intorno alla media**. La verifica avviene con il controllo della simmetria e della curtosi. Le conseguenze della **non normalità degli errori** spesso non sono gravi. Solamente una fortissima asimmetria ha effetti sul livello di significatività del **test F** e del **test t** che **sono robusti rispetto a questa condizione**; la correlazione e la regressione ne risentono maggiormente.

**L'omogeneità delle varianze o omoschedasticità** (omoschedalità in altri testi) viene verificata mediante i test già illustrati. Quando essa non è realizzata, si ha una perdita d'informazione sull'efficacia dei trattamenti, poiché il confronto tra le medie e la stima degli effetti aggiunti sono fondati sull'assunto che tutti i gruppi abbiano la stessa **varianza naturale o varianza vera** ( $s^2$ ). Inoltre, se le varianze non sono omogenee si determina una variazione del peso relativo dei gruppi sul valore della varianza d'errore.

Quando si rifiuta l'ipotesi di omoschedasticità, si può classificare l'**eteroscedasticità** come **regolare o irregolare**.

La eteroscedasticità è detta **irregolare** quando non si evidenzia alcun rapporto tra media e varianza; può derivare da cause aberranti, come la presenza di un dato anomalo, o da una non corretta impostazione dell'esperimento. In questi casi si deve verificare se si tratta di sbagli commessi dallo sperimentatore (come nella trascrizione dei dati) o di variazioni reali. Nel primo caso si dovrebbe ripetere l'esperimento, se non è possibile individuare la causa e apportare la correzione. Nel secondo caso si può procedere alla trasformazione dei dati, con uno dei metodi che verranno di seguito presentati.

La **eteroschedasticità è regolare** quando esiste una relazione di tipo noto, come nella distribuzione poissoniana, o comunque evidenziabile con i metodi della statistica descrittiva. In questo caso si opera la trasformazione dei dati, che per la poissoniana è la trasformazione in radice quadrata.

**Le trasformazioni dei dati si propongono 3 scopi principali:**

- 1 - **stabilizzare le varianze,**
- 2 - **linealizzare le relazioni tra variabili,**
- 3 - **normalizzare le distribuzioni,**

**e 2 scopi secondari:**

- 1 - **semplificare l'elaborazione di dati** che presentano caratteristiche non gradite,

## 2 - rappresentare i dati in una scala ritenuta più adatta.

Le trasformazioni alle quali più frequentemente si ricorre sono:

- la lineare,
- la logaritmica,
- le potenze, che comprendono le radici e soprattutto la radice quadrata e cubica, la trasformazione
- reciproca, la quadratica,
- le angolari,
- i probit, i logit, i normit.

**La trasformazione lineare** consiste nel **cambiamento della scala** o dell'origine delle misure, per facilitare la loro comprensione rispetto alla misura precedente. E' il caso della trasformazione della temperatura da gradi Celsius a Fahrenheit, o viceversa. Altre trasformazioni lineari sono ottenute sottraendo un valore costante a tutti i dati, come diminuire di 120 i valori di un campione che varia da 120 a 140; in questi casi, servono soprattutto per semplificare i calcoli.

La trasformazione lineare risulta inutile quando si intende modificare le caratteristiche fondamentali della distribuzione. A questo scopo assumono maggiore importanza le trasformazioni non lineari, in cui a distanze uguali nella prima scala non corrispondono distanze uguali nella seconda.

Tutte le trasformazioni di seguito riportate, le più ricorrenti nelle applicazioni della statistica, non sono lineari.

**La trasformazione in ranghi**, poco citata nei testi, è una tecnica molto semplice e sempre più frequentemente raccomandata da autori importanti. Con essa, quando i dati sono abbastanza numerosi, si ricostruiscono le condizioni di validità e si possono applicare tutti i test parametrici. E' il caso soprattutto di disegni sperimentali complessi, a tre o più fattori con eventuale interazione od analisi gerarchica, per i quali non esistono alternative alla statistica parametrica. Questo accorgimento permette anche di superare **uno dei limiti fondamentali della statistica non parametrica, che offre test di significatività ma che raramente e con difficoltà è adattabile ai problemi, non meno importanti, di stima accurata dei parametri.**

### **La trasformazione logaritmica**

$$Y = \log_a X$$

di solito avviene con base 10 o con base naturale (e), anche se non sono infrequenti quelli con base 2. Serve per ottenere relazioni lineari, quando si analizzano curve di crescita esponenziali; in variabili continue, è utile per stabilizzare le varianze quando crescono in modo regolare all'aumentare della media; nel caso di effetti moltiplicativi, quando si vuole ritornare ad effetti additivi.

Abitualmente viene applicata al conteggio di popolazioni che hanno un incremento esponenziale in funzione del tempo. La tabella sottostante mostra come dati che possono variare da 2 a 20000 riducano il loro campo di variazione da 0,30 a 4,30 con logaritmi a base 10.

X	2	20	200	2000	20000
Y	0,30	1,30	2,30	3,30	4,30

Quando i coefficienti di variazione di gruppi a confronto sono approssimativamente costanti, le varianze aumentano in modo direttamente proporzionale alle medie; di conseguenza, confrontando i due gruppi A e B, tra i loro dati esiste la relazione

$$X_{Ai} = cX_{Bi}$$

dove c è la costante della proporzione.

La trasformazione dei dati con i logaritmi

$$\log X_{Ai} = \log c + \log X_{Bi}$$

rende le varianze omogenee, poiché i dati avranno una media differente ma la stessa forma di distribuzione.

**La trasformazione logaritmica può essere applicata solamente a valori positivi**, in quanto non esistono i logaritmi di valori negativi. Quando si hanno valori nulli, poiché  $\log 0 = -\infty$  (meno infinito), la trasformazione avviene con l'accorgimento di **aggiungere una costante** ( $c = 1$  oppure  $c = 0,5$ ) a tutti i dati (non solo a quelli nulli)

$$Y = \log(X + c)$$

In varie situazioni, la trasformazione logaritmica ha effetti multipli e serve contemporaneamente a stabilizzare la varianza, per ridurre ad effetti additivi un effetto moltiplicativo, a normalizzare la distribuzione.

La **trasformazione in radice quadrata**

$$Y = \sqrt{X}$$

è uno dei casi più frequenti di **trasformazioni mediante potenze**, in cui  $c = 1/2$ .

$$Y = X^c$$

E' utile in particolare per omogeneizzare le varianze; è applicata a conteggi, quindi a **valori sempre positivi o nulli**, che seguono la distribuzione poissoniana.

In batteriologia, ematologia, fitosociologia è il caso di conteggi microbiologi oppure di animali o piante dispersi su una superficie. Poiché la varianza ( $npq$ ) è proporzionale alla media ( $np$ ), la condizione di omoschedasticità è violata in partenza.

Quasi sempre i dati sono rappresentati da piccoli numeri, poiché all'aumentare della media la distribuzione poissoniana tende alla normale.

Quando si ha la presenza di almeno uno zero è consigliabile (per tutti i dati) la trasformazione

$$Y = \sqrt{X + 0,5}$$

che risulta appropriata per valori piccoli, con medie inferiori a 1, in cui la semplice trasformazione in radice quadrata determinerebbe un ampliamento delle distanze tra i valori minori.

#### La **trasformazione in radice cubica**

$$Y = \sqrt[3]{X}$$

viene utilizzata per popolazioni che vivono in uno spazio tridimensionale. Come in ecologia per la distribuzione di animali sul terreno si usa la radice quadrata, in idrobiologia per conteggi di plancton che non risentano della crescita esponenziale di tali popolazioni si ricorre abitualmente alla trasformazione in radice cubica.

Nell'analisi di popolazioni che vivono in una spazio tridimensionale, si usa la trasformazione logaritmica quando ha la prevalenza la differenza stagionale, in specie che hanno esplosioni demografiche, per cui si possono avere campioni con poche unità ed altri con varie migliaia di individui; si usa quella in radice cubica se i dati presentano differenze minori e la distribuzione è asimmetrica.

#### La **trasformazione reciproca**

$$Y = \frac{1}{X}$$

è particolarmente utile nell'analisi di tempi, come per confronti sulla sopravvivenza dopo somministrazione di un tossico ad elevata letalità o di reazione a stimoli. Di norma, la maggior parte delle reazioni cadono in un intervallo relativamente ristretto e con distribuzione approssimativamente normale; ma esistono anche individui che hanno tempi di reazione molto alti, con valori che alterano profondamente la distribuzione con una simmetria a destra.

Quando uno o più individui non manifestano reazioni, il tempo diventa infinito: è impossibile fare la somma, calcolare la media e tutti le altre misure da essa derivate. La trasformazione reciproca, che attribuisce alla variabile  $Y = \infty$  il valore zero, permette la stima di tutti parametri. Con essa, valori elevati di  $X$  corrispondono a valori di  $Y$  prossimi allo zero ed aumenti molto elevati in  $X$  producono effetti trascurabili in  $Y$ .

Per l'interpretazione sui risultati conviene ritornare alla scala originale, come per la media armonica.

La trasformazione reciproca serve per stabilizzare la varianza, quando essa aumenta in modo molto pronunciato rispetto alla media.

#### La **trasformazione quadratica**

$$Y = X^2$$

è utile in situazioni opposte a quelle fino ad ora presentate, cioè quando la varianza tende a decrescere all'aumentare di  $X$  e la distribuzione dei dati ha una forte asimmetria negativa.

Sono fenomeni rari nella ricerca ambientale e pertanto il suo uso è poco diffuso.

La **trasformazione angolare** o in **arcoseno**

$$Y = \arcsen \sqrt{\frac{P}{100}}$$

o **trasformazione seno inverso**

$$Y = \text{sen}^{-1} \sqrt{\frac{X}{n}}$$

è utilizzata nel caso di percentuali (p) o frazioni (x), che seguono una distribuzione di tipo binomiale e quindi hanno un valore della varianza (pq) determinato da quella della media (p).

Per l'uso dei test parametrici, percentuali e frazioni presentano alcuni problemi, che richiedono analisi preliminari, poiché sono utilizzati per rendere le osservazioni indipendenti dalle dimensioni del campione. Per esempio, tra un primo esperimento che abbia fornito 3 risposte positive su 4 tentativi, un secondo che ne abbia dato 81 su 100 tentativi ed un terzo con 248 su 300 si può effettuare il confronto ricorrendo ai loro rapporti (rispettivamente 0,75 per il primo; 0,80 per il secondo; 0,82667 per il terzo) oppure mediante percentuale (75%, 80% e 82,667%). Ma ognuno di questi dati ha una "attendibilità" diversa ed un intervallo fiduciale differente; di conseguenza, non possono essere elaborati insieme.

**Quando si dispone di percentuali e rapporti, occorre preliminarmente verificare su quali dimensioni del campione sono stati calcolati. L'analisi con test parametrici è accettabile solamente se le dimensioni sono relativamente simili:** non è possibile elaborare insieme percentuali stimate su poche unità con altre stimate su un centinaio di individui od oltre. In statistica 3/4 non è uguale a 15/20, se con il primo si intendono 3 risposte positive su 4 individui ed con il secondo 15 risposte positive su 20.

**Con questi dati è possibile solamente un test non parametrico, poiché la reale informazione posseduta da una serie di tali valori è quella di rango.**

Una volta che sia stato chiarito questo aspetto, occorre passare alla trasformazione angolare. Una percentuale con media **p** ha una varianza uguale a **p(1-p)**: ha valori massimi per **p** prossimo a **0,5** o **50%** e ha valori progressivamente minori per **p** che tende a **0** oppure a **1** (**0%** oppure **100%**). La trasformazione angolare ha la caratteristica opposta: determina variazioni maggiori agli estremi che al centro della scala, riconducendo i rapporti tra le varianze a valori di omoschedasticità.

La tabella sottostante, che riporta alcuni valori, permette un confronto tra le due serie.

percentuale(p)	$\arcsen \sqrt{\frac{p}{100}}$
1	5,7
2	8,1
3	10,0
5	10,9
10	18,4
20	26,6
30	33,2
40	39,2
50	45,0
60	50,8
70	56,8
80	63,4
90	71,6
95	77,1
97	80,0
98	81,9
99	84,3

La **trasformazione seno inverso iperbolico**

$$Y = \sqrt{k} \operatorname{sen} h^{-1} \sqrt{\frac{X}{k}}$$

occupa una posizione intermedia tra la trasformazione logaritmica da applicare in variabili poissoniane altamente disturbate e la trasformazione in radice quadrata per variabili poissoniane standard.

La **trasformazione tangente iperbolica inversa**

$$Y = 1/2 \log_e \frac{1+r}{1-r} = \tan h_r^{-1}$$

è analoga alla trasformazione logaritmica ed è applicata a variabili che variano da -1 a +1; è utile per normalizzare la distribuzione dei coefficienti di correlazione (r) che, come vedremo, sono distribuiti normalmente solo per valori prossimi allo zero, mentre sono altamente asimmetrici per valori prossimi a +1 e -1.

La **trasformazione log-log**

$$Y = \log_e (-\log_e p)$$

e la **trasformazione log-log complementare**

$$Y = \log_e (-\log_e (1-p))$$

si applicano a percentuali di sopravvivenza, nello studio dei tempi di eliminazione di un gruppo di cavie in dosaggi biologici.

La **trasformazione probit**

$$P = \frac{1}{\sqrt{2\pi s}} \int_{-\infty}^{z-5} \exp\left(-\frac{(X-m)^2}{2s^2}\right) dx$$

(**probits** da *probability units*) è definita come la devinata normale equivalente, aumentata di 5.

Nello studio della relazione dose-risposta, la percentuale di individui che rispondono all'effetto causato dalla dose viene di solito rappresentato con una curva cumulata. Essa ha forma sigmoide, se la curva della distribuzione originaria è normale, con la conseguenza che a parità di errore nella dose l'errore nella risposta non è costante, ma varia secondo il punto in cui incontra perpendicolarmente la sigmoide. Per un errore costante nella risposta, occorre trasformarla in una retta.

La curva percentuale cumulata può essere linearizzata in vari modi; uno consiste nel sostituire ai valori di p dell'ordinata quelli corrispondenti all'ascissa della distribuzione normale standardizzata

$$Y' = \frac{x - \mu}{\sigma}$$

A causa della simmetria della distribuzione normale, il 50% dei valori Y' è negativo e l'altro 50% è positivo. E' noto e può anche essere osservato sulla tabella della distribuzione normale che meno di 2 valori su 10.000 hanno un valore di Y' inferiore a -3,5. Aggiungendo quindi a tutti i valori trasformati in Y' la quantità 5, si eliminano tutti i valori negativi. Questi valori trasformati mediante la relazione

$$Y = 5 + Y' = 5 + \frac{X - \mu}{\sigma}$$

sono i **probits**.

Nei suoi effetti, questa trasformazione è analoga a quella angolare, in quanto i valori verso gli estremi della distribuzione sono più dilatati di quelli collocati nella parte centrale. Il campo di variazione della scala probit tende all'infinito e la scala dei probit si distingue da quella angolare soprattutto nei valori prossimi a 0 e a 1.

La trasformazione in probits, rendendo lineare la sigmoide di una cumulata tratta dalla distribuzione normale, permette di trattare la stima dei parametri della distribuzione normale ( $\mu$  e  $\sigma$ ) come quello dei parametri di una regressione lineare (intercetta  $\alpha$  e coefficiente angolare  $\beta$ ). Ma la stima corretta dei parametri della retta richiede che i punti sperimentali abbiano la stessa varianza; di conseguenza i valori dei probits dovrebbero essere ponderati.

L'effetto linearizzante della trasformazione probit è stato ampiamente utilizzato nelle **carte di probabilità**, usate per verificare in modo semplice e con un grafico se una distribuzione era normale. La diffusione dei calcolatori, che permettono stime rapide ed esatte dei valori di asimmetria e curtosi di una serie di dati campionari, ha reso superflui questi metodi grafici.

La **trasformazione normit**

$$P = \frac{1}{\sqrt{2p}} \int_{-\infty}^z \exp\left(-\frac{u^2}{2}\right) du$$

è un'altra trasformazione di percentuali cumulative basate sull'integrale di probabilità della curva normale. Fornisce valori diversi dai probits.

La **trasformazione logit** viene anche essa applicata a osservazioni percentuali ed è ottenuta con

$$Y = \log_e \frac{p}{1-p}$$

L'effetto di questa trasformazione **logistica** o logit è simile a quella probit e può determinare analisi del tutto uguali, in particolare nello studio del dosaggio con risposte quantali.

L'attuale diffusione dell'informatica, che ha superato le difficoltà derivanti dalla complessità dei calcoli e dal tempo richiesto nei calcoli manuali, ha annullato la necessità di linearizzare le distribuzioni. Di conseguenza, le trasformazioni probit e logit sono sempre meno usate.

### **10.11 LA SCELTA DELLA TRASFORMAZIONE IDONEA: IL METODO DI BOX-COX**

La scelta del tipo di trasformazione, da applicare ad una serie di dati campionari per rispettare le condizioni di validità dei test parametrici, spesso è conosciuta a priori, sulla base di quanto noto sulle caratteristiche del fenomeno. Le trasformazioni possibili e quelle suggerite nelle diverse situazioni, riportate nel paragrafo precedente, sono derivate da queste esperienze. Ma quando si analizza un fattore nuovo o collocato in un ambiente diverso, si possono ottenere **forme di distribuzione dei dati che non sono identificabili con facilità; di conseguenza, è difficile individuare la trasformazione più appropriata**, per ricostruire le condizioni di validità statistica richieste dai test inferenziali parametrici.

Il problema si pone soprattutto quando un fenomeno può essere misurato in modi diversi. Per esempio, le diverse velocità di un gruppo di soggetti possono essere valutate sia misurando il tempo (in minuti) impiegato a percorrere mille metri, sia dividendo la distanza per i minuti necessari ad ognuno.

**Le due serie di dati non hanno la stessa forma di distribuzione e quindi l'analisi statistica potrebbe condurre ad inferenze differenti.**

Per scegliere quella più adeguata esistono due criteri:

- il primo dipende dalla conoscenza scientifica dell'argomento: **è la misura che meglio valuta il fenomeno e lo rende più comprensibile;**

- il secondo è di tipo tecnico-statistico: **è la misura che ha una forma di distribuzione dei dati più rispettosa delle condizioni di validità del test.** Ma come effettuare questa scelta?

**Per la ricerca applicata, è uno dei problemi più importanti e che ricorrono con frequenza maggiore:** da esso dipendono sia la potenza del test, sia la validità stessa delle inferenze.

Nel 1964, **G. E. P. Box** e **D. R. Cox**, con l'articolo "*An analysis of transformation (with Discussion)*" (pubblicato su *J. R. Stat. Soc., Series B* 26, pp. 211-252) hanno proposto un metodo iterativo, lungo e complesso anche se relativamente semplice dal punto di vista concettuale, soggetto in seguito a numerose elaborazioni. Trascurato nei decenni scorsi, tale metodo ora è divenuto relativamente semplice e di vasta applicazione, con la diffusione dei computer; in particolare da quando è inserito nei programmi informatici più sofisticati è quindi possibile applicare direttamente ad ogni distribuzione di dati. Una distribuzione dei dati di forma qualsiasi, con asimmetria positiva o negativa, platicurtica o leptocurtica, quasi sempre può essere modificata in una curva "**pressoché normale**" con la trasformazione di Box-Cox.

Applicando l'analisi della varianza, la **trasformazione migliore è quella che rende minima la devianza d'errore rispetto alla devianza spiegata, poiché l'allontanamento dalla normalità determina appunto anche una crescita dei residui.**

Per ogni insieme di  $X$  osservati, con la relazione

$$X^l = \frac{X^l - 1}{l\bar{X}^{l-1}}$$

dove  $\bar{X}$  è la media geometrica di tutti i dati,

si stima un valore di  $l$  che indica la trasformazione più adeguata, sulla base del concetto che

**il valore ricercato di  $l$  (the maximum likelihood value) è quello che rende il test più significativo e quindi massimo il valore di  $F$ , per i fattori prescelti.**

Una proposta più recente utilizza una formula che richiede il calcolo della funzione  **$L$**  (nota come **log-likelihood function**), con

$$L = -\frac{n}{2} \ln s_{TRAS}^2 + (l-1) \frac{n}{N} \sum \ln X$$

in cui

$s^2_{\text{TRAS}}$  è la varianza del campione (dopo trasformazione della X) nel caso di un solo gruppo,  
o della varianza d'errore nel caso di più gruppi,

$v$  sono i gdl della varianza precedente (del gruppo o quella d'errore in caso di più gruppi),

$N$  è il numero totale di dati,

mentre il secondo termine dell'equazione è la somma dei **log** naturali (**ln**) delle X originali.

**Con essa si stima il valore di l che rende massimo il valore di L.**

Inoltre è **possibile calcolare l'intervallo fiduciale di l**, entro il quale è conveniente scegliere la trasformazione più adeguata. Questa tecnica permette di prestare molta attenzione all'interpretazione ecologica od ambientale del fenomeno e ai suoi aspetti teorici; è un elemento rilevante nella ricerca applicata, poiché è troppo riduttivo prendere in considerazione solo gli aspetti tecnici, che permettono di normalizzare una serie di dati sperimentali. La stima di l non avviene direttamente, ma attraverso i limiti fiduciali di  $S_l$ , derivato dalla somma dei residui.

Nel testo di George E. P. **Box**, William G. **Hunter** e J. Stuart **Hunter** "*Statistics for Experimenters. An introduction to Design, Data Analysis and Model Building*", pubblicato da John Wiley & Sons, New York, 1978, p. 653 (a pag. 239), per la probabilità  $\alpha$  si propone di stimare S

$$S = S_l \left( 1 + \frac{t_{n,\alpha/2}^2}{n} \right)$$

che definisce il limite massimo della deviazione standard minima.

Secondo questi autori, nella successiva analisi della varianza applicata ai dati trasformati, **la devianza e la varianza d'errore perderebbero 1 gdl**, appunto perché vincolate alla condizione di essere le minori possibili nei confronti del fattore considerato.

**Il valore di l individuato corrisponde all'esponente a cui elevare la variabile da trasformare**

$$(X' = X^l).$$

Con un elenco sufficientemente dettagliato, si ricorda che a:

$l = 3$ , corrisponde una trasformazione con elevamento al cubo  $x^3$  (poiché la curva ha un  $g_1$  negativo);

$l = 2$ , indica una trasformazione con elevamento al quadrato  $x^2$  (curva con  $g_1$  meno negativo);

$l = 1$ , una trasformazione lineare che non modifica la curva della distribuzione  $x$ , (curva già normale);

$l = 1/2$ , una trasformazione con radice quadrata  $\sqrt{x}$  ( $g_1$  leggermente positivo);

$l = 1/3$ , una trasformazione con radice cubica  $\sqrt[3]{x}$  ( $g_1$  positivo);

$l = 0$ , una trasformazione logaritmica  $\ln x$  ( $g_1$  fortemente positivo), con la scelta ulteriore della base;

$l = -1/3$ , una trasformazione reciproca, con  $x$  sotto radice cubica  $\frac{1}{\sqrt[3]{x}}$  ( $g_1$  positivo);

$l = -1/2$ , una trasformazione reciproca, con  $x$  sotto radice quadrata  $\frac{1}{\sqrt{x}}$  ( $g_1$  positivo);

$l = -1$ , una trasformazione reciproca di  $x$ , con  $\frac{1}{x}$  ( $g_1$  positivo);

$l = -2$ , una trasformazione reciproca con  $x$  al quadrato  $\frac{1}{x^2}$  ( $g_1$  positivo);

$l = -3$ , una trasformazione reciproca con  $x$  al cubo  $\frac{1}{x^3}$  ( $g_1$  positivo).

E' importante sottolineare che il caso  $l = 0$  corrisponde alla trasformazione **log**

$$X^0 = \bar{X} \log X$$

dove la media  $\bar{X}$  è quella geometrica.

(La media geometrica di  $n$  valori di  $x$  è ottenuta facendo la media aritmetica dei **log x**, con successiva trasformazione del risultato in antilog).

Una caratteristica molto importante della trasformazione Box-Cox è che **la nuova media dei valori trasformati corrisponde alla mediana dei valori originali**. I dati originali avevano una distribuzione asimmetrica, tanto da richiedere una trasformazione: la loro nuova distribuzione è simmetrica e centrata sulla stima più corretta della precedente tendenza centrale, che è appunto la sua mediana.

La complessità dei problemi da risolvere per scegliere la trasformazione più adeguata e il dibattito che sempre si pone sulla reale validità dell'analisi attuata possono essere meglio illustrati con la discussione ampia di un esempio, tratto dalla letteratura (dati e grafici sono tratti dal testo di George E. P. **Box**, William G. **Hunter** e J. Stuart **Hunter** "*Statistics for Experimenters. An introduction to Design, Data Analysis and Model Building*", pp. 228-240).

ESEMPIO. Per verificare gli effetti di 3 sostanze tossiche (A, B, C) sulla sopravvivenza di cavie di età diversa (I, II, III, IV), ad ognuno dei 12 gruppi (3 trattamenti x 4 blocchi) sono state assegnati 4 individui. Per ognuno di essi è stato misurato il tempo di sopravvivenza, tradotto in una grandezza unitaria equivalente a 10 ore.

I risultati sono riportati nella tabella sottostante

Età BLOCCHI	Sostanze Tossiche TRATTAMENTI											
	A				B				C			
I	0,31	0,45	0,46	0,43	0,36	0,29	0,40	0,23	0,22	0,21	0,18	0,23
II	0,8	1,10	0,88	0,72	0,92	0,61	0,49	1,24	0,30	0,37	0,38	0,29
III	0,43	0,45	0,63	0,76	0,44	0,35	0,31	0,40	0,23	0,25	0,24	0,22
IV	0,45	0,71	0,66	0,62	0,56	1,02	0,71	0,38	0,30	0,36	0,31	0,33

(in cui il valore 0,31 della prima cavia, appartenente alla classe d'età I e alla quale è stato somministrato il tossico A, indica che essa è sopravvissuta 3,1 giorni).

E' un disegno fattoriale a due fattori con repliche (3 trattamenti x 4 blocchi con 4 repliche per ogni esperimento e quindi con 48 dati), che permette di verificare l'eventuale significatività sia di ognuno dei due fattori, sia della loro interazione.

L'analisi della varianza (ovviamente ottenuta con un programma informatico) fornisce i seguenti risultati

Fonte di variazione	Devianza	Gdl	Varianza	F	P
<b>Totale</b>	<b>3,005</b>	<b>47</b>	---	---	---
<b>Tra gruppi</b>	<b>2,204</b>	<b>11</b>	<b>0,200</b>	<b>9,010</b>	<b>.000</b>
Tra tossici	1,033	2	0,517	23,222	.000
Tra età	0,921	3	0,307	13,806	.000
Interazione	0,250	6	0,0417	1,874	.112
<b>Entro gruppi (errore)</b>	<b>0,801</b>	<b>36</b>	0,0222	---	---

che permettono di rifiutare l'ipotesi nulla, relativamente al confronto tra tossici e tra età; per l'interazione si può sostenere una significatività tendenziale ( $P = .112$ ), che potrebbe forse essere dimostrata con un aumento delle dimensioni del campione.

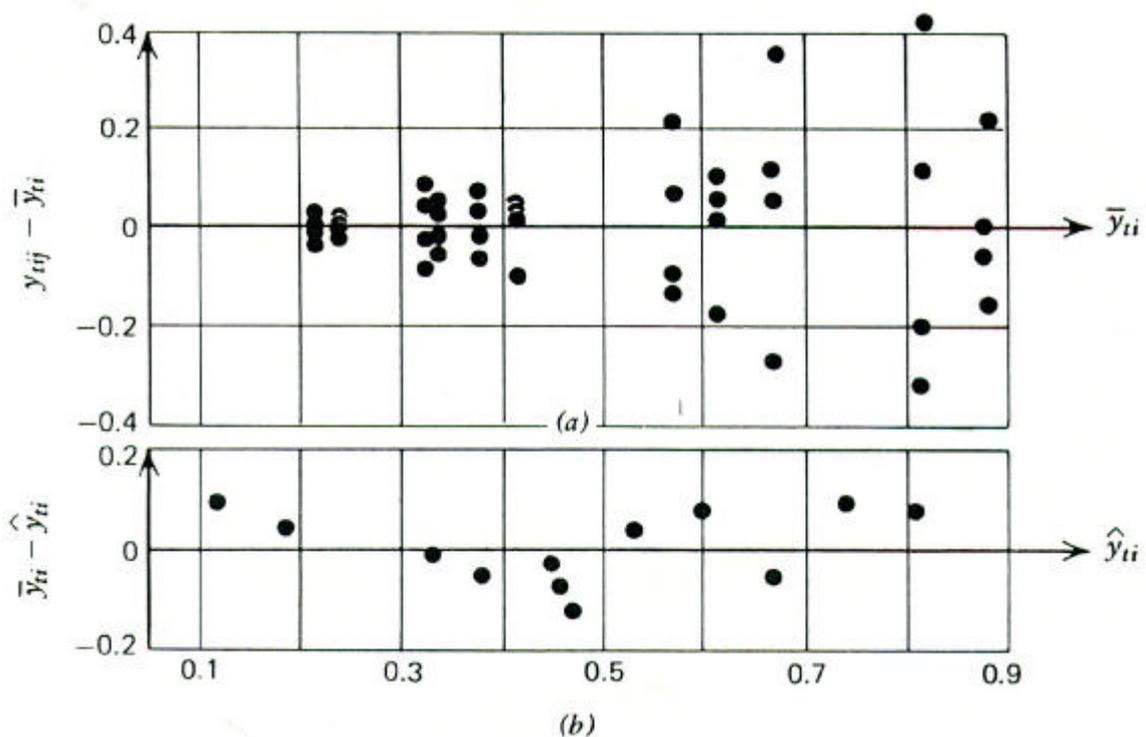
### **Ma l'analisi effettuata è valida?**

Per valutare se sono state rispettate le condizioni di validità, è utile analizzare i residui. Il modo più semplice è quello della loro rappresentazione grafica, che può riguardare sia la variabilità entro gruppi che l'interazione. A questo scopo è utile **costruire due grafici** (che i programmi informatici più sofisticati permettono di stampare con facilità):

- il primo (vedi grafico **a**) può essere ottenuto riportando sull'asse delle ascisse (trasferito al centro) la media di ogni gruppo (in questo caso quello di casella  $\bar{X}_{ij}$ ) e sull'asse delle ordinate gli scarti di ognuna delle **n** (4) repliche da essa ( $X_{ijk} - \bar{X}_{ij}$ );

- il secondo (vedi grafico **b**) è costruito riportando sull'asse delle ascisse le medie attese  $\hat{X}_{ij}$  in ogni gruppo (con  $\hat{X}_{ij} = \bar{X}_i + \bar{X}_j - \bar{\bar{X}}$ ) e sull'asse delle ordinate le differenze tra le medie osservate e queste medie attese ( $\bar{X}_{ij} - \hat{X}_{ij}$ ).

(Nelle figure successive, tratte dal testo citato, la variabile è indicata con Y)



Dall'analisi dei due grafici appare con evidenza che

- 1- gli scarti di ognuna delle 4 repliche dalla media del loro gruppo aumentano al crescere del valore della media;
- 2- gli scarti tra le medie osservate e quelle attese tendono ad una relazione di tipo curvilineo, all'aumentare del valore delle medie.

Per conclusioni condivise sul primo punto, occorrerebbe effettuare i confronti tra varianze. Ma le analisi inferenziali sulla omogeneità delle varianze (test di Hartley, Cochran, Bartlett) sono molto tolleranti: non rifiutare l'ipotesi nulla non significa che essa sia vera, in particolare quando i dati sono pochi.

Di conseguenza, è lecito il sospetto che l'analisi della varianza applicata in precedenza non sia valida, in quanto potrebbe non essere rispettata la condizione di omoschedasticità. Si impone quindi una trasformazione dei dati.

**Ma quale è la trasformazione più adeguata?** Il fatto che la varianza entro casella o errore cresca all'aumentare della media suggerisce di utilizzare una trasformazione per  $g_1$  positivo (forte asimmetria destra); ma esse sono tante, da quella in radice quadrata a quella logaritmica, oppure il reciproco.

Per meglio comprendere gli effetti delle trasformazioni, un primo tentativo può essere effettuato con la radice quadrata. I valori diventano quelli riportati nella tabella successiva

**Trasformazione in radice quadrata ( $\sqrt{X}$  arrotondata alla seconda cifra decimale)**

Età BLOCCHI	Sostanze Tossiche TRATTAMENTI											
	A				B				C			
I	0,56	0,67	0,68	0,66	0,60	0,54	0,63	0,48	0,47	0,46	0,42	0,48
II	0,91	1,05	0,94	0,85	0,96	0,78	0,70	1,11	0,55	0,61	0,62	0,54
III	0,66	0,67	0,79	0,87	0,66	0,59	0,56	0,63	0,48	0,50	0,49	0,47
IV	0,67	0,84	0,81	0,79	0,75	1,01	0,84	0,62	0,55	0,60	0,56	0,57

e l'analisi della varianza applicata ad essi fornisce i seguenti risultati

Fonte di variazione	Devianza	Gdl	Varianza	F	P
<b>Totale</b>	<b>1,365</b>	<b>47</b>	---	---	---
<b>Tra gruppi</b>	<b>1,071</b>	<b>11</b>	<b>0,0974</b>	<b>11,938</b>	<b>.000</b>
Tra tossici	0,561	2	0,280	34,389	.000
Tra età	0,431	3	0,144	17,601	.000
Interazione	0,079	6	0,013	1,62	.169
<b>Entro gruppi (errore)</b>	<b>0,294</b>	<b>36</b>	<b>0,00815</b>	---	---

Da essi emerge che:

1- il test F tra gruppi (F = 11,9 con 11 gdl), quello tra tossici che interessa maggiormente (F = 34,4 con 2 gdl) e quello tra età (F = 17,6 con 3 gdl) sono tutti più significativi di quanto risultassero in precedenza, con i dati originari;

2 – il test F per l'interazione ( $F = 1,62$  con 6 gdl) è meno significativo di quanto suggerito dall'analisi precedente.

**I risultati sono migliori; ma questa è la trasformazione più adeguata oppure ne esistono altre preferibili?**

E' semplice dimostrare che, con la trasformazione reciproca, i dati diventano

**Trasformazione in reciproco ( $\frac{1}{X}$  arrotondata alla seconda cifra decimale)**

Età BLOCCHI	Sostanze Tossiche TRATTAMENTI											
	A				B				C			
I	3,23	2,22	2,17	2,33	2,78	3,45	2,50	4,35	4,55	4,76	5,56	4,35
II	1,22	0,91	1,14	1,39	1,09	1,64	2,04	0,81	3,33	2,70	2,63	3,45
III	2,33	2,22	1,59	1,32	2,27	2,86	3,23	2,50	4,35	4,00	4,17	4,55
IV	2,22	1,41	1,52	1,61	1,79	0,98	1,41	2,63	3,33	2,78	3,23	3,03

e l'analisi della varianza fornisce risultati

Fonte di variazione	Devianza	Gdl	Varianza	F	P
<b>Totale</b>	<b>65,505</b>	<b>47</b>	---	---	---
<b>Tra gruppi</b>	<b>56,862</b>	<b>11</b>	<b>5,169</b>	<b>21,531</b>	<b>.000</b>
Tra tossici	34,877	2	17,439	72,635	.000
Tra età	20,414	3	6,805	28,343	.000
Interazione	1,571	6	0,262	1,090	.387
<b>Entro gruppi (errore)</b>	<b>8,643</b>	<b>36 (35)</b>	<b>0,240</b>	---	

ancor più significativi per i due fattori, ma che escludono la significatività, anche solo tendenziale, della loro interazione:

- il test F tra gruppi fornisce un valore pari a 31,531 (contro 11,938 precedente e 9,010 del primo caso);

- il test F tra tossici fornisce un valore pari a 72,635 (contro 34,389 precedente e 23,222 del primo caso);
- il test F tra età fornisce un valore pari a 28,343 (contro 17,601 precedente e 13,806 del primo caso);
- il test F dell'interazione fornisce un valore pari a 1,090 (contro 1,623 precedente e 1,874 del primo caso).

La figura successiva, che riporta

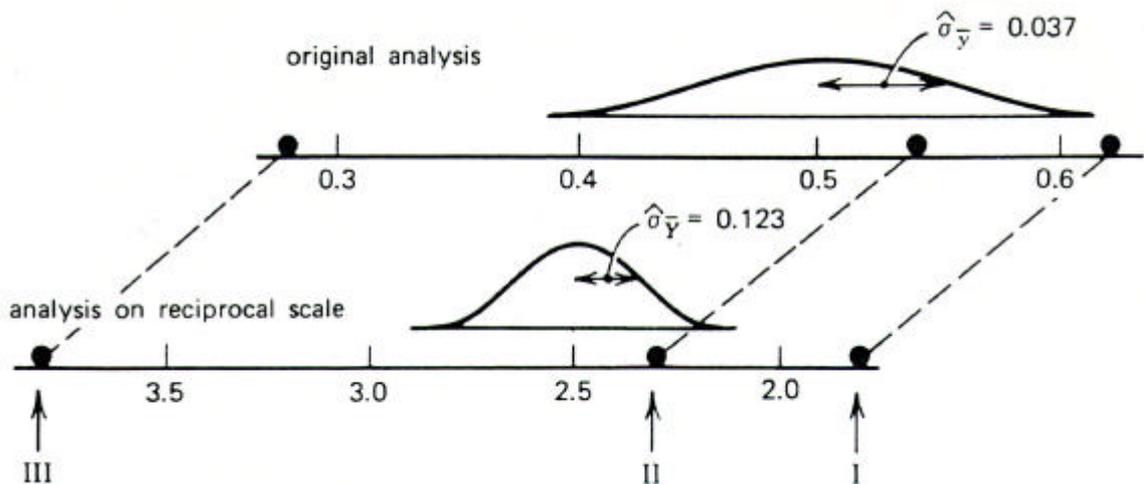
- nella parte superiore, i dati originari con la media dei tre tossici e la loro deviazione standard

$$s_{X_{originari}} = \sqrt{\frac{0,0222}{16}} = 0,037$$

- nella parte inferiore, i dati trasformati in reciproco con la media dei tre tossici e la loro deviazione standard

$$s_{X_{trasformati}} = \sqrt{\frac{0,240}{16}} = 0,123$$

rapportati alla stessa scala



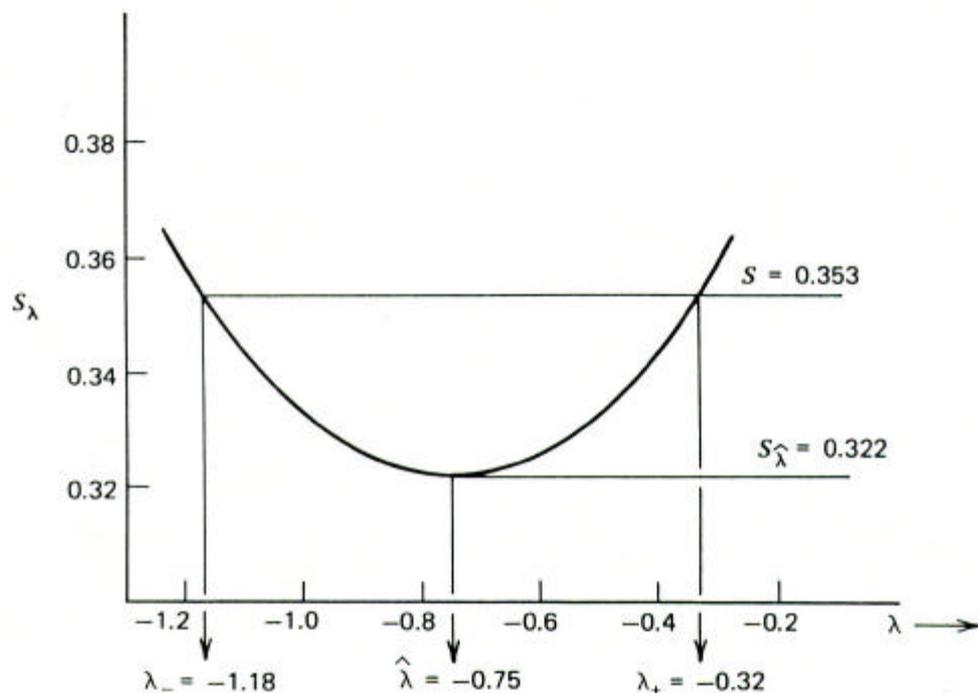
dimostra visivamente gli effetti della trasformazione sulla riduzione della varianza d'errore; nel caso specifico, sulla deviazione standard delle tre medie a confronto.

Per scegliere la trasformazione più adeguata all'esempio riportato, con una procedura sistematica che consideri tutte le possibilità migliori, **Box e al.** nel loro testo considerano solamente gli effetti di riga e di colonna, quindi una serie di valori  $S_1$ , derivati dalla somma dei quadrati dei residui con 42 gdl.

Nella tabella sottostante, per ogni valore di  $l$  è riportato il corrispondente valore di  $S_1$  da essi stimato:

$l$	-2,5	-2,0	-1,6	-1,4	-1,2	-1,0	-0,8	-0,6	-0,4	-0,2	0,0	0,5	1,0
$S_1$	1,333	0,664	0,463	0,401	0,359	0,333	0,323	0,326	0,343	0,375	0,424	0,635	1,051

Da essi è stato ricavato il grafico



che riporta il valore di  $l$  in ascissa e di  $S_\lambda$  in ordinata,

dal quale risulta che, con i criteri precedentemente definiti, la trasformazione più adeguata è

$$l = -0,75 \text{ corrispondente al valore minimo di } S_1 = 0,322$$

Questa risposta solleva 2 problemi:

- il valore di  $l = -0,75$  è una risposta campionaria e non è accettabile impostare la trasformazione solo su un risultato sperimentale, poiché sarebbe differente nei vari casi affrontati;
- una trasformazione con elevamento alla potenza  $-0,75$  è insolita e priva di significato specifico, mentre l'esperienza ha dimostrato che in questi casi (tempi di risposta ad uno stimolo) quella adeguata è la trasformazione reciproca.

La stima dell'intervallo fiduciale permette di giungere ad una risposta generale.

Per  $\alpha = 0.05$  e con  $n = 42$ , poiché  $t_{42, 0.025} = 2,021$

si ottiene un valore di S

$$S = 0,322 \cdot \left( 1 + \frac{2,021^2}{42} \right) = 0,353$$

pari a 0,353.

Di conseguenza, è accettabile un valore  $S_1$  fino al limite di 0,353. Sulla figura precedente, simmetrica rispetto al valore centrale, corrispondono valori di  $l$  che sono compresi tra  $-1,18$  e  $-0,32$ . Poiché  $l = -1$  è compreso in questo intervallo fiduciale, la trasformazione reciproca è adatta ai dati sperimentali raccolti, in pieno accordo con la teoria sulle misure di tempo.

In questo esempio, che descrive una realtà complessa ma frequente nella ricerca ambientale, con la trasformazione si è voluto risolvere contemporaneamente i problemi derivanti da più cause:

- la **non additività** dei due fattori considerati (per la presenza di una interazione tendenzialmente significativa),
- la **non omoschedasticità** dei gruppi a confronto,
- la **non normalità** della distribuzione dei dati.

Il metodo di Box-Cox permette di stimare il valore di  $l$  più adeguato a risolvere ognuno di questi problemi, sia separatamente che insieme.

Ad essi è da aggiungere il caso in cui i dati presentano una variabilità elevata, cioè quando il rapporto

$$X_{\text{massimo}} / X_{\text{minimo}}$$

è grande, indicativamente maggiore di tre.

## 10.12. LETTURA DI DUE TABULATI DI PROGRAMMI INFORMATICI

Sono riportati due tabulati di calcolatore su esempi già discussi nel capitolo.

### ESEMPIO con dati su località e stagione

**analisi della varianza a due criteri di classificazione con repliche,**

**con verifica dell'interazione.**

General Linear Models Procedure

1)

Class	Levels	Values
STAGIONE	3	autunno estate primaver
LOCALITA	4	a b c d
Number of observations in data set = 24		

**2)**

Dependent Variable: INSETTI

Source	DF	Sum of Squares	F Value	Pr > F
Model	11	3052.12500000	13.40	0.0001
Error	12	248.50000000		
Corrected Total	23	3300.62500000		
	R-Square	C.V.	INSETTI Mean	
	0.924711	6.779353	67.1250000	

Source	DF	Type I SS	F Value	Pr > F
STAGIONE	2	1870.75000000	45.17	0.0001
LOCALITA	3	1084.12500000	17.45	0.0001
STAGIONE*LOCALITA	6	97.25000000	0.78	0.5995

Source	DF	Type III SS	F Value	Pr > F
STAGIONE	2	1870.75000000	45.17	0.0001
LOCALITA	3	1084.12500000	17.45	0.0001
STAGIONE*LOCALITA	6	97.25000000	0.78	0.5995

**3)**

Tukey's Studentized Range (HSD) Test for variable: INSETTI

NOTE: This test controls the type I experimentwise error rate.

Alpha= 0.05 Confidence= 0.95 df= 12 MSE= 20.70833  
 Critical Value of Studentized Range= 3.773  
 Minimum Significant Difference= 6.07

Comparisons significant at the 0.05 level are indicated by '\*\*\*'.

STAGIONE Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit	
autunno - estate	4.555	10.625	16.695	***
autunno - primavera	15.555	21.625	27.695	***
estate - autunno	-16.695	-10.625	-4.555	***
estate - primavera	4.930	11.000	17.070	***
primavera - autunno	-27.695	-21.625	-15.555	***
primavera - estate	-17.070	-11.000	-4.930	***

4)

Scheffe's test for variable: INSETTI

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than Tukey's for all pairwise comparisons.

Alpha= 0.05 Confidence= 0.95 df= 12 MSE= 20.70833  
 Critical Value of F= 3.88529  
 Minimum Significant Difference= 6.3426

Comparisons significant at the 0.05 level are indicated by '\*\*\*'.

STAGIONE Comparison	Simultaneous	Difference Between Means	Simultaneous	
	Lower Confidence Limit		Upper Confidence Limit	
autunno - estate	4.282	10.625	16.968	***
autunno - primavera	15.282	21.625	27.968	***
estate - autunno	-16.968	-10.625	-4.282	***
estate - primavera	4.657	11.000	17.343	***
primavera - autunno	-27.968	-21.625	-15.282	***
primavera - estate	-17.343	-11.000	-4.657	***

5)

Dunnett's T tests for variable: INSETTI

NOTE: This tests controls the type I experimentwise error for comparisons of all treatments against a control.

Alpha= 0.05 Confidence= 0.95 df= 12 MSE= 20.70833  
 Critical Value of Dunnett's T= 2.502  
 Minimum Significant Difference= 5.6938

Comparisons significant at the 0.05 level are indicated by '\*\*\*'.

STAGIONE Comparison	Simultaneous	Difference Between Means	Simultaneous	
	Lower Confidence Limit		Upper Confidence Limit	
estate - autunno	-16.319	-10.625	-4.931	***
primavera - autunno	-27.319	-21.625	-15.931	***

6)

Student-Newman-Keuls test for variable: INSETTI

NOTE: This test controls the type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha= 0.05 df= 12 MSE= 20.70833

Number of Means            2            3  
 Critical Range    4.9574623 6.0700016

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	STAGIONE
A	77.875	8	autunno
B	67.250	8	estate
C	56.250	8	primavera

7)

Tukey's Studentized Range (HSD) Test for variable: INSETTI

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 12 MSE= 20.70833  
 Critical Value of Studentized Range= 3.773  
 Minimum Significant Difference= 6.07

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	STAGIONE
A	77.875	8	autunno
B	67.250	8	estate
C	56.250	8	primave

8)

Scheffe's test for variable: INSETTI

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than REGWF for all pairwise comparisons

Alpha= 0.05 df= 12 MSE= 20.70833  
 Critical Value of F= 3.88529  
 Minimum Significant Difference= 6.3426

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	STAGIONE
A	77.875	8	autunno
B	67.250	8	estate
C	56.250	8	primavera

9)

Tukey's Studentized Range (HSD) Test for variable: INSETTI

NOTE: This test controls the type I experimentwise error rate.

Alpha= 0.05 Confidence= 0.95 df= 12 MSE= 20.70833  
 Critical Value of Studentized Range= 4.199  
 Minimum Significant Difference= 7.8

Comparisons significant at the 0.05 level are indicated by '\*\*\*'.

LOCALITA Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit	
c - b	-1.800	6.000	13.800	
c - a	7.700	15.500	23.300	***
c - d	8.200	16.000	23.800	***
b - c	-13.800	-6.000	1.800	
b - a	1.700	9.500	17.300	***
b - d	2.200	10.000	17.800	***
a - c	-23.300	-15.500	-7.700	***
a - b	-17.300	-9.500	-1.700	***
a - d	-7.300	0.500	8.300	
d - c	-23.800	-16.000	-8.200	***
d - b	-17.800	-10.000	-2.200	***
d - a	-8.300	-0.500	7.300	

10)

Scheffe's test for variable: INSETTI

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than Tukey's for all pairwise comparisons.

Alpha= 0.05 Confidence= 0.95 df= 12 MSE= 20.70833  
 Critical Value of F= 3.49029  
 Minimum Significant Difference= 8.5017

Comparisons significant at the 0.05 level are indicated by '\*\*\*'.

LOCALITA Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit	
c - b	-2.502	6.000	14.502	
c - a	6.998	15.500	24.002	***
c - d	7.498	16.000	24.502	***
b - c	-14.502	-6.000	2.502	
b - a	0.998	9.500	18.002	***
b - d	1.498	10.000	18.502	***
a - c	-24.002	-15.500	-6.998	***
a - b	-18.002	-9.500	-0.998	***
a - d	-8.002	0.500	9.002	
d - c	-24.502	-16.000	-7.498	***
d - b	-18.502	-10.000	-1.498	***
d - a	-9.002	-0.500	8.002	

11)

Dunnett's T tests for variable: INSETTI

NOTE: This tests controls the type I experimentwise error for comparisons of all treatments against a control.

Alpha= 0.05 Confidence= 0.95 df= 12 MSE= 20.70833  
 Critical Value of Dunnett's T= 2.683  
 Minimum Significant Difference= 7.0489

Comparisons significant at the 0.05 level are indicated by '\*\*\*'.

LOCALITA Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit	
c - a	8.451	15.500	22.549	***
b - a	2.451	9.500	16.549	***
d - a	-7.549	-0.500	6.549	

**12)**

Student-Newman-Keuls test for variable: INSETTI

NOTE: This test controls the type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha= 0.05 df= 12 MSE= 20.70833

Number of Means	2	3	4
Critical Range	5.7243844	7.0090341	7.8000392

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	LOCALITA
A	76.500	6	c
B	70.500	6	b
C	61.000	6	a
C	60.500	6	d

**13)**

Tukey's Studentized Range (HSD) Test for variable: INSETTI

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 12 MSE= 20.70833  
 Critical Value of Studentized Range= 4.199  
 Minimum Significant Difference= 7.8

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	LOCALITA
A	76.500	6	c
A			
A	70.500	6	b
B	61.000	6	a
B			
B	60.500	6	d

14)

Scheffe's test for variable: INSETTI

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than REGWF for all pairwise comparisons

Alpha= 0.05 df= 12 MSE= 20.70833  
 Critical Value of F= 3.49029  
 Minimum Significant Difference= 8.5017

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	LOCALITA
A	76.500	6	c
A			
A	70.500	6	b
B	61.000	6	a
B			
B	60.500	6	d

Level of STAGIONE	Level of LOCALITA	N	-----INSETTI----- Mean	SD
autunno	a	2	74.5000000	6.36396103
autunno	b	2	78.5000000	4.94974747
autunno	c	2	85.0000000	4.24264069
autunno	d	2	73.5000000	4.94974747
estate	a	2	61.0000000	5.65685425
estate	b	2	73.0000000	5.65685425
estate	c	2	77.0000000	4.24264069
estate	d	2	58.0000000	2.82842712
primaver	a	2	47.5000000	3.53553391
primaver	b	2	60.0000000	4.24264069
primaver	c	2	67.5000000	3.53553391
primaver	d	2	50.0000000	2.82842712

**ESEMPIO su alimentazione, genotipo dei maschi e delle femmine del capitolo 10:  
 analisi gerarchica della varianza a 3 fattori**

General Linear Models Procedure

1)

Class	Levels	Values
ALIMENTAZIONE	3	I II III
MASCHIO	9	1 2 3 4 5 6 7 8 9
FEMMINA	23	A B C D E F G H I L M N O P Q R S T U V X Y Z

Number of observations in data set = 106

2)

Dependent Variable: PESO

Source	DF	Sum of Squares	F Value	Pr > F
Model	22	1901.70691824	11.35	0.0001
Error	83	632.06666667		
Corrected Total	105	2533.77358491		

R-Square	C.V.	PESO Mean
0.750543	8.805391	31.3396226

Source	DF	Type I SS	F Value	Pr > F
ALIMENT	2	660.43723225	43.36	0.0001
MASCHIO	6	418.19046521	9.15	0.0001
FEMMINA	14	823.07922078	7.72	0.0001

3)

Student-Newman-Keuls test for variable: PESO

NOTE: This test controls the type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha= 0.05 df= 83 MSE= 7.615261  
WARNING: Cell sizes are not equal.  
Harmonic Mean of cell sizes= 32.90066

Number of Means            2            3  
Critical Range    1.3532596    1.6237331

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	ALIMENTI
A	34.4722	36	III
B	30.6957	46	II
C	27.8750	24	I

4)

Tukey's Studentized Range (HSD) Test for variable: PESO

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 83 MSE= 7.615261  
Critical Value of Studentized Range= 3.375  
Minimum Significant Difference= 1.6237  
WARNING: Cell sizes are not equal.  
Harmonic Mean of cell sizes= 32.90066

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	ALIMENT
A	34.4722	36	III
B	30.6957	46	II
C	27.8750	24	I

5)

Scheffe's test for variable: PESO

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than REGWF for all pairwise comparisons

Alpha= 0.05 df= 83 MSE= 7.615261  
 Critical Value of F= 3.10651  
 Minimum Significant Difference= 1.6959  
 WARNING: Cell sizes are not equal.  
 Harmonic Mean of cell sizes= 32.90066

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	ALIMENT
A	34.4722	36	III
B	30.6957	46	II
C	27.8750	24	I

6)

Dunnett's T tests for variable: PESO

NOTE: This tests controls the type I experimentwise error for comparisons of all treatments against a control.

Alpha= 0.05 Confidence= 0.95 df= 83 MSE= 7.615261  
 Critical Value of Dunnett's T= 2.695

Comparisons significant at the 0.05 level are indicated by '\*\*\*'.

MASCHIO Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit	
7 - 1	9.1728	12.5152	15.8575	***
8 - 1	8.7166	12.1333	15.5501	***
9 - 1	7.9312	11.0667	14.2021	***
3 - 1	5.2209	8.5000	11.7791	***
5 - 1	4.7286	8.4762	12.2238	***
2 - 1	5.1979	8.3333	11.4688	***
6 - 1	4.9897	7.9524	10.9151	***
4 - 1	2.9140	6.8333	10.7526	***

7)

Student-Newman-Keuls test for variable: PESO

NOTE: This test controls the type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha= 0.05 df= 83 MSE= 7.615261  
 WARNING: Cell sizes are not equal.  
 Harmonic Mean of cell sizes= 10.27597

Number of Means            2            3            4            5  
 Critical Range    2.4214306 2.9053974 3.1919379 3.3950041

Number of Means            6            7            8            9  
 Critical Range    3.5517011 3.678912 3.7857504 3.877692

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	MASCHIO
A	35.182	11	7
A			
A	34.800	10	8
A			
B	33.733	15	9
B			
B	31.167	12	3
B			
B	31.143	7	5
B			
B	31.000	15	2
B			
B	30.619	21	6
B			
C			
C	29.500	6	4
D	22.667	9	1

8)

Tukey's Studentized Range (HSD) Test for variable: PESO

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 83 MSE= 7.615261  
 Critical Value of Studentized Range= 4.504  
 Minimum Significant Difference= 3.8777  
 WARNING: Cell sizes are not equal.  
 Harmonic Mean of cell sizes= 10.27597

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	MASCHIO	
A	35.182	11	7	
A				
B	34.800	10	8	
B				
B	A C	33.733	15	9
B				
B	D C	31.167	12	3
B				
B	D C	31.143	7	5
B				
B	D C	31.000	15	2
B				
D				
D	C	30.619	21	6
D				
D		29.500	6	4
E	22.667	9	1	

9)

Scheffe's test for variable: PESO

NOTE: This test controls the type I experimentwise error rate  
but generally has a higher type II error rate than REGWF  
for all pairwise comparisons

Alpha= 0.05 df= 83 MSE= 7.615261  
Critical Value of F= 2.05201  
Minimum Significant Difference= 4.9327  
WARNING: Cell sizes are not equal.  
Harmonic Mean of cell sizes= 10.27597

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	MASCHIO
A	35.182	11	7
A			
A	34.800	10	8
A			
B A	33.733	15	9
B A			
B A	31.167	12	3
B A			
B A	31.143	7	5
B A			
B A	31.000	15	2
B A			
B A	30.619	21	6
B			
B	29.500	6	4
C	22.667	9	1

10)

Student-Newman-Keuls test for variable: PESO

NOTE: This test controls the type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha= 0.05 df= 83 MSE= 7.615261  
 WARNING: Cell sizes are not equal.  
 Harmonic Mean of cell sizes= 4.339623

Number of Means	2	3	4	5	6
Critical Range	3.7261254	4.470859	4.9117908	5.2242715	5.4653986
Number of Means	7	8	9	10	11
Critical Range	5.6611522	5.8255563	5.9670372	6.0910485	6.2013191
Number of Means	12	13	14	15	16
Critical Range	6.3005091	6.3905809	6.473023	6.5489903	6.6193966
Number of Means	17	18	19	20	21
Critical Range	6.6849769	6.7463311	6.8039548	6.8582624	6.9096034
Number of Means	22	23			
Critical Range	6.9582756	7.0045344			

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	FEMMINA	
A	37.333	3	Z	
A				
A	37.333	6	T	
A				
B	36.750	4	D	
B				
B	A C	35.200	5	U
B	A C			
B	A C	35.000	5	G
B	A C			
B	A C	34.400	5	V
B	A C			
B	A C	34.333	6	X
B	A C			
B	A C	34.000	3	N
B	A C			
B	D A C	33.600	5	E
B	D A C			
E	B D A C	32.833	6	Q
E	B D A C			
E	B D A C	32.600	5	S
E	B D A C			
E	B D A C	32.000	3	L
E	B D A C			
E	B D A C	31.333	6	Y
E	B D A C			
E	B D A C	31.200	5	P
E	B D C			
E	B D F C	30.400	5	R
E	D F C			
E	D F C	29.750	4	F
E	D F C			
E	D F C	29.000	4	M
E	D F			
E	G D F	27.600	5	O
E	G F			
E	G F H	27.000	3	I
E	G F H			
E	G F H	26.667	3	H
G	F H			
G	F H	25.000	6	C
G	H			
G	H	23.200	5	B
	H			
	H	22.000	4	A

11)

Tukey's Studentized Range (HSD) Test for variable: PESO

NOTE: This test controls the type I experimentwise error rate,  
but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 83 MSE= 7.615261  
Critical Value of Studentized Range= 5.288  
Minimum Significant Difference= 7.0045  
WARNING: Cell sizes are not equal.  
Harmonic Mean of cell sizes= 4.339623

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	FEMMINA
A	37.333	3	Z
A			
A	37.333	6	T
A			
B A	36.750	4	D
B A			
B A C	35.200	5	U
B A C			
B A C	35.000	5	G
B A C			
B D A C	34.400	5	V
B D A C			
B D A C	34.333	6	X
B D A C			
E B D A C	34.000	3	N
E B D A C			
E B D A C F	33.600	5	E
E B D A C F			
E B D A C F	32.833	6	Q
E B D A C F			
E B D A C F	32.600	5	S
E B D A C F			
E B D A G C F	32.000	3	L
E B D A G C F			
E B D A G C F	31.333	6	Y
E B D A G C F			
E B D A G C F	31.200	5	P
E B D A G C F			
E B D A G C F	30.400	5	R
E B D G C F			
E B D H G C F	29.750	4	F
E D H G C F			
E I D H G C F	29.000	4	M
E I D H G F			
E I D H G F	27.600	5	O
E I H G F			
E I H G F	27.000	3	I
I H G F			
I H G F	26.667	3	H
I H G			
I H G	25.000	6	C
I H			
I H	23.200	5	B
I			
I	22.000	4	A

12)

Scheffe's test for variable: PESO

NOTE: This test controls the type I experimentwise error rate  
but generally has a higher type II error rate than REGWF  
for all pairwise comparisons

Alpha= 0.05 df= 83 MSE= 7.615261  
Critical Value of F= 1.67191  
Minimum Significant Difference= 11.362  
WARNING: Cell sizes are not equal.  
Harmonic Mean of cell sizes= 4.339623

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	FEMMINA
A	37.333	3	Z
A			
A	37.333	6	T
A			
A	36.750	4	D
A			
B A	35.200	5	U
B A			
B A	35.000	5	G
B A			
B A C	34.400	5	V
B A C			
B A C	34.333	6	X
B A C			
B A C	34.000	3	N
B A C			
B A C	33.600	5	E
B A C			
B D A C	32.833	6	Q
B D A C			
B D A C	32.600	5	S
B D A C			
B D A C	32.000	3	L
B D A C			
B D A C	31.333	6	Y
B D A C			
B D A C	31.200	5	P
B D A C			
B D A C	30.400	5	R
B D A C			
B D A C	29.750	4	F
B D A C			
B D A C	29.000	4	M
B D A C			
B D A C	27.600	5	O
B D A C			
B D A C	27.000	3	I
B D A C			
B D A C	26.667	3	H
B D C			
B D C	25.000	6	C
D C			
D C	23.200	5	B
D			
D	22.000	4	A

## CAPITOLO XI

### TEST NON PARAMETRICI

#### PER PIU' CAMPIONI

##### 11.1. TEST NON PARAMETRICI ANALOGHI ALL'ANALISI DELLA VARIANZA

**Tutti i test non parametrici hanno in comune una sola caratteristica: non richiedono la normalità della distribuzione della popolazione.** Presentano quindi un vantaggio indubbio rispetto a quelli parametrici, sia quando la non-normalità è evidente, sia quando la normalità della distribuzione dei dati è solo probabile. In esperimenti nuovi e con pochi dati, non è possibile dimostrare che la condizione di normalità è rispettata, data la bassa potenza dei test relativi: non essere in grado di rifiutare l'ipotesi nulla sulla normalità ed omoschedasticità, a causa del ridotto numero di repliche, non significa che essa sia vera.

**Il test F ed il test t bidirezionale sono metodi statisticamente robusti, rispetto a questa condizione: la non-normalità, pure quando causata da forte asimmetria, determina errori ridotti, nel calcolo della significatività.** Alcuni autori stimano che una forte asimmetria può fissare la variazione delle probabilità  $\alpha$  tra 0.04 e 0.07 quando quella nominale è 0.05 e tra 0.005 e 0.02 quando  $\alpha$  nominale è 0.01. E' un'approssimazione che permette di accettare i risultati, poiché anche i test non parametrici non sempre conducono a stime più precise.

Le trasformazioni dei dati (presentate e discusse nella parte finale del capitolo precedente) possono rappresentare una soluzione. Ma se, dopo la loro applicazione, le caratteristiche della distribuzione non migliorano in modo sensibile, avvicinandosi sufficientemente a quelle richieste dai test parametrici, i metodi **non parametrici sono indubbiamente da preferire.**

Le situazioni più frequenti, in cui è consigliato scegliere un test non parametrico, si realizzano quando:

- 1 - **i campioni mantengono una distribuzione estremamente asimmetrica;**
- 2 - **il test è a una coda e l'alternativa sarebbe il test t unilaterale** (che è più sensibile all'asimmetria);
- 3 - **i campioni a confronto hanno un numero di dati molto differente** (poiché soprattutto in statistica parametrica influisce sulla potenza del test);
- 4 - **le distribuzioni non sono omoschedastiche** (condizione d'invalidità severa, per i test parametrici);
- 5 - **sono presenti valori fortemente anomali (*outliers*)** (perdita di valore della media e quindi delle misure da essa derivate, come la varianza);
- 6 - **è stata utilizzata una scala ordinale oppure qualitativa** (e quindi è doveroso utilizzare l'informazione realmente fornita dai dati).

Per campioni sufficientemente grandi, i test parametrici possono essere utilizzati anche con dati rappresentati da conteggi di variabili binarie: abitualmente, **hanno una distribuzione non normale in piccoli campioni, ma la approssimano abbastanza bene con campioni sufficientemente grandi**, come già ripetutamente dimostrato per la distribuzione binomiale, la poissoniana, la ipergeometrica e i test  $\chi^2$  in tabelle 2 x 2.

Anche i test non parametrici per più campioni, proposti in tempi e situazioni diverse da autori differenti per affrontare problemi specifici, possono essere didatticamente classificati sulla base dell'organizzazione dell'esperimento. Nei testi di statistica applicata è molto frequente la suddivisione in **test per k campioni indipendenti** e **test per k campioni dipendenti**, analoghi rispettivamente all'**analisi della varianza ad un criterio** e all'**analisi a due criteri di classificazione**.

I test per **k campioni indipendenti** maggiormente utilizzati nella ricerca e più frequentemente riportati nei programmi informatici per verificare l'ipotesi sulla forma della distribuzione sono:

- il  $\chi^2$  e il **test G in tabelle m x n** (già presentati in modo dettagliato nel capitolo specifico), mentre quelli sull'eguaglianza della tendenza centrale sono
  - l'estensione del **test della mediana** e il **test di Nemenyi**,
  - il test di **Kruskal-Wallis**, detto anche analisi non parametrica della varianza per un criterio di classificazione, che può essere visto come l'estensione del test di Wilcoxon–Mann-Whitney, fondato sui ranghi;
- per verificare le ipotesi sulla omogeneità della variabilità vi è
- il **test per l'eterogeneità della varianza** che, come estensione di quello per 2 campioni, nella fase finale applica il test di Kruskal- Wallis.

I test per **k campioni dipendenti** che verificano le ipotesi sulla uguaglianza della mediana sono:

- il test **Q di Cochran** e il test di **Bowker**, estensione di quello di Mac Nemar
- il test di **Friedman**, denominato pure analisi della varianza per ranghi a due criteri di classificazione.

## 11.2. ESTENSIONE DEL TEST DELLA MEDIANA

L'estensione a k campioni (con  $k > 2$ ) del test della mediana (*the median test for several samples*) serve per **verificare l'ipotesi se le tendenze centrali sono significativamente differenti**.

L'ipotesi nulla, che prende in considerazione **le mediane** ( $me$ ) delle popolazioni dalle quali sono state estratti i vari campioni (A, B, ..., K), è che esse siano tutte uguali

$$H_0: me_A = me_B = \dots = me_K$$

mentre l'ipotesi alternativa, solo bilaterale come in tutti i confronti multipli effettuati simultaneamente, è

**$H_1$ : le mediane dei  $k$  gruppi non sono tutte uguali**

E' un test molto semplice, che appunto per questo non ha un autore definito. E' fondato sulle tecniche di analisi di una tabella  $k \times 2$  con il  $c^2$  o il test  $G$  quando il numero di osservazioni è sufficientemente grande e sulla distribuzione ipergeometrica quando il numero di osservazioni è limitato. E' utilizzato in modo appropriato **quando le misure sono approssimate e possono essere ordinate per rango solo in modo parziale, con molti valori identici, in particolare se collocati agli estremi..** Il test si avvale solo di una parte dell'informazione contenuta in una misura quantitativa e quindi accetta che molti dati abbiano un valore identico.

E' utile soprattutto nel caso in cui i **dati** sono **misurati con precisione non costante oppure sono troncati verso uno o ambedue gli estremi**. Nella raccolta dei dati ambientali può succedere di utilizzare strumenti di misura tarati con precisione solamente per i valori centrali, più frequenti; ma che non sono in grado di valutare le misure collocate verso gli estremi, più rare, troppo piccole per essere rilevate dallo strumento o troppo grandi per essere determinate con la stessa precisione di quelle intorno alla norma.

Si deve ricorrere necessariamente al test della mediana anche nell'**analisi di distribuzioni in cui le misure estreme siano state raggruppate in classi aperte**, formate da valori minori ( $< X$ ) e/o maggiori ( $> X$ ) di una quantità determinata.

**Per i calcoli manuali**, il fattore limitante all'uso del test della mediana per  $k$  campioni indipendenti è quello del  $c^2$  o del test  $G$ ; il **numero di osservazioni deve essere sufficientemente elevato**.

Nel caso di  $k$  campioni con **un numero di osservazioni molto ridotto**, si può utilizzare il **metodo esatto, fondato sulla distribuzione ipergeometrica**, analogo al metodo esatto di Fisher in tabelle  $2 \times 2$ ; ma richiede molti calcoli e quindi è praticamente applicabile solo con programmi informatici.

**Il metodo dell'estensione del test della mediana** per il confronto tra le tendenze centrali di più campioni può essere schematizzato in alcuni punti fondamentali, seguendone l'applicazione ad un esempio.

1 – Dopo aver raccolto i dati di  $k$  campioni indipendenti, con un numero  $r$  di osservazioni che può essere diverso, come nell'analisi della varianza ad un criterio di classificazione,

GRUPPO		
A	B	C
< 1	3,7	2,1
<1	2,8	2,5
3,8	0,9	2,9
2,1	2,2	>10
--	2,5	8,7
--	--	1,6

2 - ordinare per ranghi tutte le osservazioni dei  $k$  gruppi a confronto, come se fossero un gruppo unico, mantenendo per ogni valore l'informazione del gruppo di appartenenza

<1	<1	0,9	1,6	2,1	2,1	2,2	<b>2,5</b>	2,5	2,8	2,9	3,7	3,8	8,7	>10
A	A	B	C	A	C	B	<b>B</b>	C	B	C	B	A	C	C

3- Identificare la mediana di questa distribuzione unica,

che con dati dell'esempio è l'8° valore, il 2,5 del gruppo B (in grassetto).

Poiché il dato successivo ha un valore identico, collocare la mediana tra il 7° e l'8° valore.

4 - Contare quante sono le osservazioni di ogni gruppo che hanno valore inferiore ( $n_1$ ) e quante quelle che hanno valore superiore (od uguale, in questo caso) ( $n_2$ ) alla mediana;

GRUPPI	< mediana	≥ mediana
A	3	1
B	2	3
C	2	4

l'eventuale valore corrispondente alla mediana può essere classificato in uno dei due sottogruppi indifferentemente (in questo caso è stato contato con quelli superiori alla mediana);

con  $k$  gruppi si ottiene una distribuzione di frequenza in una tabella  $k \times 2$ , come quella appena riportata.

5 - **Se è vera l'ipotesi nulla** ( $H_0$ : i vari gruppi a confronto sono estratti dalla stessa popolazione o da campioni con la stessa tendenza centrale), **ogni gruppo dovrebbe avere lo stesso numero di osservazioni prima e dopo la mediana**; se i due gruppi non hanno la stessa frequenza, la stima delle frequenze attese può essere fatta come nella tabella sottoriportata

GRUPPI	< mediana	≥ mediana	Totale
A	$4 \times 7 / 15 = 1,87$	$4 \times 8 / 15 = 2,13$	<b>4</b>
B	$5 \times 7 / 15 = 2,33$	$5 \times 8 / 15 = 2,67$	<b>5</b>
C	$6 \times 7 / 15 = 2,80$	$6 \times 8 / 15 = 3,20$	<b>6</b>
Totale	<b>7</b>	<b>8</b>	<b>15</b>

**Se l'ipotesi nulla è falsa**, almeno un gruppo dovrebbe avere una prevalenza significativa di osservazioni con valore minore o maggiore della mediana.

6 - Nel caso di **campioni grandi**, l'accordo tra la distribuzione osservata e la distribuzione attesa può essere analizzata con il test  $\chi^2$  corrispondente (o il test G), con **gdl** uguali a **k - 1** (k = numero di gruppi).

7- Se **i campioni sono piccoli** si ricorre al metodo esatto, disponendo di un programma informatico adeguato.

Indicando con

GRUPPI	< mediana	≥ mediana	Totale
A	<b>a</b>	<b>b</b>	<b>n<sub>1</sub></b>
B	<b>c</b>	<b>d</b>	<b>n<sub>2</sub></b>
C	<b>e</b>	<b>f</b>	<b>n<sub>3</sub></b>
Totale	<b>n<sub>4</sub></b>	<b>n<sub>5</sub></b>	<b>N</b>

dove

**a, b, c, d, e, f,** sono le frequenze inferiori e superiori alla mediana nei **k** gruppi,

**n<sub>1</sub>, n<sub>2</sub>, n<sub>3</sub>, n<sub>4</sub>, n<sub>5</sub>, N** sono rispettivamente i totali marginali e il totale generale

la formula generale per calcolare la probabilità esatta della risposta specifica ottenuta, derivata dalla **distribuzione ipergeometrica** è

$$P_e = \frac{n_1!n_2!n_3!n_4!n_5!}{a!b!c!d!e!f!N!}$$

nel caso di tre gruppi;

è **facilmente estensibile a k gruppi** mediante

$$P_e = \frac{\text{Prodotto dei fattoriali dei totali marginali!}}{\text{Prodotto dei fattoriali delle frequenze di casella e del totale generale}}$$

Con questa formula, si stima la probabilità di avere solo la risposta sperimentale.

Come nel metodo esatto di Fisher, **la probabilità calcolata deve essere sommata con quelle di tutte le risposte più estreme**. Poiché in questo caso si tratta di un test bilaterale (come sempre con **k** campioni), **le risposte più estreme possono essere individuate con facilità dal valore della loro probabilità esatta: sono risposte più estreme tutte quelle che hanno una probabilità inferiore a quella calcolata per la risposta sperimentale.**

ESEMPIO. Si intende verificare se esiste una differenza significativa nella densità (numero di individui entro una superficie unitaria) di 5 specie vegetali (A, B, C, D, E). A questo scopo è stato raccolto un numero variabile di campioni, in aree di dimensione diversa, stimandone la concentrazione media in modo approssimato.

A	4	3	<1	7	<1	1	2	3	1
B	2	9	7	6	9	8	7	--	--
C	3	4	2	4	5	2	<1	3	--
D	8	7	9	>9	7	--	--	--	--
E	5	4	6	5	5	7	--	--	--

Le 5 aree hanno una concentrazione mediana significativamente differente?

Risposta.

1 - I rapporti sono stati calcolati su **k** (con **k > 2**) campioni di superficie non costanti e sono espressi con misure approssimate, che forniscono solo la dimensione del fenomeno, non una misura di una scala ad intervalli o di rapporti; di conseguenza, il test appropriato è l'estensione della mediana per **k** campioni.

L'ipotesi nulla è che tutti i campioni abbiano la stessa mediana

$$H_0: me_A = me_B = me_C = me_D = me_E$$

con ipotesi alternativa che almeno una sia differente o

$$H_1: \text{non tutte le mediane sono uguali}$$

2 - Come prima elaborazione dei dati, è utile costruire una distribuzione ordinata di tutti i valori, conservando per ognuno l'informazione del gruppo di appartenenza, allo scopo di

- determinare la mediana comune,
- contare per ognuno dei 5 gruppi a confronto quanti sono i valori inferiori e quelli superiori

<1	<1	<1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5	
A	A	C	A	A	A	B	C	C	A	A	C	C	A	C	C	E	C	E	E	E	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	Segue

6	6	7	7	7	7	7	7	7	8	8	9	9	9	>9
B	E	A	B	B	D	D	E	B	D	B	B	D	D	
22	23	24	25	26	27	28	29	30	31	32	33	34	35	

Nell'esempio, i dati sono 35: la mediana è il 18° valore e quindi 5. Poiché i valori identici alla mediana sono più di uno ed appartengono a vari gruppi, risulta utile aggregare il valore della mediana al gruppo dei valori maggiori di essa (>).

3 - Per ognuno dei 5 gruppi a confronto, si contano quanti sono i valori minori e quanti sono uguali o maggiori alla mediana. Si ottiene la tabella (5 x 2) seguente.

Gruppi	<	> e =	Totale
A	8	1	9
B	1	6	7
C	7	1	8
D	0	5	5
E	1	5	6
Totale	17	18	35

4 - Se è vera l'ipotesi nulla che tutti i gruppi hanno la stessa tendenza centrale, la distribuzione attesa è facilmente intuibile: ogni gruppo dovrebbe avere metà (nell'esempio esattamente 17/35) dei suoi valori prima della mediana e l'altra metà (esattamente 18/35) dei suoi valori dopo la mediana.

**La significatività delle differente distribuzione dei 5 gruppi è verificata con un test  $\chi^2$  con 4 df.**

Con i dati dell'esempio, il numero di dati potrebbe essere ritenuto insufficiente per una sua applicazione valida per il test  $\chi^2$  o il test G, poiché tutte le 10 caselle delle frequenze attese hanno valori inferiori a 5. Sarebbe quindi più corretto l'uso della distribuzione ipergeometrica.

E' una ulteriore conferma del fatto che, quando si utilizza in modo ridotto la quantità d'informazione contenuta nei dati raccolti, per raggiungere con il test la potenza sufficiente è necessario disporre di campioni di dimensioni relativamente grandi.

Nel caso del test chi quadrato, quando si hanno gruppi con un numero di osservazioni particolarmente ridotto è utile procedere ad una aggregazione delle classi; con i dati dell'esempio, sarebbe conveniente unire i 4 gruppi minori B e C, D e E in due soli gruppi secondo le loro affinità. Si ricostruiscono le condizioni di validità, che nella distribuzione attesa richiedono almeno 5 osservazioni per casella, anche se tale operazione comporta un'ulteriore perdita di informazione e non permette l'analisi dettagliata per ogni singola specie.

I programmi di computer di norma forniscono almeno tre risultati, tra loro non identici:

- il valore del  $c^2$  con gdl 4, (sovente, un secondo  $c^2$  con la correzione di Yates),
- il valore del test G sempre con 4 gdl, (sovente, una seconda versione con la correzione di Williams o di Mantel-Haenzel),
- la probabilità esatta, stimata con la distribuzione ipergeometrica.

### 11.3. CENNI SUL TEST DI NEMENYI

Il test di Nemenyi può essere visto come un caso particolare del test della mediana e serve per verificare le stesse ipotesi, ricorrendo ad un metodo e a una logica differenti da quella del chi quadrato.

L'ipotesi nulla resta

$$H_0: me_A = me_B = \dots = me_K$$

e l'ipotesi alternativa, solo bilaterale, è ancora

$$H_1: \text{le mediane non sono tutte uguali}$$

Il test di Nemenyi richiede due condizioni sperimentali più restrittive del precedente test sulla mediana:

- **i k gruppi a confronto devono avere lo stesso numero d'osservazioni,**
- **che deve essere grande.**

Ha pure il **grave svantaggio di richiedere tabelle di valori critici difficilmente reperibili (qui non riportati).**

Il lavoro originale di **P. Nemenyi** è citato in alcuni testi di statistica, ma non è mai stato pubblicato: è una tesi di dottorato presentata nel 1963 presso l'Università di Princeton (dal titolo *Distribution-free multiple comparisons*). Tuttavia, in una rassegna che vuole essere abbastanza completa, il test è da ricordare, in quanto **presente in alcuni programmi informatici da molti anni**, poiché il Dipartimento di Statistica della Università di Princeton, dotato di una struttura di ricerca di fama mondiale che produceva e diffondeva programmi informatici di alto livello, lo inserì nel suo circuito internazionale.

La metodologia, differente dal test della mediana solo nella parte conclusiva, richiede che:

1- nella programmazione dell'esperimento si raccolga un numero d'osservazioni identico in tutti i gruppi a confronto

$$n_1 = n_2 = \dots = n_k$$

come nell'esempio seguente con 4 gruppi (che in realtà hanno un numero d'osservazioni troppo limitato, per un'applicazione corretta)

GRUPPO			
A	B	C	D
<1	3,7	2,1	3,8
<1	2,8	2,5	2,2
3,8	0,9	2,9	5,2
2,1	2,2	>10	>10
3,2	2,5	8,7	>10

2 - Dopo aver definito l'ipotesi nulla

$$H_0: me_A = me_B = \dots = me_K$$

con ipotesi alternativa bilaterale

**H<sub>1</sub>: le mediane non sono tutte uguali**

3 - ordinare per ranghi tutte le osservazioni dei **k** gruppi a confronto come se fossero un gruppo unico, mantenendo per ogni valore l'informazione del gruppo di appartenenza

<1	<1	0,9	2,1	2,1	2,2	2,2	2,5	2,5	2,8	<b>mediana</b> segue
A	A	B	A	C	B	D	B	C	B	

2,9	3,2	3,7	3,8	3,8	5,2	8,7	>10	>10	>10
C	A	B	A	D	D	C	C	D	D

4 - Successivamente, **identificare la mediana** di questa distribuzione unica (nella tabella precedente è a metà tra il 10° e l'11° valore) e, per ogni gruppo,

GRUPPI	< mediana	> <b>mediana</b>	Totale
A	3	<b>f<sub>A</sub> = 2</b>	5
B	4	<b>f<sub>B</sub> = 1</b>	5
C	2	<b>f<sub>C</sub> = 3</b>	5
D	1	<b>f<sub>D</sub> = 4</b>	5

contare quante sono le osservazioni che hanno valore superiore alla mediana (l'eventuale valore corrispondente alla mediana è classificato in uno dei due gruppi, anche se non dovrebbero risultare sbilanciati).

Nella distribuzione dei valori sopra e sotto la mediana, costruita come nel test relativo, **si prendono in considerazione solo le frequenze superiori alla mediana**; si utilizza solo una serie di frequenze  $f_1, f_2, \dots, f_k$  (quelle riportate il grassetto nella tabella precedente).

5 - Con  $k$  gruppi e quindi  $k$  frequenze di valori superiori alla mediana, sono possibili  $k(k - 1)/2$  confronti 2 a 2, tra le frequenze assolute  $f_i$ .

Da esse si deriva l'indice  $h$ , determinato dalla **differenza massima in valore assoluto**

$$h = \max |f_i - f_j|$$

Con i dati dell'esempio, fra i 6 possibili confronti 2 a 2, si sceglie quello di B con D

$$h = |1 - 4| = 3$$

che fornisce la differenza assoluta massima  $h = 3$ .

6 - Attraverso **tavole dei valori critici**, è possibile stimare la probabilità di ottenere per caso, nella condizione che l'ipotesi nulla sia vera, uno scarto uguale o superiore a quello stimato.

I programmi informatici riportano la probabilità  $\alpha$ . Tale valore permette di decidere se è possibile rifiutare l'ipotesi nulla.

**Per gli stessi scopi, in letteratura si trovano altri test.** Fra questi, è possibile ricordare:

- il test di **Rijkoort**, proposto nel 1952, che si fonda su modalità simili alla somma dei ranghi utilizzata nel test di T Wilcoxon;
- il test proposto congiuntamente da **Brown e Mood** nel 1951, affine a quello della mediana;
- il test di **Bhapkar** del 1961;
- il test di **Deshpandé** ed il test di **Sugiura**, elaborati in modo indipendente nel 1965 e riproposti insieme con correzioni nel 1968, che utilizzano il calcolo delle precedenze e quindi sono una estensione del test U.

**Poiché sono test per ora presenti in pochissimi programmi informatici ed analoghi a quelli già illustrati in modo dettagliato, per la loro presentazione si rimanda a testi specifici.**

#### **11.4. ANALISI DELLA VARIANZA PER RANGHI AD UN CRITERIO DI CLASSIFICAZIONE: IL TEST DI KRUSKAL-WALLIS**

Quando si utilizzano misure con una scala ordinale, quindi tutti i dati possono essere disposti in ranghi con un numero nullo o comunque ridottissimo di valori uguali, è utile ricorrere ad **un test più potente del test della mediana**. La quantità di informazione contenuta in ogni osservazione è superiore a

quella utilizzata nel test precedente; di conseguenza, diviene più probabile verificare la significatività della differenza nella tendenza centrale, pure disponendo di un numero inferiori di dati. E' lo stesso concetto espresso nel confronto tra test dei segni e test di Wilcoxon-Mann-Whitney, nel caso di due campioni.

Il test proposto nel 1952 da Kruskal e Wallis (**Kruskal-Wallis One-Way ANOVA by Ranks** o più semplicemente **the Kruskal-Wallis test**, con l'articolo di W.H. Kruskal e W. A.. Wallis *Use of ranks in one criterion variance analysis* pubblicato da *Amer. Statist. Ass.*, 1952, vol. 47, pp. 583–621 e con l'articolo di W. H. Kruskal *A non parametric test for the several sample problem* pubblicato su *Ann. Math. Statist.*, 1952, vol. 23, pp. 525-540) è l'equivalente non parametrico dell'analisi della varianza ad un criterio di classificazione.

E' uno dei più potenti per verificare l'ipotesi nulla  $H_0$  se  $k$  gruppi indipendenti provengano dalla stessa popolazione oppure da popolazioni che abbiano **la medesima mediana**.

Anche la **metodologia del test di Kruskal-Wallis** è molto semplice e può essere schematizzata in alcuni passaggi.

1 – Per verificare l'ipotesi nulla che tutti i campioni abbiano la stessa mediana

$$H_0: me_A = me_B = me_C = me_D = me_E$$

con ipotesi alternativa che almeno una sia differente o

$$H_1: \text{non tutte le mediane sono uguali}$$

come nell'analisi della varianza ad un criterio di classificazione,

i dati dei  $k$  gruppi a confronto

GRUPPO			
A	B	C	D
76	47	55	36
85	52	46	18
22	63	71	29
67	---	42	46
72	---	---	---
81	---	---	---

possono essere riportati in una tabella e i gruppi a confronto avere un diverso numero d'osservazioni.

2 - Tutte le osservazioni dei  $k$  gruppi devono essere considerate come una serie unica e convertite in ranghi, mantenendo la stessa forma della tabella;

<b>GRUPPO</b>			
<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
15	8	10	4
17	9	6,5	1
2	11	13	3
12	---	5	6,5
14	---	---	---
16	---	---	---

Se sono presenti misure uguali, a ciascuna di esse deve essere assegnato il loro rango medio.

3 - Calcolare

<b>GRUPPO</b>				
	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
	15	8	10	4
	17	9	6,5	1
	2	11	13	3
	12	--	5	6,5
	14	--	--	--
	16	--	--	--
<b><math>R_i</math></b>	<b>76</b>	<b>28</b>	<b>34,5</b>	<b>14,5</b>
<b><math>n_i</math></b>	<b>6</b>	<b>3</b>	<b>4</b>	<b>4</b>
<b><math>\bar{r}_i</math></b>	<b>12,67</b>	<b>9,33</b>	<b>8,63</b>	<b>3,63</b>

- la somma dei ranghi di ogni gruppo ( $R_i$ ) e quella totale ( $R$ ),
- il numero di osservazioni di ogni gruppo ( $n_i$ ) e totale ( $N$ ),

- da cui la media di ogni gruppo ( $\bar{r}_i$ )

Con

$$\mathbf{R} = 76 + 28 + 34,5 + 14,5 = \mathbf{153}$$

e

$$N = 6 + 3 + 4 + 4 = 17$$

la media generale  $\bar{r}$

$$\bar{r} = \mathbf{R} / N = \mathbf{153} / \mathbf{17} = \mathbf{9}$$

risulta uguale a 9.

4 - Se i campioni provengono dalla stessa popolazione o da popolazioni con la stessa tendenza centrale (**H<sub>0</sub> vera**), queste medie aritmetiche dei ranghi di ogni gruppo ( $\bar{r}_i = \mathbf{R}_i / \mathbf{n}_i$ ) dovrebbero essere statisticamente simili tra loro e alla media generale.

Da questo concetto è possibile derivare la formula per il calcolo di un indice (**g**), che dipende dalle differenze tra le medie dei gruppi e la media generale dei ranghi (nella formula sottostante è nascosta dalla stima della media attraverso la somma totale dei ranghi e che ovviamente dipende da **N**)

$$\mathbf{g} = \frac{12}{N(N+1)} \cdot \sum_{i=1}^k n_i \left[ \frac{\bar{r}_i - (N+1)}{2} \right]^2$$

La quantità

$$\mathbf{N(N+1) / 12}$$

è la varianza (riportata al denominatore nelle formule generali), che dipende solo da **N**, mentre la media degli **n<sub>i</sub>** ranghi ( $r_{ij}$ ) è

$$\bar{r}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} r_{ij}$$

Come nell'analisi della varianza, il parametro **g** (vari testi lo indicano con **KW**, iniziali dei due autori) può essere calcolato a partire dalle somme, con **una formula abbreviata** che offre anche il vantaggio di evitare le approssimazioni dovute alle medie

$$g = \left( \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} \right) - 3(N+1)$$

dove:

**n<sub>i</sub>** è il numero di dati del campione o gruppo i-esimo,

**N** è il numero totale di osservazioni dei k campioni,

**k** è il numero di campioni a confronto,

$R_i$  è la somma dei ranghi del campione o gruppo  $i$ -esimo,  
e la sommatoria è estesa a tutti i  $k$  gruppi.

Di conseguenza, con una formula che evidenzia in modo più evidente le sue componenti, anche se i calcoli richiedono un tempo maggiore, l'ipotesi di uguaglianza fra mediane è basata sulla funzione

$$g = \frac{12}{N(N+1)} \sum_{i=1}^k n_i (\bar{r}_i - \bar{r})^2 = \frac{12}{N(N+1)} \sum_{i=1}^k n_i \left( \bar{r}_i - \frac{N+1}{2} \right)^2$$

dove

$\bar{r}$  è la media generale dei ranghi che, con  $N$  dati, corrisponde a  $\frac{N+1}{2}$

$\bar{r}_i$  è la media dei ranghi del gruppo  $i$

**Nel caso di campioni piccoli con valori ripetuti, è conveniente dare la preferenza a questa formula.**

Il **parametro  $g$  si distribuisce approssimativamente come la distribuzione  $\chi^2$  con gdl  $k-1$**  (dove  $k$  è il numero di gruppi a confronto), quando le dimensioni del campione rispettano le condizioni minime richieste per la validità del chi quadrato (numero totale di osservazioni  $N$  non eccessivamente ridotto e numero minimo di dati per gruppo  $n_i$  non inferiore a 5).

**L'approssimazione alla distribuzione chi quadrato è tanto migliore quanto maggiore è il numero ( $k$ ) di gruppi ed il numero di osservazioni entro ogni gruppo è alto, maggiore di 5.**

Quando il **numero di gruppi a confronto è ridotto (uguale a 3)** ed il **numero di osservazioni entro ogni gruppo è basso (inferiore a 5)** la distorsione dalla distribuzione  $\chi^2$  è elevata; di conseguenza, **per la significatività di  $g$  (o KW) si fa ricorso a tabelle specifiche**, predisposte da Kruskal con Wallis (1952).

**Sono tavole limitate a casi molto particolari, in quanto valgono solo per analisi con 3 gruppi e dimensioni non superiori a 5 osservazioni in ogni gruppo.**

Esse iniziano da dimensioni minime di 2, 1 e 1, nei 3 gruppi. **Sono dimensioni nettamente inferiori a quelle richieste per il test  $F$** : per campioni molto piccoli è quindi preferibile ricorrere al test non parametrico.

**Tabella dei valori critici di g (o KW) del test di Kruskal -Wallis,  
per confronti fra 3 campioni con un numero ridotto di osservazioni( £ 5).**

I 3 campioni devono essere ordinati per dimensioni in modo decrescente

(l'ultima riga coincide con il  $\chi^2$  per 2 df)

<b>n<sub>i</sub> per campione</b>			<b>valori critici alle probabilità riportate</b>				
n <sub>1</sub>	n <sub>2</sub>	n <sub>3</sub>	a =.10	a =.05	a =.01	a =.005	a =.001
2	2	2	4,57				
3	2	2	4,50	4,71			
3	3	2	4,56	5,36			
3	3	3	4,63	5,60		7,20	
4	2	2	4,46	5,33			
4	3	2	4,51	5,45	6,45	7,00	
4	3	3	4,71	5,73	6,75	7,32	8,02
5	2	2	4,36	5,16	6,53		
5	3	2	4,65	5,25	6,82	7,18	
5	3	3	4,53	5,65	7,08	7,51	8,24
5	4	2	4,54	5,27	7,12	7,57	8,11
5	4	3	4,55	5,63	7,44	7,91	8,50
5	4	4	4,62	5,62	7,76	8,14	9,00
5	5	2	4,62	5,34	7,27	8,13	8,68
5	5	3	4,54	5,71	7,54	8,24	9,06
5	5	4	4,53	5,64	7,77	8,37	9,32
5	5	5	4,56	5,78	7,98	8,72	9,68
<b>N grande e k = 3</b>			<b>4,61</b>	<b>5,99</b>	<b>9,21</b>	<b>10,60</b>	<b>13,82</b>

Si rifiuta l'ipotesi nulla, alla probabilità  $\alpha$  riportata nella tabella, quando il valore  $g$  (oppure **KW**) calcolato è uguale o superiore a quello critico riportato nella tabella.

L'ultima riga riportata in questa tabella coincide con i valori critici della distribuzione chi quadrato per gradi di libertà 2.

**A differenza dell'analisi della varianza, il test di Kruskal-Wallis può essere utilizzato anche quando un gruppo ha una sola osservazione.** Dal testo *Non parametric statistical methods* di M. Hollander e D. Wolfe (John Wiley & Sons, New York) del 1973 è tratto questo esempio, sulle percentuali di acqua contenuti in cinque sostanze diverse:

Percentuali di acqua contenuti in 5 campioni

A	B	C	D	E
7,8	5,4	8,1	7,9	7,1
8,3	7,4	6,4	9,5	
7,6	7,1		10,0	
8,4				
8,3				

Dopo trasformazione in ranghi ed aver calcolato la somma dei ranghi per colonna

	A	B	C	D	E
	7	1	9	8	3,5
	10,5	5	2	13	
	6	3,5		14	
	12				
	10,5				
R <sub>j</sub>	46	9,5	11	35	3,5

Recentemente è stato evidenziato, come mostra una lettura attenta dei valori riportati nella tabella, che il valore di  $g$  non è monotono: il suo andamento è prima crescente e poi decrescente, è asimmetrico e caratterizzato da numerosi valori modali. Pertanto, sono state evidenziate perplessità riguardo alla sua effettiva capacità di permettere la verifica di ipotesi sull'uguaglianza di mediane e sono state proposte alcune modifiche, che tuttavia sono ancora poco utilizzate.

Il test di **Kruskal-Wallis**, fondato sui ranghi, è analogo all'analisi della varianza ad un criterio di classificazione, come il test di **Wilcoxon-Mann-Whitney**, fondato ugualmente sui ranghi, è analogo al test  $t$  di Student. Lo stesso confronto vale per la sua efficienza asintotica relativa.

L'**efficienza asintotica relativa** del test **KW** rispetto al test **F** è quindi identica a quelle del test **WMW** rispetto al **t**:

- quando la distribuzione dei dati è Normale ha un valore uguale a  $0,95 (3/\pi)$ ,
- quando la distribuzione dei dati è Rettangolare ha un valore uguale a 1,
- quando la distribuzione dei dati è Esponenziale Doppia ha un valore uguale a  $1,50 (3/2)$ .

**ESEMPIO1.** L'ozono si forma da  $O_2$  in presenza di  $NO_2$  e di radiazione solare. A concentrazioni elevate, causa congestione polmonare; il limite di accettabilità in Italia è fissato dalla legge in  $200 \text{ gm}^{-3}$  (0,1 ppm).

Durante una giornata estiva, in quattro zone di una città (A, B, C, D) si sono rilevate le concentrazioni di  $O_3$ .

A	B	C	D
150	120	200	195
140	115	190	210
145	30	185	220
160	155	180	205
165	130	--	175
170	--	--	430
125	--	--	--

Si vuole verificare se esiste una differenza significativa tra le mediane della concentrazione di  $O_3$  nelle quattro zone.

Risposta.

E' noto che i valori di concentrazione di una sostanza nell'aria sovente hanno valori anomali, a causa delle correnti e della disposizione delle fonti. Con pochi dati e in una ricerca nuova, sono ignote le caratteristiche statistiche della popolazione da cui sono estratti i dati campionari.

Nell'esempio riportato, anche la semplice lettura o rappresentazione grafica dei dati sono in grado di evidenziare la non-normalità dei dati di alcune zone e la loro non omoscedasticità. Nel gruppo D, la presenza del valore 430 determina una varianza sensibilmente maggiore ed una distribuzione lontana dalla normalità (come sarebbe necessario dimostrare con test adeguati). Non è quindi possibile applicare l'analisi della varianza parametrica, ma si impone il ricorso al test di Kruskal-Wallis.

I valori devono essere sostituiti dal loro rango, calcolato su tutte le osservazioni dei **k** gruppi a confronto.

Da essi, si calcola la somma dei ranghi (**R<sub>i</sub>**) ed il numero di osservazioni (**n<sub>i</sub>**) di ogni gruppo o campione.

	A	B	C	D
	8	3	18	17
	6	2	16	20
	7	1	15	21
	10	9	14	19
	11	5	--	13
	12	--	--	22
	4	--	--	--
<b>R<sub>i</sub></b>	<b>58</b>	<b>20</b>	<b>63</b>	<b>112</b>
<b>n<sub>i</sub></b>	<b>7</b>	<b>5</b>	<b>4</b>	<b>6</b>

Con **N = 22** e **k = 4** si ottiene un valore di **g**

$$g = \frac{12}{22 \cdot 23} \cdot \left( \frac{58^2}{7} + \frac{20^2}{5} + \frac{63^2}{4} + \frac{112^2}{6} \right) - 3 \cdot 23$$

$$g = 0,0237 \cdot (480,6 + 80 + 992,3 + 2090,7) - 69 = 17,35$$

uguale a 17,35.

La tabella dei valori critici con **3 gdl** riporta

- 7,82 alla probabilità  $\alpha = 0.05$ ,
- 11,34 alla probabilità  $\alpha = 0.01$ ,
- 16,27 alla probabilità  $\alpha = 0.001$ .

Pertanto, si può rifiutare l'ipotesi nulla, con una probabilità di commettere un errore di I° tipo inferiore a 0.001.

Per la sua applicazione corretta, **il test di Kruskal-Wallis richiede che la misura utilizzata sia continua**. Di conseguenza, non si dovrebbero avere valori identici; ma nella pratica sperimentale, per l'approssimazione della scala o dello strumento può succedere che alcune siano uguali. In questo caso, **con valori identici che occupano lo stesso rango la varianza campionaria è ridotta e diviene opportuno correggere il valore di g**.

**La correzione per misure ripetute (ties)** aumenta il valore di **g**; quindi incrementa la probabilità di trovare differenze significative tra le mediane dei gruppi a confronto. Tuttavia l'effetto della correzione è quasi sempre trascurabile, quando le misure identiche sono meno di un quarto delle osservazioni e sono distribuite tra più ranghi.

Per ottenere il valore di g' corretto, si deve dividere la quantità **g** calcolata per un fattore di correzione **C**

$$C = 1 - \frac{\sum_{i=1}^p c_i (c_i^2 - 1)}{N \cdot (N^2 - 1)}$$

dove:

- p** è il numero di raggruppamenti con ranghi ripetuti,
- c** è il numero di ranghi ripetuti nel raggruppamento i-esimo,
- N** è il numero totale di osservazioni nei **k** campioni a confronto.

**Un altro metodo di correzione per i ties (che apporta variazioni maggiori sul risultato) è spiegato nella parte finale del successivo paragrafo 6**

ESEMPIO 2. In una ricerca sulla qualità della vita, in tre quartieri (X, Y, Z) della stessa città sono stati ottenuti i punteggi di seguito riportati, con la medesima impostazione tabellare di un'analisi della varianza ad 1 criterio di classificazione.

X	Y	Z
7	4	2
7	4	2
6	4	1
5	7	3
8	5	--

Esistono differenze significative tra le loro mediane?

Risposta.

Si devono sostituire i punteggi con i ranghi relativi e calcolare le somme, come nella tabella sottostante:

	X	Y	Z
	12	6	2,5
	12	6	2,5
	10	6	1
	8,5	12	4
	14	8,5	--
<b>R<sub>i</sub></b>	<b>56,5</b>	<b>38,5</b>	<b>10</b>
<b>n<sub>i</sub></b>	<b>5</b>	<b>5</b>	<b>4</b>

Da esse, con la formula

$$g = \left( \frac{12}{N(N+1)} \sum_i^k \frac{R_i^2}{n_i} \right) - 3(N+1)$$

si stima il valore di **g**

$$= \left[ \frac{12}{14 \cdot (14+1)} \cdot \left( \frac{56,5^2}{5} + \frac{38,5^2}{5} + \frac{10^2}{4} \right) \right] - 3 \cdot (14+1)$$

$$= \left[ \frac{12}{210} \cdot (638,45 + 296,45 + 25) \right] - 45 = (0,05714 \times 959,9) - 45 = 9,848$$

che risulta uguale a **9,848**.

Per 3 campioni di dimensioni 5, 5, 4 alla probabilità  $\alpha = 0.001$  il valore critico riportato nella tabella è uguale a 9,32. Di conseguenza, con probabilità inferiore a 0.001 si rifiuta l'ipotesi nulla: esiste una differenza significativa tra le 3 mediane a confronto.

I dati presentano un numero elevato di ripetizioni:

- il valore **2** e il valore **5** compaiono 2 volte;
- il valore **4** e il valore **7** compaiono 3 volte.

Il fattore di correzione C

$$C = 1 - \frac{\sum_{i=1}^g c_i (c_i^2 - 1)}{N \cdot (N^2 - 1)}$$

con i dati dell'esempio risulta

$$1 - \frac{[2 \cdot (2^2 - 1) + 2 \cdot (2^2 - 1) + 3 \cdot (3^2 - 1) + 3 \cdot (3^2 - 1)]}{14 \cdot (14^2 - 1)} = 1 - \frac{2 \cdot 3 + 2 \cdot 3 + 3 \cdot 8 + 3 \cdot 8}{14 \cdot 195}$$

$$1 - \frac{60}{2730} = 1 - 0,02198 = 0,97802$$

uguale a 0,97802

ed il valore corretto di **g (g')**

$$g' = 9,848 / 0,97802 = 10,069$$

diviene **10,069 rispetto al 9,848** precedente.

Anche questo esempio dimostra che **il fattore di correzione aumenta il valore di g** ma per entità trascurabili. **Con questi dati la correzione non era necessaria, poiché il valore stimato era già significativo**; tuttavia è stato applicato, per fornire una stima corretta di **g**.

Come in tutte le correzioni per i *ties*, nell'indice di correzione **un solo valore ripetuto molte volte ha un peso relativo maggiore di molti valori ripetuti poche volte**.

Come semplice dimostrazione si può stimare che, se nelle 14 osservazioni vi fosse stato un solo valore ripetuto 5 volte, il fattore di correzione C sarebbe stato

$$C = 1 - \frac{5 \cdot (5^2 - 1)}{14 \cdot (14^2 - 1)} = 1 - \frac{5 \cdot 24}{14 \cdot 195} = 1 - \frac{120}{2730} = 1 - 0,04396 = 0,95604$$

uguale a 0,95604

ed il valore corretto di **g**

$$g' = 9,848 / 0,95604 = 10,301$$

uguale a **10,301** fornendo uno scarto superiore al precedente.

L'esempio è solo teorico, in quanto **con tanti dati identici la validità del test è molto dubbia.**

### 11.5. CONFRONTI MULTIPLI NELL'ANALISI DELLA VARIANZA PER RANGHI, CON K CAMPIONI INDIPENDENTI

Secondo alcuni autori di testi di Statistica non parametrica (come esempio vedi: G. Landenna e D. Marasini, *Metodi statistici non parametrici*, edito da il Mulino, Bologna, nel 1990, che a pag. 234 scrive) *un diverso modo per verificare l'ipotesi nulla  $H_0: me_A = me_B = \dots = me_K$ , in alternativa all'ipotesi  $H_1$  in cui non tutte le mediane sono uguali, è quello di ricorrere ai confronti multipli che, nel caso di rifiuto di  $H_0$ , consentono anche la identificazione delle popolazioni con mediane diverse ovvero dei trattamenti i cui effetti hanno provocato il rifiuto medesimo.*

Secondo la maggioranza degli autori questo concetto non è espresso con la dovuta chiarezza.

Anche nella statistica non parametrica per **k** campioni, solo quando

- **prima** è stata rifiuta l'ipotesi nulla con il test di Kruskal-Wallis (più avanti si vedrà anche il test di Jonckheere),
- **successivamente** è possibile ricorrere ai **confronti multipli, per individuare quali sono i gruppi che hanno una tendenza centrale tra loro significativamente differente.**

Il motivo è che le due risposte potrebbero non coincidere, date le differenze dei metodi, anche se rispetto ai test parametrici quelli non parametrici offrono il vantaggio di avere la varianza definita dal numero di dati, non dalle variazioni campionarie.

Con il test di Kruskal-Wallis si potrebbe non rifiutare l'ipotesi nulla, mentre i confronti multipli potrebbero evidenziare almeno una differenza significativa. Per evitare tali contraddizioni logiche, è **prassi ormai unanime accettare i confronti multipli solo dopo il rifiuto dell'ipotesi nulla con il test di Kruskal-Wallis.**

Una **procedura molto lunga, possibile solo con programmi informatici**, ma concettualmente molto semplice, è fondata sugli stessi principi del test di **casualizzazione (permutation test)** già illustrato sia per due campioni dipendenti che indipendenti.

Con **N** dati suddivisi in **k** gruppi, è teoricamente facile

- analizzare tutte le possibili risposte, determinate dalle diverse collocazioni dei valori nei **k** gruppi;

- stabilita la probabilità  $\alpha$ , sono collocati nella zona di rifiuto dell'ipotesi nulla le distribuzioni dei ranghi che danno le differenze massime tra i ranghi.

Metodi operativi concettualmente più complessi, ma operativamente più rapidi, sono quelli analoghi alle procedure presentate nei confronti multipli.

Secondo la logica già esposta nella presentazione del **t di Bonferroni**, il valore della probabilità deve essere diviso per il numero di confronti possibili. Ad esempio, scegliendo la probabilità complessiva di  $\alpha = 0.05$ , per effettuare tutti i confronti tra  $k$  gruppi, occorre utilizzare il valore di  $z$  (o di  $t$  quando i campioni sono molto piccoli) corrispondente ad un valore di  $\alpha$  uguale a

$$\alpha = 0.05 / [k \cdot (k - 1)]$$

Per evitare di effettuare tutti i possibili confronti tra coppie di medie con  $k$  gruppi (pari a  $C_k^2$ ), con una procedura analoga a quella del test parametrico **T** di Tukey, si calcolano le differenze in valore assoluto  $|\bar{r}_a - \bar{r}_b|$  tra le medie dei ranghi di tutti i gruppi.

Di esse sono significative, alla probabilità  $\alpha$ , tutte quelle che sono uguali o maggiori della quantità **D**, data da

$$D = z_{\alpha/c} \cdot \sqrt{\frac{N \cdot (N + 1)}{12} \cdot \left(\frac{1}{n_a} + \frac{1}{n_b}\right)}$$

dove

**N** è il numero complessivo di dati considerando tutti i gruppi,

**n<sub>a</sub>** e **n<sub>b</sub>** sono il numero di dati nei due gruppi (chiamati a e b) a confronto,

**c** è il numero di possibili confronti, che con  $k$  gruppi è uguale a  $\frac{k \cdot (k - 1)}{2}$ ,

**z** alla probabilità  $\alpha/c$  è tratta dalla distribuzione normale.

Se i gruppi a confronto hanno tutti lo stesso numero d'osservazioni, si calcola un solo valore, chiamato **differenza minima significativa (least significant difference)**. Sono significative tutte le differenze tra coppie di medie di ranghi che risultano superiori alla quantità **D** calcolata.

ESEMPIO. In alcuni tratti di 5 corsi d'acqua è stata misurata la quantità di tensioattivi anionici (misurata in  $\text{mg l}^{-1}$ ) presenti

A	B	C	D	E
0,50	0,58	0,43	0,90	1,06
0,86	0,42	0,70	1,13	1,98
0,90	0,62	0,75	0,80	1,42
0,23	0,60	0,58	0,95	1,48
0,55	0,48	0,89	0,82	0,85
0,75	0,60	0,75	0,80	0,90
--	--	0,75	0,60	1,90

Dopo aver verificato se esistono differenze significative tra le mediane dei diversi corsi d'acqua, in caso positivo, individuare tra quali corsi tali differenze sono significative.

Risposta.

I primi passi dell'analisi sono:

- trasformare i valori nei loro ranghi,
- annotare quanti sono i valori identici e quante le loro repliche.

	A	B	C	D	E
	5	7,5	3	25	28
	22	2	13	29	33
	25	12	15,5	18,5	30
	1	10	7,5	27	31
	6	4	23	20	21
	15,5	10	15,5	18,5	25
	--	--	15,5	10	32
<b>R<sub>i</sub></b>	74,5	45,5	93,0	148,0	200
<b>n<sub>i</sub></b>	6	6	7	7	7
<b><math>\bar{r}_i</math></b>	12,42	7,58	13,29	21,14	28,57

E' inoltre utile ricordare che, quando sono presenti valori identici, il rango da attribuire è il loro valore medio.

Con un numero elevato di gruppi e di osservazioni entro ogni gruppo è facile incorrere in errori nell'attribuzione dei ranghi. **Una verifica rapida dei calcoli** è data dalla corrispondenza tra somma totale dei ranghi e numero totale di osservazioni: tra essi esiste l'uguaglianza

$$\text{Somma di tutti i ranghi} = \frac{N \cdot (N + 1)}{2}$$

Nell'esempio, l'operazione di attribuzione dei ranghi e le loro somme per gruppo sono state effettuate in modo corretto, poiché,

$$(74,5 + 45,5 + 93 + 148 + 200) = 561 = \frac{33 \cdot 34}{2}$$

Per il calcolo di  $g$  è utile ricordare che  $k = 5$  e  $N = 33$ ; pertanto si ottiene un valore

$$g = \frac{12}{33 \cdot 34} \cdot \left( \frac{74,5^2}{6} + \frac{49,5^2}{6} + \frac{93^2}{7} + \frac{148^2}{7} + \frac{200^2}{7} \right) - 3 \cdot 34$$

$$g = \frac{12}{1122} \cdot (925,0 + 345,0 + 1235,6 + 3129,1 + 5714,3) - 102$$

$$g = 0,0107 \cdot 11349 - 102 = 21,704$$

di  $g$  uguale a 21,704.

Con 5 gruppi, la significatività è fornita dalla tabella sinottica dei valori critici del  $\chi^2$  per 4 gdl:

- alla probabilità  $\alpha = 0.05$  è uguale a 9,49,
- alla probabilità  $\alpha = 0.01$  è uguale a 13,28,
- alla probabilità  $\alpha = 0.001$  è uguale a 20,52.

A causa della presenza di misure ripetute, può rivelarsi vantaggioso **correggere il valore di  $g$**  (non in questo caso, in quanto già significativo ad una probabilità inferiore a 0.001).

Con i dati dell'esempio, poiché sono presenti i seguenti valori identici riportati nella tabella

per	2	volte compare il valore	0,58	che ha rango medio	7,5
"	3	"	0,60	"	10
"	4	"	0,75	"	15,5
"	2	"	0,80	"	18,5
"	3	"	0,90	"	25

il fattore di correzione C è

$$C = 1 - \frac{2 \cdot (2^2 - 1) + 3 \cdot (3^2 - 1) + 4 \cdot (4^2 - 1) + 2 \cdot (2^2 - 1) + 3 \cdot (3^2 - 1)}{33 \cdot (33^2 + 1)}$$

$$C = 1 - \frac{2 \cdot 3 + 3 \cdot 8 + 4 \cdot 15 + 2 \cdot 3 + 3 \cdot 8}{33 \cdot 1088} = 0,996658$$

uguale a 0,996658.

Il **valore corretto di g** diventa

$$\frac{21,704}{0,996658} = 21,777$$

uguale a 21,777 che, ovviamente, è ancor più significativo del valore stimato in precedenza.

**Rifiutata l'ipotesi nulla** (tutti i gruppi o campioni siano estratti dalla stessa popolazione o da popolazioni con la medesima mediana) **e quindi accettata l'ipotesi alternativa** (non tutte le mediane dei gruppi a confronto sono uguali), si può mettere in evidenza quali sono le medie dei ranghi che hanno una differenza significativa.

Per confronti semplici tra tutte le medie dei ranghi, con 5 campioni il numero di differenze può essere stimato con

$$C_2^5 = \frac{5!}{(5-2)!2!} = 10 \quad \text{oppure} \quad \frac{k \cdot (k-1)}{2} = \frac{5 \cdot 4}{2} = 10$$

e risulta uguale a 10.

Esse possono essere riportate in valore assoluto in una matrice triangolare con il relativo numero di osservazioni

		A(6)	B(6)	C(7)	D(7)	E(7)
		12,42	7,58	13,29	21,14	28,57
A	12,42					
B	7,58	<b>4,84</b>				
C	13,29	<b>0,87</b>	<b>5,71</b>			
D	21,14	<b>8,72</b>	<b>13,56</b>	<b>7,85</b>		
E	28,57	<b>16,15*</b>	<b>20,99*</b>	<b>15,28*</b>	<b>7,43</b>	

Alla probabilità complessiva  $\alpha_T = 0.05$  per 10 confronti simultanei, la probabilità  $\alpha$  di ogni confronto è uguale a 0.005; per un test bilaterale (quindi alla probabilità  $\alpha = 0.0025$  in una coda della distribuzione) sulla tavola della distribuzione normale ad essa corrisponde un valore di  $z$  uguale a 2,81 (esattamente 2,807).

Per un test a due code, la differenza minima significativa  $D$  (con  $N = 33$ ) per il confronto tra A e B che hanno entrambe 6 osservazioni

$$D = 2,807 \cdot \sqrt{\frac{33 \cdot 34}{12} \cdot \left(\frac{1}{6} + \frac{1}{6}\right)} = 2,807 \cdot \sqrt{93,5 \cdot 0,33} = 15,66$$

è uguale a 15,66.

Per i 6 confronti tra le medie dei ranghi dei campioni A e B, che hanno 6 osservazioni, con i campioni C, D e E che hanno 7 osservazioni,

$$D = 2,807 \cdot \sqrt{\frac{33 \cdot 34}{12} \cdot \left(\frac{1}{6} + \frac{1}{7}\right)} = 2,807 \cdot \sqrt{93,5 \cdot 0,309} = 15,10$$

è uguale a 15,10.

Per i 3 confronti tra i gruppi C, D e E che hanno 7 osservazioni.

$$D = 2,807 \cdot \sqrt{\frac{33 \cdot 34}{12} \cdot \left(\frac{1}{7} + \frac{1}{7}\right)} = 2,807 \cdot \sqrt{93,5 \cdot 0,286} = 14,51$$

ed è uguale a 14,51.

Confrontando i valori di  $D$  stimati con le differenze riportate nella tabella triangolare precedente, **alla probabilità complessiva  $\alpha_T = 0.05$  risultano significative solamente 3 delle 10 differenze calcolate tra i cinque gruppi** (i valori in corsivo e con un asterisco):

- la differenza tra la media dei ranghi del gruppo A e quella del gruppo E,
- la differenza tra la media dei ranghi del gruppo B e quella del gruppo E,
- la differenza tra la media dei ranghi del gruppo C e quella del gruppo E.

Osservando i tre diversi intervalli calcolati (15,66 - 15,10 - 14,51) è didatticamente utile sottolineare come l'intervallo diminuisca in modo non trascurabile anche all'aumento di una sola osservazione, quando i campioni sono numericamente così ridotti.

Con lo stesso numero di gruppi, **l'intervallo minimo significativo varia anche in rapporto al numero di confronti che si vogliono fare.**

Se, in modo analogo al test di Scheffé, si intende effettuare oltre ai confronti semplici anche **confronti complessi tra le medie combinate di alcuni gruppi**, il valore di  $z$  cambia (cresce in rapporto all'aumentare del numero totale di confronti possibili, come evidenziato dalla tabella relativa).

Ad esempio, se si intendono fare 15 confronti mantenendo costante la probabilità complessiva di  $\alpha_T = 0.05$ , la probabilità  $\alpha$  di ogni confronto diventa  $(0.05/15)$  uguale a 0.0033 per un test unilaterale e 0.00167 per un test bilaterale, al quale corrisponde un valore di  $z$  uguale a 2,94.

Nel caso di **confronti tra alcuni trattamenti ed un controllo** (per esempio, se si intendesse confrontare le medie dei trattamenti B, C, D, E solamente con la media del campione A, ritenuto la situazione normale o standard) il numero di confronti si riduce notevolmente; è pari a  $k-1$ , quando  $k$  è il numero totale di gruppi.

Il metodo è utilizzato in medicina quando si intendono valutare gli effetti di alcuni farmaci rispetto al solo placebo; è frequente nella ricerca tossicologica ed ambientale, per confrontare più situazioni a rischio con quella normale.

Con 5 gruppi o campioni, di cui 1 è il placebo o controllo e 4 sono i farmaci o le situazioni da verificare, il numero di confronti è 4.

Dopo aver applicato l'analisi della varianza di Kruskal-Wallis per dimostrare l'esistenza di almeno una eventuale differenza tra i 5 gruppi, è possibile verificare la significatività di ognuna delle differenze tra le mediane dei  $k-1$  trattamenti rispetto alla mediana o tendenza centrale del controllo.

**Il procedimento** è identico al precedente, con 2 sole variazioni.

1 - Alla medesima probabilità di rifiutare l'ipotesi nulla, il valore di  $z$  è minore, perché inferiore è il numero di possibili confronti; di conseguenza, sarà minore anche il valore della differenza minima significativa, a parità del numero di osservazioni impiegate nei rispettivi gruppi.

Per esempio alla probabilità complessiva  $\alpha_T = 0.05$  la probabilità  $\alpha$  di ogni confronto diviene:

- $0.05/4$  uguale a 0.0125 ed il valore di  $z$  corrispondente uguale a 2,24 per un test ad una coda,
- $0.025/4$  uguale a 0.00625 ed il valore di  $z$  corrispondente uguale a 2,50 per un test bilaterale.

2 - Con lo stesso numero  $N$  di osservazioni totali, è possibile ottenere una maggiore efficienza - potenza dei 4 confronti, non programmando gruppi con lo stesso numero di dati ma ponendo un maggiore numero di osservazioni nel gruppo o campione di controllo.

Infatti, il suo numero di osservazioni viene utilizzato nei calcoli di tutte le  $k-1$  differenze minime significative.

Ad esempio, con 30 osservazioni e 5 gruppi invece di attribuire 6 dati per gruppo risulta più conveniente assegnare 5 osservazioni ai 4 trattamenti e 10 osservazioni al controllo.

I 4 confronti saranno tutti tra due gruppi rispettivamente di 10 (il controllo) e 5 (un trattamento) osservazioni; essi risultano più potenti di 4 confronti effettuati sempre tra due gruppi di 6 osservazioni (sia il controllo che un trattamento).

### 11.6. TEST PER L'ETEROGENEITA' DELLA VARIANZA CON K CAMPIONI

Nella ricerca biologica, ecologica ed ambientale, sono frequenti le situazioni in cui l'attenzione del ricercatore è rivolta alla **variabilità dei dati**, più che alla loro tendenza centrale. E' il caso di misure d'inquinamento che in zone differenti possono avere una variabilità diversa, pure con una tendenza centrale simile. Anche se le mediane sono tutte sotto i limiti di legge, dove la varianza risulta maggiore è più urgente intervenire, poiché singole osservazioni possono superarli con frequenza più alta. Inoltre, **se k serie di dati hanno varianze non uguali, appartengono a popolazioni differenti.**

E' un problema già discusso nel caso di 2 campioni indipendenti, che può essere **facilmente esteso a k campioni, come il test WMW** (Wilcoxon-Mann-Whitney) **ha la sua generalizzazione nel test di KW** (Kruskal-Wallis).

Prendendo come esempio un articolo pubblicato su *Applied Statistics* nel 1989 (di D. V. **Hinkley**, *Modified profile likelihood in trasformed linear models*, vol. 38, pp. 495-506), la procedura presentata da **P. Sprent** nel volume *Applied nonparametric statistical methods*, (second Edition, Chapman & Hall, London, 1993, pp. 155-156) prevede che

1 - per la verifica dell'ipotesi nulla

$$H_0: s^2_A = s^2_B = \dots = s^2_K$$

contro l'ipotesi alternativa

$$H_1: \text{le } s^2 \text{ dei gruppi a confronto non sono tutte uguali}$$

in un caso **con pochi dati** (meno di quelli richiesti dal metodo di Moses per 2 campioni indipendenti), situazione non rara nella ricerca ambientale,

A	2	8	8	4
B	8	7	14	--
C	33	59	48	56
D	60	101	67	--

2 - dopo il **calcolo delle medie** ( $\bar{x}_i$ ) **di ogni gruppo**

Gruppo	X <sub>ij</sub>				medie $\bar{x}_i$
A	2	8	8	4	<b>5,50</b>
B	8	7	14	--	<b>9,67</b>
C	33	59	48	56	<b>49,00</b>
D	60	101	67	--	<b>76,00</b>

3 - si debbano stimare **le deviazioni, in valore assoluto**, di ogni dato dalla media del suo gruppo

$$|x_{ij} - \bar{x}_i|$$

ottenendo una nuova tabella di dati trasformati come la seguente

	$ x_{ij} - \bar{x}_i $			
A	3,50	2,50	2,50	1,50
B	1,67	2,67	4,33	--
C	16,00	10,00	1,00	7,00
D	16,00	25,00	9,00	--

4 - che deve essere ulteriormente **modificata nei ranghi** relativi, considerando tutto il campione

A	<b>7</b>	<b>4,5</b>	<b>4,5</b>	<b>2</b>
B	<b>3</b>	<b>6</b>	<b>8</b>	--
C	<b>12,5</b>	<b>11</b>	<b>1</b>	<b>9</b>
D	<b>12,5</b>	<b>14</b>	<b>10</b>	--

5 - se **l'ipotesi nulla è vera** (variabilità uguale in ogni gruppo), i ranghi di ognuno di essi dovrebbero essere distribuiti casualmente e quindi avere medie uguali, sia tra loro, sia alla media generale.

Se **l'ipotesi nulla è falsa**, almeno un gruppo dovrebbe avere la media dei ranghi significativamente diversa dagli altri

E' la stessa condizione (sulle medie dei ranghi e mediane dei valori) verificata dal test di **Kruskal-Wallis** (che può essere applicato sui ranghi dell'ultima tabella).

6 - Per giungere alla stima di **g** con la formula abbreviata

$$g = \left( \frac{12}{N(N+1)} \sum_i^k \frac{R_i^2}{n_i} \right) - 3(N+1)$$

dapprima si calcolano i totali (**R<sub>i</sub>**) dei ranghi e il numero di osservazioni (**n<sub>i</sub>**) entro ogni gruppo

	A	B	C	D
<b>R<sub>i</sub></b>	<b>18</b>	<b>17</b>	<b>33,5</b>	<b>36,5</b>
<b>n<sub>i</sub></b>	<b>4</b>	<b>3</b>	<b>4</b>	<b>3</b>

e con  $N = 14$

$$g = \left[ \frac{12}{14 \cdot (14+1)} \cdot \left( \frac{18^2}{4} + \frac{17^2}{3} + \frac{33,5^2}{4} + \frac{36,5^2}{3} \right) \right] - 3 \cdot (14+1) =$$

$$= \left[ \frac{12}{210} \cdot (81,00 + 96,33 + 280,56 + 444,08) \right] - 45 = (0,05714 \times 901,97) - 45 = 6,54$$

si ottiene un valore di **g** uguale a 6,54 .

Il risultato deve essere confrontato con i valori critici del  $c^2$ .

Per **k** = 4 gruppi, i **gdl** sono 3; il valore critico alla probabilità  $\alpha = 0.05$  è uguale a 7,815. Non è possibile rifiutare l'ipotesi nulla: i vari gruppi non hanno una variabilità significativamente differente.

**Poiché esistono valori identici, è possibile apportare la correzione relativa.**

Ricorrendo alla formula già illustrata

$$C = 1 - \frac{\sum_{i=1}^p c_i (c_i^2 - 1)}{N \cdot (N^2 - 1)}$$

dove:

**p** è il numero di raggruppamenti con ranghi ripetuti,  
**c** è il numero di ranghi ripetuti nel raggruppamento *i*-esimo,  
**N** è il numero totale di osservazioni nei **k** campioni a confronto.

Con **p** = 2 e **c** = 2 si ottiene un termine di correzione **C**

$$C = 1 - \frac{(2 \times 3) + (2 \times 3)}{14 \times 195} = 1 - 0,0044 = 0,9956$$

uguale a 0,9956 che non modifica sostanzialmente il valore di **g corretto**

$$g \text{ corretto} = 6,54 / 0,9956 = 6,57$$

risultando uguale a 6,57 con arrotondamento, rispetto al 6,54 precedente.

Poiché il campione è piccolo ed esistono valori identici, come pubblicato anche da Hinkley, è conveniente **usare un'altra correzione** che, scritta nei suoi passaggi operativi, è

$$g = \frac{(N-1) \cdot (Sp - C)}{Sr - C}$$

dove

$$Sp = \sum_{i=1}^k \left( \frac{s_i^2}{n_i} \right)$$

con  $S_i = \sum r_{ik}^2$  (sommatoria del quadrato dei ranghi dei **k** gruppi) e

**n<sub>i</sub>** = numero di dati di un gruppo

$$Sr = \sum_{i,k} r_{ik}^4$$

(sommatoria di tutti i ranghi elevati alla quarta)

$$C = \sum_{i=1}^k s_i^2 / N$$

Con i dati dell'esempio,

$$s_1 = 7^2 + 4,5^2 + 4,5^2 + 2^2 = 93,5$$

$$s_2 = 3^2 + 6^2 + 8^2 = 109,0$$

$$s_3 = 12,5^2 + 11^2 + 1^2 + 9^2 = 359,25$$

$$s_4 = 12,5^2 + 14^2 + 10^2 = 452,25$$

con

$$s_1 + s_2 + s_3 + s_4 = N \times (N+1) \times (2N+1) / 6 \quad \text{solo quando non esistono valori identici}$$

infatti  $93,5 + 109,0 + 359,25 + 452,25 = 1014$

mentre  $14 \times 15 \times 29 / 6 = 1015$

Dai valori  $s_1, s_2, s_3, s_4$  si ottiene  $S_p$  mediante

$$S_p = (93,5^2 / 4) + (109^2 / 3) + (359,25^2 / 4) + (452,25^2 / 3) = 106.587,724$$

che risulta uguale a 106.587,724

mentre  $S_r$  è dato dalla somma di tutti i 14 ranghi alla quarta

$$S_r = 7^4 + 4,5^4 + 4,5^4 + 2^4 + 12,5^4 + 14^4 + 10^4 = 127.157,25$$

e risulta uguale a 127.157,25

e  $C$  è dato dal quadrato della somma dei  $k$   $s_i$  diviso  $N$

$$C = 1014^2 / 14 = 73.442,571$$

e risulta uguale a 73.442,571.

Da essi si stima il valore corretto di  $g$

$$g = \frac{(N-1) \cdot (S_p - C)}{S_r - C} = \frac{(14-1) \cdot (106.587,724 - 73.442,571)}{127.157,25 - 73.442,571} = \frac{430.886,989}{53.714,679} = 8,0217$$

che risulta uguale a 8,0217.

Il nuovo valore risulta significativo alla probabilità  $\alpha = 0.05$  e rovescia le conclusioni precedenti, evidenziando una differenza significativa nella variabilità dei 4 gruppi di dati.

## 11.7. IL TEST Q DI COCHRAN

Con un esperimento organizzato come i blocchi randomizzati già discussi nell'analisi della varianza a due criteri di classificazione, quando le **risposte sono di tipo binario** nella statistica non parametrica è possibile ricorrere al test **Q** di **W. G. Cochran** presentato nel 1954 su *Biometrika* (*The comparisons of percentages in matched samples*, vol. 37, pp. 256-266), chiamato anche analisi della varianza a due vie per ranghi (**Two-Way ANOVA by Ranks**).

Il test serve per **verificare se il numero, la proporzione o la frequenza totale di successi o insuccessi di N prove ripetute** (da cui anche il nome di analisi della varianza a misure ripetute) **differiscono in modo significativo tra le varie situazioni a confronto**. In vari testi è presentato come l'estensione a  $k$  campioni del test di McNemar per 2 campioni dipendenti.

Il test **Q di Cochran** è utilizzato quando si devono valutare i risultati di sondaggi o giudizi qualitativi binari (sufficiente o insufficiente; accettabile o non-accettabile; migliorato o non migliorato) espressi dalle stesse persone su una serie di tre o più provvedimenti o situazioni.

Le valutazioni devono essere rappresentate mediante una votazione binaria, indicata solo mediante 0 e 1.

Per esempio, è possibile chiedere ad un gruppo di **N** persone se la situazione ambientale di **K** quartieri è ritenuta accettabile o no, se a loro parere nell'ultimo anno la qualità della vita sia migliorata oppure no, se gli interventi dell'amministrazione sono ritenuti sufficienti o insufficienti, indicando la risposta per ogni quartiere.

I risultati devono essere riportati in una tabella a due entrate come quella sottostante

INDIVIDUI	QUARTIERI			
	A	B	---	K
1	0	0	---	0
2	0	1	---	1
3	1	1	---	1
---	---	---	---	---
N	1	0	---	1

in cui, di norma,

- sulle righe sono indicati i valori attribuiti da ogni individuo e
- nelle colonne sono riportate le varie situazioni (in questo esempio i quartieri).

Il test serve per verificare se le valutazioni assegnate alle varie colonne sono simili o statisticamente differenti: se le percentuali o proporzioni di 1 (oppure di 0) calcolate per colonna sono uguali (**H<sub>0</sub>**) oppure significativamente differenti (**H<sub>1</sub>**).

Con **r righe e c colonne**, la formula di **Cochran**

$$Q = \frac{(k-1) \cdot (k \cdot \sum C_j^2 - (\sum C_j)^2)}{k \cdot \sum R_i - \sum R_i^2}$$

dove,

**k** è il numero di colonne,

**C<sub>j</sub>** è il numero totale di successi nella colonna j-esima,

**R<sub>i</sub>** è il numero totale di successi nelle riga i-esima,

determina il valore **Q**.

Quando il numero di righe non è troppo piccolo, **Q** segue

la **distribuzione  $\chi^2$  con gdl k-1**.

Per l'uso dei valori critici riportati nella tabella del chi quadrato, **il campione è ritenuto di dimensioni accettabili quando il numero totale di osservazioni (N individui per k situazioni) è complessivamente uguale o maggiore di 24 e contemporaneamente il numero di righe (N) non è inferiore a 4**.

Per esperimenti di dimensioni minori, non presentati in questa trattazione, possono essere ottenute stime esatte dei valori critici mediante le permutazioni.

E' tuttavia diffusa la prassi di **utilizzare comunque la distribuzione chi quadrato**, poiché i calcoli con le permutazioni richiedono molto tempo e in conclusione determinano valori approssimativamente simili a quelli del chi quadrato.

Come potrà essere messo in evidenza con l'esempio, le righe composte solo da valori identici, siano essi tutti 0 oppure tutti 1 (come le risposte dell'individuo 1 e dell'individuo 3 nella tabella precedente) non influiscono sul valore dell'indice **Q**. Esso risente solo delle valutazioni differenti, che variano il numero di risposte positive tra colonne.

Di conseguenza, **per calcolare le dimensioni reali di una tabella si devono considerare solo le righe con valori che non siano tutti 1 oppure tutti 0**.

Anche in questo caso, **la metodologia è spiegata in modo semplice seguendo la soluzione di un problema**.

**ESEMPIO.** In una città con gravi problemi di traffico è stata profondamente modificata la circolazione del trasporto privato, cambiando sensi unici e divieti, sia di accesso, sia di sosta. Anche se la situazione in complesso appare migliorata, gli abitanti di alcuni quartieri non si ritengono soddisfatti, ritenendo che nella loro zona la nuova situazione non presenti gli stessi vantaggi evidenti in altri, riguardo all'inquinamento acustico e a quello dell'aria, al controllo del traffico e alla qualità della vita in genere.

A 15 esperti di vari settori della qualità della vita, è stato chiesto un giudizio sulla situazione di 4 quartieri (I, II, III, IV).

Essi devono rispondere se, rispetto all'anno precedente, in ogni quartiere la situazione è migliorata (1) oppure no (0) (come prima o peggiorata).

.

La tabella riporta le 15 risposte, indicando con 1 quelle positive e con 0 quelle non positive (di parità oppure negative, poiché la risposta può essere solo binaria).

	Quartieri			
Esperti	I	II	III	IV
1	0	1	1	1
2	0	1	1	0
3	1	1	1	1
4	1	1	0	0
5	0	0	0	0
6	0	1	1	1
7	1	0	0	0
8	1	1	0	0
9	0	0	0	1
10	0	0	0	0
11	1	0	0	0
12	1	1	1	0
13	1	1	0	0
14	1	1	1	0
15	0	1	1	1
<b>C<sub>j</sub></b>	<b>8</b>	<b>10</b>	<b>7</b>	<b>5</b>

Si può sostenere che i 4 quartieri hanno avuto variazioni differenti, per cui in essi la qualità della vita è migliorata in modo significativamente diverso?

Risposta.

Per l'applicazione della formula di Cochran, occorre dapprima calcolare

- il totale di ogni colonna ( $C_j$ ),
- il totale di ogni riga ( $R_i$ ),
- e quello dei suoi quadrati ( $R_i^2$ ),

Esperti	Quartieri				$R_i$	$R_i^2$
	I	II	III	IV		
1	0	1	1	1	3	9
2	0	1	1	0	2	4
3	1	1	1	1	4	16
4	1	1	0	0	2	4
5	0	0	0	0	0	0
6	0	1	1	1	3	9
7	1	0	0	0	1	1
8	1	1	0	0	2	4
9	0	0	0	1	1	1
10	0	0	0	0	0	0
11	1	0	0	0	1	1
12	1	1	1	0	3	9
13	1	1	0	0	2	4
14	1	1	1	0	3	9
15	0	1	1	1	3	9
$C_j$	8	10	7	5	30	80

che risultano uguali a

$$\sum C_j = 30 \quad \sum R_i = 30 \quad \sum R_i^2 = 80$$

Applicando ai dati la formula di Cochran

$$Q = \frac{(4-1) \cdot (4 \cdot (8^2 + 10^2 + 7^2 + 5^2) - 30^2)}{4 \cdot 30 - 80} = \frac{3 \cdot (952 - 900)}{40} = \frac{156}{40} = 3,9$$

si ottiene un valore di  $Q$  uguale a 3,9.

La sua significatività deve essere verificata con i valori critici della distribuzione  $\chi^2$  con 3 gdl, che alla probabilità  $\alpha = 0.05$  fornisce il valore 7,82.

Il valore calcolato è inferiore non solo a quello critico della probabilità  $\alpha = 0.05$  ma anche a quello della probabilità  $\alpha = 0.10$ .

Non è dimostrato che nei 4 quartieri la situazione sia cambiata in modo significativamente diverso; l'alto valore di  $\alpha$  (superiore a 0.10) permette di concludere che la qualità della vita nei 4 quartieri è migliorata in modo relativamente uniforme.

**Quando il numero di colonne o risposte è uguale a 2, le N risposte in un test di Cochran, riportate nelle righe, possono essere riassunte in una tabella 2 x 2 e quindi analizzate come test di Mc Nemar, sempre attraverso la distribuzione chi quadrato.**

#### **11.8. ESTENSIONE DEL TEST DI McNEMAR A UNA TABELLA QUADRATA N X N:**

##### **IL TEST DI BOWKER**

Quando, **con due campioni dipendenti** (quindi molti testi lo collocano in tale capitolo), le risposte non sono binarie ma hanno tre o più modalità, si costruisce una tabella quadrata N x N per analizzare la simmetria ai lati della diagonale. E' il test proposto da A. H. **Bowker** nel 1948 (con l'articolo "A test for symmetry in contingency tables" pubblicato su *J. Amer. Statist. Assoc.*, 43, pp. 572-4) per verificare, come permette il test di Mc Nemar, l'ipotesi che nel confronto tra due situazioni sia avvenuto un cambiamento significativo nella frequenza delle varie modalità. Tra i testi a maggior diffusione, questo metodo è presentato da **P. Sprent** nel volume *Applied nonparametric statistical methods*, (second Edition, Chapman & Hall, London, 1993, pp. 239-240).

Nella ricerca ambientale, si ricorre al test di Bowker nel caso in cui si confrontano le risposte multiple (non binarie) delle stesse persone prima e dopo uno stimolo o i differenti effetti di due sostanze tossiche, somministrate allo stesso gruppo di cavie. E' un approccio diverso dal Q di Cochran, quando i livelli sono ordinali.

Si supponga che un'amministrazione comunale abbia deciso la chiusura del centro storico al traffico automobilistico.

Prima della delibera è stato fatto un sondaggio nominativo su un campione di individui residenti nella zona storica, deducendo da una serie di risposte se erano favorevoli, incerti o contrari. Dopo due mesi di applicazione del divieto, il test è stato ripetuto sugli stessi individui.

La sottostante tabella 3 x 3 riporta il numero di individui per ognuna delle 9 combinazioni di risposte.

		DOPO		
		Favorevoli	Incerti	Contrari
PRIMA	Favorevoli	85	16	35
	Incerti	28	34	25
	Contrari	54	39	74

La diagonale fornisce il numero di coloro che hanno mantenuto lo stesso parere, mentre le altre caselle riportano il numero di coloro che lo hanno cambiato. Si vuole sapere se esistono differenze statisticamente significative tra prima e dopo.

Una indicazione preliminare, solo descrittiva, può essere fornita dai totali marginali

		DOPO			TOTALI
		Favorevoli	Incerti	Contrari	
PRIMA	Favorevoli	85	<b>16 (a)</b>	<b>35 (b)</b>	136
	Incerti	<b>28 (a)</b>	34	<b>25 (c)</b>	87
	Contrari	<b>54 (b)</b>	<b>39 (c)</b>	74	167
TOTALI		167	89	134	390

dai quali risulta che i favorevoli sono aumentati da 136 a 167 e quelli incerti sono aumentati da 87 a 89 mentre i contrari sono diminuiti da 167 a 134.

Ma, trattandosi di due campioni dipendenti, come per il test di Mc Nemar **la risposta per l'inferenza è fornita da quelli che hanno cambiato parere** (in grassetto nella tabella).

**Il test presentato da Bowker nel 1948 propone di confrontare le frequenze distribuite in modo simmetrico rispetto alla diagonale: se  $H_0$  è vera, dovrebbero essere approssimativamente uguali.**

**I concetti ed il metodo sono semplici** e possono essere descritti in 4 passaggi logici.

1 - Se non è intervenuto nessun mutamento sistematico nelle opinioni espresse dagli intervistati ( $H_0$  vera),

- a) coloro che prima erano Incerti e dopo Favorevoli (28) dovrebbero essere equivalenti a quelli che prima erano Favorevoli e dopo Incerti (16);
- b) coloro che prima erano Contrari e dopo Favorevoli (54) dovrebbero essere equivalenti a quelli che prima erano Favorevoli e dopo sono Contrari (35);
- c) coloro che prima erano Contrari e dopo Incerti (39) dovrebbero essere equivalenti a quelli che prima erano Incerti e dopo Contrari (25).

In modo più formale, l'ipotesi nulla è

$$H_0: p_{ij} = p_{ji}$$

contro l'ipotesi alternativa

$$H_1: p_{ij} \neq p_{ji}$$

con esclusione dei valori della diagonale.

2 – Il confronto è effettuato mediante un  $\chi^2$  che considera la differenza tra la frequenza assoluta di una casella non collocata sulla diagonale e quella ad essa simmetrica, con una formula

$$c^2 = \sum \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}}$$

dove

**la sommatoria è estesa a tutti gli  $i$  da 1 a  $n-1$  ma solo con  $j > i$ ,**

e che rappresenta solo la generalizzazione di quella abbreviata di Mc Nemar, fondata sulle diagonale secondaria (b, c), senza termine di correzione

$$c^2 = \sum \frac{(b-c)^2}{b+c}$$

Con i dati dell'esempio, si ottiene

$$c^2 = \frac{(28-16)^2}{28+16} + \frac{(54-35)^2}{54+35} + \frac{(39-25)^2}{39+25} = \frac{144}{44} + \frac{361}{89} + \frac{225}{64} = 3,27 + 4,06 + 3,52 = 10,85$$

un valore uguale a 10,85

3 – Sotto l'ipotesi nulla di simmetria dei dati, la distribuzione asintotica (teoricamente, per un numero  $n$  di dati tendente all'infinito; in pratica per  $n$  sufficientemente grande) è un chi quadrato con  $gdl$

$$gdl = \frac{n \cdot (n-1)}{2}$$

Con i dati dell'esempio, i gdl

$$gdl = \frac{3 \cdot (3-1)}{2} = 3$$

sono 3, corrispondenti alle 3 somme fatte o 3 coppie di valori a confronto (indicate con a, b, c, nella tabella).

4 - Poiché il valore del chi quadrato con 3 gdl

- alla probabilità  $\alpha = 0.05$  è uguale a 7,815,
- alla probabilità  $\alpha = 0.01$  è uguale a 11,345,

si rifiuta l'ipotesi nulla, con una probabilità  $\alpha$  inferiore a 0.05.

ESEMPIO. Per valutare l'impatto di una discarica sulla popolazione residente in un comune, prendendo un campione nominativo è stata fatta un'indagine sulla percezione di odori sgradevoli, fornendo 4 livelli di risposta: mai, di rado, spesso, sempre.

Dopo un anno, l'indagine è stata ripetuta sugli stessi individui con i seguenti risultati complessivi:

		DOPO			
		Mai	Di rado	Spesso	Sempre
PRIMA	Mai	25	<b>86</b>	<b>97</b>	<b>25</b>
	Di rado	<b>65</b>	78	<b>54</b>	<b>33</b>
	Spesso	<b>58</b>	<b>47</b>	68	<b>28</b>
	Sempre	<b>12</b>	<b>39</b>	<b>17</b>	14

Si può sostenere che l'intervento ha modificato la situazione in modo statisticamente significativo?

Risposta.

Indicando la posizione dei dati (esclusa la diagonale) con  $X_{ij}$  come nella tabella

		DOPO			
		Mai	Di rado	Spesso	Sempre
PRIMA	Mai	25	<b>86</b> $X_{1,2}$	<b>97</b> $X_{1,3}$	<b>25</b> $X_{1,4}$
	Di rado	<b>65</b> $X_{2,1}$	78	<b>54</b> $X_{2,3}$	<b>33</b> $X_{2,4}$
	Spesso	<b>58</b> $X_{3,1}$	<b>47</b> $X_{3,2}$	68	<b>28</b> $X_{3,4}$
	Sempre	<b>12</b> $X_{4,1}$	<b>39</b> $X_{4,2}$	<b>17</b> $X_{4,3}$	14

Il metodo si fonda su quattro passaggi logici:

1- per verificare l'ipotesi nulla

$$H_0: p_{ij} = p_{ji}$$

contro l'ipotesi alternativa

$$H_1: p_{ij} \neq p_{ji}$$

per tutti gli  $i$  da 1 a  $n-1$  con  $j > i$ ,

il confronto è tra i valori a destra della diagonale con quelli a sinistra, che occupano la posizione simmetrica (esempio  $X_{1,2}$  contro  $X_{2,1}$  e  $X_{2,4}$  contro  $X_{4,2}$ ...);

2 - con la formula

$$c^2 = \sum \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}}$$

e sommatoria per tutti gli  $i$  da 1 a  $n-1$  con  $j > i$

si calcola un valore del chi quadrato

$$c^2 = \frac{(86-65)^2}{86+65} + \frac{(97-58)^2}{97+58} + \frac{(25-12)^2}{25+12} + \frac{(54-47)^2}{54+47} + \frac{(33-39)^2}{33+39} + \frac{(28-17)^2}{28+17} =$$

$$= \frac{441}{151} + \frac{1521}{155} + \frac{169}{37} + \frac{49}{101} + \frac{36}{72} + \frac{121}{45} = 2,92 + 9,81 + 4,57 + 0,49 + 0,50 + 2,69 = 20,98$$

che risulta uguale a 20,98;

3 - con il rapporto

$$gdl = \frac{n \cdot (n-1)}{2}$$

applicato alla matrice 4 x 4

$$\text{gdl} = \frac{4 \cdot (4 - 1)}{2} = 6$$

si stimano 6 gdl, che corrispondono al numero di coppie di dati a confronto;

4 – poiché il valore critico del chi quadrato con 6 gdl

- alla probabilità  $\alpha = 0.05$  è uguale a 12,592 e

- alla probabilità  $\alpha = 0.01$  è uguale a 16,815

si rifiuta l'ipotesi nulla, con una probabilità  $\alpha$  inferiore a 0.01.

### **11.9. ANALISI DELLA VARIANZA PER RANGHI, A 2 CRITERI DI CLASSIFICAZIONE: TEST DI FRIEDMAN, CON UNA E CON K REPLICHE**

Il test di Cochran si applica a dati dicotomizzati. Spesso è possibile disporre di misure più precise, su una scala quantitativa continua, che deve essere almeno di tipo ordinale ma con pochi valori identici.

Quando i dati raccolti sono valori di una scala ordinale oppure misure più sofisticate come i valori di una scala d'intervalli o di rapporti, ma non sussistono i requisiti di validità per utilizzare l'ANOVA, si può ricorrere al test presentato nel 1937 da **M. Friedman** (con l'articolo *The use of ranks to avoid the assumptions of normality implicit in the analysis of variance* pubblicato su *Journal Amer. Statist. Assoc.*, vol. 32, pp. 675-701).

**E' uno dei test non parametrici più potenti, in un disegno sperimentale a 2 criteri di classificazione** o a blocchi randomizzati. Come tale è riportato in quasi tutte le librerie statistiche informatiche, che comprendano anche pochi test non parametrici.

Serve per determinare la probabilità se le diverse situazioni a confronto, riportate nelle colonne, abbiano la stessa mediana. Con un linguaggio tecnicamente più preciso, per verificare se è possibile sostenere che sono campioni estratti dalla stessa popolazione o da popolazioni differenti, ma con la medesima tendenza centrale ( $H_0$ ).

In termini formali, l'ipotesi nulla è

$$H_0: Me_A = Me_B = \dots = Me_K$$

e l'ipotesi alternativa è

$$H_1: \text{non tutte le } k \text{ mediane sono uguali.}$$

Per una presentazione chiara del problema, è utile che i dati siano riportati in una tabella, nella quale i valori delle varie righe fanno riferimento agli stessi soggetti e le colonne ai medesimi casi di studio, come nell'analisi della varianza a due criteri di classificazione di cui è la versione non parametrica.

Ma, a differenza dell'ANOVA parametrica a due criteri, **l'ipotesi nulla è una sola e verte sulle k situazioni o casi o trattamenti, mentre i soggetti o individui sono considerati soltanto come repliche.**

**NEL CASO DI UN SOLO DATO PER CASELLA**, si ottiene una tabella

SOGGETTI O INDIVIDUI	SITUAZIONI O CASI			
	A	B	---	K
1	23	150	---	8
2	10	12	---	5
---	---	---	---	---
N	12	13	---	9
<b>Mediane</b>	<b>Me<sub>A</sub></b>	<b>Me<sub>B</sub></b>		<b>Me<sub>K</sub></b>

Se l'ipotesi riguarda le mediane degli **N** individui, è sufficiente scambiare le righe con le colonne. Le due analisi non possono essere condotte in modo simultaneo, come nell'analisi della varianza a due criteri; ma nulla vieta che, interessando entrambi i problemi, possano essere condotte entrambe in tempi successivi, in modo indipendente.

In questo disegno sperimentale, si ricorre al test di Friedman fondamentalmente in 2 situazioni:

- 1- quando sono state utilizzate misure continue in scale d'intervallo o di rapporto, di conseguenza con nessuna o pochissime misure ripetute, ma non sono rispettate le assunzioni di validità dell'ANOVA;
- 2- quando sono state usate misure discrete, semi-quantitative o di rango anche se rappresentate in forma simbolica.

**Il procedimento** è semplice e richiede pochi passaggi.

- 1 - data una tabella a doppia entrata

SOGGETTI O INDIVIDUI	SITUAZIONI O CASI		
	A	B	C
1	23	50	18
2	12	10	5
3	23	28	19
4	12	13	9

2 - trasformare i punteggi o le misure in ranghi entro la stessa riga, assegnando 1 al punteggio minore e progressivamente valori maggiori fino a **k**, uguale al numero di colonne, al punteggio maggiore della medesima riga;

successivamente, sommare per colonna i valori dei ranghi ( $T_i$ )

SOGGETTI O INDIVIDUI	SITUAZIONI O CASI		
	A	B	C
1	2	3	1
2	3	2	1
3	2	3	1
4	2	3	1
<b>Totali (<math>T_i</math>)</b>	<b>9</b>	<b>11</b>	<b>4</b>

3 - Se l'ipotesi nulla  $H_0$  è vera, nelle colonne a confronto i ranghi minori e quelli maggiori dovrebbero essere distribuiti casualmente;

pertanto, le somme dei ranghi nelle **k** colonne ( $T_i$  osservati) dovrebbero essere tra loro tutte equivalenti ed uguali ad un valore atteso, che dipende dal numero di osservazioni,

$$T_i \text{ (attesi)} = \frac{N \cdot (k + 1)}{2}$$

dove **N** è il numero di righe.

Con i dati dell'esempio, **N** = 4 e **k** = 3

$$T_i \text{ (attesi)} = \frac{4 \cdot (3+1)}{2} = 8$$

la somma di ogni colonna dovrebbe essere uguale o almeno prossima a 8, considerando le variazioni casuali.

4 - Se l'ipotesi nulla  $H_0$  è falsa, in almeno una colonna si concentrano i ranghi minori o maggiori; di conseguenza, tale somma tende ad essere significativamente differente dal valore  $T_i$  atteso.

Per decidere se queste somme ( $T_i$  osservati) sono significativamente differenti dell'atteso, si calcola la statistica  $Fr$

$$Fr = \sum_{i=1}^k \left( T_i - \frac{N(k+1)}{2} \right)^2$$

che è la sommatoria dei quadrati degli scarti tra i  $k$  totali osservati e i corrispondenti attesi.

E' ovvio che tale valore di  $Fr$

- tenderà a 0 nel caso di accordo tra totali osservati e totali attesi ( $H_0$  vera),
- mentre tenderà ad essere grande al crescere dello scarto tra essi ( $H_0$  falsa).

Con i dati dell'esempio, si ottiene un valore di  $Fr$

$$Fr = (9 - 8)^2 + (11 - 8)^2 + (4 - 8)^2 = 1^2 + 3^2 + 4^2 = 1 + 9 + 16 = 26$$

uguale a 26.

Nel caso di **piccoli campioni**

**(dimensioni inferiori a 30-35 osservazioni, che variano tra  $k = 3$  e  $N \leq 15$  e tra  $k = 4$  e  $N \leq 9$ )**

la significatività della statistica  $Fr$  è fornita da tabelle specifiche (riportate nella pagina seguente).

Per  $k = 3$  e  $N = 4$ , i valori critici sono

24 alla probabilità  $\alpha = 0.10$

26 alla probabilità  $\alpha = 0.05$

32 alla probabilità  $\alpha = 0.01$

Poiché il valore calcolato è stato 26, che coincide con la probabilità  $\alpha = 0.05$ , si può rifiutare l'ipotesi nulla.

**VALORI CRITICI DI  $F_r$  DEL TEST DI FRIEDMAN  
PER L'ANALISI DELLA VARIANZA NON PARAMETRICA A DUE CRITERI  
IN PICCOLI CAMPIONI**

<b>k</b>	<b>N</b>	<b>0.10</b>	<b>0.05</b>	<b>0.01</b>	<b>0.001</b>
<b>3</b>	<b>3</b>	18	18	---	---
	<b>4</b>	24	26	32	---
	<b>5</b>	26	32	42	50
	<b>6</b>	32	42	54	72
	<b>7</b>	38	50	62	86
	<b>8</b>	42	50	72	98
	<b>9</b>	50	56	78	114
	<b>10</b>	50	62	96	122
	<b>11</b>	54	72	104	146
	<b>12</b>	62	74	114	150
	<b>13</b>	62	78	122	168
	<b>14</b>	72	86	126	186
	<b>15</b>	74	96	134	194
<b>4</b>	<b>2</b>	20	20	---	---
	<b>3</b>	33	37	45	---
	<b>4</b>	42	52	64	74
	<b>5</b>	53	65	83	105
	<b>6</b>	64	76	102	128
	<b>7</b>	75	91	121	161
	<b>8</b>	84	102	138	184

Per **campioni di grandi dimensioni ( $k \geq 5$ )**, è stato proposto un indice  $c_F^2$  che si distribuisce approssimativamente come il  $\chi^2_{(k-1)}$  con **gdl k-1**.

Può essere stimato mediante la formula

$$\chi^2_F = \frac{12}{Nk(k+1)} \sum_{i=1}^k \left( T_i - \frac{N(k+1)}{2} \right)^2$$

in cui

la seconda parte è data dagli scarti al quadrato tra somma osservata ed attesa,

mentre la prima dipende dall'errore standard, determinato numero di dati, trattandosi di ranghi.

Per **semplificare i calcoli, sono state proposte formule abbreviate**, tra le quali una di quelle ricorrenti nei testi di statistica è

$$c_F^2 = \frac{12 \cdot \sum T_i}{N \cdot k \cdot (k+1)} - 3N \cdot (k+1)$$

dove:

**N** è il numero di righe od osservazioni in ogni campione (tutte con il medesimo numero di dati),

**k** è il numero di colonne o campioni a confronto,

**T<sub>i</sub>** è la somma dei ranghi di ogni colonna e la sommatoria  $\Sigma$  è estesa a tutte le colonne.

I risultati possono essere confrontati con **i valori critici di seguito riportati, tratti dalla distribuzione chi quadrato con df k-1**

VALORI CRITICI DEL $c_F^2$				
<b>k</b>	<b>a = 0.10</b>	<b>a = 0.05</b>	<b>a = 0.01</b>	<b>a=0.001</b>
<b>5</b>	7,78	9,49	13,28	18,46
<b>6</b>	9,24	11,07	15,09	20,52
<b>7</b>	10,64	12,59	16,81	22,46
<b>8</b>	12,02	14,07	18,48	24,32
<b>9</b>	13,36	15,51	20,09	26,12
<b>10</b>	14,68	16,92	21,67	27,88

o comunque presi direttamente da una distribuzione chi quadrato.

NEL CASO DI k DATI PER CASELLA, come le misure d'inquinamento raccolte in un corso d'acqua,

- in **k** stazioni (situazioni o località, riportate nei trattamenti)
- in **N** giorni di rilevazione, identici per ogni stazione (riportati nei blocchi)
- con **r** repliche, uguali per ogni stazione e giorno (riportate all'incrocio tra trattamenti e blocchi) e quindi con **esperimenti bilanciati** come nella tabella sottostante, per verificare la differenza nelle tendenze centrali d'inquinamento tra le stazioni, con ipotesi nulla

$$H_0: Me_A = Me_B = \dots = Me_K$$

e ipotesi alternativa

**H<sub>1</sub>: non tutte le k mediane sono uguali**

GIORNI	STAZIONI O LOCALITA'			
	A	B	---	K
1	23	150	---	8
	28	123	---	15
2	10	12	---	5
	15	14	---	4
---	---	---	---	---
	---	---	---	---
N	12	13	---	9
	10	15	---	6
Mediane	Me <sub>A</sub>	Me <sub>B</sub>		Me <sub>K</sub>

si deve

1- trasformare i dati in ranghi in modo indipendente per ogni blocco, considerando entro essi contemporaneamente i **kr** dati,

2 - calcolare i totali **T<sub>i</sub>** di ogni colonna o trattamento degli **Nr** ranghi,

3 - stimare il valore **c<sub>F</sub><sup>2</sup>**

$$c_F^2 = \frac{12}{N \cdot k \cdot r^2 \cdot (k \cdot r + 1)} \cdot \sum \left[ T_i - \frac{N \cdot r \cdot (k \cdot r + 1)}{2} \right]^2$$

4 - la cui significatività è data dalla distribuzione  $c^2_{(k-1)}$ .

Nel caso di un solo dato o replica per casella (quindi  $r = 1$ ), la formula precedente coincide con

$$c_F^2 = \frac{12}{Nk(k+1)} \sum_{i=1}^k \left( T_i - \frac{N(k+1)}{2} \right)^2$$

Il test di **Friedman** è analogo all'analisi della varianza a due criteri di classificazione.

L'**efficienza asintotica relativa** del test **Friedman** rispetto al test **F** di **Fisher-Snedecor**

- quando la distribuzione dei dati è Normale ha un valore uguale a  $0,95k/(k+1) = (3/\pi) \cdot k/(k+1)$ ,

- quando la distribuzione dei dati è Rettangolare ha un valore uguale a  $1k/(k+1)$ ,

- quando la distribuzione è Esponenziale Doppia ha un valore uguale a  $1,5k/(k+1) = (3/2) \cdot k/(k+1)$ .

Quando la distribuzione dei dati è normale, il test non parametrico ha una potenza leggermente inferiore, all'aumentare del numero di gruppi; mentre ha una potenza superiore quando ci si allontana dalla normalità. Resta il vantaggio, già ripetuto per i test non parametrici, che le loro conclusioni non possono essere rifiutate.

**ESEMPIO 1 (PICCOLI CAMPIONI E 1 DATO PER CASELLA).** Per il confronto sulla qualità della vita in 4 quartieri (A, B, C, D) sono stati raccolti punteggi o misure su 5 caratteristiche, come riportato nella tabella

	QUARTIERI			
	A	B	C	D
<b>Rumore diurno (in dBA)</b>	80	75	70	76
<b>Rumore notturno (in DbA)</b>	74	69	70	71
<b>Inquinamento dell'aria</b>	9	7	5	6
<b>Traffico</b>	7	8	7	6
<b>Carenza di Parcheggi</b>	8	6	5	6

dove, per ogni fattore analizzato, il punteggio maggiore corrisponde ad un livello di disagio crescente.

Si vuole verificare se esistono differenze significative tra i 4 quartieri nei livelli mediani di disagio.

Risposta.

**Per fornire una sintesi di parametri così diversi, espressi in unità di misura differenti e a volte soggettive, è necessario trasformare i valori nei loro ranghi.**

Le misure ed i punteggi sono quindi trasformati come nella tabella sottostante

	A	B	C	D
Rumore diurno (in dBA)	4	2	1	3
Rumore notturno (in dBA)	4	1	2	3
Inquinamento dell'aria	4	3	1	2
Traffico	2,5	4	2,5	1
Carenza di Parcheggi	4	2,5	1	2,5
<b>Totali Ranghi <math>T_i</math></b>	<b>18,5</b>	<b>12,5</b>	<b>7,5</b>	<b>11,5</b>

nella quale sono riportati i totali dei ranghi per colonna ( $T_i$ ).

Applicando la formula per piccoli campioni

$$Fr = \sum_{i=1}^k \left( T_i - \frac{N(k+1)}{2} \right)^2$$

in cui

$$\left( \frac{N(k+1)}{2} \right) = \left( \frac{5(4+1)}{2} \right) = \frac{25}{2} = 12,5$$

si ottiene un valore di Fr

$$Fr = (18,5 - 12,5)^2 + (12,5 - 12,5)^2 + (7,5 - 12,5)^2 + (11,5 - 12,5)^2 = 36 + 0 + 25 + 1 = 62$$

uguale a 62.

Per  $k = 4$  e  $N = 5$

alla probabilità  $\alpha = 0.05$  il valore critico riportato nella tabella precedente è 65.

Poiché il valore calcolato è 62, non è possibile rifiutare l'ipotesi nulla: non è dimostrata una differenza significativa tra le mediane della qualità della vita nei 4 quartieri.

ESEMPIO 2 (**GRANDI CAMPIONI E 1 DATO PER CASELLA**). Prodotto dalle combustioni contenenti zolfo, il biossido di zolfo  $\text{SO}_2$  è uno dei fattori di inquinamento dell'aria più dannosi, a breve raggio. Secondo la legislazione italiana, il limite di accettabilità è fissato in  $80 \text{ gm}^{-3}$  (equivalenti a 0,03 ppm) come mediana delle concentrazioni medie giornaliere di un anno.

<b>Emissioni medie giornaliere delle 6 zone</b>						
<b>Giorno</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>
<b>1</b>	55	75	32	60	36	48
<b>2</b>	57	70	69	65	41	52
<b>3</b>	60	24	43	68	98	54
<b>4</b>	52	85	51	58	33	64
<b>5</b>	57	69	72	56	28	53
<b>6</b>	59	45	107	64	38	55
<b>7</b>	58	73	33	35	66	54
<b>8</b>	58	67	75	71	35	56
<b>9</b>	57	48	51	67	55	36
<b>10</b>	59	79	48	81	52	63
<b>11</b>	88	70	53	64	43	65
<b>12</b>	61	40	42	71	81	55
<b>13</b>	57	45	38	65	39	53
<b>14</b>	59	76	40	67	38	51
<b>15</b>	55	73	57	69	42	56

Un'esposizione continua alla concentrazione di 0,03-0,05 ppm determina un peggioramento delle condizioni dei pazienti bronchitici; una esposizione di solo 20 secondi alla concentrazione 0,3-1 ppm può portare all'alterazione dell'attività cerebrale; più di 6 ore di esposizione ad oltre 20 ppm possono causare la saturazione delle vie e dei tessuti polmonari, con eventuale paralisi e/o morte.

Durante il periodo invernale, con una rilevazione continua, per 15 giorni sono state misurate le medie giornaliere presso 6 aree sia industriali che residenziali di una città.

Si vuole verificare se la tendenza centrale delle emissioni del periodo è significativamente diversa tra le 6 zone di monitoraggio.

Risposta.

I dati raccolti formano un campione di grandi dimensioni e le misure riportate utilizzano una scala d'intervallo o di rapporto. Tuttavia, la distribuzione è fortemente asimmetrica, in alcuni gruppi: anche senza una misura oggettiva di valutazione delle simmetrie e di confronto della varianza, è possibile osservare che l'intervallo di variazione di alcuni gruppi è molto è differente.

Nel caso del problema, soprattutto è l'uso di valori medi che vieta il ricorso all'ANOVA parametrica.

E' vantaggioso ed appropriato utilizzare il test di Friedman. A tal fine,

- si trasformano i dati in ranghi entro la stessa riga;
- successivamente, si calcolano le somme per colonna, come nella tabella sottostante

Ricordando che  $N = 15$  e  $k = 6$ ,

dai dati della tabella successiva, in cui i valori sono stati trasformati in ranghi entro la stessa riga, si stima il valore  $C_F^2$

$$C_F^2 = \frac{12 \cdot (58^2 + 64^2 + 44^2 + 69^2 + 35^2 + 45^2)}{15 \cdot 6 \cdot 7} - 3 \cdot 15 \cdot 7$$

$$C_F^2 = \frac{12 \cdot (3364 + 4096 + 1936 + 4761 + 1255 + 2025)}{630} - 315$$

$$C_F^2 = \frac{12 \cdot 17407}{630} - 315 = \frac{208884}{630} - 315 = 331,57 - 315 = 16,57$$

che risulta uguale a 16,57.

Emissioni medie giornaliere (gm <sup>-3</sup> ) delle 6 zone						
giorno	A	B	C	D	E	F
1	4	6	1	5	2	3
2	3	6	5	4	1	2
3	4	1	2	5	6	3
4	3	6	2	4	1	5
5	4	5	6	3	1	2
6	4	2	6	5	1	3
7	4	6	1	2	5	3
8	3	4	6	5	1	2
9	5	2	3	6	4	1
10	3	5	1	6	2	4
11	6	5	2	3	1	4
12	4	1	2	5	6	3
13	5	3	1	6	2	4
14	4	6	2	5	1	3
15	2	6	4	5	1	3
<b>T<sub>i</sub></b>	<b>58</b>	<b>64</b>	<b>44</b>	<b>69</b>	<b>35</b>	<b>45</b>

Il valore critico del  $\chi^2$  con **df 5 (k = 6)**

- alla probabilità  $\alpha = 0.05$  è uguale a 11,07
- mentre alla probabilità  $\alpha = 0.01$  è uguale a 15,09
- e alla probabilità  $\alpha = 0.001$  è uguale a 20,52.

Di conseguenza, con probabilità inferiore a 0.01 di commettere un errore di I° tipo, si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa: le mediane dei 6 gruppi a confronto non sono tra loro tutte statisticamente uguali.

ESEMPIO 3 (CON **k REPLICHE O MISURE RIPETUTE PER CASELLA**). Per valutare la significatività delle differenze nei livelli d'inquinamento tra 4 stazioni collocate lungo un corso d'acqua, data la grande variabilità stagionale nelle portate d'acqua, i campioni sono stati raccolti alle stesse date, effettuando 2 prelievi in 5 giorni diversi

GIORNI	STAZIONI O LOCALITA'			
	A	B	C	D
I	20 28	33 31	34 39	40 41
II	55 58	61 54	69 63	73 70
III	18 22	24 19	15 23	28 26
IV	14 18	17 13	21 23	19 24
V	37 33	41 38	53 54	48 51

Esistono differenze significative nei livelli d'inquinamento delle 4 zone o stazioni?

Risposta.

GIORNI	STAZIONI O LOCALITA'			
	A	B	C	D
I	1	4	5	7
	2	3	6	8
II	2	4	6	8
	3	1	5	7
III	2	6	1	7
	4	3	5	8
IV	2	3	6	5
	4	1	7	8
V	2	4	7	5
	1	3	8	6
<b>Totali Ti</b>	<b>23</b>	<b>32</b>	<b>56</b>	<b>69</b>

Dopo

- **aver trasformato i valori nei ranghi relativi entro lo stesso blocco**, con una impostazione grafica leggermente diversa da quella della tabella precedente, per meglio evidenziare che si tratta di repliche nelle stessa stazione e alla stessa data,
- aver calcolato i totali dei ranghi per colonna,

**il cui totale  $ST_i$**

$$(23 + 32 + 56 + 69) = 180$$

è uguale a 180 e che, come verifica di non aver commesso errori nella trasformazione in ranghi e nelle somme successive, **deve essere uguale a**

$$ST_i = \frac{N \cdot k \cdot r \cdot (1 + k \cdot r)}{2}$$

(con i dati dell'esempio

$$ST_i = \frac{5 \cdot 4 \cdot 2 \cdot (1 + 4 \cdot 2)}{2} = \frac{40 \cdot 9}{2} = \mathbf{180}$$

è infatti uguale a 180)

- con la formula generale

$$c_F^2 = \frac{12}{N \cdot k \cdot r^2 \cdot (kr + 1)} \cdot \sum \left[ T_i - \frac{Nr(kr + 1)}{2} \right]^2$$

dai dati dell'esempio, dopo aver stimato che

il totale atteso di ogni colonna è

$$\frac{N \cdot r \cdot (k \cdot r + 1)}{2} = \frac{5 \cdot 2 \cdot (4 \cdot 2 + 1)}{2} = \mathbf{45}$$

uguale a 45

- si calcola il valore di  $c_F^2$

$$\begin{aligned} c_F^2 &= \frac{1}{5 \cdot 4 \cdot 2^2 \cdot (4 \cdot 2 + 1)} \cdot [(23 - 45)^2 + (32 - 45)^2 + (56 - 45)^2 + (69 - 45)^2] = \\ &= \frac{1}{80 \cdot 9} \cdot (484 + 169 + 121 + 576) = \frac{1}{720} \cdot 1350 = \mathbf{1,875} \end{aligned}$$

che risulta uguale a 1,875 con 3 gdl.

Poiché alla probabilità  $\alpha = 0.05$  il valore critico del chi quadrato con 3 gdl è uguale a 7,815 non è possibile rifiutare l'ipotesi nulla.

**Correzione per valori identici o ranghi ripetuti (ties).**

Con dati discreti e scale semiquantitative od ordinali, i valori identici possono essere numerosi. La varianza della distribuzione campionaria diventa minore; si rende necessario apportare una correzione al valore di  $C_F^2$ , che ne aumenta il valore. La correzione non ha effetti quando non esistono ripetizioni; di conseguenza, alcuni testi la inseriscono nella formula generale (che ovviamente risulterà differente da quella precedentemente riportata), anche se in questo modo il calcolo manuale di  $C_F^2$  è più lungo e più frequentemente può determinare errori.

Si ottiene il **valore corretto di  $C_F^2$**  ponendo al denominatore non più  $N \cdot k \cdot (k + 1)$  ma tutta la formula seguente

$$N \cdot k \cdot (k + 1) + \frac{N \cdot k - \sum_{i=1}^N \sum_{j=1}^p r_{ij}^3}{k - 1}$$

dove:

**N** è il numero di righe o osservazioni per gruppo,

**k** è il numero di colonne o gruppi,

**p** è il numero di dati con lo stesso valore, e quindi con lo stesso rango, nella medesima riga;

**r<sub>ij</sub>** è la dimensione dei ranghi ripetuti.

Nel calcolo della correzione, con **r<sub>ij</sub> = 1** vengono inclusi anche i dati con valori che compaiono una sola volta nella stessa riga. Pertanto essi contribuiscono per un valore  $r_{ij}^3 = 1$  ( $1^3 = 1$ ), mentre

- i valori che compaiono 2 volte contribuiscono per un valore  $r_{ij}^3 = 8$  ( $2^3 = 8$ ),
- quelli che compaiono 3 volte per un valore  $r_{ij}^3 = 27$  ( $3^3 = 27$ ), ecc....

**ESEMPIO 4 (CON CORREZIONE PER TIES).** Al fine di ottenere un quadro possibilmente completo della qualità dell'aria, in 6 zone di una città (centro storico (A), 3 aree periferiche (B, C, D) e 2 zone industriali (E, F)) sono state collocate stazioni di rilevamento per misurare con continuità tutti i parametri indicati dalla legge.

Sono stati misurati gli inquinanti: anidride solforosa, ossidi di azoto, ossido di carbonio, fluoro, piombo, polveri, ozono, idrocarburi. Per ognuno di essi è stato valutato il grado di rispetto delle norme di legge. Inoltre sono stati rilevati alcuni parametri meteorologici che influiscono sulla qualità dell'aria: velocità e direzione del vento, temperatura, umidità, irraggiamento solare.

Per una più facile e diffusa comprensione della situazione da parte della popolazione e degli amministratori, i tecnici hanno divulgato i dati con cadenza settimanale ed hanno elaborato un criterio che definisce il "giudizio sintetico" di qualità dell'aria in 5 classi:

buona (++), discreta (+), mediocre (=), scadente (-), pessima (--).

I dati riportati in tabella sono i risultati di 10 settimane.

settimane	ZONA					
	A	B	C	D	E	F
1	++	+	+	-	=	-
2	+	=	+	-	=	=
3	--	+	+	++	-	-
4	-	=	=	+	+	-
5	=	+	+	-	=	--
6	+	=	+	-	-	-
7	=	=	--	+	-	++
8	+	+	++	-	=	-
9	+	-	--	+	=	=
10	++	+	++	-	=	-

Si vuole verificare se tra le 6 zone esiste una differenza significativa nei giudizi mediani della qualità dell'aria.

Risposta.

L'analisi statistica richiede preliminarmente la trasformazione dei giudizi sintetici in una scala di rango. E' un campione di grandi dimensioni, con misure ordinali o semiquantitative.

Un modo razionale di trasformazione dei simboli in ranghi potrebbe essere l'assegnazione del valore 1 alla zona che ha la migliore qualità dell'aria e progressivamente un valore maggiore al crescere dell'inquinamento, fino ad assegnare 6 alla zona la cui qualità dell'aria è stata giudicata la peggiore.

Nulla cambia per l'analisi statistica se si agisse nel modo opposto, assegnando valori da 6 a 1.

Nel caso di valutazioni identiche entro la stessa riga, si deve assegnare lo stesso valore stimato come media aritmetica dei ranghi occupati.

Per facilitare il calcolo della **correzione per misure ripetute i ties**, è utile riportare anche la somma dei valori  $r_{ij}^3$

settimane	ZONA						$\sum_j^p r_{ij}^3$
	A	B	C	D	E	F	
1	1	2,5	2,5	5,5	4	5,5	$1 + 8 + 1 + 8 = 18$
2	1,5	4	1,5	6	4	4	$8 + 27 + 1 = 36$
3	6	2,5	2,5	1	4,5	4,5	$1 + 8 + 1 + 8 = 18$
4	5,5	3,5	3,5	1,5	1,5	5,5	$8 + 8 + 8 = 24$
5	3,5	1,5	1,5	5	3,5	6	$8 + 8 + 1 + 1 = 18$
6	1,5	3	1,5	5	5	5	$8 + 1 + 27 = 36$
7	3,5	3,5	6	2	5	1	$8 + 1 + 1 + 1 + 1 = 12$
8	2,5	2,5	1	5,5	4	5,5	$1 + 8 + 1 + 8 = 18$
9	1,5	5	6	1,5	3,5	3,5	$8 + 8 + 1 + 1 + = 18$
10	1,5	3	1,5	5,5	4	5,5	$8 + 1 + 1 + 8 = 18$
<b>T<sub>i</sub></b>	<b>29,0</b>	<b>31,0</b>	<b>27,5</b>	<b>38,5</b>	<b>39,0</b>	<b>46,0</b>	$\sum_{i=1}^N \sum_j^p r_{ij}^3 = 216$

La somma dei valori per i ranghi ripetuti ( $\sum_{i=1}^N \sum_{j=1}^p r_{ij}^3$ ) è uguale a 216.

Nella formula generale

$$\chi^2_F = \frac{12}{Nk(k+1)} \sum_{i=1}^k \left( T_i - \frac{N(k+1)}{2} \right)^2$$

al denominatore al posto di  $Nk(k+1)$

si sostituisce la formula con i ties

$$N \cdot k \cdot (k+1) + \frac{N \cdot k - \sum_{i=1}^N \sum_{j=1}^p r_{ij}^3}{k-1}$$

dove:

**N** è il numero di righe o osservazioni per gruppo,

**k** è il numero di colonne o gruppi,

**p** è il numero di dati con lo stesso valore, e quindi con lo stesso rango, nella medesima riga,

$r_{ij}$  è la dimensione dei ranghi ripetuti.

Con i dati dell'esempio, il valore  $c_F^2$  corretto per i ties è

$$c_F^2 = \frac{12 \cdot (29^2 + 31^2 + 27,5^2 + 38,5^2 + 39^2 + 46^2)}{10 \cdot 6 \cdot 7 + \left(\frac{10 \cdot 6 - 216}{6-1}\right)} - 3 \cdot 10 \cdot 7$$

e svolgendo i calcoli

$$c_F^2 = \frac{12 \cdot (841 + 961 + 756,25 + 1482,25 + 1521 + 1516)}{420 + \left(\frac{-156}{5}\right)} - 210$$

si arriva a calcolare

$$c_F^2 = \frac{12 \cdot 7675,5}{420 - 31,2} - 210 = \frac{92130}{388,8} - 210 = 236,96 - 210 = 26,96$$

un risultato uguale a 26,96.

Il valore di  $c_F^2$  calcolato (26,96) è nettamente superiore al valore critico della tabella  $\chi^2$  per 5 df, anche alla probabilità  $\alpha = 0.001$  (20,52): si rifiuta l'ipotesi nulla.

#### 11.10. I CONFRONTI MULTIPLI TRA MEDIE DI RANGHI NELL'ANALISI DELLA VARIANZA NON PARAMETRICA, A DUE CRITERI DI CLASSIFICAZIONE

Anche nell'analisi della varianza non parametrica a due criteri di classificazione, **dopo aver rifiutato l'ipotesi nulla i confronti multipli permettono di individuare quali sono i trattamenti che risultano tra loro differenti**, alla probabilità prefissata.

Sulla base dei concetti del Bonferroni sul test **t**, quando si devono effettuare più confronti, per mantenere costante la probabilità complessiva la singola probabilità di ogni differenza deve diminuire in rapporto ai confronti che si vogliono effettuare.

Con **k** gruppi, i confronti possibili, tra loro non indipendenti, sono

$$\frac{k \cdot (k-1)}{2}$$

Se la probabilità scelta per tutte le differenze tra coppie di **medie dei ranghi** complessivamente è  $\alpha_T = 0.05$ , per ogni singolo confronto la probabilità  $\alpha$  diventa

$$\mathbf{a} = \frac{0.05}{\frac{k \cdot (k-1)}{2}}$$

Tuttavia, dato che di norma si considerano le differenze in valore assoluto, i confronti sono bilaterali. Di conseguenza, la probabilità tra ogni confronto deve essere dimezzata.

Con una probabilità complessiva prefissata  $\mathbf{a}_T = 0.05$  per  $k$  gruppi, la probabilità corrispondente  $\mathbf{a}$  per ogni confronto è

$$\alpha = \frac{0.05}{k \cdot (k-1)}$$

Quando il numero di dati e di gruppi è sufficientemente grande, si ricorre alla distribuzione normale. Sono significative **le differenze tra le medie dei ranghi** di ogni trattamento che, in valore assoluto, sono uguali o superano la quantità **D**

$$D \geq z_{\frac{\mathbf{a}}{k \cdot (k-1)}} \cdot \sqrt{\frac{k \cdot (k+1)}{6N}}$$

dove:

$\mathbf{a}$  è la probabilità complessiva prefissata,

$\mathbf{k}$  è il numero di gruppi, tra i quali sono possibili  $k \cdot (k-1)$  confronti,

$\mathbf{N}$  è il numero di righe od osservazioni per ogni campione.

Quando entro i  $\mathbf{k}$  gruppi esiste **un trattamento usato come controllo**, per cui i risultati degli altri  $\mathbf{k-1}$  trattamenti vengono confrontati solamente con il controllo, la significatività di ognuna delle  $\mathbf{k-1}$  differenze medie possono essere verificate con la stessa metodologia.

Di conseguenza, la probabilità  $\mathbf{a}$  del confronto di ogni singolo trattamento con il controllo risulta più alta di quella stimata per i confronti precedenti:

- il numero di confronti diventa  $\mathbf{k-1}$ , per cui il valore di  $\mathbf{a}_T$  complessivo deve essere diviso per  $\mathbf{k-1}$ ;
- inoltre, poiché sono confronti unilaterali, non è richiesto di dimezzare ulteriormente la probabilità.

In conclusione, con una probabilità complessiva  $\mathbf{a}_T = 0.05$ , la probabilità  $\mathbf{a}$  di ogni confronto diviene uguale a

$$\alpha = \frac{0.05}{(k-1)}$$

ESEMPIO. In un esempio precedente, si è dimostrato che esiste una differenza significativa tra le mediane delle emissioni gassose giornaliere di 6 zone di una città, rilevate per 15 giorni. Nella tabella sottostante, sono state riportate le medie dei ranghi dell'esempio.

Emissioni giornaliere di 6 zone						
	A	B	C	D	E	F
medie dei ranghi	3,86	4,26	2,93	4,60	2,33	3,00

Con confronti multipli a posteriori, si vuole verificare:

- 1 - tra quali zone esiste una differenza significativa alla probabilità  $\alpha_T$  complessiva uguale a 0.05;
- 2 - rispetto a quali altre zone la E, collocata in periferia e caratterizzata dai valori minimi della città, ha una media dei ranghi significativamente inferiore.

Risposte.

- 1 - Tra le medie dei ranghi delle 6 zone è possibile calcolare 15 differenze ( $\frac{6 \cdot 5}{2}$ ), che sono utilmente riportate in valore assoluto in una matrice triangolare

	A	B	C	D	E	F
	3,86	4,26	2,93	4,60	2,33	3,00
B	4,26	<b>0,40</b>				
C	2,93	<b>0,93</b>	<b>1,33</b>			
D	4,60	<b>0,74</b>	<b>0,34</b>	<b>1,67</b>		
E	2,33	<b>1,53</b>	<b>1,93</b>	<b>0,60</b>	<b>2,27</b>	
F	3,00	<b>0,86</b>	<b>1,26</b>	<b>0,07</b>	<b>1,60</b>	<b>0,67</b>

Per una probabilità complessiva prefissata  $\alpha_T = 0.05$  in un test a due code, con 15 medie a confronto simultaneo la probabilità  $\alpha$  per ogni media è  $0.05/(15 \times 2)$  uguale a **0.00167**.

Alla probabilità  $\alpha = 0.00167$  nella distribuzione normale corrisponde un valore di  $z$  uguale a 2,935 (approssimativamente la metà fra  $z = 2,93$  della probabilità 0.0017 e  $z = 2,94$  della probabilità 0.0016).

Applicando la formula

$$D \geq z \cdot \frac{\alpha}{k \cdot (k-1)} \cdot \sqrt{\frac{k \cdot (k+1)}{6N}}$$

con  $N = 15$  e  $k = 6$ ,

la differenza minima significativa **D**

$$D \geq 2,935 \cdot \sqrt{\frac{6 \cdot 7}{6 \cdot 15}} = 2,935 \cdot \sqrt{0,467} = 2,935 \cdot 0,683 = 2,01$$

è uguale a 2,01.

Confrontando il valore calcolato con le 15 differenze riportate nella tabella triangolare, risulta significativa solamente la differenza (uguale a 2,27) tra la media dei ranghi del gruppo D (uguale a 4,60) e quella del gruppo E (uguale a 2,33).

2 - Le differenze tra la zona di controllo **E** (con inquinamento minimo) e le altre 5 zone (**A, B, C, D, F**) sono riportate nella tabella

	A	B	C	D	F
medie dei ranghi	3,86	4,26	2,93	4,60	3,00
<b>differenze da E (2,33)</b>	<b>1,53</b>	<b>1,93*</b>	<b>0,60</b>	<b>2,27*</b>	<b>0,67</b>

Per effettuare questi 5 confronti unilaterali alla probabilità complessiva  $\alpha_T = 0.05$ , per ogni confronto singolo la probabilità specifica  $\alpha$  diviene

$$\alpha = \frac{0.05}{5} = 0.01$$

Ad essa corrisponde un valore di **z** uguale a **2,33 in una coda della distribuzione**.

Con gli stessi dati utilizzati per il campione precedente

**N = 15** e **k = 6**

il valore **D** è

$$D \geq 2,33 \cdot \sqrt{\frac{6 \cdot 7}{6 \cdot 15}} = 2,33 \cdot \sqrt{0,467} = 2,33 \cdot 0,683 = 1,591$$

uguale a 1,591.

Alla probabilità complessiva  $\alpha_T = 0.05$  risultano significativamente maggiori dell'inquinamento mediano della zona E (2,33) quelli rilevati nella zona B (4,26) e nella zona D (4,60) poiché la loro differenza (segnata con l'asterisco) è superiore al valore minimo calcolato.

## CAPITOLO XII

### REGRESSIONE LINEARE SEMPLICE

#### 12.1. REGRESSIONE O CORRELAZIONE?

Nei capitoli precedenti, sono state trattati i più importanti e diffusi test statistici per una **singola variabile quantitativa**. La distribuzione normale è utile per analizzare le caratteristiche di una distribuzione, il test t di Student e l'ANOVA servono per confrontare le differenze tra le medie di due o più campioni, ma limitatamente ad una sola variabile quantitativa. Nell'analisi della varianza a due o a più criteri di classificazione sono presi in considerazione contemporaneamente più fattori casuali, come i vari trattamenti e blocchi con le loro interazioni, ma sempre relativi alla medesima ed unica variabile misurata.

Quando si prendono in considerazione congiuntamente **due o più variabili quantitative** (per quelle qualitative, dette anche categoriali, si vedranno le misure di associazione), oltre alle precedenti analisi su ognuna è possibile

- **esaminare anche il tipo e l'intensità delle relazioni che sussistono tra esse.**

Per esempio, quando per ogni individuo si misurano contemporaneamente due variabili, come il peso e l'altezza, è possibile verificare se esse variano simultaneamente e come: se gli individui più alti sono anche mediamente più pesanti. E' possibile chiedersi

- **quale relazione matematica esista tra peso ed altezza nel campione analizzato,**
- **se la tendenza calcolata sia significativa, presente anche nella popolazione oppure debba essere ritenuta solo apparente,** effetto probabile di variazioni casuali del campione.

L'analisi congiunta di due variabili può offrire al ricercatore anche l'opportunità di

- **predire il valore di una variabile quando sia nota l'altra** (ad esempio, come valutare in un gruppo d'individui quale sia il peso medio al variare dell'altezza).

Nel caso della rilevazione simultanea di due variabili, è possibile ricorrere all'analisi della regressione oppure a quella della correlazione, da considerare tra loro alternative, seppure fondate su tecniche simili e in qualche caso interscambiabili.

Si ricorre all'**analisi della regressione** quando si intende sviluppare un modello statistico che può essere usato per predire i valori di **una variabile, detta dipendente o più raramente predetta ed individuata come effetto**, sulla base dei valori dell'**altra variabile, detta indipendente o esplicativa, individuata come causa**.

L'**analisi della correlazione** viene usata per misurare l'intensità dell'associazione tra **due variabili quantitative che variano congiuntamente, senza essere legate direttamente da causa-effetto**, ma

facilmente mediate da una terza variabile. Può avvenire anche per un'evoluzione temporale simile o per una differenziazione spaziale analoga.

E' importante saper sempre **distinguere tra associazione o casualità, tra evoluzione temporale simile o legame di causa-effetto**, per applicare l'analisi corretta al problema che si vuole risolvere.

## 12.2. DESCRIZIONE DI DISTRIBUZIONI BIVARIATE

Quando per ciascuna unità di un campione o di una popolazione si rilevano due caratteri, si ha una **distribuzione** che è detta **doppia o bivariata**. I dati possono essere riportati **in forma tabellare**, sia **grafica**.

Se il numero di dati è ridotto, la distribuzione doppia può essere rappresentata in una tabella che riporta in modo dettagliato tutti i valori delle due variabili, indicate con X e Y nel caso della regressione e con  $X_1$  e  $X_2$  nel caso della correlazione, come in quella seguente:

Unità	Carattere X o $X_1$	Carattere Y o $X_2$
1	$X_1$	$Y_1$
2	$X_2$	$Y_2$
---	---	---
i	$X_i$	$Y_i$
---	---	---
n	$X_n$	$Y_n$

Se il numero di osservazioni è grande, non è più possibile né conveniente fornire un lungo elenco nominativo: la sua lettura sarebbe troppo dispersiva e renderebbe ardua l'interpretazione delle tendenze fondamentali dei dati campionari. Si ricorre quindi ad una sintesi tabellare, chiamata **distribuzione doppia di frequenze**, come quella riportata nella pagina seguente.

Si suddividono le unità del collettivo in **modalità o classi** per entrambi i caratteri ( $X_i$  e  $Y_j$ ); successivamente

- si riportano quelle del primo carattere nella **testata** e
- quelle del secondo nella **colonna madre**

evidenziando, nelle caselle collocate al loro incrocio,

- il **numero di misure** che appartengono contestualmente ad entrambe le **modalità** ( $n_{ij}$ ).

		TESTATA							Totali
		$X_1$	$X_2$	$X_3$	...	$X_i$	...	$X_n$	
COLONNA	$Y_1$	$a_{11}$	$a_{12}$	$a_{13}$	...	$a_{1i}$	...	$a_{1n}$	$N_1$
	$Y_2$	$a_{21}$	$a_{22}$	$a_{23}$	...	$a_{2i}$	...	$a_{2n}$	$N_2$
	$Y_3$	$a_{31}$	$a_{32}$	$a_{33}$	...	$a_{3i}$	...	$a_{3n}$	$N_3$
MADRE	...	...	...	...	...	...	...	...	...
	$Y_j$	$a_{j1}$	$a_{j2}$	$a_{j3}$	...	$a_{ji}$	...	$a_{jn}$	$N_j$
	...	...	...	...	...	...	...	...	...
	$Y_m$	$a_{m1}$	$a_{m2}$	$a_{m3}$	...	$a_{mi}$	...	$a_{mn}$	$N_m$
	<b>Totali</b>	$M_1$	$M_2$	$M_3$	...	$M_i$	...	$M_n$	$T$

I totali delle righe ( $N_j$ ) e delle colonne ( $M_j$ ) rappresentano due distribuzioni semplici e sono dette **distribuzioni marginali** della distribuzione doppia. E' ovvio che i due collettivi (i totali di riga e i totali di colonna) devono avere lo stesso numero di unità, coincidente con il totale generale T.

Le frequenze riportate in una colonna o in una riga qualsiasi, come le frequenze nella colonna delle varie Y con  $X_2$  o quelle nella riga delle varie X con  $Y_3$ , sono dette **distribuzioni parziali** della distribuzione doppia.

Nello schema di tabella sopra riportata, sono presenti due distribuzioni marginali e 10 distribuzioni parziali, 5 per riga e 5 per colonna.

Quando il numero di modalità è molto grande, si può ricorrere al **raggruppamento in classi**. E' effettuato in modo indipendente per le due variabili, con i metodi già descritti per una sola variabile nel I capitolo sulla statistica descrittiva. Non è assolutamente richiesto che il numero di classi o il passo siano uguali per la testata e per la colonna madre: le due variabili sono trattate in modo simultaneo, ma come due distribuzioni indipendenti.

Una **distribuzione doppia di quantità** può essere rappresentata graficamente in vari modi. I metodi più frequentemente utilizzati sono due:

- gli **istogrammi**, quando si riportano le frequenze dei raggruppamenti in classi; il metodo è uguale a quello utilizzato anche nelle distribuzioni di conteggi con dati qualitativi, come già dimostrato nelle tabelle  $m \times n$  del test  $\chi^2$ ;
- il **diagramma di dispersione** (chiamato nei programmi informatici *scatter plot*) quando si rappresentano con **punti** in un piano cartesiano le singole coppie di misure osservate. Si ottiene una **nuvola di punti**, che descrive in modo visivo la relazione tra le due variabili. E' particolarmente utile con valori continui, quando i dati sono misurati in scale d'intervalli o di rapporti.

ESEMPIO 1 (tabella ed istogramma). Lo studio e la classificazione tassonomica di varie specie di Macrobiotidi si fonda sia su aspetti qualitativi sia sui rapporti tra gli arti ed i loro segmenti. Di norma, si ha una bassa variabilità entro la stessa specie e una forte variabilità tra le specie.

Per 45 animali presumibilmente della stesso gruppo *Macrobiotus hufelandi*, ma con forti dubbi sull'attribuzione alla specie per le difficoltà di classificazione dovute alla compresenza di giovani ed adulti difficilmente distinguibili, sono state misurate al microscopio le dimensioni (in  $\mu\text{m}$ ) di varie parti dello scheletro; tra esse le dimensioni della prima e della seconda placca, che sono state riportate nella tabella seguente.

Animali	I Placca	II Placca
1	31	22
2	31	21
3	28	20
4	33	24
--	--	--
45	32	23

Prima di analizzare la relazione esistente tra le due misure con i metodi dell'inferenza, è utile fornire una descrizione tabellare o grafica dei valori rilevati.

Per evitare una o più pagine di numeri di difficile interpretazione, l'elevato numero di osservazioni impone il ricorso ad una rappresentazione più sintetica, che può essere ottenuta con una tabella. Per i valori della prima variabile (riportata in testata) e della seconda (riportata nella colonna madre), si formano le distribuzioni di frequenza, con modalità analoghe a quelle della statistica univariata.

Quando le caselle sono eccessivamente numerose per essere riportate in una tabella di dimensioni unitarie, come quella successiva, si ricorre al raggruppamento in classi di una sola variabile o di entrambe.

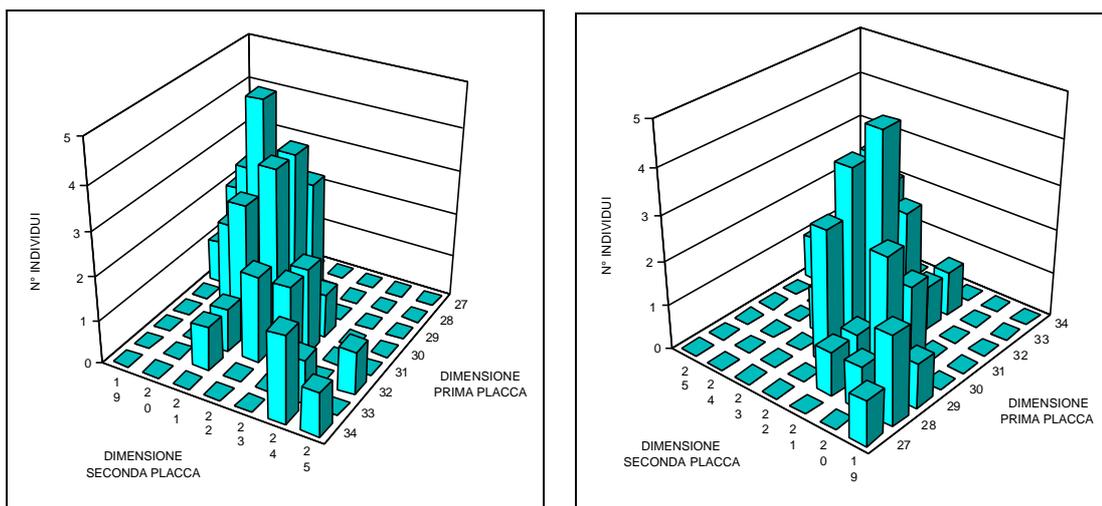
		Dimensione della I placca (in $\mu\text{m}$ )								Totale
		27	28	29	30	31	32	33	34	
Dimensione della II Placca (in $\mu\text{m}$ )	19	1	2	1	0	0	0	0	0	4
	20	0	1	3	2	0	0	0	0	6
	21	0	1	1	5	3	1	1	0	12
	22	0	0	3	4	4	2	0	0	13
	23	0	0	0	1	2	2	0	0	5
	24	0	0	0	0	0	0	1	2	3
	25	0	0	0	0	0	1	0	1	2
	Totale	1	4	8	12	9	6	2	3	45

Per esempio, se le misure in  $\mu\text{m}$  della prima placca avessero avuto un campo di variazione da 26 a 60 (35  $\mu\text{m}$ ) e quelle della seconda placca da 15 a 49 (35  $\mu\text{m}$ ), non sarebbe stato conveniente formare una tabella di dimensioni 35 x 35, con 1225 caselle. Il numero di caselle sarebbe stato di gran lunga più alto del numero di individui che formano il campione, con il risultato che la quasi totalità delle caselle sarebbero state vuote ed alcune avrebbero avuto frequenze molto ridotte, spesso una sola osservazione.

Per ottenere un effetto di raggruppamento dei dati ed una distribuzione tendenzialmente normale, si devono formare 4 o 5 raggruppamenti, sia per una variabile che per l'altra, con 16-25 caselle in totale.

La **tabella** fornisce una prima serie di informazioni elementari, presentate nel capitolo sulla statistica descrittiva. Con le misure riportate nell'esempio, la semplice lettura della tabella evidenzia come a valori crescenti della prima variabile corrispondano valori in aumento anche nella seconda: i dati risultano concentrati lungo la diagonale, sottolineando come la distribuzione facilmente non sia casuale. Se le due variabili fossero totalmente indipendenti, le 45 misure risulterebbero disperse in tutta la tabella, con un addensamento a forma circolare od ellissoidale verso il centro, in rapporto ai valori delle loro varianze.

Gli **istogrammi di una distribuzione bivariata**, di semplice realizzazione con programmi informatici, presentano alcuni inconvenienti rispetto a quelli che sono utilizzati per una sola variabile (vedi cap. I).



Le figure evidenziano le caratteristiche fondamentali, che appaiono di lettura più facile ed immediata rispetto alla tabella, con le altezze che sono proporzionali alle frequenze riportate nelle caselle.

In realtà, come la foto di un quartiere formato da grattacieli ripresi da una strada diversa, non è una rappresentazione oggettiva; spesso si deve scegliere un angolo di visuale, che mostra i dati in prima linea, ma nasconde quelli collocati dietro gli istogrammi più alti.

I due istogrammi bidimensionali appena riportati sono stati costruiti con gli stessi valori. Appaiono diversi e sarebbe possibile costruirne altri ancora, visivamente differenti da questi e tra loro, partendo da diversi partendo da angolazioni differenti (è stata invertita la scala per la I placca).

**E' quindi una rappresentazione da evitare se, partendo dagli stessi dati, è possibile fornire impressioni differenti sulle loro caratteristiche statistiche.**

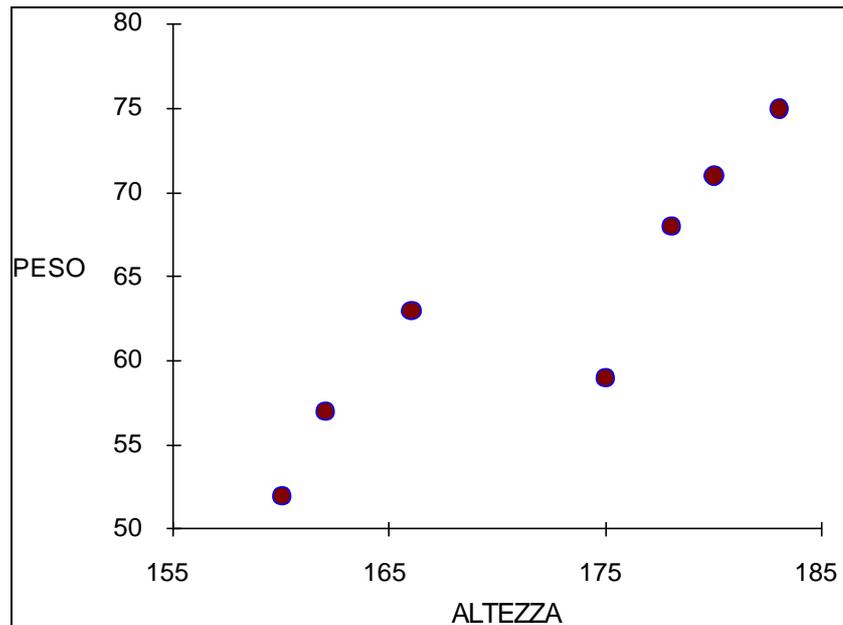
ESEMPIO 2 (diagramma di dispersione). Quando i dati sono espressi in una scala continua è conveniente una rappresentazione grafica mediante diagramma di dispersione. La coppia di dati riferiti ad ogni individuo sono riportati su un grafico bidimensionale ed indicati da un punto, le cui coordinate corrispondono al valore di X sull'asse delle ascisse e di Y su quella delle ordinate.

Con le misure di peso (in Kg.) e di altezza (in cm.) di 7 giovani, come riportato nella tabella,

Individui	1	2	3	4	5	6	7
Peso (Y)	52	68	75	71	63	59	57
Altezza (X)	160	178	183	180	166	175	162

è possibile costruire il diagramma, detto diagramma di dispersione, che evidenzia con maggiore chiarezza ed immediatezza la relazione esistente tra le due variabili,

- sia nella sua **tendenza generale, indicata da una retta** (al crescere di una variabile aumenta anche l'altra),
- sia nella individuazione dei **dati che se ne distaccano** (come l'individuo 6 di altezza 175 cm. e 59 Kg. di peso).



Il grafico può essere costruito anche con centinaia o migliaia di punti. La chiarezza del grafico dipende solo dalle loro dimensioni, che ovviamente devono essere inversamente proporzionali al loro numero. Il disegno risulta ugualmente chiaro ed il grafico è leggibile, sia quando i dati sono pochi che molto numerosi.

Nel caso di rappresentazione di dati riportati originariamente in tabella di frequenza, nel diagramma di dispersione **il diverso numero di ricorrenze può essere indicato da simboli convenzionali**. Una modalità usata con frequenza è la costruzione di più cerchi concentrici o con superfici differenti, in cui il numero di cerchi e/o le loro dimensioni sono proporzionali al **numero di dati** che si vuole rappresentare.

### 12.3. MODELLI DI REGRESSIONE

Il diagramma di dispersione fornisce una descrizione visiva, completa e dettagliata della relazione esistente tra due variabili; tuttavia, la sua interpretazione resterebbe soggettiva. Come per la forma di una singola distribuzione, è necessario tradurre le caratteristiche evidenziate dal grafico in numeri, che permettano a tutti di giungere alle medesime conclusioni a partire dagli stessi dati.

La funzione matematica che può esprimere in modo oggettivo la relazione di causa-effetto tra due variabili è chiamata **equazione di regressione** o **funzione di regressione della variabile Y sulla variabile X**.

Il termine **regressione** fu introdotto verso la metà dell'ottocento dall'inglese Sir **Francis Galton** (1822-1911) che, nei suoi studi di eugenica, voleva verificare se la statura dei figli potesse essere predeterminata sulla base di quella dei genitori, esprimendo questa corrispondenza in una legge matematica. In circa 200 coppie in cui aveva potuto misurare l'altezza del padre e quella di un figlio maschio di circa 20 anni, egli osservò che l'altezza dei padri e quella dei figli avevano medie uguali, entrambe 68 pollici (un pollice è uguale a 2,54 cm.) e deviazione standard uguali ( $s = 2,5$  in pollici); ma fu colpito dal fatto che i padri con altezza superiore a 73 pollici avevano figli con un'altezza media minore di 73 pollici, mentre i padri che avevano un'altezza inferiore a 63 pollici avevano figli con un'altezza media maggiore di 63. Chiamo questo fenomeno, per cui padri alti e bassi hanno figli più vicini alla media del gruppo, **regressione verso la media**.

Se avesse analizzato l'altezza dei padri in rapporto a quella dei figli, avrebbe simmetricamente trovato che i figli più bassi e quelli più alti hanno genitori con un'altezza più vicina alla media del loro gruppo. Non cogliendo questo secondo aspetto, anche perché illogico nella relazione tra causa ed effetto, fu colpito dal supposto fenomeno di maggiore omogeneità dei figli rispetto ai genitori ed affermò che la dispersione o variabilità della statura tende a regredire, scendendo da valori estremi verso la media. Per un genetista, o meglio uno studioso di eugenetica che si poneva il problema di come ottenere individui "migliori", il fenomeno fu visto come negativo: era una regressione della popolazione verso una uniformità, che non permette di selezionare i migliori.

Per comprendere esattamente il significato della ricerca statistica nel contesto storico del periodo, ad esemplificazione del legame che spesso esiste tra analisi dei dati e loro interpretazione estesa ai valori sociali e culturali, è importante rileggere alcuni passi sulla biografia di Francis Galton, che tra i suoi allievi ebbe anche Karl Pearson, a sua volta maestro di Fisher e di Gosset, vero cognome di Student (vedi: *Statistica non parametrica*, seconda edizione, di **Sidney Siegel e John Castellan**, McGraw-Hill, Milano, a pag. XXXII e segg.).

Galton, di famiglia nobile inglese, nono figlio di un famoso e facoltoso banchiere, era cugino di Darwin. *L'origine della specie* ebbe su di lui una profonda impressione. L'opera del filosofo Herbert Spencer (*Principles of Biology*, 1872, alla quale viene attribuito l'inizio del "darwinismo sociale"), che

asseriva che la selezione naturale avviene attraverso una vera e propria “lotta al coltello”, dalla quale sopravvive solo il più forte, colui che riesce a dominare tutti gli altri, lo convinse ad interessarsi della trasmissione dei caratteri da una generazione all’altra.

E’ noto che, a quei tempi, l’opera di Mendel non fosse conosciuta. Anche se il monaco di Brno aveva pubblicato la sua ricerca nel 1865, quella ricerca sul *pisum sativum* (i piselli da orto) era passata del tutto inosservata nell’ambiente scientifico dell’epoca. La riscoperta avvenne solo nel 1910. Fu nell’ambito del darwinismo sociale che nacque la dialettica ereditarietà - ambiente. In quei decenni e per altri ancora, si dibatteva su quanto nell’evoluzione delle “facoltà mentali” fosse dovuto all’ambiente e quanto all’ereditarietà.

Galton era convinto che i “caratteri morali” si ereditassero, così come certe malattie. Cercò quindi di dimostrare che un uomo, “alla nascita e per nascita”, è già tutto quello che sarà in avvenire: cercò di dimostrare una teoria tanto di moda nel suo ambiente, la teoria detta del “sangue blu”, secondo la quale non si ereditano solo i titoli nobiliari e le proprietà, ma anche le caratteristiche fisiche come l’altezza e il colore degli occhi; nello stesso modo, per le stesse leggi, si ereditano l’intelligenza e i valori morali, il talento e l’onestà.

Passando dagli individui alle popolazioni, Galton era anche convinto che le “razze umane” e le “classi sociali inferiori”, non potessero elevare le loro caratteristiche mentali e morali attraverso l’educazione, in quanto non possedevano le qualità biologiche per attuare tale evoluzione. Esse potevano solo essere “migliorate”, attraverso “incroci” con coloro che già possedevano tali caratteri. Fu quindi tra i fondatori, per gli spetti concettuali e metodologici, dell’eugenetica, finalizzata a migliorare le “razze e le classi inferiori” attraverso misure tese ad evitare il diffondersi dei caratteri ereditari indesiderati. Alcune norme sui matrimoni, presupposto all’aver figli, e sulla migrazioni risentono di queste convinzioni, diffuse nelle classi economiche, sociali, culturali e politiche dominanti.

In seguito, dal suo significato originario di "ritornare indietro" verso la media e verso “la mediocrità”, il termine regressione assunse solo quello neutro di funzione che esprime matematicamente la relazione tra la **variabile attesa o predetta o teorica**, indicata con Y, e la **variabile empirica od attuale**, indicata con X.

La forma più generale di una equazione di regressione è

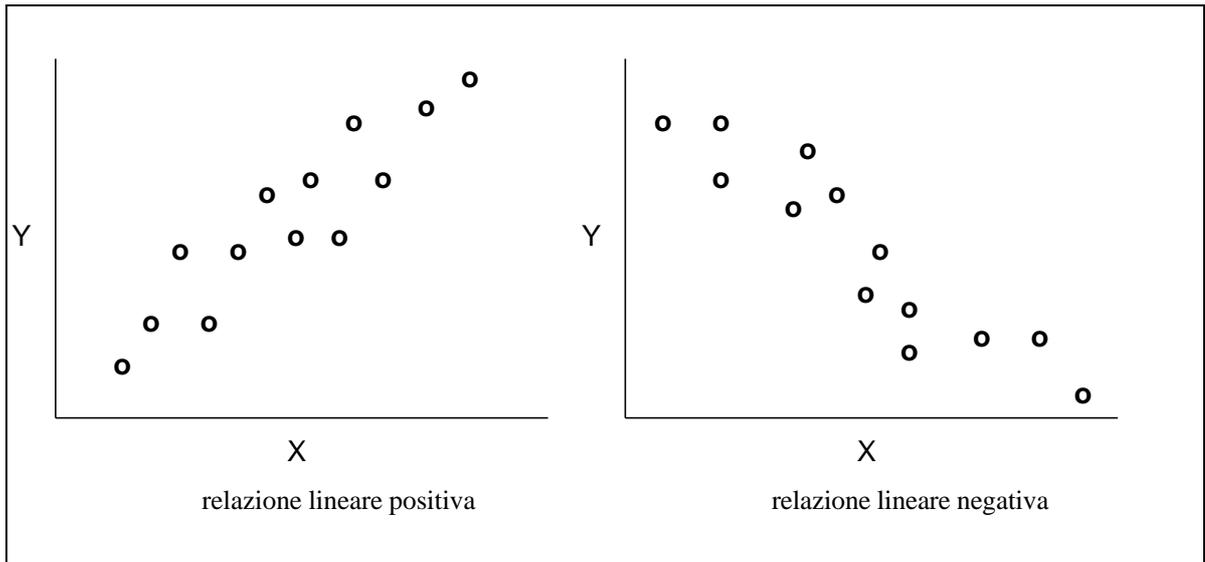
$$Y = a + bX + cX^2 + dX^3 + eX^4 + \dots$$

dove il secondo membro è un polinomio intero di X.

L'approssimazione della curva teorica ai dati sperimentali è tanto migliore quanto più elevato è il numero di termini del polinomio: con **n** punti, una curva di grado **n-1** passa per tutti i punti.

Ma l’ecologo e l’ambientalista non possono limitarsi alla ricerca della funzione matematica che meglio descrive i dati raccolti con un solo campione: devono soprattutto fornirne una interpretazione logica del fenomeno, con argomenti tratti dalla disciplina.

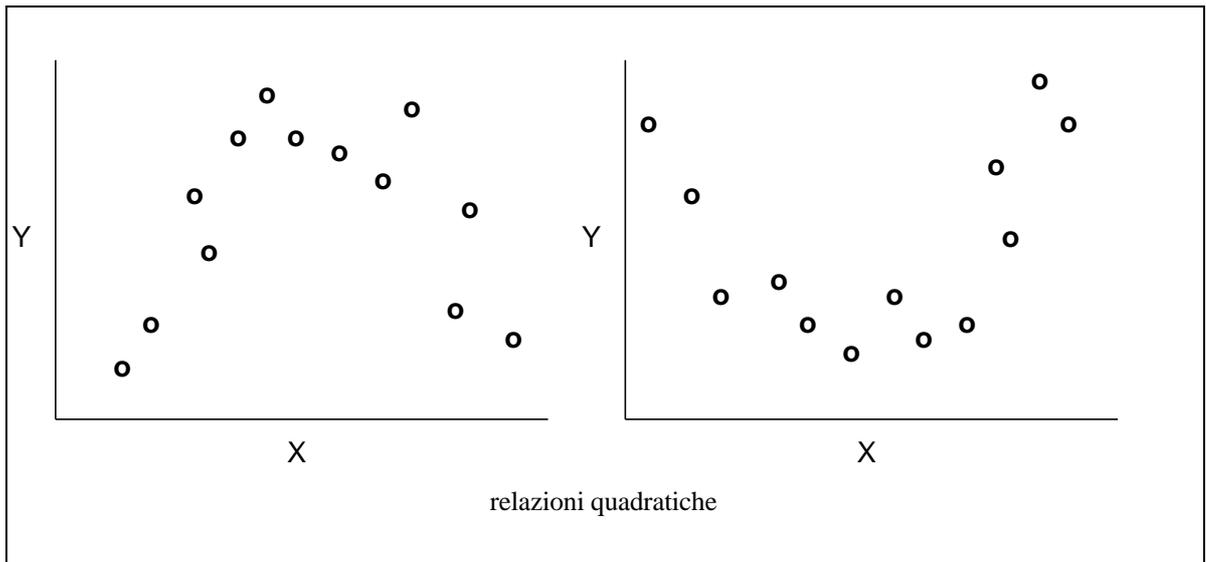
Quasi sempre l'interpretazione dell'equazione di regressione è tanto più attendibile e generale quanto più la curva è semplice, come quelle di primo o di secondo grado. Regressioni di ordine superiore sono quasi sempre legate alle variazioni casuali; sono effetti delle situazioni specifiche del campione raccolto e solo molto raramente esprimono relazioni reali, non accidentali, tra le due variabili. Di conseguenza, anche l'ecologo e l'ambientalista, come tutti coloro che ricorrono alla statistica applicata nell'ambito della loro disciplina, utilizzano quasi esclusivamente regressioni lineari (di primo ordine) o le regressioni curvilinee (di secondo ordine).



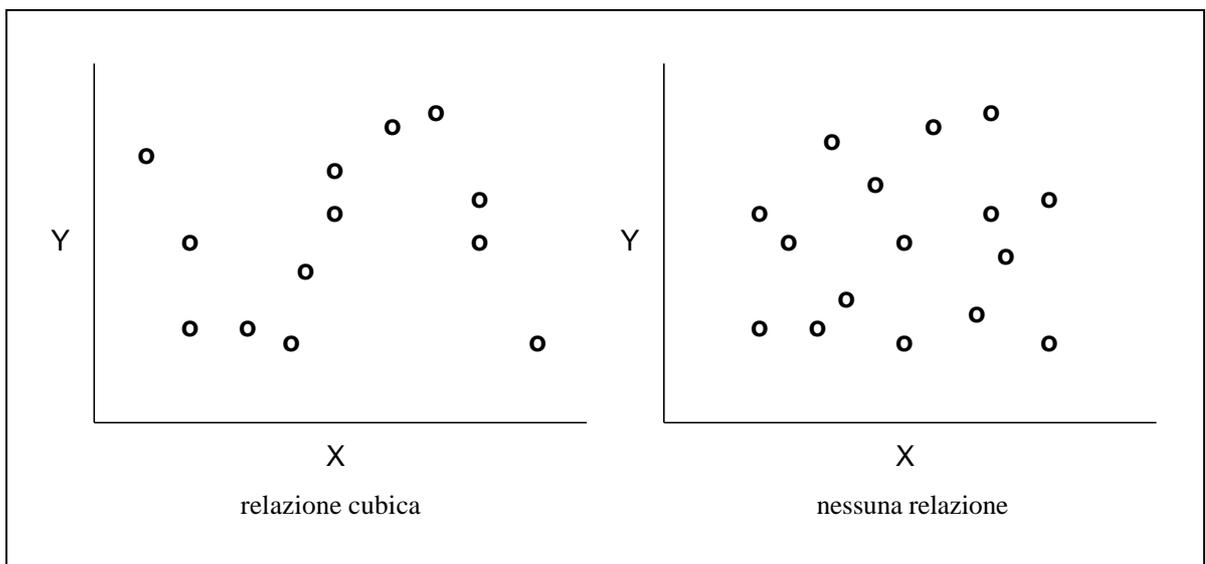
La regressione lineare può essere **positiva** o **negativa**:

- nel primo caso, all'aumento dei valori di una variabile corrisponde un aumento anche nell'altra;
- nel secondo, all'aumento dell'una corrisponde una diminuzione dell'altra.

E' la relazione più semplice e frequente tra due variabili quantitative.



Oltre alle forme a parabola rappresentate nell'ultima figura, la regressione curvilinea di secondo grado può seguire vari altri modelli, come l'iperbole, l'esponenziale e la logaritmica. Sono fenomeni frequenti in natura e di interpretazione semplice: una sostanza può determinare effetti positivi a dosi basse ed effetti fortemente decrescenti o stabili a dosi in aumento.



Le curve e le relazioni cubiche (di terzo ordine) e quelle di ordine superiore rappresentano rapporti tra due variabili che sono eccessivamente complessi per una interpretazione generale. Ad esempio, come suggerisce la prima delle due ultime figure riportate (relazione cubica), è molto raro trovare una sostanza che a dosi crescenti determina una contrazione della seconda variabile nelle fasi iniziali, per causare un suo aumento in una seconda fase e successivamente una nuova diminuzione, continuando sempre ad aumentare la dose.

#### 12.4. LA REGRESSIONE LINEARE SEMPLICE

La relazione matematica più semplice tra due variabili è la regressione lineare semplice, rappresentata dalla retta

$$\hat{Y}_i = a + bX_i$$

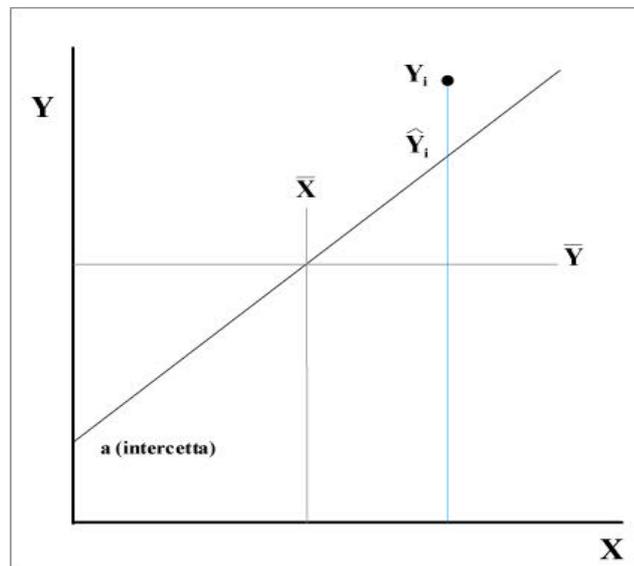
dove

- $\hat{Y}_i$  è il valore stimato per il valore  $X$  dell'osservazione  $i$ ,
- $X_i$  è il valore empirico di  $X$  per l'osservazione  $i$ ,
- $a$  è l'intercetta della retta di regressione,
- $b$  è il coefficiente angolare ed indica la quantità di cui varia  $Y$  al variare di una unità di  $X$ .

La rappresentazione grafica evidenzia che il termine costante  $a$ , chiamato **intercetta**, fissa la posizione della retta rispetto all'asse delle ordinate:

- $a$  è il valore di  $Y$ , quando  $X$  è uguale a 0.

Due rette che differiscano solo per il valore di  $a$ , quindi con  $b$  uguale, sono tra loro parallele.



Come evidenzia il diagramma cartesiano precedente, ogni punto sperimentale ha una componente di errore  $e_i$ , che rappresenta lo scarto verticale del valore osservato dalla retta (quindi tra la  $Y$  osservata e quella proiettata perpendicolarmente sulla retta). Utilizzare un qualsiasi punto sperimentale per stimare  $a$  porterebbe ad avere tante stime diverse quanti sono i punti sperimentali, tutti affetti appunto da un errore diverso. Di conseguenza, come punto di riferimento per stimare  $a$  e costruire la retta, viene

utilizzato il punto identificato dai valori medi di Y e di X ( $\bar{Y}$  e  $\bar{X}$ ), che rappresenta il baricentro della distribuzione (dal quale essa passerà sempre).

Nel calcolo della retta di regressione, l'intercetta **a** è stimata a partire da **b** e dalle medie delle variabili X e Y sulla base della relazione

$$a = \bar{Y} - b\bar{X}$$

Di conseguenza, l'unica reale incognita è il valore del coefficiente angolare **b**.

Per calcolare la retta che meglio approssima la distribuzione dei punti, è utile partire dall'osservazione che ogni punto osservato  $Y_i$  si discosta dalla retta di una certa quantità  $e_i$  detta errore o **residuo**

$$Y_i = a + bX_i + e_i$$

Ognuno di questi valori  $e_i$  **può essere positivo** oppure **negativo**:

- è positivo quando il punto  $Y_i$  sperimentale è sopra la retta (come nella figura precedente),
- è negativo quando il punto  $Y_i$  sperimentale è sotto la retta.

Per costruire la retta che descrive la distribuzione dei punti, i principi ai quali riferirsi possono essere differenti e da essi derivano metodi diversi.

Gli statistici hanno scelto il **metodo dei minimi quadrati**. La retta scelta è quella che riduce al minimo **la somma dei quadrati degli scarti di ogni punto dalla sua proiezione verticale** (parallelo all'asse delle Y). E' un valore del tutto identico alla devianza e permette analisi simili a quelle dell'ANOVA.

In modo più formale, indicando con

- $Y_i$  il valore osservato od empirico e con
- $\hat{Y}_i$  il corrispondente valore sulla retta,

si stima come **migliore interpolante**, quella che minimizza la sommatoria **del quadrato** degli scarti dei valori osservati ( $Y_i$ ) rispetto a quelli stimati sulla retta ( $\hat{Y}_i$ )

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \text{minimo}$$

Poiché

$$e_i = Y_i - (a + bX_i)$$

è possibile scrivere

$$\sum e_i^2 = \sum (Y_i - (a + bX_i))^2 = \text{minimo}$$

e da essa

$$\sum e_i^2 = \sum (Y_i - \bar{Y} - b(X_i - \bar{X}))^2 = \text{minimo}$$

Eguagliando a zero le derivate parziali, si trova il valore di **b** che minimizza tale sommatoria

$$\sum (X - \bar{X})^2 \cdot \left[ b - \frac{\sum (X - \bar{X}) \cdot (Y - \bar{Y})}{\sum (X - \bar{X})^2} \right]^2 + \sum (Y - \bar{Y})^2 - \frac{[\sum (X - \bar{X}) \cdot (Y - \bar{Y})]^2}{\sum (X - \bar{X})^2}$$

Dopo semplificazione, il valore di **b** risulta uguale al

**rapporto della covarianza di X e Y con la devianza di X,**

che è più facile ricordare come

$$b = \frac{\text{Cod}_{xy}}{\text{Dev}_x}$$

La **covarianza** è un concetto non ancora incontrato nel corso di statistica, poiché serve nello studio di due variabili: stima come **X** e **Y** variano congiuntamente, rispetto al loro valore medio. E' definita come la **sommatoria dei prodotti degli scarti di X rispetto alla sua media e di Y rispetto alla sua media:**

$$\text{Cod}_{xy} = \sum (X - \bar{X}) \cdot (Y - \bar{Y})$$

Come per la devianza, anche la covarianza ha una formula empirica od abbreviata che permette un calcolo più rapido

$$\text{Cod}_{xy} = \sum (x \cdot y) - \frac{\sum x \cdot \sum y}{n}$$

e preciso a partire dai dati campionari.

Infatti evita l'uso delle medie, che sono quasi sempre valori approssimati e impongono di trascinare nei vari calcoli alcuni decimali.

In conclusione, il coefficiente angolare **b** è calcolato dalle coppie dei dati sperimentali X e Y come

$$b = \frac{\sum (X - \bar{X}) \cdot (Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

che ne definisce il significato,

oppure dalla equivalente formula rapida od empirica

$$b = \frac{\sum (x \cdot y) - \frac{\sum x \cdot \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

Dopo aver calcolato **b**, si stima **a**:

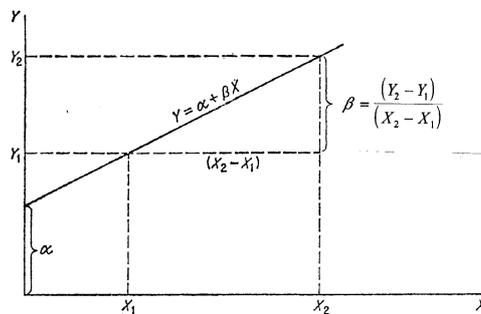
$$a = \bar{Y} - b\bar{X}$$

Noti i valori dell'**intercetta a** e del coefficiente **angolare b**, è possibile procedere alla **rappresentazione grafica della retta**.

Anche a questo scopo, è importante ricordare che la retta passa sempre dal baricentro del diagramma di dispersione, individuato dal punto d'incontro delle due medie campionarie  $\bar{X}$  e  $\bar{Y}$ .

Di conseguenza, è sufficiente calcolare il valore di  $\hat{Y}_i$  corrispondente ad uno solo e qualsiasi valore di  $X_i$  (ovviamente diverso dalla media), per tracciare la retta che passa per questo punto calcolato e per il punto d'incontro tra le due medie.

**Se non sono stati commessi errori di calcolo, qualsiasi altro punto  $\hat{Y}_i$  stimato nella rappresentazione grafica deve risultare collocato esattamente sulla retta tracciata.** E' un procedimento semplice, che può essere utilizzato per verificare la correttezza di tutti i calcoli effettuati.



ESEMPIO. Per sette giovani donne, indicate con un numero progressivo, è stato misurato il peso in Kg e l'altezza in cm.

<b>Individui</b>	1	2	3	4	5	6	7
<b>Peso (Y) in Kg.</b>	52	68	75	71	63	59	57
<b>Altezza (X) in cm.</b>	160	178	183	180	166	175	162

Calcolare la retta di regressione che evidenzi la relazione tra peso ed altezza.

Risposta.

Occorre individuare quale è la variabile indipendente, che deve essere indicata con **X**, e quale la variabile dipendente, indicata con **Y**. Tra queste due serie di misure, la variabile indipendente è l'altezza e la variabile dipendente è il peso. Infatti ha significato stimare quanto dovrebbe pesare un individuo in rapporto alla sua altezza, ma non viceversa.

Successivamente, dalle 7 coppie di dati si devono calcolare le quantità

$$\sum (X \cdot Y) = 76945; \quad \sum X = 1204; \quad \sum Y = 445; \quad \sum X^2 = 207598; \quad n = 7$$

che sono necessari per

- la stima del **coefficiente angolare b**

$$b = \frac{\sum (x \cdot y) - \frac{\sum x \cdot \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{76945 - \frac{1204 \cdot 445}{7}}{207598 - \frac{1204^2}{7}} = 0,796$$

che risulta uguale a **0,796**

- la stima dell'**intercetta a**

$$a = \bar{Y} - b\bar{X} = 63,571 - 0,796 \cdot 172 = -73,354$$

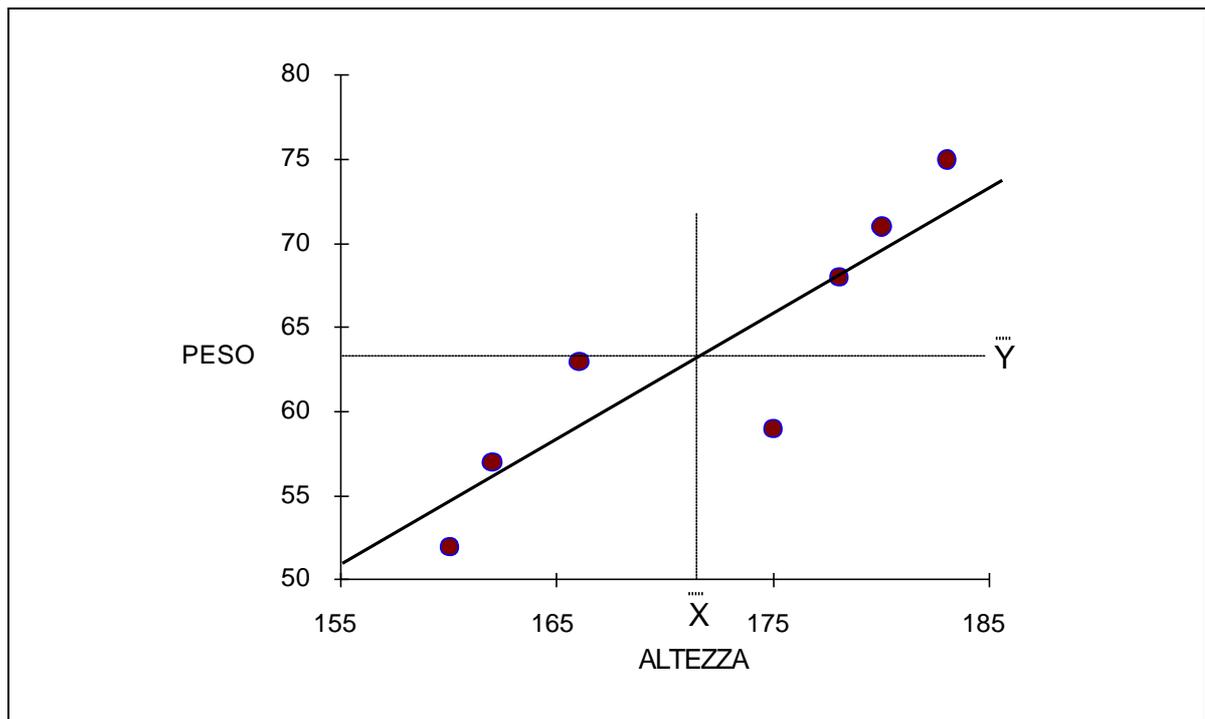
che risulta uguale a **-73,354**.

Si è ottenuta la **retta di regressione**

$$\hat{Y}_i = -73,354 + 0,796 \cdot X_i$$

con la quale è possibile stimare i punti sulla retta, corrispondenti a quelli sperimentalmente rilevati.

**Per tracciare la retta è sufficiente calcolare un solo altro punto**, oltre quello noto individuato dall'incrocio delle due medie, che identifica il baricentro della distribuzione. Di norma, ma non necessariamente, è scelto entro il campo di variazione delle  $X_i$  empiriche; successivamente, si deve prolungare il segmento che per estremi ha il punto stimato ed il baricentro della distribuzione, come nella figura di seguito riportata.



Qualsiasi altro valore di  $\hat{Y}_i$ , stimato a partire da un generico  $X_i$ , sarà collocato su questa retta, se non sono stati commessi errori di calcolo in una fase qualsiasi del procedimento.

Nel sua interpretazione biologica, il valore calcolato di **b** indica che in media gli individui che formano il campione aumentano di **0,796 Kg. al crescere di 1 cm. in altezza.**

E' quindi ovvio che, se l'altezza delle 7 giovani fosse stata misurata in metri (1,60; 1,78; ...), il coefficiente angolare **b** sarebbe risultato uguale a 79,6 (cento volte il valore precedentemente stimato), indicando l'incremento di 79,6 kg. per l'aumento di 1 metro in altezza.

Nello stesso modo e simmetricamente, se il peso fosse stato stimato in ettogrammi (520, 680, ...) e l'altezza sempre in centimetri, il coefficiente angolare **b** sarebbe risultato uguale a 7,96 indicando un aumento medio del peso di hg. 7,96 per un aumento di 1 cm in altezza.

Sono concetti utili, quando si devono confrontare due o più coefficienti angolari di rette di regressione e fornire interpretazioni di tipo ecologico od ambientale.

## 12.5. VALORE PREDITTIVO DELLA REGRESSIONE

La retta di regressione è sovente usata a scopi predittivi, per stimare una variabile conoscendo il valore dell'altra. Ma è necessario procedere con cautela: in questa operazione spesso viene dimenticato che, **sotto l'aspetto statistico, qualsiasi previsione o stima di Y è valida solamente entro il campo di variazione sperimentale della variabile indipendente X.**

Questo campo di variazione comprende solo i valori osservati della X, usati per la stima della regressione. Per valori minori o maggiori, non è assolutamente dimostrato che la relazione trovata tra le due variabili persista e sia dello stesso tipo.

L'ipotesi che la relazione stimata si mantenga costante anche per valori esterni al campo d'osservazione è totalmente arbitraria; **estrapolare i dati all'esterno del reale campo d'osservazione è un errore di tecnica statistica**, accettabile solamente nel contesto specifico della disciplina studiata, a condizione che sia motivato da una maggiore conoscenza del fenomeno. In alcuni casi, questo metodo è utilizzato appunto per dimostrare come la legge lineare trovata non possa essere valida per valori inferiori o superiori, stante l'assurdità della risposta.

Nell'esempio del paragrafo precedente, la relazione trovata tra Y e X con la retta di regressione è valida solamente entro un'altezza compresa tra 160 e 183 centimetri. E' da ritenere statisticamente errato usare la retta stimata per predire valori di Y in funzione di valori di X che siano minori di 160 o maggiori di 183 centimetri.

Come dimostrazione semplice di tale principio, nei vari testi di statistica sono riportati esempi anche divertenti, ma è possibile usare la retta calcolata.

Una bambina alla nascita di norma ha un'altezza (lunghezza) di circa 50 centimetri.

Che peso dovrebbe avere, se la relazione precedente fosse applicabile anche al suo caso?

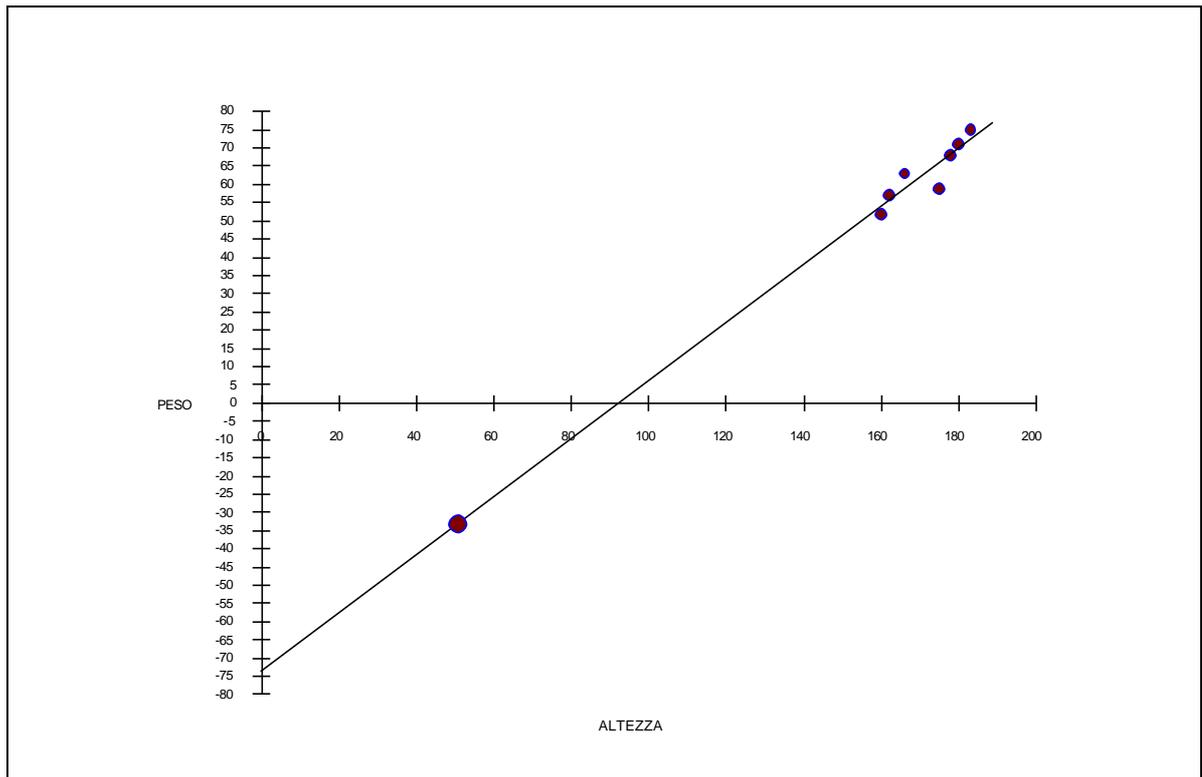
**La prosecuzione della retta stimata**

$$\hat{Y}_i = -73,354 + 0,796 \cdot X_i$$

per una **lunghezza ( $X_i$ ) uguale a 50 cm.**

$$-73,354 + 0,796 \cdot 50 = -33,554$$

fornisce un **peso medio ( $Y_i$ ) uguale a Kg. -33,554.**



E' una risposta chiaramente assurda, evidenziata nella figura, poiché la **relazione lineare calcolata per giovani da 160 a 183 cm. non può essere estesa a dimensioni diverse.** E' intuitivo che gli effetti saranno tanto più distorti, quanto maggiore è la distanza dai limiti sperimentali utilizzati per il calcolo della regressione.

Anche nella ricerca ambientale, l'**evoluzione temporale** e la **diffusione spaziale** di un fenomeno sono casi ricorrenti di uso della regressione lineare a fini predittivi. I dati, se ordinati secondo il periodo, sono chiamati **serie temporali o storiche**, mentre sono chiamate **serie territoriali**, quando ordinate sulla base della distanze dal luogo di rilevazione. Sono particolarmente importanti per lo studio dell'aumento (o della diminuzione) dei tassi di inquinamento ad iniziare da un certo momento oppure per analizzare la diffusione geografica di un inquinante a partire da una fonte.

Una serie temporale può essere scomposta in 4 componenti:

- la **componente di fondo**, detta **trend**, che ne rappresenta l'evoluzione più importante, a lungo termine;
- le **oscillazioni periodiche, stagionali, o cicliche** che si ripetono con regolarità ad intervalli costanti;
- le **variazioni casuali**, non riconducibili a nessuna causa costante;
- gli **eventi eccezionali**, che sono in grado di modificare le tendenze di medio o di lungo periodo.

Per esse e per le serie territoriali, tra i metodi specifici è utilizzata la regressione, in particolare per predire la tendenza di fondo.

Per approfondimenti sull'argomento delle serie storiche o territoriali, si rinvia a trattazioni specifiche.

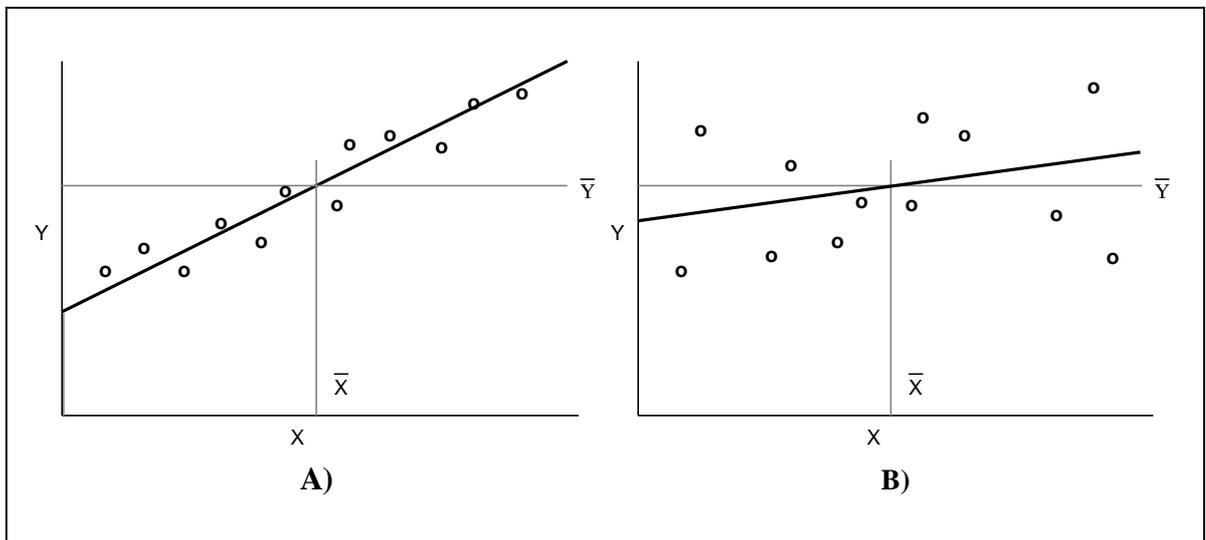
## 12.6. SIGNIFICATIVITÀ' DEI PARAMETRI $b$ E $a$ DELLA RETTA DI REGRESSIONE

Con le formule presentate, è sempre possibile ottenere la retta che meglio si adatta ai dati rilevati, con qualunque forma di dispersione dei punti.

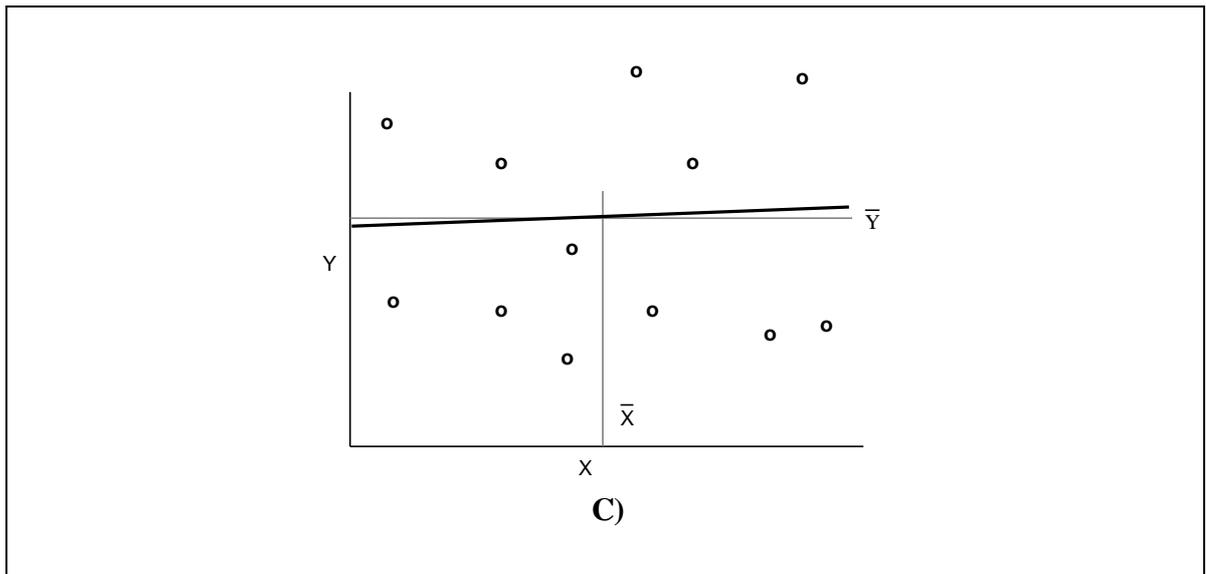
Tuttavia, allo statistico il semplice calcolo della retta non è sufficiente. Essa potrebbe indicare

- una **relazione reale** tra le due variabili, se la dispersione dei punti intorno alla retta è ridotta,
- una **relazione casuale o non significativa**, quando la dispersione dei punti intorno alla retta è approssimativamente uguale a quella intorno alla media.

Le tre figure successive (A, B, C) possono illustrare questi concetti con semplicità e chiarezza.



La figura A riporta una retta che, visivamente, esprime la relazione tra le due variabili: i punti hanno distanze dalla retta di regressione sensibilmente minori di quelle dalla media ( $\bar{Y}$ ). Conoscendo X, il valore stimato di Y può avvicinarsi molto a quello reale, rappresentato dal punto.



All'opposto, la figura C evidenzia una situazione in cui la retta calcolata non è un miglioramento effettivo della distribuzione dei punti rispetto alla media. In questo caso, la retta calcolata può essere interpretata come una variazione casuale della media: con questi dati, la retta ha una pendenza positiva; ma con un altro campione estratto dalla stessa popolazione o con l'aggiunta di un solo dato della stessa popolazione si potrebbe stimare un coefficiente angolare (**b**) negativo.

Il caso B raffigura una situazione di maggiore incertezza sulla significatività della retta calcolata; **la semplice rappresentazione grafica risulta insufficiente per decidere se all'aumento di X i valori di Y tendano realmente a crescere.**

E' sempre necessario ricorrere a metodi che, a partire dagli stessi dati, conducano tutti alle stesse conclusioni. Sono i test di inferenza. Per rispondere alle domande poste, occorre valutare la significatività della retta, cioè se il coefficiente angolare **b** si discosta da zero in modo significativo.

Il coefficiente angolare **b** è relativo al campione. La sua generalizzazione nella popolazione è indicata con  $\beta$  (beta) e la sua significatività è saggiata mediante la verifica dell'ipotesi nulla  $H_0$

$$H_0: \beta = 0$$

Rifiutando l'ipotesi nulla e senza altre indicazioni, si accetta l'ipotesi alternativa a due code  $H_1$

$$H_1: b \neq 0$$

Affermare che **b** è uguale a zero, nella regressione lineare significa che, al

**variare di X, Y resta costante, uguale al valore dell'intercetta a.**

Di conseguenza, **non esiste alcun legame di regressione o predittivo tra X e Y**, poiché la prima cambia mentre la seconda, che dovrebbe essere da essa determinata, resta costante.

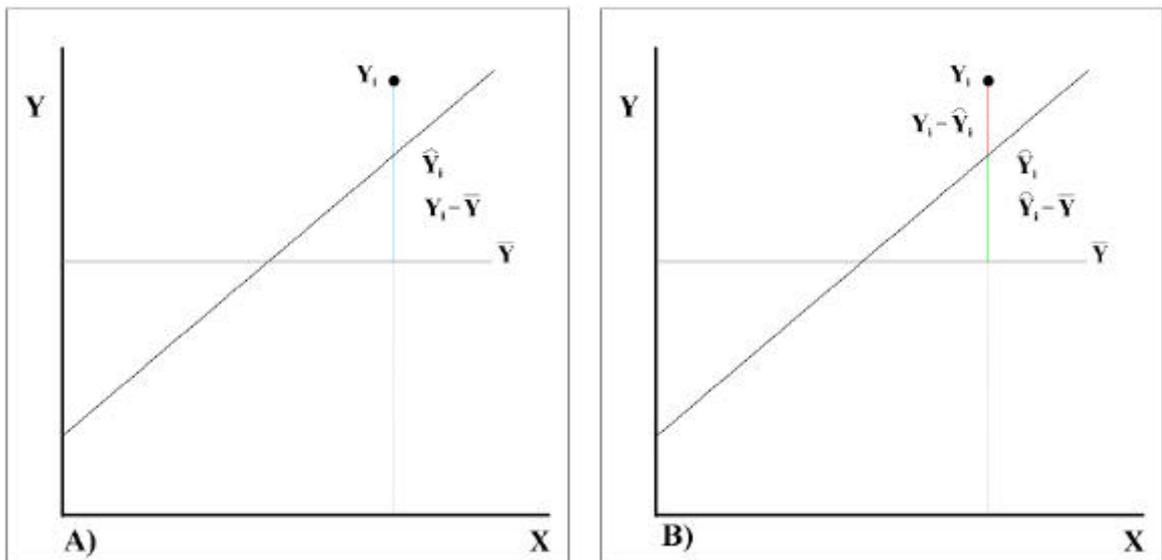
Rifiutando l'ipotesi nulla, implicitamente si accetta l'ipotesi alternativa  $H_1$  che  $b$  sia diverso da zero: al variare di  $X$  si ha una corrispondente variazione sistematica di  $Y$ . Di conseguenza, si afferma che la regressione esiste, perché conoscendo  $X$  si ha informazione non nulla sul valore di  $Y$ .

Per la verifica della significatività della retta calcolata, un metodo semplice e didatticamente utile alla comprensione del significato statistico della regressione è il **test F**, fondato sulla scomposizione delle devianze e dei relativi gdl.

Indicando con

- $Y_i$  il punto sperimentale,
- con  $\hat{Y}_i$  la sua proiezione (parallela all'asse delle ordinate) sulla retta,
- con  $\bar{Y}$  la media,

come nelle figure sottostanti A e B,



a partire dalla somma dei **quadrati delle distanze tra i tre punti** ( $Y$ ,  $\hat{Y}$  e  $\bar{Y}$ ) si definiscono **tre devianze**, come nell'analisi della varianza ad un criterio:

- la **devianza totale**, con gdl **n-1**,
- la **devianza della regressione** o devianza dovuta alla regressione, con gdl **1**,
- la **devianza d'errore** o devianza dalla regressione o residuo, con gdl **n-2**.

secondo le formule di seguito riportate con i relativi gdl:

- **Devianza totale**  $\sum (Y_i - \bar{Y})^2$  con gdl **n-1** (Fig. A)
- **Devianza della regressione**  $\sum (\hat{Y}_i - \bar{Y})^2$  con gdl **1** (Fig. B, parte inferiore)
- **Devianza d'errore**  $\sum (Y_i - \hat{Y}_i)^2$  con gdl **n-2** (Fig. B, parte superiore)

Queste formule definiscono **il significato** delle 3 devianze. Potrebbero essere usate per stimare i valori, ma richiedono calcoli lunghi e forniscono risultati approssimati, poiché fondati sulle medie e sui valori della retta, che non sono quasi mai valori esatti e impongono l'uso di decimali.

Per effettuare **in modo più rapido e preciso i calcoli**, si utilizzano le **formule abbreviate**:

- **Devianza totale**  $\sum Y^2 - \frac{(\sum Y)^2}{n}$  con gdl **n-1**

- **Devianza della regressione**  $\frac{\text{Cod}_{xy}^2}{\text{Dev}_x}$  con gdl **1**

ricordando che, sempre con le formule abbreviate,

$$\text{Cod}_{(x,y)} = \sum (x \cdot y) - \frac{\sum x \cdot \sum y}{n}$$

e

$$\text{Dev}_x = \sum X^2 - \frac{(\sum X)^2}{n}$$

Successivamente, per differenza, si calcola la devianza d'errore:

- **Devianza d'errore = (Devianza totale – Devianza della regressione )** con gdl **n-2**

Dal rapporto della devianza della regressione con i suoi gdl si stima **la varianza della regressione**; dal rapporto della devianza d'errore con i suoi gdl si ottiene **la varianza d'errore**.

**Se l'ipotesi nulla è vera, la varianza d'errore e la varianza della regressione stimano le stesse grandezze e quindi dovrebbero essere simili.**

Se invece esiste regressione ( $H_0$  falsa), la varianza della regressione è maggiore di quella d'errore.

Il rapporto tra queste due varianze determina il valore del test  $F$  con gdl  $1$  e  $n-2$

$$F_{(1, n-2)} = \frac{\text{Varianza della regressione}}{\text{Varianza d'errore}}$$

Teoricamente, quando l'ipotesi nulla è falsa, si ottengono valori significativamente maggiori di  $1$ .

In pratica, se il valore di  $F$  calcolato è inferiore al valore tabulato, relativo alla probabilità prefissata e ai gdl corrispondenti, si accetta l'ipotesi nulla: non si ha una regressione lineare statisticamente significativa.

Al contrario, se il valore calcolato di  $F$  supera il valore tabulato, si rifiuta l'ipotesi nulla e pertanto si accetta l'ipotesi alternativa: la regressione lineare tra le due variabili è significativa.

Gli stessi concetti possono essere espressi con termini più tecnici.

- Se  $b = 0$ , la varianza dovuta alla regressione e quella d'errore sono stime indipendenti e non vizzate della variabilità dei dati.
- Se  $b \neq 0$ , la varianza d'errore è una stima non viziata della variabilità dei dati, mentre la varianza dovuta alla regressione è stima di una grandezza maggiore.
- Di conseguenza, il rapporto tra le varianze (varianza d'errore/varianza della regressione) con d.f. rispettivamente  $1$  e  $n-2$  è da ritenersi utile alla verifica dell'ipotesi  $b = 0$ .

Il test applicato è detto anche **test di linearità**. Infatti, rifiutare l'ipotesi nulla non significa affermare che tra  $X$  e  $Y$  non esista alcuna relazione, ma solamente che **non esiste una relazione di tipo lineare** tra le due variabili. **Potrebbe esistere una relazione di tipo differente**, come quella curvilinea, di secondo grado o di grado superiore.

ESEMPIO. Con le misure di peso ed altezza rilevati su 7 giovani donne

Individui	1	2	3	4	5	6	7
Peso (Y) in Kg.	52	68	75	71	63	59	57
Altezza (X) in cm.	160	178	183	180	166	175	162

è stata calcolata la retta di regressione

$$\hat{Y} = -73,354 + 0,796 X$$

Valutare la sua significatività mediante il test F.

Risposta.

Valutare se esiste regressione tra le due variabili con il test F equivale a verificare l'ipotesi nulla

$$H_0: b = 0$$

con ipotesi alternativa

$$H_1: b \neq 0$$

Dopo i calcoli preliminari dei valori richiesti dalle formule abbreviate

$$\sum (X \cdot Y) = 76945 \quad \sum X = 1204 \quad \sum X^2 = 207598 \quad \sum Y = 445 \quad \sum Y^2 = 28693$$

$$n = 7$$

precedentemente riportate, si ottengono le tre devianze:

$$- \text{SQ totale} = 28693 - \frac{445^2}{7} = 28693 - 28289,285 = 403,715$$

$$- \text{SQ della regressione} = \frac{(76945 - \frac{1204 \cdot 445}{7})^2}{207598 - \frac{1204^2}{7}} = \frac{(76945 - 76540)^2}{207598 - 207088} = \frac{164025}{510} = 321,618$$

$$- \text{SQ d'errore} = 403,715 - 321,618 = 82,097$$

Per presentare in modo chiaro i risultati, è sempre utile riportare sia le tre devianze e i **df** relativi, sia le varianze rispettive, in una tabella riassuntiva,

	<b>Devianza</b>	<b>DF</b>	<b>Varianza</b>	<b>F</b>	<b>P</b>
<b>Totale</b>	<b>403,715</b>	<b>6</b>	----	---	
<b>Regressione</b>	<b>321,618</b>	<b>1</b>	<b>321,62</b>	<b>19,59</b>	<b>&lt;0.01</b>
<b>Errore</b>	<b>82,097</b>	<b>5</b>	<b>16,42</b>	---	

che fornisce tutti gli elementi utili al calcolo e all'interpretazione di **F**.

Con i dati dell'esempio, il valore di F

$$F_{(1,5)} = \frac{321,62}{16,42} = 19,59$$

risulta uguale a **19,59** con df **1** e **5**.

I valori critici riportati nelle tavole sinottiche di **F** per df **1** e **5** sono

- **6,61** alla probabilità  $\alpha = 0.05$
- **16,26** alla probabilità  $\alpha = 0.01$ .

Il valore calcolato è superiore a quello tabulato alla probabilità  $\alpha = 0.01$ . Pertanto, con probabilità inferiore a **0.01** (di commettere un errore di I tipo, cioè di rifiutare l'ipotesi nulla quando in realtà è vera), si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa: **nella popolazione** dalla quale è stato estratto il campione di 7 giovani donne, **esiste un relazione lineare** tra le variazioni in altezza e quelle in peso.

La verifica della **significatività della retta** o **verifica dell'esistenza di una relazione lineare tra le due variabili** può essere attuata anche mediante **il test t di Student**, con risultati perfettamente equivalenti al test F. Come già dimostrato per il confronto tra le medie di due campioni dipendenti od indipendenti, anche

**nel test di linearità il valore di t con df n-2 è uguale alla radice quadrata di F con df 1 e n-2**

$$t_{(n-2)} = \sqrt{F_{(1,n-2)}}$$

Il test **t** è fondato su calcoli che sono didatticamente meno chiari di quelli del test **F**, per la comprensione dei parametri riportati nelle formule; ma per l'inferenza offre due vantaggi

- **può essere più facilmente applicato anche a test unilaterali,**

$$H_1: b < 0 \quad \text{oppure} \quad H_1: b > 0$$

- **permettere il confronto con qualsiasi valore ( $b_0$ ), (non solo 0 come con il test F) quindi verificare l'ipotesi nulla**

$$H_0: b = b_0$$

ovviamente sempre con ipotesi alternative **H<sub>1</sub>** bilaterali oppure unilaterali.

I test unilaterali non solo **sono più potenti** di quelli bilaterali, ma spesso sono anche più adeguati ai fini della ricerca. Ad esempio, sulla relazione lineare tra altezza e peso fino ad ora utilizzato, è più logico un test unilaterale (all'aumentare dell'altezza il peso aumenta) che non un test bilaterale

(all'aumentare dell'altezza il peso varia), potendo a priori escludere come accettabile il risultato che all'aumentare dell'altezza il peso medio possa diminuire.

Il **test t** è fondato sul rapporto tra il valore del coefficiente angolare **b** ed il suo errore standard **s<sub>b</sub>**.

La formula generale può essere scritta come

$$t_{(n-2)} = \frac{b - b_0}{s_b}$$

dove

- **b<sub>0</sub>** è il valore atteso,
- **s<sub>b</sub>** è determinato dalla radice quadrata del rapporto tra la dispersione dei dati sperimentali (Y) intorno alla retta di regressione ( $\hat{Y}$ ) e la devianza totale di X.

$$s_b = \sqrt{\frac{\text{Varianza} \cdot d'errore \cdot della \cdot retta}{\text{Devianza} \cdot della \cdot X}} = \sqrt{\frac{s_e^2}{\sum (X_i - \bar{X})^2}}$$

Nella verifica della significatività della regressione **b** è uguale a **0**; ma **b** può assumere qualsiasi valore di confronto o ipotizzato (**b<sub>0</sub>**) e quindi la formula può essere utilizzata per

**verificare la significatività dello scostamento da qualunque valore atteso.**

Un caso relativamente frequente nella ricerca applicata consiste nel verificare se il coefficiente angolare campionario **b** può essere in disaccordo con la teoria che Y aumenti di una unità all'aumentare di una di X, cioè se  $\beta = 1$ . Si ricorre a questo confronto, ad esempio, quando si raffrontano i risultati di due metodi di valutazione che dovrebbero dare gli stessi valori.

E' importante osservare che l'errore standard di **b** (**s<sub>b</sub>**) diminuisce, quindi il valore di **t** diventa più significativo, all'aumentare della devianza di **X**. L'osservazione ha applicazioni importanti nella programmazione degli esperimenti, per la scelta dei valori campionari di X.

Si supponga di dover valutare la regressione tra peso ed altezza. Si pone un problema di scegliere gli individui, ai fini di trovare una regressione significativa. Molti sono incerti se sia preferibile

- scegliere individui di altezza media, con la motivazione che rappresentano il caso "tipico",
- scegliere individui che coprano tutto il campo di variazione dell'altezza?

Per ottenere più facilmente la significatività della pendenza della retta,

**è sempre vantaggioso utilizzare per la variabile X un campo di variazione molto ampio,**

con più misure collocate ai valori estremi. Infatti

- **se la devianza di X è grande**, il valore di **S<sub>b</sub>** è piccolo;

- di conseguenza **il valore di t è grande e più facilmente significativo.**

La varianza d'errore della retta  $s_e^2$  con df **n-2** è chiamata anche **errore standard della stima**; è data da

$$s_e^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{n - 2}$$

E' fondata sui valori attesi e quindi il suo calcolo richiede vari passaggi. Può essere stimata con le formule presentate nel test **F**, dove la devianza d'errore è ottenuta in modo rapido per differenza tra la devianza totale e quella dovuta alla regressione.

Quando è nota la retta, è possibile calcolare la devianza dovuta alla regressione direttamente dai valori sperimentali di **X** e **Y** mediante

$$\text{Devianza della regressione} = \sum Y_i^2 - a \cdot \sum Y_i - b \cdot \sum (X_i \cdot Y_i)$$

ESEMPIO 1. Con le stesse 7 misure di peso ed altezza degli esercizi precedenti, stimare la significatività della regressione mediante il test **t** di Student.

Risposta.

E' vantaggioso e più logico ricorrere ad un test unilaterale, quindi verificare se il peso aumenta in modo significativo al crescere dell'altezza. Tuttavia, in questo caso e solo con lo scopo di confrontare il risultato del test **t** con quello del precedente test **F**, è stato preferito un test bilaterale.

Ricordando dai calcoli precedenti che

$$b = 0,796 \quad s_e^2 = 16,42 \quad n = 7 \quad \text{Devianza di X} = 510 \quad S_b^2 = \frac{16,42}{510}$$

$$s_b = 0,1794$$

il valore di  $t_5$

$$t_5 = \frac{0,796}{0,1794} = 4,437$$

risulta uguale a **4,437**.

Come già messo in evidenza in varie altre occasioni, il test **F** ed il test **t** danno il medesimo risultato. Infatti,

$$F_{1,5} = 19,59 \quad \text{corrisponde a} \quad t_5 = \sqrt{19,59} = 4,426$$

(La piccola differenza tra 4,437 e 4,426 dipende dai vari arrotondamenti usati nelle due differenti serie di calcoli.)

ESEMPIO 2. Con una ricerca bibliografica, è stato trovato che il coefficiente angolare  $b_0$  della retta di regressione tra altezza ( $X$ ) e peso ( $Y$ ) in una popolazione è risultato uguale a **0,950**.

Il valore di **0,796** calcolato sulle **7** giovani se ne discosta in modo significativo?

Risposta.

E' un test bilaterale, in quanto chiede semplicemente se il valore calcolato  $b$  si discosta in modo significativo da un valore atteso, dove

$$H_0: \beta = 0,950 \quad e \quad H_1: \beta \neq 0,950$$

Applicando la formula

$$t_{(n-2)} = \frac{b - \mathbf{b}}{s_b}$$

si trova

$$t_{(5)} = \frac{0,796 - 0,950}{0,1794} = \frac{-0,154}{0,1794} = -0,858$$

un valore di  $t$  uguale a **-0.858** con **5** df.

E' un rapporto inferiore all'unità, quindi senza dubbio non significativo. Di conseguenza, si deve concludere che non è dimostrata l'esistenza di una differenza tra il coefficiente angolare riportato sulla pubblicazione e quello sperimentalmente calcolato con i 7 dati.

Quando non è possibile rifiutare l'ipotesi nulla in merito al coefficiente angolare  $b$  (pertanto la retta campionaria non può essere assunta come significativa di una relazione lineare tra le due variabili), la risposta ai diversi valori di  $X$  è fornita dalla media di  $Y$ , della quale può essere utile la conoscenza della **varianza** e della **deviazione standard**.

Con la simbologia ormai consueta, la varianza ( $s_{\bar{Y}}^2$ ) e la deviazione standard ( $s_{\bar{Y}}$ ) della media  $\bar{Y}$ , sono rispettivamente

$$s_{\bar{Y}}^2 = \frac{s_e^2}{n} \quad e \quad s_{\bar{Y}} = \frac{s_e}{\sqrt{n}}$$

Nella ricerca ambientale, oltre alla significatività del coefficiente angolare  $b$  spesso è importante verificare anche

- la significatività dell'**intercetta (a)** (rispetto a zero)
- la significatività della sua **differenza da un valore atteso o prefissato**.

Il concetto è identico al confronto tra una media campionaria  $\bar{X}$  e la media reale  $m$  o della popolazione.

Il confronto è verificato ricorrendo ancora alla distribuzione **t**, con una formula analoga a quella per la media  $\bar{X}$  e per il coefficiente angolare **b**.

Un caso frequente è quando l'origine della retta dovrebbe coincidere con l'origine degli assi; quindi con  $X = 0$  si dovrebbe  $Y = 0$ , cioè una risposta media di **a** che non si discosta significativamente da 0.

Il test può comunque essere applicato al confronto con qualsiasi valore atteso dell'intercetta e l'ipotesi alternativa **H<sub>1</sub>** può essere sia unilaterale che bilaterale.

Per la **significatività dell'intercetta a**, si verifica l'ipotesi nulla

$$H_0: a = 0$$

mentre per il **confronto dell'intercetta a con un generico valore atteso a<sub>0</sub>** si verifica l'ipotesi nulla

$$H_0: a = a_0$$

dove **a** è il valore della popolazione dalla quale è stato estratto il campione che ha permesso il calcolo di **a**.

Il test è effettuato con il calcolo di un valore di **t**, con gdl **n-2** in quanto fondato sulla varianza d'errore della retta; è dato da

$$t_{(n-2)} = \frac{a - a_0}{s_a}$$

dove

- **s<sub>a</sub>** è l'errore standard dell'intercetta

ed è stimato come

$$s_a = \sqrt{s_e^2 \cdot \left( \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right)}$$

con **s<sub>e</sub><sup>2</sup>** che indica la **varianza d'errore** della retta

(già utilizzata per stimare la significatività del coefficiente angolare **b**).

ESEMPIO. Utilizzando gli stessi 7 dati della relazione peso-altezza, in cui

$$a = -73,357 \quad s_e^2 = 16,101 \quad n = 7 \quad \text{Devianza di X} = 510 \quad \bar{X} = 172$$

stimare se l'intercetta **a** si discosta in modo significativo da zero.

Risposta

Per verificare l'ipotesi nulla

$$H_0: a = 0$$

con ipotesi alternativa

$$H_1: a \neq 0$$

poiché l'errore standard di  $a$

$$s_a = \sqrt{16,101 \cdot \left( \frac{1}{7} + \frac{172^2}{510} \right)} = 30,599$$

è uguale a **30,599**

si ottiene un valore del  $t$  di Student

$$t_5 = \frac{-73,357}{30,599} = -2,397$$

uguale a **-2,397** con **5** df.

Per un test bilaterale, il **valore critico di  $t$  con 5 df** alla probabilità  $\alpha = 0.05$  è uguale a **2,571**.

Di conseguenza, l'intercetta calcolata non risulta significativamente diversa da zero.

In realtà il valore è così vicino alla significatività che, con un numero maggiore di dati, il test sarebbe risultato significativo.

**Per una interpretazione più attenta e meno affrettata del risultato, si pone il problema di stimare della potenza del test effettuato, prima di affermare con sufficiente sicurezza che il coefficiente angolare  $b$  oppure, come in questo caso, l'intercetta  $a$  non sono significativamente differenti da zero oppure da un qualunque valore atteso.**

## 12.7. POTENZA A PRIORI E A POSTERIORI NELLA REGRESSIONE LINEARE

Come risulterà evidente alla fine del capitolo successivo, la regressione lineare semplice e la correlazione lineare semplice hanno

- **finalità differenti,**
- **condizioni di validità differenti,**
- **nei test di significatività rispondono a domande differenti.**

Nella **verifica della significatività**

- con la retta di regressione, l'ipotesi nulla verte sul valore del coefficiente angolare  **$b$**

$$H_0: b = 0$$

- nella correlazione, l'ipotesi nulla verte sul valore del coefficiente  **$r$**

$$H_0: r = 0$$

Come sarà illustrato nei paragrafi successivi, nel confronto con un qualsiasi valore teorico,

- con il **coefficiente angolare  $b$** , si verifica l'ipotesi nulla

$$H_0: b = b_0$$

- con il **coefficiente di correlazione  $r$**  si verifica l'ipotesi nulla

$$H_0: r = r_0$$

Tuttavia, la regressione e la correlazione possono essere calcolate sulle stesse coppie di dati.

Per ambedue,

- la significatività può essere stimata sia con il test **F** sia con il test **t**,
- i test hanno gli stessi gradi di libertà,
- le ipotesi alternative possono essere ugualmente sia bilaterali che unilaterali.,
- **i risultati della significatività sono identici**: un test **t** e un test **F** su **b** forniscono lo stesso valore di quello applicato su **r**.

Sulla base di queste affinità, anche

- **la potenza a priori**, cioè il numero di dati (**n**) che servono affinché il coefficiente angolare **b** o il coefficiente di correlazione **r** risultino significativamente differenti da zero o da un valore prefissato,
- **la potenza a posteriori (1-b)**, cioè la probabilità di rifiutare correttamente l'ipotesi nulla in un test sulla significatività di un coefficiente angolare **b** oppure un coefficiente di correlazione **r**,  
**risultano uguali.**

Di conseguenza, **per il calcolo della potenza della regressione si utilizza la procedura per la correlazione** (che è spiegata nel capitolo successivo), dopo aver ricavato **r** dai dati della regressione o dai suoi indici. Questo valore **r** può essere **ricavato** a partire

- sia dal **coefficiente di determinazione  $R^2$**  (come sarà spiegato in un paragrafo successivo) con

$$r = \sqrt{R^2}$$

- sia dal **coefficiente angolare  $b$**  con

$$r = b \cdot \sqrt{\frac{\sum (X_i - \bar{X})^2}{\sum (Y_i - \bar{Y})^2}}$$

Nei paragrafi successivi di questo capitolo, verranno presentati i metodi per **confrontare due coefficienti angolari  $b_1$  e  $b_2$** , per verificare se appartengono alla stessa popolazione con coefficiente angolare **b**.

Anche per la correlazione, nel capitolo successivo saranno presentati i metodi per **confrontare due coefficienti di correlazione**  $r_1$  e  $r_2$ , per verificare se appartengono alla stessa popolazione con coefficiente di correlazione  $r$ .

In modo analogo al caso precedente, anche per questo test è possibile stimare

- **la potenza a priori**, cioè il numero di dati ( $n$ ) che servono affinché i coefficienti angolari  $b_1$  e  $b_2$  oppure i coefficienti di correlazione  $r_1$  e  $r_2$  risultino significativamente differenti tra loro,
- **la potenza a posteriori (1-b)** di un test di confronto tra due coefficienti angolari oppure due coefficienti di correlazione.

Anche per **il calcolo della potenza di un test sul confronto tra due coefficienti di regressione, si utilizza la procedura equivalente per la correlazione** (spiegata nel capitolo successivo).

## 12.8. INTERVALLO DI CONFIDENZA DEI PARAMETRI $b$ E $a$

L'uso della **retta di regressione a fini predittivi** richiede che possa essere stimato l'**errore di previsione**

- del coefficiente angolare  $b$
- dell'intercetta  $a$ .

I **limiti di confidenza** sono utili anche per eventuali confronti con un parametro prefissato, quindi ai fini dell'inferenza, come già fatto per la media campionaria  $\bar{X}$ . Infatti un qualsiasi valore  $b_0$  oppure  $b_0$ , se non è compreso entro i limiti di **limiti di confidenza di una retta  $b$** , è significativamente differente da esso; questa analisi coincide con i risultati di un **test t bilaterale**, alla stessa **probabilità  $\alpha$** .

Calcolato un valore  **$b$ , il coefficiente angolare della popolazione ( $b$ ) con probabilità  $\alpha$**  si trova entro i limiti fiduciali  $L_1$  e  $L_2$

$$L_1 = b - t_{(n-2, \alpha/2)} \cdot s_b$$

$$L_2 = b + t_{(n-2, \alpha/2)} \cdot s_b$$

spesso scritto anche

$$\beta = b \pm t_{(n-2, \alpha/2)} \cdot s_b$$

dove

- $s_b$  è l'errore standard di  $b$

$$s_b = \sqrt{\frac{s_e^2}{\sum (X_i - \bar{X})^2}}$$

Per l'intercetta **a**, il valore reale **a**

si trova entro l'intervallo

$$\mathbf{a} = a \pm t_{(n-2, \alpha/2)} \cdot s_a$$

dove

-  $s_a$  è l'errore standard di **a**

$$s_a = \sqrt{s_e^2 \cdot \left( \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right)}$$

con

-  $s_e^2$  che indica la varianza d'errore.

ESEMPIO.

Ricorrendo agli stessi dati su altezza e peso, con i quali sono stati calcolati la retta e la sua significatività, si è ottenuto

$$b = 0,796 \quad s_b = 0,1794 \quad t_{(5,0.025)} = 2,571 \quad t_{(5,0.005)} = 4,032 \quad a = -73,357 \quad s_a = 30,599$$

Stimare alla probabilità  $\alpha = \mathbf{0.05}$  e a quella  $\alpha = \mathbf{0.01}$

- l'intervallo di confidenza del coefficiente angolare **b**
- l'intervallo di confidenza dell'intercetta **a**.

Risposta.

A) L'intervallo di confidenza del coefficiente angolare **b** alla probabilità del 95% è

$$0,796 - 2,571 \cdot 0,1794 \leq \beta \leq 0,796 + 2,571 \cdot 0,1794$$

cioè

$$0,335 \leq \beta \leq 1,257$$

$L_1$  è uguale a **0,335** e  $L_2$  è uguale a **1,257**.

Si può anche scrivere che il valore della popolazione con probabilità del 95% è compreso tra il limite fiduciale inferiore  $L_1 = \mathbf{0,335}$  e il limite fiduciale superiore  $L_2 = \mathbf{1,257}$ .

Alla probabilità del 99% è

$$0,796 - 4,032 \cdot 0,1794 \leq \beta \leq 0,796 + 4,032 \cdot 0,1794$$

cioè

$$0,073 \leq \mathbf{b} \leq 1,519$$

il valore reale  $\mathbf{b}$  è compreso tra  $\mathbf{L}_1 = 0,073$  e  $\mathbf{L}_2 = 1,519$ .

B) L'**intervallo di confidenza per l'intercetta**  $\alpha$  alla probabilità del 95% è

$$-73,357 - 2,571 \cdot 30,599 \leq \alpha \leq -73,357 + 2,571 \cdot 30,599$$

$$-152,027 \leq \alpha \leq 5,313$$

compreso tra  $\mathbf{L}_1 = -152,027$  e  $\mathbf{L}_2 = 5,313$ .

Alla probabilità del 99% è

$$-73,357 - 4,032 \cdot 30,599 \leq \alpha \leq -73,357 + 4,032 \cdot 30,599$$

$$-196,732 \leq \alpha \leq 50,018$$

compreso tra  $\mathbf{L}_1 = -196,732$  e  $\mathbf{L}_2 = 50,018$ .

Anche in queste applicazioni, come già evidenziato per la media della popolazione ( $\mu$ ) rispetto alla media campionaria ( $\bar{X}$ ), l'intervallo fiduciale costruito attorno al valore campionario cresce, se si aumenta la probabilità che esso comprenda il valore reale.

(Per semplicità didattica e come aiuto alla esecuzione di tutti i calcoli richiesti dalle formule, è stato utilizzato un numero di dati molto limitato, nettamente inferiore a quello che si usa nella ricerca; di conseguenza, i parametri della retta hanno intervalli molto ampi).

## 12.9. LIMITI DI CONFIDENZA PER I VALORI MEDI DI $\hat{Y}_i$ STIMATI

Nella ricerca applicata all'analisi e alla gestione dell'ambiente, si rilevano utili tre diversi casi di stima dell'intervallo fiduciale:

- di **tutta la retta**, come nel paragrafo precedente;
- del **valore medio di Y stimato** ( $\hat{Y}_i$ ), corrispondente ad un dato valore di X; è il caso in cui si somministra una sostanza tossica ad un gruppo di cavie e si vuole stimare quale sarà l'**effetto medio** sulla loro crescita, supposto che esista complessivamente una relazione lineare tra dose e accrescimento;
- di **un singolo valore di Y stimato**, sempre corrispondente ad un dato valore di X; è il caso in cui si voglia predire **la risposta di un singolo soggetto**, come succede al medico che per un paziente voglia stimare la risposta individuale possibile alla somministrazione di una dose X di un farmaco.

L'intervallo di uno specifico valore medio di  $\hat{Y}_i$  o valore atteso di  $\hat{Y}_i$  corrispondente ad un singolo valore di  $X_i$ , come può essere l'intervallo fiduciale del peso medio di giovani donne alte 1,70 cm.,

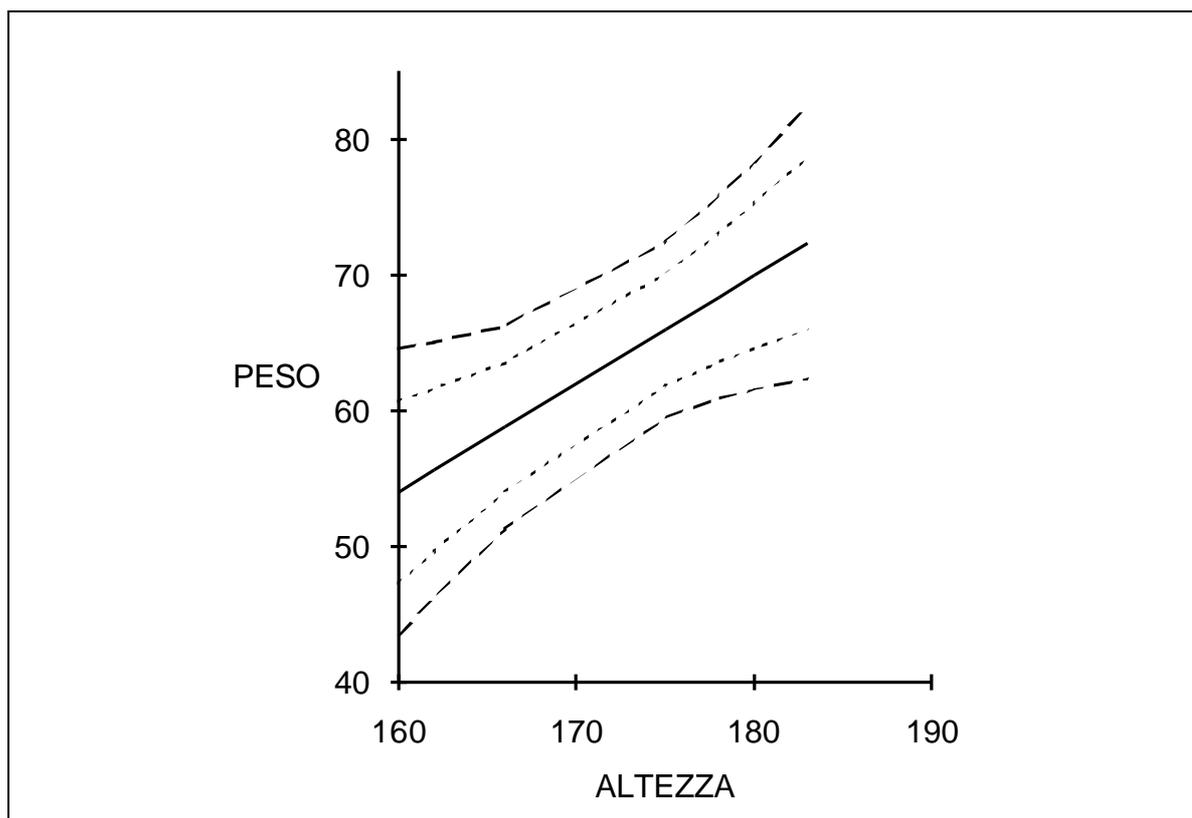
è stimato mediante la formula

$$\hat{Y}_i \pm t_{(n-2, \alpha/2)} \cdot s_e \cdot \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$$

dove

- $\hat{Y}_i$  è il valore previsto o medio di Y per un dato valore di X,
- $s_e$  è la radice quadrata della varianza d'errore stimata sulla retta **b**,
- **n** è la dimensione del campione,
- $X_i$  è il valore di X del quale si analizza a risposta media  $\hat{Y}_i$ ,
- $\sum (X_i - \bar{X})^2$  è la devianza di X.

Questa formula spiega come l'ampiezza dell'intervallo di confidenza dipenda da vari fattori.



Intervalli di confidenza per valori medi di  $\hat{Y}_i$  al 5% (linee a punti)  
e all'1% (linee tratteggiate)

Per una data probabilità  $\alpha$ ,

- **aumenta al crescere della varianza d'errore,**

- **diminuisce all'aumentare del numero n di osservazioni**, per l'effetto congiunto del valore di  $t_{n-2, \alpha/2}$  e del rapporto  $1/n$ ,
- **diminuisce al crescere della devianza di X**,
- **varia in funzione dei valori di X**, con valori minimi quando  $X_i$  è vicino alla sua media e valori massimi quando  $X_i$  ha distanza massima dalla media.

E' importante evidenziare questa ultima caratteristica. A differenza di quanto succede per l'intervallo di tutta la retta, l'intervallo fiduciale del valore medio atteso  $\hat{Y}_i$  **non è costante, ma varia con una funzione iperbolica in rapporto alla vicinanza di  $X_i$  alla sua media**  $(X_i - \bar{X})^2$ . Di conseguenza, quando si fanno previsioni su valori di  $\hat{Y}_i$  per  $X_i$  molto distanti dalla media, si stima un intervallo di confidenza maggiore di quelli ad essa vicini.

Considerando i 7 dati dell'esempio ricorrente sulla relazione tra peso e altezza, è stato calcolato l'intervallo fiduciale degli  $Y_i$  stimati per ogni valore X rilevato; è possibile anche il confronto con il valore Y campionario.

Altezza	Peso	Valori attesi di Y con il loro intervallo di confidenza					
		$\alpha = 0.05$			$\alpha = 0.01$		
X	Y	$L_1$	$\hat{Y}_i$	$L_2$	$L_1$	$\hat{Y}_i$	$L_2$
160	52	47,291 ≤ 54,018 ≤ 60,495			43,468 ≤ 54,018 ≤ 64,568		
178	68	63,582 ≤ 68,348 ≤ 73,114			60,873 ≤ 68,348 ≤ 75,823		
183	75	65,968 ≤ 72,328 ≤ 78,688			62,353 ≤ 72,328 ≤ 82,303		
180	71	64,596 ≤ 69,940 ≤ 75,284			61,560 ≤ 69,940 ≤ 78,321		
166	63	54,029 ≤ 58,795 ≤ 63,561			51,320 ≤ 58,795 ≤ 66,270		
175	59	61,827 ≤ 65,960 ≤ 70,093			59,478 ≤ 65,960 ≤ 72,442		
162	57	49,605 ≤ 55,611 ≤ 61,617			46,192 ≤ 55,611 ≤ 65,030		

Nella tabella sono riportati

- i valori medi di  $\hat{Y}_i$  (al centro)

- i relativi intervalli di confidenza ( $L_1$ , valore medio,  $L_2$ ) alla probabilità  $\alpha = 0.05$  e  $\alpha = 0.01$ , per alcuni valori di  $X_i$  elencati in ordine casuale:

I valori di  $L_1$  e  $L_2$ , insieme con la figura, evidenziano

- la minore dispersione del valore medio di  $Y_i$  stimato ( $\hat{Y}_i$ ) quando il valore di  $X_i$  è prossimo alla sua media,
- la maggiore dispersione delle stime alla probabilità  $\alpha = 0.01$  rispetto a quelle della probabilità  $\alpha = 0.05$ .

La stima dell'errore standard di ogni valore medio permette anche **il confronto tra un valore medio calcolato per una specifica quantità  $X_i$  ed un valore medio ipotizzato**, attraverso il test  $t$  con df  $n-2$  ed ipotesi  $H_1$  sia bilaterali che unilaterali

$$t_{(n-2)} = \frac{\hat{Y}_{\text{calcolato}} - \bar{Y}_{\text{ipotizzato}}}{\sqrt{s_e^2 \cdot \left( \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right)}}$$

Nel calcolo di singoli valori medi, l'**errore standard di a** ( $s_a$ ) è uguale a quello di **b**. E' infatti semplice osservare che per  $x_i = 0$  si ottiene  $\hat{Y}_i = a$ .

$$s_a = s_e \cdot \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$$

scritto spesso come

$$s_a = \sqrt{s_e^2 \cdot \left( \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right)}$$

in quanto  $X_i = 0$

Tuttavia l'intervallo fiduciale di  $a$  è quasi sempre molto grande: la sua distanza dal valore medio è massima, quindi  $(X_i - \bar{X})^2$  oppure  $\bar{X}^2$  sono valori molto grandi.

Ponendo l'attenzione sul valore medio di  $Y_i$  stimato ( $\hat{Y}_i$ ) per una specifica quantità  $X_i$ , in varie occasioni il ricercatore è motivato ad **aumentare il numero di osservazioni**, appunto per quella

quantità. Poiché la varianza d'errore può essere ipotizzata costante, l'aumento delle osservazioni aumenta la precisione.

L'errore standard per quel valore di  $Y_i$  stimato ( $\hat{Y}_i$ ) cioè  $s_{Y_i}$  diventa

$$s_{Y_i} = \sqrt{s_e^2 \cdot \left( \frac{1}{m} + \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right)}$$

dove

-  $m$  è il numero di osservazioni aggiuntive per lo specifico  $X_i$

### 12.10. LIMITI DI CONFIDENZA PER SINGOLI VALORI DI $\hat{Y}_i$ STIMATI

Un'altra esigenza frequente nella ricerca è la previsione dell'**intervallo di confidenza per una singola risposta di  $Y_i$** ; in altri termini, quale è la dispersione dei singoli valori di  $Y$ , per un dato valore  $X_i$ .

L'intervallo precedente era analogo a quello stimato con l'errore standard di una media; questo è analogo a quello calcolato con la deviazione standard.

L'intervallo di confidenza per singoli valori ha una forma simile a quella del valore medio. Ovviamente è più ampio; ha infatti lo scopo di stimare un valore individuale e non una risposta media.

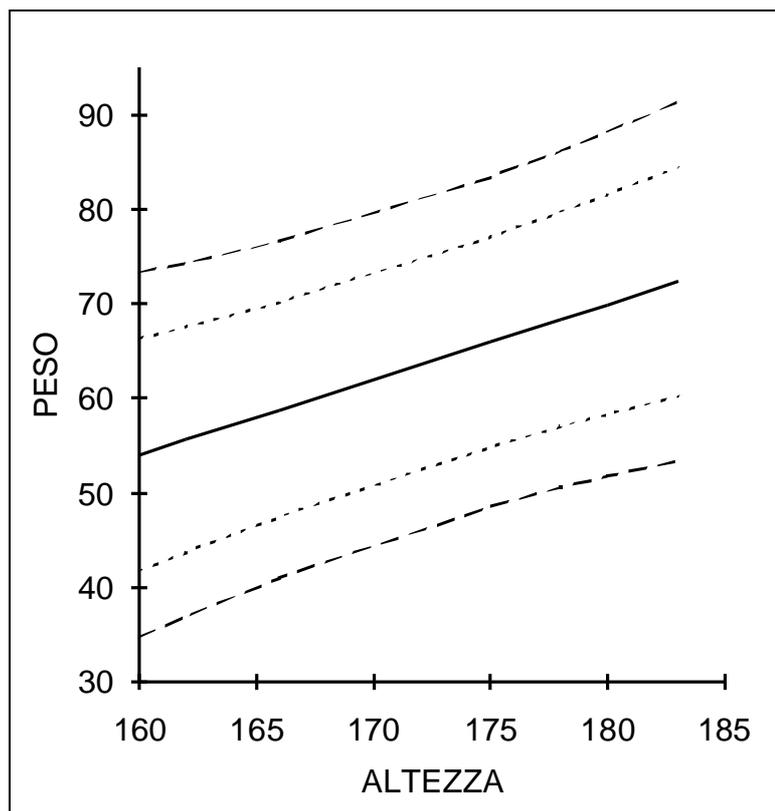
**L'intervallo di previsione di un singolo valore di  $Y$**  per un dato valore  $X_i$  può essere stimato mediante

$$\hat{Y}_i \pm t_{(n-2, \alpha/2)} \cdot s_e \cdot \sqrt{1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$$

con la consueta simbologia,

usata anche per i valori medi di  $Y$  nel paragrafo precedente.

La figura successiva è costruita con i dati della tabella, per gli stessi valori  $X_i$  già utilizzati per il calcolo dell'intervallo fiduciale dei valori medi  $\hat{Y}_i$  riportati nel paragrafo precedente.



Intervalli di confidenza per singoli valori di  $\hat{Y}_i$  al 5% (linee a punti) e all'1% (linee tratteggiate)

La formula infatti è identica, con la sola differenza del  $1+$  nella parte sotto radice

Altezza	Peso	Valori attesi di Y con il loro intervallo di confidenza					
		$\alpha = 0.05$			$\alpha = 0.01$		
X	Y	$L_1$	$\hat{Y}_i$	$L_2$	$L_1$	$\hat{Y}_i$	$L_2$
160	52	41,702 ≤ 54,018 ≤ 66,334			34,703 ≤ 54,018 ≤ 73,332		
178	68	56,984 ≤ 68,348 ≤ 79,712			50,526 ≤ 68,348 ≤ 86,170		
183	75	60,208 ≤ 72,328 ≤ 84,447			53,321 ≤ 72,328 ≤ 91,335		
180	71	58,322 ≤ 69,940 ≤ 81,558			51,719 ≤ 69,940 ≤ 88,161		
166	63	47,431 ≤ 58,795 ≤ 70,159			40,973 ≤ 58,795 ≤ 73,617		
175	59	54,846 ≤ 65,960 ≤ 77,074			48,531 ≤ 65,960 ≤ 83,389		
162	57	43,674 ≤ 55,611 ≤ 67,548			36,890 ≤ 55,611 ≤ 74,332		

## 12.11. ERRORI DELLE VARIABILI E LIMITI DI TOLLERANZA

Per analizzare la regressione di Y su X; **non si prende in considerazione alcuna forma di variabilità casuale dei valori di X**, ma solo quella di Y.

In alcuni esperimenti, è evidente che questa concetto è corretto: la variabile X non varia, se non in modo trascurabile, solo per errori di misura. E' il caso dello studio delle relazioni dose-effetto, come in esperimenti di tossicologia: la dose (X) di tossico somministrata è misurata in modo preciso, con l'unico errore trascurabile dato dallo strumento, mentre la risposta (Y) è fisiologica e quindi presenta la variabilità individuale a stimoli di pari entità.

In altri esperimenti, come nella regressione peso-altezza fino ad ora utilizzata quale esempio, può apparire, a chi non ha esperienza statistica, che i valori di X abbiano variazioni casuali, in quanto affetti sia da errori di misura sia da perturbazioni aleatorie, in modo non diverso dalle Y corrispondenti.

In realtà, nella trattazione classica della regressione lineare, gli errori casuali di X non sono considerati: **viene analizzata solamente la variabilità delle Y, in quanto è la “risposta” o effetto che viene valutata** e che può presentare dispersione rispetto allo “stimolo” o causa.

Recentemente, sono stati proposti approcci più complessi che intendono valutare il valore vero di X stimato ( $\hat{X}$ ) al posto del valore di X osservato; ma tali argomenti esulano dalle finalità della presente trattazione.

Nello studio della regressione, l'ipotesi di omoscedalità (l'ipotesi di indipendenza dell'errore od omogeneità della varianza) e quella di normalità riguardano solo le Y.

Quando le condizioni di validità, che verranno di seguito discusse, non sono rispettate, è possibile ricorrere a

- una regressione lineare non parametrica (che sarà presentata in un capitolo successivo),
- **una valutazione ed una descrizione della regressione mediante i limiti di tolleranza.**

**I limiti di tolleranza** sono un metodo utile per rappresentare l'evoluzione temporale di un fenomeno (X uguale al tempo) oppure la risposta dose-effetto, quando si abbia

- una **distribuzione di valori non normali**, non simmetrici rispetto alla tendenza centrale,
- **variabilità diversa** al variare della X,
- presenza di **outliers**.

Questa metodologia permette di individuare l'evoluzione della tendenza centrale, cioè quali valori di Y rientrano nella norma e quali se ne discostano in modo rilevante, sulla base della loro frequenza. La tecnica può essere applicata a qualunque tipo di regressione.

**L'intervallo di tolleranza** può essere considerato la versione non parametrica dell'intervallo di confidenza, come presentato nel capitolo di statistica descrittiva:

- **la mediana** (per ogni tempo o dose di X) individua la **tendenza centrale**,
- **i quartili, i decili oppure i centili** descrivono la **variabilità** di una distribuzione di dati, misurati con una scala che sia almeno di tipo ordinale.

Nello stesso modo, dati che siano stati rappresentati graficamente in un diagramma di dispersione possono essere raggruppati in classi, secondo il valore della X, suddivisa in intervalli di dimensioni che possono anche essere variabili. Utilizzando le singole osservazioni per ogni dose di X, si richiede un numero molto alto di dati, in rapporto al numero di classi e alla precisione con la quale si vuole descrivere la dispersione (quartili oppure centili).

Per ogni gruppo che sia formato da un numero sufficiente di dati, è semplice individuare la mediana, i quartili e i centili, tra i quali di norma sono utilizzati quelli che escludono il 20%, il 10% o il 5% dei valori in una o in entrambe le code della distribuzione. Le quantità sono scelte in rapporto ai fenomeni che si intendono evidenziare. Unendo con un tratteggio gli stessi indicatori dei gruppi diversi, si ottiene un disegno simile a quello della retta con i suoi intervalli fiduciali.

Sono rappresentazioni bidimensionali che descrivono graficamente l'evoluzione della tendenza centrale e le caratteristiche fondamentali della dispersione dei dati, per ogni raggruppamento effettuato.

**E' una tecnica descrittiva, non inferenziale**, che è applicata soprattutto per mostrare l'evoluzione geografico-temporale di una variabile, quali una serie annuale di valori d'inquinamento, rilevati giornalmente in vari punti di una città. Altra applicazione frequente è la relazione dose-risposta, quando la risposta Y individuale non è simmetrica attorno ai valori medi e/o la variabilità non è costante, ma varia in funzione dello stimolo X.

#### 12.12. IL COEFFICIENTE DI DETERMINAZIONE: $R^2$ E $R^2_{adj}$ .

Il **coefficiente di determinazione** (*coefficient of determination*)  $R^2$  (*R square* indicato in alcuni testi e in molti programmi informatici anche con **R** oppure  $r^2$ ) è la proporzione di variazione totale che è spiegata dalla variabile dipendente.

E' il rapporto della devianza dovuta alla regressione sulla devianza totale:

$$R^2 = \frac{\text{Devianza della regressione}}{\text{Devianza totale}} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

Espresso a volte in percentuale, più spesso con un indice che varia da 0 a 1,  $R^2$  serve per misurare **quanto della variabile dipendente Y sia predetto dalla variabile indipendente X** e quindi per valutare l'utilità dell'equazione di regressione ai fini della previsione sui valori della Y.

Il valore del coefficiente di determinazione è tanto più elevato quanto più la retta passa vicino ai punti, fino a raggiungere 1 quando tutti i punti sperimentali sono collocati esattamente sulla retta. In tale caso, infatti, ogni  $Y_i$  può essere predetto con precisione totale dal corrispondente valore di  $X_i$ . Nella ricerca ambientale, è prassi diffusa che la determinazione possa essere ritenuta buona, (in linguaggio tecnico, il modello ha un buon *fitting* con i valori sperimentali), quando  $R^2$  supera 0,6 (o 60%).

In realtà è una indicazione molto approssimativa. La valutazione del valore di  $R^2$  è in stretto rapporto con la disciplina studiata, tanto che **i sociologi spesso ritengono alto un valore di  $R^2$  uguale a 0,30 mentre i fisici stimano basso un  $R^2$  pari a 0,98.**

$R^2$  è una misura che ha **scopi descrittivi** del campione raccolto; non è legata ad inferenze statistiche, ma a scopi pratici, specifici dell'uso della regressione come metodo per prevedere Y conoscendo X.

Per meglio spiegare il concetto, è utile un esempio.

In una città, per valutare l'inquinamento atmosferico sono state prese misure della concentrazione di  $NO_2$  con una serie di rilevatori. Per evitare variazioni indotte dal traffico (quale la sosta prolungata di un automezzo con i motori accesi) e manomissioni da parte dei passanti, tali strumenti di rilevazione sono stati collocati a 12 metri di altezza. All'obiezione che per le persone comunque sono rilevanti le concentrazioni che esse respirano, presenti ad un'altezza dal suolo tra metri 1 e 1,8 e non certo a 12 metri, con alcune misure di confronto è stata fatta una regressione lineare. Mediante essa, a partire dal valore dell'inquinamento a 12 metri (X), era stimato il valore presente a 1,5 metri (Y).

L'errore di stima appariva trascurabile per una disciplina ambientale, poiché  $R^2$  risultava uguale a circa 0,92. Purtroppo, in molte stazioni, il valore stimato risultava intorno a 9,6 quando i limiti di legge, oltre i quali scattano misure di riduzione del traffico, erano posti a 10. Con un errore di 0,08 o 8% nello stimare il valore al suolo (in valore assoluto pari a 0,77 rispetto al valore stimato di 9,6) diventa impossibile decidere se i limiti di legge sono stati effettivamente superati oppure no.

In questo caso, il valore di  $R^2$ , seppure oggettivamente molto alto, era troppo piccolo per ottenere una stima abbastanza precisa. Ovviamente, se i valori stimati fossero stati sensibilmente inferiori o maggiori, anche un  $R^2$  inferiore a 0,9 sarebbe stato un risultato ottimo.

Per la significatività della retta, alcuni utilizzano la significatività di  $r$  ( $r = \sqrt{R^2}$ ), il coefficiente di correlazione lineare (che sarà spiegato nel capitolo successivo), anche se sarebbe corretto utilizzare il test **F** o il test **t** di Student, spiegati nei paragrafi precedenti.

Seppure i risultati siano coincidenti, la procedura non è appropriata, sia per le differenze concettuali tra regressione e correlazione, sia per le diverse condizioni di validità. (L'argomento è discusso in modo più approfondito nel capitolo successivo, dedicato alla correlazione).

Benché i test di statistica evidenzino la funzione descrittiva di  $R^2$ , sia a scopi predittivi, sia per la significatività, in alcune condizioni ad esso viene attribuito anche un **significato inferenziale**.

Il valore di  $R^2$  è riferito ai dati del campione, ma in varie situazioni viene utilizzato per indicare l'errore che si commette nello stimare Y sulla base di generici valori di X; quindi **non è utilizzato solo per descrivere il caso sperimentale, ma è esteso come capacità predittiva alla relazione tra le due variabili**.

Quando si costruisce un modello di distribuzione dei dati (in questo caso i valori di Y conoscendo il valore di X), è importante valutare se i dati del campione sono in accordo con la teoria. In altri termini si deve **verificare la bontà dell'adattamento** (*goodness of fit*).

Per il **modello lineare**, è stato proposto un  $R^2$  corretto, chiamato  **$R^2$  aggiustato** ( *$R^2$  adjusted*) indicato nei programmi informatici con  $R^2_{adj}$

**ottenuto dalla formula generale**

$$R^2_{adj} = R^2 - \frac{p \cdot (1 - R^2)}{N - p - 1}$$

dove

- $N$  è il numero di coppie di dati od individui misurati
- $p$  è il numero di variabili (nel caso della regressione lineare semplice  $p = 1$ ).

Nel caso della **regressione lineare semplice**,

la formula semplificata diventa

$$R^2_{adj} = R^2 - \frac{1 - R^2}{N - 2}$$

In altri testi, la formula per il calcolo di  $R^2_{adj}$  da  $R^2$

è riportata come

$$R^2_{adj} = 1 - \frac{(1 - R^2) \cdot (N - 1)}{g.d.l. \text{ dell'errore}}$$

dove, nella regressione lineare semplice,

i gdl dell'errore sono  $N-2$

Dalla semplice lettura della formula si evidenzia che  **$R^2$  aggiustato è sempre minore di  $R^2$  sperimentale**, come appunto necessario in caso di inferenza.

ESEMPIO. Con le 7 osservazioni su peso ed altezza, il coefficiente di determinazione  $r^2$  è

$$r^2 = \frac{321,618}{403,715} = 0,797$$

uguale a 0,797.

Questo risultato indica che, noto il valore dell'altezza, **nel caso dei 7 dati utilizzati** il valore del peso del peso è stimato mediante la retta di regressione con una approssimazione di circa l'80 per cento (79,7%); il restante 0,2 (rapportato a 1) oppure 20% è determinato dalla variabilità individuale di scostamento dalla retta.

Per una applicazione di quella retta a scopi predittivi, estesa all'universo delle relazioni tra peso ed altezza nella popolazione dalla quale sono stati estratti i 7 individui del campione, è più corretto utilizzare  $R^2_{adj}$ .

Il valore di  $R^2_{adj}$

- sia con la prima formula

$$R^2_{adj} = R^2 - \frac{1-R^2}{N-2} = 0,797 - \frac{1-0,797}{5} = 0,797 - \frac{0,203}{5} = 0,797 - 0,0406 = 0,7564$$

- sia con la seconda

$$R^2_{adj} = 1 - \frac{(1-R^2) \cdot (N-1)}{g.d.l. \text{ dell'errore}} = 1 - \frac{(1-0,797) \cdot 6}{5} = 1 - \frac{0,203 \cdot 6}{5} = 1 - 0,2436 = 0,7564$$

risulta uguale a 0,7564.

### 12.13. LA PREDIZIONE INVERSA

Stimata la retta nella relazione corretta di causa-effetto, anche nella ricerca ambientale non è rara la richiesta di ricorrere alla stima inversa. Soprattutto quando si valuta l'effetto di un tossico oppure di un farmaco o di una qualsiasi sostanza ( $\mathbf{X}$ ) con effetti ( $\mathbf{Y}$ ) fisiologici, spesso è importante stimare quale è la dose  $\mathbf{X}_i$  che determina uno specifico effetto  $\mathbf{Y}_i$ .

Dopo aver identificato l'effetto  $\mathbf{Y}$  (quale il 50% oppure il 90% dei decessi; la crescita di  $\mathbf{Y}_i$  cm. o hg. in individui sottoposti a dieta con integratori alimentari) che si ritiene importante sotto l'aspetto biologico, ecologico od ambientale, si vuole risalire alla dose  $\mathbf{X}_i$  corrispondente, che lo ha determinato.

E' la **predizione inversa** (*inverse prediction*); la sua formula può essere ricavata facilmente

dalla formula generale della retta

$$\hat{Y}_i = a + bX_i$$

ed è

$$\hat{X}_i = \frac{Y_i - a}{b}$$

Nel caso della relazione tra peso ed altezza, la predizione inversa diventa la stima dell'altezza  $\hat{X}_i$  media ( $\hat{X}_i$  è una media, come i valori collocati sulla retta) per una persona di peso  $Y_i$ .

ESEMPIO 1. Nel caso della relazione tra peso ed altezza, con retta

$$\hat{Y}_i = -73,354 + 0,796 \cdot X_i$$

stimare l'altezza (teorica o media) di una giovane donna con peso uguale a 60 Kg.

Risposta

Con

$$\hat{X}_i = \frac{Y_i - a}{b}$$

dove

$$- Y_i = 60 \quad a = -73,354 \quad b = 0,796$$

si ottiene

$$\hat{X}_i = \frac{60 - (-73,354)}{0,796} = \frac{133,354}{0,796} = 167,5$$

un'altezza di 167,5 (in cm. poiché è la scala con la quale sono stati stimati gli altri parametri).

Più complessa è la stima **dell'intervallo di confidenza**; a differenza di quello di  $\hat{Y}_i$ , l'intervallo di confidenza di  $\hat{X}_i$  non è simmetrico ed è dato da

$$\bar{X} + \frac{b \cdot (Y_i - \bar{Y})}{K} \pm \frac{t_{\alpha,n}}{K} \cdot \sqrt{s_e^2 \cdot \left[ \frac{(Y_i - \bar{Y})^2}{\sum (X_i - \bar{X})^2} + K \left( 1 + \frac{1}{n} \right) \right]}$$

dove

- oltre alla simbologia consueta,
- $\bar{X}$  è il valore predetto o stimato per un dato  $Y_i$ ,

- **K** è un valore che dipende dalla probabilità **a** e dai df della varianza d'errore; può essere stimato con il valore critico **t** alla probabilità **a bilaterale** e con df **n= n-2**

$$K = b^2 - t_{a,n}^2 \cdot s_b^2$$

oppure con il valore critico **F** alla stessa probabilità **a** e con df **n<sub>1</sub> = 1** e **n = n-2**

$$K = b^2 - F_{a,1,n} \cdot s_b^2$$

dove

$$s_b = \sqrt{\frac{s_e^2}{\sum (X_i - \bar{X})^2}}$$

ESEMPIO 2. Calcolare alla probabilità **a = 0.05** l'intervallo di confidenza del valore **Y<sub>i</sub>** precedentemente stimato, ricordando che per **Y<sub>i</sub> = 60 Kg.** si era ottenuto **X<sub>i</sub> = 167,5 cm.**

Risposta

Con i dati del problema  $\bar{X} = 167,5$   $Y_i = 60$   $\alpha = 0.05$

- si devono dapprima trovare i dati che servono, per la formula presentata, nei calcoli precedenti

$$n = 7 \quad \bar{Y} = 63,57 \quad s_e^2 = 16,42 \quad \sum (X_i - \bar{X})^2 = 510 \quad s_b^2 = 0,1794$$

- e nelle tabelle dei valori critici, dove

per **a = 0.05** in una distribuzione **bilaterale** con **df n = 5** il valore di **t = 2,571**

e/o per **a = 0.05** con **df n<sub>1</sub> = 1** e **n<sub>2</sub> = 5** il valore di **F = 6,61** (ricordando che **2,571<sup>2</sup> = 6,61**)

Successivamente si calcola **K**

$$K = 0,796^2 - 6,61 \cdot 0,1794^2 = 0,634 - 0,213 = 0,421$$

ed infine l'intervallo di confidenza

dove **L<sub>1</sub>** risulta

$$L_1 = 167,5 + \frac{0,796 \cdot (60 - 63,57)}{0,421} - \frac{2,571}{0,421} \cdot \sqrt{16,42 \cdot \left[ \frac{(60 - 63,57)^2}{510} + 0,421 \cdot \left( 1 + \frac{1}{7} \right) \right]}$$

$$L_1 = (167,5 + (-6,75)) - 6,11 \cdot \sqrt{16,42 \cdot (0,025 + 0,481)}$$

$$L_1 = 174,25 - 6,11 \cdot 2,88 = 174,25 - 17,60 = 156,65$$

uguale a **156,65**

e  $L_2$  risulta

$$L_2 = 167,5 + \frac{0,796 \cdot (60 - 63,57)}{0,421} + \frac{2,571}{0,421} \cdot \sqrt{16,42 \cdot \left[ \frac{(60 - 63,57)^2}{510} + 0,421 \cdot \left( 1 + \frac{1}{7} \right) \right]}$$

$$L_2 = (167,5 + (-6,75)) + 6,11 \cdot \sqrt{16,42 \cdot (0,025 + 0,481)}$$

$$L_2 = 174,25 + 6,11 \cdot 2,88 = 174,25 + 17,60 = 191,85$$

uguale a **191,85**.

Rispetto al valore medio di 167,5 cm. l'intervallo di confidenza al 95% di probabilità è compreso tra 156,65 e 191,85 (Osservare come non sia simmetrico rispetto al valore centrale)

Se gli individui con lo stesso valore di Y prescelto (60 Kg di peso) sono numerosi, pari a  $m$ , la formula per il calcolo della media di X non varia, mentre la stima dell'intervallo fiduciale diventa più complessa.

#### 12.14. CONFRONTO TRA DUE O PIU' RETTE DI REGRESSIONE

**I coefficienti angolari delle rette di regressione possono essere posti a confronto, con metodi del tutto simili a quelli utilizzati per le medie.**

Anche sotto l'aspetto concettuale, **le rette sono medie**, in quanto

- **indicano la risposta media di  $Y_j$  per un dato valore di  $X_j$ .**

Questi **test**, detti **di parallelismo** in quanto le rette con lo stesso coefficiente angolare sono parallele, servono per verificare la significatività delle differenze tra due o più coefficienti di regressione, mediante la distribuzione t o la distribuzione F.

Quando si dispone di dati di regressione che sono stati classificati in due o più gruppi, all'interno di ognuno di essi è possibile ammettere l'esistenza di una regressione lineare di Y su X. Come esempio, può essere considerato il caso del confronto della relazione tra peso ed altezza calcolata in un gruppo di maschi ed uno di femmine, oppure tra più gruppi classificati per classe d'età (giovani di 20 anni, adulti di 40 anni, anziani di 70 anni).

I dati di **p gruppi**, ognuno con **m osservazioni** sperimentali, possono essere riportati in una tabella come la seguente, utile per la presentazione dei dati e per la comprensione delle formule.

Le rette di regressione possono differire per

- la pendenza o **coefficiente angolare**,
- la posizione o **intercetta**.

Nella ricerca ambientale è più **frequente la richiesta di confronti sulla pendenza**, per il loro significato specifico di variazione unitaria di Y in rapporto alla differenza di 1 unità di X; di conseguenza, in questa trattazione dei concetti fondamentali di statistica applicata alla ricerca e alla gestione ambientali, vengono presi in considerazione solamente i coefficienti angolari.

Nel caso di più gruppi, i dati originari possono essere riportati in una tabella come quella sottostante, nella quale per ogni gruppo sono riportati i valori della variabile X e quelli corrispondenti, riferiti allo stesso individuo, della variabile Y.

	Gruppi									
	1		2		...	j		...	p	
	<b>X<sub>1</sub></b>	<b>Y<sub>1</sub></b>	<b>X<sub>2</sub></b>	<b>Y<sub>2</sub></b>		<b>X<sub>j</sub></b>	<b>Y<sub>j</sub></b>		<b>X<sub>p</sub></b>	<b>Y<sub>p</sub></b>
Ind.										
1	x <sub>11</sub>	y <sub>11</sub>	x <sub>12</sub>	y <sub>12</sub>	...	x <sub>1j</sub>	y <sub>1j</sub>	...	x <sub>1p</sub>	y <sub>1p</sub>
2	x <sub>21</sub>	y <sub>21</sub>	x <sub>22</sub>	y <sub>22</sub>	...	x <sub>2j</sub>	y <sub>2j</sub>	...	x <sub>2p</sub>	y <sub>2p</sub>
3	x <sub>31</sub>	y <sub>31</sub>	x <sub>32</sub>	y <sub>32</sub>	...	x <sub>3j</sub>	y <sub>3j</sub>	...	x <sub>3p</sub>	y <sub>3p</sub>
---	...	...	...	...	...	...	...	...	...	...
I	x <sub>i1</sub>	y <sub>i1</sub>	x <sub>i2</sub>	y <sub>i2</sub>	...	x <sub>ij</sub>	y <sub>ij</sub>	...	x <sub>ip</sub>	y <sub>ip</sub>
---	...	...	...	...	...	...	...	...	...	...
M	x <sub>m1</sub>	y <sub>m1</sub>	x <sub>m2</sub>	y <sub>m2</sub>		x <sub>mj</sub>	y <sub>mj</sub>		x <sub>mp</sub>	y <sub>mp</sub>
Medie	$\bar{X}_1$	$\bar{Y}_1$	$\bar{X}_2$	$\bar{Y}_2$	...	$\bar{X}_j$	$\bar{Y}_j$	...	$\bar{X}_p$	$\bar{Y}_p$

Il confronto per il parallelismo tra due rette di regressione, con coefficienti angolari **b<sub>1</sub>** e **b<sub>2</sub>** e con un numero di osservazioni **n<sub>1</sub>** e **n<sub>2</sub>**, può essere realizzato con il **test t di Student**, per verificare l'ipotesi nulla

$$H_0: b_1 = b_2$$

con ipotesi alternative che possono essere sia bilaterali che unilaterali,

$$\mathbf{H}_1: b_1 \neq b_2; \quad \mathbf{H}_1: b_1 < b_2 \quad \text{oppure} \quad \mathbf{H}_1: b_1 > b_2$$

nello stesso modo del confronto tra due medie campionarie.

Questo test  $t$  ha un numero di **gdl** uguale a  $(n_1 - 2) + (n_2 - 2)$ , in quanto utilizza le varianze d'errore delle due rette; spesso i gdl complessivi sono indicati come  $N - 4$  dove  $N$  è il numero totale di osservazioni dei due gruppi.

Il valore di  $t_{(N-4)}$  è calcolato con

$$t_{(N-4)} = \frac{b_1 - b_2}{es_{(b_1-b_2)}}$$

dove

$es_{(b_1-b_2)}$  è l'errore standard della differenza tra le due rette ed è ottenuto da

$$es_{(b_1-b_2)} = \sqrt{s_{(b_1-b_2)}^2 \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

con

$$s_{(b_1-b_2)}^2 = \frac{\sum (Y_{1i} - \hat{Y}_1)^2 + \sum (Y_{2i} - \hat{Y}_2)^2}{n_1 - 2 + n_2 - 2} \cdot \left( \frac{1}{\sum (X_{1i} - \bar{X}_1)^2} + \frac{1}{\sum (X_{2i} - \bar{X}_2)^2} \right)$$

**Se l'ipotesi nulla non viene respinta, è accettabile assumere che i due coefficienti angolari  $b_1$  e  $b_2$  siano approssimativamente uguali.**

Di conseguenza, in vari casi si richiede di utilizzare

- **il coefficiente angolare medio  $b_c$** ,

che è ottenuto nel modo più rapido dal rapporto tra la somma delle due covarianze con quella delle due devianze di X

$$b_c = \frac{\sum (X_{1i} - \bar{X}_1) \cdot (Y_{1i} - \bar{Y}_1) + \sum (X_{2i} - \bar{X}_2) \cdot (Y_{2i} - \bar{Y}_2)}{\sum (X_{1i} - \bar{X}_1)^2 + \sum (X_{2i} - \bar{X}_2)^2}$$

La sua varianza  $s^2(b_c)$  è uguale alla  $s_{(b_1-b_2)}^2$  riportata sopra.

**ESEMPIO.** Calcolare la regressione lineare semplice tra peso ( $\mathbf{Y}$ ) misurato in Kg ed altezza ( $\mathbf{X}$ ) misurata in cm su un gruppo di 8 maschi ed uno di 8 femmine

MASCHI		
Individui	X	Y
1	160	59
2	175	72
3	180	81
4	190	87
5	182	82
6	186	84
7	166	67
8	171	68

FEMMINE		
Individui	X	Y
1	165	50
2	170	62
3	174	68
4	162	54
5	169	60
6	178	69
7	160	53
8	164	55

Dopo aver verificato se i parametri  $\alpha$  e  $\beta$  di ognuna delle due rette si discostano in modo significativo da 0, valutare se esiste una differenza significativa tra i due coefficienti angolari; in caso di parallelismo, stimare la retta comune.

Risposta

La maggior parte dei programmi informatici più diffusi, forniscono la stima dei parametri e la loro significatività con tabulati che riportano quasi sempre le informazioni seguenti.

Le prime sono riferite al gruppo dei maschi

STIMA E SIGNIFICATIVITA' DEI PARAMETRI **a** e **b** DEI MASCHI

Parametro	Df	Valore	Err. Standard	t per $H_0$ : parametro = 0	Prob.
<b>a</b>	1	- 93,639	11,973	-7,821	0.0002
<b>b</b>	1	0,957	0,068	14,106	0.001

ANALISI DELLA VARIANZA PER LA SIGNIFICATIVITA' DI **b** DEI MASCHI

	Devianza	Df	Varianza	F	Prob.
Totale	688,000	7			
Della regressione	667,860	1	667,860	198,967	0.0001
D'errore	20,140	6	3,357		

$$R^2 = 0,957 \quad \text{Devianza totale X} = 729,5 \quad \text{Codevianza XY} = 698,132$$

e le successive al gruppo delle femmine

STIMA E SIGNIFICATIVITA' DEI PARAMETRI **a** e **b** DELLE FEMMINE

Parametro	Df	Valore	Err. Standard	t per H <sub>0</sub> : parametro = 0	Prob.
<b>a</b>	1	- 119,142	29,701	-4,011	0.0070
<b>b</b>	1	1,061	0,177	5.997	0,0010

ANALISI DELLA VARIANZA PER LA SIGNIFICATIVITA' DI **b** DELLE FEMMINE

	Devianza	Df	Varianza	F	Prob.
Totale	348,875	7			
Della regressione	298,995	1	298,995	35,965	0.0010
D'errore	49,880	6	8,313		

$$R^2 = 0,857 \quad \text{Devianza totale X} = 265,5 \quad \text{Codevianza XY} = 281,696$$

Dall'analisi dei risultati è possibile dedurre che **le due rette di regressione si discostano in modo significativo da 0, sia per l'intercetta che per il coefficiente angolare**; inoltre, il valore predittivo della retta ( $R^2$ ) è molto alto e quindi come la devianza d'errore sia molto piccola rispetto a quella totale.

E' possibile osservare come per la significatività del coefficiente angolare **b**

- il valore di t e quello di F coincidano, ricordando che  $t_{(6)}^2 = F_{(1,6)}$ .

Indicando con 1 il gruppo dei maschi e con 2 il gruppo delle femmine, **la significatività della differenza tra i due coefficienti angolari** può essere valutata con una formula che, per meglio sottolineare e più facilmente ricordare i concetti relativi, può essere riportata come

$$t_{(N-4)} = \frac{b_1 - b_2}{\sqrt{s^2 \left( \frac{1}{Dev.tot.X_1} + \frac{1}{Dev.tot.X_2} \right)}}$$

dove  $s^2$

$$s^2 = \frac{Dev.d'err.Y_1 + Dev.d'err.Y_2}{(n_1 - 2) + (n_2 - 2)}$$

Applicata ai dati dell'esempio, in cui i due campioni sono stati scelti con lo stesso numero d'osservazioni ma che possono avere dimensioni diverse, la significatività della differenza tra i due coefficienti angolari mediante il test **t**

$$t_{(12)} = \frac{0,957 - 1,061}{\sqrt{\left(\frac{20,14 + 49,88}{6 + 6}\right) \cdot \left(\frac{1}{729,5} + \frac{1}{265,5}\right)}} = \frac{-0,104}{\sqrt{\frac{70,02}{12} (0,00137 + 0,00377)}} =$$

$$= \frac{-0,104}{\sqrt{0,02999}} = \frac{-0,104}{0,173} = -0,601$$

ha un valore uguale a -0,601.

Per 12 df, il **t** risulta non significativo; è addirittura inferiore a 0,695 che corrisponde alla probabilità del 50%.

Nonostante il numero ridotto di osservazioni, sulla base di questa probabilità a così alta si può affermare che le due rette:

- non solo non hanno un valore di **b** diverso,
- ma **sono parallele**;

pertanto può essere utile calcolare il coefficiente angolare comune.

Il **coefficiente angolare comune  $b_c$  tra le rette 1 e 2** è dato da

$$b_c = \frac{Cod.XY_1 + Cod.XY_2}{Dev.tot.X_1 + Dev.tot.X_2}$$

che risulta una media ponderata dei due **b**, in quanto codevianze e devianze non sono indipendenti dal numero di osservazioni dei due gruppi.

La sua **varianza  $s^2_{(b)}$**  è

$$s^2_{(b)} = \frac{s^2}{Dev.tot.X_1 + Dev.tot.X_2}$$

dove  $s^2$  è

$$s^2 = \frac{Dev.d'err.Y_1 + Dev.d'err.Y_2}{(n_1 - 2) + (n_2 - 2)}$$

come già riportato.

Applicato ai dati dell'esempio, **il coefficiente angolare  $b_c$  comune alle due rette**

$$b_c = \frac{698,132 + 281,696}{729,5 + 265,5} = \frac{979,828}{995} = 0,985$$

risulta uguale a 0,985

e la sua **varianza  $s^2_{bc}$**

$$s^2_{bc} = \frac{\frac{20,14 + 49,88}{6 + 6}}{729,5 + 265,5} = \frac{5,835}{995} = 0,00586$$

uguale a 0,00586.

**La significatività delle differenze tra più coefficienti di regressione** può essere verificata mediante

l'analisi della varianza,

con ipotesi nulla

$$H_0: \beta_1 = \beta_2 = \dots = \beta_j = \dots = \beta_p$$

ed ipotesi alternativa

$$H_1: b_1, b_2, \dots, b_j, \dots, b_p \text{ non sono tutti uguali}$$

nella condizione che

- le varianze d'errore dei vari gruppi siano omogenee.

Se si assume come vera l'ipotesi nulla che sono tra loro tutti uguali, i coefficienti angolari a confronto rappresentano variazioni casuali dell'unico vero coefficiente angolare, la cui stima migliore è fornita dal **coefficiente di regressione comune  $b_c$** , calcolato come rapporto tra la sommatoria delle codevianze e quella delle devianze totali di X:

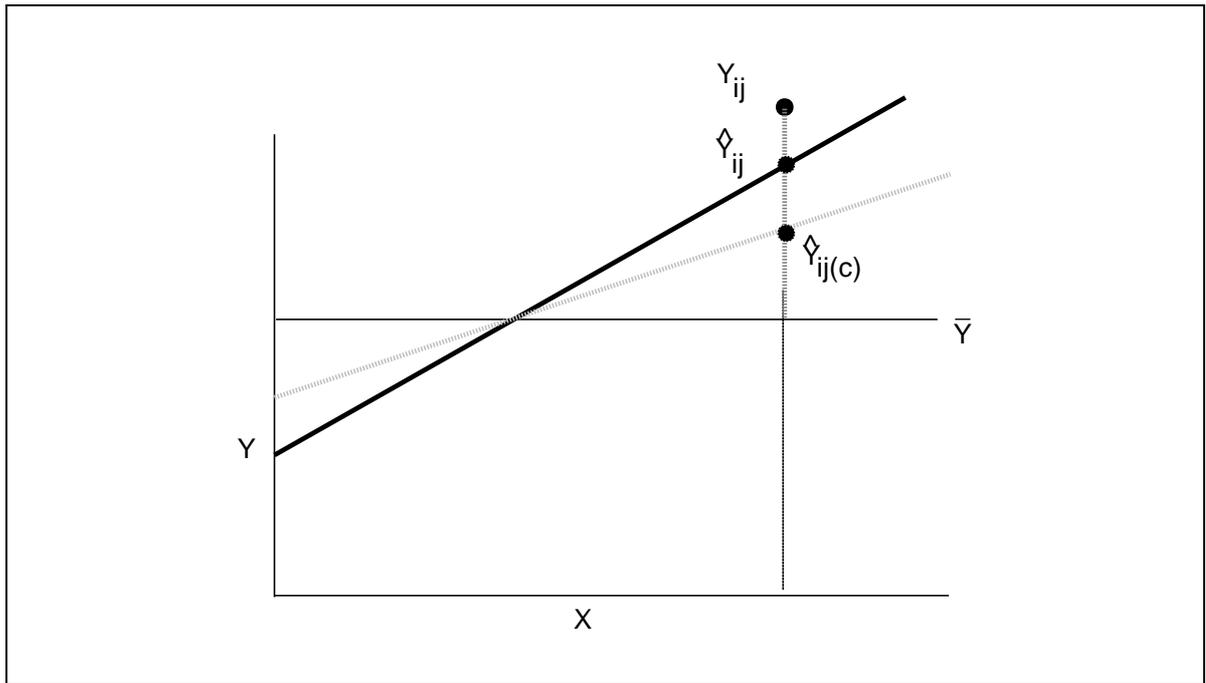
$$b_c = \frac{\sum \sum (X_{ij} - \bar{X}_j) \cdot (Y_{ij} - \bar{Y}_j)}{\sum \sum (X_{ij} - \bar{X}_j)^2}$$

A questo coefficiente angolare comune è associata una quota di devianza della Y ( $Dev_c$  = devianza comune) pari a

$$Dev_c = \frac{\left( \sum \sum (X_{ij} - \bar{X}_j) \cdot (Y_{ij} - \bar{Y}_j) \right)^2}{\sum \sum (X_{ij} - \bar{X}_j)^2}$$

Per ogni j-esimo gruppo la retta di regressione è data da

$$\hat{Y}_{ij} = \bar{Y}_j + b_j \cdot (X_{ij} - \bar{X}_j)$$



Se le varie rette a confronto possono essere considerate tra loro parallele, nello stesso modo la stima del valore medio comune della Y ( $\hat{Y}_{ijc}$ ) per  $X_1$  è data da

$$\hat{Y}_{ijc} = \bar{Y}_j + b_c \cdot (X_{ij} - \bar{X}_j)$$

Come riportato nel grafico precedente,

**lo scostamento di ogni singola osservazione  $Y_{ij}$  dalla media del proprio gruppo  $\bar{Y}_j$**

può essere diviso in tre quote:

- del punto dalla retta del suo gruppo ( $Y_{ij} - \bar{Y}_{ij}$ ),
- della retta del gruppo da quella comune ( $\hat{Y}_{ij} - \hat{Y}_{ijc}$ ),
- della retta comune dalla media generale ( $\hat{Y}_{ijc} - \bar{Y}_j$ )

$$Y_{ij} - \bar{Y}_j = (Y_{ij} - \bar{Y}_{ij}) + (\hat{Y}_{ij} - \hat{Y}_{ijc}) + (\hat{Y}_{ijc} - \bar{Y}_j)$$

Le rispettive devianze, ossia la somma dei quadrati di questi scarti, con  $k$  gruppi o rette a confronto e con un numero totale di osservazioni pari a  $N$  possono essere ripartite nello stesso modo:

1 - devianza totale **entro gruppi** con gdl  **$N-k$**   
corrispondente alla somma delle devianze totali di ogni gruppo,

2 - devianza **residua intorno alle rette separate** con gdl **N-2k**  
 corrispondente alla somma delle devianze d'errore di ogni retta,

3 - devianza **della regressione di ogni retta** con gdl **k**  
 ottenuta sottraendo la 2 alla 1,

4- devianza **dovuta alla retta comune**, con gdl **1**  
 o di parallelismo, data dal rapporto tra il quadrato della somma delle codevianze e le devianze di X,

5- devianza **dovuta alle differenze tra rette**, con gdl **k-1**  
 o di scostamento dal parallelismo, ottenuta sottraendo la 4 alla 3.

Indicando

- la somma dei quadrati degli scarti di  $X_{ij}$  rispetto alla sua media  $\bar{X}_j$  nel gruppo j-esimo con

$$(Sx^2)_j = \sum_i (X_{ij} - \bar{X}_j)^2$$

- la somma dei quadrati degli scarti di  $Y_{ij}$  rispetto alla sua media  $\bar{Y}_j$  con

$$(Sy^2)_j = \sum_i (Y_{ij} - \bar{Y}_j)^2$$

- la somma dei prodotti degli scarti di X e Y rispetto alle loro medie con

$$(\sum xy)_j = \sum (X_{ij} - \bar{X}_j) \cdot (Y_{ij} - \bar{Y}_j)$$

il calcolo delle devianze è mostrato con semplicità.

Con le formule abbreviate, si ottiene

- la devianza entro gruppi con

$$\sum_j (Sy^2)_j S_i(Sy^2)_i$$

- la devianza dovuta alla retta comune o al coefficiente angolare comune con

$$\frac{\left(\sum_i (Sxy)_i\right)^2}{\sum_i (Sx^2)_i}$$

- la devianza dovuta alle differenze tra coefficienti o alle differenze tra rette:

$$\sum_i \frac{(Sxy)_i^2}{(Sx^2)_i} - \frac{\left(\sum_i (Sxy)_i\right)^2}{\sum_i (Sx^2)_i}$$

- la devianza residua intorno alle rette separate:

$$\sum_i (Sy^2)_i - \sum_i \frac{(Sxy)_i^2}{(Sx^2)_i}$$

L'analisi della varianza per verificare la significatività delle differenze tra i coefficienti di regressione è un test **F** con gdl **k-1** e **N-2k**. E' ottenuto mediante il rapporto tra la varianza  $s_A^2$  delle differenze tra coefficienti di regressione lineare (detta anche **varianza di scostamento dalla regressione**) e la varianza  $s^2$  del residuo intorno alle rette separate (o **varianza d'errore**)

$$F_{k-1, n-2k} = \frac{s_A^2}{s^2}$$

ESEMPIO. Per tre gruppi di persone, classificati in giovani, adulti ed anziani, dei quali sono stati rilevati l'altezza (X) in cm ed il peso (Y) in Kg con i valori riportati nelle tre tabelle seguenti

GIOVANI		
Ind.	X	Y
1	190	86
2	170	68
3	165	67
4	178	72
5	185	78
6	193	87
7	188	80
8	174	70

ADULTI		
Ind.	X	Y
1	185	85
2	180	83
3	172	74
4	177	75
5	160	70
6	168	72
7	165	71
---	---	---

ANZIANI		
Ind.	X	Y
1	180	70
2	172	65
3	158	60
4	175	70
5	163	61
6	167	59
7	175	68
8	184	75

calcolare separatamente i parametri delle tre regressioni lineari semplici. Dopo aver verificato se le tre rette possono essere considerate parallele, in caso affermativo calcolare il coefficiente di regressione comune.

Risposta. Ognuno dei tre coefficienti angolari stimati separatamente nei tre gruppi a confronto risulta significativamente diverso da 0:

STIMA E SIGNIFICATIVITA' DEI PARAMETRI **a** e **b** NEI GIOVANI

Parametro	Df	Valore	Err. Standard	t per H <sub>0</sub> : parametro = 0	Prob.
<b>A</b>	1	-59,929	14,535	-4,123	0.0062
<b>B</b>	1	0,754	0,0805	9,365	0.0001

ANALISI DELLA VARIANZA PER LA SIGNIFICATIVITA' DI **b** NEI GIOVANI

	Devianza	Df	Varianza	F	Prob.
Totale	438,000	7			
Della regressione	409,955	1	409,955	87,705	0.0001
D'errore	28,045	6	4,674		

$$R^2 = 0,936 \quad \text{Devianza totale } X = 721,875 \quad \text{Codevianza } XY = 544,294$$

è uguale a 0,754 nel gruppo di giovani.

STIMA E SIGNIFICATIVITA' DEI PARAMETRI **a** e **b** NEGLI ADULTI

Parametro	Df	Valore	Err. Standard	t per H <sub>0</sub> : parametro = 0	Prob.
<b>A</b>	1	-32,344	19,106	-1,693	0.151
<b>B</b>	1	0,627	0,111	5,662	0.0024

ANALISI DELLA VARIANZA PER LA SIGNIFICATIVITA' DI **b** NEGLI ADULTI

	Devianza	Df	Varianza	F	Prob.
Totale	211,428	6			
Della regressione	182,903	1	182,903	32,060	0.0024
D'errore	28,525	5	5,705		

$R^2 = 0,865$       Devianza totale X = 465,714      Codevianza XY = 292,003  
a 0,627 nel gruppo degli adulti.

**STIMA E SIGNIFICATIVITA' DEI PARAMETRI a e b NEGLI ANZIANI**

Parametro	Df	Valore	Err. Standard	t per $H_0$ : parametro = 0	Prob.
<b>A</b>	1	-38841	17,138	-2,266	0.0640
<b>B</b>	1	0,610	0,0997	6,124	0.0009

**ANALISI DELLA VARIANZA PER LA SIGNIFICATIVITA' DI b NEGLI ANZIANI**

	Devianza	Df	Varianza	F	Prob.
Totale	228,000	7			
Della regressione	196,557	1	196,557	35,508	0.0009
D'errore	31,443	6	5,240		

$R^2 = 0,862$       Devianza totale X = 527,500      Codevianza XY = 321,775  
e a 0,610 nel gruppo degli anziani.

**Per verificare se le tre rette si discostano in modo significativo dal parallelismo, si calcolano**

**1 - la devianza totale entro gruppi**

$$438,000 + 211,429 + 228,000 = 877,429$$

che risulta uguale a 877,429 con gdl

$$7 + 6 + 7 = 20$$

uguali a 20;

**2 - la devianza dei residui**

$$28,045 + 28,524 + 31,443 = 88,012$$

che risulta uguale a 88,012 con gdl

$$6 + 5 + 6 = 17$$

uguali a 17;

### 3 - la devianza di regressione di ogni retta

$$877,429 - 88,012 = 789,417$$

che risulta uguale a 789,417 con gdl

$$20 - 17 = 3$$

uguali a 3;

4- la devianza della regressione comune ottenuta con  $\frac{(\sum Cod. XY)^2}{\sum Dev. X}$

$$\frac{(544,294 + 292,003 + 321,775)^2}{721,875 + 465,714 + 527,500} = \frac{1158,072^2}{1715,089} = 781,960$$

che risulta uguale a 781,960 con 1 gdl;

### 5 - la devianza di scostamento dal parallelismo

$$789,417 - 781,960 = 7,457$$

che risulta uguale a 7,457 ed ha 2 gdl (3 - 1).

La sua varianza è uguale a 3,7285 (7,457 / 2) e la varianza d'errore è uguale a 5,177 (88,012 / 17).

Il rapporto

$$F_{(1,17)} = \frac{781,960}{5,177} = 151,045$$

che risulta uguale a 151,045 con df 1 e 17 risulta altamente significativo; di conseguenza la retta comune si discosta in modo significativo da 0.

Il rapporto

$$F_{(2,17)} = \frac{3,7285}{5,177} < 1$$

risulta minore di 1; le tre rette sono parallele, non si discostano dal parallelismo in modo significativo.

Il coefficiente angolare comune  $b_c$  è dato da

$$b_c = \frac{\sum Cod. XY}{\sum Dev. X}$$

e con i dati

$$b_c = \frac{544,294 + 292,003 + 321,775}{721,875 + 465,714 + 527,500} = \frac{1158,072}{1715,089} = 0,675$$

risulta uguale a 0,675.

Per il calcolo della retta comune si può ricorrere a

$$\hat{Y}_{ijc} = \bar{Y}_j + b_c \cdot (X_{ij} - \bar{X}_j)$$

dopo aver stimato la media generale di X e Y, per le N coppie di valori complessivamente rilevati nei K gruppi a confronto.

## 12.15. CONFRONTI MULTIPLI TRA PIU' COEFFICIENTI ANGOLARI

Rifiutata l'ipotesi nulla

$$H_0: b_1 = b_2 = \dots = b_p$$

si tratta di verificare tra quali coppie dei **p** coefficienti angolari **b** la differenza sia significativa.

**La risposta può venire dai confronti multipli a posteriori, con metodi del tutto analoghi a quelli descritti per il confronto tra p medie, quali**

- **il metodo di Tukey per confronti semplici,**
- **il metodo di Scheffé per confronti complessi,**
- **il metodo di Dunnett per il confronto di un controllo con p trattamenti.**

Tra **p** coefficienti angolari (**b<sub>1</sub>, b<sub>2</sub>, ..., b<sub>p</sub>**) è possibile verificare la significatività della differenza tra due qualsiasi **b<sub>1</sub>** e **b<sub>2</sub>** con ipotesi nulla

$$H_0: b_1 = b_2$$

e ipotesi alternativa bilaterale

$$H_1: b_1 \neq b_2$$

con il **metodo di Tukey**

$$q_{a,n,p} = \frac{b_1 - b_2}{s_{b_1 - b_2}}$$

dove

- **q** è il valore critico riportato nella tabella del q studentizzato o una sua evoluzione, che considera anche i passi di distanza tra i ranghi dei valori a confronto,
- **a** è la probabilità prefissata per la significatività,
- **v** sono i gdl di  $s_{b_1 - b_2}$
- **p** è il numero di gruppi a confronto.

Il valore di  $s_{b_1 - b_2}$  è

$$s_{b_1 - b_2} = \sqrt{\frac{p \cdot s_e^2}{2 \cdot \left[ \sum (X_{1i} - \bar{X}_1)^2 + \sum (X_{2i} - \bar{X}_2)^2 \right]}}$$

con gdl corrispondenti a quelli della varianza d'errore, nell'analisi della varianza tra i  $p$  coefficienti angolari a confronto.

Nel test di **Scheffé** e nel test di **Dunnnett**, varia solo la stima della probabilità  $\alpha$  di ogni confronto.

## **12.16. ANALISI DELLA RELAZIONE DOSE-EFFETTO NEL CASO DI Y RIPETUTE:**

### **TEST PER LA LINEARITA' E CALCOLO DELLA RETTA DI REGRESSIONE**

Nella ricerca di laboratorio, spesso si richiede di saggiare la risposta biologica di animali o di vegetali a dosi crescenti di un farmaco o di un tossico. Si impostano esperimenti nei quali vengono somministrate **quantità progressivamente crescenti**, per verificare come varia la loro risposta media.

Il calcolo della retta di regressione dose-risposta prevede almeno due test preliminari, allo scopo di verificare:

- 1 - se esiste una differenza significativa tra il risultato medio dei gruppi a confronto;
- 2 - se la retta di regressione lineare semplice può rappresentare in modo corretto la relazione tra la quantità della dose e le relative risposte medie.

Nel caso in cui queste due risposte siano positive, occorre calcolare la retta che esprime la relazione quantitativa.

Il primo test è un'**analisi della varianza ad 1 criterio di classificazione**; il secondo è il **test per la linearità**. Sono due analisi possibili solamente quando i dati prevedono repliche delle  $Y$ , dette **disposizioni**, per valori di  $X$  predeterminati e costanti.

Per una lettura semplice ed una prima indicazione delle loro caratteristiche, i dati possono essere rappresentati in forma tabellare, con le medesime modalità delle repliche dei trattamenti nell'analisi della varianza ad un criterio di classificazione.

Il calcolo della retta, che esprime la regressione dei valori medi di ogni disposizione, richiede che l'esperimento rispetti 3 vincoli:

- 1) il numero  $n_i$  di repliche in ogni gruppo sia uguale;
- 2) le distanze tra le dosi, misurate nel loro valore reale oppure dopo trasformazione (di solito in log), siano costanti;
- 3) il numero di gruppi sia almeno 4-5, per stimare quale è il tipo di curva (di primo ordine o di ordine superiore) più adeguata a descrivere la risposta media al crescere della dose.

Ogni osservazione ( $Y_{ij}$ ) entro disposizioni, per determinati valori di  $X_i$ , si discosta dalla retta ( $\hat{Y}_{i1}$ ) di una quantità che può essere suddivisa in 2 parti:

- 1 - lo scostamento dell'osservazione ( $Y_{ij}$ ) dalla media della sua disposizione ( $\bar{Y}_i$ ),
- 2 - lo scostamento di quest'ultima dalla retta ( $\hat{Y}_{i1}$ )

$$Y_{ij} - \hat{Y}_i = (Y_{ij} - \bar{Y}_i) + (\bar{Y}_i - \hat{Y}_i)$$

Le devianze relative mantengono la stessa relazione.

**La devianza residua della retta di regressione**  $(Y_{ij} - \hat{Y}_i)^2$  è scomponibile

- **in una prima devianza, dovuta alla dispersione dei singoli valori intorno alle medie della loro**

**disposizione**  $\sum (Y_{ij} - \bar{Y}_i)^2$

- **in una seconda devianza, dovuta agli scarti delle medie di disposizione dalla retta di**

**regressione**  $\sum (\bar{Y}_i - \hat{Y}_i)^2$  :

$$\sum (Y_{ij} - \hat{Y}_i)^2 = \sum (Y_{ij} - \bar{Y}_i)^2 + \sum (\bar{Y}_i - \hat{Y}_i)^2$$

Quando l'analisi della varianza, applicata ovviamente ai soli valori della Y, porta alla conclusione che esiste una differenza altamente significativa tra le medie dei gruppi (classificati sulla base dello stesso valore di X<sub>j</sub>), esiste la condizione logica per verificare, mediante una ulteriore specifica analisi della varianza, se al variare della dose la risposta media cambia in modo lineare.

Per questi primi calcoli, si richiedono 3 stime:

1- la **devianza dovuta alla regressione** (df = 1), mediante la solita formula  $\frac{(Cod. XY)^2}{Dev. X}$  ricordando

che in questo caso il valore di X (la dose) è attribuito a tutti i singoli valori di Y entro la stessa disposizione;

2 - la **devianza tra gruppi** (df = k-1) stimata mediante l'analisi della varianza ad 1 criterio di classificazione;

3 - la **devianza delle medie dalla regressione** (df = k-2) ottenuta dalla relazione

**Devianza delle medie dalla regressione = Devianza tra gruppi - Devianza della regressione;**

4 - la **devianza residua entro disposizioni** (df = n-k) calcolata da

**Devianza d'errore o residuo = Devianza totale - Devianza tra gruppi**

ricordando che,

con la consueta simbologia,

- k = numero di disposizioni
- n = numero totale di osservazioni.

Le devianze da calcolare, con le formule abbreviate e i df relativi, sono riportate nella tabella sottostante:

DEVIANZA	FORMULE ABBREVIATE	DF
<b>Totale</b>	$\sum Y^2 - \frac{(\sum Y)^2}{n}$	<b>n-1</b>
<b>Tra gruppi (disposizioni)</b>	$\sum (\sum Y_i)^2 - \frac{(\sum Y)^2}{n}$	<b>k-1</b>
<b>Della regressione</b>	$\frac{(\sum (X \cdot Y) - \frac{\sum X \cdot \sum Y}{n})^2}{\sum X^2 - \frac{(\sum X)^2}{n}}$	<b>1</b>
<b>Delle medie dalla regressione</b>	<b>Tra gruppi – Della regress .</b>	<b>k-2</b>
<b>Residuo (entro gruppi)</b>	<b>Totale – Tra gruppi</b>	<b>n-k</b>

Mediante le 3 ultime varianze (della regressione, delle medie dalla regressione e residuo entro gruppi), si determinano 2 test F.

#### Il **primo F**

$$F_{1,n-k} = \frac{\text{Varianza dovuta alla regressione}}{\text{Varianza residua entro gruppi}}$$

con df **1** e **n-k**

per verificare l'ipotesi nulla  $\beta = 0$  ovvero se la regressione lineare sia significativa;

#### il **secondo F**:

$$F_{k-2,n-k} = \frac{\text{Varianza delle medie dalla regressione}}{\text{Varianza residua entro gruppi}}$$

con df **k-2** e **n-k**

per verificare l'ipotesi se esistano curve di ordine superiore che siano in grado di rappresentare in modo significativamente migliore la relazione tra dose e risposte medie.

Nel caso in cui questo secondo test risulti significativo, occorre scegliere il tipo di curva più appropriato. Questi ulteriori metodi ed il calcolo dei parametri delle curve di secondo ordine o di

ordine superiore non rientrano tra gli argomenti di questo capitolo (abituamente nei testi di statistica sono presentati dopo la trattazione della regressione multipla).

ESEMPIO. E' dimostrato che l'inquinamento da cromo in dosi subletali agisce in modo negativo sull'accrescimento somatico di molte specie acquatiche.

Con un esperimento di laboratorio, si vuole stabilire la relazione che intercorre tra la concentrazione della sostanza (a dosi crescente in modo lineare: 5, 10, 15, 20, 25) e la risposta biologica in alcuni gruppi di crostacei della stessa specie, dei quali vengono fornite le dimensioni dopo una settimana dalla schiusa delle uova. (Per semplificare i calcoli senza alterare le analisi, è stata fatta una trasformazione lineare dei dati, con la sottrazione di 20 mm a tutte le misure delle Y).

RISPOSTE	Dose (X)					TOTALI
	5	10	15	20	25	
1	10,5	8,4	7,7	5,3	4,6	
2	11,3	8,6	6,9	4,3	5,6	
3	12,1	9,2	5,8	4,8	3,9	
4	11,4	9,1	7,2	5,0	4,8	
$\sum Y_i$	45,3	35,3	27,6	19,4	18,9	146,5
$\bar{Y}_i$	11,325	8,825	6,900	4,850	4,725	7,325
$\sum (Y_i)^2$	507,56	313,70	192,38	94,62	89,36	1197,92

Verificare se l'effetto risente della somministrazione di dosi diverse; in caso positivo, stimare se la retta è adeguata a descrivere la relazione dose-effetto.

Risposta.

Per rispondere ai quesiti proposti, la prima verifica è fornita dall'analisi della varianza ad un criterio di classificazione (benché il calcolo della regressione richieda che il numero di osservazioni (Y) sia uguale in tutti i gruppi). Se il test F non risulta significativo e pertanto non permette di rifiutare l'ipotesi nulla, si deve giungere alla conclusione logica che al variare della dose le risposte medie dei gruppi a confronto non manifestano differenze significative. Di conseguenza, la media generale rappresenta la

stima lineare migliore dell'effetto medio delle varie dosi ed è inutile procedere al calcolo della retta di regressione.

Con i dati dell'esempio, l'analisi della varianza

	DEVIANZA	DF	VARIANZA
Totale	130,938	19	
Tra gruppi	$\sum (\sum Y_i)^2 - \frac{(\sum Y)^2}{n} = 125,265$	4	31,316
Errore	Totale - Tra gruppi = 5,673	15	0,378

permette di calcolare un test F

$$F_{4,15} = \frac{31,316}{0,378} = 84,5$$

che risulta uguale a 84,5 con df 4 e 15.

Ricordando che il valore tabulato di F per  $\alpha = 0.01$  è uguale a 4,89 si deve concludere che le risposte differiscono in modo significativo al variare della dose.

La risposta successiva che occorre fornire è se la retta rappresenta una stima accettabile dell'effetto biologico al crescere della dose, mediante **il test per la linearità**, oppure se è più adeguata una curva di grado superiore.

Ricorrendo alle formule già presentate, con i dati dell'esempio

	DEVIANZA	DF	VARIANZA
Totale	130,938	19	6,891
Tra gruppi (dispos.)	125,265	4	31,316
Dovuta alla regr.	117,992	1	117,992
Delle medie dalla regr.	$125,265 - 117,992 = 7,273$	3	2,424
Residuo entro dispos.	$130,938 - 125,265 = 5,673$	15	0,378

per stimare l'effetto della retta rispetto alla media generale

$$F_{1,15} = \frac{117,992}{0,378} = 312,148$$

si ottiene

- un **primo valore di F** uguale a 312,148 con df 1 e 15.

Esso dimostra che **la regressione lineare semplice è altamente significativa**: la retta passa molto più vicino alle medie dei 5 gruppi che non la media generale. Di conseguenza è molto vantaggioso calcolare, la retta per evidenziare la relazione tra dose e risposta media.

Con un **secondo test F** determinato dal rapporto tra la varianza “delle medie dalla regressione lineare” e la varianza d’errore,

$$F_{3,15} = \frac{2,424}{0,378} = 6,413$$

- si ottiene un valore uguale a 6,413 con df 3 e 15 che risulta significativo, seppure in modo meno evidente del precedente: **una curva di grado superiore si avvicina alle medie delle 5 dosi in modo significativo.**

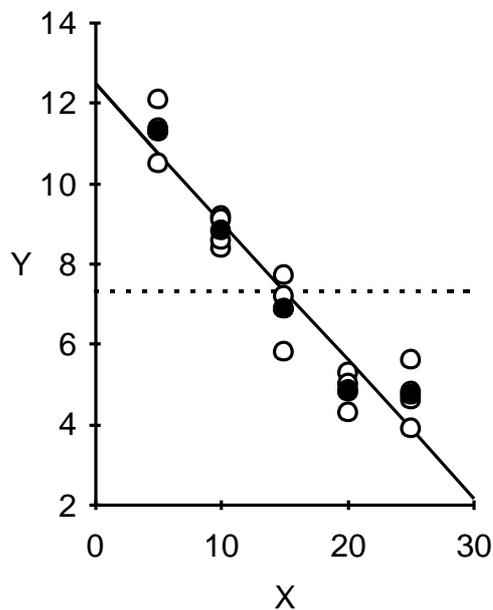


Grafico dei dati dell'esempio con retta di regressione.

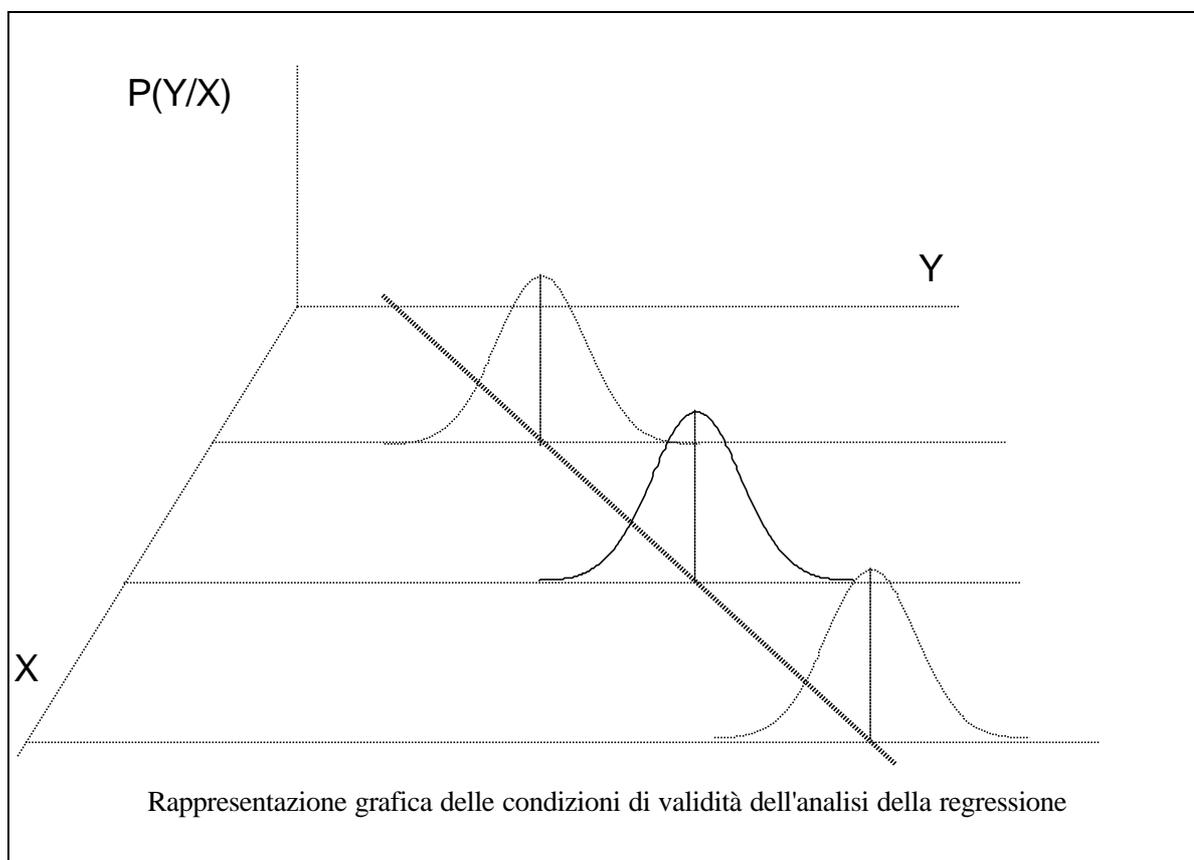
I pallini pieni rappresentano le medie dei gruppi (disposizioni) e il segmento tratteggiato la media generale della variabile Y.

La sua scomposizione per curve di 2°, di 3° e di 4° grado, ognuno con 1 df, potrebbe non risultare significativo. L'interpretazione biologica sarebbe che ognuna di queste curve di ordine superiore, pure passando più vicino alle medie dei 5 gruppi della curva di ordine minore, non rappresenta un miglioramento significativo.

### 12.17. CONDIZIONI DI VALIDITA' DELLA REGRESSIONE, ANALISI DEI RESIDUI E TRASFORMAZIONI

Le condizioni di validità dell'analisi della regressione sono analoghe a quelle già evidenziate per l'analisi della varianza ed il test t di Student: normalità, omoschedasticità, indipendenza dell'errore.

La condizione di **normalità** richiede che **il valore di Y sia normalmente distribuito per ogni valore di X**. E' un'ipotesi già illustrata quando si è discusso della variabilità delle Y e che è facilmente comprensibile nel caso delle Y ripetute per lo stesso valore di X. Come il test t, anche **l'analisi della regressione è robusta**, nel caso di deviazione dalla normalità: fino a quando la distribuzione dei valori di Y per lo stesso valore di X non si differenzia in modo estremo dalla normale, le probabilità calcolate non sono eccessivamente distorte e le inferenze sono ritenute valide. Tale ipotesi di distribuzione normale dei dati coincide con quella di normalità degli errori, cioè degli scarti dal valore medio.



La rappresentazione grafica precedente illustra il concetto di omoschedasticità, mostrando la stessa forma di distribuzione delle Y per le 3 serie di valori di X.

La condizione di **omoschedasticità** richiede che le **varianze delle disposizioni** siano costanti per tutti i valori di X, da quelli minori a quelli maggiori raccolti con il campione: i valori di Y devono avere varianze omogenee per ogni valore di X sperimentato. Sovente succede che all'aumentare delle X si abbia un aumento della varianza delle Y; come già esposto ampiamente in precedenza, le trasformazioni dei dati possono ricostruire la condizione di validità per l'inferenza.

**L'analisi grafica dei residui** permette di evidenziare in modo semplice se il modello di regressione è adeguato ai dati sperimentali e se esistono violazioni delle ipotesi di validità. Sono tecniche elementari, che richiedono un numero di dati non troppo limitato. Di conseguenza, comportano molto tempo per il calcolo e hanno potuto diventare di ampia applicazione con la diffusione dei computer e l'uso di programmi informatici.

I valori residui  $e_i$

$$e_i = Y_i - \hat{Y}_i$$

dati dalla differenza tra valori osservati ( $Y_i$ ) e valori previsti sulla retta ( $\hat{Y}_i$ ) sono posti su un asse orizzontale, da non confondere con la media anche se coincidente, che rappresenta la retta di regressione per  $\beta = 0$ .

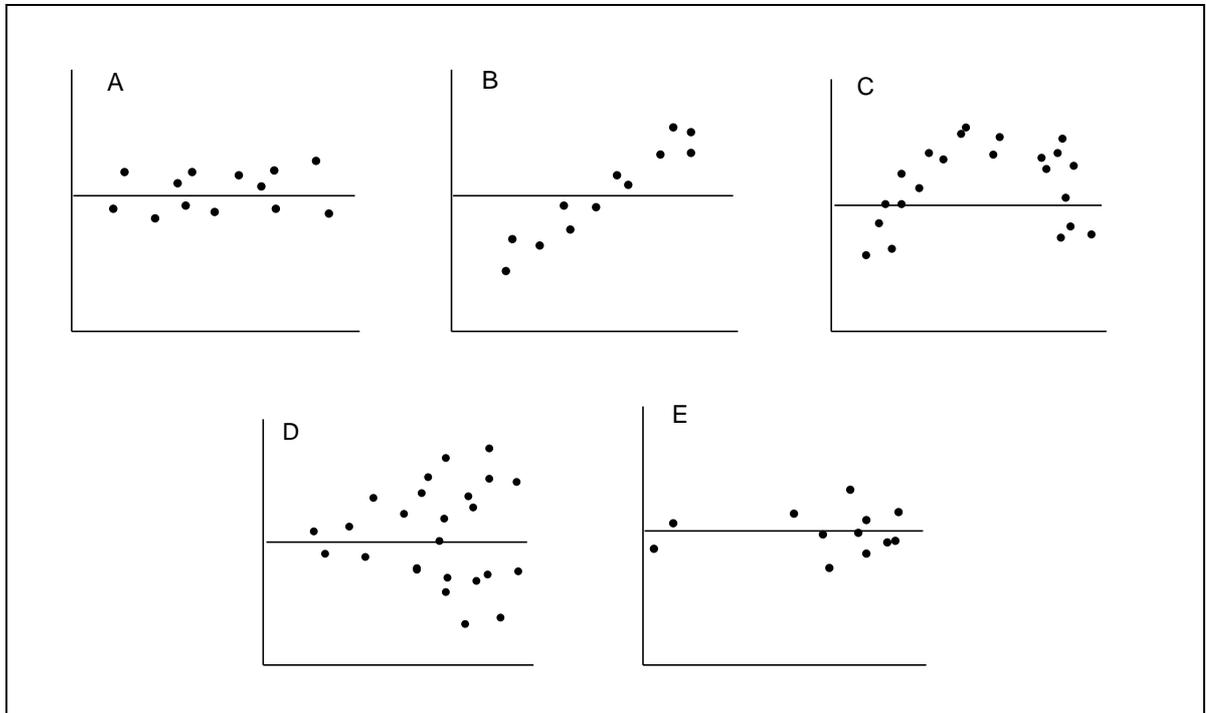
Dopo standardizzazione, ma è possibile anche utilizzare il valore calcolato, i residui ( $e_i$ ) sono collocati in un diagramma cartesiano in cui l'ordinata riporta gli scarti rispetto alla retta e l'ascissa indica il valore corrispondente della variabile indipendente X.

L'ipotesi di **omoschedasticità** è realizzata quando i punti che li rappresentano occupano un'area omogenea lungo tutta la retta; al contrario, si parla di varianze eterogenee quando i punti si allontanano dalla retta in modo non costante. Di norma, si parla di effetto a ventaglio: la variabilità dei residui cresce all'aumentare della X.

La figura A rappresenta la situazione corretta, attesa quando le condizioni di validità sono pienamente rispettate. La figura D evidenzia un progressivo aumento della varianza: per ottenere una inferenza attendibile, occorre trasformare le Y con formule che riducano i valori elevati (logaritmica, in radice quadrata, il reciproco, ...).

L'ipotesi di **normalità** è realizzata quando i residui hanno una distribuzione che può essere approssimata alla distribuzione normale: **gli scarti grandi e piccoli, quelli positivi e i negativi**

dovrebbero essere all'incirca uguali come numero, simmetrici per posizione ed in successione casuale, senza la presenza di valori anomali verso un estremo.



Il grafico rappresenta alcune delle situazioni più diffuse di distribuzione dei residui.

La figura E rappresenta un caso di mancato rispetto della condizione di normalità degli errori. Purtroppo nel caso delle Y ripetute, un numero limitato di repliche (4-6 dati) per lo stesso valore di X non permette di verificare compiutamente l'ipotesi. A parere di vari studiosi, si può presumere che l'analisi della regressione sia corretta quando non si evidenzia una rilevante violazione dell'ipotesi di normalità.

L'**indipendenza** delle osservazioni dipende dal tipo di campionamento, ma è sovente messa in discussione quando i dati sono rilevati in successione cronologica: si può avere un fenomeno di **autocorrelazione** temporale, a causa dell'inerzia o stabilità dei valori osservati, per cui ogni valore è influenzato da quello precedente e determina in parte rilevante quello successivo. Ad esempio, se nell'arco di una giornata si rileva la temperatura ad intervalli costanti di alcuni minuti, si ottiene una successione di valori crescenti fino al momento in cui viene raggiunta la temperatura massima del giorno e poi una successione di valori decrescenti: ogni valore non è casuale, nell'ambito della variabilità dei valori giornalieri, ma risente del valore precedente.

Le figure B e C indicano che la retta calcolata non descrive adeguatamente la dispersione dei dati. Nel caso B, il coefficiente angolare è stimato in modo non corretto per l'influenza di un altro fattore sistematico e lineare; nel caso C, si evidenzia che una curva di secondo grado sarebbe più adeguata della retta.

### Le trasformazioni di Y

Quando le distribuzioni dei dati non rispettano le condizioni di validità, è possibile ricorrere alle trasformazioni.

Sono già state ampiamente discusse nella loro presentazione generale; nel caso della regressione, interessano di solito la variabile Y e quelle più frequenti nella ricerca ambientale sono:

1) la radice quadrata,

$$Y' = \sqrt{Y}$$

quando i dati hanno una **distribuzione poissoniana**, sono cioè **conteggi**; **con frequenze molto basse**, a essa, da parte di molti ricercatori, viene preferita

$$Y' = \sqrt{Y + 0,5}$$

cioè l'aggiunta di una **costante 0,5** soprattutto, ma non necessariamente, quando si ha la presenza di **osservazioni nulle**;

per **stabilizzare la varianza**, nel caso di **crescita moderata all'aumentare di X**, viene usata anche

$$Y' = \sqrt{Y + \frac{3}{8}}$$

oppure

$$Y' = \sqrt{Y} + \sqrt{Y+1}$$

nel caso in cui **Y < 2**

2) l'arcoseno

$$Y' = \arcsin \sqrt{Y}$$

quando i valori hanno una **distribuzione binomiale**, come **proporzioni e percentuali** (Y è la **percentuale**);

con **percentuali molto basse o alte** (vicine a 0% oppure a 100%) è stata proposta la trasformazione

$$Y' = \arcsin \sqrt{\frac{Y + \frac{3}{8}}{n + \frac{3}{4}}}$$

dove

- **Y** è la frequenza assoluta
- **n** sono le dimensioni del campione (**p = Y/n**)  
mentre è stata indicata

$$Y' = \frac{1}{2} \left[ \arcsin \sqrt{\frac{Y}{n+1}} + \arcsin \sqrt{\frac{Y+1}{n+1}} \right]$$

quando **le percentuali sono lontane dai valori estremi**;

- 3) la **trasformazione logaritmica** (con qualsiasi base)

$$Y' = \log Y$$

soprattutto quando si devono **omogeneizzare le varianze, che aumentano molto al crescere di X**;  
con presenza di valori nulli si ricorre a

$$Y' = \log(Y + 1)$$

## CAPITOLO XIII

### CORRELAZIONE E COVARIANZA

#### 13.1 LA CORRELAZIONE

La regressione lineare è finalizzata all'analisi della dipendenza tra due variabili, delle quali

- una (**Y**) è a priori definita come dipendente o effetto,
- l'altra (**X**) è individuata come indipendente o causa.

L'interesse della ricerca è rivolta essenzialmente all'**analisi delle cause** o allo **studio predittivo delle quantità medie di Y**, che si ottengono come risposta al variare di X.

Spesso anche nella ricerca ambientale la relazione di causa-effetto non ha una direzione logica o precisa: potrebbe essere ugualmente applicata nei due sensi, come nell'analisi della relazione tra l'altezza del marito e quella della moglie, sul peso in coppie di fratelli gemelli, tra la quantità di due diverse sostanze inquinanti. Altre volte, la causa può essere individuata in un terzo fattore, che agisce simultaneamente sui primi due, in modo diretto od indiretto, determinando i valori di entrambi e le loro variazioni, come la quantità di polveri sospese nell'aria e la concentrazione di benzene, entrambi dipendenti dall'intensità del traffico. In altre ancora, l'interesse può essere limitato ad una **misura sintetica di come due variabili variano congiuntamente**, per poi andare alla ricerca delle eventuali cause, se la risposta fosse statisticamente significativa. In tutti questi casi, è corretto utilizzare la correlazione.

Più estesamente, è chiamato **coefficiente di correlazione prodotto-momento di Pearson** (*Pearson product-moment correlation coefficient*), perché nella sua espressione algebrica è stato presentato per la prima volta da Karl Pearson (1857-1936) in un lavoro del 1895. Il termine correlazione era già presente nella ricerca statistica anche se Galton (1822-1911) parla di *co-relation*; Galton è stato il primo ad usare il simbolo **r** (chiamato *reversion*), ma per indicare il coefficiente angolare **b** nei suoi studi sull'ereditarietà. La pratica di indicare il coefficiente di correlazione con **r** diventa generale a partire dal 1920. Dalla scuola francese, sovente si utilizzano i termini "**coefficiente di correlazione di Bravais-Pearson**", per ricordare il loro connazionale Bravais (1846), che aveva presentato alcuni concetti importanti di tale metodo cinquanta anni prima di K. Pearson.

Un esempio divertente che spiega chiaramente l'attenzione da porre nella **scelta tra condurre un'analisi della regressione o della correlazione**, dopo che sia stato raccolto un gruppo di dati campionari, è tratto da quanto evidenziato da un ricercatore dei paesi nordici.

Durante i mesi invernali, nelle case in cui è presente un neonato la temperatura viene mantenuta più alta della norma; le cicogne, attratte dal maggior calore, nei periodi più rigidi con maggiore frequenza scelgono i camini di queste case, per nidificare o soffermarsi più a lungo. E' semplice

suddividere un'ampia area rurale in zone con una popolazione equivalente e contare per ognuna il numero dei camini con cicogne (**X**) e quello dei bambini neonati (**Y**). Ricorrere all'analisi della regressione su queste due variabili implica una relazione di causa-effetto. Un tentativo di spiegazione di tale legame, già implicito nella regressione anche se non dichiarato, conduce anche involontariamente alla conclusione che i bambini (se indicati con **Y**) sono portati dalle cicogne (quando indicate con **X**); addirittura con **b** si arriva ad indicare quanti bambini sono portati mediamente da ogni cicogna.

Con la correlazione si afferma solamente che le due variabili variano in modo congiunto, eventualmente per analisi successive alla ricerca delle cause.

**L'analisi della correlazione misura solo il grado di associazione spaziale o temporale dei due fenomeni; ma lascia liberi nella scelta della motivazione logica, nel rapporto tra i due fenomeni. Il coefficiente r è una misura dell'intensità dell'associazione tra due variabili.**

Nella correlazione, le due variabili vengono indicate con **X<sub>1</sub>** e **X<sub>2</sub>**, non più con **X** (causa) e **Y** (effetto), per rendere evidente l'**assenza del concetto di dipendenza funzionale**. (Purtroppo, in vari lavori sono usati ugualmente X e Y, senza voler implicare il concetto della regressione).

L'indice statistico (**+r oppure -r**) misura

- **il tipo (con il segno + o -)**
- **ed il grado (con il valore assoluto)**

**di interdipendenza tra due variabili.**

Il **segno** indica il tipo di associazione:

- positivo quando le due variabili aumentano o diminuiscono insieme,
- negativo quando all'aumento dell'una corrisponde una diminuzione dell'altra.

Il **valore assoluto** varia da 0 a 1:

- è massimo (1) quando c'è una perfetta corrispondenza lineare tra **X<sub>1</sub>** e **X<sub>2</sub>**;
- tende a ridursi al diminuire della corrispondenza ed è zero quando essa è nulla.

**L'indicatore della correlazione r è fondato sulla Codevianza e la Covarianza delle due variabili.**

La Codevianza e la Covarianza tra **X<sub>1</sub>** e **X<sub>2</sub>** ( $Cod_{X_1/X_2}$  e  $Cov_{X_1/X_2}$ ) hanno la proprietà vantaggiosa di contenere queste due informazioni sul tipo (segno) ed il grado (valore) di associazione; ma presentano anche lo svantaggio della regressione, poiché il loro valore risente in modo determinante della scala con la quale le due variabili **X<sub>1</sub>** e **X<sub>2</sub>** sono misurate.

Quantificando il peso in chilogrammi oppure in grammi e l'altezza in metri oppure in centimetri, si ottengono valori assoluti di Codevianza con dimensioni diverse, appunto perché fondati sugli scarti dalle medie ( $x_i - \bar{x}$ ):

$$\text{Cod}_{X_1/X_2} = \sum (x_{1i} - \bar{x}_1) \cdot (x_{2i} - \bar{x}_2)$$

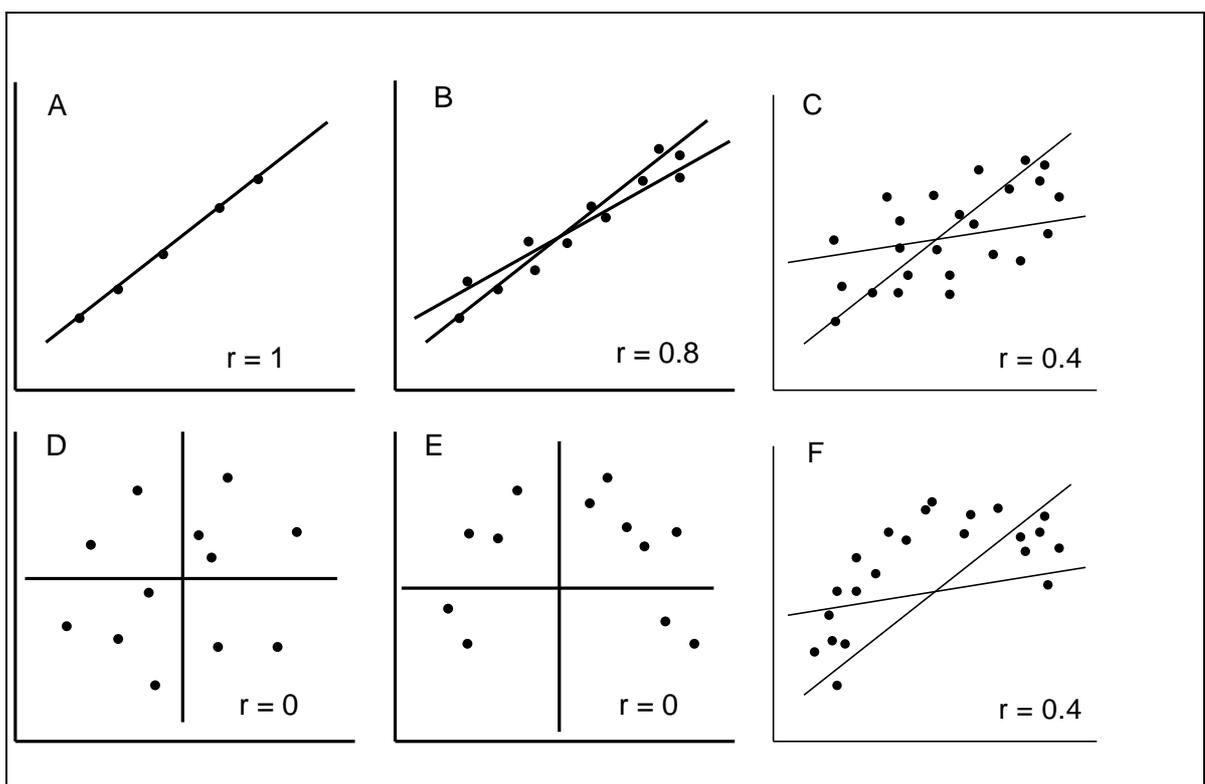
E' possibile **pervenire a valori direttamente comparabili**, qualunque sia la dimensione dei due fenomeni, cioè ottenere valori **adimensionali**, solo ricorrendo ad unità standard, quale appunto la variazione tra  $-1$  e  $+1$ . Si perviene ad essa,

**mediante il rapporto tra la codevianza e la media geometrica delle devianze di  $X_1$  e  $X_2$ .**

$$r = \frac{\sum (X_1 - \bar{X}) \cdot (X_2 - \bar{X})}{\sqrt{\sum (X_1 - \bar{X})^2 \cdot \sum (X_2 - \bar{X})^2}}$$

Per comprendere il significato geometrico dell'indice  $r$  correlazione e derivarne la formula, un approccio semplice è il **confronto tra le due rette di regressione**, calcolate dai valori di  $X_1$  e  $X_2$ :

- la prima calcolata con  $X_1$  usata come variabile dipendente e  $X_2$  come variabile indipendente;
- la seconda scambiando le variabili, quindi utilizzando  $X_2$  come dipendente e  $X_1$  come indipendente.



Le due rette, riportate in ognuna dei 6 figure precedenti, sono calcolate sulle stesse coppie di osservazioni, scambiando appunto  $X_1$  e  $X_2$ . Esse

- si intersecano nel **baricentro della distribuzione**, il punto che rappresenta il valore medio di  $X_1$  e di  $X_2$ ;
- ma non sono identiche o coincidenti (eccetto nella figura A, in cui  $r = 1$ ), poiché entrambe tendono ad avvicinarsi alla media della variabile assunta come dipendente.

Quando le due rette sono tra loro perpendicolari (figura D e figura E) con angoli di  $90^\circ$  e coincidono con le due medie, le due variabili sono indipendenti e tra loro non esiste alcuna correlazione ( $r = 0$ ); inversamente, quando le due rette tendono ad avvicinarsi con un angolo minore, il valore assoluto della correlazione tende ad aumentare (figura C e figura B). Il valore massimo ( $r = 1$ ) viene raggiunto quando le due rette coincidono e l'angolo tra esse è nullo (figura A).

Il segno della correlazione dipende dal coefficiente angolare delle due rette: è positivo, se il loro coefficiente angolare è positivo, mentre è negativo quando il coefficiente angolare è negativo. Pertanto il valore di  $r$  può variare tra  $+1$  e  $-1$ .

(Tra le figure non sono stati riportati valori di  $r$  negativi: la distribuzione dei punti avrebbe evidenziato una diminuzione dei valori della ordinata al crescere di quelli in ascissa e quindi le due rette avrebbero avuto una inclinazione verso il basso all'aumentare dell'ascissa.)

E' importante ricordare che **un valore assoluto basso o nullo di correlazione non deve essere interpretato come assenza di una qualsiasi forma di relazione tra le due variabili:**

- è assente solo una relazione di tipo lineare,
- ma tra esse **possono esistere relazioni di tipo non lineare**, espresse da curve di ordine superiore, tra le quali la più semplice e frequente è quella di secondo grado.

**L'informazione contenuta in  $r$  riguarda solamente la quota espressa da una relazione lineare.**

Per derivare la formula di  $r$  da quanto già evidenziato sulla regressione lineare semplice, è utile ricordare che essa può essere vista come **la media geometrica dei due coefficienti di regressione lineare.**

Infatti, indicando con

- $b_{x1/x2}$  la prima retta di regressione,
- $b_{x2/x1}$  la seconda,

il coefficiente di correlazione  $r$  può essere stimato

come

$$r = \sqrt{b_{x1/x2} \cdot b_{x2/x1}}$$

Poiché

$$b_{(i/j)} = \frac{Cod.ij}{Dev.i}$$

e dato che le due Codevianze sono identiche,

$$r = \sqrt{\frac{Cod.ij}{Dev.i} \cdot \frac{Cod.ji}{Dev.j}}$$

dopo semplificazione,

nella formulazione estesa con la consueta simbologia

si ottiene

$$r = \frac{\sum (X_1 - \bar{X}) \cdot (X_2 - \bar{X})}{\sqrt{\sum (X_1 - \bar{X})^2 \cdot \sum (X_2 - \bar{X})^2}}$$

Per calcolare il coefficiente di correlazione da una serie di rilevazioni, si possono presentare due casi distinti:

- il primo, con **poche osservazioni**, quando i dati sono forniti come **coppie distinte di valori**;
- il secondo, con **molte osservazioni**, quando i dati sono stati raggruppati in **classi di frequenza**.

**La formula sopra riportata è applicabile nel caso di osservazioni singole.**

ESEMPIO. In 18 laghi dell'Appennino Tosco-Emiliano sono state misurate la conducibilità e la concentrazione di **anioni + cationi**, ottenendo le coppie di valori riportati nella tabella

Laghi	Conducibilità (X <sub>1</sub> )	Anioni + Cationi (X <sub>2</sub> )
SILLARA INF.	20	0,328
SILLARA SUP.	22	0,375
SCURO CERR.	22	0,385
VERDAROLO	26	0,445
SQUINCIO	24	0,445
SCURO PARMENSE	28	0,502
PALO	27	0,503
ACUTO	26	0,520
SCURO	29	0,576
COMPIONE INF.	35	0,645
GEMIO INF.	33	0,650
PETRUSCHIA	37	0,675
GEMIO SUP.	34	0,680
SANTO PARMENSE	35	0,746
BICCHIERE	37	0,764
BALLANO	39	0,845
BACCIO	41	0,936
VERDE	45	0,954

Calcolare il coefficiente di correlazione tra queste due variabili

- in modo diretto e
  - mediante i due coefficienti angolari,
- per meglio comprendere l'equivalenza delle due formule.

Risposta.

In modo diretto, con la formula che utilizza le singole coppie di valori

$$r = \frac{\sum (X_{1i} - \bar{X}_1) \cdot (X_{2i} - \bar{X}_2)}{\sqrt{\sum (X_{1i} - \bar{X}_1)^2 \cdot \sum (X_{2i} - \bar{X}_2)^2}}$$

si ottiene un valore di **r**

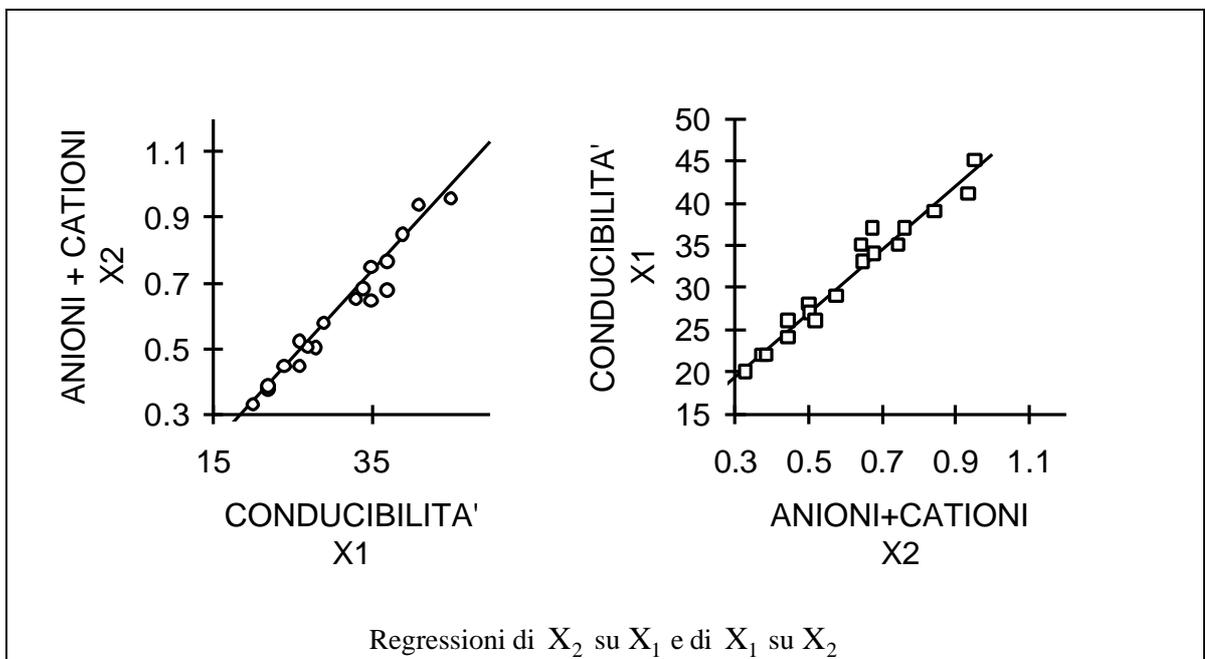
$$r = \frac{22,893}{\sqrt{0,606 \cdot 887,778}} = \frac{22,893}{23,194} = \mathbf{0,987}$$

uguale a 0,987.

Utilizzando i coefficienti angolari delle due regressioni, che dai calcoli risultano

$$b_{X_2/X_1} = 0,026; \quad b_{X_1/X_2} = 37,521$$

e che sono rappresentati nelle due figure seguenti



il coefficiente di correlazione

$$r = \sqrt{b_{X_2/X_1} \cdot b_{X_1/X_2}} = \sqrt{0,026 \cdot 37,521} = 0,9876$$

risulta uguale a 0,9876

con una differenza, dalla stima precedente, determinata dagli arrotondamenti.

Nel caso di **osservazioni raggruppate in classi**, il metodo per calcolare l'indice di correlazione resta sostanzialmente invariato da quello presentato nel capitolo sulla statistica descrittiva.

Per ogni classe, come valore rappresentativo viene assunto il valore centrale; le differenze tra questi valori centrali di ogni classe ed il valore centrale di tutta la distribuzione devono essere moltiplicate per il numero di osservazioni.

Per semplificare i calcoli e per una esatta comprensione del fatto che **le variazioni di scala non incidono assolutamente sul valore di r (che è adimensionale)** è possibile utilizzare non i valori osservati ma gli scarti delle grandezze da una qualsiasi origine arbitraria. Di norma è quella centrale, in quanto determina scarti minimi e simmetrici.

La classe di frequenza centrale o prossima al centro viene indicata con zero e le altre con il numero progressivo, positivo a destra e negativo a sinistra, di distanze unitarie da essa.

Per esempio, la distribuzione di  $X_1$  in 7 classi

50-69	70-89	90-109	110-129	130-149	150-169	170-189
-------	-------	--------	---------	---------	---------	---------

che potrebbe utilizzare i valori centrali relativi (60, 80, 100, 120, 140, 160, 180) per il calcolo dell'indice r di correlazione può essere utilmente trasformata in una scala unitaria

-3	-2	-1	0	+1	+2	+3
----	----	----	---	----	----	----

mentre la distribuzione della variabile  $X_2$  in 6 classi

20-29	30-39	40-49	50-59	60-69	70-79
-------	-------	-------	-------	-------	-------

può essere trasformata in

-3	-2	-1	0	+1	+2
----	----	----	---	----	----

un'altra distribuzione arbitraria equivalente, seppure non simmetrica come la precedente.

E' intuitivo che, con questi nuovi dati, i prodotti e le somme necessarie alla stima del coefficiente di correlazione  $r$  risultano molto semplificati, per un calcolo manuale. Sono quindi tecniche del passato, superate dalle nuove possibilità offerte dall'informatica, con la quale non si pongono problemi di semplificazione dei calcoli. Restano però importanti i concetti: **l'indice di correlazione  $r$  tra due variabili è adimensionale**, fornisce lo stesso valore al variare delle scale di misura.

Ritornando al concetto dell'invarianza del valore di  $r$  rispetto al tipo di scala, nulla muterebbe nel suo valore se la prima o la seconda distribuzione fossero trasformate in una scala ancora differente, come la seguente

0	1	2	3	4	5	...	N
---	---	---	---	---	---	-----	---

Con **dati raggruppati in distribuzioni di frequenze**, il coefficiente di correlazione  $r$  può essere ottenuto con la solita formula

$$r = \frac{Cod. X_1 X_2}{\sqrt{Dev. X_1} \cdot \sqrt{Dev. X_2}}$$

in cui la Codevianza di  $X_1$  e  $X_2$  è data da

$$Cod. X_1 X_2 = \sum f_{X_1 X_2} \cdot d_{X_1} \cdot d_{X_2} - \frac{(\sum f_{X_1} \cdot d_{X_1}) \cdot (\sum f_{X_2} \cdot d_{X_2})}{N}$$

e le due devianze da

$$Dev_{x_1} = \sum f_{x_1} \cdot d_{x_1}^2 - \frac{(\sum f_{x_1} \cdot d_{x_1})^2}{\sum f_{x_1}}$$

la prima

e da

$$Dev_{x_2} = \sum f_{x_2} \cdot d_{x_2}^2 - \frac{(\sum f_{x_2} \cdot d_{x_2})^2}{\sum f_{x_2}}$$

la seconda,

dove

- $d_{X_1}$  e  $d_{X_2}$  sono gli scarti, misurati su una scala arbitraria, dei valori delle classi dall'origine scelta;
- $f_{X_1}$  ed  $f_{X_2}$  sono le frequenze dei valori di  $X_1$  e di  $X_2$  entro ciascuna classe;
- $f_{X_1 X_2}$  sono le frequenze delle coppie  $X_1$ - $X_2$  entro ciascuna coppia di classi.

ESEMPIO. Da una serie di rilevazioni effettuate su un campione d'acqua di 17 laghi (differente dall'elenco precedente)

Laghi	Conducibilità $X_1$	Anioni + Cationi $X_2$
SCURO PR	28	0,502
COMPIONE INF.	35	0,645
SANTO PARMENSE	35	0,746
COMPIONE SUP.	45	0,815
BALLANO	39	0,845
BACCIO	41	0,936
PRADACCIO	53	1,218
OSIGLIA	54	1,259
SANTO MODENESE	79	1,382
NERO PIACENTINO	61	1,530
BUONO	71	1,771
NINFA	96	2,162
PRANDA	108	2,192
CERRETANO	99	2,272
SCAFFAILOLO	108	2,317
PADULE CERRETO	122	2,563
LAME	110	2,616

- costruire la relativa tabella a doppia entrata di distribuzione delle frequenze e
- calcolare da essa il coefficiente di correlazione semplice  $r$ .

Risposta

Con i dati del campione è possibile costruire la tabella a doppia entrata, come quella di seguito riportata

	$X_1$	21-36	37-52	53-68	69-84	85-100	101-116	117-133		
$X_2$	$F_{X_1X_2}$	-3	-2	-1	0	1	2	3	$f_{X_2}$	Medie
0,5-0,8	-3	3 (27)							3	0,631
0,8-1,1	-2		3 (12)						3	0,865
1,1-1,4	-1			2 (2)	1				3	1,286
1,4-1,7	0			1					1	1,530
1,7-2,0	1				1				1	1,771
2,0-2,3	2					2 (4)	1 (4)		3	2,208
2,3-2,6	3						1 (4)	1 (9)	2	2,440
2,6-2,9	4						1 (8)		1	2,616
$F_{X_1}$		3	3	3	2	2	3	1	17	
Medie		32,66	41,66	56,00	75,00	97,50	108,66	122,00		

Nel riquadro interno della tabella sono riportate le  $f_{x_1x_2}$  e (tra parentesi) i prodotti  $f_x \cdot d_{x_1} \cdot d_{x_2}$  che saranno utilizzati per il calcolo della covarianza. Non sono state riportate le frequenze nulle.

I vari passaggi necessari per stimare la Devianza di  $X_1$  dai dati della distribuzione in classi sono riportati nella tabella successiva

$X_1$	$f_x$	$d_x$	$f_x \cdot d_x$	$f_x \cdot d_x^2$
21-37	3	-3	-9	27
37-53	3	-2	-6	12
53-69	3	-1	-3	3
69-85	2	0	0	0
85-101	2	+1	2	2
101-117	3	+2	6	12
117-133	1	+3	3	9
TOTALE	17		-7	65

Con la formula abbreviata

$$\sqrt{Dev_x} = \sqrt{\sum f_x \cdot d_x^2 - \frac{(\sum f_x \cdot d_x)^2}{\sum f_x}} = \sqrt{65 - \frac{(-7)^2}{17}} = 7,88$$

si ottiene la radice quadrata della devianza di  $X_1$ , utile ai calcoli successivi, che è uguale a 7,88.

Seguendo le stesse modalità, il calcolo della Devianza di  $X_2$  e della sua radice quadrata (i cui passaggi sono riportati nella tabella successiva)

$X_2$	$f_x$	$d_x$	$f_x \cdot d_x$	$f_x \cdot d_x^2$
0,5-0,8	3	-3	-9	27
0,8-1,1	3	-2	-6	12
1,1-1,4	3	-1	-3	3
1,4-1,7	1	0	0	0
1,7-2,0	1	+1	1	1
2,0-2,3	3	+2	6	12
2,3-2,6	2	+3	6	18
2,6-2,9	1	+4	4	16
TOTALE	17		-1	89

$$\sqrt{89 - \frac{(-1)^2}{17}} = 9,43$$

fornisce un risultato di 9,43.

Dalle due tabelle è possibile ottenere i dati necessari alla stima della Codevianza

$$Cod_{X_1X_2} = \sum f_{X_1X_2} \cdot d_{X_1} \cdot d_{X_2} - \frac{(\sum f_{X_1} \cdot d_{X_1}) \cdot (\sum f_{X_2} \cdot d_{X_2})}{N} = 72 - \frac{(-7)(-1)}{17} = 71,588$$

dove  $N = \sum f_{X_1} = \sum f_{X_2}$

che risulta uguale a 71,588

Il coefficiente di correlazione r

$$r = \frac{Cod_{X_1X_2}}{\sqrt{Dev_{X_1}} \cdot \sqrt{Dev_{X_2}}} = \frac{71,588}{9,431 \cdot 7,88} = 0,963$$

risulta uguale a 0,963.

E' semplice verificare empiricamente, come dimostrano i calcoli successivi, che anche cambiando i valori di  $d_{X_1}$  e  $d_{X_2}$  il coefficiente di correlazione non cambia.

	$X_1$	21-36	37-52	53-68	69-84	85-100	101-116	117-133		
$X_2$	$F_{X_1X_2}$	0	1	2	3	4	5	6	$f_{X_2}$	Medie
0,5-0,8	0	3							3	0,631
0,8-1,1	1		3 (3)						3	0,865
1,1-1,4	2			2 (8)	1 (6)				3	1,286
1,4-1,7	3			1 (6)					1	1,530
1,7-2,0	4				1 (12)				1	1,771
2,0-2,3	5					2 (40)	1 (25)		3	2,208
2,3-2,6	6						1 (30)	1 (36)	2	2,440
2,6-2,9	7						1 (35)		1	2,616
$F_{X_1}$		3	3	3	2	2	3	1	17	
Medie		32,66	41,66	56,00	75,00	97,50	108,66	122,00		

Si può infatti notare le  $f_{X_1}$  ed  $f_{X_2}$  sono rimaste inalterate, mentre sono cambiate le  $f_{X_1X_2}$

La radice quadrata della Devianza di  $X_1$

i cui passaggi sono riportati nella tabella successiva

$X_1$	$f_x$	$d_x$	$f_x \cdot d_x$	$f_x \cdot d_x^2$
21-37	3	0	0	0
37-53	3	1	3	3
53-69	3	2	6	12
69-85	2	3	6	18
85-101	2	4	8	32
101-117	3	5	15	75
117-133	1	6	6	36
TOTALE	17		44	176

$$\sqrt{\text{Dev}_x} = \sqrt{\sum f_x \cdot d_x^2 - \frac{(\sum f_x \cdot d_x)^2}{\sum f_x}} = \sqrt{176 - \frac{(44)^2}{17}} = \sqrt{176 - 113,882} = \sqrt{62,118} = 7,881$$

risulta uguale a 7,881

e la radice quadrata della devianza di  $X_2$

$X_2$	$f_x$	$d_x$	$f_x \cdot d_x$	$f_x \cdot d_x^2$
0,5-0,8	3	0	0	0
0,8-1,1	3	1	3	3
1,1-1,4	3	2	6	12
1,4-1,7	1	3	3	9
1,7-2,0	1	4	4	16
2,0-2,3	3	5	15	75
2,3-2,6	2	6	12	72
2,6-2,9	1	7	7	49
TOTALE	17		50	236

$$\sqrt{Dev_{X_2}} = \sqrt{\sum f_{X_2} \cdot d_{X_2}^2 - \frac{(\sum f_{X_2} \cdot d_{X_2})^2}{\sum f_{X_2}}} = \sqrt{236 - \frac{(50)^2}{17}} = \sqrt{236 - 147,059} = \sqrt{88,941} = 9,431$$

risulta uguale a 9,431.

La Codevianza di  $X_1$  e  $X_2$

$$Cod_{X_1X_2} = \sum f_{X_1X_2} \cdot d_{X_1} \cdot d_{X_2} - \frac{(\sum f_{X_1} \cdot d_{X_1}) \cdot (\sum f_{X_2} \cdot d_{X_2})}{N} =$$

dove  $N = \sum f_{X_1} = \sum f_{X_2}$

$$= 201 - \frac{44 \cdot 50}{17} = 201 - 129,412 = 71,588$$

risulta uguale a 71,588.

Essendo rimaste invariate sia la Codevianza che le Devianze, il coefficiente di correlazione semplice  $r$  non può che rimanere identico.

### 13.2. CONDIZIONI DI VALIDITA' E SIGNIFICATIVITA' DI $r$ CON $r = 0$ E CON $r \neq 0$

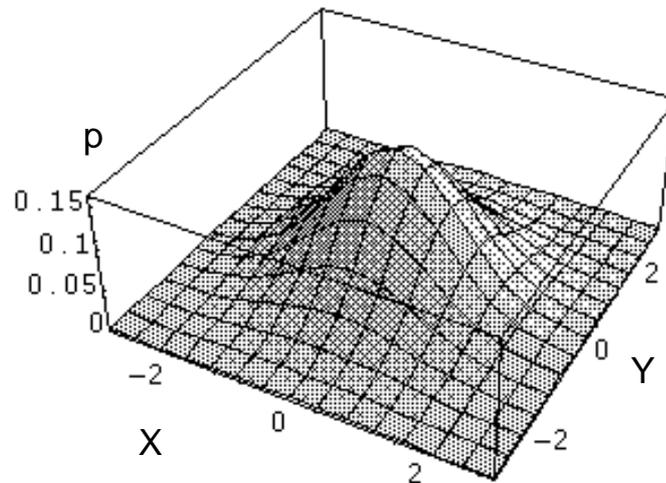
Le **condizioni di validità** della correlazione, il cui indice nel caso di una popolazione è indicato con  $r$  (rho), sono le stesse della regressione. Tuttavia, mentre nella regressione sono applicate solo alla variabile  $Y$ , nel caso della correlazione, che utilizza indistintamente entrambe le variabili, **richiede che sia  $X_1$  che  $X_2$  siano distribuite in modo approssimativamente normale.**

Con due variabili, l'ipotesi di normalità della distribuzione pretende la **distribuzione normale bivariata**, che è un'estensione a tre dimensioni della curva normale.

Mentre la superficie di una distribuzione univariata è determinata in modo compiuto da due parametri (media  $m$  e deviazione standard  $s$ ), la superficie normale bivariata è determinata da cinque parametri:

- media e deviazione standard della variabile  $X_1$ ,
- media e deviazione standard della variabile  $X_2$ ,
- coefficiente di correlazione ( $r$ ) tra  $X_1$  e  $X_2$ .

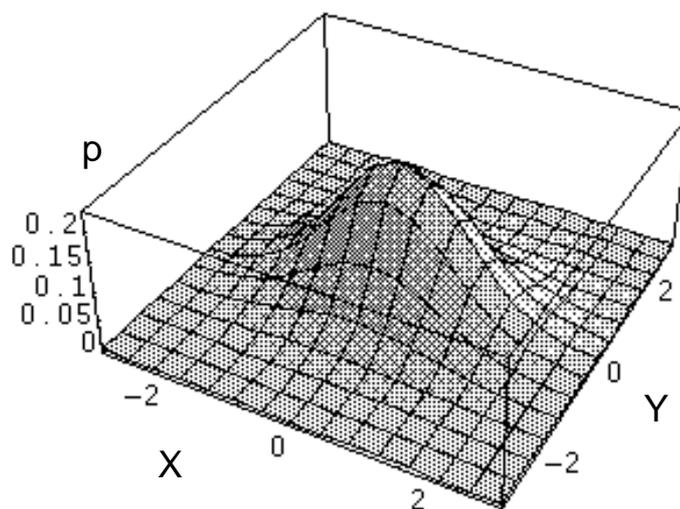
La sua rappresentazione grafica, nel caso in cui non esista correlazione ( $r = 0$ ) tra le due variabili ed esse abbiano varianza uguale, determina una figura come quella sottoriportata.



Distribuzione normale biviariata  
 X e Y sono due variabili **indipendenti** ( $\rho = 0$ ) di **uguale varianza** ( $\sigma_X^2 = \sigma_Y^2 = 1$ )

La distribuzione normale biviariata assume la forma di una collina di forma circolare, che degrada nello stesso modo su tutti i versanti; la pendenza dipende dal valore della varianza.

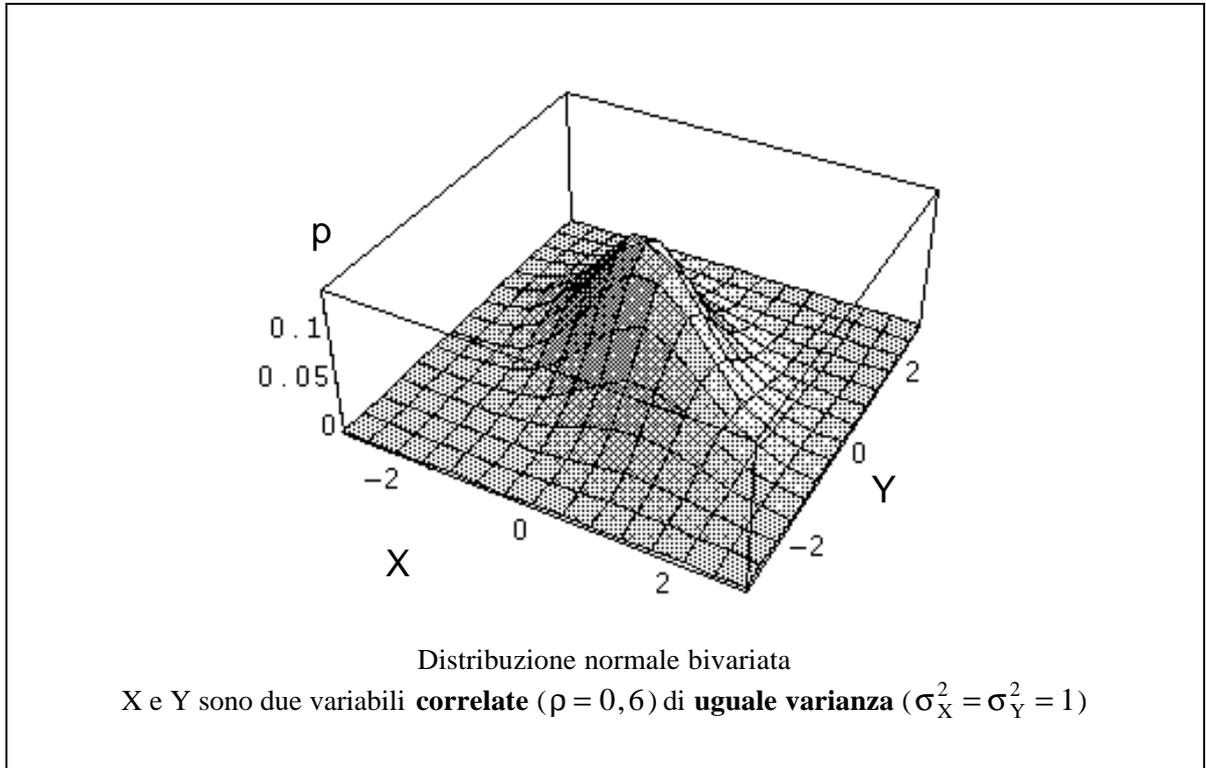
Quando le varianze sono diverse, sempre nel caso che non esista correlazione, la rappresentazione



Distribuzione normale biviariata  
 X e Y sono due variabili **indipendenti** ( $\rho = 0$ ) con **varianze diverse** ( $\sigma_X^2 = 1$ ;  $\sigma_Y^2 = 0,7$ )

grafica assume la forma di una collina a pendenze diverse, con un declino più rapido dove la varianza è minore.

Quando esiste correlazione, come nella figura successiva, la distribuzione bivariata tende ad assumere la forma di una cresta di montagna, distribuita lungo la diagonale, da un angolo a quello opposto. La cresta è tanto più sottile quanto più alto è il valore  $r$  della correlazione.



Con  $r = 1$  la rappresentazione grafica diventa un piano perpendicolare alla base, posto in diagonale rispetto alle ascisse e alle ordinate. Il segno della correlazione determina solo la direzione di tale piano rispetto alla base.

Dopo il calcolo di un **coefficiente di correlazione  $r$**  sempre valido come indice che misura la relazione tra due variabili in quanto solo descrittivo come il calcolo di una media o di una varianza, può porsi il **duplice problema** della sua **significatività**, cioè di verificare

- a) l'ipotesi nulla  $H_0: r = 0$  (**non significativamente diverso da zero**)
- b) l'ipotesi nulla  $H_0: r = r_0$  (**non significativamente diverso da un qualsiasi valore prefissato, ma diverso da zero**)

con ipotesi alternativa bilaterale oppure unilaterale in entrambi i casi.

A differenza dei test sulla media e sul coefficiente angolare **b** (oppure l'intercetta **a**), che possono assumere qualsiasi valore e quindi essere sempre distribuiti normalmente rispetto al valore della

popolazione, un test di significatività pone problemi differenti di validità se intende verificare l'ipotesi nulla

a)  $r = 0$

b)  $r \neq 0$ .

Nel primo caso ( $r = 0$ ), i valori campionari  $r$  possono essere assunti come distribuiti in modo **approssimativamente normale e simmetrico** rispetto alla correlazione della popolazione ( $r$ ).

Nel secondo caso ( $r \neq 0$ ), i valori campionari  $r$  si distribuiscono in modo sicuramente asimmetrico intorno alla correlazione della popolazione ( $r$ ) e in modo tanto più accentuato quanto più essa si allontana da zero e si avvicina a uno dei due estremi ( $-1$  o  $+1$ ). E' intuitivo che, considerando ad esempio risultati positivi, con un valore reale di  $r = 0,9$  il valore campionario  $r$  non potrà mai superare 1, mentre potrebbe essere 6 se non 5 oppure 4, in funzione del numero di dati

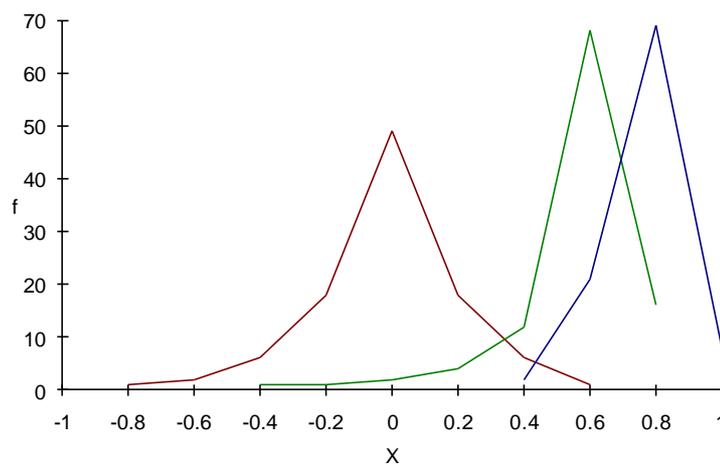


Grafico delle distribuzioni campinarie di 3 coefficienti di correlazione.  
La distribuzione è simmetrica solo quando il suo valore atteso ( $\rho$ ) è zero.

In questo secondo caso, occorre procedere ad una trasformazione di  $r$ , per rispettare la condizioni di validità.

**VALORI CRITICI IN TEST BILATERALE  
DEL COEFFICIENTE DI CORRELAZIONE SEMPLICE  $r$**

(DF = N-2) CON IPOTESI  $H_0: r = 0$

(I VALORI RIPORTATI CORRISPONDONO A QUELLI  
DELLE PROBABILITA' 0.025, 0.005, 0.0005, IN UN TEST UNILATERALE)

DF	a=0.05	a=0.01	a=0.001
1	0,9969	0,9999	1,0000
2	0,9500	0,9900	0,9990
3	0,8783	0,9587	0,9911
4	0,8114	0,9172	0,9741
5	0,7545	0,8745	0,9509
6	0,7067	0,8343	0,9249
7	0,6664	0,7977	0,8983
8	0,6319	0,7646	0,8721
9	0,6021	0,7348	0,8471
10	0,5760	0,7079	0,8233
11	0,5529	0,6835	0,8010
12	0,5324	0,6614	0,7800
13	0,5139	0,6411	0,7604
14	0,4973	0,6226	0,7419
15	0,4821	0,6055	0,7247
16	0,4683	0,5897	0,7084
17	0,4555	0,5751	0,6932
18	0,4438	0,5614	0,6788
19	0,4329	0,5487	0,6652
20	0,4227	0,5368	0,6524
21	0,4132	0,5256	0,6402
22	0,4044	0,5151	0,6287
23	0,3961	0,5052	0,6177
24	0,3882	0,4958	0,6073
25	0,3809	0,4869	0,5974
26	0,3739	0,4785	0,5880
27	0,3673	0,4705	0,5790
28	0,3610	0,4629	0,5703
29	0,3550	0,4556	0,5620
30	0,3494	0,4487	0,5541

DF	a=0.05	a=0.01	a=0.001
35	0,3246	0,4182	0,5189
40	0,3044	0,3932	0,4896
45	0,2875	0,3721	0,4647
50	0,2732	0,3541	0,4433
55	0,2609	0,3385	0,4245
60	0,2500	0,3248	0,4079
65	0,2405	0,3127	0,3911
70	0,2319	0,3017	0,3798
75	0,2242	0,2919	0,3678
80	0,2172	0,2830	0,3569
85	0,2108	0,2748	0,3468
90	0,2050	0,2673	0,3376
95	0,1996	0,2604	0,3291
100	0,1946	0,2540	0,3211
110	0,1857	0,2425	0,3069
120	0,1779	0,2324	0,2943
130	0,1710	0,2235	0,2832
140	0,1648	0,2155	0,2733
150	0,1593	0,2083	0,2643
160	0,1543	0,2019	0,2562
170	0,1497	0,1959	0,2488
180	0,1455	0,1905	0,2420
190	0,1417	0,1855	0,2357
200	0,1381	0,1809	0,2299
300	0,113	0,148	0,188
400	0,098	0,128	0,164
500	0,088	0,115	0,146
600	0,080	0,105	0,134
700	0,074	0,097	0,124
800	0,069	0,091	0,116
900	0,065	0,086	0,109
1000	0,062	0,081	0,104

**VALORI CRITICI IN TEST UNILATERALE**  
**DEL COEFFICIENTE DI CORRELAZIONE SEMPLICE  $r$**

(DF = N-2) CON IPOTESI  $H_0: r = 0$

(I VALORI RIPORTATI CORRISPONDONO A QUELLI  
DELLE PROBABILITA' 0.10, 0.02, 0.002, IN UN TEST BILATERALE)

DF	a=0.05	a=0.01	a=0.001
1	0,988	1,000	1,000
2	0,900	0,980	0,998
3	0,805	0,934	0,986
4	0,729	0,882	0,963
5	0,669	0,833	0,935
6	0,621	0,789	0,905
7	0,582	0,750	0,875
8	0,549	0,715	0,847
9	0,521	0,685	0,820
10	0,497	0,658	0,795
11	0,476	0,634	0,772
12	0,457	0,612	0,750
13	0,441	0,592	0,730
14	0,426	0,574	0,711
15	0,412	0,558	0,694
16	0,400	0,542	0,678
17	0,389	0,529	0,662
18	0,378	0,515	0,648
19	0,369	0,503	0,635
20	0,360	0,492	0,622
21	0,352	0,482	0,610
22	0,344	0,472	0,599
23	0,337	0,462	0,588
24	0,330	0,453	0,578
25	0,323	0,445	0,568
26	0,317	0,437	0,559
27	0,311	0,430	0,550
28	0,306	0,423	0,541
29	0,301	0,416	0,533
30	0,296	0,409	0,526

DF	a=0.05	a=0.01	a=0.001
35	0,275	0,381	0,492
40	0,257	0,358	0,463
45	0,243	0,338	0,439
50	0,231	0,322	0,449
55	0,220	0,307	0,401
60	0,211	0,295	0,385
65	0,202	0,284	0,371
70	0,195	0,274	0,358
75	0,189	0,264	0,347
80	0,183	0,257	0,336
85	0,178	0,249	0,327
90	0,173	0,242	0,318
95	0,168	0,236	0,310
100	0,164	0,230	0,303
110	0,156	0,220	0,289
120	0,150	0,210	0,277
130	0,144	0,202	0,267
140	0,139	0,195	0,257
150	0,134	0,189	0,249
160	0,130	0,183	0,241
170	0,126	0,177	0,234
180	0,122	0,172	0,228
190	0,119	0,168	0,222
200	0,116	0,164	0,216
300	0,095	0,134	0,177
400	0,082	0,116	0,154
500	0,074	0,104	0,138
600	0,067	0,095	0,126
700	0,062	0,088	0,116
800	0,058	0,082	0,109
900	0,055	0,077	0,103
1000	0,052	0,073	0,098

Quando l'ipotesi nulla è

$$H_0: r = 0$$

la significatività del coefficiente angolare  $r$  può essere verificata con 3 modalità, che ovviamente forniscono risultati identici:

- 1 – la tabella dei valori di  $r$ , in funzione di  $a$  e dei gdl (oppure del numero  $n$  di osservazioni),
- 2 – il test  $F$  di Fisher-Snedecor,
- 3 – il test  $t$  di Student.

La prima modalità utilizza le tabelle sinottiche del valore di  $r$ , con gradi di libertà  $n-2$ , come sono stati riportati nelle pagine precedenti. Di conseguenza, è evidente che occorrono almeno 3 coppie d'osservazioni ( $DF = 1$ ).

La semplice lettura dei valori critici alle probabilità  $a = 0.05$ ,  $a = 0.01$  e  $a = 0.001$

DF	$a = 0.05$	$a = 0.01$	$a = 0.001$
3	0,8783	0,9587	0,9911
200	0,1381	0,1809	0,2299
1000	0,062	0,081	0,104

mostra come sia errata l'affermazione semplicistica, riportata su alcuni testi, che un valore di correlazione  $r = 0,3$  sia indicativamente basso e un valore  $r = 0,5$  sia alto.

La significatività della correlazione è fortemente influenzata dai DF, in modo molto più marcato di quanto avviene nelle distribuzioni  $t$  ed  $F$ .

Dal semplice confronto delle due serie riportate nella tabellina precedente, risulta evidente che,

- con **pochi dati**, potrebbe non essere significativo alla probabilità  $a = 0.05$  un valore di  $r$  apparentemente alto quale **0,85**;
- con **molti dati**, potrebbe essere altamente significativo, alla probabilità  $a = 0.001$ , anche un valore apparentemente basso, quale **0,25**.

Pochi testi riportano i valori critici di  $r$ , **validi per verificare l'ipotesi nulla  $H_0: r = 0$** ; quasi sempre si deve ricorrere alla distribuzione  $F$  o a quella  $t$  che tutti i testi, anche elementari, riportano. Pure i programmi informatici, insieme con il valore di  $r$ , riportano la probabilità di  $F$  e/o di  $t$ .

Ricorrendo ai concetti spiegati nella regressione lineare semplice, anche nella verifica dell'ipotesi nulla relativa alla correlazione

$$H_0: r = 0$$

il test  $F$ , con gdl **1** e  **$n-2$** ,

$$F_{1,n-2} = \frac{\frac{r^2}{1}}{\frac{1-r^2}{n-2}}$$

è dato dal rapporto tra

- la **varianza dovuta alla regressione** (la devianza  $r^2 / 1$  df) e
- la **varianza d'errore** (la devianza d'errore  $1 - r^2 / n-2$  df)

La formula semplificata diventa

$$F_{1,n-2} = \frac{r^2 \cdot (n-2)}{1-r^2}$$

Con il test t, che ha df **n-2**,  
ricordando nuovamente che

$$t_{n-2} = \sqrt{F_{1,n-2}}$$

la formula abitualmente utilizzata

è

$$t_{(n-2)} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Con il test **F**, è possibile

- sia la verifica dell'ipotesi alternativa **H<sub>1</sub>** bilaterale

$$H_1: r \neq 0$$

- sia la verifica dell'ipotesi alternativa **H<sub>1</sub>** unilaterale

$$H_1: r > 0 \quad \text{oppure} \quad H_1: r < 0$$

assumendo sempre in una distribuzione bilaterale al posto delle probabilità 0.05, 0,01 e 0.001 rispettivamente le probabilità 0.10, 0.02, 0.002, come nelle tabelle precedenti sui valori critici di **r**. Ma è di più difficile comprensione, per chi non abbia ancora abbastanza familiarità con i test statistici, perché la distribuzione F con pochi gdl, come di solito nella pratica sperimentale, è asimmetrica.

La distribuzione **t**, in quanto simmetrica come la distribuzione **z**, permette di meglio comprendere la scelta delle probabilità in rapporto alla direzione dell'ipotesi alternativa. Per molti è quindi preferibile al test F, in particolare in test unilaterali, pure fornendo valori identici ai due metodi prima presentati.

ESEMPIO 1. La tavola sinottica di  $r$  per test bilaterali, con **df 15** alla probabilità  $\alpha = 0.05$ , riporta il valore di **0,4821**.

Verificare la corrispondenza con il valori critici

a) della distribuzione **F** e

b) della **t** di Student,

che possono essere rintracciati nelle tabelle relative.

Risposta.

a) Con  $r = 0,4821$  e  $n = 17$

la verifica dell'ipotesi nulla

$$H_0: r = 0$$

con ipotesi alternativa bilaterale

$$H_1: r \neq 0$$

mediante il test **F**

$$F_{1, n-2} = \frac{r^2 \cdot (n-2)}{1-r^2}$$

fornisce un risultato

$$F_{1,15} = \frac{0,4821^2 \cdot 15}{1-0,4821^2} = \frac{3,4486}{0,768} = 4,539$$

uguale a 4,539.

b) Mediante il test **t**

$$t_{(n-2)} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

fornisce

$$t_{(15)} = \frac{0,4821 \cdot \sqrt{15}}{\sqrt{1-0,4821^2}} = \frac{0,4821 \cdot 3,873}{\sqrt{0,7676}} = \frac{1,867}{0,876} = 2,13$$

un risultato uguale a 2,13.

E' semplice verificare, sulle tabelle dei valori critici di **F** e di **t**, che i due risultati corrispondono esattamente ai valori riportati per la probabilità  $\alpha = 0.05$  in una distribuzione bilaterale e che

$$2,13^2 = 4,539$$

a meno delle approssimazioni dei calcoli.

Per un test di significatività del coefficiente di correlazione r rispetto ad un qualsiasi valore di  $r_0$  diverso da zero, quindi per verificare l'ipotesi nulla

$$H_0: r = r_0$$

a causa dei motivi prima illustrati il valore di  $r$  deve essere trasformato. Tra le diverse proposte, è ancora molto diffusa l'utilizzazione di quella di R. A. Fisher presentata nel 1915 nel dibattito sui grandi campioni (vedi: *Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika*, 10: 507-521) e nel 1921 per i piccoli campioni (vedi: *On the "probable error" of a coefficient of correlation deduced a small sample. Metron* 1: 3-32).

Il valore di  $r$  è trasformato in un valore  $z$  (zeta minuscolo) mediante

$$z = 0.5 \cdot \ln \left( \frac{1+r}{1-r} \right)$$

Con essa,

- i valori positivi di  $r$ , che ovviamente variano da  $0$  a  $+1$ , cadono tra  $0$  e  $+\infty$
- i valori negativi di  $r$ , che ovviamente variano da  $0$  a  $-1$ , cadono tra  $0$  e  $+\infty$

in modo simmetrico. In realtà, nella pratica sperimentale dove i valori di  $r$  asintoticamente vicini a  $1$  sono rari, la variazione cade in un intervallo minore di poche unità, in modo simmetrico intorno alla zero.

Ad esempio

- $r = +0,88$

$$z = 0,5 \cdot \ln \left( \frac{1+0,88}{1-0,88} \right) = 0,5 \cdot \ln \frac{1,88}{0,12} = 0,5 \cdot \ln 15,66 = 0,5 \cdot 2,75 = +1,375$$

diventa  $z = 1,375$

- $r = +0,98$

$$z = 0,5 \cdot \ln \left( \frac{1+0,98}{1-0,02} \right) = 0,5 \cdot \ln \frac{1,98}{0,02} = 0,5 \cdot \ln 99 = 0,5 \cdot 4,595 = +2,2975$$

diventa  $z = +2,2975$

mentre

- $r = -0,88$

$$z = 0,5 \cdot \ln \left( \frac{1+(-0,88)}{1-(-0,88)} \right) = 0,5 \cdot \ln \frac{0,12}{1,88} = 0,5 \cdot \ln 0,0638 = 0,5 \cdot (-2,75) = -1,375$$

diventa  $z = -1,375$

-  $r = -0,98$

$$z = 0,5 \cdot \ln\left(\frac{1 + (-0,98)}{1 - (-0,98)}\right) = 0,5 \cdot \ln\frac{0,02}{1,98} = 0,5 \cdot \ln 0,0101 = 0,5 \cdot (-4,595) = -2,2975$$

diventa  $z = -2,2975$

Anche il valore teorico od atteso di confronto ( $r_0$ ) è trasformato nello stesso modo e viene indicato con  $Z$  (zeta minuscolo dell'alfabeto greco).

La verifica di una differenza significativa tra un generico valore campionario  $r$  e il valore atteso  $r_0$ , con ipotesi nulla

$$H_0: r = r_0$$

ed ipotesi alternativa bilaterale oppure unilaterale, è quindi effettuata con la distribuzione normale  $Z$

$$Z = \frac{z - Z}{S_z}$$

dove

- $Z$  (maiuscola) è il valore che serve per stimare la probabilità  $a$  nella distribuzione normale,
- $z$  (minuscola) è il valore di  $r$  trasformato,
- $Z$  (zeta greca, minuscola) è il valore di  $r_0$  trasformato,
- $S_z$  è l'**errore standard** di questa differenza (poiché  $r$  e  $r_0$  sono valori medi), dato approssimativamente da

$$\sigma_z = \sqrt{\frac{1}{n-3}}$$

ESEMPIO 2. Sulla base di numerosi campionamenti, su una rivista scientifica si afferma che la correlazione tra la presenza quantitativa della specie A e della specie B è positiva e pari a 0,85. Da una rilevazione campionaria con 30 osservazioni, il valore di  $r$  è risultato uguale a **+0,71**.

C'è motivo di ritenere che in questo caso si abbia un valore correlazione significativamente diversa?

Risposta

Per verificare l'ipotesi nulla

$$H_0: r = +0,85$$

con ipotesi alternativa bilaterale

$$H_1: r \neq +0,85$$

per applicare la formula

$$Z = \frac{z - Z}{s_z}$$

- dapprima si deve trasformare in  $z$  il valore  $r = +0,71$

$$z = 0,5 \cdot \ln\left(\frac{1+0,71}{1-0,71}\right) = 0,5 \cdot \ln\left(\frac{1,71}{0,29}\right) = 0,5 \cdot \ln 5,8965 = 0,5 \cdot 1,7744 = +0,887$$

ottenendo  $z = +0,887$

- successivamente si deve trasformare in  $z$  il valore  $r_0 = +0,85$

$$z = 0,5 \cdot \ln\left(\frac{1+0,85}{1-0,85}\right) = 0,5 \cdot \ln\left(\frac{1,85}{0,15}\right) = 0,5 \cdot \ln 12,3333 = 0,5 \cdot 2,5123 = +1,256$$

ottenendo  $z = +1,256$

- e, con  $n = 30$ , si calcola l'errore standard  $s_z$

$$s_z = \sqrt{\frac{1}{30-3}} = \sqrt{0,037} = 0,192$$

Per la significatività della differenza tra valore osservato ( $r = +0,71$ ) e valore arreso ( $r_0 = +0,85$ ), si ottiene

$$Z = \frac{0,887 - 1,256}{0,192} = -\frac{0,369}{0,192} = -1,92$$

un valore di  $Z = -1,92$ .

In una distribuzione normale bilaterale è associato ad una probabilità  $\alpha = 0.055$ ; di conseguenza, il test non risulta significativo.

Se il **test** fosse stato **unilaterale**, cioè se vi fosse stato motivo di chiedersi se il valore calcolato fosse significativamente minore di quello stimato,

con ipotesi alternativa unilaterale fosse stata

$$H_0: r < r_0$$

il procedimento di calcolo sarebbe stato identico. Sarebbe variata solo la lettura della probabilità  $\alpha$ , che in una distribuzione unilaterale sarebbe risultata uguale a 0.027 e quindi avrebbe determinato un test significativo.

### 13.3. SIGNIFICATIVITA' DELLA RETTA CON $R^2$ ?

Alla fine del capitolo sono riportati alcuni output di programmi informatici sulla regressione lineare semplice. Insieme con le risposte sulla significatività dei parametri

- **a** (intercetta),
- **b** (coefficiente angolare),
- è riportato il valore di  $R^2$  (*R-square*).

Vari ricercatori, per valutare la significatività della retta di regressione utilizzano non il relativo test **t** o il test **F**, il cui valore è sempre riportato come illustrato nel capitolo precedente, ma semplicemente riportano il valore di **r** come

$$r = \sqrt{R^2}$$

stimandone la significatività.

Il risultato numerico è identico a quello effettuato sulla retta, poiché il valore di **F**,

- sia nel test per la retta con coefficiente angolare **b**,
- sia in quello per la correlazione **r**

è dato dal rapporto tra la devianza della regressione e la devianza d'errore,

$$F = \frac{\text{Varianza della regressione}}{\text{Varianza d'errore}}$$

seppure il concetto sovente sia nascosto nelle formule abbreviate, di solito utilizzate.

Ad esempio, con le misure di peso ed altezza rilevati su 7 giovani donne

<b>Peso (Y) in Kg.</b>	52	68	75	71	63	59	57
<b>Altezza (X) in cm.</b>	160	178	183	180	166	175	162

è stata calcolata la retta di regressione

$$\hat{Y} = -73,354 + 0,796 X$$

La significatività del coefficiente angolare **b** per verificare l'ipotesi nulla

$$H_0: b = 0$$

con ipotesi alternativa bilaterale

$$H_1: b \neq 0$$

può essere derivata dalla tabella riassuntiva (vedi tabulati nell'ultimo paragrafo, diversi dai calcoli manuali riportati nel capitolo precedente, per le approssimazioni),

	Devianza	Gdl	Varianza
<b>Totale</b>	<b>403,715</b>	<b>6</b>	
<b>Della regressione</b>	<b>323,208</b>	<b>1</b>	<b>323,208</b>
<b>Errore</b>	<b>80,506</b>	<b>5</b>	<b>16,101</b>

che fornisce tutti gli elementi utili al calcolo di **F**, ottenendo un valore che

$$F_{(1,5)} = \frac{323,208}{16,101} = 20,07$$

risulta uguale a **20,07** con df **1** e **5**.

Utilizzando gli stessi dati (come il precedente fornito dal tabulato del computer nell'ultimo paragrafo), il valore di **R<sup>2</sup>** (R-square) risulta uguale a **0,8006** e **R<sup>2</sup><sub>adj</sub>** (Adj R-sq) uguale a **0,7607**.

La significatività del test **F** per verificare l'ipotesi nulla

$$H_0: r = 0$$

con ipotesi alternativa

$$H_1: r \neq 0$$

mediante la formula

$$F_{1, n-2} = \frac{r^2 \cdot (n-2)}{1-r^2}$$

fornisce un **F** con df **1** e **5**

$$F_{1,5} = \frac{0,8006 \cdot 5}{1-0,8006} = \frac{4,003}{0,1994} = 20,07$$

uguale a **20,07** e identico al precedente.

Ma, nonostante il risultato identico, i due metodi sottendono scopi differenti e hanno condizioni di validità differenti; di conseguenza, **usare la significatività di r al posto di b è errato**.

Negli ultimi anni, il coefficiente di correlazione ha assunto un ruolo nettamente più limitato rispetto al passato, quando sovente era preferito alla regressione lineare semplice: la sua genericità, cioè il

non richiedere specificatamente una relazione di causa-effetto, veniva interpretata come maggiore possibilità di adattamento alla varietà delle condizioni ambientali. Più recentemente, si preferisce la regressione, in quanto dovrebbe indurre il ricercatore a ragionare con attenzione sui rapporti tra le due variabili, alla ricerca della relazione di causa effetto e alla sua direzione.

**I fattori principali che attualmente limitano l'uso della correlazione rispetto alla regressione lineare**, per cui anche i test di significatività non sono interscambiabili, sono almeno 5:

**1 - le differenze nelle condizioni di validità della correlazione rispetto alla regressione:** nella prima devono essere realizzate in entrambe le variabili  $X_1$  e  $X_2$ , mentre nella seconda solo per la variabile  $Y$ ;

**2 - il diverso significato di relazione tra le due variabili**, che nella correlazione è solo di covarianza lineare e non di causa - effetto;

**3 - la quantità di informazione contenute nelle analisi e nei test di significatività:** nella correlazione è più ridotto, rispetto all'informazione data da **a, b,  $r^2$**  della regressione;

**4 - la maggiore complessità della verifica di differenze da valori teorici che non siano nulli e dei confronti tra risultati differenti nella correlazione**, a causa della sua asimmetria nella distribuzione per valori distanti da zero;

**5 - l'assenza di significato ai fini predittivi** della correlazione.

Attualmente, la correlazione viene preferita alla regressione solo quando non si vuole dichiarare, in quanto priva di significato, una relazione di causa - effetto tra le due variabili considerate.

#### **13.4. INTERVALLO DI CONFIDENZA DI $r$**

Pure nel caso della correlazione, la stima dell'intervallo di confidenza di un parametro richiede che i campioni siano distribuiti in modo simmetrico, rispetto al valore vero o della popolazione. Ma, come già evidenziato nel paragrafo precedente, a differenza di quanto avviene per la media campionaria  $\bar{x}$  rispetto a  $\mu$  e per il coefficiente angolare  $\mathbf{b}$  rispetto a  $\mathbf{b}$ , i valori campionari  $\mathbf{r}$  sono distribuiti normalmente solo quando il valore di  $r$  è piccolo (teoricamente zero) e i campioni sono abbastanza grandi (teoricamente infiniti).

Quando il valore di  $r$  si allontana da zero, la distribuzione dei valori  $\mathbf{r}$  campionari è sempre asimmetrica. Di conseguenza, pure conoscendo il valore di  $\mathbf{r}$

e la sua **varianza**

$$s_r^2 = \frac{1 - r^2}{n - 2}$$

oppure il suo **errore standard**  $s_r$ ,

$$s_r = \sqrt{\frac{1-r^2}{n-2}}$$

non è corretto calcolare l'errore fiduciale di  $r$   
attraverso

$$\mathbf{r} = r \pm t_{\alpha/2, n-2} \cdot s_r$$

applicando alla correlazione  
quanto è possibile per la retta

$$\mathbf{b} = b \pm t_{\alpha/2, n-2} \cdot s_b$$

A causa dell'asimmetria dei valori campionari attorno alla media,

- **i limiti di confidenza**, che utilizzano il valore di  $t$  e l'errore standard, **non sono adatti a verificare la significatività di un qualsiasi valore di correlazione rispetto ad un valore teorico atteso**, a differenza di quanto avviene per la media di un campione o il coefficiente angolare di una retta di regressione;
- **non è possibile nemmeno utilizzare il test  $t$  per il confronto tra due indici di regressione  $r_1$  e  $r_2$** .

L'uso di questo test inferenziale è limitato al solo caso in cui si vuole verificare l'assenza di correlazione, espressa dall'ipotesi nulla  $H_0: r = 0$ .

Per l'uso generale, con qualsiasi valore di  $r$ , dell'intervallo di confidenza, sono stati proposti vari metodi. Ne possono essere citati quattro:

- a) **la trasformazione di  $r$  in  $z$**  proposta da **Fisher**, valida soprattutto per grandi campioni,
- b) il precedente **metodo di Fisher**, ma **con l'uso della distribuzione  $t$  di Student al posto della distribuzione normale**, più cautelativa per piccoli campioni,
- c) **la procedura proposta da M. V. Muddapur** nel 1988, che utilizza la distribuzione **F** (vedi: *A simple test for correlation coefficient in a bivariate normal population*. **Sankhy: Indian J. Statist. Ser. B.** 50: 60-68),
- d) **la procedura proposta da S Jeyaratnam** nel 1992, analoga alla precedente, ma con l'uso della distribuzione **t** (vedi: *Confidence intervals for the correlation coefficient*. **Statist. Prob. Lett.** 15: 389-393).

Gli ultimi due, citati anche da J. H. Zar nel suo test del 1999 (*Biostatistical Analysis*, fourth edition, Prentice Hall, New Jersey, a pagg. 383-384), offrono il vantaggio di stimare **un intervallo generalmente minore** di quello di Fisher, oltre all'aspetto pratico di **non richiedere trasformazioni di  $r$**  e quindi di essere **più rapidi**.

Questi test, come qualsiasi intervallo fiduciale, possono essere utilizzati anche per la verifica dell'ipotesi sulla differenza tra due medie, in un test bilaterale.

A) Il **metodo di Fisher** stima il limite inferiore  $L_1$  e il limite superiore  $L_2$  dell'intervallo di confidenza attraverso le relazioni

$$L_1 = z - Z_{\alpha/2} \cdot s_z$$

$$L_2 = z + Z_{\alpha/2} \cdot s_z$$

dove

- $z$  è il valore campionario di  $r$  trasformato, attraverso la relazione

$$z = 0,5 \cdot \ln\left(\frac{1+r}{1-r}\right)$$

- $Z_{\alpha/2}$  è il valore della  $Z$  nella distribuzione normale alla probabilità  $\alpha/2$ , prescelta per definire l'intervallo,
- $s_z$  è l'**errore standard** (approssimato)

di  $r$  trasformato in  $z$

$$s_z = \sqrt{\frac{1}{n-3}}$$

che dipende da  $n$ .

Successivamente, i due valori stimati  $L_1$  e  $L_2$ , calcolati ovviamente in una scala  $z$ , devono essere riportati sulla scala di  $r$ , con la trasformazione

$$r = \frac{e^{2z} - 1}{e^{2z} + 1}$$

In questo ultimo passaggio, si perde la simmetria di  $L_1$  e  $L_2$  rispetto al valore centrale, quando  $r \neq 0$ . L'asimmetria intorno ad  $r$  risulta tanto più marcata quanto più esso si avvicina a +1 oppure a -1.

B) Più recentemente, in vari testi è proposta la misura più cautelativa, che fornisce un intervallo maggiore in rapporto alla dimensione campionaria  $n$ , con la distribuzione **t di Student**.

Il limite inferiore  $L_1$  e il limite superiore  $L_2$  dell'intervallo di confidenza sono calcolati attraverso le relazioni

$$\mathbf{L}_1 = z - t_{\alpha/2, n} \cdot \mathbf{S}_z$$

$$\mathbf{L}_2 = z + t_{\alpha/2, n} \cdot \mathbf{S}_z$$

dove

-  $t_{\alpha/2, n}$  è il valore alla probabilità  $\alpha/2$  prescelta per definire l'intervallo, con  $n = n - 2$ ,

mentre tutti gli altri parametri restano identici a quelli appena presentati.

Nulla cambia rispetto al metodo classico di Fisher, per quanto riguarda

- dapprima la trasformazione di  $\mathbf{r}$  in  $\mathbf{z}$
- successivamente le trasformazioni dei valori  $\mathbf{L}_1$  e  $\mathbf{L}_2$  in scala  $\mathbf{r}$ .

C) Il metodo proposto da **Muddapur**, meno noto del metodo classico di Fisher, più recente e più raro in letteratura, ma più rapido in quanto **non richiede alcuna trasformazione**,

con  $\mathbf{L}_1$

$$\mathbf{L}_1 = \frac{[(1+F) \cdot r] + (1-F)}{(1+F) + [(1-F) \cdot r]}$$

e  $\mathbf{L}_2$

$$\mathbf{L}_1 = \frac{[(1+F) \cdot r] - (1-F)}{(1+F) - [(1-F) \cdot r]}$$

dove

-  $\mathbf{F}$  è il valore corrispondente nella distribuzione  $\mathbf{F}$  di Fisher alla probabilità  $\alpha$  per un test bilaterale e con df  $n_1 = n_2 = n - 2$

fornisce una stima uguale a quella classica di Fisher.

D) Il metodo proposto da **Jeyaratnam**, che può essere letto come una variante del precedente, una sua formula abbreviata in quanto ancor più rapido

con  $\mathbf{L}_1$

$$\mathbf{L}_1 = \frac{r - \sqrt{\frac{t_{\alpha, n}^2}{t_{\alpha, n}^2 + n}}}{1 - r \sqrt{\frac{t_{\alpha, n}^2}{t_{\alpha, n}^2 + n}}}$$

e  $\mathbf{L}_2$

$$L_2 = \frac{r + \sqrt{\frac{t_{a,n}^2}{t_{a,n}^2 + n}}}{1 + r \sqrt{\frac{t_{a,n}^2}{t_{a,n}^2 + n}}}$$

dove

- $t$  è il valore corrispondente nella distribuzione  $t$  di Student alla probabilità  $a$  per un test bilaterale (come in tutti gli intervalli fiduciali) e con  $df$   $n = n-2$
- $n = n-2$ .

#### ESEMPIO

Con 30 coppie di dati, è stato calcolato il coefficiente di correlazione lineare semplice  $r = 0,71$ .

Entro quale intervallo si colloca il valore reale  $\rho$  della popolazione ( $r$ ), alla probabilità  $a = 0,05$ ?

Calcolare i valori estremi  $L_1$  e  $L_2$  dell'intervallo fiduciale con i 4 diversi metodi.

Risposta

A) Con il metodo classico di **Fisher**

1 – dopo aver trasformato  $r = 0,71$  in  $z$

con

$$z = 0,5 \cdot \ln\left(\frac{1+0,71}{1-0,71}\right) = 0,5 \cdot \ln 5,8965 = 0,5 \cdot 1,7744 = 0,8872$$

ottenendo  $z = 0,8872$

2 – si stima il suo errore standard  $s_z$ , ovviamente su scala  $z$ , con la formula approssimata

$$s_z = \sqrt{\frac{1}{30-3}} = \sqrt{0,037} = 0,1924$$

ottenendo  $s_z = 0,1924$

3 – Successivamente, con  $Z_{a/2} = 1,96$  (valore della distribuzione normale standardizzata alla probabilità  $a = 0,05$  bilaterale) si stimano i due limiti dell'intervallo  $L_1$  e  $L_2$ :

con

$$L_1 = 0,8872 - 1,96 \cdot 0,1924 = 0,8872 - 0,3771 = 0,5101$$

si ottiene  $L_1 = 0,5101$

e con

$$L_2 = 0,8872 + 1,96 \cdot 0,1924 = 0,8872 + 0,3771 = \mathbf{1,2643}$$

si ottiene  $L_1 = \mathbf{1,2643}$

4 – Per il confronto di  $L_1$  e  $L_2$  con  $r = \mathbf{0,71}$ , è necessario ritrasformare i due valori  $z$  ottenuti nei corrispondenti valori in scala  $r$ . Ricordando che  $e = \mathbf{2,718}$

per  $z = \mathbf{0,5101}$  con

$$r = \frac{2,718^{2 \cdot 0,5101} - 1}{2,718^{2 \cdot 0,5101} + 1} = \frac{2,718^{1,0202} - 1}{2,718^{1,0202} + 1} = \frac{2,7735 - 1}{2,7735 + 1} = \frac{1,7735}{3,7735} = \mathbf{0,470}$$

si ottiene  $L_1 = \mathbf{0,470}$

e per  $z = \mathbf{1,2643}$  con

$$r = \frac{2,718^{2 \cdot 1,2643} - 1}{2,718^{2 \cdot 1,2643} + 1} = \frac{2,718^{2,5286} - 1}{2,718^{2,5286} + 1} = \frac{12,5327 - 1}{12,5327 + 1} = \frac{11,5327}{13,5327} = \mathbf{0,852}$$

si ottiene  $L_2 = \mathbf{0,852}$

Con il metodo classico di Fisher, l'intervallo di confidenza di  $r = \mathbf{0,71}$  calcolato su  $\mathbf{30}$  copie di dati è compreso tra i limiti  $\mathbf{0,470}$  e  $\mathbf{0,852}$  con probabilità  $\alpha = \mathbf{0,05}$ .

B) Utilizzando sempre il metodo di **Fisher**, ma con la **distribuzione t** al posto della distribuzione  $z$ ,

1 – i primi due passaggi sono identici ai punti 1 e 2 precedenti, nei quali si era ottenuto

$$z = \mathbf{0,8872} \quad \text{e} \quad s_z = \mathbf{0,1925}$$

2 – Dopo aver scelto il valore di  $t$  che, con  $\alpha = \mathbf{0,05}$  bilaterale e  $df \ n = \mathbf{28}$  è

$$t_{0,025, 28} = \mathbf{2,048}$$

3 – si ottiene (sempre in scala  $z$ )

$$L_1 = 0,8872 - 2,048 \cdot 0,1925 = 0,8872 - 0,3942 = \mathbf{0,493}$$

un valore di  $L_1 = \mathbf{0,493}$

e

$$L_2 = 0,8872 + 2,048 \cdot 0,1925 = 0,8872 + 0,3942 = \mathbf{1,2814}$$

un valore di  $L_2 = \mathbf{1,2814}$

4 – Infine, come nella fase 4 precedente, si riportano questi due valori **z** in scala **r**:

per **z = 0,493** con

$$r = \frac{2,718^{2 \cdot 0,493} - 1}{2,718^{2 \cdot 0,493} + 1} = \frac{2,718^{0,986} - 1}{2,718^{0,986} + 1} = \frac{2,6802 - 1}{2,6802 + 1} = \frac{1,6802}{3,6802} = \mathbf{0,457}$$

si ottiene **L<sub>1</sub> = 0,457**

e per **z = 1,2814** con

$$r = \frac{2,718^{2 \cdot 1,2814} - 1}{2,718^{2 \cdot 1,2814} + 1} = \frac{2,718^{2,5628} - 1}{2,718^{2,5628} + 1} = \frac{12,9686 - 1}{12,9686 + 1} = \frac{11,9686}{13,9686} = \mathbf{0,857}$$

si ottiene **L<sub>2</sub> = 0,857**.

Con il metodo di Fisher nel quale sia utilizzata la distribuzione **t** con df **n-2**, l'intervallo di confidenza di **r = 0,71** calcolato su **30** copie di dati è compreso tra i limiti **0,457** e **0,857** con probabilità **a = 0.05**.

C) Con il metodo proposto da **Muddapur**,

1 - dapprima si trova il valore di **F** alla probabilità **a = 0.05** bilaterale con df **n<sub>1</sub> = n<sub>2</sub> = 28**; rilevato in una tabella molto più dettagliata di quella riportata nelle dispense esso risulta uguale a **2,13**;

2 – successivamente si calcolano

**L<sub>1</sub>** con

$$L_1 = \frac{(1 + 2,13) \cdot 0,71 + (1 - 2,13)}{(1 + 2,13) + (1 - 2,13) \cdot 0,71} = \frac{2,2223 - 1,13}{3,13 - 0,8023} = \frac{1,0923}{2,3277} = \mathbf{0,469}$$

ottenendo **L<sub>1</sub> = 0,469**

e **L<sub>2</sub>** con

$$L_1 = \frac{(1 + 2,13) \cdot 0,71 - (1 - 2,13)}{(1 + 2,13) - (1 - 2,13) \cdot 0,71} = \frac{2,2223 + 1,13}{3,13 + 0,8023} = \frac{3,3523}{2,9323} = \mathbf{0,853}$$

ottenendo **L<sub>1</sub> = 0,853**

D) Con il metodo proposto da **Jeyaratnam**,

1 - dapprima si trova il valore di **t** alla probabilità **a = 0.05** bilaterale con df **n = 28**; esso risulta uguale a **2,13**;

2 – successivamente si calcolano

$L_1$  con

$$L_1 = \frac{0,71 - \sqrt{\frac{2,048^2}{2,048^2 + 28}}}{1 - 0,71 \cdot \sqrt{\frac{2,048^2}{2,048^2 + 28}}} = \frac{0,71 - \sqrt{\frac{4,1943}{32,1943}}}{1 - 0,71 \cdot \sqrt{\frac{4,1943}{32,1943}}} =$$
$$L_1 = \frac{0,71 - \sqrt{0,1303}}{1 - 0,71 \cdot \sqrt{0,1303}} = \frac{0,71 - 0,361}{1 - 0,71 \cdot 0,361} = \frac{0,349}{0,744} = 0,469$$

ottenendo  $L_1 = 0,469$

e  $L_2$  con

$$L_2 = \frac{0,71 + \sqrt{\frac{2,048^2}{2,048^2 + 28}}}{1 + 0,71 \cdot \sqrt{\frac{2,048^2}{2,048^2 + 28}}} = \frac{0,71 + \sqrt{\frac{4,1943}{32,1943}}}{1 + 0,71 \cdot \sqrt{\frac{4,1943}{32,1943}}} =$$
$$L_2 = \frac{0,71 + \sqrt{0,1303}}{1 + 0,71 \cdot \sqrt{0,1303}} = \frac{0,71 + 0,361}{1 + 0,71 \cdot 0,361} = \frac{1,071}{1,256} = 0,853$$

ottenendo  $L_2 = 0,853$

Questo calcolo diventa molto più rapido se dapprima, separatamente, si stima la parte sotto radice,

$$\sqrt{\frac{2,048^2}{2,048^2 + 28}} = \sqrt{\frac{4,1943}{32,1943}} = \sqrt{0,1303} = 0,361$$

che risulta uguale a **0,361**:

$$L_1 = \frac{0,71 - 0,361}{1 - 0,71 \cdot 0,361} = \frac{0,349}{0,744} = 0,469$$

$$L_2 = \frac{0,71 + 0,361}{1 + 0,71 \cdot 0,361} = \frac{1,071}{1,256} = 0,853$$

I risultati dei 4 metodi, con i dati dell'esempio, sono riportati nella tabella sottostante:

METODO	L <sub>1</sub>	L <sub>2</sub>
Fisher	0,470	0,852
Fisher, con distribuzione t	0,457	0,857
Muddapur	0,469	0,853
Jeyaratnam	0,469	0,853
r = 0,71    n = 30    α = 0.05		

E' sufficiente il semplice confronto, per verificare la loro corrispondenza. I calcoli sono stati fatti alla quarta cifra decimale per evitare arrotondamenti e meglio porre a confronto i risultati.

### 13.5. POTENZA A PRIORI E A POSTERIORI PER LA SIGNIFICATIVITA' DI r

Le ipotesi nulle  $H_0: r = 0$  e  $H_0: r = r_0$  sono rifiutate correttamente, senza commettere l'errore b di II Tipo, quando nella realtà  $r \neq 0$  nel primo caso e  $r \neq r_0$  nel secondo. Ma, in tali condizioni, la distribuzione è asimmetrica. Pertanto, per il calcolo della potenza  $1-b$  in un test di significatività del coefficiente di correlazione  $r$ , è necessario ricorrere alla **trasformazione di r in z di Fisher**.

I metodi di stima di  $1-b$  presentano alcune differenze se

- a) l'ipotesi nulla è  $H_0: r = 0$
- b) l'ipotesi nulla è  $H_0: r = r_0$

A) Nel primo caso, nella verifica dell'ipotesi  $H_0: r = 0$ , la stima della **potenza a posteriori  $1-b$**  è ottenuta dalla relazione:

$$Z_b = (z_{a,n} - z) \cdot \sqrt{n-3}$$

dove

- $Z_b$  è il valore che, nella **distribuzione normale unilaterale**, permette di ricavare direttamente (senza trasformazione) la probabilità  $b$ ;
- $z_{a,n}$  è il **valore critico di r**, da prendere in una delle due tabelle dei valori critici di  $r$ , per la probabilità  $a$  prefissata in una distribuzione **bilaterale o unilaterale** e con  $df \ n = n - 2$ , **trasformato** in  $z$  con la formula di Fisher;
- $z$  è il valore di **r sperimentale**, trasformato in  $z$  con la formula di Fisher;
- $n$  è il numero di coppie di dati, con i quali è stato calcolato  $r$ ,

- ricordando che sia il valore critico di  $r$  sia il valore sperimentale di  $r$  devono essere trasformati con la solita formula

$$z = 0,5 \cdot \ln\left(\frac{1+r}{1-r}\right)$$

Sempre per rifiutare l'ipotesi nulla  $H_0: r = 0$  quando il valore atteso è  $r_0 \neq 0$ ,

la potenza a priori o numero minimo  $n$  di dati perché il valore campionario  $r$  sia significativo è

$$n = \left(\frac{Z_b + Z_a}{z_0}\right)^2 + 3$$

dove

- $Z_b$  e  $Z_a$  sono i valori di  $z$  nella distribuzione normale per le probabilità  $a$  e  $b$  (senza trasformazione di Fisher),
- $z_0$  (zeta greca minuscola) è il valore  $r_0$  atteso, trasformato in  $z$  secondo Fisher.

**ESEMPIO 1** Con **20** dati è stato calcolato un valore di  $r = 0,35$  che non è risultato significativamente diverso da zero, alla probabilità  $a = 0,05$ . Infatti, nelle due tabelle relative è semplice osservare che il valore critico alla stessa probabilità  $a = 0,05$  in un test bilaterale con **df = 18** è **0,4438**.

Si chiede:

- quale è la potenza di questo test?
- quante coppie di dati ( $n$ ) occorre raccogliere, per rifiutare l'ipotesi nulla  $H_0: r = 0$  nel 90% dei casi, al livello di significatività del 5%?

Risposte

- Applicando la formula

$$Z_b = (z - z_{a,n}) \cdot \sqrt{n-3}$$

dove

- $Z_b$  è il valore nella **distribuzione normale unilaterale** che permette di ricavare direttamente la probabilità  $b$ ;
- $z$  è il valore sperimentale di  $r$  (uguale a **0,35**), che trasformato in  $z$  con il metodo di Fisher

$$z = 0,5 \cdot \ln\left(\frac{1+0,35}{1-0,35}\right) = 0,5 \cdot \ln\left(\frac{1,35}{0,65}\right) = 0,5 \cdot \ln 2,077 = 0,5 \cdot 0,73 = 0,365$$

risulta uguale a **0,365**

- $z_{a,n}$  è il **valore critico** di **r**, alla probabilità **a = 0.05 bilaterale** con **n = n - 2 = 18** (nella tabella relativa è **0,4438**); trasformato in **z** con il metodo di Fisher

$$z_{0,05,18} = 0,5 \cdot \ln\left(\frac{1+0,4438}{1-0,4438}\right) = 0,5 \cdot \ln\left(\frac{1,4438}{0,5562}\right) = 0,5 \cdot \ln 2,593 = 0,5 \cdot 0,953 = 0,477$$

risulta uguale a **0,477**.

Con i dati dell'esempio,

$$Z_b = (0,365 - 0,477) \cdot \sqrt{20 - 3} = -0,112 \cdot 4,123 = -0,46$$

$Z_b$  risulta uguale a **0,46**.

Nella distribuzione normale **unilaterale**, ad essa corrisponde una probabilità **b = 0,323**. Il segno indica solo la coda della distribuzione.

Di conseguenza, la **potenza 1-b** del test è  $1 - 0,323 = 0,677$  o **67,7%**

- b) Per stimare quanti dati sono necessari per un test con **b = 0.10** e **a = 0.05 bilaterale**, affinché un valore atteso di **r = 0,35** risulti significativamente **diverso da zero**, usando la formula

$$n = \left( \frac{Z_b + Z_a}{z_0} \right)^2 + 3$$

dove con

- $Z_b = 1,28$  (in una distribuzione normale **unilaterale** per una probabilità uguale a **0.10**),
- $Z_a = 1,96$  (in una distribuzione normale **bilaterale** per una probabilità uguale a **0.05**),
- $z_0 = 0,365$  ottenuto dalla trasformazione di **r = 0,35**

mediante

$$z = 0,5 \cdot \ln\left(\frac{1+0,35}{1-0,35}\right) = 0,5 \cdot \ln\left(\frac{1,35}{0,65}\right) = 0,5 \cdot \ln 2,077 = 0,5 \cdot 0,73 = 0,365$$

si ottiene

$$n = \left( \frac{1,28 + 1,96}{0,365} \right)^2 + 3 = 8,877^2 + 3 = 78,8 + 3 = 81,8$$

un valore di  $n = 81,8$ . E' necessario rilevare almeno **82** coppie di dati.

B) Nel secondo caso, quando l'ipotesi nulla è  $H_0: r = r_0$  dove  $r_0 \neq 0$ ,

la formula per calcolare **la potenza a posteriori 1-b** è leggermente più complessa di quella precedente, diventando

$$Z_b = (|z - z_0| - z_{\alpha n}) \cdot \sqrt{n - 3}$$

dove, mantenendo uguali gli altri parametri,

- in  $|z - z_0|$
  - $z$  è il valore campionario di  $r$  trasformato in  $z$  con la formula di Fisher
  - $z_0$  (zeta greca minuscola) è la stessa trasformazione del valore  $r_0$  atteso o ipotizzato,
- ricordando che i tre valori entro parentesi, cioè (1) il valore sperimentale di  $r$ , (2) il valore atteso o teorico di confronto  $r_0$ , (3) il valore critico di  $r$  con **df n-2** devono essere trasformati con la formula di Fisher.

**ESEMPIO 2** La correlazione tra le variabili  $X_1$  e  $X_2$  è stata valutata in  $r = 0,15$ . Con **100** dati, un ricercatore ha calcolato  $r = 0,35$  e ha motivo di credere che, nella nuova condizione sperimentale, la correlazione sia significativamente maggiore.

Calcolare la potenza del test, per una significatività  $\alpha = 0.05$ .

Risposta

1) Con  $r = 0,35$  il valore di  $z$

$$z = 0,5 \cdot \ln \left( \frac{1 + 0,35}{1 - 0,35} \right) = 0,5 \cdot \ln \left( \frac{1,35}{0,65} \right) = 0,5 \cdot \ln 2,077 = 0,5 \cdot 0,73 = 0,365$$

è uguale a **0,365**.

2) Con  $r_0 = 0,15$  il valore di  $z_0$

$$z = 0,5 \cdot \ln\left(\frac{1+0,15}{1-0,15}\right) = 0,5 \cdot \ln\left(\frac{1,15}{0,85}\right) = 0,5 \cdot \ln 1,353 = 0,5 \cdot 0,302 = 0,151$$

è uguale a **0,151**,

3) Il valore critico di **r** alla probabilità **a = 0.5** in una distribuzione unilaterale con **df = 98** non è riportato nelle 2 tabelle specifiche. Per interpolazione tra **0,168 con df 95** e **0,164 con df 100** può essere stimato uguale a **0,165**. Da esso si ricava

il valore di  $z_{0,05,98}$

$$z_{0,05,98} = 0,5 \cdot \ln\left(\frac{1+0,165}{1-0,165}\right) = 0,5 \cdot \ln\left(\frac{1,165}{0,835}\right) = 0,5 \cdot \ln 1,395 = 0,5 \cdot 0,333 = 0,166$$

che risulta uguale a **0,166**.

4) Da questi tre valori e con **n = 100**, si ottiene

$$Z_b = ((0,365 - 0,151) - 0,166) \cdot \sqrt{100 - 3} = (0,214 - 0,166) \cdot \sqrt{97} = 0,048 \cdot 9,849 = 0,47$$

un valore di  $Z_b = 0,47$

In una distribuzione normale unilaterale, a **z = 0,47** corrisponde una probabilità di **0,316**.

Con questi dati, il valore di **b** è uguale a **0,316** e la potenza **1-b** del test richiesto è **1-0,316 = 0,684**.

### **13.6. DIFFERENZA TRA DUE COEFFICIENTI DI CORRELAZIONE IN CAMPIONI INDIPENDENTI E CALCOLO DEL COEFFICIENTE COMUNE**

Il confronto tra due coefficienti di correlazione **r<sub>1</sub>** e **r<sub>2</sub>**, calcolati su **due campioni indipendenti**, per verificare

l'ipotesi nulla

$$\mathbf{H_0: r_1 = r_2}$$

con ipotesi alternativa sia bilaterale

$$\mathbf{H_1: r_1 \neq r_2}$$

che unilaterale

$$\mathbf{H_1: r_1 < r_2} \quad \text{oppure} \quad \mathbf{H_1: r_1 > r_2}$$

pone sempre il problema della forte asimmetria dei valori campionari alla quale si aggiunge quella della non omogeneità delle due varianze.

La **simmetria** e la **omoschedasticità** sono ricostruite mediante la trasformazione di Fisher, per cui il test diventa

$$Z = \frac{z_1 - z_2}{s_{z_1 - z_2}}$$

dove

- Z è il valore della distribuzione normale, unilaterale o bilaterale in rapporto all'ipotesi alternativa,
- $z_1$  e  $z_2$  sono  $r_1$  e  $r_2$  trasformati con la formula di Fisher,
- $s_{z_1 - z_2}$  è l'**errore standard della differenza** precedente, ottenuta con

$$s_{z_1 - z_2} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

come **formula generale** e

con la **formula semplificata**

$$s_{z_1 - z_2} = \sqrt{\frac{2}{n_1 - 3}}$$

quando i due campioni sono bilanciati ( $n_1 = n_2$ )

Con poche decine di dati, all'uso della distribuzione normale alcuni preferiscono l'uso della **tabella t** con **df = N-4**, in quanto più cautelativa. Permane il problema che solitamente i testi per la distribuzione t riportano solo le tavole sinottiche; quindi risulta impossibile stimare la probabilità in modo più preciso.

Nel caso di **due campioni dipendenti**, come possono essere i coefficienti di correlazione del padre e quello della madre con una caratteristica di un figlio, calcolate su varie coppie di genitori, la procedura è differente.

Se i due coefficienti di correlazione  $r_1$  e  $r_2$  non risultano significativamente differenti, anche sulla base della potenza del test e delle dimensioni del campione si può concludere che  $r_1 = r_2$ . In tali condizioni, spesso è utile calcolare un **coefficiente di correlazione comune o pesato** (*common or weighted correlation coefficient*); è la misura più corretta della correlazione tra le variabili  $X_1$  e  $X_2$ , in quanto media ponderata dei risultati dei due esperimenti.

Sempre per i problemi di simmetria e omoschedasticità, per ottenere il coefficiente di regressione comune  $r_w$

- dopo la trasformazione di  $r_1$  e  $r_2$  rispettivamente in  $z_1$  e  $z_2$
- si calcola il valore medio  $z_w$

con

$$z_w = \frac{(n_1 - 3) \cdot z_1 + (n_2 - 3) \cdot z_2}{(n_1 - 3) + (n_2 - 3)}$$

che, nel caso di  $n_1 = n_2$ , può essere semplificata in

$$z_w = \frac{z_1 + z_2}{2}$$

- Infine si trasforma  $z_w$  in  $r_w$

con

$$r = \frac{e^{2z} - 1}{e^{2z} + 1}$$

Recentemente, sono stati proposti altri metodi, che dovrebbero dare stime più precise di  $r_w$ .

#### ESEMPIO

a) Calcolare la significatività della differenza tra i due coefficienti di correlazione

- $r_1 = 0,22$  con  $n_1 = 30$  e
- $r_2 = 0,31$  con  $n_2 = 50$

b) Calcolare inoltre il coefficiente di correlazione ponderato  $r_w$ .

Risposte

A) Dopo aver trasformato  $r_1 = 0,22$

$$z_1 = 0,5 \cdot \ln\left(\frac{1+0,22}{1-0,22}\right) = 0,5 \cdot \ln\left(\frac{1,22}{0,78}\right) = 0,5 \cdot \ln 1,564 = 0,5 \cdot 0,447 = 0,224$$

in  $z_1 = 0,224$

e  $r_2 = 0,31$

$$z_1 = 0,5 \cdot \ln\left(\frac{1+0,31}{1-0,31}\right) = 0,5 \cdot \ln\left(\frac{1,31}{0,69}\right) = 0,5 \cdot \ln 1,899 = 0,5 \cdot 0,641 = 0,321$$

in  $z_2 = 0,321$

- si calcola l'errore standard della differenza  $z_1 - z_2$

$$s_{z_1 - z_2} = \sqrt{\frac{1}{30-3} + \frac{1}{50-3}} = \sqrt{0,037 + 0,021} = \sqrt{0,058} = 0,24$$

ottenendo  $s_{z_1 - z_2} = 0,24$ .

Da essi si ricava

$$Z = \frac{0,224 - 0,321}{0,24} = \frac{-0,097}{0,24} = -0,40$$

un valore di  $Z = -0,40$ .

Nella **distribuzione normale bilaterale**, poiché la domanda verte solo sulla significatività della differenza tra i due coefficienti di correlazione  $r_1$  e  $r_2$ , a  $Z = -0,40$  corrisponde una probabilità  $\alpha = 0,689$ . Pure senza un'analisi della potenza del test, illustrata nel paragrafo successivo, la probabilità è così alta che si può concludere non esiste alcuna differenza tra i due coefficienti angolari.

Nel caso di un test unilaterale, rispetto alla procedura illustrata l'unica differenza consiste nella stima della probabilità  $\alpha$ , non diversamente dal test t su due medie o due coefficienti angolari.

B) Per ottenere il coefficiente ponderato  $r_w$ , con

-  $z_1 = 0,224$  e  $n_1 = 30$

-  $z_2 = 0,321$  e  $n_2 = 50$

- dapprima si stima  $z_w$

$$z_w = \frac{(30-3) \cdot 0,224 + (50-3) \cdot 0,321}{(30-3) + (50-3)} = \frac{6,048 + 15,087}{74} = \frac{21,135}{74} = 0,286$$

che risulta uguale a **0,286**

- infine con la trasformazione

$$r_w = \frac{2,718^{20,286} - 1}{2,718^{20,286} + 1} = \frac{2,718^{0,572} - 1}{2,718^{0,572} + 1} = \frac{1,77 - 1}{1,77 + 1} = \frac{0,77}{2,77} = 0,278$$

si ricava  $r_w = 0,278$ .

Il coefficiente di correlazione comune o ponderato tra

-  $r_1 = 0,22$  con  $n_1 = 30$  e

-  $r_2 = 0,31$  con  $n_2 = 50$

è  $r_w = 0,278$ .

### 13.7. POTENZA A PRIORI E A POSTERIORI DEL TEST PER LA SIGNIFICATIVITA' DELLA DIFFERENZA TRA DUE COEFFICIENTI DI CORRELAZIONE

La potenza a posteriori 1-b del test di significatività tra due coefficienti di correlazione in campioni indipendenti è valutata mediante

$$Z_b = \frac{|z_1 - z_2|}{s_{z_1 - z_2}} - Z_a$$

- dove, con la solita simbologia,

-  $z_1$  e  $z_2$  sono  $r_1$  e  $r_2$  trasformati con la formula di Fisher,

-  $s_{z_1 - z_2}$  è l'errore standard della differenza precedente, ottenuta

con

$$s_{z_1 - z_2} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

-  $Z_b$  è il valore di  $Z$  che permette di ricavare la probabilità  $b$  in una **distribuzione normale unilaterale**,

-  $Z_a$  è il valore ricavato dalle tabella sulla base della probabilità  $a$  prefissata nella stessa distribuzione normale, ma **unilaterale o bilaterale in rapporto alla direzione dell'ipotesi  $H_1$** .

La potenza a priori o numero minimo di coppie di osservazioni (n) può essere valutata

con

$$n = 2 \left( \frac{Z_a + Z_b}{z_1 - z_2} \right)^2 + 3$$

utilizzando la simbologia consueta.

Come nel test  $t$ , la quantità  $n$  indica il numero minimo di dati per ognuno dei due coefficienti di correlazione a confronto, affinché l'ipotesi nulla  $r_1 = r_2$  possa essere rifiutata alla probabilità  $\alpha$  prefissata e con la potenza  $1-\beta$  desiderata.

Il bilanciamento di due campioni indipendenti permette di raggiungere la potenza massima del test, utilizzando il numero minimo di osservazioni per ogni gruppo. Ma nella pratica sperimentale, non sempre le rilevazioni nei due gruppi hanno lo stesso costo morale od economico: somministrare sostanze tossiche aumenta la mortalità delle cavie rispetto al placebo; le analisi del controllo e del trattato possono richiedere procedure differenti che esigono tempi di durata differente; tra due aree a confronto nelle quali prelevare i campioni, una può essere sul posto e l'altra molto distante. In vari settori della ricerca applicata, per valutare le trasformazioni intervenute nel periodo, è prassi richiedere il confronto di dati ancora da raccogliere rispetto a dati storici, il cui campione ovviamente non può essere ampliato.

Spesso, può essere utile diminuire al minimo le osservazioni di un gruppo oppure utilizzare il piccolo campione già raccolto, aumentando ovviamente quelle dell'altro gruppo, affinché il test non perda la potenza desiderata.

La media armonica permette di stimare

- quante osservazioni deve avere il campione **2** ( $n_2$ ),
- una volta che sia stato stimato il numero minimo di dati ( $n$ ) e
- prefissato il numero di dati che si intende raccogliere o già raccolti per il campione **1** ( $n_1$ ) mediante la relazione

$$n_2 = \frac{n_1(n+3) - 6n}{2n_1 - n - 3}$$

#### ESEMPIO 1

Il precedente test per la significatività della differenza tra i due coefficienti di correlazione

- $r_1 = 0,22$  con  $n_1 = 30$
- $r_2 = 0,31$  con  $n_2 = 50$

non ha permesso di rifiutare l'ipotesi nulla.

Si chiede

- a) Quale era la potenza di questo test per una significatività con  $\alpha = 0.05$ ?
- b) Quanti dati per gruppo sarebbero necessari affinché nel **80%** dei casi il test risulti significativo alla probabilità  $\alpha = 0.05$

Risposte

A) Dapprima si trasformano  $r_1 = 0,22$

$$z_1 = 0,5 \cdot \ln\left(\frac{1+0,22}{1-0,22}\right) = 0,5 \cdot \ln\left(\frac{1,22}{0,78}\right) = 0,5 \cdot \ln 1,564 = 0,5 \cdot 0,447 = 0,224$$

in  $z_1 = 0,224$

e  $r_2 = 0,31$

$$z_1 = 0,5 \cdot \ln\left(\frac{1+0,31}{1-0,31}\right) = 0,5 \cdot \ln\left(\frac{1,31}{0,69}\right) = 0,5 \cdot \ln 1,899 = 0,5 \cdot 0,641 = 0,321$$

in  $z_2 = 0,321$

Successivamente si calcola l'errore standard della differenza  $z_1 - z_2$

$$s_{z_1 - z_2} = \sqrt{\frac{1}{30-3} + \frac{1}{50-3}} = \sqrt{0,037 + 0,021} = \sqrt{0,058} = 0,24$$

ottenendo  $s_{z_1 - z_2} = 0,24$

e nella tabella della distribuzione normale bilaterale per  $\alpha = 0,05$  si rileva  $Z = 1,96$ .

Infine con

$$Z_b = \frac{|0,224 - 0,321|}{0,24} - 1,96 = 0,40 - 1,96 = -1,56$$

si ottiene  $Z_b = -1,56$

Ad un valore di  $Z = 1,56$  in una distribuzione normale unilaterale corrisponde una probabilità uguale 0,059; ma il valore negativo indica che essa si trova a sinistra della distribuzione, molto distante dal valore dell'ipotesi nulla; di conseguenza, la potenza del test è particolarmente bassa, pari appunto al 5,9%.

B) Una conferma di questa **potenza a posteriori** molto bassa può venire anche dalla stima della **potenza a priori** o numero minimo di dati necessari affinché il test sulla differenza tra i due coefficienti di correlazione risulti significativo. Con una potenza così bassa, ovviamente il numero ( $n$ ) richiesto risulterà molto alto.

Poiché,

- per una probabilità  $\alpha = 0,05$  bilaterale, il valore di  $Z_\alpha$  è uguale a **1,96**

- mentre, per una probabilità  $b = 0.20$  unilaterale, il valore di  $Z_b$  è uguale a **0,84** e con  $z_1 = 0,224$  e  $z_2 = 0,321$  il numero minimo di dati per ogni per ogni gruppo

$$n = 2 \left( \frac{1,96 + 0,84}{0,224 - 0,321} \right)^2 + 3 = \left( \frac{2,8}{0,097} \right)^2 + 3 = 833,2 + 3 = 837,2$$

è  $n = 838$  (sempre arrotondato all'unità superiore).

Il numero stimato è molto maggiore di quello dei dati raccolti nei due campioni (30 e 50). In vari settori della ricerca applicata, nei quali ogni misura campionaria ha costi non trascurabili, un numero così elevato indica l'impossibilità pratica di dimostrare la significatività della differenza tra i due coefficienti di correlazione. Pertanto, si può ritenere che tra essi non esista una differenza significativa, come d'altronde indicava il valore di probabilità  $a$  così alto.

#### ESEMPIO 2

Un ricercatore dispone di un campione di **40** osservazioni, raccolte in una rilevazione di alcuni anni prima, nel quale il coefficiente di correlazione lineare semplice sulle quantità di due componenti chimici di un alimento è risultato uguale a **0,19**. Egli si aspetta che, per le trasformazioni intervenute nell'ultimo periodo nella coltivazione e nella conservazione dei cibi, tale correlazione sia aumentata. Con un campione di **30** misure ha infatti trovato un valore di  $r = 0,48$ .

Calcolare:

- Quanti dati servono, per un esperimento con 2 campioni bilanciati, affinché il test risulti significativo alla probabilità  $\alpha = 0.05$  con un rischio di commettere un errore di II Tipo pari a una probabilità  $b = 0.20$ ?
- Poiché il campione storico è di **40** dati, quanti ne deve raccogliere con il secondo campione per rispettare le probabilità  $a$  e  $b$  precedenti?

Risposte

Dopo aver effettuato il test di significatività, poiché se esso risultasse positivo il problema sarebbe già risolto, si stima il numero di dati minimo per ognuno dei due gruppi a confronto.

- Dapprima si trasforma  $r_1 = 0,19$  in  $z_1$

$$z_1 = 0,5 \cdot \ln \left( \frac{1 + 0,19}{1 - 0,19} \right) = 0,5 \cdot \ln \left( \frac{1,19}{0,81} \right) = 0,5 \cdot \ln 1,469 = 0,5 \cdot 0,385 = 0,192$$

ottenendo  $z_1 = 0,192$

e  $r_2 = 0,48$  in  $z_2$

$$z_1 = 0,5 \cdot \ln\left(\frac{1+0,48}{1-0,48}\right) = 0,5 \cdot \ln\left(\frac{1,48}{0,52}\right) = 0,5 \cdot \ln 2,846 = 0,5 \cdot 1,046 = 0,523$$

ottenendo  $z_2 = 0,523$ .

- Successivamente, dalla distribuzione normale si ricava

il valore di  $Z$  per  $\alpha = 0.05$  unilaterale (perché il test chiede se vi è stato un aumento significativo)

ottenendo  $Z_a = 1,645$

e il valore di  $Z$  per  $\beta = 0.20$  unilaterale ottenendo  $Z_b = 0,84$ .

- Infine si ricava  $n$

$$n = 2 \left( \frac{1,645 + 0,84}{0,192 - 0,523} \right)^2 + 3 = \left( \frac{2,485}{0,331} \right)^2 + 3 = 7,508^2 + 3 = 56,4 + 3 = 59,4$$

che risulta uguale a 60.

Poiché è risultato un campione ( $n = 60$ ) non molto più grande di  $n_1 = 40$ , è possibile stimare il numero di dati necessario nel secondo ( $n_2$ ) per mantenere costante le probabilità richieste.

Applicando

$$n_2 = \frac{n_1(n+3) - 6n}{2n_1 - n - 3}$$

risulta

$$n_2 = \frac{40 \cdot (60 + 3) - 6 \cdot 60}{2 \cdot 40 - 60 - 3} = \frac{40 \cdot 63 - 360}{80 - 57} = \frac{2160}{23} = 93,9$$

che il secondo campione deve contenere almeno  $n_2 = 94$

E' una risposta che, per il numero non eccessivamente elevato di dati richiesti rispetto al campione già raccolto, rende possibile l'esperimento. Ad esempio, se il numero ( $n$ ) di dati richiesto per ogni gruppo fosse stato oltre il doppio degli  $n_1 = 40$  già raccolti con il primo campione, il numero di dati da raccogliere con il secondo ( $n_2$ ) sarebbe stato molto grande, tale da rendere molto costoso, se non impossibile, l'esperimento. Inoltre con due campioni così sbilanciati, le condizioni di potenza e significatività del test sarebbero state profondamente alterate.

Per altri concetti sul bilanciamento di due campioni indipendenti, si rinvia al capitolo sul test  $t$ , in quanto la procedura è simile.

**13.8. TEST PER LA DIFFERENZA TRA PIU' COEFFICIENTI DI CORRELAZIONE  
COEFFICIENTE DI CORRELAZIONE COMUNE  $r_w$  E SUA SIGNIFICATIVITA'**

Per il confronto simultaneo tra più coefficienti di correlazione indipendenti  $r_1, r_2, \dots, r_k$ , cioè per verificare l'ipotesi nulla

$$H_0: r_1 = r_2 = \dots = r_k$$

(seguendo le indicazioni di Jerrold H. Zar nel suo test del 1999 *Biostatistical Analysis*, fourth ed. Prentice Hall, New Jersey, pp. 390-394)

si stima un valore  $C_{k-1}^2$  con gdl  $k-1$

$$C_{k-1}^2 = \sum_{i=1}^k (n_i - 3) \cdot z_i^2 - \frac{\left[ \sum_{i=1}^k (n_i - 3) \cdot z_i \right]^2}{\sum_{i=1}^k (n_i - 3)}$$

dove

- $k$  è il numero di coefficienti di correlazione campionari  $r$  a confronto simultaneo
- $n_i$  è il numero di dati campionari di ogni  $r_i$
- $z_i$  è il valore di ogni  $r_i$  trasformato nel corrispondente valore  $z_i$  con la formula

$$z = 0,5 \cdot \ln \left( \frac{1+r}{1-r} \right)$$

Se il test non risulta significativo, si può concludere che i diversi valori  $r_i$  sono stime campionarie dello stesso valore reale  $r$ ; di conseguenza, come sua stima migliore, è utile calcolare il **coefficiente di correlazione comune** o **medio ponderato**  $r_w$  (*common r* or *weighted mean of r*)

con

$$z_w = \frac{\sum_{i=1}^k (n_i - 3) \cdot z_i}{\sum_{i=1}^k (n_i - 3)}$$

Successivamente, per un confronto con gli  $r_1, r_2, \dots, r_k$  originali,  $z_w$  può essere ritrasformato nella scala  $r$  (ricordando ancora che, in valore assoluto,  $r$  varia da  $0$  a  $1$ , mentre  $z$  varia da  $0$  a  $\infty$ ) attraverso

$$r = \frac{e^{2z} - 1}{e^{2z} + 1}$$

dove

- $e$  è la costante neperiana (approssimata con 2,718).

A questo valore comune  $r_w$ , con un test molto più potente rispetto a quelli di ogni singolo  $r_i$ , poiché ha il vantaggio di essere calcolato su tutte le  $N$  coppie di dati, in quanto

$$N = \sum_{i=1}^k n_i$$

può essere applicato

- sia il test per la verifica dell'ipotesi nulla

$$H_0: r = 0$$

che  $r_w$  non sia **significativamente differente da zero**

mediante la formula proposta da J. Neyman nel 1959 (vedi capitolo: *Optimal asymptotic test of composite hypothesis*, pp. 213-234. Nel libro di U. Grenander (ed.) **Probability and Statistics: The Harald Cramér Volume**, John Wiley, New York) e ripresa da S. R. Paul nel 1988

$$Z = \frac{\sum_{i=1}^k (n_i \cdot r_i)}{\sqrt{N}}$$

- sia il test per la verifica dell'ipotesi nulla

$$H_0: r = r_0$$

che  $r_w$  non sia **significativamente differente da un valore prefissato  $r_0$**

mediante la formula proposta da S. R. Paul nel 1988 (vedi, anche per la citazione precedente, l'articolo: *Estimation of and testing significance for a common correlation coefficient. Communic. Statist. - Theor. Meth.* 17: 39-53, nel quale fa una presentazione e disamina generale di questi metodi)

$$Z = (z_w - z_0) \cdot \sqrt{\sum_{i=1}^k (n_i - 3)}$$

dove

- $z_w$  è il valore  $r_w$  dopo trasformazione di Fisher,
- $z_0$  è il valore atteso o di confronto  $r_0$  dopo trasformazione di Fisher.

In questi due test sul coefficiente comune  $r_w$ , le ipotesi alternative  $H_1$  possono essere bilaterali oppure unilaterali. L'unica differenza consiste nella stima della probabilità  $\alpha$ : se in una distribuzione normale bilaterale oppure unilaterale.

Se i  $k$  campioni a confronto sono **dipendenti**, quali ad esempio i valori  $r_i$  di correlazione della variabile  $X_A$  con le variabili  $X_B$ ,  $X_C$ ,  $X_D$  calcolati sugli stessi prelievi, **questa tecnica non è corretta**. Si dovrà ricorrere alla statistica multivariata, con la correlazione multipla.

#### ESEMPIO 1.

In una ricerca sulla correlazione lineare semplice tra le quantità di due conservanti  $X_A$  e  $X_B$  contenuti in campioni di alimenti, con tre verifiche indipendenti sono stati ottenuti i seguenti risultati:

$$1 - r_1 = 0,48 \quad \text{con} \quad n_1 = 120$$

$$2 - r_2 = 0,31 \quad \text{con} \quad n_2 = 100$$

$$3 - r_3 = 0,48 \quad \text{con} \quad n_3 = 150$$

Con questi dati,

- verificare se tra questi  $r_i$  esiste una differenza significativa;
- in caso di non rifiuto della ipotesi nulla, calcolare il coefficiente comune  $r_w$ ;
- verificare se il coefficiente di correlazione comune  $r_w$  è significativamente maggiore di zero;
- verificare se  $r_w$  si discosta significativamente dal valore  $r_0 = 0,32$  indicato come valore reale in una pubblicazione presa a riferimento.

Risposte

A) Dopo aver trasformato i valori di  $r_i$  nei corrispondenti valori  $z_i$

$$z_1 = 0,5 \cdot \ln\left(\frac{1+0,48}{1-0,48}\right) = 0,5 \cdot \ln\left(\frac{1,48}{0,52}\right) = 0,5 \cdot \ln 2,846 = 0,5 \cdot 1,046 = 0,523$$

$$z_2 = 0,5 \cdot \ln\left(\frac{1+0,31}{1-0,31}\right) = 0,5 \cdot \ln\left(\frac{1,31}{0,69}\right) = 0,5 \cdot \ln 1,899 = 0,5 \cdot 0,641 = 0,320$$

$$z_3 = 0,5 \cdot \ln\left(\frac{1+0,37}{1-0,37}\right) = 0,5 \cdot \ln\left(\frac{1,37}{0,63}\right) = 0,5 \cdot \ln 2,175 = 0,5 \cdot 0,777 = 0,389$$

ottenendo:  $z_1 = 0,523$ ;  $z_2 = 0,320$ ;  $z_3 = 0,389$

si calcola un valore del  $c^2$  con **gdl= 2**  
applicando

$$c_{k-1}^2 = \sum_{i=1}^k (n_i - 3) \cdot z_i^2 - \frac{\left[ \sum_{i=1}^k (n_i - 3) \cdot z_i \right]^2}{\sum_{i=1}^k (n_i - 3)}$$

Scindendo per comodità di calcolo

$$c^2 = A - B$$

dove

$$A = \sum_{i=1}^k (n_i - 3) \cdot z_i^2$$

e

$$B = \frac{\left[ \sum_{i=1}^k (n_i - 3) \cdot z_i \right]^2}{\sum_{i=1}^k (n_i - 3)}$$

con i dati dell'esempio

$$A = (120 - 3) \cdot 0,523^2 + (100 - 3) \cdot 0,320^2 + (150 - 3) \cdot 0,389^2$$

$$A = 117 \cdot 0,274 + 97 \cdot 0,102 + 147 \cdot 0,151 = 32,058 + 9,894 + 22,197 = 64,149$$

risultano **A = 64,149**

e

$$B = \frac{[(120 - 3) \cdot 0,523 + (100 - 3) \cdot 0,320 + (150 - 3) \cdot 0,389]^2}{(120 - 3) + (100 - 3) + (150 - 3)}$$

$$B = \frac{(61,191 + 31,04 + 57,183)^2}{117 + 97 + 147} = \frac{149,417^2}{361} = \frac{22325,44}{361} = 61,843$$

**B = 61,843**

determinando

$$c^2 = 64,149 - 61,843 = 2,306$$

un valore di  $c^2_2 = 2,306$ .

Poiché il valore critico alla probabilità  $\alpha = 0,05$  con **df = 2** è uguale a **5,991** non si può rifiutare l'ipotesi nulla. Una lettura più attenta della tabella dei valori critici evidenzia che il valore calcolato (**2,306**) è inferiore anche a quello riportato per la probabilità  $\alpha = 0,25$  che risulta uguale a **2,773**.

Si può quindi concludere che i tre coefficienti di correlazione  $r_1$ ,  $r_2$  e  $r_3$  sono statisticamente uguali.

B) Con questi risultati dell'analisi sulla significatività della differenza tra i  $k$  valori campionari, è utile calcolare il coefficiente di correlazione comune  $r_w$ , come stima migliore del valore  $r$  della popolazione.

Con  $z_1 = 0,523$ ;  $z_2 = 0,320$ ;  $z_3 = 0,389$  e  $n_1 = 120$ ,  $n_2 = 100$ ,  $n_3 = 150$

il valore di  $z_w$  comune

$$z_w = \frac{(120-3) \cdot 0,523 + (100-3) \cdot 0,320 + (150-3) \cdot 0,389}{(120-3) + (100-3) + (150-3)}$$

$$z_w = \frac{61,191 + 31,04 + 57,183}{117 + 97 + 147} = \frac{149,417}{361} = 0,414$$

risulta uguale a **0,414**.

Ritrasformato in  $r_w$

$$r_w = \frac{2,718^{2 \cdot 0,414} - 1}{2,718^{2 \cdot 0,414} + 1} = \frac{2,718^{0,828} - 1}{2,718^{0,828} + 1} = \frac{2,289 - 1}{2,289 + 1} = \frac{1,289}{3,289} = 0,392$$

risulta uguale a **0,392**.

C) Per la verifica dell'ipotesi nulla  $H_0: r = 0$

con ipotesi alternativa unilaterale  $H_1: r > 0$

**si utilizzano le varie osservazioni campionarie**, per cui

da  $r_1 = 0,48$ ;  $r_2 = 0,31$ ;  $r_3 = 0,37$  e  $n_1 = 120$ ,  $n_2 = 100$ ,  $n_3 = 150$

si ottiene

$$Z = \frac{(120 \cdot 0,48) + (100 \cdot 0,31) + (150 \cdot 0,37)}{\sqrt{120 + 100 + 150}} = \frac{57,6 + 31,0 + 55,5}{\sqrt{370}} = \frac{144,1}{19,24} = 7,49$$

un valore di  $Z$  uguale a **7,49**.

Poiché nella **distribuzione normale unilaterale** il valore ottenuto è così grande da non essere nemmeno riportato nella tabella, ad esso è associata una probabilità  $\alpha$  estremamente piccola. Di conseguenza, si può concludere che il valore medio  $r_w$  è significativamente maggiore di zero.

D) Per la verifica dell'ipotesi nulla  $H_0: r = r_0$

con ipotesi alternativa bilaterale  $H_1: r \neq r_0$

cioè che  $r_w = 0,392$  sia **significativamente differente da  $r_0 = 0,32$**

attraverso

$$Z = (z_w - z_0) \cdot \sqrt{\sum_{i=1}^k (n_i - 3)}$$

dopo aver calcolato il coefficiente di correlazione comune in scala  $\mathbf{z}$ , cioè  $\mathbf{z}_w = \mathbf{0,414}$

si trasforma nella stessa scala il valore  $r_0 = \mathbf{0,32}$

ottenendo

$$z_{r_0} = z_0 = 0,5 \cdot \ln\left(\frac{1+0,32}{1-0,32}\right) = 0,5 \cdot \ln\frac{1,32}{0,68} = 0,5 \cdot \ln 1,941 = 0,5 \cdot 0,663 = 0,331$$

Infine si calcola  $\mathbf{Z}$

$$Z = (0,414 - 0,331) \cdot \sqrt{117 + 97 + 147} = 0,083 \cdot 19 = 1,58$$

ottenendo un valore uguale a 1,58.

In una **distribuzione normale bilaterale**, a  $\mathbf{Z} = \mathbf{1,58}$  corrisponde una probabilità  $\alpha = \mathbf{0,114}$ . Non è possibile rifiutare l'ipotesi nulla:  $\mathbf{r}_w = \mathbf{0,392}$  non è significativamente differente da  $\mathbf{r}_0 = \mathbf{0,32}$ .

### Correzioni per il bias della trasformazione di Fisher

La trasformazione di  $\mathbf{r}$  in  $\mathbf{z}$  con la formula di Fisher determina un errore piccolo, ma sistematico e in eccesso, nel valore di  $\mathbf{z}$ . E' un risultato noto da tempo, già evidenziato da H. **Hotelling** nel 1953 (vedi articolo: *New light on the correlation coefficient and its trasformation. Journal Royal Statistical Society B* 15, pp. 193-232) e discusso dallo stesso R. A. **Fisher** nel 1958 (vedi testo: *Statistical Methods for Research Workers*, 13<sup>th</sup> edition, Hafner, New York, 146 pp.).

Ne deriva, in particolare quando si ricorre a test per la significatività della differenza tra  $\mathbf{k}$  campioni, utilizzando ad esempio la formula già presentata

$$c_{k-1}^2 = \sum_{i=1}^k (n_i - 3) \cdot z_i^2 - \frac{\left[ \sum_{i=1}^k (n_i - 3) \cdot z_i \right]^2}{\sum_{i=1}^k (n_i - 3)}$$

che gli errori si sommano e la differenza tra valore calcolato e valore reale non è più trascurabile.

L'errore è tanto maggiore quanto più alto è  $\mathbf{k}$ .

Per la correzione del *bias* nel valore di  $\mathbf{z}$ , ottenuto con la trasformazione di  $\mathbf{r}$ , le proposte sono numerose. Tra le più frequenti per aggiustare  $\mathbf{z}$  possono esserne citate due:

A) **La formula di Hotelling** che, nell'articolo già citato del 1953, propone di sottrarre  $\mathbf{c}$

$$c = \frac{3z + r}{4(n-1)}$$

al valore  $z$  stimato.

Ad esempio, con  $r = 0,48$  calcolato su  $n = 120$

$$z_1 = 0,5 \cdot \ln\left(\frac{1+0,48}{1-0,48}\right) = 0,5 \cdot \ln\left(\frac{1,48}{0,52}\right) = 0,5 \cdot \ln 2,846 = 0,5 \cdot 1,046 = 0,523$$

si ottiene un valore di  $z = 0,523$

Il valore corretto, ottenuto con la sottrazione della correzione

$$c = \frac{3 \cdot 0,523 + 0,48}{4 \cdot (120 - 1)} = \frac{2,009}{476} = 0,004$$

$c = 0,004$  diventa  $z' = z - c = 0,523 - 0,004 = 0,519$ .

B) **La formula di Fisher**, proposta nel testo del 1955 per la correzione, è

$$c = \frac{r}{2(n-1)}$$

Con lo stesso esempio, la correzione è

$$c = \frac{0,48}{2 \cdot (120 - 1)} = \frac{0,48}{236} = 0,002$$

$c = 0,002$  e il valore corretto diventa  $z' = z - c = 0,523 - 0,002 = 0,521$ .

Per il calcolo della significatività della differenza tra più coefficienti, invece di correggere ogni valore  $z_i$  e usare la formula generale già presentata, è possibile utilizzare direttamente

**la formula corretta di Paul**,

$$c_{Paul}^2 = \sum_{i=1}^k \frac{n_i \cdot (r_i - r_w)^2}{(1 - r_i \cdot r_w)^2}$$

che, come la precedente, utilizza la distribuzione  $c^2$  con  $df = k-1$ .

Ad esempio, applicata agli stessi dati della formula generale che ha stimato un valore di  $c^2_2 = 2,306$ , con

$r_1 = 0,48$ ;  $r_2 = 0,31$ ;  $r_3 = 0,37$  e  $n_1 = 120$ ,  $n_2 = 100$ ,  $n_3 = 150$  e  $r_w = 0,392$

si ottiene

$$c_{Paul}^2 = \frac{120 \cdot (0,48 - 0,392)^2}{[1 - (0,48 \cdot 0,392)]^2} + \frac{100 \cdot (0,31 - 0,392)^2}{[1 - (0,31 \cdot 0,392)]^2} + \frac{150 \cdot (0,37 - 0,392)^2}{[1 - (0,37 \cdot 0,392)]^2}$$

$$c_{Paul}^2 = \frac{120 \cdot 0,088^2}{(1 - 0,188)^2} + \frac{100 \cdot 0,082^2}{(1 - 0,122)^2} + \frac{150 \cdot 0,022^2}{(1 - 0,180,145)^2}$$

$$c_{Paul}^2 = \frac{0,924}{0,659} + \frac{0,67}{0,771} + \frac{0,075}{0,731} = 1,402 + 0,869 + 0,103 = 2,374$$

un valore pari a 2,374.

Sempre S. R. **Paul** nella sua ampia presentazione del 1988 (*Estimation of and testing significance for a common correlation coefficient. Communic. Statist. - Theor. Meth.* 17: 39-53) suggerisce che quando r è minore di 0,5 (indicazione approssimata)

- per il calcolo del coefficiente medio  $z_w$

al posto di

$$z_w = \frac{\sum_{i=1}^k (n_i - 3) \cdot z_i}{\sum_{i=1}^k (n_i - 3)}$$

sia usato

$$z_w = \frac{\sum_{i=1}^k (n_i - 1) \cdot z_i}{\sum_{i=1}^k (n_i - 1)}$$

- e nell'inferenza con ipotesi nulla  $H_0: r = r_0$  per calcolare  $Z$

al posto di

$$Z = (z_w - z_0) \cdot \sqrt{\sum_{i=1}^k (n_i - 3)}$$

sia usata

$$Z = (z_w - z_0) \cdot \sqrt{\sum_{i=1}^k (n_i - 1)}$$

### 13.9. CENNI SUI CONFRONTI MULTIPLI TRA PIU' r

Come nel caso delle medie, effettuato il test per la verifica dell'uguaglianza tra più coefficienti di correlazione, se si arriva alla conclusione che non tutti i coefficienti di correlazione sono tra loro uguali, si pone il problema di sapere tra quali la differenza sia significativa.

E' possibile pervenire alla soluzione sia

- con **confronti a priori**, attraverso la tecnica dei contrasti ortogonali,
- con **confronti a posteriori** o multipli, in modo analogo al test SNK o alla procedura di Tukey.

Senza riprendere dettagliatamente i concetti generali, già ampiamente discussi nei confronti tra **k** medie, è utile ricordare che i confronti a priori godono del vantaggio rilevante di utilizzare per ogni contrasto la stessa probabilità  $\alpha$  del test generale, mentre i confronti a posteriori, sulla base del principio del Bonferroni, tendono ad utilizzare per ogni confronto una probabilità  $\alpha/k$ , dove **k** è il numero di confronti possibili.

Per i **confronti a priori**, con la tecnica dei contrasti o confronti ortogonali,

- dapprima si stima un valore critico  $S_a$  per la probabilità  $\alpha$  prefissata con

$$S_a = \sqrt{c_{a,k-1}^2}$$

- successivamente si calcola **S**

$$S = \frac{\left| \sum_i c_i z_i \right|}{SE}$$

dove

$$SE = \sqrt{\sum_i c_i^2 s_{zi}^2}$$

- $c_i$  è il coefficiente del contrasto,
- $z_i$  è la trasformazione di **r** in **z**.

Sono significativi i contrasti il cui valore **S** supera il valore critico  $S_a$ .

Per i **confronti a posteriori**, i metodi proposti sono numerosi.

Una procedura semplice prevede che, sempre dopo l'analisi complessiva con il  $c^2$  che deve risultare significativa, tra

due coefficienti di correlazione  $r_A$  e  $r_B$  calcolati su campioni di dimensioni  $n_A$  e  $n_B$  esista una differenza significativa se il valore **q**

$$q = \frac{z_A - z_B}{\sqrt{\frac{1}{2} \left( \frac{1}{n_A - 3} + \frac{1}{n_B - 3} \right)}}$$

è superiore a quello riportato nella tabella dei valori critici del q studentizzato con indici  $\alpha$ ,  $\infty$ , k dove

- $\alpha$  è la probabilità complessiva,
- $\infty$  è il numero totale di dati utilizzati per il confronto dei k gruppi,
- p è il numero di passi tra i due valori r a confronto, dopo che sono stati tutti ordinati per rango.

Anche nel caso dei coefficienti di correlazione, il confronto tra k gruppi può essere tra un controllo e k-1 valori sperimentali. Il metodo è analogo a quello spiegato per le medie, poiché il numero di confronti possibili non è più  $k(k-1)/2$  ma scende a **k-1**.

### **13.10. LA CORRELAZIONE PARZIALE O NETTA DI PRIMO ORDINE E DI ORDINE SUPERIORE; LA CORRELAZIONE SEMIPARZIALE**

Quando si analizzano le relazioni tra più variabili, la correlazione tra due di esse risente anche delle relazioni esistenti con le altre. Sovente, nella ricerca è richiesto di valutare l'associazione tra due variabili, eliminando l'influsso delle altre:

- è la **correlazione parziale o netta** (*partial correlation*),
- mentre quella discussa nei paragrafi precedenti è la **correlazione semplice o totale**.

Per esempio, nel caso in cui si intenda valutare le correlazioni tra 3 variabili come altezza, peso e circonferenza toracica, le relazioni esistenti tra circonferenza toracica ed altezza sono influenzate in modo rilevante da quelle esistenti tra ognuna di queste due con il peso. Nello stesso modo, la correlazione tra altezza e diametro del tronco di un albero risente della correlazione di entrambi con la sua età.

**La correlazione parziale o netta è la stima della correlazione tra due variabili, dopo l'eliminazione degli effetti dovuti all'eventuale associazione con la terza (o il restante gruppo di k variabili).**

Un metodo teoricamente possibile per valutare la correlazione netta tra due variabili sarebbe la misura della correlazione semplice o totale, mantenendo costante la terza variabile. Ma questa procedura presenta vari inconvenienti, facilmente identificabili:

- **la necessità di ripetere più volte i calcoli;**
- **l'impossibilità di estendere e generalizzare le conclusioni**, poiché per ogni valore della terza variabile si avrebbe una correlazione con un valore campionario differente;
- **un forte aumento della numerosità del campione** e quindi sia dei costi che dei tempi richiesti dalla ricerca.

Nel linguaggio statistico, per misurare la correlazione parziale o netta tra due variabili si distinguono **correlazioni di vari gradi od ordini**, in rapporto al numero di variabili complessivamente utilizzate, ricordando che il concetto di correlazione riguarda la relazione esistente tra due.

- Quando si dispone solamente delle osservazioni relative a due variabili (come in precedenza), **la correlazione è detta di grado zero o di ordine zero**;
  - quando le variabili osservate sono tre, la correlazione tra due senza l'influenza della terza è detta **correlazione di 1° grado o di 1° ordine**;
- con quattro variabili, eliminata l'influenza di due,
- **la correlazione è di 2° grado o di 2° ordine**;
- con N variabili, eliminata l'influenza delle altre N-2,
- **la correlazione tra due variabili è di grado od ordine (N-2)esimo.**

**Nel caso di tre variabili**, quando sono stati calcolati i coefficienti di correlazione semplice o totale, **il coefficiente di correlazione parziale o netta** (scritta come  $r_{12,3}$  e detta **correlazione tra le variabili  $X_1$  e  $X_2$  al netto degli effetti della variabile  $X_3$** ) è data da

$$r_{12,3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2) \cdot (1 - r_{23}^2)}}$$

con **gdl N-3**, dove

- $r_{12,3}$  è la correlazione parziale tra le variabili 1 e 2, a meno (o al netto) degli effetti della 3;
- $r_{12}$ ,  $r_{13}$ ,  $r_{23}$  sono le correlazioni semplici tra le rispettive coppie di variabili.

Per la stima della correlazione netta, **le condizioni di validità** sono essenzialmente due:

- le correlazioni di ordine zero devono essere lineari;
- il numero N di osservazioni di ogni correlazione di ordine zero deve essere sempre superiore di alcune unità al numero delle variabili, poiché **il numero di gdl della correlazione parziale con k variabili è uguale a N-k.**

ESEMPIO. Si considerino i 18 laghi dell'Appennino tosco-emiliano, già utilizzati precedentemente, per ognuno dei quali è stata misurata

- la conducibilità ( $X_1$ ),
- la concentrazione di Ione Calcio + Ione Magnesio ( $X_2$ ),
- la concentrazione di Solfati + Carbonati ( $X_3$ ).

I valori sono riportati nella tabella successiva

Laghi	Conducibilità ( $X_1$ )	Ione Calcio + Ione Magnesio ( $X_2$ )	Solfati + Carbonati ( $X_3$ )
SILLARA INF.	20	0,063	0,137
SILLARA SUP.	22	0,077	0,149
SCURO CERRETO	22	0,078	0,095
VERDAROLO	26	0,125	0,156
SQUINCIO	24	0,120	0,107
SCURO PARMENSE	28	0,144	0,191
PALO	27	0,143	0,228
ACUTO OVEST	26	0,115	0,212
SCURO RIGOSO	29	0,185	0,244
COMPIONE INF.	35	0,194	0,322
GEMIO INF.	33	0,193	0,301
PETRUSCHIA	37	0,218	0,304
GEMIO SUP.	34	0,207	0,312
SANTO PARMENSE	35	0,254	0,311
BICCHIERE	37	0,250	0,352
BALLANO	39	0,315	0,354
BACCIO	41	0,364	0,415
VERDE	45	0,338	0,459

Calcolare i tre coefficienti di correlazione semplice o totale. Successivamente, al fine di valutare con maggiore precisione la correlazione esistente tra queste variabili, stimare i tre coefficienti di correlazione parziale.

Risposta.

I tre coefficienti di correlazione semplice o totale (calcolati con un programma informatico) sono risultati:

$$r_{12} = 0,9628$$

$$r_{13} = 0,9704$$

$$r_{23} = 0,9388$$

Poiché la tabella di valori di correlazione semplice

con **16 df** ( $N-2 = 18-2$ ) e alla probabilità  $\alpha = \mathbf{0.001}$ , riporta il valore 0,7484

si deve concludere che i tre coefficienti di correlazione semplice sono tutti altamente significativi.

L'analisi con la correlazione netta permette di valutare se questi valori di correlazione tra coppie di variabili sono rafforzati dalla comune correlazione con la terza, per cui le correlazioni reali tra coppie di variabili sono minori.

Applicando la precedente formula

$$r_{12,3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2) \cdot (1 - r_{23}^2)}}$$

le stime dei 3 coefficienti di correlazione parziale danno i risultati seguenti:

-  $r_{12,3}$

$$r_{12,3} = \frac{0,9628 - 0,9704 \cdot 0,9388}{\sqrt{(1 - 0,9707^2) \cdot (1 - 0,9388^2)}} = \frac{0,9628 - 0,9110}{\sqrt{0,0583 \cdot 0,1187}} = \frac{0,0518}{0,0832} = 0,623$$

è uguale a 0,623;

-  $r_{13,2}$

$$r_{13,2} = \frac{0,9704 - 0,9628 \cdot 0,9388}{\sqrt{(1 - 0,9628^2) \cdot (1 - 0,9388^2)}} = \frac{0,9704 - 0,9039}{\sqrt{0,073 \cdot 0,1187}} = \frac{0,0665}{0,0931} = 0,714$$

è uguale a 0,714;

-  $r_{23,1}$

$$r_{23,1} = \frac{0,9388 - 0,9628 \cdot 0,9704}{\sqrt{(1 - 0,9628^2) \cdot (1 - 0,9704^2)}} = \frac{0,9388 - 0,9343}{\sqrt{0,0730 \cdot 0,0583}} = \frac{0,0045}{0,0652} = 0,069$$

è uguale a 0,069.

I valori critici sono riportati nella solita tabella dei coefficienti di correlazione;

questi **coefficienti di correlazione con df 15** ( $N-k = 18-3$ )

- alla probabilità  $\alpha = \mathbf{0.05}$  danno un valore di **r** uguale a **0,4821**
- mentre alla probabilità  $\alpha = \mathbf{0.01}$  è uguale a **0,6055**
- e alla probabilità  $\alpha = \mathbf{0.001}$  è **0,7247**.

E' semplice osservare che i 3 coefficienti netti calcolati risultano minori di quelli semplici; quindi è ovvio dedurre che la correlazione totale tra due variabili era aumentata dalla comune correlazione con la terza variabile.

In merito alla loro significatività, in modo più dettagliato,

- il valore di  $r_{12,3}$  risulta significativo con probabilità inferiore a 0.01
- il valore di  $r_{13,2}$  è significativo con probabilità inferiore a 0.001
- il valore di  $r_{23,1}$  non è significativo; anzi è molto distante dalla significatività e prossimo alla totale assenza di correlazione.

Nell'analisi dei coefficiente di correlazione parziale o netta si possono realizzare due situazioni:

- se **il valore parziale è maggiore di quello semplice** o addirittura diventa significativo, si deve dedurre che l'altro fattore nasconde la correlazione che effettivamente esiste;
- se **il coefficiente parziale è minore di quello semplice** o addirittura perde la significatività, si può dedurre che la terza variabile considerata è fortemente correlata con entrambe e le fa variare congiuntamente, **senza che tra esse esista una relazione diretta.**

La significatività della regressione risente in modo marcato di due fattori:

- il numero di osservazioni
- il campo di variazione di  $X_1$  e  $X_2$ .

L'importanza del primo fattore è evidenziata dalla semplice lettura della tabella dei valori critici, che diminuiscono in modo rilevante all'aumentare del numero di gradi di libertà, come già sottolineato per la regressione lineare semplice.

Per il secondo fattore, si può osservare che

- quando i valori delle due variabili hanno un intervallo molto limitato, il coefficiente di correlazione ha un valore assoluto molto basso, difficilmente significativo;
- al contrario, quando anche solo una variabile ha un intervallo di variazione molto ampio, il coefficiente di correlazione è molto alto.

**Per la corretta programmazione di un esperimento, è quindi conveniente**

- **raccogliere in precedenza informazioni sulla loro variabilità e**
- **impostare la raccolta dei dati in modo che essa sia grande.**

Inoltre, poiché l'interpretazione della correlazione tra due variabili è fortemente influenzata

- sia dal numero di dati,
- sia dal campo di variazione,

è difficile confrontare la significatività di due coefficienti di correlazione con dati raccolti in condizioni diverse.

Con metodi molto simili alla **correlazione parziale di primo ordine**, gli stessi principi possono essere estesi a 4 o più variabili, con la **correlazione parziale di secondo ordine e a quella di ordine superiore**.

**Prendendo in considerazione 4 variabili** ( $X_1, X_2, X_3, X_4$ ),

ognuna con lo stesso numero **N di osservazioni**,

si devono calcolare i 6 coefficienti di correlazione semplice o totale ( $r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34}$ ).

**La correlazione parziale di secondo ordine tra due di queste variabili  $X_1$  e  $X_2$ , mantenendo costanti le altre due  $X_3$  e  $X_4$**  (scritta  $r_{12,34}$ ),

può essere calcolata con la formula

$$r_{12,34} = \frac{r_{12,3} - r_{14,3} \cdot r_{24,3}}{\sqrt{(1 - r_{14,3}^2) \cdot (1 - r_{24,3}^2)}}$$

che utilizza tre correlazioni parziali di primo ordine ( $r_{12,3}; r_{14,3}; r_{24,3}$ ).

**Con più di 4 variabili** ( $X_1, X_2, X_3, X_4, \dots, X_n$ ), la formula generale, per calcolare la correlazione parziale tra la variabile  $X_1$  e  $X_2$  con le variabili  $X_3$  e  $X_G$  (il gruppo di tutte le altre) mantenute costanti, diventa

$$r_{12,3G} = \frac{r_{12,G} - r_{13,G} \cdot r_{23,G}}{\sqrt{(1 - r_{13,G}^2) \cdot (1 - r_{23,G}^2)}}$$

Per i calcoli è necessario utilizzare un programma informatico, dato l'alto numero di operazioni da effettuare e quindi l'elevata probabilità di commettere errori.

#### ESEMPIO

In dodici laghi sono state rilevate 4 parametri chimici ( $SO_4, HCO_3, NH_4, K_2OC$ ) tra loro legati da correlazioni di intensità e segno diverso.

Dalla serie di 12 dati per le 4 variabili:

Variabili	X <sub>1</sub> SO <sub>4</sub>	X <sub>2</sub> HCO <sub>3</sub>	X <sub>3</sub> NH <sub>4</sub>	X <sub>4</sub> K <sub>2</sub> OC
Laghi	0,100	0,418	0,005	54
Osiglia	0,067	0,407	0,008	54
Riondo	0,073	0,281	0,002	39
Bicchiere	0,110	0,242	0,004	37
Gemio Inferiore	0,106	0,195	0,003	33
Gemio Superiore	0,108	0,204	0,003	34
Pradaccio	0,131	0,455	0,003	53
Santo Parmense	0,081	0,230	0,003	35
Scuro Parmense	0,067	0,124	0,004	28
Scuro Rigoso	0,102	0,142	0,001	29
Verdarolo	0,069	0,087	0,000	26
Verde	0,092	0,367	0,003	45

- calcolare  $r_{12,34}$

cioè la correlazione tra le variabili X<sub>1</sub> e X<sub>2</sub> al netto delle variabili X<sub>3</sub> e X<sub>4</sub>.

Risposta.

Da questa serie di 12 dati campionari, tra le 4 variabili è possibile ricavare le 6 correlazioni semplici riportate nella tabella:

	Correlazioni semplici tra vettori			
	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
X <sub>1</sub>	----			
X <sub>2</sub>	0.327	----		
X <sub>3</sub>	-0.106	0.596	----	
X <sub>4</sub>	0.261	0.986	0.0670	----

Ognuna di esse ha 10 gdl; la loro significatività può essere verificata con i metodi già presentati.

Per calcolare  $r_{12,34}$ , occorre prima stimare  $r_{12,3}$   $r_{14,3}$   $r_{24,3}$

1) Con

$$r_{12,3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2) \cdot (1 - r_{23}^2)}}$$

si ottiene

$r_{12,3} = 0,4889$  con 9 gdl corrispondente ad una probabilità  $\alpha = 0.127$

2) Con

$$r_{14,3} = \frac{r_{14} - r_{13} \cdot r_{43}}{\sqrt{(1 - r_{13}^2) \cdot (1 - r_{43}^2)}}$$

si ottiene

$r_{14,3} = \mathbf{0,4498}$  con 9 gdl corrispondente ad una probabilità  $\alpha = 0.165$

3) Con

$$r_{24,3} = \frac{r_{24} - r_{23} \cdot r_{43}}{\sqrt{(1 - r_{23}^2) \cdot (1 - r_{43}^2)}}$$

si ottiene

$r_{24,3} = \mathbf{0,9847}$  con 9 gdl corrispondente ad una probabilità  $\alpha = 0.000$

l'unica delle tre che risulti significativa.

Da queste 3 correlazioni parziali di primo ordine, mediante

$$r_{12,34} = \frac{r_{12,3} - r_{14,3} \cdot r_{24,3}}{\sqrt{(1 - r_{14,3}^2) \cdot (1 - r_{24,3}^2)}}$$

si ricava il valore desiderato, che risulta uguale a

$r_{12,34} = \mathbf{0,2954}$  con 8 gdl corrispondente ad una probabilità  $\alpha = 0.407$

Mentre la correlazione parziale  $r_{12,3}$  valuta gli effetti tra due variabili ( $X_1$  e  $X_2$ ) dopo che entrambe sono state aggiustate per la regressione con la terza variabile ( $X_3$ ), per cui essa è la correlazione tra due variabili aggiustate ( $X_{1,3}$  e  $X_{2,3}$ ),

la **correlazione semiparziale** (*semipartial or part correlation*)  $r_{1(2,3)}$

- valuta la correlazione tra la variabile  $X_1$  e la variabile  $X_2$ ,
- dopo che solo la  $X_2$  è stata aggiustata per la variabile  $X_3$  (indicata con  $X_{2,3}$ ).

La correlazione semiparziale è quindi la correlazione tra una variabile non aggiustata ( $X_1$ ) ed una seconda variabile aggiustata per la terza ( $X_{2,3}$ )

ed è calcolata con

$$r_{1(2,3)} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1 - r_{23}^2}}$$

**Un esempio di correlazione parziale riportato nei testi è la correlazione tra la perdita di peso in un gruppo N pazienti e il tempo che ognuno di essi dedica alla ginnastica, aggiustato per il consumo di calorie che l'esercizio, più o meno faticoso, comporta. Nella ricerca ambientale può essere il caso di salinità, ossigeno e temperatura**

Con 3 variabili ( $X_1, X_2, X_3$ ) sono teoricamente possibili 6 correlazioni parziali: oltre alla precedente  $r_{1(2,3)}$ , si hanno

$$r_{1(3,2)} = \frac{r_{13} - r_{12} \cdot r_{32}}{\sqrt{1 - r_{32}^2}}$$

$$r_{2(1,3)} = \frac{r_{21} - r_{23} \cdot r_{13}}{\sqrt{1 - r_{13}^2}}$$

ed in modo analogo le altre tre  $r_{2(3,1)}, r_{3(1,2)}, r_{3(2,1)}$

anche se spesso hanno reale significato ecologico od ambientale solo una o due di esse. Non è quindi necessario calcolarle tutte, ma è bene limitare l'analisi a quelle che possono essere interpretate.

**Anche nella correlazione semiparziale, una variabile può essere aggiustata per un numero più alto di variabili, non solo per una terza.**

**La correlazione semiparziale di secondo ordine  $r_{1(2,34)}$  può essere stimata**

- sia partendo dalle correlazioni parziali di primo ordine con

$$r_{1(2,34)} = \frac{(r_{12,3} - r_{14,3} \cdot r_{24,3}) \cdot \sqrt{1 - r_{13}^2}}{\sqrt{1 - r_{24,3}^2}}$$

- sia partendo da quella di secondo ordine con

$$r_{1(2,34)} = r_{12,34} \frac{\sqrt{1 - r_{13}^2}}{\sqrt{1 - r_{14,3}^2}}$$

Le correlazioni parziali e semiparziali hanno la loro applicazione principale nello studio delle **inter-relazioni lineari** che esistono fra tre o più variabili. E' un concetto di estrema importanza nella

teoria della **regressione multipla**, che è alla base della statistica multivariata. Lo studio delle relazioni tra variabili è quindi meglio affrontato con queste metodologie.

### 13.11. ANALISI DELLA COVARIANZA (ANCOVA)

Combinando l'analisi della varianza con la regressione, si ottiene l'analisi della covarianza (ANCOVA). Il caso più semplice è quello di un disegno sperimentale completamente randomizzato, con una sola covariata (X), in cui la regressione tra le due variabili permette di ridurre la varianza d'errore della Y.

Con la simbologia consueta, mentre nell'analisi della varianza completamente randomizzata il modello è

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

nell'analisi della covarianza diviene

$$Y_{ij} = \mu + \alpha_i + \beta(x_{ij} - \bar{x}) + \varepsilon_{ij}$$

Nell'analisi della covarianza, è possibile utilizzare una variabile concomitante X, definita covariata, che spiega parte della variabilità della variabile dipendente Y. Consente un calcolo più preciso degli effetti dei trattamenti, riducendo la variabilità casuale.

Per esempio, si supponga di voler stimare l'effetto di tre tossici ( $\alpha_i$ ) sul peso ( $Y_i$ ) delle cavie. L'analisi della varianza ad un solo criterio richiede che si formino tre gruppi inizialmente identici per caratteristiche degli individui; dopo un periodo di somministrazione del tossico, le differenze tra i pesi medi dei tre gruppi permettono di valutare l'effetto dei tossici sulla crescita ponderale delle cavie. Ma avere tre gruppi di cavie con le stesse dimensioni iniziali spesso è difficile, soprattutto se l'esperimento non avviene in laboratorio ma in natura, con animali catturati.

Un altro caso sovente citato nei testi di statistica applicata è il confronto tra il peso ( $Y_i$ ) di organi di animali sottoposti a trattamenti ( $\alpha_i$ ) diversi, in cui il peso dell'organo dipende sia dall'effetto del trattamento che dalle dimensioni ( $X_i$ ) dell'animale. E' possibile eliminare dal peso dell'organo quella quantità che dipende delle dimensioni dell'animale [ $\beta(x_{ij} - \bar{x})$ ], per valutare in modo più preciso la quota che può essere attribuita al trattamento.

I modi più semplici e diffusi per eseguire l'analisi della covarianza, in particolare quando si attuano calcoli manuali, sono due:

- il **metodo delle Y ridotte**,
- il **metodo delle Y aggiustate**.

Il **metodo delle Y ridotte** richiede alcuni passaggi, che possono essere riassunti in **5 punti**.

1 - Si applica l'**analisi della varianza alle Y**, per verificare l'ipotesi nulla

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

se le medie delle popolazioni dalle quali sono estratti i k campioni a confronto sono uguali, con ipotesi alternativa

$H_1$  che non tutte le  $\mu$  sono uguali.

Con le formule abbreviate abituali si calcolano:

- la **devianza totale delle Y**

$$\sum Y_{ij}^2 - \frac{(\sum Y_{ij})^2}{n} \quad \text{con df } \mathbf{n-1}$$

- la **devianza tra trattamenti delle Y**

$$\sum \frac{(\sum Y_i)^2}{n_i} - \frac{(\sum Y_{ij})^2}{n} \quad \text{con df } \mathbf{k-1}$$

- e, per differenza, la **devianza d'errore delle Y**

$$\text{Dev. Y totale} - \text{Dev. Y tra tratt.} \quad \text{con df } \mathbf{n-k}$$

dove

$n$  è il numero totale di dati,

$n_i$  il numero di dati entro ogni gruppo,

$k$  il numero di gruppi a confronto.

Stimata la varianza tra trattamenti e la varianza d'errore, si perviene al test F

$$F_{(k-1, n-k)} = \frac{s^2 \text{tratt.}}{s^2 \text{errore}}$$

che **raramente risulta significativo, se l'effetto della regressione di Y su X è elevato.**

2 - Per correggere i dati calcolati, si valuta l'**effetto della regressione** ricordando che è stimata dal rapporto

$$\frac{Cod_{\cdot XY}^2}{Dev_{\cdot X}}$$

A questo scopo si devono calcolare:

- la **codevianza XY totale**

$$\sum (X_{ij} \cdot Y_{ij}) - \frac{(\sum X_{ij}) \cdot (\sum Y_{ij})}{n} \quad \text{con df } \mathbf{n-1}$$

- la **codevianza XY tra trattamenti**

$$\sum \frac{(\sum X_{ij} \cdot \sum Y_{ij})}{n_i} - \frac{(\sum X_{ij}) \cdot (\sum Y_{ij})}{n} \quad \text{con df } \mathbf{k-1}$$

- e, per differenza, la **codevianza XY d'errore**

$$\text{Cod. XY totale} - \text{Cod. XY tra trattamenti} \quad \text{con df } \mathbf{n-k}$$

(A questo proposito, è importante ricordare che le **Codevianze possono risultare sia positive che negative**. Per esempio, con una Codevianza XY totale positiva si può ottenere anche una Codevianza XY tra trattamenti che risulti negativa; di conseguenza, la Codevianza XY d'errore può risultare maggiore di quella totale.)

3 - Sempre per stimare l'effetto della regressione, è preliminare effettuare il calcolo delle devianze delle X, con le stesse modalità seguite per la Y:

- la **devianza totale delle X**

$$\sum X_{ij}^2 - \frac{(\sum X_{ij})^2}{n} \quad \text{con df } \mathbf{n-1}$$

- la **devianza tra trattamenti delle X**

$$\sum \frac{(\sum X_i)^2}{n_i} - \frac{(\sum X_{ij})^2}{n} \quad \text{con df } \mathbf{k-1}$$

- e, per differenza, la **devianza d'errore delle X**

$$\text{Dev. X totale} - \text{Dev. X tra tratt.} \quad \text{con df } \mathbf{n-k}$$

4 - Con i valori calcolati al punto 2 e al punto 3, si stimano le **Devianze dovute alla regressione** del coefficiente **b** comune; servono solo la devianza totale e quella d'errore, ottenute da:

- **devianza totale dovuta alla regressione**

$$\frac{(\text{Cod. XY totale})^2}{\text{Dev. X totale}} \quad \text{con df } \mathbf{n-1}$$

- e la **devianza d'errore dovuta alla regressione**

$$\frac{(\text{Cod. XY.d' errore})^2}{\text{Dev. X.d' errore}} \quad \text{con df } \mathbf{n-k}$$

5 - E' così possibile ottenere le **devianze delle Y ridotte** o devianze dovute alle deviazioni della regressione, sottraendo alla devianza totale e alla devianza d'errore delle Y, calcolate al punto 1, quelle rispettive calcolate al punto 4.

Si stimano:

- **la devianza delle Y ridotte totale**

$$\text{Dev. Y totale} - \text{Dev. Y totale della regressione} \quad \text{con df } \mathbf{(n-1) - 1}$$

che nell'operazione perde un altro df

- **e la devianza delle Y ridotte d'errore**

$$\text{Dev. Y d'errore} - \text{Dev. Y d'errore della regressione} \quad \text{con df } \mathbf{(n-k)-1}$$

che rispetto alla devianza d'errore delle Y, calcolata al punto 1, ha 1 df in meno.

La **devianza delle Y ridotte tra trattamenti** è ottenuta per differenza tra queste ultime due immediatamente precedenti:

$$\text{Dev. delle Y ridotte totale} - \text{Dev. delle Y ridotte d'errore} \quad \text{con df } \mathbf{k-1}$$

che mantiene i suoi **df = k-1**.

Calcolate la varianza tra trattamenti e la varianza d'errore sulle Y ridotte, **il test F, che considera l'effetto della regressione sui valori delle Y**, è dato dal loro rapporto

$$F_{(k-1, n-k-1)} = \frac{s^2 \text{ tra trattamenti delle Y ridotte}}{s^2 \text{ d'errore delle Y ridotte}}$$

con df **k-1** e **(n-k)-1**.

ESEMPIO. A tre gruppi di cavie sono state somministrate tre sostanze tossiche (A, B, C) che, alterando il metabolismo, determinano un forte aumento ponderale.

Poiché sono stati utilizzati animali di dimensioni diverse, per valutare correttamente gli effetti sul peso (Y) deve essere considerata anche la lunghezza (X) delle cavie.

### TRATTAMENTI

A		B		C	
X	Y	X	Y	X	Y
25	18	15	18	19	16
23	16	12	15	21	19
19	13	17	20	18	18
24	16	11	12	17	15
21	14	19	22	19	17
---	---	16	18	---	---

(Il metodo non richiede che i tre gruppi abbiano lo stesso numero di osservazioni, come nell'analisi della varianza ad 1 criterio di classificazione. Per facilitare il calcolo manuale, pesi ed altezze sono stati riportati in valori trasformati, che non modificano i risultati; inoltre, sempre per facilitare l'esecuzione dei calcoli, sono state scelti campioni molto piccoli).

Effettuare l'analisi della varianza e della covarianza, per valutare compiutamente l'effetto delle tre sostanze sul peso finale delle cavie.

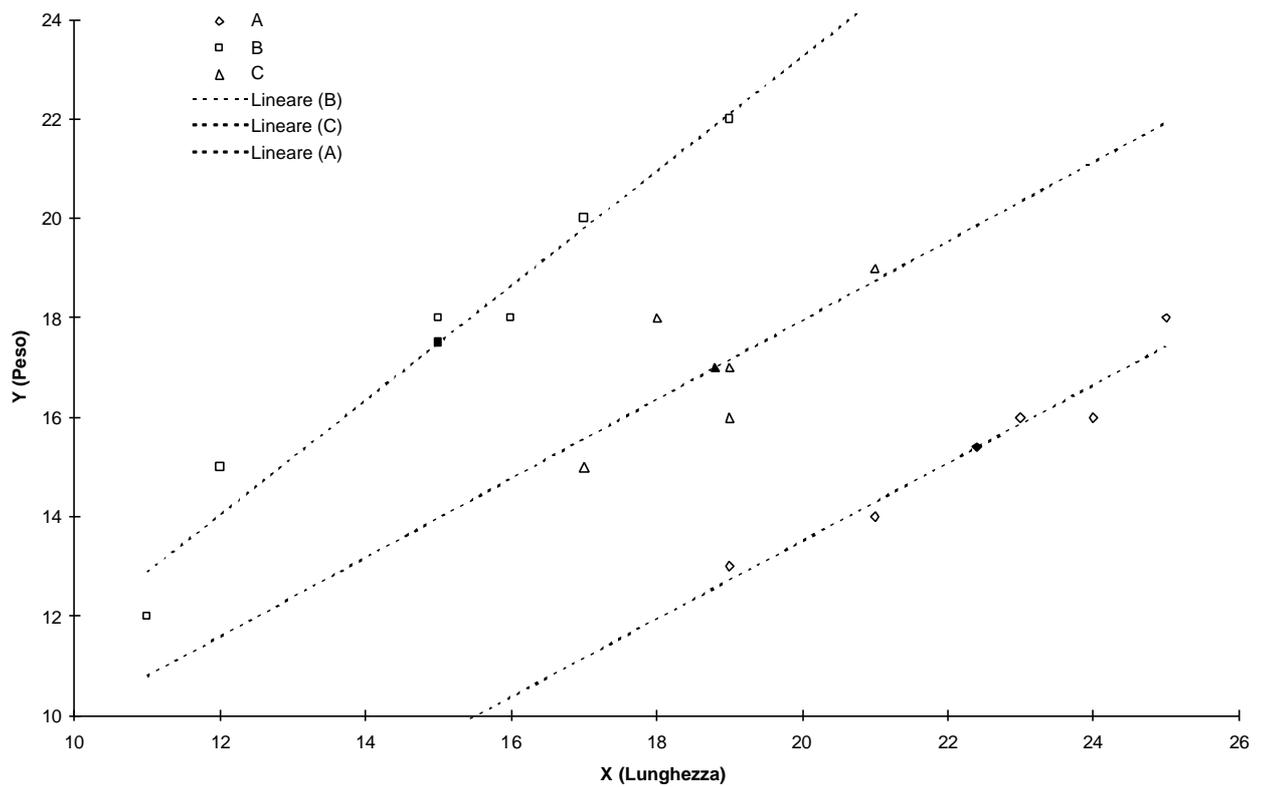
Risposta.

Prima di procedere alle analisi, è sempre di elevata utilità una rappresentazione grafica dei dati e delle medie a confronto.

Il **diagramma di dispersione dei 3 gruppi** (riportato nella pagina seguente) mostra che le differenze tra le tre medie dei valori campionari di Y sono ridotte e che la regressione lineare tra lunghezza X e peso Y per ogni gruppo è evidente, con coefficienti angolari simili.

Per l'interpretazione dei risultati e per i calcoli successivi con le formule abbreviate, è utile disporre delle seguenti somme relative ai dati campionari:

Somma $X_A = 112$	Somma $X_B = 90$	Somma $X_C = 94$	Somma $X = 296$
Somma $Y_A = 77$	Somma $Y_B = 105$	Somma $Y_C = 85$	Somma $Y = 267$
$n_A = 5$	$n_B = 6$	$n_C = 5$	$n = 16$
media $X_A = 22,40$	media $X_B = 15,00$	media $X_C = 18,80$	media $X = 18,50$
media $Y_A = 15,40$	media $Y_B = 17,50$	media $Y_C = 17,00$	media $Y = 16,6875$
Somma $X^2_A = 2532$	Somma $X^2_B = 1396$	Somma $X^2_C = 1776$	Somma $X^2 = 5704$
Somma $Y^2_A = 1201$	Somma $Y^2_B = 1901$	Somma $Y^2_C = 1445$	Somma $Y^2 = 4557$
Somma $XY_A = 1743$	Somma $XY_B = 1628$	Somma $XY_C = 1605$	Somma $XY = 4976$



Seguendo lo stesso schema precedentemente descritto, le analisi da condurre ed i calcoli da effettuare possono essere raggruppati in 5 fasi.

1 - Per l'analisi della varianza ad 1 criterio di classificazione sui valori di Y (peso), si devono stimare i valori di

- la devianza Y totale

$$4557 - \frac{267^2}{16} = 4557 - 4455,56 = 101,56$$

che risulta uguale a 101,56 ed ha 15 df,

- la devianza Y tra trattamenti

$$\frac{77^2}{5} + \frac{105^2}{6} + \frac{85^2}{5} - \frac{267^2}{16} = 1185,8 + 1837,5 + 1445 - 4455,56 = 12,74$$

che è uguale a 12,74 ed ha 2 df

- la devianza Y d'errore

$$101,44 - 12,74 = 88,7$$

che è uguale a 88,7 ed ha 13 df (15-2).

I risultati possono essere presentati in una tabella

	Devianze Y	df	Varianze	F	Prob.
Totale	101,44	15			
Tra tratt.	12,74	2	6,37	<1	---
Errore	88,70	13	6,823		

dalla quale risulta evidente che il valore di F è inferiore a 1; pertanto, le differenze tra le medie campionarie non sono assolutamente significative.

Il risultato può dipendere dal fatto di non avere considerato che i tre gruppi di animali hanno lunghezze differenti; considerando la regressione tra peso ed altezza, le differenze tra il peso medio dei tre campioni di cavie potrebbe risultare significativo.

2 - Per tenere in considerazione l'effetto della regressione, occorre calcolare le codevianze tra X e Y; quindi:

- la **codevianza XY totale**

$$4976 - \frac{296 \cdot 267}{16} = 4976 - 4939,5 = 36,5$$

che risulta uguale a **36,5** con **13 df**,

- la **codevianza XY tra trattamenti**

$$\frac{112 \cdot 77}{5} + \frac{90 \cdot 105}{6} + \frac{94 \cdot 85}{5} - \frac{296 \cdot 267}{16} = 1724,8 + 1575 + 1598 - 4939,5 = -41,7$$

che ha un valore negativo (**- 41,7**) con **df 2**

- la **codevianza XY d'errore**

$$36,5 - (-41,7) = 78,2$$

che risulta maggiore di quella totale (**78,2**) ed ha **13 df** (15 - 2).

3 - Per procedere alle stime richieste è necessario calcolare anche le devianze di X:

- la **devianza X totale**

$$5704 - \frac{296^2}{16} = 5704 - 5476 = 228$$

che è uguale a **228** con **15 df**,

- la **devianza X tra trattamenti**

$$\frac{112^2}{5} + \frac{90^2}{6} + \frac{94^2}{5} - \frac{296^2}{16} = 2508,8 + 1350 + 1767,2 - 5476 = 150$$

risulta uguale a **150** con **df 2**,

- la **devianza X d'errore**

$$228 - 150 = 78$$

uguale a **78** con **13** df (15 -2).

4 - Le devianze dovute alla **regressione b comune**, necessarie alla stima della Y ridotte, sono:

- la **devianza totale della regressione**

$$\frac{36,5^2}{228} = 5,84$$

che risulta uguale a **5,84** con **15** df,

- la **devianza d'errore della regressione**

$$\frac{78,2^2}{78} = 78,4$$

che risulta uguale a **78,4** con **13** df.

5 - In conclusione, **le devianze dovute alle deviazioni dalla regressione o devianze delle Y ridotte** sono:

- la **devianza Y ridotte totale**

$$101,44 - 5,84 = 95,6$$

uguale a **95,6** con **14** df a causa della perdita di un df (16 - 1 - 1),

- la **devianza Y ridotte d'errore**

$$88,7 - 78,4 = 10,3$$

uguale a **10,3** con df **12**,

- e, per differenza, la **devianza Y ridotte tra trattamenti**

$$95,6 - 10,3 = 85,3$$

che risulta uguale a **85,3** con **2** df.

Con questi ultimi dati, è possibile applicare l'analisi della varianza dei valori di Y che considerano l'effetto di regressione sulla X

	Devianze Y ridotte	df	Varianze	F	Prob.
Totale	95,6	14			
Tra tratt.	85,3	2	42,65	49,69	<0.0001
Errore	10,3	12	0,85833		

e permettono un test F

$$F_{(2,12)} = \frac{42,65}{0,85833} = 49,69$$

che risulta altamente significativo.

Il **metodo delle Y aggiustate** è più complesso di quello presentato e comporta una serie più lunga di calcoli. Per gli scopi del presente manuale è quindi inutile la sua presentazione, soprattutto in considerazione del fatto che per i calcoli si richiede l'uso di programmi informatici standard e lo statistico applicato si limita alla interpretazione dei risultati.

L'analisi della covarianza ha avuto ampi sviluppi, in **due direzioni** diverse.

- Da una parte, in analogia con l'analisi della varianza, permette di considerare contemporaneamente vari fattori e le loro interazioni, in riferimento ad una sola covariata. Se l'analisi è limitata a due sole variabili, con la consueta simbologia il modello additivo è

$$Y_{ijk} = \mu + \alpha_i \gamma_j + \alpha \gamma_{ij} + \beta(X_{ijk} - \bar{X}) + \epsilon_{ijk}$$

- Dall'altra, con la covarianza multipla si possono seguire più covariate ( $X_1, X_2, \dots, X_n$ ). Il modello più semplice, due covariate in una analisi della varianza ad 1 solo criterio,

è

$$Y_{ijk} = \mu + \alpha_i + \beta_1(X_{1ij} - \bar{X}_1) + \beta_2(X_{2ij} - \bar{X}_2) + \epsilon_{ijk}$$

Dalla loro combinazione, più variabili e più covariate, si ottengono modelli additivi che appaiono molto complessi. Superati con l'informatica i problemi di calcolo, attualmente il limite alla complessità del disegno sperimentale sono posti solamente dall'interpretazione dei risultati, particolarmente complessa nel caso di più interazioni.

Per la trattazione dell'analisi della covarianza a disegni complessi, più variabili e più covariate con eventuale interazione, si invia a test specifici.

### 13.12. LETTURA DI TRE TABULATI DI PROGRAMMI INFORMATICI

Il tabulato 1) utilizza i dati dell'esempio sulla regressione lineare, in cui la variabile dipendente, che deve sempre essere indicata espressamente, è il peso.

Nella parte inferiore del riquadro, sono riportati i parametri della retta di regressione: l'intercetta ed il coefficiente angolare, con i relativi errori standard.

Nella quinta colonna sono indicati i valori del t di Student, per la verifica dell'ipotesi nulla  $H_0$  che il parametro in oggetto sia significativamente diverso da 0.

La sesta ed ultima colonna, sempre nella parte inferiore a destra del riquadro 1), mostra il valore di probabilità, per un test bilaterale.

Nella parte superiore del riquadro è riportata l'analisi della varianza, con tutti i valori relativi ai parametri indicati. Il valore di F è il quadrato di quello del t di Student ed, ovviamente, le due probabilità coincidono.

Sotto l'analisi della varianza sono riportati altri indicatori utili ad eventuali confronti ed interpretazioni ulteriori dei risultati:

**Root MSE** è la radice quadrata della varianza (Mean Square, sovente tradotto in italiano con quadrato medio);

**Dep mean** è la media della variabile dipendente;

**C. V.** è il coefficiente di variazione (sempre della variabile dipendente);

**R-square** è il valore di  $R^2$ , o R oppure  $r^2$  già trattato nella discussione sul valore predittivo della retta;

**Adj. R-sq** (simboleggiato sovente con  $\bar{R}^2$ ) è il valore di R Adjusted, che considera l'effetto dei gdl ed è calcolato come

$$\bar{R}^2 = 1 - \frac{(1 - R^2) \cdot (n - 1)}{dfe}$$

dove

**n** è il numero di dati,

**dfe** sono i gdl della varianza d'errore.

1)

Dependent Variable: PESO

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	323.20784	323.20784	20.073	0.0065
Error	5	80.50644	16.10129		
Total	6	403.71429			

Root MSE	4.01264	R-square	0.8006
Dep Mean	63.57143	Adj R-sq	0.7607
C.V.	6.31202		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	-73.354062	30.59903924	-2.397	0.0618
ALTEZZA	1	0.796078	0.17768273	4.480	0.0065

Nel riquadro 2) è riportata l'analisi della correlazione. Sono stati utilizzati gli stessi dati dell'esempio precedente, relativi alla regressione lineare tra peso ed altezza in 7 giovani, per facilitare il confronto tra i due risultati. Sovente, i programmi della regressione forniscono analisi delle caratteristiche della distribuzione delle due serie di dati, presentati nel capitolo della statistica descrittiva ed utili alla verifica delle condizioni di validità della correlazione e della regressione, che sono molto più dettagliate di quelle riportate nel riquadro sottostante

2)

Correlation Analysis

1 'WITH' Variables: ALTEZZA  
1 'VAR' Variables: PESO

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
ALTEZZA	7	172.00	9.2195	1204	160	183
PESO	7	63.57	8.2028	445	52	75

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 7

	PESO
ALTEZZA	0.89475 0.0065

I risultati indicati nella parte superiore del riquadro 2) non hanno bisogno di ulteriori spiegazioni. Nella parte inferiore, è riportato il valore di correlazione  $r$  di Pearson e la probabilità relativa alla sua significatività, come verifica dell'ipotesi nulla  $H_0: \rho = 0$

Nei riquadri 3) e 4) sono descritti i risultati dell'analisi della covarianza,

3)

```

General Linear Models Procedure
Class Level Information

Class      Levels      Values
GRUPPO      3      A B C
Number of observations in data set = 16

Dependent Variable: PESO

Source      DF      Sum of Squares      Mean Square      F Value      Pr > F
Model      3      91.1380128      30.3793376      35.40      0.0001
Error      12      10.2994872      0.8582906
Corrected Total 15      101.4375000

R-Square      C.V.      Root MSE      PESO Mean
0.898465      5.551699      0.92644      16.6875

Source      DF      Type I SS      Mean Square      F Value      Pr > F
GRUPPO      2      12.7375000      6.3687500      7.42      0.0080
LUNGHE      1      78.4005128      78.4005128      91.34      0.0001

Source      DF      Type III SS      Mean Square      F Value      Pr > F
GRUPPO      2      85.2948111      42.6474055      49.69      0.0001
LUNGHE      1      78.4005128      78.4005128      91.34      0.0001

```

con i dati dell'esempio sul peso.

Sono stati utilizzati 16 dati campionari, suddivisi in tre gruppi ed indicati con le lettere A, B e C.

Sempre nel riquadro 3) sono riportati i risultati di varie analisi della varianza.

La parte superiore fornisce la varianza d'errore e la parte inferiore le varianze relative ai confronti delle medie dei 3 gruppi (df = 2) con il metodo delle Y ridotte e la stima della significatività della regressione lineare (df = 1).

4)

General Linear Models Procedure

Tukey's Studentized Range (HSD) Test for variable: PESO

NOTE: This test controls the type I experimentwise error rate.

Alpha= 0.05 Confidence= 0.95 df= 12 MSE= 0.858291  
Critical Value of Studentized Range= 3.773

Comparisons significant at the 0.05 level are indicated by '\*\*\*'.

GRUPPO Comparison	Simultaneous	Difference Between Means	Simultaneous	
	Lower Confidence Limit		Upper Confidence Limit	
B - C	-0.9966	0.5000	1.9966	
B - A	0.6034	2.1000	3.5966	***
C - B	-1.9966	-0.5000	0.9966	
C - A	0.0369	1.6000	3.1631	***
A - B	-3.5966	-2.1000	-0.6034	***
A - C	-3.1631	-1.6000	-0.0369	***

Nel riquadro 4) sono riportati i confronti multipli tra le tre medie, con il metodo di Tukey.

Per ogni coppia di medie è riportata la differenza, sia positiva che negativa, con i limiti dell'intervallo di confidenza. La differenza risulta significativa alla probabilità prefissata (nel tabulato uguale a 0.05) quando l'intervallo fiduciale, che ovviamente comprende la differenza media, esclude lo 0.

## CAPITOLO XIV

### MISURE DI TENDENZA NON PARAMETRICA E DI ASSOCIAZIONE TRA VARIABILI

Nella ricerca ambientale, sovente l'interesse è rivolto alla stima di parametri (media, varianza, coefficiente angolare ed intercetta, correlazione, ...). Scopo principale di queste analisi è fornire tutte le informazioni, per descrivere la natura ed il tipo sia della tendenza o valore centrale, sia della variabilità di un insieme di osservazioni. Mentre **per la statistica descrittiva di un campione o dell'intera popolazione non esistono condizioni di validità** od ipotesi che debbano essere rispettate, **nell'inferenza e nella valutazione di parametri di una popolazione il procedimento è ritenuto valido solo quando è possibile:**

- 1 - specificare un modello di distribuzione dei dati,**
- 2 - stimare i modelli di dispersione dei valori in modo efficiente,**
- 3 - raccogliere informazioni campionarie con precisione, utilizzando la medesima scala.**

In termini più tecnici, per l'inferenza e la stima di parametri occorre che

- a) siano rispettate le condizioni di validità sulla distribuzione normale e l'omogeneità delle varianze, eventualmente mediante la trasformazione dei dati,
- b) i valori siano misurati almeno con la precisione di una scala d'intervalli o di rapporti.

Oltre ai casi in cui non sono rispettate le condizioni richieste dai test parametrici, si ricorre ai test non parametrici quando si richiede lo **studio di caratteristiche particolari della distribuzione di una o due variabili**, quali **causalità, tendenza, indipendenza, simmetria, bontà di adattamento, associazione**.

Tra i test già esposti per verificare almeno una di queste caratteristiche, è possibile ricordare

- il test di McNemar, utile per verificare la significatività delle differenze tra le frequenze di due campioni dipendenti,
- il test  $\chi^2$  di Pearson ed il log likelihood ratio (G), per l'indipendenza tra due o più distribuzioni osservate (tabelle 2 x 2 e M x N) o la bontà di adattamento tra una distribuzione osservata ed una attesa.

Quando 2 distribuzioni sono normali ed omoschedastiche, per verificare la significatività di differenze tra le medie, il test **t** di Student e il test **F** di Fisher sono i più potenti. Tra i test non parametrici,

l'efficienza relativa di quelli basati sui ranghi è approssimativamente 0,95; ma appartengono alla statistica non parametrica anche i test di permutazione, detti anche di casualizzazione, che, sempre per confronti sulla media, hanno potenza massima, quindi uguale a 1. Quando le distribuzioni non sono normali, l'efficienza relativa dei test sui ranghi è minore in valore assoluto, ma non è mai inferiore a 0,86 e può essere notevolmente superiore a quella dei test parametrici **t** o **F**, i quali in quelle condizioni hanno una forte perdita di potenza.

I metodi non parametrici sono utilizzati prevalentemente per condurre test di significatività. Possono essere utilizzati anche per la stima di parametri (cioè calcolo della tendenza centrale e della variabilità, con i loro intervalli fiduciali); ma sono necessari procedimenti complessi, che non sono presentati in questa trattazione.

Le condizioni di validità dei test non parametrici non sono mai restrittive; di conseguenza, i risultati sono generali e solo raramente possono essere messi in discussione. **Nella ricerca ambientale, spesso è utile confermare i risultati ottenuti da un test basato sulla teoria della distribuzione normale, mediante un appropriato test non parametrico.** Se i risultati coincidono, vengono confutate le eventuali obiezioni sul mancato rispetto delle condizioni di validità, tanto più legittime quanto più le dimensioni del campione sono ridotte. Inoltre, il confronto tra le due probabilità permette di verificare la robustezza del test parametrico.

In questo contesto, appare logico il consiglio di alcuni statistici, tra cui Armitage:

**“Sei incerto se usare un test parametrico o non-parametrico? Usali entrambi!”.**

Anche per le misure di tendenza e di associazione, sono stati proposti vari test. Nelle discipline sociali ed economiche sono usati il “corner test” di Olmstead e Tukey, il “quadrant test” di Blomqvist, il metodo “quadrant layer ranks” proposto da vari statistici (tra cui gli italiani Pesarin, Fassò e Baldessari), il “test di cograduazione” di Gini. Nel presente corso, l'esposizione è limitata a quelli maggiormente utilizzati in biologia e nelle scienze ambientali.

Tra i test di maggiore utilità nella ricerca ambientale, sono da ricordare **i test di tendenza**. La loro applicazione classica è la verifica, ad esempio, dell'ipotesi che lungo il corso di un fiume si abbia una **tendenza monotona all'aumento (o alla diminuzione)** dei livelli d'inquinamento. La raccolta dei dati può essere fatta, schematicamente, in tre modi diversi, ai quali corrispondono tre test diversi, per verificare la stessa ipotesi:

- il primo caso è quando lungo il corso sono stati fatti oltre una decina di rilevazioni, in punti diversi: si applica il test di **Cox e Stuart**;
- il secondo è quando sono state individuate almeno quattro o cinque stazioni, in ognuna delle quali sono state fatte alcune rilevazioni, come nell'analisi della varianza ad un criterio: si applica il test di **Jonckheere o Jonckheere-Terpstra** ;

- il terzo, quando nella quattro o cinque stazioni prefissate sono state fatte rilevazioni alle stesse date, come nell'analisi della varianza a due criteri di classificazione: si applica il test di **Page**.

#### 14.1. IL TEST DI COX E STUART

Con un campione di **dati registrati in successione temporale** (come il livello medio d'inquinamento atmosferico rilevato giornalmente per un mese) **oppure indicatori d'inquinamento raccolti lungo un gradiente geografico** (come le stazioni collocate lungo il corso di un fiume o campioni raccolti a profondità diverse in un lago), **il test di Cox e Stuart permette di verificare**, per la variabile considerata, che deve essere misurata con una **scala continua per non avere valori identici** (o almeno pochissimi) **se esiste una tendenza monotona all'aumento oppure alla diminuzione**.

Con il loro metodo proposto nel 1955 (*Some quick test for trend in location and dispersion*, pubblicato su *Biometrika*, 42, pp. 80-95), **D. R. Cox e A. Stuart** permettono di **valutare se nel complesso dei dati, tra i valori iniziali e quelli finali, esiste un incremento o una diminuzione sistematica, pure in presenza di ampie irregolarità locali o casuali e di un allontanamento rilevante dalla linearità**. Il test non risulta significativo se i dati hanno una fase d'incremento e una fase altrettanto lunga di decremento o viceversa.

In vari test di statistica, è collocato tra i test per un campione, non assumendo l'altro parametro alcuna importanza, purché i dati siano riportati in serie, secondo l'ordine spaziale o temporale della rilevazione; in altri, è classificato tra i test per dati bivariati, analogo alla regressione lineare semplice o alla regressione curvilinea della statistica parametrica, in quanto utile a verificare la tendenza di  $Y$  o  $X_1$  al variare sistematico di  $X_2$ .

Con  $N$  osservazioni indipendenti

$$X_1, X_2, \dots, X_N$$

in serie ordinata (nel tempo, nello spazio o comunque da un'origine)

l'ipotesi nulla è che non esiste trend

$$H_0: \text{la posizione delle osservazioni da 1 a } N \text{ è costante}$$

contro una delle ipotesi alternative,

$$H'_1: \text{la serie di osservazioni ha valori medi non costanti (ipotesi bilaterale)}$$

oppure

$$H''_1: \text{le osservazioni tendono alla crescita (ipotesi unilaterale crescente)}$$

oppure

$$H'''_1: \text{le osservazioni tendono alla diminuzione (ipotesi unilaterale decrescente)}.$$

Il metodo proposto, estremamente semplice e fondato sulla distribuzione binomiale, richiede alcuni passaggi logici, spiegati con facilità attraverso l'applicazione ad un esempio.

Si supponga di avere la seguente serie di misure (qui arrotondate all'unità), raccolte in successione temporale o geografica

17	16	14	17	21	22	21	23	22	18	10	23
----	----	----	----	----	----	----	----	----	----	----	----

Per applicare e comprendere il test, è utile seguire i seguenti passaggi:

1 - **Individuare l'osservazione centrale** (con  $N = 12$  tra la sesta e la settima misura, quindi tra 22 e 21), **che corrisponde alla mediana** del tempo o dello spazio della serie di rilevazioni;

2 - Se il numero  $N$  d'osservazioni è pari (come nell'esempio), separare il campione in due gruppi, ognuno con  $N/2$  dati,

Prima della mediana	17	16	14	17	21	22
Dopo la mediana	21	23	22	18	10	23

distinguendo la serie di quelli antecedenti la mediana da quelli successivi.

3 - Abbinare il primo valore dei dati che precedono la mediana al primo valore di quelli successivi alla mediana, il secondo del primo gruppo con il secondo del secondo gruppo, fino all'ultimo dato, calcolando le  **$N/2$  differenze, come riportato nella tabella seguente:**

Prima della mediana	17	16	14	17	21	22
Dopo la mediana	21	23	22	18	10	23
<b>Segno della differenza</b>	+	+	+	+	-	+

**si ottengono  $N/2$  segni**, che possono essere **positivi (+) o negativi (-)**, se tra le coppie di dati a confronto si è realizzato un aumento o una diminuzione.

4) Se è vera l'ipotesi nulla  $H_0$ , questa serie di  $N/2$  differenze avrà tendenza centrale nulla (in termini più discorsivi, i segni positivi e i segni negativi tenderanno ad avere la stessa frequenza), mentre avrà tendenza centrale positiva (maggioranza significativa di segni positivi) se la serie di rilevazioni ha valori crescenti e avrà tendenza centrale negativa (maggioranza significativa di segni negativi) se la serie di rilevazioni ha valori decrescenti.

5) Se il numero  $N$  di osservazioni è dispari, come nella serie successiva,

17	16	14	17	21	22	<b>20</b>	21	23	22	18	10	23
----	----	----	----	----	----	-----------	----	----	----	----	----	----

eliminare il valore  $(N + 1)/2$  corrispondente alla mediana del tempo o dello spazio (il valore 20 che occupa la settima posizione su 13 dati); successivamente costruire nello stesso modo del punto 3 i due gruppi, calcolando poi le  $(N-1)/2$  differenze:

Prima della mediana	17	16	14	17	21	22
Dopo la mediana	21	23	22	18	10	23
<b>Segno della differenza</b>	+	+	+	+	-	+

6) Applicare il **test dei segni** alla serie di differenze; utilizzando  
 - la **distribuzione binomiale, nel caso di campioni piccoli,**

$$P_{(r)} = C_N^r p^r q^{N-r}$$

in cui

$N$  è il numero totale di segni

$r$  quello del segno meno frequente

e ricorrendo alla distribuzione cumulata, per ottenere la probabilità complessiva da quella distribuzione a quella più estrema nella stessa direzione,

- la **distribuzione normale per campioni grandi**

$$z = \frac{x - n \cdot p}{\sqrt{n \cdot p \cdot q}}$$

secondo le modalità già illustrate nel paragrafo del test dei segni per un campione.

Poiché **l'analisi è condotta sulle  $N/2$  differenze**, la serie di rilevazioni del campione deve essere il doppio di quanto necessario alla significatività della distribuzione binomiale: di conseguenza, **serve almeno una serie di 12 osservazioni (6 differenze)** per raggiungere la significatività, come è possibile dedurre dalla semplice lettura della tabella della distribuzione binomiale cumulata.

Nel caso in cui una **differenza risulti uguale a 0**, si perde un'osservazione, in quanto essa non può essere conteggiata né tra **N** né tra **r**. E' il motivo per cui (come già evidenziato) si richiede che la scala, con la quale i valori o punteggi sono stimati, sia continua: i valori identici dovrebbero essere rari.

### PROBABILITA' CUMULATE DELLA DISTRIBUZIONE BINOMIALE

$$C_N^r \cdot p^r \cdot q^{N-r}$$

N = numero di osservazioni

r = numero minore tra segni positivi e negativi

**N**

<b>r</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>
<b>0</b>	0.016	0.008	0.004	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<b>1</b>	0.109	0.062	0.035	0.020	0.011	0.006	0.003	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000
<b>2</b>	0.344	0.227	0.144	0.090	0.055	0.033	0.019	0.011	0.006	0.004	0.002	0.001	0.001	0.000	0.000
<b>3</b>	0.656	0.500	0.363	0.254	0.172	0.113	0.073	0.046	0.029	0.018	0.011	0.006	0.004	0.002	0.001
<b>4</b>	0.891	0.773	0.637	0.500	0.377	0.274	0.194	0.133	0.090	0.059	0.038	0.024	0.015	0.010	0.006
<b>5</b>	0.984	0.938	0.856	0.746	0.623	0.500	0.387	0.291	0.212	0.151	0.105	0.072	0.048	0.032	0.021
<b>6</b>	1.00	0.992	0.965	0.910	0.828	0.726	0.613	0.500	0.395	0.304	0.227	0.166	0.119	0.084	0.058
<b>7</b>		1.00	0.998	0.980	0.945	0.887	0.806	0.710	0.605	0.500	0.402	0.314	0.240	0.180	0.132
<b>8</b>			1.00	0.996	0.989	0.967	0.927	0.867	0.788	0.696	0.598	0.500	0.407	0.324	0.252
<b>9</b>				1.00	0.999	0.994	0.981	0.954	0.910	0.849	0.773	0.686	0.593	0.500	0.412
<b>10</b>					1.00	0.999	0.997	0.989	0.971	0.941	0.895	0.834	0.760	0.676	0.588
<b>11</b>						1.00	1.00	0.998	0.994	0.982	0.962	0.928	0.881	0.820	0.748
<b>12</b>							1.00	1.00	0.999	0.996	0.989	0.976	0.952	0.916	0.868
<b>13</b>								1.00	1.00	1.00	0.998	0.994	0.985	0.968	0.942
<b>14</b>									1.00	1.00	1.00	0.999	0.996	0.990	0.979
<b>15</b>										1.00	1.00	1.00	0.999	0.998	0.994
<b>16</b>											1.00	1.00	1.00	1.00	0.999
<b>17</b>												1.00	1.00	1.00	1.00
<b>18</b>													1.00	1.00	1.00
<b>19</b>														1.00	1.00
<b>20</b>															1.00

Se la serie dei segni è composta da 2 sole successioni di lunghezza equivalente, come la seguente

+ + + + + - - - -

formata da 5 + e 4 -,

il test ovviamente non risulterà significativo. Infatti è logico dedurre che dopo una fase di crescita si è avuta una fase di contrazione, ma non si è realizzata una tendenza monotona solo crescente o solo decrescente, come il test di Cox e Stuart permette di verificare.

ESEMPIO. In una stazione di rilevazione dell'inquinamento, collocata in un corso d'acqua dopo un centro abitato, durante due mesi di siccità sono stati effettuati 20 prelievi, per verificare l'ipotesi di un aumento dell'inquinamento d'origine organica nel tempo.

I valori che possono essere

- sia indicatori complessi o punteggi della presenza di sostanze inquinanti, quindi una scala di rango
  - sia misure di concentrazione per litro, su una scala di rapporti
- sono stati

Tempo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Indice	12	13	15	22	16	13	17	17	18	15	14	15	16	13	18	18	18	20	19	19

Con questi dati campionari, si può sostenere che esiste un aumento significativo nella concentrazione delle sostanze inquinanti, durante il periodo in oggetto?

Risposta.

E' un'ipotesi unilaterale crescente, con

$H_0$ : la posizione delle osservazioni da 1 a N è costante

e

$H_1$ : la serie dei valori dimostra un aumento

La stima della probabilità può essere ottenuta con alcuni passaggi:

1 - Separare la serie complessiva in

- valori prima della mediana e
- valori dopo la mediana,

mantenendo per ogni gruppo lo stesso ordine della rilevazione o serie iniziale

< Mediana	12	13	15	22	16	13	17	17	18	15
> Mediana	14	15	16	13	18	18	18	20	19	19
Segno della differenza	+	+	+	-	+	+	+	+	+	+

2 - Benché i valori siano approssimati, in tutti i casi è possibile

calcolare il segno delle differenze: risultano 9+ e 1-

3 - Se fosse vera l'ipotesi nulla  $H_0$ , la probabilità di trovare per caso questa risposta e risposte più estreme è data dalla somma delle relative distribuzioni binomiali; in questo caso

$$P_{(9)} = C_{10}^9 0,5^9 0,5^1$$

e

$$P_{(10)} = C_{10}^0 0,5^{10} 0,5^0$$

Se la somma delle probabilità  $P_{(9)} + P_{(10)}$  risulta inferiore a 0.05 è dimostrata una tendenza significativa all'aumento della variabile X.

Senza perdere tempo nel calcolo della probabilità, **la tabella dei valori cumulati**

per  $N = 10$  e  $r = 1$  fornisce una probabilità  $\alpha = 0.011$ ,

che permette di rifiutare l'ipotesi nulla ed affermare che esiste una tendenza significativa all'aumento dei valori.

Se il **test** fosse stato **bilaterale**, sarebbe stato necessario moltiplicare questa probabilità per due: si sarebbe ottenuta una probabilità  $\alpha$  uguale a 0.022 che avrebbe ugualmente permesso di affermare che esiste una variazione monotona significativa (senza indicarne la direzione).

#### 14.2. TEST DI JONCKHEERE O JONCKHEERE-TERPSTRA

##### PER ALTERNATIVE ORDINATE IN K CAMPIONI INDIPENDENTI

Nel caso di **k campioni indipendenti**, come nell'analisi della varianza ad un criterio di classificazione, quando si suppone che siano ordinati secondo il valore delle loro mediane (non importa se in modo crescente o decrescente), con il **test di Jonckheere** è possibile **verificare l'ipotesi se i vari campioni o gruppi abbiano tendenze centrali in accordo con la sequenza fissata a priori**.

E' chiamato anche test di **Jonckheere-Terpstra o delle alternative ordinate**, in quanto proposto quasi contemporaneamente ed in modo indipendente da T. J. Terpstra nel 1952 (nell'articolo *The asymptotic normality and consistency of Kendall's test against trend when ties are present in one ranking* pubblicato su *Indag. Math.* 14, pp. 327-333) e da A. R. Jonckheere nel 1954, con un articolo su *Biometrika* (vol. 41, pp. 133-145) intitolato "A distribution-free k-sample test against ordered alternative".

Un altro test che verifica la stessa ipotesi, con modalità differenti e che non ha avuto lo stesso successo di applicazioni, è stato proposto nel 1984 da **Kepner e Robinson**, con un articolo intitolato "A distribution free rank test for ordered alternatives in randomized complete block designs" sulla rivista "Journal Amer. Statist. Ass."

**L'ipotesi nulla  $H_0$  è che le mediane (me) dei k gruppi a confronto siano uguali,**

$$H_0: me_1 = me_2 = me_3 = \dots = me_k$$

mentre l'**ipotesi alternativa**  $H_1$  è che esse siano allineate in base al valore, con un ordine fissato a priori, che può essere il seguente

$$H_1: me_1 \leq me_2 \leq me_3 \leq \dots \leq me_k$$

oppure quello in direzione opposta.

Se si rifiuta l'ipotesi nulla, si accetta che tra i valori centrali dei gruppi a confronto **esista almeno una differenza**.

**Il test richiede che i campioni siano misurati con una scala ordinale o più precisa, come quelle di intervallo oppure di rapporti, in modo da non avere valori con lo stesso rango.**

Per una presentazione semplice dei dati e per facilitare l'interpretazione dei risultati, di norma i valori vengono disposti in una tabella identica a quella dell'analisi della varianza ad un solo criterio di classificazione:

GRUPPI						
1	2	3	...	j	...	k
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	...	$X_{1,j}$	...	$X_{1,k}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	...	$X_{2,j}$	...	$X_{2,k}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	...	$X_{3,j}$	...	$X_{3,k}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$X_{i,1}$	$X_{i,2}$	$X_{i,3}$	...	$X_{i,j}$	...	$X_{i,k}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$X_{n_1,1}$	$X_{n_2,2}$	$X_{n_3,3}$	...	$X_{n_j,j}$	...	$X_{n_k,k}$

**con la quale condivide l'ipotesi nulla.**

Nella tabella precedente, i **k campioni a confronto sono disposti in ordine progressivo**, in accordo con l'ipotesi a priori formulata sui valori delle mediane rispettive.

Il test **utilizza il metodo delle precedenze**; è un **procedimento lungo**, quando si confrontano più gruppi, poiché analizza tutte le loro combinazioni semplici 2 a 2.

Per ogni coppia di gruppi, si deve contare il numero di precedenze, ossia quante volte ogni osservazione dell'i-esimo gruppo precede le osservazioni del gruppo j-esimo, secondo il procedimento proposto da Mann e Whitney con il test U.

Nel caso di **due valori uguali (ties)**, il punteggio delle precedenze deve essere **aumentato di 0,5**.

L'applicazione del metodo ad un esempio permette di fornire una spiegazione semplice e chiara.

Si supponga di dover analizzare 3 gruppi (A, B e C), composti rispettivamente 3, 4 e 3 osservazioni, come nella tabella sottostante,

1 - **in cui**, come primo passo, **i valori devono essere ordinati in modo crescente entro ogni gruppo**:

Gruppo A	Gruppo B	Gruppo C
7,5	9,1	10,1
8,9	9,9	15,4
12,3	14,3	16,2
	18,2	

Con 3 gruppi (A, B, e C) i confronti possibili sono 3 (A-B, A-C, B-C).

2 – Successivamente occorre **riportare il numero di precedenze relativo ad ogni dato**

	A-B	A-C	B-C
	4	3	3
	4	3	3
	2	2	2
			0
Somme	10	8	8

di un gruppo rispetto a tutti i dati del gruppo a confronto;

nella tabella precedente,

- il primo 4 del confronto A-B indica che il valore 7,5 del gruppo A precede nell'ordine naturale i 4 valori di B;
- il valore 2 (terza posizione) del confronto A-B indica che il valore 12,3 del gruppo A precede nell'ordine naturale 2 valori di B (14,3 e 18,2);

- mentre nel confronto B-C il quarto numero è 0, perché il valore 18,2 del gruppo B non precede nessun valore del gruppo C (essendo maggiore di tutti i dati del gruppo C).

3- Il terzo passo è **la somma di tutte le precedenze**, che fornisce la stima dell'**indice J**.

Con i dati dell'esempio, J risulta uguale a 26 (10 + 8 + 8).

4- **Quando è vera l'ipotesi nulla, il valore di J tende ad essere uguale ad un valore medio  $m_J$** , determinato

- dal numero totale **N** e
- dal numero  **$n_i$**  di dati
- in ognuno dei **k** gruppi,

secondo la relazione

$$m_J = (N^2 - \sum_{i=1}^k n_i^2) / 4$$

Per **piccoli campioni**, la distribuzione campionaria di **J** è fornita da tabelle, riportate nelle 3 pagine seguenti.

Nella prima colonna delle tabelle è riportato il numero di dati che formano i **k** campioni a confronto. L'ordine delle loro dimensioni è ininfluente.

Per esempio, con 3 gruppi in cui il primo ha 6 osservazioni, il secondo 8 ed il terzo 6, si usa la terza riga della terza tabella, che riporta i valori di **J** per **k** uguali a 6, 6, 8. (85, 90, 99, 103).

Si rifiuta l'ipotesi nulla alla probabilità prefissata, quando il valore calcolato è uguale o superiore ( $\geq$ ) a quello riportato nelle tabelle.

Ritornando ai dati dell'esempio con 3, 3 e 4 dati, il valore critico

- alla probabilità  $\alpha = 0.05$  è **26** e
- alla probabilità  $\alpha = 0.01$  è **29**.

Il valore calcolato, uguale a 26, permette di rifiutare l'ipotesi nulla alla probabilità  $\alpha = 0.05$ .

**Valori critici di J nel test di Jonckheere**  
per 3 gruppi con osservazioni da 3 a 8  
(PRIMA PARTE)

Dimensioni del campione	a			
	0.10	0.05	0.01	0.005
3 3 3	20	22	24	25
3 3 4	24	26	29	30
3 3 5	28	30	33	35
3 3 6	32	34	38	40
3 3 7	36	38	42	44
3 3 8	40	42	47	49
3 4 4	29	31	34	36
3 4 5	33	35	39	41
3 4 6	38	40	44	46
3 4 7	42	45	49	52
3 4 8	47	50	55	57
3 5 5	38	41	45	47
3 5 6	43	46	51	53
3 5 7	48	51	57	59
3 5 8	53	57	63	65
3 6 6	49	52	57	60
3 6 7	54	58	64	67
3 6 8	60	64	70	73
3 7 7	61	64	71	74
3 7 8	67	71	78	81
3 8 8	74	78	86	89

**Valori critici di J nel test di Jonckheere**  
per 3 gruppi con osservazioni da 4 a 8 e da 5 a 8  
(SECONDA PARTE)

Dimensioni del campione	a			
	0.10	0.05	0.01	0.005
4 4 4	34	36	40	42
4 4 5	39	41	45	48
4 4 6	44	47	51	54
4 4 7	49	52	57	60
4 4 8	54	57	63	66
4 5 5	44	47	52	55
4 5 6	50	53	58	61
4 5 7	56	59	65	68
4 5 8	61	65	71	75
4 6 6	56	60	66	69
4 6 7	62	66	73	76
4 6 8	68	73	80	83
4 7 7	69	73	81	84
4 7 8	76	80	88	92
4 8 8	83	88	97	100
5 5 5	50	54	59	62
5 5 6	57	60	66	69
5 5 7	63	67	73	76
5 5 8	69	73	80	84
5 6 6	63	67	74	77
5 6 7	70	74	82	85
5 6 8	77	81	89	93
5 7 7	77	82	90	94
5 7 8	85	89	98	102
5 8 8	92	98	107	111

**Valori critici di J nel test di Jonckheere**

da 3 a 6 gruppi con osservazioni da 2 a 5

(TERZA PARTE)

Dimensioni Del campione	a			
	0.10	0.05	0.01	0.005
6 6 6	71	75	82	86
6 6 7	78	82	91	94
6 6 8	85	90	99	103
6 7 7	86	91	100	103
6 7 8	94	99	109	113
6 8 8	102	108	118	122
7 7 7	94	99	109	113
7 7 8	102	108	119	123
7 8 8	111	117	129	133
8 8 8	121	127	139	144
2 2 2 2	18	19	21	22
2 2 2 2 2	28	30	33	34
2 2 2 2 2 2	40	43	46	49
3 3 3 3	37	39	43	45
3 3 3 3 3	58	62	68	70
3 3 3 3 3 3	85	89	97	101
4 4 4 4	63	66	72	76
4 4 4 4 4	100	105	115	119
4 4 4 4 4 4	146	153	166	171
5 5 5 5	95	100	109	113
5 5 5 5 5	152	159	173	178
5 5 5 5 5 5	223	233	251	258

Per **campioni grandi**, la distribuzione campionaria di **J** è approssimativamente normale, con la media  $\mu_J$  già presentata

$$m_j = (N^2 - \sum_{i=1}^k n_i^2) / 4$$

e varianza  $\sigma_J^2$  uguale a

$$s_j^2 = \frac{N^2 \cdot (2N + 3) - \sum_{i=1}^k (n_i^2 \cdot (2n_i + 3))}{72}$$

dove

**N** è il numero complessivo di dati raccolti

$n_j$  è il numero di dati del j-esimo gruppo.

Poiché le alternative sono ordinate, **il test è sempre a una coda**.

#### ESEMPIO 1 (PER GRANDI CAMPIONI).

Si intende verificare se le aree più vicine ad un forno inceneritore sono maggiormente esposte alla ricaduta di polveri e di sostanze inquinanti, con un gradiente asintotico, nonostante la variabilità nella direzione e nell'intensità dei venti. A questo scopo, è stata misurata la quantità di sostanze inquinanti ricadute, in rapporto alla distanza dalla sorgente di emissione.

(La quantità può essere espressa come misura relativa, il rapporto rispetto alla quantità totale emessa, oppure come misura assoluta, cioè i milligrammi per unità di tempo d'esposizione in rapporto alla superficie unitaria. Trattandosi di un test non parametrico possono essere utilizzate sia misure ordinali sia una distribuzione di valori con la presenza di dati anomali, che la rendono fortemente asimmetrica, oppure dati in cui le varianze dei gruppi a confronto siano significativamente differenti).

L'area è stata suddivisa in 4 zone (A, B, C, D) concentriche, di raggio progressivamente maggiore passando dal gruppo A al gruppo D.

Per abbreviare e semplificare i calcoli,

- **i gruppi sono già stati ordinati** secondo l'ordine crescente atteso, espresso nell'ipotesi  $H_1$

$$H_1 : me_D \leq me_C \leq me_B \leq me_A$$

- i dati della quantità d'inquinamento raccolti in ogni zona sono già ordinati in ordine crescente entro ogni gruppo, come nella tabella seguente

ZONE				
	D	C	B	A
	12	28	31	35
	15	30	36	40
	18	38	39	52
	20	48	44	67
	38	60	54	78
	47	66	57	83
	48	70	63	88
	51	71	77	101
	90		87	119
	108		123	
			124	
$N_j$	<b>10</b>	<b>8</b>	<b>11</b>	<b>9</b>
	<b>N = 38</b>			

Per i calcoli successivi, servono

- i 4 valori di  $n_j$  (10, 8, 11, 9), corrispondenti al numero di osservazioni entro ogni gruppo e
- il loro numero totale  $N = 38$

Risposta.

Per verificare se la quantità di polveri ricadute è significativamente maggiore nelle aree presso la fonte d'emissione, l'**ipotesi nulla**  $H_0$  è che all'aumento della distanza si abbia una quantità ( $q$ ) costante

$$H_0 : q_A = q_B = q_C = q_D$$

mentre l'**ipotesi alternativa**  $H_1$  è che campioni raccolti in aree più distanti dall'inceneritore abbiano **quantità ( $q$ ) progressivamente minori** di polveri, anche se non è quantificabile il tipo di relazione (lineare, quadratica, ecc. ...)

$$H_1 : q_D \neq q_C \neq q_B \neq q_A$$

Il primo passo dell'analisi richiede il calcolo delle precedenze, confrontando tutte le possibili coppie di zone campionate; con 4 gruppi sono 6

D>C   D>B   D>A   C>B   C>A   B>A.

come riportate nella tabella sottostante

CALCOLO DELLE PRECEDENZE						
	D>C	D>B	D>A	C>B	C>A	B>A
	8	11	9	11	9	9
	8	11	9	11	9	8
	8	11	9	9	8	8
	8	11	9	7	7	7
	5,5	9	8	5	6	6
	5	7	7	4	6	6
	4,5	7	7	4	5	6
	4	7	7	4	5	5
	0	2	2			3
	0	2	1			0
						0
$U_{ij}$	<b>51</b>	<b>78</b>	<b>68</b>	<b>55</b>	<b>55</b>	<b>58</b>

Nella prima colonna, il confronto **D>C** riporta 4 volte il punteggio 8: i primi 4 valori del gruppo **D** sono minori degli 8 valori del gruppo **C**. Il punteggio 5,5 deriva dal fatto che il valore 38 del gruppo **D** precede 5 valori del gruppo **C** (48, 60, 66, 70,71) ed è appaiato dal valore 38 del gruppo **C**.

Anche da questo approccio,

- è evidente che i valori rilevati devono essere misurati in una scala continua per ridurre i dati appaiati;
- **nel caso di valori uguali, il punteggio aumenta di 0,5** dopo aver contato quanti dati sono superiori a quello di confronto.

Il valore di **J** è dato dalla somma di tutte le precedenze

$$J = 51 + 78 + 68 + 55 + 55 + 58 = 365$$

e risulta uguale a 365.

Se l'ipotesi nulla fosse vera, questo valore dovrebbe tendere alla media  $\mu_j$

$$\mathbf{m}_j = \frac{N^2 - \sum n_j^2}{4} = \frac{38^2 - (10^2 + 8^2 + 11^2 + 9^2)}{4} = \frac{1444 - (100 + 64 + 121 + 81)}{4} = \frac{1078}{4} = 269,5$$

e con i dati del campione essere uguale a 269,5

mentre la sua varianza  $\sigma_j^2$  è

$$\begin{aligned} \sigma_j^2 &= \frac{N^2 \cdot (2N + 3) - \sum n_j^2 \cdot (2n_j + 3)}{72} = \\ &= \frac{38^2 \cdot (76 + 3) - (10^2 \cdot (20 + 3) + 8^2 \cdot (16 + 3) + 11^2 \cdot (22 + 3) + 9^2 \cdot (18 + 3))}{72} = \\ &= \frac{114076 - (2300 + 1216 + 3025 + 1701)}{72} = \frac{105834}{72} = \mathbf{1469,9} \end{aligned}$$

uguale a 1469,9

e la deviazione standard  $\sigma_j$

$$\sigma_j = \sqrt{1469,9} = \mathbf{38,34}$$

risulta uguale a 38,34.

Da questi dati si calcola il valore di  $\mathbf{z}$

$$z = \frac{J - \mathbf{m}_j}{\mathbf{s}_j} = \frac{365 - 269,5}{38,34} = \frac{95,5}{38,34} = 2,49$$

che risulta uguale a 2,49.

Ricordando che **nella distribuzione normale per un test unilaterale al valore di z uguale a 2,49 corrisponde una probabilità P = 0.006**, si rifiuta l'ipotesi nulla.

Di conseguenza, si accetta l'ipotesi alternativa che la tendenza centrale dei 4 gruppi tende a decrescere allontanandosi dalla fonte d'inquinamento. Tra le 4 aree esiste almeno una differenza nella quantità d'inquinamento, in accordo con l'ordine prefissato nell'ipotesi alterativa  $H_1$ .

**Il test di Jonckheere è utile soprattutto nel caso del confronto tra rapporti o percentuali, quando calcolati su dimensioni di campioni sensibilmente differenti.** Con tali dati infatti non è possibile applicare i test parametrici, poiché i valori a confronto non hanno tutti la stessa attendibilità.

La trasformazione delle percentuali mediante arcoseno (vedi paragrafo sulle trasformazioni) serve solamente per rendere omogenee le varianze di gruppi che hanno medie differenti; ma per test parametrici le percentuali a confronto debbono essere calcolate su totali simili, come già ripetutamente ricordato in vari esempi.

#### ESEMPIO 2 (PER PICCOLI CAMPIONI).

Nei test di tossicità cronica sono usate varie specie acquatiche, in particolare Ciprinidi, Salmoidi, Cladoceri e Ditteri. Gli animali vengono esposti a sostanze tossiche per tempi generalmente lunghi, al fine di consentire che siano sottoposti all'azione del principio attivo non solo gli adulti, ma anche gli stadi larvali e giovanili.

Tra le stime di tossicità, è frequente l'impiego

- del rapporto tra il numero di individui sopravvissuti sul totale degli immessi,
- la percentuale di femmine che arrivano alla riproduzione,
- la quota di uova che si schiudono.

Con una serie di esperimenti, è stato calcolato il numero di uova rimaste sterili su quelle deposte, per stimare l'effetto di una sostanza ritenuta tossica sulla loro schiusa.

DOSE		
1	2	3
9/84	100/108	8/52
4/35	12/91	7/38
3/22	9/57	8/41
5/76	9/125	3/41
5/29	12/64	15/91
2/15		11/86
11/87		

Si intende verificare se all'aumento della dose (1, 2, 3) si ha una quota crescente di uova sterili (misurate con  $x/y$ , dove  $x$  è il numero di uova sterili e  $y$  è il numero totale di uova deposte).

Risposta.

Per verificare l'ipotesi nulla

$$H_0 : q_1 = q_2 = q_3$$

mentre l'ipotesi alternativa  $H_1$

$$H_1 : q_1 \neq q_2 \neq q_3$$

1- la prima operazione consiste nel calcolare i rapporti e riportare i loro valori

DOSE		
1	2	3
0,107	0,093	0,153
0,114	0,132	0,184
0,136	0,158	0,195
0,066	0,072	0,073
0,172	0,187	0,165
0,133		0,128
0,126		

2 - Successivamente si devono ordinare i valori entro ogni gruppo in modo crescente, per facilitare le successive operazioni di confronto tra i gruppi

DOSE		
1	2	3
0,066	0,072	0,073
0,107	0,093	0,128
0,114	0,132	0,153
0,126	0,158	0,165
0,133	0,187	0,184
0,136		0,195
0,172		

3 - Con 3 gruppi, si hanno 3 combinazioni di confronti a coppie:

1 contro (>) 2,

1 contro (>) 3,

2 contro (>) 3.

Per ognuna di esse si devono contare le precedenze, riportate, con la somma per colonna e totale, nella tabella sottostante

	1>2	1>3	2>3
	5	6	6
	3	5	5
	3	5	4
	3	5	3
	2	4	1
	2	4	
	1	2	
R <sub>i</sub>	19	31	19

4 - Il valore di **J**, dato dalla somma di tutte le precedenze,

$$\mathbf{J = 19 + 31 + 19 = 69}$$

risulta uguale a 69.

Il valore di **J** tabulato per 3 gruppi con 5, 6 e 7 osservazioni alla probabilità  $\alpha = 0.05$  è 74.

Il valore di **J** calcolato (69) è inferiore a quello tabulato; di conseguenza, non si può rifiutare l'ipotesi nulla. All'aumentare della dose di sostanza inquinante, con questi dati non è dimostrato che si abbia un aumento significativo della quota di uova sterili.

In letteratura si trovano **applicazioni del test di Jonckheere-Terpstra**

- **a campioni estremamente piccoli,**
- **con gruppi in cui compare addirittura una sola osservazione,**

contando sul fatto che il test ha una potenza elevata, superiore a quelli che si fondano sui ranghi.

Ad esempio, un articolo pubblicato nel 1989 su Applied Statistics (vol. 38, pp. 495-506) di D. V. Hinkley (*Modified profile likelihood in trasformed linear models*), per valutare l'aumento delle rotture meccaniche alla crescita della velocità (20, 25, 30 35 miglia orarie), presenta l'analisi dei seguenti dati

Velocità (mph)			
20	25	30	35
48	33	60	85
	59	101	107
	48	67	
	56		

in cui si hanno solo 10 osservazioni, suddivise in 4 gruppi, dei quali uno con una sola osservazione. Il caso risulta sufficiente per rifiutare l'ipotesi nulla al livello di significatività uguale a 0,11%.

### 14.3. IL TEST DI PAGE PER ALTERNATIVE ORDINATE

Come già discusso all'inizio della presentazione dei test sulla tendenza, il test di Cox e Stuart, il test di Jonckheere-Terpstra e il test di Page servono per analizzare lo stesso problema, ma con un disegno sperimentale diverso:

- hanno l'ipotesi nulla e l'ipotesi alternativa simili,
- ma sono applicati in condizioni sperimentali differenti.

Si ricorre al test di **Cox e Stuart**, quando si dispone di una serie di osservazioni, di solito raccolte nel tempo in modo casuale oppure quando si tratta di rilevazioni effettuate lungo un fiume, per verificare l'aumento del livello d'inquinamento.

Occorre scegliere il test di **Jonckheere-Terpstra**, quando per il campionamento sono state scelte poche stazioni e le rilevazioni sono state ripetute senza una programmazione dettagliata, eventualmente in date diverse o in condizioni differenti; di conseguenza, il numero di osservazioni per stazioni che può essere diverso. I dati sono impostati in una tabella uguale a quella dell'analisi della varianza ad un criterio di classificazione, con la quale condivide l'ipotesi nulla.

Il **test di Page** serve quando le rilevazioni sono condotte in modo programmato, come possono essere 5 misure d'inquinamento condotte nelle stazioni prefissate durante le stesse giornate. E' il caso di **k campioni dipendenti** e i dati sono impostati in una tabella a doppia entrata, come nell'analisi della varianza a due criteri di classificazione.

Con il test, detto anche per alternative ordinate (*distribution-free test for ordered alternatives based on Friedman rank sums*), proposto da **E. B. Page nel 1963** con l'articolo intitolato "*Ordered hypotheses for multiple treatments: a significance test for linear ranks*" pubblicato dalla rivista "*Journal Amer. Statist. Ass.*" è possibile verificare l'ipotesi che **le mediane dei gruppi a confronto siano ordinate secondo una sequenza specificata a priori**.

**Il test è solo unilaterale e quindi occorre definire l'ordine naturale dei valori delle mediane.**

Anche il test di Page serve per verificare l'ipotesi nulla

$$H_0: me_1 = me_2 = me_3 = \dots = me_K$$

contro l'ipotesi alternativa unilaterale

$$H_1: me_1 \leq me_2 \leq me_3 \leq \dots \leq me_K$$

o con un ordine delle mediane opposto.

Accettare l'ipotesi alternativa  $H_1$  significa che **esiste almeno una disuguaglianza (<) valida in senso stretto**. La **procedura è simile a quella già applicata nel test di Friedman** (del quale condivide anche l'ipotesi nulla) ed è **fondata sulla somma dei ranghi di ogni gruppo**. La scala utilizzata per determinare i valori raccolti con il campione deve essere almeno ordinale.

I dati sono utilmente rappresentati in una tabella a doppia entrata, con **i trattamenti (sui quali è fondata l'ipotesi nulla)** riportati in colonna e i blocchi nelle righe.

		T R A T T A M E N T I						
		1	2	3	...	j	...	k
B L O C C H I	B	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	...	$X_{1,j}$	...	$X_{1,k}$
	L	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	...	$X_{2,j}$	...	$X_{2,k}$
	O	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	...	$X_{3,j}$	...	$X_{3,k}$
	C	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	C	$X_{i,1}$	$X_{i,2}$	$X_{i,3}$	...	$X_{i,j}$	...	$X_{i,k}$
	H	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	I	$X_{m,1}$	$X_{m,2}$	$X_{m,3}$	...	$X_{m,j}$	...	$X_{m,k}$

Ovviamente,

- tutti i trattamenti devono avere lo stesso numero di osservazioni;
- entro la stessa riga, i dati sono riferiti allo stesso individuo o alla medesima situazione.

Per spiegare la metodologia, è utile ricorrere ad un esempio. Si supponga di avere voluto **sperimentare gli effetti di una sostanza tossica sulla crescita di una specie vegetale**. A questo

scopo, sono stati collocati i semi in 5 colture con concentrazione crescente (36, 54, 72, 108, 144) del principio attivo e il test è stato ripetuto da 3 ricercatori (A, B, C).

Nella tabella sottostante sono riportati i valori medi delle 4 repliche fatte da ogni ricercatore per ogni concentrazione (l'uso di valori medi, vieta l'uso della statistica parametrica, se non si conoscono le varianze relative):

Ricercatore	CONCENTRAZIONE (g/l)				
	36	54	72	108	144
A	7,63	8,14	7,76	7,17	7,46
B	8,00	8,15	7,73	7,57	7,68
C	7,93	7,87	7,74	7,80	7,21

E' possibile affermare che l'aumento della concentrazione della sostanza tossica (da 36 a 144) inibisce la crescita della specie?

**La metodologia inferenziale del test di Page** richiede alcuni passaggi.

1 – Dapprima formulare l'ipotesi nulla

$$H_0: me_{36} = me_{54} = me_{72} = me_{108} = me_{144}$$

secondo la quale le mediana della crescita alle 5 diverse concentrazioni sono uguali

e l'ipotesi alternativa unilaterale

$$H_1: me_{144} \leq me_{108} \leq me_{72} \leq me_{54} \leq me_{36}$$

secondo la quale la crescita della specie aumenta con il diminuire della concentrazione.

**2 – L'ipotesi alternativa definisce l'ordine con il quale devono essere riportati i dati nella tabella.**

Con i dati dell'esempio, è necessario invertire l'ordine delle colonne dati (i trattamenti a confronto), che diventa

	CONCENTRAZIONE (g/l)				
Ricercatore	144	108	72	54	36
A	7,46	7,17	7,76	8,14	7,63
B	7,68	7,57	7,73	8,15	8,00
C	7,21	7,80	7,74	7,87	7,93

3 – Come nel test di Friedman, i valori devono essere **trasformati nel loro rango, entro la stessa riga**. Si assegna il rango minore (1) al punteggio più basso e si aumenta di 1 per ogni punteggio successivo della stessa riga fino a **k**, uguale al numero di trattamenti o gruppi a confronto. Nel caso di due o più valori uguali entro la stessa riga, secondo la procedura abituale, il punteggio viene calcolato sulla media dei loro ranghi.

Con i dati dell'esempio:

	CONCENTRAZIONE (g/l)				
Ricercatore	144	108	72	54	36
A	2	1	4	5	3
B	2	1	3	5	4
C	1	3	2	4	5

4 - Il test utilizza i totali dei ranghi  $R_j$ , effettuati per colonna.

	CONCENTRAZIONE (g/l)				
	144	108	72	54	36
$R_j$	5	5	9	14	12

Se è vera l'ipotesi nulla  $H_0$ , i punteggi saranno distribuiti a caso entro la stessa riga e pertanto le somme dei ranghi ( $R_j$ ) per colonna tenderanno ad essere uguali. Questi valori dipenderanno dal numero  $N$  di dati presenti in ogni colonna.

5 - Nel test di Page, la sintesi degli  $R_j$  è fornita da  $L$ , definita come la somma dei ranghi di ogni colonna, moltiplicata per la posizione attribuita alla colonna, come nella formula seguente

$$L = \sum j \cdot R_j = 1 \cdot R_1 + 2 \cdot R_2 + 3 \cdot R_3 + \dots + k \cdot R_k$$

Con i dati dell'esempio

$$L = 1 \times 5 + 2 \times 5 + 3 \times 9 + 4 \times 14 + 5 \times 12 = 5 + 10 + 27 + 56 + 60 = 158$$

$L$  risulta uguale a 158.

6 - Nel caso di **piccoli campioni**, la significatività di  $L$  è fornita dalle tabelle. L'entrata è determinata dal numero di colonne o gruppi ( $k$ ) e dal numero di dati ( $N$ ) in ogni colonna.

Con i dati dell'esempio,

in cui  $k = 5$  e  $N = 3$ ,

la tabella dei valori critici alla probabilità

- $\alpha = 0.05$  riporta un valore uguale a **150**
- $\alpha = 0.01$  riporta un valore uguale a **155**
- $\alpha = 0.001$  riporta un valore uguale a **160**

7 - Nel caso di **grandi campioni**, la statistica  $L$  si distribuisce in modo approssimativamente normale

$$z = \frac{L - \mu_L}{\sigma_L}$$

dove, in tabelle di

$k$  trattamenti o colonne ,

$N$  righe o osservazioni per colonna,

la media attesa  $\mu_L$  è

$$\mu_L = \frac{N \cdot k \cdot (k+1)^2}{4}$$

### Valori critici di L per il test di Page

(Sono significativi i valori uguali o maggiori di quelli riportati alle probabilità prefissate)

N	k = 3			K = 4			k = 5		
	a			a			a		
	0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
2	28			58	60		103	106	109
3	41	42		84	87	89	150	155	160
4	54	55	56	111	114	117	197	204	210
5	66	68	70	137	141	145	244	251	259
6	79	81	83	163	167	172	291	299	307
7	91	93	96	189	193	198	338	346	355
8	104	106	109	214	220	225	384	393	403
9	116	119	121	240	246	252	431	441	451
10	128	131	134	266	272	278	477	487	499
11	141	144	147	292	298	305	523	534	546
12	153	156	160	317	324	331	570	581	593
13	165	169	172						
14	178	181	185						
15	190	194	197						
16	202	206	210						
17	215	218	223						
18	227	231	235						
19	239	243	248						
20	251	256	260						

(continua)

### Valori critici di L per il test di Page

(Sono significativi i valori uguali o maggiori di quelli riportati alle probabilità prefissate)

N	k = 6			K = 7			k = 8		
	a			a			a		
	0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
2	166	173	178	252	261	269	362	376	388
3	244	252	260	370	382	394	532	549	567
4	321	331	341	487	501	516	701	722	743
5	397	409	420	603	620	637	869	893	917
6	474	486	499	719	737	757	1037	1063	1090
7	550	563	577	835	855	876	1204	1232	1262
8	625	640	655	950	972	994	1371	1401	1433
9	701	717	733	1065	1088	1113	1537	1569	1603
10	777	793	811	1180	1205	1230	1703	1736	1773
11	852	869	888	1295	1321	1348	1868	1905	1943
12	928	946	965	1410	1437	1465	2035	2072	2112

(continua)

### Valori critici di L per il test di Page

(Sono significativi i valori uguali o maggiori di quelli riportati alle probabilità prefissate)

N	k = 9			k = 10		
	a			a		
	0.05	0.01	0.001	0.05	0.01	0.001
2	500	520	544	670	696	726
3	736	761	790	987	1019	1056
4	971	999	1032	1301	1339	1382
5	1204	1236	1273	1614	1656	1704
6	1436	1472	1512	1927	1972	2025
7	1668	1706	1750	2238	2288	2344
8	1900	1940	1987	2549	2602	2662
9	2131	2174	2223	2859	2915	2980
10	2361	2407	2459	3169	3228	3296
11	2592	2639	2694	3478	3541	3612
12	2822	2872	2929	3788	3852	3927

e  $s_L$  è

$$s_L = \sqrt{\frac{N(k^3 - k)^2}{144(k-1)}}$$

Sostituendo in

$$z = \frac{L - \mu_L}{\sigma_L}$$

e semplificando,

si ottiene la formula abbreviata

$$z_L = \frac{12 \cdot L - 3 \cdot N \cdot k \cdot (k+1)^2}{k \cdot (k^2 - 1)} \cdot \sqrt{\frac{k-1}{N}}$$

che permette un calcolo più rapido e semplice del valore di Z.

Poiché è fornito il verso delle differenze, come nel test di Jonckheere, **il test di Page è un test ad una coda o unilaterale**. Pertanto, il valore critico significativo alla probabilità  $\alpha = 0.05$  è  $z = 1,645$ .

**TIES**. I dati dovrebbero essere misurati su una scala continua, per cui **non dovrebbero esistere valori con lo stesso rango**. Non sono state proposte correzioni per i ties. Tuttavia un numero molto limitato di valori sono accettabili, poiché non alterano sensibilmente il risultato e forniscono una risposta più cautelativa.

Nel **caso di 2 soli gruppi (k = 2), l'alternativa al test di Page è il test U di Mann-Whitney per 2 campioni dipendenti, nell'ipotesi unidirezionale** (Infatti verifica la direzione della tendenza centrale dei due gruppi).

**Nelle situazioni in cui i trattamenti hanno un ordine naturale, il test di Page è preferibile a quello di Friedman**, che risponde solo alla domanda se le mediane dei vari trattamenti sono differenti.

Una volta che sia stata rifiutata l'ipotesi nulla, per ottenere una informazione più completa sulla serie delle mediane, è possibile **verificare tra quali esista una differenza significativa, ricorrendo ai confronti multipli basati sulla somma dei ranghi di Friedman**.

ESEMPIO 1 (PER GRANDI CAMPIONI).

Lungo il corso d'acqua che attraversa una città, sono state collocate 6 stazioni (A, B, C, D, E, F) di rilevazione dell'inquinamento. In ognuna delle 6 stazioni, per 15 giorni è stata fatta una misura del carico inquinante.

I valori campionati, classificati per stazione e per giorno di rilevazione, sono riportati nella tabella a due entrate sottostante:

GIORNI	STAZIONI					
	A	B	C	D	E	F
1	20	18	24	22	29	38
2	32	37	34	31	39	38
3	18	23	19	25	23	26
4	9	7	14	11	12	11
5	29	37	32	59	40	45
6	38	25	27	47	45	45
7	8	15	7	12	15	13
8	18	13	22	26	23	22
9	32	36	37	35	48	40
10	23	25	26	25	32	56
11	6	8	12	9	10	10
12	24	18	20	27	25	27
13	13	18	14	14	19	26
14	18	26	19	19	29	32
15	14	12	25	56	54	75

Si intende verificare se, con l'accumulo degli scarichi urbani, lungo il percorso (dalla stazione A alla F) si ha un aumento significativo del carico inquinante.

Risposta.

Dapprima si devono definire l'ipotesi nulla, che è fondata sulle mediane

$$H_0: me_A = me_B = me_C = me_D = me_E = me_F$$

e l'ipotesi alternativa, relativa al quesito del problema

$$H_1: me_A \leq me_B \leq me_C \leq me_D \leq me_E \leq me_F$$

Successivamente, poiché i trattamenti sono già nell'ordine naturale, si devono trasformare i valori in ranghi, considerando in modo indipendente ogni riga, come nella tabella sottostante:

GIORNI	STAZIONI					
	A	B	C	D	E	F
1	2	1	4	3	5	6
2	2	4	3	1	6	5
3	1	3,5	2	5	3,5	6
4	2	1	6	3,5	5	3,5
5	1	3	2	6	4	5
6	3	1	2	6	4,5	4,5
7	2	5,5	1	3	5,5	4
8	2	1	3,5	6	5	3,5
9	1	3	4	2	6	5
10	1	2,5	4	2,5	5	6
11	1	2	6	3	4,5	4,5
12	3	1	2	5,5	4	5,5
13	1	4	2,5	2,5	5	6
14	1	4	2,5	2,5	5	6
15	2	1	3	5	4	6
R <sub>j</sub>	25,0	37,5	47,5	56,5	71,5	82,5

Successivamente, calcolare le somme dei ranghi per colonna ( $R_j$ ), come riportato nell'ultima riga della tabella precedente.

Dalla somma dei ranghi per colonna ( $R_j$ ) si ottiene il valore di  $L$

$$L = (1 \cdot 25,0) + (2 \cdot 37,5) + (3 \cdot 47,5) + (4 \cdot 56,5) + (5 \cdot 71,5) + (6 \cdot 82,5) = \\ = 25 + 75 + 142,5 + 226 + 357,5 + 495 = \mathbf{1321}$$

che risulta uguale a 1321

e da esso si stima il valore di  $z_L$  relativo

$$z_L = \frac{(12 \cdot 1321) - (3 \cdot 15 \cdot 6 \cdot 7^2)}{6 \cdot (6^2 - 1)} \cdot \sqrt{\frac{5}{15}} = \frac{15852 - 13230}{210} \cdot \sqrt{0,333} = 12,48 \cdot 0,577 = \mathbf{7,20}$$

che risulta uguale a 7,20.

Il valore di  $z_L$  è particolarmente elevato, superiore a quelli riportati nelle tabella della distribuzione normale: corrisponde ad una probabilità  $\alpha$  inferiore a 1 su un milione.

Pertanto, si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa: dalla stazione A alla stazione F si ha un aumento altamente significativo del carico inquinante, anche se non è possibile dire nulla sulla curva di crescita.

L'eventuale ricorso ai confronti multipli con il test di Friedman permette di verificare tra quali stazioni esiste una differenza significativa.

#### ESEMPIO 2 (PER PICCOLI CAMPIONI).

In un centro urbano sono stati attivati 5 punti di osservazione (A, B, C, D, E) per rilevare l'inquinamento dovuto al traffico e per ogni zona sono state effettuate 4 osservazioni, a distanza di 4 ore ognuna, dalle ore 6 alle ore 18.

Per esporre alla cittadinanza la situazione dell'inquinamento atmosferico in modo chiaro e sintetico, l'ufficio comunale ha fornito le misure su una scala ordinale, con giudizi della qualità dell'aria riportati in modo simbolico,

- pessima,
- insufficiente,
- = sufficiente,
- + buona,
- ++ ottima

come nella tabella sottostante

STAZIONE	ORE			
	6	10	14	18
a	-	--	+	=
b	--	=	-	=
c	-	=	+	+
d	=	+	=	++
e	-	+	=	+

Si vuole verificare se, nel corso di una giornata con blocco del traffico, dal mattino alla sera vi è stata una diminuzione nei tassi d'inquinamento.

Risposta.

I dati devono essere

- trasformati in ranghi entro la stessa riga e
- sommati per colonna ( $R_j$ ),

STAZIONE	ORE			
	6	10	14	18
a	2	1	4	3
b	1	3,5	2	3,5
c	1	2	3,5	3,5
d	1,5	3	1,5	4
e	1	3,5	3	3,5
<b><math>R_j</math></b>	<b>6,5</b>	<b>13,0</b>	<b>14,0</b>	<b>17,5</b>

al fine di calcolare il valore di  $L$

$$L = 1 \times 6,5 + 2 \times 13 + 3 \times 14 + 4 \times 17,5 = 144,5$$

che risulta uguale a 144,5.

Per  $k = 4$  e  $N = 5$ ,

la tabella sinottica dei **valori critici della  $L$  di Page** riporta

- alla probabilità  $\alpha = 0.05$  il valore 137
- alla probabilità  $\alpha = 0.01$  il valore 141
- alla probabilità  $\alpha = 0.001$  il valore 145.

Il valore di  $L$  calcolato (144,5) è compreso tra quello relativo alla probabilità  $\alpha = 0.01$  e quello alla probabilità  $\alpha = 0.001$ .

Di conseguenza, si rifiuta l'ipotesi nulla con probabilità inferiore a 0.01 e maggiore di 0.001 di commettere un errore di I tipo.

Si accetta l'ipotesi alternativa: esiste un miglioramento delle condizioni dell'aria al trascorrere delle ore, durante la giornata di blocco del traffico.

### ESEMPIO 3

(VALORI CRITICI PER PICCOLI CAMPIONI E DISTRIBUZIONE Z PER GRANDI).

E' ovvio che, quando il campione non è troppo piccolo, in tutti i test non parametrici esista corrispondenza tra le probabilità indicate dai valori critici riportate nelle tabelle sinottiche e quelle stimate con la distribuzione normale.

E' una relazione che spesso sfugge allo studente e al ricercatore, che non abbiano acquisito sufficiente familiarità con la statistica. A questo scopo, è didatticamente utile l'esercizio di valutare le probabilità corrispondenti ai valori critici.

**Per il test di Page** dalla tabella dei valori critici ricaviamo che

per **k** uguale a 10 ed **N** uguale a 12,

- alla probabilità  $\alpha = 0.05$  il valore critico è 3788,
- alla probabilità  $\alpha = 0.01$  è 3852 e
- alla probabilità  $\alpha = 0.001$  è 3927.

Poiché (con **k** uguale a 10 ed **N** uguale a 12) il campione è di dimensioni relativamente grandi, le probabilità calcolate con la distribuzione normale, ovviamente solo in una coda come richiede il test, dovrebbero essere molto simili.

Infatti, ricordando la formula generale

$$z_L = \frac{12 \cdot L - 3 \cdot N \cdot k \cdot (k+1)^2}{k \cdot (k^2 - 1)} \cdot \sqrt{\frac{k-1}{N}}$$

con **N** = 12 e **k** = 10 si ottiene

$$z = \frac{12 \cdot L - 3 \cdot 12 \cdot 10 \cdot (10+1)^2}{10 \cdot (10^2 - 1)} \cdot \sqrt{\frac{10-1}{12}}$$

nella quale l'unica incognita è il valore di **L**, fornito dalle tabelle per piccoli campioni.

Per la probabilità  $\alpha = 0.05$  il valore critico (già riportato) di **L** è 3788;

con esso il risultato di **z** diviene

$$z = \frac{(12 \cdot 3788) - (360 \cdot 11^2)}{10 \cdot 99} \cdot \sqrt{\frac{9}{12}} = \frac{45456 - 43560}{990} \cdot \sqrt{0,75} = \frac{1896}{990} \cdot 0,866 = 1,66$$

uguale a 1,66.

Nella tabella dei valori di **z** in una coda della distribuzione normale a 1,66 corrisponde la probabilità di 0.0485. E' un valore molto vicino a 0.05.

Per la probabilità  $\alpha = 0.01$  a **L** si sostituisce il valore critico di 3852;

il risultato di **z** diviene

$$z = \frac{(12 \cdot 3852) - (360 \cdot 11^2)}{10 \cdot 99} \cdot \sqrt{\frac{9}{12}} = \frac{46224 - 43560}{990} \cdot \sqrt{0,75} = \frac{2664}{990} \cdot 0,866 = 2,33$$

uguale a 2,33 al quale, in una coda della distribuzione normale, corrisponde la probabilità di 0.099.

Per la probabilità  $\alpha = 0.001$  a **L** si sostituisce il valore critico di 3927;

il risultato di **z** diviene

$$z = \frac{(12 \cdot 3927) - (360 \cdot 11^2)}{10 \cdot 99} \cdot \sqrt{\frac{9}{12}} = \frac{47124 - 43560}{990} \cdot \sqrt{0,75} = \frac{3564}{990} \cdot 0,866 = 3,11$$

uguale a 3,11 al quale corrisponde la probabilità di 0.00099.

**In assenza dei valori critici, il ricorso alla distribuzione normale fornisce risposte attendibili anche per campioni non troppo piccoli.**

#### 14.4. LE MISURE D'ASSOCIAZIONE O D'INDIPENDENZA

Quando i dati sono classificati per dati categoriali o qualitativi secondo due variabili, le **frequenze (di solito assolute e quindi espresse come conteggi)** sono riportate in una **tabella di contingenza m x n**, con dimensioni minime 2 x 2. Gli indici per quantificare le relazioni tra le due variabili sono chiamate **misure di associazione e le ipotesi relative verificano l'esistenza o meno dell'indipendenza.**

Una dimensione corrisponde ad una variabile classificatoria che dovrebbe essere **esplicativa** (come la dose di un farmaco, la località nella quale si è raccolto un campione di alcune specie animali o vegetali) e l'altra dimensione essere una **risposta** (come l'effetto del farmaco che può essere nullo, moderato o forte; le varie specie raccolte).

Per le due variabili, i gruppi possono essere formati sulla base di dati

- **qualitativi o nominali**, come l'elenco delle località e quello delle specie;
- **ordinali o di rango**, come l'intensità della risposta al farmaco (nulla, moderata, forte) o la classificazione delle specie in classi d'età (giovani, adulti, vecchi) o livelli di sviluppo;
- **di intervalli o di rapporti** (come l'età o le dimensioni) raggruppati in classi, **con intervalli differenti oppure costanti.**

Da queste tre tipi di classificazioni, si possono derivare tabelle a due entrate con suddivisione

- nominale per ambedue le variabili,
- nominale per una ed ordinale per l'altra,
- ordinale per ambedue le variabili,
- nominale per una e intervallare per l'altra,
- in tutte le combinazioni possibili, fino a intervallare per entrambe.

Non esiste una misura ideale dell'associazione o concordanza tra le due variabili, valida per tutte le situazioni. Sono quindi stati proposti diversi indici, che appartengono alla statistica descrittiva; pertanto, non esiste la necessità di formulare ipotesi e di verificare che esistano le condizioni di validità. Dove si stima la varianza dell'indice, è possibile calcolare l'intervallo fiduciale e quindi, sull'ipotesi che l'indice abbia una distribuzione normale, giungere all'inferenza.

La scala di misura utilizzata, l'ipotesi formulata, le caratteristiche dei dati richiedono indici di associazione differenti, anche quando i dati sono riportati con le stesse modalità in una **tabella bivariata, come quella già presentata per il test  $\chi^2$** . Una classificazione scolastica, utile per ordinare la presentazione degli indici più frequentemente utilizzati, propone una suddivisione per misure **nominali, ordinali e in classi d'intervalli**.

Un modo per classificazione gli indici di associazione o indipendenza, semplice e pertanto seguito in questa presentazione, è quello della distinzione in

- tabelle 2 x 2
- tabelle m x n

proposto nel testo di **Graham J. G. Upton** (The Analysis of Cross-tabuled Data, John Wiley & Sons, Chichester – New York, 1978, reprinted April 1980).

#### 14.5. ASSOCIAZIONE IN TABELLE 2 X 2 O FRA VARIABILI DICOTOMICHE:

**IL Q E L'Y DI YULE, IL  $j$ , IL  $D_{sim}$ . E IL  $D_{xy}$  DI SOMERS, IL  $t_c$ ;**

**IL  $f$  E IL C DI PEARSON, IL V DI CRAMER, IL  $D_T$  DI TSCHUPROW**

In una tabella 2 x 2, della quale viene riportato lo schema con la consueta simbologia,

	+	-	<b>Totale</b>
<b>Gruppo A</b>	<i>a</i>	<i>b</i>	<i>n<sub>1</sub></i>
<b>Gruppo B</b>	<i>c</i>	<i>d</i>	<i>n<sub>2</sub></i>
<b>Totale</b>	<i>n<sub>3</sub></i>	<i>n<sub>4</sub></i>	<i>N</i>

la significatività dell'associazione può essere misurata attraverso il chi quadrato o il test G, con tutte le loro varianti di correzioni per la continuità; nel caso di grandi campioni è possibile utilizzare la distribuzione normale, eventualmente con una correzione per la continuità quando è inferiore a 200 osservazioni; se il campione è piccolo, si ricorre al metodo esatto di Fisher.

Dalla tabella, per l'analisi delle relazioni tra le due variabili possono essere ricavate anche misure su **il tipo e l'intensità dell'associazione**;

a questo scopo, si ricorre all'analisi delle due diagonali, in cui

- **a-d è la diagonale principale**
- **b-c è la diagonale secondaria.**

Per convenzione, alla associazione è attribuito

- **segno positivo, quando le frequenze sono più alte nelle due celle della diagonale principale (a-d);**
- **segno negativo, quando le frequenze sono più alte nelle due celle della diagonale secondaria (b-c).**

Definire un'associazione positiva o negativa in questo modo **è puramente convenzionale**, poiché è sufficiente invertire la posizione delle due righe oppure delle due colonne per ottenere un'associazione di tipo opposto.

Il concetto di indipendenza o di associazione può essere fatto derivare da quello di equilibrio o squilibrio tra le due modalità di una variabile categoriale. Stimato a partire dai totali marginali, **l'equilibrio tra le modalità di una dicotomia è massimo quando ciascuna ha lo stesso numero di dati**; in questa situazione si ha anche il **massimo di varianza**, poiché si ha il massimo di probabilità di errore quando si vuole indovinare se un dato appartiene ad una categoria oppure all'altra. Il concetto può essere compreso più facilmente partendo dalla situazione opposta.

Se nella zona A tutti i laghi hanno un inquinamento elevato e nella zona B tutti hanno livelli d'inquinamento bassi, come nella tabella seguente,

	Inquinamento		
	Alto	Basso	Totale
Zona A	50	0	50
Zona B	0	60	60
Totale	50	60	110

è facile indovinare, sulla semplice appartenenza alla zona, se il lago ha un livello d'inquinamento alto o basso. Nulla cambierebbe se fosse l'opposto,

	Inquinamento		
	Alto	Basso	Totale
Zona A	0	50	50
Zona B	60	0	60
Totale	60	50	110

cioè con i valori massimi collocati sulla diagonale secondaria.

**L'associazione emerge con la massima chiarezza, quando le frequenze sono distribuite nelle due celle appartenenti alla stessa diagonale.**

Al contrario, quando le due dicotomie sono esattamente equilibrate

	Inquinamento		
	Alto	Basso	Totale
Zona A	30	30	60
Zona B	25	25	50
Totale	55	55	110

la probabilità di indovinare se il lago abbia un livello d'inquinamento alto o basso, sulla base della zona di appartenenza, è minima e quindi la varianza d'errore è massima.

Un metodo generalmente valido, in tabelle di qualsiasi dimensione, per valutare il **tipo di associazione in ogni casella** (non più in complesso come prima) è il confronto tra la frequenza osservata e quella attesa, sulla base della nota relazione:

$$freq. \text{ attesa} = \frac{\text{totale di riga} \times \text{totale di colonna}}{\text{totale generale}}$$

Se la **frequenza osservata** è

- **maggiore** di quella attesa, l'associazione è **positiva**
- **minore** di quella attesa, l'associazione è **negativa**.

La scuola francese le definisce rispettivamente **attrazione** e **repulsione**.

Per valutare **l'associazione in tabelle 2 x 2, quindi fra due dicotomie**, sono stati proposti decine di coefficienti. Qui vengono riportati quelli di uso più comune nei programmi informatici più diffusi. Una classificazione, riportata in vari testi, li distingue in

- **coefficienti bidirezionali,**
- **coefficienti unidirezionali**

anche se vari coefficienti, presentati come uni-direzionali e ritenuti tali da molti autori di testi di statistica, per altri non lo sono affatto.

I **coefficienti bidirezionali sono fondati** sul cosiddetto **prodotto incrociato** (*cross-product*)

$$\text{prodotto incrociato} = ad - bc$$

- che può avere valore positivo, negativo o nullo,
- in corrispondenza di una associazione positiva, negativa od inesistente.

I vari indici proposti differiscono nel modo di normalizzare il prodotto incrociato, attraverso il denominatore della frazione, come sarà possibile verificare con un confronto complessivo, ovviamente dopo la presentazione di ognuno di essi.

**Udny Yule** nel 1900 (con l'articolo *On the association of attributes in statistics*, pubblicato su *Phil. Trans., A*, 194, pp. 257-319) ha proposto il **coefficiente di associazione Q**, in onore dello statista belga **Quetelet**

$$Q = \frac{ad - bc}{ad + bc}$$

che probabilmente è la misura di associazione più utilizzata.

Quando **N** è sufficientemente grande, la distribuzione di **Q** è normale, con varianza  $s_Q^2$

$$s_Q^2 = \frac{1}{4} \cdot (1 - Q^2)^2 \cdot \left( \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right)$$

Il valore di **Q** può variare tra

- **-1**, che indica un'associazione completa negativa e
- **+1** che indica un'associazione completa positiva,
- 
- con **0** che indica assenza di associazione o presenza di indipendenza totale.

**La misura della varianza permette l'inferenza sul valore di Q ricorrendo all'intervallo fiduciale con la distribuzione z, alla probabilità  $\alpha$  specificata ( $z_\alpha$ ), come possibile nel caso di grandi campioni:**

$$\text{Intervallo fiduciale di } Q = Q \pm z_\alpha \sqrt{S_Q^2}$$

Con un test bilaterale si può rifiutare l'ipotesi nulla, se il valore di confronto (di norma **0** quando si intende valutare se il valore di **Q** è significativo) è escluso dall'intervallo stimato.

In precedenza, è stato più volte ripetuto come l'ipotesi di indipendenza o associazione tra due variabili deve essere testata attraverso il  $\chi^2_{(1)}$  o il test  $G^2$ .

Utilizzando anche questo ultimo metodo, può avvenire che

**il test con l'intervallo fiduciale di Q e il test  $\chi^2_{(1)}$  danno risposte differenti,**

poiché la misura dell'associazione è calcolata in modi differenti.

Come riportano vari testi, **la teoria statistica suggerisce di dare maggiore credito al test  $\chi^2_{(1)}$ .**

I coefficienti bidirezionali servono quando si vuole analizzare la reciproca influenza tra le due variabili categoriali, in modo analogo a quanto avviene nella correlazione.

Coefficienti bidirezionali possono servire per valutare l'associazione della presenza o assenza di una specie animale rispetto ad un'altra, quando tra esse non esiste predazione o simbiosi, come nella tabella:

		Specie A		
		Presenza	Assenza	Totale
Specie B	Presenza	32	48	80
	Assenza	13	57	70
Totale		45	105	150

Per verificare se esiste una differenza significativa nella frequenza della presenza della specie A e della specie B si ottiene un valore del chi quadrato, con **1** gdl, uguale a **8,163**. Poiché la tavola sinottica del  $\chi^2_{(1)}$  riporta

- il valore critico di **3,84** alla probabilità  $\alpha = 0.05$ ,
- il valore critico di **6,64** alla probabilità  $\alpha = 0.01$ ,

si rifiuta l'ipotesi nulla ed implicitamente si è accettata l'ipotesi alternativa.

Il valore dell'indice **Q**, stima dell'associazione tra presenza della specie A e della specie B, è

$$Q = \frac{32 \cdot 57 - 48 \cdot 13}{32 \cdot 57 + 48 \cdot 13} = \frac{1824 - 624}{1824 + 624} = \frac{1200}{2448} = 0,49$$

uguale a **0,49**.

La sua varianza  $s_Q^2$  è

$$s_Q^2 = \frac{1}{4} \cdot (1 - 0,49)^2 \cdot \left( \frac{1}{32} + \frac{1}{48} + \frac{1}{13} + \frac{1}{57} \right)$$

$$s_Q^2 = 0,25 \cdot 0,7599^2 \cdot (0,0313 + 0,0200 + 0,0769 + 0,0175) = 0,25 \cdot 0,5774 \cdot 0,1457 = 0,021$$

uguale a 0,021

Poiché il campione può essere considerato di grandi dimensioni (anche se vari autori pongono questo limite per  $N \geq 200$ ), è corretto utilizzare la distribuzione normale, nel quale per un **test bilaterale**

- alla probabilità  $\alpha = 0,05$  il valore critico è **1,96**
- alla probabilità  $\alpha = 0,01$  il valore critico è **2,58**.

Alla probabilità  $\alpha = 0,05$

$$\text{Intervallo fiduciale di } Q = 0,49 \pm 1,96 \sqrt{0,021} = 0,49 \pm 0,284$$

l'intervallo fiduciale di  $Q$  è compreso tra 0,206 e 0,774

mentre alla probabilità  $\alpha = 0,01$

$$\text{Intervallo fiduciale di } Q = 0,49 \pm 2,58 \sqrt{0,021} = 0,49 \pm 0,374$$

l'intervallo fiduciale di  $Q$  è compreso tra 0,116 e 0,864.

Sempre **U. Yule** nel 1912 (con l'articolo *On the methods of measuring association for ordinal variables*, pubblicato da *Journal of the Royal Statistical Society* n. 75 pp. 579-642) ha proposto un secondo **indice di associazione Y**, di uso meno comune,

ottenuto da

$$Y = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

Con i dati della tabella precedente,  $Y$

$$Y = \frac{\sqrt{32 \cdot 57} - \sqrt{48 \cdot 13}}{\sqrt{32 \cdot 57} + \sqrt{48 \cdot 13}} = \frac{42,71 - 24,98}{42,71 + 24,98} = \frac{17,73}{67,69} = 0,262$$

risulta uguale a 0,262.

Un terzo metodo molto usato è **il coefficiente j**, in cui la normalizzazione (per normalizzazione si intende porre una variabile in rapporto con un'altra; in questo caso, rendere indipendente il numeratore dal numero di dati utilizzati) è ottenuta ponendo al denominatore la radice quadrata del prodotto dei 4 totali marginali:

$$j = \frac{ad - bc}{\sqrt{n_1 \cdot n_2 \cdot n_3 \cdot n_4}}$$

Con i dati dell'esempio, j

$$j = \frac{32 \times 57 - 48 \times 13}{\sqrt{80 \times 70 \times 45 \times 105}} = \frac{1824 - 624}{\sqrt{26460000}} = \frac{1200}{5143,94} = 0,233$$

risulta uguale a 0,233.

Un quarto coefficiente è il **D<sub>sim.</sub>** di Robert **Somers** (ricordando che Somers ha proposto anche un **D** asimmetrico, nettamente differente da questo, oltre a proporre un **D** uni-direzionale e un **D** bidirezionale, per variabili ordinali, non qualitative o dicotomiche come queste, trattate nei paragrafi successivi), con

$$D_{sim.} = \frac{ad - bc}{(ad) + (bc) + \frac{(a + d) \cdot (b + c)}{2}}$$

Sempre con i dati dello stesso esempio, **D<sub>sim.</sub>**

$$D_{sim.} = \frac{32 \times 57 - 48 \times 13}{(32 \times 57) + (48 \times 13) + \frac{(32 + 57) \cdot (48 + 13)}{2}}$$

$$D_{sim.} = \frac{1824 - 624}{1824 + 624 + \frac{89 \cdot 61}{2}} = \frac{1200}{1824 + 624 + 2714,5} = \frac{1200}{5162,5} = 0,232$$

risulta uguale a 0,232.

Un quinto è il **coefficiente t<sub>c</sub>**, con

$$t_c = \frac{4 \cdot (ad - bc)}{N^2}$$

Con i dati dell'esempio, t<sub>c</sub>

$$t_c = \frac{4 \cdot (32 \cdot 57 - 48 \cdot 13)}{150^2} = \frac{4800}{22500} = 0,213$$

risulta uguale 0,213.

Per un confronto empirico, senza entrare nella discussione sulle caratteristiche di ogni indice, è semplice osservare che con

- Q = 0,490
- Y = 0,262
- $\phi = 0,233$
- $D_{sim.} = 0,232$
- $t_c = 0,213$

i valori di associazione stimati sulla stessa tabella 2 x 2 sono tra loro simili, eccetto il **Q** di Yule.

**I coefficienti unidirezionali** servono per rilevare l'influenza di una variabile sull'altra. La tabella viene costruita mettendo

- la **variabile indipendente sulle righe** e quindi
- la **variabile dipendente sulle colonne**.
- E' il caso in cui si vuole analizzare l'associazione tra livello d'inquinamento e frequenze di malattie polmonari: l'alto inquinamento della zona può essere visto come la causa delle malattie polmonari (per un successivo confronto tra i risultati, sono stati utilizzati gli stessi dati della tabella precedente):

	<b>Personne <u>con</u> malattie polmonari</b>	<b>Personne <u>senza</u> malattie polmonari</b>	<b>Totale</b>
<b>Zona ad <u>alto</u> inquinamento</b>	32 <b>a</b>	48 <b>b</b>	<b>80</b> <b>n<sub>1</sub></b>
<b>Zona a <u>basso</u> inquinamento</b>	13 <b>c</b>	57 <b>d</b>	<b>70</b> <b>n<sub>2</sub></b>
<b>Totale</b>	<b>45</b> <b>n<sub>3</sub></b>	<b>105</b> <b>n<sub>4</sub></b>	<b>150</b> <b>N</b>

**Tra i coefficienti uni-direzionali**, è utile ricordare il D asimmetrico, indicato con **D<sub>xy</sub>** anch'esso attribuito a R. H. Somers (con l'articolo del 1962, *A new asymmetric measure of association for*

*ordinal variables*, pubblicato su *American Sociological Review* 27, n.6, pp.700-811), benché alcuni autori ritengano che la prima proposta sia da attribuire all'americano Pierce nel 1884:

esso normalizza per i due totali di riga  $\mathbf{n}_1$  e  $\mathbf{n}_2$  con la formula

$$\mathbf{D}_{xy} = \frac{ad - bc}{(a+b) \cdot (c+d)}$$

scritta in modo leggermente diverso da  $\mathbf{n}_1 \times \mathbf{n}_2$ .

Si può osservare che il D asimmetrico ( $\mathbf{D}_{xy}$ ) non assomiglia al D simmetrico ( $\mathbf{D}_{sim.}$ ), quanto piuttosto al  $j$ .

Con i dati dell'esempio,  $\mathbf{D}_{xy}$

$$\mathbf{D}_{xy} = \frac{32 \times 57 - 48 \times 13}{(32 + 48) \cdot (13 + 57)} = \frac{1824 - 624}{80 \cdot 70} = \frac{1200}{5600} = 0,214$$

risulta uguale a 0,214.

Per scegliere tra  $\mathbf{Q}$ ,  $j$ ,  $\mathbf{D}_{sim.}$ ,  $t_c$  il coefficiente bidirezionale che meglio si adatta a descrivere l'associazione tra due variabili e valutare le distorsioni  $\mathbf{D}_{xy}$  nelle varie situazioni, è conveniente ricordare che:

- quando entrambe le coppie di totali marginali ( $n_1 = n_2$ ,  $n_3 = n_4$ ) sono equilibrate, tutti e cinque gli indici danno stime corrette;
- quando una variabile è equilibrata e l'altra meno, come nel caso dell'esempio (con 80 e 70 quasi simili, mentre 105 e 45 differiscono sensibilmente)  $\mathbf{Q}$  fornisce una sovrastima;
- la  $\mathbf{Q}$  di Yule è preferibile a tutte le altre misure, se una diagonale è semivuota;
- la  $j$  è preferibile alla  $\mathbf{Q}$ , quando una sola coppia è equilibrata;
- la  $t_c$  fornisce le risposte più accettabili, quando tre celle sono semivote; in termini più tecnici, in caso di associazione d'angolo;
- la  $\mathbf{D}_{xy}$  è sicuramente inaffidabile, se le celle vuote o semivote sono una oppure tre.

**Emerge con chiarezza che non esiste un solo coefficiente dicotomico affidabile in tutti i casi.**

Come più volte ripetuto, con **categorie qualitative** il grado di associazione o di relazione tra due variabili è fornito dal  $c^2$  di Pearson.

**Tuttavia, il valore del  $c^2$  osservato dipende**

- non solo dallo **scostamento delle frequenze osservate da quelle attese** (fenomeno che si vuole analizzare),
- ma pure **dalle dimensioni del campione** e
- **dal numero di gdl** (che in questo caso risultano fattori di disturbo).

Di conseguenza se, **a scopo descrittivo**, si intende **estendere il confronto a due o più tabelle** (ad esempio al fine di valutare il grado associazione tra specie e località), esse dovrebbero avere lo stesso

numero di dati e lo stesso numero di livelli nelle due variabili; è una condizione estremamente limitativa, per la ricerca sperimentale.

Inoltre, **il valore del  $c^2$  non mostra con immediatezza il grado di associazione**, come succede con l'indice di correlazione  $r$  di Pearson (che può variare da  $-1$  a  $+1$  ed è  $0$  quando non esiste associazione).

Per

- **eliminare l'influenza delle dimensioni del campione e**
- **dei gradi di libertà e**
- **per restringere l'intervallo di distribuzione dei valori tra  $0$  e  $1$** , un campo di variazione che permette un confronto immediato tra coefficienti calcolati su campioni diversi,

è stato proposto da Pearson l'indice  $f$  (phi)

$$f = \sqrt{\frac{c^2}{N}}$$

Per ottenere valori compresi sempre nell'intervallo  $0-1$ , **Pearson** ha proposto il **coefficiente C**

$$C = \sqrt{\frac{c^2}{c^2 + N}}$$

chiamato **coefficiente di contingenza**, (raramente indicato anche con  $D_p$ ) che soffre del limite di non raggiungere mai  $1$ . Il suo valore massimo dipende dalle dimensioni della tabella, ossia dal numero di righe e di colonne (ad es. raggiunge come massimo  $0,87$  in una tabella  $4 \times 4$ ).

Tra le varie correzioni proposte, per il suo uso frequente è da ricordare anche il **coefficiente V di Cramér** (*Cramér's statistic*) proposto appunto dallo statistico svedese Harald Cramér nel 1946 (nel volume *Mathematical Methods of Statistics*, stampato dalla *Princeton University Press*)

$$V = \sqrt{\frac{c^2}{N \cdot (k - 1)}}$$

dove

$k$  è il numero minimo di colonne o di righe nella tabella di contingenza  $m \times n$ .

**Il V di Cramér raggiunge il valore massimo di 1 in tabelle di qualsiasi dimensioni.**

Ovviamente, come è banale dedurre dalla formula, in tabelle  $2 \times 2$  e in quelle  $n \times 2$ , i coefficienti  $f$  di Pearson e  $V$  di Cramér assumono valori identici.

Analogo come concetti e per lo stesso uso, è stato proposto anche **il coefficiente  $D_T$  (a volte indicato anche con  $T$ ) di Tschuprow** (di nazionalità russa e quindi con cognome scritto in cirillico; Tschuprow è la translitterazione tedesca; in italiano alcuni autori di testi translitterano in Sciuprov):

$$D_T = \sqrt{\frac{c^2}{N\sqrt{(r-1) \cdot (c-1)}}}$$

dove

$c$  e  $r$  sono rispettivamente il numero di colonne e di righe

$N$  il numero di osservazioni.

Caratteristica di questo indice è che può raggiungere 1 (quindi il valore massimo) qualunque sia il numero di righe e di colonne della tabella di contingenza.

Nella tabella 2 x 2 coincide con il  $f$  di Pearson.

**Per questi coefficienti non esistono valori critici, in quanto hanno solo significato descrittivo.**

Sebbene possano essere utilizzate per confrontare l'intensità dell'associazione in tabelle diverse, tutte queste misure basate sul  $c^2$  sono di facile interpretazione solo quando il valore è prossimo a 0, cioè esiste indipendenza tra le due variabili e non si ha associazione tra esse.

ESEMPIO. Nei laghi, la quantità di fosforo è il fattore di norma più importante nel fenomeno della eutrofizzazione. In funzione della sua concentrazione, un lago è classificato in una delle seguenti 5 categorie: ultraoligotrofo, oligotrofo, mesotrofo, eutrofo, ipereutrofo.

La stessa definizione può essere data sulla base della quantità di clorofilla o della trasparenza dell'acqua (che dipendono direttamente dalla quantità di biomassa), dalla quantità di azoto, dalla presenza di gruppi caratteristici, dalla frequenza di fioriture algali, dalla distribuzione verticale della biomassa planctonica, dal numero e dal tipo di specie contemporaneamente presenti od assenti.

Per verificare il grado di associazione tra due variabili qualitative, per 66 laghi è stato contato il numero in cui i fattori A e B (che possono essere due specie o due altri qualsiasi fattori qualitativi) sono presenti(+) od assenti (-) in modo congiunto.

LAGO	FATTORE	
	A	B
1	A+	B+
2	A+	B+
3	A-	B-
4	A+	B-
5	A+	B+
6	A-	B+
---	---	---
65	A+	B+
66	A-	B-

Il lungo elenco è stato riassunto in una tabella 2 x 2, differente da quella del  $\chi^2$  ed analoga a quella di McNemar:

		A		tot
		+	-	
B	+	36	5	41
	-	9	16	25
	tot	45	21	66

Essa evidenzia che in 36 laghi i due fattori sono presenti contemporaneamente, in 5 è presente il fattore B ma assente il fattore A, in 9 è presente il solo fattore A ed assente il B, mentre in 16 laghi sono assenti contemporaneamente sia A che B.

Per valutare il grado di associazione tra le due variabili e stimare la significatività, dapprima si calcola il valore del  $\chi^2$ , che in questo caso ha 1 gdl. Apportando la correzione per campioni con meno di 100 osservazioni si ottiene

$$\chi^2 = \frac{(|36 \cdot 16 - 5 \cdot 9| - 33)^2 \cdot 66}{41 \cdot 25 \cdot 45 \cdot 21} = \frac{(531 - 33)^2 \cdot 66}{968825} = \frac{248004 \cdot 66}{968825} = 16,898$$

un valore del  $\chi^2$  uguale a 16,898 con 1 df.

Successivamente, si deve stimare la significatività dell'associazione. Il valore del  $\chi^2$  è nettamente superiore a quello tabulato anche alla probabilità  $\alpha = 0.001$  (uguale a 10,83); pertanto si rifiuta l'ipotesi nulla.

E' dimostrata una elevatissima significatività dell'associazione tra queste 2 variabili qualitative: fattore A e fattore B tendono ad essere presenti od assenti in modo congiunto.

Al fine di **permettere il confronto tra matrici di dimensioni differenti e per tenere in considerazione il numero di osservazioni** si può calcolare l'indice **V di Cramér**

$$V = \sqrt{\frac{16,898}{66 \cdot (2 - 1)}} = \sqrt{0,256} = 0,51$$

Il coefficiente di associazione V risulta uguale a 0,51; la sua significatività è quella del  $\chi^2$  con 1 gdl, come precedentemente stimato.

Senza il termine di correzione, il valore del  $\chi^2$  sarebbe stato

$$\chi^2 = \frac{(36 \cdot 16 - 5 \cdot 9)^2 \cdot 66}{41 \cdot 25 \cdot 45 \cdot 21} = \frac{531^2 \cdot 66}{968625} = \frac{281961 \cdot 66}{968625} = \mathbf{19,212}$$

uguale a 19,212 ed il corrispondente V di Cramér sarebbe risultato

$$V = \sqrt{\frac{19,212}{66 \cdot (2 - 1)}} = \sqrt{0,291} = 0,5395$$

uguale a 0,5395.

E' una ulteriore dimostrazione empirica dell'effetto cautelativo determinato dalla correzione per campioni, con un numero non sufficientemente elevato d'osservazioni.

Il valore V di Cramér permette il confronto tra campioni di dimensioni diverse come numero di osservazioni e/o numero di specie considerate (quindi dimensioni della tabella e gdl).

#### **14.6. IL CROSS-PRODUCT RATIO (CPR)**

Il valore dell'associazione tra due variabili qualitative o nominali dipende dalla formula del chi-quadrato,

$$\chi^2 = \sum \frac{(Oss.-Att.)^2}{Att.}$$

Di conseguenza, esso aumenta quando lo scarto tra osservato ed atteso è moltiplicato per una quantità **k**, anche se le frequenze delle varie classi restano uguali sia in percentuale che nel loro rapporto. Infatti, **moltiplicando con un fattore k sia le frequenze osservate che quelle attese**

$$\begin{aligned} \frac{(kOss.-kAtt.)^2}{kAtt.} &= \frac{k^2(Oss.-Att.)^2}{kAtt.} = \frac{k^2}{k} \frac{(Oss.-Att.)^2}{Att.} = \\ &= k \frac{(Oss.-Att.)^2}{Att.} \end{aligned}$$

**il valore del  $c^2$  aumenta di un identico fattore k.**

E' una situazione che, secondo vari statistici, rende poco chiari il significato e l'interpretazione della misura d'associazione. Pertanto, sono state proposte tavole di equivalenza per tabelle con lo stesso grado d'associazione.

Uno di questi indici d'associazione, che non variano quando tutti i dati della tabella sono moltiplicati per una costante, è il **cross- product ratio** o **CPR**.

Utilizzando la simbologia già utilizzata nel calcolo del  $c^2$  in tabelle 2 x 2, il valore del CPR è

$$CPR = \frac{a \cdot d}{b \cdot c}$$

Con i dati dell'ultimo esempio, il CPR

$$CPR = \frac{36 \cdot 16}{5 \cdot 9} = \frac{576}{45} = 12,8$$

risulta uguale a 12,8.

Il campo di variazione del CPR è da **0** a **∞**, in cui 1 corrisponde all'indipendenza o totale assenza di associazione tra i due fattori.

La distribuzione del CPR non è simmetrica. Infatti, un'associazione di pari intensità ma di segno contrario, una associazione negativa rispetto alla precedente supposta positiva, è determinata da valori di **a x d** maggiori di quelli **b x c**, in cui il risultato diviene 1/CPR.

Nell'esempio, il valore di associazione negativa corrispondente a quella stimata sarebbe 1/12,8 e quindi uguale a 0,078.

Di conseguenza sono state proposte **trasformazioni del cross-product ratio**, come la seguente

$$CPR' = \frac{2}{p} \arctan \ln CPR$$

che riconduce la variazione dei valori tra -1 e +1, con 0 che indica assenza di associazione.

**Anche per il CPR non esistono tabella di significatività; per la sua stima, si ricorre a quella del chi-quadrato.**

#### **14.7. ASSOCIAZIONE PER VARIABILI CATEGORIALI IN TABELLE M x N:**

##### **LA PRE, IL I SIMMETRICO ED ASIMMETRICO DI GOODMAN E KRUSKAL, LE MISURE FONDATE SUL CHI QUADRATO**

Le tabelle M x N sono costruite per tre grandi tipi di variabili: nominali, ordinali e d'intervallo.

- Nel primo caso, si parla di **associazione** e gli indici sono presentati in questo paragrafo;
- nel secondo caso, si parla di **cograduazione** e gli indici verranno presentati nel paragrafo successivo;
- nell'ultimo caso, la relazione tra le due variabili è chiamata **correlazione**.

I risultati ottenuti con i metodi precedenti, tutti fondati sul  $\chi^2$ , restano difficili da interpretare, anche dopo le trasformazioni proposte. In particolare quando i valori sono distanti da zero e quindi non si ha indipendenza tra le due variabili qualitative, non è chiaro il tipo di associazione. Per renderlo più evidente, nel 1954 Goodman e Kruskal hanno introdotto il concetto di **Riduzione Proporzionale nell'Errore**, abbreviato in **PRE** dai termini inglesi (vedi: Goodman L. A. and Kruskal E. H. *Measures of association for cross classification* su *J. Amer. Statist. Assoc.* n. 49, pp. 732 – 764).

Nella previsione delle frequenze con cui compare **una variabile nominale o categoriale**, è possibile utilizzare sia la sola conoscenza di quella variabile sia anche la conoscenza delle altre; la PRE è data dal rapporto fra le due misure dell'errore.

Si supponga di avere una tavola di contingenza 3 x 3

Livello d'inquinamento	Specie prevalente			Totale di riga
	A	B	C	
	<b>26</b>	<b>8</b>	<b>0</b>	<b>34</b>
ALTO	<i>0,181</i>	<i>0,056</i>	<i>0,000</i>	<i>0,237</i>
	<b>10</b>	<b>30</b>	<b>17</b>	<b>57</b>
MEDIO	<i>0,069</i>	<i>0,208</i>	<i>0,118</i>	<i>0,395</i>
	<b>0</b>	<b>6</b>	<b>47</b>	<b>53</b>
BASSO	<i>0,000</i>	<i>0,042</i>	<i>0,326</i>	<i>0,368</i>
	<b>36</b>	<b>44</b>	<b>64</b>	<b>144</b>
Totale di colonna	<i>0,250</i>	<i>0,306</i>	<i>0,444</i>	<i>1,000</i>

in cui è stato riportato quante volte in 144 laghi di una regione, classificati secondo il livello d'inquinamento (alto in 34 laghi, medio in 57 e basso in 53), sono state trovate come prevalenti le specie A (36 volte), B (44) e C (64), di difficile classificazione e note in letteratura per vivere in zone con un tasso d'inquinamento rispettivamente alto, medio e basso.

(In grassetto è riportato quante volte sono state trovate come prevalenti le 3 specie in laghi a differente livello d'inquinamento; in corsivo, la proporzione relativa.)

La specie che risulta prevalente con frequenza maggiore, **la categoria modale** tra le specie, è la C con 64 presenze su 144 casi, pari ad una frequenza relativa di 0,444.

**La stima della probabilità di una classificazione non corretta  $P_1$  delle specie prevalenti, utilizzando solo le informazioni sulla loro morfologia, è 1 meno la probabilità della categoria modale**

$$P_1 = 1 - 0,444 = 0,556$$

Poiché le specie sono prevalenti in ambienti diversi, **per ridurre l'errore e migliorare la loro classificazione è possibile utilizzare anche l'informazione sul livello d'inquinamento.**

Per ogni gruppo di laghi, classificato sulla base del livello d'inquinamento, la categoria risultante è quella con frequenza maggiore per quel livello; in altri termini, in laghi con alto inquinamento è prevista come prevalente la specie A, in laghi con inquinamento medio la specie B e in quelli con inquinamento basso la specie C. **La probabilità di errore nella classificazione delle specie prevalenti, quando viene usato anche il livello d'inquinamento del lago, è data dalla somma**

delle probabilità di tutte le celle sulla stessa riga e colonna delle celle in questione, esclusa la probabilità relativa alla cella stessa:

$$P_2 = 0,056 + 0,069 + 0,118 + 0,042 = 0,285$$

Utilizzando anche l'informazione derivante dalla classificazione del livello d'inquinamento, la probabilità d'errore nella classificazione della specie diminuisce da  $P_1 = 0,556$  a  $P_2 = 0,285$ .

**Il lambda ( $\lambda$ ) di Goodman e Kruskal misura la riduzione proporzionale nell'errore, sulla base della relazione**

$$\lambda = \frac{P_1 - P_2}{P_1}$$

Con i dati dell'esempio,

$$\lambda = \frac{0,556 - 0,285}{0,556} = 0,487$$

si ottiene esattamente una riduzione pari a 48,7% dell'errore nella classificazione della specie prevalente, quando si utilizza anche l'informazione derivante dal livello d'inquinamento.

Una formula abbreviata per calcolare lo stesso valore è

$$\lambda = \frac{\sum_{j=1}^k f_{\max} - C_{\max}}{N - C_{\max}}$$

dove

**$f_{\max}$**  è la frequenza maggiore in ogni colonna

**$C_{\max}$**  è il totale per colonna maggiore

**$N$**  è il totale generale

Con i dati dell'esempio,

$$\lambda = \frac{(26 + 30 + 47) - 64}{144 - 64} = \frac{39}{80} = 0,4875$$

si ottiene 0,4875 senza gli arrotondamenti prima necessari.

**Il valore di  $\lambda$  varia sempre da 0 a 1** Il valore 0, che si ottiene quando le frequenze entro ogni casella sono distribuite a caso (calcolabili attraverso il prodotto dei totali di riga e di colonna diviso il totale generale), indica che la variabile indipendente non aggiunge informazioni nella previsione della variabile dipendente e che pertanto non può essere utile nella sua classificazione. Un valore uguale a 1 indica che esiste corrispondenza perfetta e quindi che la variabile dipendente è classificata

correttamente anche dalla variabile indipendente (le specie A, B e C sono rispettivamente presenti sempre e soltanto in laghi con inquinamento alto, basso e medio).

Come già espresso in altre occasioni, non esiste corrispondenza biunivoca tra il valore 0 del lambda e l'associazione tra le due variabili: quando le due variabili sono indipendenti lambda è uguale a 0; ma quando lambda risulta uguale a 0 non sempre si ha indipendenza statistica. **L'indice lambda deve essere usato solo in condizioni particolari di analisi dell'associazione: quando i valori di una variabile qualitativa sono utilizzati per prevedere quelli dell'altra variabile** (anche se, come nell'esempio, una variabile è qualitativa e l'altra di rango).

Come è stato proposto il livello d'inquinamento di un lago per predire la presenza della specie prevalente, nello stesso modo è possibile utilizzare la presenza della specie prevalente per indicare il livello d'inquinamento. E' quindi possibile **calcolare un altro valore di lambda, scambiando le righe con le colonne, il previsore con la variabile predetta**. Salvo casi fortuiti, di norma i diversi approcci danno risultati differenti. Il **lambda** presentato è **asimmetrico**; è quindi importante scegliere la variabile dipendente adatta.

In vari casi, come nell'esempio, non è possibile o semplice distinguere tra variabile dipendente ed indipendente. Viene quindi utilizzato un **lambda simmetrico**, in cui le variabili di riga e di colonna hanno le stesse frequenze.

Per spiegare questi concetti con una serie di esempi dettagliati, vengono riportati quelli già utilizzati da **Graham J. G. Upton** (*The Analysis of Cross-tabuled Data*, John Wiley & Sons, Chichester – New York, 1978, reprinted April 1980, da pag. 30 a 32).

Data una tabella M x N, con le variabili qualitative A e B, come quella riportata,

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	Totale
A <sub>1</sub>	10	5	18	20	53
A <sub>2</sub>	8	16	5	13	42
A <sub>3</sub>	11	7	3	4	25
Totale	29	28	26	37	120

$\lambda_b$  stima la diminuzione relativa della probabilità d'errore nell'indovinare la categoria B, utilizzando anche la classificazione di A invece del solo totale marginale di B.

I dati che servono (indicati con l'asterisco e in grassetto nella tabella successiva) sono

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	Totale
A <sub>1</sub>	10	5	18	<b>20*</b>	53
A <sub>2</sub>	8	<b>16*</b>	5	13	42
A <sub>3</sub>	<b>11*</b>	7	3	4	25
Totale	29	28	26	<b>37*</b>	<b>120*</b>

- i valori maggiori in ognuna delle 3 righe A (A<sub>1</sub> = 20; A<sub>2</sub> = 16; A<sub>3</sub> = 11),
- il totale maggiore fra le 4 colonne B (37),
- il totale generale (120).

Da essi si ricava  $\lambda_b$

$$\lambda_b = \frac{(20+16+11)-37}{120-37} = \frac{10}{83} = 0,120$$

che risulta uguale a 0,120.

Invertendo i concetti sulla previsione di A, i dati che servono per stimare  $\lambda_a$  (indicati con l'asterisco e in grassetto nella tabella successiva) sono

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	Totale
A <sub>1</sub>	10	5	<b>18*</b>	<b>20*</b>	<b>53*</b>
A <sub>2</sub>	8	<b>16*</b>	5	13	42
A <sub>3</sub>	<b>11*</b>	7	3	4	25
Totale	29	28	26	37	<b>120*</b>

- i valori maggiori in ognuna delle 4 colonne B (B<sub>1</sub> = 11; B<sub>2</sub> = 16; B<sub>3</sub> = 18; B<sub>4</sub> = 20),
- il totale maggiore fra le 3 righe A (53),
- il totale generale (120).

Da essi si ricava  $\lambda_a$

$$\lambda_a = \frac{(11+16+18+20)-53}{120-53} = \frac{12}{67} = 0,179$$

che risulta uguale a 0,179.

Da  $\lambda_a$  e  $\lambda_b$  si ricava  $\lambda$

$$\lambda = \frac{12+10}{67+83} = \frac{22}{150} = 0,15$$

che risulta uguale a 0,15 un valore compreso tra  $\lambda_a$  e  $\lambda_b$ .

Il lambda utilizza i casi presenti nella matrice. **Goodman e Kruskal hanno proposto anche altri metodi, come il t (tau), che per calcolare le probabilità d'errore utilizzano i totali di riga o di colonna.** Questo metodo richiede calcoli più lunghi, nei quali è più facile commettere errori. Di conseguenza, esso non è presentato in modo dettagliato.

Quando i dati sono distribuiti in modo indipendente, anche il valore di **tau** è pari a 0.

Per verificare se il valore ottenuto di **L** o di **t** è significativo si deve ricorrere ad una distribuzione campionaria complessa, che è approssimativamente normale quando **N** è relativamente grande.

Tra gli altri limiti per verificare la significatività, è da considerare pure che non è possibile testare l'ipotesi che la PRE sia uguale a 0, quindi rifiutare l'ipotesi nulla, oppure che sia uguale a 1, e quindi accettare l'ipotesi di una associazione completa.

Per stimare la significatività del **L** e/o **t** calcolato, si utilizza il valore del  $c^2$  calcolato sulla matrice originaria di dimensioni **m x n**, che ovviamente **ha gdl pari a (m-1) x (n-1)**.

Anche nell'ultima tabella riportata è possibile applicare **le misure di associazione che utilizzano il chi quadrato.**

Poiché esso risulta uguale a 19,55 utilizzando le stesse formule presentate nel paragrafo per le tabelle 2 x 2 si ottiene:

- l'indice **f** (phi) di Pearson

$$f = \sqrt{\frac{c^2}{N}} = \sqrt{\frac{19,55}{120}} = \sqrt{0,163} = 0,403$$

risulta uguale a 0,403

- il **coefficiente C**

$$C = \sqrt{\frac{c^2}{c^2 + N}} = \sqrt{\frac{19,55}{19,55 + 120}} = \sqrt{0,14} = 0,374$$

risulta uguale a 0,374

- il coefficiente **V di Cramér**

$$V = \sqrt{\frac{c^2}{N \cdot (k-1)}} = \sqrt{\frac{19,55}{120 \cdot (3-1)}} = \sqrt{0,0815} = 0,285$$

risulta uguale a 0,285

- il coefficiente **D<sub>T</sub> (o T) di Tschuprow**

$$D_T = \sqrt{\frac{c^2}{N \sqrt{(r-1) \cdot (c-1)}}} = \sqrt{\frac{19,55}{120 \sqrt{2 \cdot 3}}} = \sqrt{\frac{19,55}{293,84}} = \sqrt{0,0665} = 0,258$$

risulta uguale a 0,258.

#### 14.8. COGRADUAZIONE PER VARIABILI ORDINALI IN TABELLE M x N:

**IL g di GOODMAN E KRUSKALL, il t<sub>K</sub> DI KENDALL, IL d<sub>ba</sub> E IL d<sub>ab</sub> DI SOMERS**

In una **tabella M x N, in cui le due variabili siano di tipo ordinale**, l'associazione viene chiamata con il nome specifico di **cograduazione**. Anche in questo caso, il punto di riferimento sono le due diagonali.

Quando la tabella M x N è impostata come quella successiva

	Valore Basso	VARIABILE 1		Valore Alto
Valore Basso	<i>a</i>			<i>b</i>
VARIABILE 2				
Valore Alto	<i>c</i>			<i>d</i>

- la diagonale dai valori bassi verso quelli alti (**a – d**) è chiamata **diagonale della cograduazione**
- la diagonale opposta (**c – b**) è chiamata **diagonale della contro-graduazione**.

L'indice di **cograduazione** più diffuso è il **g di Goodman e Kruskal** (1954) che ha una formula analoga al coefficiente bidirezionale Q di Yule per tabelle 2 x 2

$$Q = \frac{ad - bc}{ad + bc}$$

ovviamente estesa ad una tabella M x N

$$g = \frac{S - D}{S + D}$$

La difficoltà maggiore nella comprensione dell'indice è capire come sono calcolati S e D.

Più della definizione

- S = somma delle coppie cograduate: numero totale di coppie di osservazioni in cui si abbiano sia  $i > i'$  e  $j > j'$  oppure entrambi  $i < i'$  e  $j < j'$

- D = somma delle coppie contr graduate: numero totale di coppie di osservazioni in cui si abbiano sia  $i > i'$  e  $j < j'$  oppure entrambi  $i < i'$  e  $j > j'$

è utile un esempio a partire dalla tabella

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	Totale
A <sub>1</sub>	10	5	18	20	53
A <sub>2</sub>	8	16	5	13	42
A <sub>3</sub>	11	7	3	4	25
Totale	29	28	26	37	120

in cui, A e B in questo caso devono essere due variabili ordinali, ranghizzate in modo crescente.

S è determinato attraverso la somma di più prodotti  $S_i$ , che nel caso specifico della tabella in oggetto sono sei, dove

$$1) S_1 = 10 (16 + 5 + 13 + 7 + 3 + 4) = 10 (48) = 480$$

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	Totale
A <sub>1</sub>	10*	5	18	20	53
A <sub>2</sub>	8	16*	5*	13*	42
A <sub>3</sub>	11	7*	3*	4*	25
Totale	29	28	26	37	120

$$2) S_2 = 5 (5 + 13 + 3 + 4) = 5 (25) = 125$$

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	Totale
A <sub>1</sub>	10	<b>5*</b>	18	20	53
A <sub>2</sub>	8	16	<b>5*</b>	<b>13*</b>	42
A <sub>3</sub>	11	7	<b>3*</b>	<b>4*</b>	25
Totale	29	28	26	37	120

$$3) S_3 = 18 (13 + 4) = 18 (17) = 306$$

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	Totale
A <sub>1</sub>	10	5	<b>18*</b>	20	53
A <sub>2</sub>	8	16	5	<b>13*</b>	42
A <sub>3</sub>	11	7	3	<b>4*</b>	25
Totale	29	28	26	37	120

$$4) S_4 = 8 (7 + 3 + 4) = 8 (14) = 112$$

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	Totale
A <sub>1</sub>	10	5	18	20	53
A <sub>2</sub>	<b>8*</b>	16	5	13	42
A <sub>3</sub>	11	<b>7*</b>	<b>3*</b>	<b>4*</b>	25
Totale	29	28	26	37	120

$$5) S_5 = 16 (3 + 4) = 16 (7) = 112$$

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	Totale
A <sub>1</sub>	10	5	18	20	53
A <sub>2</sub>	8	<b>16*</b>	5	13	42
A <sub>3</sub>	11	7	<b>3*</b>	<b>4*</b>	25
Totale	29	28	26	37	120

$$6) S_6 = 5 (4) = 20$$

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	Totale
A <sub>1</sub>	10	5	18	20	53
A <sub>2</sub>	8	16	<b>5*</b>	13	42
A <sub>3</sub>	11	7	3	<b>4*</b>	25
Totale	29	28	26	37	120

$$S = \sum S_i = 480 + 125 + 306 + 112 + 112 + 20 = 1155$$

In modo esattamente simmetrico, il valore di D è ottenuto a partire dal valore in alto a destra

$$1) D_1 = 20 (8 + 16 + 5 + 11 + 7 + 3) = 20 (50) = 1000$$

$$2) D_2 = 18 (8 + 16 + 11 + 7) = 18 (42) = 756$$

$$3) D_3 = 5 (8 + 11) = 5 (19) = 95$$

$$4) D_4 = 13 (11 + 7 + 3) = 13 (21) = 273$$

$$5) D_5 = 5 (11 + 7) = 5 (18) = 90$$

$$6) D_6 = 16 (11) = 176$$

$$D = \sum D_i = 1000 + 756 + 95 + 273 + 90 + 176 = 2390$$

Da S e D si ricava  $\gamma$

$$g = \frac{1155 - 2390}{1155 + 2390} = \frac{-1235}{3545} = -0,348$$

che risulta uguale a **-0,348**; il segno negativo sta ad indicare che a valori bassi di A sono associati valori alti di B, come evidenzia la tabella dei dati.

Per la **cograduazione Kendall** ha proposto un indice  $t$ ; per distinguerlo da altri metodi che utilizzano sempre il simbolo  $t$ , su vari testi spesso è indicato con  $t_K$ .

Esso è dato da

$$t_K = \frac{2 \cdot (S - D)}{\sqrt{(S + D + T_a) \cdot (S + D + T_b)}}$$

in cui

- **S** e **D** sono uguali alla formula precedente ed ovviamente calcolati nello stesso modo,
- **T<sub>a</sub>** = numero totale di coppie di osservazioni in cui **i = i'**: è la somma dei prodotti di ogni valore per la somma di quelli che sulla stessa riga stanno alla sua destra, ovviamente a partire dalla prima colonna;
- **T<sub>b</sub>** = numero totale di coppie di osservazioni in cui **j = j'**: è la somma dei prodotti di ogni valore per la somma di quelli che stanno sotto di lui, nella stessa colonna, a partire dalla prima riga.

Dalla stessa tabella

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	Totale
A <sub>1</sub>	10	5	18	20	53
A <sub>2</sub>	8	16	5	13	42
A <sub>3</sub>	11	7	3	4	25
Totale	29	28	26	37	120

**T<sub>a</sub>** è la somma di 9 valori, dei quali vengono riportati dettagliatamente tutti i calcoli, come spiegazione chiara del metodo:

- 1)  $T_{a1} = 10 (5 + 18 + 20) = 10 (43) = 430$
- 2)  $T_{a2} = 5 (18 + 20) = 5 (38) = 190$
- 3)  $T_{a3} = 18 (20) = 360$
- 4)  $T_{a4} = 8 (16 + 5 + 13) = 8 (34) = 272$
- 5)  $T_{a5} = 16 (5 + 13) = 16 (18) = 288$
- 6)  $T_{a6} = 5 (13) = 65$
- 7)  $T_{a7} = 11 (7 + 3 + 4) = 11 (14) = 154$
- 8)  $T_{a8} = 7 (3+4) = 7 (7) = 49$
- 9)  $T_{a9} = 3 (4) = 12$

$$T_a = \sum T_a i = 430 + 190 + 360 + 272 + 288 + 65 + 154 + 49 + 12 = 1820$$

**T<sub>b</sub>** (sempre in questa tabella specifica) è la somma di 8 valori:

- 1)  $T_{b1} = 10 (8 + 11) = 10 (19) = 190$

- 2)  $T_{b2} = 8 (11) = 88$
- 3)  $T_{b3} = 5 (16 + 7) = 5 (23) = 115$
- 4)  $T_{b4} = 16 (7) = 112$
- 5)  $T_{b5} = 18 (5+3) = 18 (8) = 144$
- 6)  $T_{b6} = 5 (3) = 15$
- 7)  $T_{b7} = 20 (13 + 4) = 20 (17) = 340$
- 8)  $T_{b8} = 13 (4) = 52$

$$T_b = \sum T_b i = 190 + 88 + 115 + 112 + 144 + 15 + 340 + 52 = 1056$$

Applicando la formula di Kendall, si ottiene

$$t_k = \frac{2 \cdot (1155 - 2390)}{\sqrt{(1155 + 2390 + 1820) \cdot (1155 + 2390 + 1056)}} \frac{2 \cdot (-1235)}{\sqrt{5365 \cdot 4601}} = \frac{-2470}{4968,3} = -0,497$$

un valore di  $\tau_K$  uguale a  $-0,497$ .

La quantità di calcoli richiesti, anche se semplici, possono determinare un errore. E' quindi stata proposta una verifica della correttezza di tutti i parametri considerati

$$N^2 = 2 \cdot (S + D + T_a + T_b) + \sum_i \sum_j x_{ij}$$

dove

- $N^2$  è il quadrato della somma di tutti i dati della tabella,
- $S, D, T_a$  e  $T_b$  sono i 4 valori utilizzati per la stima dell'indice,
- $\sum_i \sum_j x_{ij}$  è la somma dei quadrati di tutti i singoli valori della tabella.

Con i dati della tabella utilizzata,

$$\sum_i \sum_j x_{ij} \text{ è}$$

	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	Totale
A <sub>1</sub>	<b>100</b>	<b>25</b>	<b>324</b>	<b>400</b>	---
A <sub>2</sub>	<b>64</b>	<b>256</b>	<b>25</b>	<b>169</b>	---
A <sub>3</sub>	<b>121</b>	<b>49</b>	<b>9</b>	<b>16</b>	---
Totale	---	---	---	---	1558

uguale a 1558; di conseguenza, con

$$N = 120; \quad S = 1155; \quad D = 2390; \quad T_a = 1820; \quad T_b = 1056$$

si dimostra

$$120^2 = 2 (1155 + 2390 + 1820 + 1056) + 1558 = 14400$$

l'uguaglianza delle due quantità (entrambe danno un risultato di 14400), a testimonianza della correttezza dei parametri calcolati.

Nel 1962 **R. H. Somers** (con lo stesso articolo citato per il D asimmetrico, *A new asymmetric measure of association for ordinal variables*, pubblicato su *American Sociological Review* 27, n.6, pp.700-811) ha proposto anche **indici ( $d_{ba}$  e  $d_{ab}$ ) di cograduazione asimmetrici o unidirezionali** per tabelle M x N, da applicare nel caso di variabili ordinali.

Quando **B è la variabile dipendente** e A la variabile indipendente, si può stimare  $d_{ba}$  con

$$d_{ba} = \frac{S - D}{S + D + T_b}$$

Questa statistica è distribuita in modo approssimativamente normale. La sua varianza è stata stimata da L. A. Goodman e E. H. Kruskal nel 1972 (con l'articolo *Measures of association for cross-classification*, pubblicata da *J. Amer. Statist. Assoc.*, n.67, pp. 415-421).

Quando **A è la variabile dipendente**, si stima  $d_{ab}$  con

$$d_{ab} = \frac{S - D}{S + D + T_a}$$

con formula simmetrica.

Utilizzando sempre la stessa tabella, con  $S = 1155$ ;  $D = 2390$ ;  $T_a = 1820$ ;  $T_b = 1056$

-  $d_{ba}$

$$d_{ba} = \frac{1155 - 2390}{1155 + 2390 + 1056} = \frac{-1235}{4601} = -0,268$$

risulta uguale a -0,268 e

-  $d_{ab}$

$$d_{ab} = \frac{1155 - 2390}{1155 + 2390 + 1820} = \frac{-1235}{5365} = -0,230$$

uguale a  $-0,230$ .

In un confronto tra questi diversi indici, **Graham J. G. Upton** (nel suo volume *The analysis of cross-tabulated data*, pubblicato da *John Wiley & Sons* Chichester nel 1978, a pag. 38) consiglia, giustificandola come pura scelta personale:

- per dati nominali, il  $I$  di Goodman e Kruskal,
- per dati ordinali, il  $g$  se le due variabili sono di importanza uguale,
- rispettivamente il  $I_b$  di Goodman e Kruskal e il  $d_{ba}$  di Somers, se la variabile B dipende dalla variabile A.

#### 14.9. STIMA DELL'ACCORDO CON SCALA NOMINALE: IL KAPPA DI COHEN

Nella ricerca ambientale, in varie situazioni ad un ricercatore è chiesto di **classificare una serie di oggetti, piante, animali o persone in gruppi qualitativi o nominali**. E' il caso della classificazione in specie di una serie di organismi tra loro molto simili, dell'attribuzione ad una determinata malattia attraverso i sintomi manifestati da alcuni individui, dell'assegnazione ad un gruppo categoriale di situazioni ambientali complesse.

Due persone diverse possono arrivare in vari casi a classificazioni identiche e in altri a classificazioni differenti: **il kappa di Cohen è una misura dell'accordo tra le risposte di due persone o della stessa persona che valuta coppie di oggetti**.

Si supponga che due ricercatori debbano classificare in tre gruppi qualitativi (**A, B, C**) 109 situazioni complesse ed abbiano ottenuto il risultato riportato in grassetto nella tabella seguente

Ricercatore 2	Ricercatore 1			Totale di riga
	A	B	C	
<b>A</b>	<b>15</b>	<b>8</b>	<b>7</b>	<b>30</b>
	<i>8,81</i>	<i>10,73</i>	<i>10,46</i>	<i>30,00</i>
<b>B</b>	<b>12</b>	<b>25</b>		<b>37</b>
	<i>10,86</i>	<i>13,24</i>	<i>12,90</i>	<i>37,00</i>
<b>C</b>	<b>5</b>	<b>6</b>	<b>31</b>	<b>42</b>
	<i>12,33</i>	<i>15,03</i>	<i>14,64</i>	<i>42,00</i>
Totale di colonna	<b>32</b>	<b>39</b>	<b>38</b>	<b>109</b>
	<i>32,00</i>	<i>39,00</i>	<i>38,00</i>	<i>109,00</i>

Anche se i due ricercatori seguissero criteri totalmente differenti, per caso alcuni giudizi sarebbero coincidenti: la loro frequenza attesa (ottenuta con il solito metodo delle tabelle m x n nel chi quadrato: totale di riga x totale di colonna / totale generale) è riportata in corsivo.

Una stima dell'accordo tra i due ricercatori può essere **la differenza tra la proporzione osservata di casi e la proporzione attesa**, nell'ipotesi che tra i due non esista alcun accordo, ma solo risposte coincidenti per solo effetto del caso.

Per rendere i dati indipendenti dal numero di osservazioni, si devono tradurre i valori osservati e quelli attesi in proporzioni

Ricercatore 2	Ricercatore 1			Totale di riga
	A	B	C	
A	<b>0,1376</b> <i>0,0808</i>	<b>0,0734</b> <i>0,0984</i>	<b>0,0642</b> <i>0,0960</i>	
B	<b>0,1101</b> <i>0,0996</i>	<b>0,2234</b> <i>0,1215</i>	<b>0,0000</b> <i>0,1184</i>	
C	<b>0,0459</b> <i>0,1131</i>	<b>0,0550</b> <i>0,1379</i>	<b>0,2844</b> <i>0,1343</i>	
Totale di colonna				<b>1,000</b> <i>1,000</i>

**La proporzione di casi valutati nello stesso modo** dai due ricercatori è determinata dalla somma dei valori osservati che si riferiscono agli stessi gruppi (0,1379 + 0,2234 + 0,2844); con i dati dell'esempio, risulta uguale a 0,6457 o 64,57%.

**La quota attribuita al caso** è data dalla somma dei valori attesi corrispondenti (0,0808 + 0,1215 + 0,1343) e risulta uguale a 0,3366 o 33,66%.

**La differenza tra le due quote** (0,6457 - 0,3366) è uguale a 0,3091 o 30,91% e **rappresenta una misura dell'accordo reale** tra i due ricercatori.

**Il kappa di Cohen standardizza questa differenza dividendola per la massima differenza possibile non casuale.** Se quella casuale è 0,3366, quella non casuale può al massimo essere (1 - 0,3366) e quindi in questo caso uguale a 0,6634.

Il valore di kappa è uguale a  $0,237 / (1 - 0,3366) = 0,357$  o 35,7%.

La formula generale per stimare  $k$  può essere espressa con

$$k = \frac{Poss. - Patt.}{1 - Patt.}$$

dove

*Poss.* è la proporzione di frequenze osservate di accordi tra i due valutatori,

*Patt.* è la proporzione di accordi attesi nella condizione che sia vera l'ipotesi nulla.

Della statistica kappa di Cohen è stata fornita anche una generalizzazione estesa a più di due persone.

Per la stima della significatività nel caso di 2 giudici e per il test in caso di  $N$  giudici si rinvia a trattazioni più complete.

#### 14.10. LETTURA DEI TABULATI DI UN PACCHETTO STATISTICO

Gli ultimi test presentati e gli indici di associazione tra variabili di tipo diverso sono stati presentati in modo schematico. Per essi la spiegazione è stata limitata ai concetti fondamentali, sufficienti per comprendere i tabulati dei programmi informatici, non per essere effettivamente operativi con calcoli manuali e l'aiuto solo di una calcolatrice da tavolo, come fatto per la quasi totalità dei test presentati.

Per guidare alla **scelta dei test più appropriati** e per la **corretta interpretazione dei risultati**, sono stati riportati i tabulati di un programma informatico ad ampia diffusione.

L'esercizio è una tabella 2 x 2, impostata per valutare il grado di associazione tra due variabili.

	<b>A</b>			
	+	-	tot	
<b>B</b>	+	<b>36</b>	<b>5</b>	41
	-	<b>9</b>	<b>16</b>	25
	tot	45	21	66

Poiché **non è fornita alcuna informazione specifica sul tipo di campionamento** (2 campioni dipendenti o indipendenti) e **sulle variabili utilizzate** (qualitativa, ordinale, di intervallo e loro combinazioni), gli stessi dati possono essere letti in modi differenti.

Per esempio, la stessa tabella potrebbe essere il confronto tra due campioni indipendenti con risposte qualitative

<b>Risposta</b>			
<b>Campioni</b>	<b>+</b>	<b>-</b>	<b>tot</b>
<b>A</b>	<b>36</b>	<b>5</b>	<b>41</b>
<b>B</b>	<b>9</b>	<b>16</b>	<b>25</b>
<b>tot</b>	<b>45</b>	<b>21</b>	<b>66</b>

su cui sarebbe da applicare il test  $\chi^2$  con la correzione di Yates o il test G oppure il metodo esatto di Fisher, in rapporto al numero di osservazioni, in totale ed entro ogni casella.

Se ottenute come risultato di un esperimento tra due campioni dipendenti con risposte qualitative, come

<b>PRIMA</b>			
	<b>+</b>	<b>-</b>	<b>tot</b>
<b>DOPO</b>			
<b>+</b>	<b>36</b>	<b>5</b>	<b>41</b>
<b>-</b>	<b>9</b>	<b>16</b>	<b>25</b>
<b>tot</b>	<b>45</b>	<b>21</b>	<b>66</b>

in cui un gruppo di 66 individui è stato classificato in due categorie binarie, prima e dopo l'esperimento, è possibile applicare il test di McNemar.

Quando 2 variabili a confronto forniscono risposte quantitative

<b>Variabile A</b>			
	<b>Alti</b>	<b>Bassi</b>	<b>tot</b>
<b>Variabile B</b>			
<b>Alti</b>	<b>36</b>	<b>5</b>	<b>41</b>
<b>Bassi</b>	<b>9</b>	<b>16</b>	<b>25</b>
<b>tot</b>	<b>45</b>	<b>21</b>	<b>66</b>

può essere utile applicare misure di correlazione.

**Le risposte che vengono fornite dai tabulati sono quindi numerose**, come le seguenti; tra esse **occorre scegliere quelle esplicative del problema posto**; cioè capire quali sono le risposte utili al problema specifico che spesso non è possibile porre al computer, poiché è stato impostato per rispondere a tutte le domande che è possibile porre con quella tabella di dati.

L'elenco di test qui riportato è il tabulato di un programma informatico, applicato ai dati prima presentati:

Test	Value	df	Exact Sig. (2-sided)	Exact Sig. (1-sided)
$\chi^2$ di Pearson	19.212	1	.000	.000
Continuity Correction	16.898	1		
Likelihood Ratio	19.489	1	.000	.000
Fisher's Exact			.000	.000
Linear by Linear Association	18.921	1	.000	.000
Mc Nemar			.424	.212

Le due tabelle seguenti sono altri due esempi di risultati di analisi che sarebbe possibile applicare agli stessi dati. Alcuni test, come la correlazione di Spearman, sono spiegati nel capitolo seguente

Data	Test	Value	Std Error	Approx. Sig.	
Nominal By Nominal	Lambda symmetric	.391	.155	.035	
		var. 1 dependent	.440	.137	.012
		var. 2 dependent	.333	.194	.155
	Goodman-Kruskal t	var.1 dependent	.291	.115	.000(a)
		var. 2 dependent	.291	.115	.000(a)
Ordinal By Ordinal	Somers' d symmetric	.539	.107	.000	
		var. 1 dependent	.562	.110	.000
		var. 2 dependent	.518	.109	.000
Nominal By Interval	Eta	var. 1 dependent	.540		
		var. 2 dependent	.540		

a)chi-square probability

Data	Test	Value	Std Error	Approx. Sig.
Nominal	Phi	.540		
By Nominal	Cramer's V	.540		
	Contingency Coefficient	.475		
Ordinal	Kendall's tau-b	.540	.107	.000
By Ordinal	Kendall's tau-c	.488	.107	.000
	Gamma	.855	.085	.000
	Spearman Correlation	.540	.107	.000(b)
Interval By Interval	Pearson's r	.540	.107	.000(b)
Measure of Agreement	Kappa	.535	.108	.000

b) based on normal approximation

		95% Confidence Interval	
Test	Value	Lower	Upper
Odds Ratio	12.800	3.698	44.309

## CAPITOLO XV

### TEST NON PARAMETRICI PER CORRELAZIONE, CONCORDANZA E REGRESSIONE LINEARE

#### 15.1. LA CORRELAZIONE NON PARAMETRICA $\rho$ (rho) DI SPEARMAN

##### CON LA DISTRIBUZIONE DI HOTELLING-PABST E IL TEST DI DANIELS

Il **procedimento di correlazione basato sui ranghi**, proposto da C. Spearman con l'articolo *The proof and measurement of association between two things* (pubblicato nel 1904 su *Amer. Journal Psychol.* n. 15) e con *A footnote for measuring correlation* (pubblicato nel 1906 *Brit. Journ. Psychol.* n.2) è tra i test più ampiamente diffusi, per lo studio dell'associazione tra due variabili quantitative. Questa metodologia è stata ampiamente discussa già negli anni '20, come può dimostrare l'articolo di W. S. Gosset (Student) "An experimental determination of the probable error of Dr. Spearman's correlation coefficients", comparso su "Biometrika" n. 13 nel 1921.

In riferimento ai test di correlazione non parametrica, in letteratura sono ricorrenti i termini di **correlazione tra ranghi, cograduazione, associazione e concordanza**, usati come sinonimi, anche se i primi due dovrebbero essere utilizzati con scale almeno ordinali e le ultime due con scale qualitative o categoriali.

Il coefficiente di correlazione di Spearman è sovente indicato con il simbolo greco  $\rho$  (**rho**); in altri testi è simboleggiato con  $r_s$ , per evidenziare la sua affinità con il test  $r$  di **Pearson** dal quale è derivato.

Il metodo richiede che **entrambe le variabili siano misurate su una scala almeno ordinale**. Può essere più potente del test  $r$  di **Pearson** anche per scale d'intervallo o di rapporto, quando le condizioni di validità del test parametrico non sono pienamente soddisfatte. Di conseguenza, come per altri test non parametrici, è consigliabile la sua utilizzazione insieme con il test parametrico, ad ulteriore dimostrazione e verifica delle conclusioni raggiunte; in particolare quando si disponga solo di pochi dati e pertanto non sia possibile dimostrare che le condizioni di validità del test parametrico sono soddisfatte in modo completo.

Il **coefficiente di correlazione per ranghi di Spearman** serve per verificare l'**ipotesi nulla dell'indipendenza tra due variabili**, nel senso che gli  $N$  valori della variabile  $Y$  hanno le stesse probabilità di associarsi con ognuno degli  $N$  valori di  $X$ . L'ipotesi alternativa di esistenza di una associazione può prevedere un risultato **positivo oppure negativo**. Nel primo caso è detta **associazione diretta**: le coppie di valori sono contemporaneamente alti o bassi sia per  $X$  che per  $Y$ ; nel secondo caso, chiamata anche **associazione indiretta**, a valori alti di  $X$  corrispondono valori bassi di  $Y$  o viceversa.

Per una illustrazione didattica chiara, i vari passaggi richiesti dal metodo proposto da **Spearman** possono essere suddivisi in 5 fasi.

1 - Dopo l'enunciazione dell'ipotesi nulla

$$H_0: r = 0$$

e di quella alternativa, che può essere

- sia bilaterale

$$H_1: r \neq 0$$

- sia unilaterale

$$H_1: r > 0 \quad \text{oppure} \quad H_1: r < 0$$

è utile presentare i dati come nella tabella seguente

Variabili	Coppie di valori osservati						
X	8	5	7	14	22	21	41
Y	12	3	2	10	25	19	22
Soggetti	A	B	C	D	E	F	G

2 - Successivamente, occorre **ordinare i ranghi della variabile X**, assegnando **1** al valore più piccolo e progressivamente valori interi maggiori, fino ad N per il valore più alto.

Se i dati della variabile **X** hanno due o più valori uguali, è necessario assegnare ad ognuno di essi come rango la media delle loro posizioni.

Variabili	Coppie di valori osservati						
X	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
Y	3	2	12	10	25	16	22
Soggetti	<b>B</b>	<b>C</b>	<b>A</b>	<b>D</b>	<b>F</b>	<b>E</b>	<b>G</b>

Anche se ininfluenza ai fini dei calcoli successivi, è utile alla comprensione della misura di correlazione porre nell'ordine naturale (da **1** a **N**) i ranghi della variabile **X** e spostare la collocazione dei valori di **Y** relativi al medesimo soggetto, come nella tabella sovrastante.

3 - Sostituire anche gli **N** valori di **Y** con i ranghi rispettivi; per valori di **Y** uguali, usare la media dei loro ranghi:

Variabili	Coppie di valori osservati						
<b>X</b>	1	2	3	4	5	6	7
<b>Y</b>	<b>2</b>	<b>1</b>	<b>4</b>	<b>3</b>	<b>7</b>	<b>5</b>	<b>6</b>
Soggetti	B	C	A	D	F	E	G

4 - Se le due distribuzioni sono correlate in modo positivo ( $r = +1$ ), i valori della variabile **X** e della **Y** relativi allo stesso soggetto saranno uguali;

se le due distribuzioni sono correlate in modo negativo ( $r = -1$ ), a valori alti di **X** saranno associati valori bassi di **Y** e viceversa;

se tra le due variabili non esiste correlazione ( $r = 0$ ), i valori di **X** e di **Y** relativi agli stessi soggetti saranno associati in modo casuale.

Per quantificare questo grado di correlazione o concordanza, Spearman ha proposto **la distanza tra le coppie dei ranghi** ( $d_{R_i}$ )

Variabili	Coppie di valori osservati						
<b>X</b>	1	2	3	4	5	6	7
<b>Y</b>	2	1	4	3	7	5	6
<b>R<sub>i</sub></b>	<b>-1</b>	<b>+1</b>	<b>-1</b>	<b>+1</b>	<b>-2</b>	<b>+1</b>	<b>+1</b>
<b>R<sub>i</sub><sup>2</sup></b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>4</b>	<b>1</b>	<b>1</b>

elevate al quadrato

$$\sum d_{R_i}^2$$

Con i dati dell'esempio, la somma delle  $d_{R_i}^2$  è uguale a **10**

5 - Quando  $r = +1$ , le coppie di osservazioni di **X** e **Y** hanno lo stesso rango e pertanto questa sommatoria è uguale a **0**.

Quando  $r = -1$ , se  $X$  è ordinato in modo crescente,  $Y$  è ordinato in modo decrescente: di conseguenza, le differenze sono massime e la sommatoria raggiunge un valore massimo determinato dal numero di coppie di osservazioni ( $N$ ).

Quando  $r = 0$ , mentre i ranghi di  $X$  sono ordinati in modo crescente quelli di  $Y$  hanno una distribuzione casuale: la sommatoria delle  $d_{R_i}^2$  tende ad un valore medio, determinato dal numero di coppie di osservazioni ( $N$ ).

La statistica

$$D = \sum d_{R_i}^2$$

è conosciuta come **test statistico di Hotelling-Pabst**; per essa sono state proposte tavole di valori critici, al fine di valutare la significatività del test.

6 - Il **coefficiente di correlazione tra ranghi ( $r$ ) di Spearman** è derivato dalla formula della correlazione di Pearson

$$r = \frac{\text{cod}_{xy}}{\text{dev}_x}$$

Applicata ai ranghi, diviene

$$\rho = \frac{\sum_i \left( R_{xi} - \frac{N+1}{2} \right) \cdot \left( R_{yi} - \frac{N+1}{2} \right)}{\frac{N(N^2 - 1)}{12}}$$

Dopo semplificazione, **la formula abbreviata** può essere scritta come

$$\rho = 1 - \frac{6 \cdot \sum d_{R_i}^2}{N^3 - N}$$

con  $N$  uguale al numero di coppie di osservazioni.

**Quando due o più valori di X o di Y sono identici (ties)** e pertanto hanno lo stesso rango, l'attribuzione dei punteggi medi riduce il valore della devianza. Con pochi valori identici, l'effetto è trascurabile. Con molti valori identici, è bene calcolare un **fattore di correzione T** sia per la variabile  $X$  ( $T_x$ ) sia per la  $Y$  ( $T_y$ )

$$T = \sum_{i=1}^g (t_i^3 - t_i)$$

dove

**g** è il numero di raggruppamenti con punteggi identici e

**t** è il numero di ranghi identici entro ogni raggruppamento.

Con queste correzioni, nel caso di molti valori identici la formula del  $\rho$  di Spearman diventa

$$\rho = \frac{N^3 - N - 6 \cdot d^2 - \frac{T_x + T_y}{2}}{\sqrt{(N^3 - N)^2 - (T_x + T_y) \cdot (N^3 - N) + T_x \cdot T_y}}$$

**La correzione determina una differenza sensibile quando uno stesso valore è uguale in molti casi; ha poca influenza quando si hanno molti valori ripetuti solo 2 volte.**

Nel caso di **piccoli campioni** ( $N < 20-25$ ), la significatività di  $\rho$  è fornita dalle tabelle dei valori critici. Nella pagina seguente sono riportati i valori critici di  $r$  sia per test ad una coda che a due code: alla probabilità  $\alpha$  prefissata **si rifiuta l'ipotesi nulla se il valore calcolato è uguale o superiore a quello stimato.**

Nel caso di **grandi campioni** ( $N > 20-25$ ), quando è valida l'ipotesi nulla d'assenza di correlazione, il valore di  $\rho$  è distribuito con media **0** e deviazione standard **1**. Per la sua significatività è stato proposto sia il ricorso alla **distribuzione z con la trasformazione**

$$z = \rho \cdot \sqrt{N - 1}$$

sia alla **distribuzione t di Student con gdl N - 2 con la trasformazione di r**

$$t_{(N-1)} = \rho \cdot \sqrt{\frac{N-2}{1-\rho^2}}$$

Tra **t** e **z**, il **test t sembra preferibile quando il campione ha meno di 50 osservazioni; per campioni di dimensioni maggiori i due modi di stima della significatività risultano equivalenti.**

### Valori critici del coefficiente r di Spearman

per test a 1 coda e test a 2 code

a

N	0.05	0.025	0.01	0.005	0.001	0.0005	1 coda
	0.10	0.05	0.02	0.01	0.002	0.001	2 code
4	1.000						
5	.900	1.000	1.000				
6	.829	.886	.943	1.000			
7	.714	.786	.893	.929	1.000	1.000	
8	.643	.738	.833	.881	.952	.976	
9	.600	.700	.783	.833	.917	.933	
10	.564	.648	.745	.794	.879	.903	
11	.536	.618	.709	.755	.845	.873	
12	.503	.587	.671	.727	.825	.860	
13	.484	.560	.648	.703	.802	.853	
14	.464	.538	.622	.675	.776	.811	
15	.443	.521	.604	.654	.754	.786	
16	.429	.503	.582	.635	.732	.765	
17	.414	.485	.566	.615	.713	.748	
18	.401	.472	.550	.600	.695	.728	
19	.391	.460	.535	.584	.677	.712	
20	.380	.447	.520	.570	.662	.696	
21	.370	.435	.508	.556	.648	.681	
22	.361	.425	.496	.544	.634	.667	
23	.353	.415	.486	.532	.622	.654	
24	.344	.406	.476	.521	.610	.642	

ESEMPIO. La concentrazione delle sostanze organiche presenti nell'acqua può essere misurata mediante il BOD (da Biological Oxygen Demand, la richiesta biochimica dell'ossigeno), il COD (da Chemical Oxygen Demand, la richiesta chimica dell'ossigeno) e il TOC (da Total Organic Carbon, il carbonio organico totale).

Lungo un corso d'acqua sono state fatte 16 rilevazioni del BOD<sub>5</sub> (a 5 giorni) e dell'azoto ammoniacale.

S'intende verificare se tra le due serie di valori esista una correlazione positiva significativa, nonostante la non normalità delle distribuzioni.

Nella tabella che segue sono stati riportate le quantità di BOB<sub>5</sub> e d'azoto ammoniacale rilevate in ognuna delle 16 stazioni campionate.

Stazione	BOD <sub>5</sub>	N
s <sub>1</sub>	5	0,7
s <sub>2</sub>	5	0,8
s <sub>3</sub>	12	5,6
s <sub>4</sub>	35	24,3
s <sub>5</sub>	11	9,7
s <sub>6</sub>	7	1,8
s <sub>7</sub>	8	1,6
s <sub>8</sub>	9	4,8
s <sub>9</sub>	9	1,7
s <sub>10</sub>	20	4,8
s <sub>11</sub>	14	5,6
s <sub>12</sub>	13	3,2
s <sub>13</sub>	16	3,6
s <sub>14</sub>	15	2,9
s <sub>15</sub>	13	3,9
s <sub>16</sub>	11	2,8

Risposta.

Il test è unilaterale, con

$$H_0: r = 0$$

e

$$H_1: r > 0$$

Il metodo può essere suddiviso in 7 fasi, delle quali le prime 5 sono indicate nella tabella:

- 1 - ordinare in modo crescente i valori del  $BOD_5$  ed attribuire i ranghi relativi (**colonne 1a e 1b**);
- 2 - trasformare in ranghi i corrispondenti valori di **N** (**colonne 2a e 2b**);
- 3 - calcolare la differenza **d** tra i ranghi (**colonna 3**);
- 4 - elevare al quadrato tali differenze (**d<sup>2</sup> nella colonna 4**);
- 5 - calcolare la somma dei quadrati delle differenze (**somma delle d<sup>2</sup> nella colonna 4**);

	1a	2a	1b	2b	3	4
stazione	$BOD_5$	N	$R(BOD_5)$	$R(N)$	d	$d^2$
s1	5	0,7	1,5	1	0,5	0,25
s2	5	0,8	1,5	2	-0,5	0,25
s6	7	1,8	3	5	-2	4
s7	8	1,6	4	3	1	1
s9	9	1,7	5,5	4	1,5	2,25
s8	9	4,8	5,5	11,5	-6	36
s16	11	2,8	7,5	6	1,5	2,25
s5	11	9,7	7,5	15	-7,5	56,25
s3	12	5,6	9	13,5	-4,5	20,25
s12	13	3,2	10,5	8	2,5	6,25
s15	13	3,9	10,5	10	0,5	0,25
s11	14	5,6	12	13,5	-1,5	2,25
s14	15	4,8	13	11,5	1,5	2,25
s13	16	3,6	14	9	5	25
s10	20	2,9	15	7	8	64
s4	35	24,3	16	16	0	0

Con i dati dell'esempio,  $(\sum d_{R_i}^2)$  è uguale a 220,5.

6 - Per  $N$  uguale a 16, calcolare il valore di  $r$

$$\rho = 1 - \frac{6 \cdot \sum d_{R_i}^2}{N^3 - N} = 1 - \frac{6 \cdot 220,5}{4096 - 16} = 1 - \frac{1323}{4080} = 1 - 0,324 = \mathbf{0,676}$$

che risulta uguale a 0,676.

7 - Nella **tabella dei valori critici**, alla probabilità  $\alpha = \mathbf{0.01}$  per un test a **1** coda il valore riportato è 0.582. Il valore calcolato è superiore: si rifiuta l'ipotesi nulla e si accetta implicitamente l'ipotesi alternativa dell'esistenza di un'associazione positiva tra le due serie di dati rilevati.

Benché il numero di dati ( $N = 16$ ) sia oggettivamente ridotto, la significatività può essere stimata sia con la distribuzione  $z$  che con la distribuzione  $t$  di Student.

Con il **test  $z$**  si ottiene

$$z = 0,676 \cdot \sqrt{16-1} = 0,676 \cdot 3,873 = \mathbf{2,62}$$

un valore di  $z$  uguale a 2,62.

Nella tabella della distribuzione normale, corrisponde alla probabilità 0.0044 nel caso di un test unilaterale. L'approssimazione è molto buona: con tabelle di  $r$  più dettagliate la differenza risulta trascurabile, di circa 2/1000.

Con il **test  $t$**  si ottiene

$$\begin{aligned} t_{(15)} &= 0,676 \cdot \sqrt{\frac{16-2}{1-0,676^2}} = 0,676 \cdot \sqrt{\frac{14}{1-0,457}} = \\ &= 0,767 \cdot \sqrt{\frac{14}{0,543}} = 0,676 \cdot \sqrt{25,783} = 0,676 \cdot 5,078 = \mathbf{3,433} \end{aligned}$$

un valore di uguale a 3,433 con 15 gdl.

Nella tabella sinottica dei valori critici del  $t$  di Student per un test unilaterale si trova tra la probabilità  $\alpha = \mathbf{0.005}$  ( $t_{15} = 2,947$ ) e  $\alpha = \mathbf{0.0005}$  ( $t_{15} = 4,073$ ).

La conclusione non è molto differente da quella ottenuta con i due metodi precedenti.

I tre risultati sono approssimativamente equivalenti.

Quando i dati sono raccolti come serie temporale o come distanza lineare da un punto origine, il test di Spearman può essere utilizzato per verificare se i dati hanno una sequenza ordinata: tale test per il trend è noto in letteratura come test di Daniels.

**15.2. IL COEFFICIENTE DI CORRELAZIONE  $\tau$  (tau) DI KENDALL, CON  $t_a$  E  $t_b$ ,  
IL  $C^2$  DI MANTEL-HAENSZEL**

Il test  $\tau$  (tau), proposto da **M. G. Kendall** nel 1938 con l'articolo "A new measure of rank correlation" (su "Biometrika" n. 30) e in modo più dettagliato nel 1948 con il volume "Rank correlation methods" (edito a Londra da C. Griffin), ha le **stesse assunzioni**, può essere utilizzato nelle **medesime condizioni e sui medesimi dati del test  $\rho$  di Spearman**. I **risultati tra i due test sono altamente correlati**, anche se matematicamente non equivalenti, come sarà spiegato; tuttavia, sovente il  $\rho$  di Spearman è preferito perché più semplice, meglio conosciuto e più simile al coefficiente parametrico  $r$  di Pearson.

Il vantaggio del test  $\tau$  deriva dalla sua **estensione** sia all'analisi dei coefficienti di **correlazione parziale o netta** sia alla misura dell'**accordo tra giudizi multipli**.

La metodologia per stimare il **t di Kendall** può essere suddivisa in 6 fasi, con le prime due uguali a quelle del test **r di Spearman**.

1 - Dopo la presentazione tabellare dei dati con due misure per ogni oggetto d'osservazione

Variabili	Coppie di valori osservati						
X	8	5	7	14	22	21	41
Y	12	3	2	10	25	19	22
Oggetti	A	B	C	D	E	F	G

occorre ordinare per ranghi la variabile **X**, assegnando il rango 1 al valore più piccolo e progressivamente un rango maggiore, fino ad **N**, al valore più grande. Se sono presenti due o più valori uguali nella variabile **X**, assegnare ad ognuno come rango la media delle loro posizioni. La **scala** comunque **dovrebbe essere continua, anche se di rango**, e quindi non avere valore identici, se non in casi eccezionali.

**E' indispensabile collocare nell'ordine naturale (da 1 a N) i ranghi della variabile X**, spostando di conseguenza i valori della **Y** relativi agli stessi soggetti

Variabili	Coppie di valori osservati						
X	1	2	3	4	5	6	7
Y	3	2	12	10	25	16	22
Oggetti	B	C	A	D	F	E	G

2 - Sostituire gli **N** valori di **Y** con i ranghi rispettivi; per valori di **Y** uguali, come al solito usare la media dei ranghi.

I ranghi di **Y** risultano distribuiti secondo il rango della variabile **X**, come nella tabella seguente:

Variabili	Coppie di valori osservati						
X	1	2	3	4	5	6	7
Y	2	1	4	3	7	5	6
Oggetti	B	C	A	D	F	E	G

**Il metodo proposto da Kendall utilizza le informazioni fornite dall'ordine della variabile Y.**

3 - Se le due distribuzioni sono correlate in modo positivo ( $r = +1$ ), anche i ranghi della variabile **Y** sono ordinati in modo crescente, **concordanti con l'ordine naturale**;

se le due distribuzioni sono correlate in modo negativo ( $r = -1$ ), i valori di **Y** risulteranno ordinati in modo decrescente e saranno **discordanti dall'ordine naturale**;

se tra le due variabili non esiste correlazione ( $r = 0$ ), **l'ordine della variabile Y risulterà casuale e il numero di ranghi concordanti e di quelli discordanti dall'ordine naturale tenderà ad essere uguale, con somma uguale a 0.**

Per quantificare il grado di correlazione o concordanza, Kendall ha proposto di **contare per la sola variabile Y quante sono le coppie di ranghi che sono concordanti e quante quelle discordanti dall'ordine naturale.**

Y	2	1	4	3	7	5	6
---	---	---	---	---	---	---	---

Per esempio,

il valore 2 è seguito da 1: non è nell'ordine naturale e pertanto contribuirà con -1; inoltre è seguito da altri 5 valori maggiori, che contribuiranno insieme con +5: il contributo complessivo del valore 2 al calcolo delle concordanze è uguale a +4.

Il valore 1 è seguito da 5 valori maggiori e contribuirà con + 5.

Il valore 4 contribuisce con -1, perché seguito dal 3, e con +3, in quanto i 3 successivi sono maggiori, per un valore complessivo di +2.

Il valore 3 contribuisce con +3.

Il valore 7 contribuisce con -2, in quanto seguito da 2 valori minori.

Il valore 5 contribuisce con +1.

Il valore 6 è l'ultimo e come non fornisce alcun contributo al calcolo delle concordanze; con esso termina il calcolo delle differenze tra concordanze e discordanze.

Nella tabella seguente è riportato il conteggio dettagliato e complessivo delle concordanze (+) e delle discordanze (-)

2	1	4	3	7	5	6	Totale
	-	+	+	+	+	+	+4
		+	+	+	+	+	+5
			-	+	+	+	+2
				+	+	+	+3
					-	-	-2
						+	+1
<b>Totale (concordanze meno discordanze)</b>							<b>+13</b>

**La misura della concordanza complessiva con la variabile X è dato dalla somma algebrica di tutte le concordanze e le discordanze.**

Il totale di concordanze e discordanze dei 7 valori dell'esempio (+4, +5, +2, +3, -2, +1) è uguale a +13.

4 - Il numero totale di concordanze e discordanze di una serie di valori deve essere rapportato al **massimo totale possibile**. Poiché i confronti sono fatti a coppie, con **N** dati il numero totale di confronti concordanti o discordanti è dato dalla combinazione di **N** elementi **2 a 2**

$$C_N^2$$

Con una serie di 7 dati come nell'esempio, il numero complessivo di confronti, quindi il massimo totale possibile di concordanze o discordanze, è

$$C_7^2 = \frac{7!}{(7-2)! \cdot 2!}$$

uguale a 21.

5 - Secondo il metodo proposto di **Kendall**, il grado di relazione o concordanza ( $\tau$ ) tra la variabile **X** e **Y** può essere quantificato dal rapporto

$$t = \frac{\text{totale}(\text{concordanze} - \text{discordanze})}{\text{massimo totale possibile}} = \frac{\text{totale}(\text{concordanze} - \text{discordanze})}{C_N^2}$$

Con i 7 dati riportati,

$$t = \frac{+13}{21} = +0,619$$

t è uguale a +0,619.

Il  **$\tau$  di Kendall** varia in modo simile al coefficiente **r di Pearson**

tra **+1**, quando la concordanza o correlazione tra X e Y è massimo,

e **-1**, quando la concordanza o correlazione tra le due variabili è negativa;

è **0** quando non esiste concordanza.

La **formula abbreviata** è

$$t = \frac{2 \cdot \text{totale}(\text{concordanze} - \text{discordanze})}{N \cdot (N - 1)}$$

dove **N** è il numero di coppie di dati.

In caso di due o più valori identici nella successione delle Y, il confronto con l'ordine naturale non determina né una concordanza né una discordanza: il loro confronto non contribuisce al calcolo di  $\tau$  e si riduce il valore di **N**.

La mancata correzione comporterebbe che il rango di variazione non sarebbe più tra **-1** e **+1**.

La formula, corretta per la presenza di valori identici, diventa

$$t = \frac{2 \cdot \text{totale}(\text{concordanze} - \text{discordanze})}{\sqrt{N \cdot (N - 1) \cdot T_x} \cdot \sqrt{N \cdot (N - 1) \cdot T_y}}$$

dove

$N$  è il numero totale di coppie di dati delle variabili  $X$  e  $Y$ ,

$$T_x = \sum (t_x^2 - t_x) \text{ dove}$$

$t_x$  è il numero di osservazioni identiche di ogni gruppo di valori identici della variabile  $X$ ,

$$T_y = \sum (t_y^2 - t_y) \text{ dove}$$

$t_y$  è il numero di osservazioni identiche di ogni gruppo di valori identici della variabile  $Y$ .

Per **piccoli campioni**, i valori critici sono forniti dalla tabella relativa, riportata nella pagina seguente.

L'esempio con i 7 dati precedentemente utilizzato risulta significativo alla probabilità  $\alpha = 0.05$  per un test ad 1 coda.

Per **grandi campioni** la significatività del  $\tau$  di Kendall può essere saggiata con la distribuzione normale

$$z = \frac{\tau - \mu_\tau}{\sigma_\tau} \quad (*)$$

Quando è vera l'ipotesi nulla (assenza di correlazione o d'associazione), per la media  $\mu_\tau$  vale l'uguaglianza

$$\mu_\tau = 0$$

(l'ordine della variabile  $Y$  è casuale e la somma totale delle sue concordanze e discordanze è nulla), mentre la varianza è data da

$$s_t^2 = \frac{2 \cdot (2N + 5)}{9N \cdot (N - 1)}$$

dove  $N$  è il numero di dati.

**Valori critici del coefficiente di correlazione semplice  $t$  di Kendall**

per test a 1 coda e a 2 code

**a**

<b>N</b>	<b>0.05</b>	0.025	<b>0.01</b>	0.005	<b>1 coda</b>
	0.10	<b>0.05</b>	0.02	<b>0.01</b>	<b>2 code</b>
<b>4</b>	1.000				
<b>5</b>	.800	.800	1.000		
<b>6</b>	.733	.867	.867	1.000	
<b>7</b>	.619	.714	.810	.810	
<b>8</b>	.571	.643	.714	.786	
<b>9</b>	.500	.556	.667	.722	
<b>10</b>	.467	.511	.600	.644	
<b>11</b>	.418	.491	.564	.600	
<b>12</b>	.394	.455	.545	.576	
<b>13</b>	.359	.436	.513	.564	
<b>14</b>	.363	.407	.473	.516	
<b>15</b>	.333	.390	.467	.505	
<b>16</b>	.317	.383	.433	.483	
<b>17</b>	.309	.368	.426	.471	
<b>18</b>	.294	.346	.412	.451	
<b>19</b>	.287	.333	.392	.439	
<b>20</b>	.274	.326	.379	.421	
<b>21</b>	.267	.314	.371	.410	
<b>22</b>	.257	.296	.352	.391	
<b>23</b>	.253	.295	.344	.378	
<b>24</b>	.246	.290	.341	.377	

Sostituendo nella (\*) e semplificando, con la **formula abbreviata** si ottiene una stima più rapida di **z** mediante la relazione

$$z = \frac{3t \cdot \sqrt{N \cdot (N - 1)}}{\sqrt{2 \cdot (2N + 5)}}$$

ESEMPIO. Rispondere mediante il  $\tau$  di Kendall alla medesima domanda di verifica della significatività dell'associazione tra le variabili X e Y, con gli stessi dati utilizzati nell'esercizio precedente sul  $\rho$  di Spearman.

	1	2	1	2	3	3	4
stazione	BOD <sub>5</sub>	N	R(BOD <sub>5</sub> )	R( N )	concord.	discord .	diff .
s1	5	0,7	1,5	1	+15	-0	+15
s2	5	0,8	1,5	2	+14	-0	+14
s6	7	1,8	3	5	+11	-2	+9
s7	8	1,6	4	3	+12	-0	+12
s9	9	1,7	5,5	4	+11	-0	+11
s8	9	4,8	5,5	11,5	+4	-4	0
s16	11	2,8	7,5	6	+9	-0	+9
s5	11	9,7	7,5	15	+1	-7	-6
s3	12	5,6	9	13,5	+1	-5	-4
s12	13	3,2	10,5	8	+5	-1	+4
s15	13	3,9	10,5	10	+3	-2	+1
s11	14	5,6	12	13,5	+1	-3	-2
s14	15	4,8	13	11,5	+1	-2	-1
s13	16	3,6	14	9	+1	-1	0
s10	20	2,9	15	7	+1	-0	+1
s4	35	24,3	16	16	-	-	-
<b>Totale differenze (concordanze - discordanze)</b>							<b>63</b>

La metodologia del  $\tau$  di Kendall richiede i seguenti passaggi (riportati nella tabella da 1 a 4):

- 1 - ordinare in modo crescente i valori del  $BOD_5$  ed attribuire i ranghi relativi;
- 2 - trasformare in ranghi i corrispondenti valori di  $N$ ;
- 3 - calcolare per ogni punteggio di  $N$  il numero di concordanze e di discordanze;
- 4 - calcolare la somma complessiva di tutte le concordanze e le discordanze.

La somma totale delle differenze tra concordanze e discordanze risulta positiva (+63).

- 5 - Tradotto nel corrispondente coefficiente mediante

$$t = \frac{2 \cdot (+63)}{16 \cdot 15} = \frac{+126}{240} = +0,525$$

si ottiene un valore di  $\tau$  uguale a +0,525.

6 - Per un test unilaterale, la tabella dei valori critici del  $\tau$  di Kendall con  $N = 16$  alla probabilità  $\alpha = 0,005$  riporta un valore di  $\tau$  uguale a 0,483. Il valore calcolato è superiore in modulo.

Di conseguenza, si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa: esiste un'associazione o correlazione positiva tra le due serie di dati, con probabilità inferiore a 0,005 di commettere un errore di I tipo.

Il campione utilizzato nell'esempio può essere ritenuto sufficientemente grande.

Pertanto, è possibile valutare la significatività del coefficiente  $\tau = +0,525$  mediante il **test z**:

$$z = \frac{3t \cdot \sqrt{N \cdot (N-1)}}{\sqrt{2 \cdot (2N+5)}} = \frac{3(+0,525) \cdot \sqrt{16 \cdot 15}}{\sqrt{2 \cdot (32+5)}} = \frac{+1,575 \cdot \sqrt{240}}{\sqrt{69}} = \frac{+24,340}{8,307} = +2,93$$

che fornisce un valore uguale a +2,93.

Nella distribuzione normale, a  $z$  uguale a 2,93 per un test ad una coda corrisponde una probabilità  $\alpha = 0,0017$  che non si discosta in modo rilevante da quella fornita dalle tabelle.

**Il  $\rho$  di Spearman e il  $\tau$  di Kendall hanno la stessa potenza nel rifiutare l'ipotesi nulla,** anche se i valori di  $\rho$  e  $\tau$  sono numericamente differenti per lo stesso campione di dati.

Stime pubblicate per **l'efficienza asintotica relativa del test t di Kendall**, rispetto al test **r di Pearson** e nel caso che l'ipotesi nulla sia vera, riportano

- quando la distribuzione dei dati è Normale un valore uguale a  $0,91 (3/\pi)^2$ ,
- quando la distribuzione dei dati è Rettangolare un valore uguale a 1,

- quando la distribuzione dei dati è Esponenziale Doppia un valore uguale a 1,27 (81/64).

Quando l'ipotesi nulla  $H_0$  è vera, le probabilità fornite dai due metodi sono molto simili; per grandi campioni, esse tendono ad essere molto simili anche mediante la distribuzione normale.

Tuttavia, quando l'ipotesi nulla  $H_0$  è falsa e quindi è vera l'ipotesi alternativa  $H_1$ , i due differenti indici sono diversamente sensibili alle distorsioni determinate dal diverso campo di variazione; di conseguenza, i risultati tendono a differire maggiormente.

Con i dati dell'esempio utilizzato, per un test unilaterale con il  $\rho$  di Spearman è stato ottenuto un valore di  $z = 2,62$  che corrisponde ad una probabilità di 0.0044 mentre con il  $\tau$  di Kendall si è ottenuto un valore di  $z = 2,93$  che corrisponde ad una probabilità di 0.0017.

La differenza tra le probabilità stimate con i due diversi indici è in assoluto inferiore al 3/1000 e quindi oggettivamente molto limitata; ma è elevata (2,59 a 1) se considerata in rapporto alle piccole probabilità stimate.

**Al momento, non è noto quale indice in generale dia il valore più corretto.** Molti preferiscono il  $\tau$  di Kendall per il metodo di calcolo, ritenuto più robusto di quello basato sulle differenze.

Altri testi presentano la stessa procedura di calcolo delle precedenze in modo più complesso e propongono 2 misure differenti ( $t_a$ ,  $t_b$ ); la scelta tra esse dipende dal numero di valori identici e quindi dalla continuità del tipo di scala utilizzato. E' possibile determinare i casi concordi, discordi o a pari merito, confrontando simultaneamente i valori di  $X$  e  $Y$  in una coppia d'oggetti. Una coppia di casi è **concorde (P)**, se per un oggetto i valori di entrambi le variabili sono più bassi o più alti rispetto ai valori dell'altro caso; è **discorde (Q)**, se per una variabile è maggiore e per l'altra minore, o viceversa; **pari merito (T)**, se hanno lo stesso valore per la variabile  $X$  ( $T_X$ ) o per la variabile  $Y$  ( $T_Y$ ).

**Il  $t_a$  è la differenza tra coppie concordi e discordi (P-Q), rapportata al numero totale di coppie d'oggetti:**

$$t_a = \frac{P - Q}{C_n^2}$$

**Se non esistono coppie con valori uguali, questa misura varia tra -1 e +1.**

**Se esistono coppie con valori uguali, il campo di variazione è più limitato e dipende dal numero di valori pari merito presenti sia nella variabile  $X$  che nella variabile  $Y$ .**

**Il  $t_b$  normalizza la differenza P-Q, prendendo in considerazione anche i valori pari merito delle due variabili in modo separato**

$$t_b = \frac{P-Q}{\sqrt{(P+Q+T_x) \cdot (P+Q+T_y)}}$$

I coefficienti di correlazione non parametrica  $\rho$  di Spearman e il di  $\tau$  Kendall sono utilizzati con variabili almeno ordinali. Per ognuna delle due variabili, è quindi sempre possibile classificare i valori dal minimo al massimo; in caso di un numero sufficientemente elevato di dati, è semplice **calcolare il coefficiente di correlazione r di Pearson sulla matrice** in cui nelle caselle sono riportate le frequenze dei valori congiunti delle due variabili (come già dimostrato nel capitolo relativo, analizzando il caso di classi di frequenza arbitrarie).

**Una misura dell'associazione lineare tra la variabile di riga e quella di colonna**, riportate in tale tabella di contingenza, è il  $c^2$  di Mantel-Haenszel che ha solo **1 gdl**:

è dato dal **prodotto del quadrato del coefficiente di correlazione r di Pearson per il numero di casi meno uno**

$$C^2_{(1)} \text{ di Mantel-Haenszel} = r^2 \text{ (di Pearson)} \times (n-1)$$

### 15.3. IL COEFFICIENTE DI CORRELAZIONE PARZIALE $\tau$ (tau) DI KENDALL

Quando si analizza la correlazione tra 2 variabili, sovente può sorgere il dubbio che la causa principale di un valore elevato possa essere attribuita non ad una correlazione effettiva tra loro, ma ad una correlazione o associazione di ognuna di loro con una terza variabile; oppure, che l'inatteso basso valore sia imputabile all'azione di segno opposto di una terza variabile, alla quale entrambe sono correlate con segno diverso.

Come già discusso nel caso della correlazione parametrica, gli effetti delle variazioni congiunte tra due variabili dovute al loro legame con una terza possono essere eliminati, quando si tenga costante quest'ultima; ma è un procedimento che poi limita le conclusioni a quel caso specifico. Non è detto che, per altri valori della terza variabile, le prime due mantengano lo stesso grado di correlazione calcolata per alcuni valori. E' un caso frequente nella ricerca ambientale, che utilizza le correlazioni statistiche come indicazione preliminare e supporto all'indagine causale.

Le sostanze contaminanti rilasciate in fiumi e laghi determinano effetti nocivi sulle popolazioni sia animali sia vegetali e sulle comunità acquatiche. Gli effetti principali possono essere raggruppati in alcune categorie: eutrofizzanti, deossigenanti, tossici, fisici come quelli sulla temperatura, chimici come quelli sul pH, patogeni.

Temperatura, eutrofizzazione e deossigenazione risultano tra loro correlati. Per misurare in modo corretto la relazione tra eutrofizzazione e deossigenazione in un gruppo di laghi, occorrerebbe fare

rilevazioni a temperatura costante; ma le conclusioni sarebbero limitate a quella specifica temperatura. Nello stesso modo, dovrebbero essere verificate le relazioni sia tra temperatura e deossigenazione con un livello di eutrofizzazione costante, sia le relazioni tra temperatura ed eutrofizzazione con un indice di deossigenazione mantenuto stabile.

Per rendere generali le conclusioni sulla correlazione tra due variabili occorrerebbe un numero molto alto di esperimenti, estesi a tutti i livelli o modalità della terza variabile.

**La correlazione parziale tra k variabili permette di valutare il grado di correlazione esistente tra ogni coppia di variabili, utilizzando solo una serie di dati.** Il caso più semplice e frequente è quello tra 3 variabili; ad esso viene limitata la presentazione della correlazione parziale non parametrica  $\tau$  di Kendall, come è già stato fatto con l'indice  $r$  di Pearson nel capitolo precedente sulla correlazione parziale parametrica.

**Il  $\tau_{YZ,X}$  (leggesi: valore di correlazione tra le variabili Y e Z, indipendente da X) di Kendall misura quanto le misure delle variabili Y e Z siano correlate, indipendentemente dal loro accordo con quelle di X.**

La correlazione parziale  $\tau_{YZ,X}$  può essere derivata dai valori delle 3 correlazioni semplici ( $\tau_{XY}$ ,  $\tau_{XZ}$  e  $\tau_{YZ}$ ), in accordo con la relazione

$$\tau_{YZ,X} = \frac{\tau_{YZ} - \tau_{XY} \cdot \tau_{XZ}}{\sqrt{(1 - \tau_{XY}^2) \cdot (1 - \tau_{XZ}^2)}}$$

L'applicazione e l'interpretazione sono identiche a quelle della correlazione netta parametrica; di conseguenza, per tali argomenti si rinvia ad essa.

**L'ipotesi nulla  $H_0$  è che non esista correlazione, mentre l'ipotesi alternativa  $H_1$  unilaterale può supporre la presenza di una correlazione sia positiva che negativa, ma chiaramente espressa, e l'ipotesi nulla  $H_1$  bilaterale verifica la semplice presenza di una correlazione, senza alcuna indicazione di segno.**

Con più variabili, il numero di confronti possibili diventa elevato. Si pone lo stesso problema di significatività già discusso per i confronti tra più medie.

Si ricorre a concetti analoghi al  $t$  di **Bonferroni** per confronti multipli a posteriori, se i confronti non sono già prestabiliti e limitati a quelli ritenuti importanti per la ricerca.

**Valori critici del coefficiente di correlazione parziale  $t_{xy,z}$  di Kendall**  
per test a 1 coda e a 2 code

**a**

N	0.05	0.025	0.01	0.005	0.001	1 coda
	0.10	0.05	0.02	0.01	0.002	2 code
4	.707	1.000				
5	.667	.802	.816	1.000		
6	.600	.667	.764	.866	1.000	
7	.527	.617	.712	.761	.901	
8	.484	.565	.648	.713	.807	
9	.443	.515	.602	.660	.757	
10	.413	.480	.562	.614	.718	
11	.387	.453	.530	.581	.677	
12	.365	.430	.505	.548	.643	
13	.347	.410	.481	.527	.616	
14	.331	.391	.458	.503	.590	
15	.319	.377	.442	.485	.570	
16	.305	.361	.423	.466	.549	
17	.294	.348	.410	.450	.532	
18	.284	.336	.395	.434	.514	
19	.275	.326	.382	.421	.498	
20	.268	.318	.374	.412	.488	

Per **piccoli campioni**, la significatività è fornita dalla **tabella riportata nella pagina precedente**.

Per **grandi campioni**, si ricorre alla **distribuzione normale**, con formula identica a quella della correlazione semplice

$$z = \frac{3t \cdot \sqrt{N \cdot (N-1)}}{\sqrt{2 \cdot (2N+5)}}$$

dove

$t_{YZ,X}$  è il valore calcolato con la formula presentata per la correlazione netta,

$N$  è il numero di dati, uguale per tutte le variabili a confronto.

Kendall ha proposto anche una **metodologia per stimare direttamente la correlazione netta, utilizzando i dati originari di tre variabili**. Il metodo è lungo e complesso; in pratica, può essere applicato con successo ricorrendo ai calcoli manuali, solo nel caso in cui  $N$  sia limitato a poche osservazioni.

Per **comprendere in modo dettagliato ed operativo questo metodo di Kendall**, si supponga di avere misurato tre variabili (X, Y, Z), su un gruppo formato da 5 rilevazioni campionarie, con i seguenti risultati:

	Variabili		
Campione	X	Y	Z
I	5	18	7
II	9	12	10
III	12	13	2
IV	21	15	11
V	54	31	36

Calcolare la correlazione parziale o netta ( $t_{YZ,X}$ ) tra Y e Z, al netto di X.

La metodologia richiede alcuni passaggi:

1- La serie dei valori deve essere ordinata in modo crescente per la variabile X (già fatto nella abella di presentazione dei dati).

2 - Trasformare i valori in ranghi, entro ogni variabile, ordinando separatamente i ranghi delle altre due variabili (**Y** e **Z**), ma senza spostare la loro collocazione in riferimento alla rilevazione della variabile **X**.

Con i dati del campione, dopo queste due operazioni si ottiene una nuova tabella:

Campione	Variabili		
	X	Y	Z
I	1	3	2
II	2	1	3
III	3	2	1
IV	4	4	4
V	5	5	5

3 - Per ognuna delle 3 variabili, tradurre l'ordine dei ranghi in concordanze e discordanze, assegnando + ad ogni coppia in ordine naturale o crescente e - ad ogni coppia in ordine non naturale o decrescente; ovviamente per la variabile X esistono solo concordanze e quindi segni +

Coppie di dati	Variabile		
	X	Y	Z
1 - I,II	+	-	+
2 - I,III	+	-	-
3 - I,IV	+	+	+
4 - I,V	+	+	+
5 - II,III	+	-	-
6 - II,IV	+	+	+
7 - II,V	+	+	+
8 - III,IV	+	-	+
9 - III,V	+	+	+
10 - IV,V	+	+	+

Ad esempio,

per la variabile Y la coppia di ranghi per la coppia I e II è decrescente (3 e 1) e pertanto è stato assegnato -,

mentre per la variabile Z la stessa coppia di ranghi (2 e 3) è crescente e pertanto è stato assegnato +.

4 - Riassumere in una tabella 2 x 2, come quella sottoriportata, l'informazione delle concordanze e delle discordanze delle variabili Y e Z, in relazione all'ordine di X.

Per esempio, nella coppia di rilevazioni I e II,

il rango della Y (-) è in disaccordo con X,

mentre quello di Z (+) è in accordo con X.

		Variabile Y		
		concord. con X	discord. con X	totale
variabile Z	concord. con X	<b>A 6</b>	<b>B 2</b>	n <sub>1</sub> 8
	discord. con X	<b>C 0</b>	<b>D 2</b>	n <sub>2</sub> 2
	totale	n <sub>3</sub> 6	n <sub>4</sub> 4	<b>N 10</b>

Con i dati dell'esempio, le coppie di rilevazioni sono complessivamente **10**, come riportato nella casella contrassegnata dal simbolo **N**.

Nella **casella A**, il numero **6** indica che in 6 coppie di rilevazioni (I,IV; I,V; II,IV; II,V; III,V; IV,V) le variabili **Y** e **Z** sono tra loro concordanti come con la variabile **X**; in altri termini, in 6 casi quando si ha il segno + nella colonna **X** si ha + sia nella colonna **Y** sia in quella **Z**.

Nella **casella B**, il numero **2** indica che in 2 coppie (I,II; III,IV) la variabile **Y** è discordante da **X**, mentre la variabile **Z** è concordante con **X**; in altri termini, mentre si ha il segno + nella colonna **X** si ha contemporaneamente il segno - nella colonna **Y** e quello + nella colonna **Z**.

Nella **casella C**, il numero **0** indica che non si ha alcuna coppia in cui variabile **Y** è concordante con **X**, mentre la variabile **Z** è discordante da **X**; in altri termini, non è presente alcun caso in cui si ha + nella colonna **X**, mentre si ha anche + nella colonna **Y** e contemporaneamente il segno - nella colonna **Z**.

Nella **casella D**, il numero **2** indica che in 2 coppie (I,II; II,III) la variabile **Y** e la variabile **Z** discordano simultaneamente dalla variabile **X**; in altri termini, nella casella D è riportato il numero

di casi in cui si ha il segno - contemporaneamente sia nella variabile **Y** sia nella variabile **Z**, mentre ovviamente si ha il segno + nella variabile **X**.

Le caselle **n<sub>1</sub>**, **n<sub>2</sub>**, **n<sub>3</sub>**, **n<sub>4</sub>** riportano i totali parziali.

Nella casella **n<sub>1</sub>** è riportato il numero di concordanze (8) tra la variabile **X** (+) e la variabile **Z** (+), senza considerare la variabile **Y**.

Nella casella **n<sub>2</sub>** è riportato il numero di discordanze (2) tra la variabile **X** (+) e la variabile **Z** (-), ignorando la variabile **Y**.

Nella casella **n<sub>3</sub>** è riportato il numero di concordanze (6) tra la variabile **X** (+) e la variabile **Y** (+), senza considerare la variabile **Z**.

Nella casella **n<sub>4</sub>** è riportato il numero di discordanze (4) tra la variabile **X** (+) e la variabile **Y** (-), senza considerare la variabile **Z**.

Una volta individuate concordanze e discordanze tra le tre variabili, il coefficiente di correlazione parziale  $\tau_{YZ,X}$  di Kendall, (la correlazione tra **Y** e **Z** tenendo costante **X**), è calcolata rapidamente mediante la relazione

$$t_{YZ,X} = \frac{A \cdot D - B \cdot C}{\sqrt{n_1 \cdot n_2 \cdot n_3 \cdot n_4}}$$

Con i dati dell'esempio,

$$t_{YZ,X} = \frac{6 \cdot 2 - 2 \cdot 0}{\sqrt{8 \cdot 2 \cdot 6 \cdot 4}} = \frac{+12}{\sqrt{384}} = \frac{+12}{19,60} = +0,61$$

si ottiene un valore di correlazione netta uguale a +0,61.

Trattandosi di un **campione piccolo**, per la sua significatività si utilizza la tabella precedente.

Nel caso di **grandi campioni**, data la complessità delle operazioni descritte, il metodo diventa di difficile applicazione, se non si possono utilizzare programmi informatici.

La significatività è fornita dal test **z**; se è stampato solo il suo valore, la probabilità relativa è data dalla distribuzione normale.

#### 15.4. CENNI SU MISURE DI CONCORDANZA TRA PIU' VALUTATORI: LA W, LA u DI KENDALL

I due coefficienti di correlazione non parametrica  $\rho$  di Spearman e la  $\tau$  di Kendall sono applicati a due serie di ranghi (o di valori trasformati in ranghi), relativi a  $N$  oggetti o individui. Quando si dispone di  $k$  serie di ranghi, riportati in una matrice  $N \times k$  (numero di oggetti  $\times$  numero di valutatori), è possibile determinare la **concordanza complessiva tra i  $k$  valutatori**.

Per tale scopo, i test proposti per le misure d'associazione sono numerosi; diversi sono stati pubblicati da M. G. Kendall nel 1962 con il volume "*Ranks correlation methods*", edito a Londra da Griffin. Tra essi, quelli più frequentemente proposti nei programmi informatici a maggior diffusione sono il **coefficiente di divergenza W di Kendall** e il coefficiente di **concordanza u di Kendall**, oltre a quelli già descritti nel capitolo precedente quali  **$t_K$  di Kendall**; ad essi sovente sono aggiunti la **statistica gamma** e la **d di Somers**.

A - La **W di Kendall** è una generalizzazione del test  $\rho$  di Spearman: **misura la divergenza** nella valutazione **tra  $k$  serie di misure ordinali**, in cui ogni serie è formata dai ranghi di  $N$  oggetti da parte di  $k$  valutatori.

Con  $k$  valutatori, mediante il  $r$  di Spearman è possibile calcolare tutte le correlazioni semplici tra loro, pari alle combinazioni 2 a 2 degli  $N$  oggetti ( $C_N^2$ ).

La media ( $\bar{r}$ ) di tutti questi coefficienti di correlazione  $r$  è

$$\bar{r} = \frac{(Wk) - 1}{k - 1}$$

dove

$W$  = indice di divergenza.

L'**indice di divergenza W** può essere calcolato direttamente da una serie di dati.

Si supponga che 4 ricercatori (I, II, III, IV) debbano stabilire un ordine tra 5 situazioni ambientali (A, B, C, D, E), per valutare il loro grado di degrado.

Nella tabella seguente, i valori dei  $k$  ricercatori attribuiti alle  $N$  situazioni ambientali (A, B, C, D, E) sono già stati trasformati in ranghi entro la stessa riga, attribuendo **1** al punteggio minore ed  **$N$**  a quello maggiore, mantenendo sempre fissa la posizione delle 5 situazioni.

Per esempio,

secondo il ricercatore I la situazione B è quella meno degradata e la D quella maggiormente degradata,

mentre il ricercatore II valuta la situazione A come migliore e la C come quella peggiore.

RICERCATORI	SITUAZIONI AMBIENTALI				
	A	B	C	D	E
I	2	1	4	5	3
II	1	2	5	4	3
III	1	2	3	5	4
IV	2	1	4	5	3
<b>R<sub>i</sub></b>	<b>6</b>	<b>6</b>	<b>16</b>	<b>19</b>	<b>13</b>
$\bar{R}_i$	<b>1,50</b>	<b>1,50</b>	<b>4,00</b>	<b>4,75</b>	<b>3,25</b>

Se fosse vera l'ipotesi nulla  $H_0$  d'assenza d'accordo tra i ricercatori (essi hanno fornito valutazioni di rango sulla base di principi totalmente differenti), le somme dei ranghi per colonna (**R<sub>i</sub>**) sarebbero tra loro uguali e le medie per colonna ( $\bar{R}_i$ ) uguali alla media generale;

viceversa, se fosse vera l'ipotesi alternativa di pieno accordo tra i ricercatori (essi forniscono la stessa valutazione), le somme (**R<sub>i</sub>**) e le medie relative ( $\bar{R}_i$ ) avrebbero differenze massime.

**Se l'ipotesi nulla ( $H_0$ ) fosse vera, l'indice di divergenza dovrebbe essere uguale a 0;**

**nel caso opposto ( $H_0$  falsa) di massima divergenza, l'indice W dovrebbe essere uguale a 1.**

E' un intervallo di variazione che può essere ottenuto, rapportando l'indice di divergenza al valore massimo possibile; nel caso delle medie dipende dal numero di casi **N** ed è uguale

$$N(N^2-1)/12$$

Il valore dell'indice di divergenza **W di Kendall** può essere ottenuto sia attraverso le medie

$$W = \frac{\sum_{i=1}^n (\bar{R}_i - \bar{R})^2}{N(N^2 - 1)/12}$$

sia, **con formula abbreviata**, direttamente dalle somme

$$W = \frac{12 \sum_{i=1}^n R_i^2 - 3k^2 N(N+1)^2}{k^2 N(N^2 - 1)}$$

ricordando che

**N** è il numero di oggetti da valutare o colonne,

**k** è il numero di valutatori o righe,

$\bar{R}_i$  è la media dei ranghi di ogni colonna,

$\bar{R}$  è la media generale dei ranghi

e la sommatoria è estesa a tutte le colonne.

**B - La u di Kendall è un coefficiente di concordanza, in un confronto appaiato.** E' una tecnica che può essere utile in particolare nella ricerca etologica, nello studio di preferenze di animali, anche se non mancano applicazioni nelle indagini a carattere ambientale od ecologico che, come punto di riferimento, hanno l'uomo.

Davanti a **N** oggetti, un animale difficilmente riesce ad esprimere preferenze complesse e stabilire un ordine di priorità tra più oggetti; per esempio, davanti a 5 cibi diversi, un animale potrebbe alimentarsi solo con uno, ignorando gli altri dopo averli solo annusati; posto di fronte alla scelta tra 5 femmine, un maschio potrebbe preferirne solo una, senza fornire al ricercatore elementi per stabilire un rango tra le altre 4.

Di conseguenza, nel suo piano sperimentale, il ricercatore deve impostare la ricerca in modo più complesso, con una serie di confronti a coppie: porre l'animale davanti alla scelta tra due oggetti, annotando la preferenza, per poi ripetere l'esperimento per tutte le coppie possibili di **N** oggetti.

Con **N** uguale a 5 oggetti, l'esperimento deve essere ripetuto 10 volte, come è possibile stimare con le combinazioni di 5 elementi 2 a 2 ( $C_N^2$ ).

Se con la ricerca s'intende verificare se **k** animali concordano nelle loro preferenze, occorre poi ripetere i 10 esperimenti per ognuno d'essi e riassumere tutti i risultati in una **matrice di preferenza**, per i **k** individui assieme.

Per esempio, con 5 cibi (A, B, C, D, E) differenti si è posto ognuno dei 4 animali (I, II, II, IV) davanti alla scelta in confronti appaiati, con i seguenti risultati

Coppie di 5 oggetti	Scelte dei 4 soggetti			
	I	II	III	IV
A,B	A	A	B	A
A,C	A	C	A	A
A,D	D	D	D	A
A,E	A	A	E	A
B,C	B	C	B	B
B,D	D	D	D	D
B,E	B	E	E	E
C,D	C	D	D	D
C,E	C	C	E	E
D,E	D	D	D	E

che possono essere riassunti in una tabella quadrata  $N \times N$ , nella quale sulle righe sono riportate le preferenze relative od ogni oggetto nei confronti appaiati.

Ovviamente, non esistono dati sulla diagonale.

#### OGGETTO PREFERITO

Oggetto preferito	Secondo elemento della coppia				
	A	B	C	D	E
A	---	3	3	1	3
B	1	---	3	0	1
C	1	1	---	1	2
D	3	4	3	---	3
E	1	3	2	1	---

Per esempio,

- nel confronto **A,C** 3 soggetti hanno dato la loro preferenza ad **A** (riassunto nel 3 riportato nella casella di riga **A** e colonna **C**) e 1 la sua preferenza a **C** (1 riportato all'incrocio tra la riga **C** e la colonna **A**);
- nel confronto **B,D** i 4 soggetti hanno dato tutti la loro preferenza a **D** (nella tabella sovrastante, è riportato 4 all'incrocio tra la riga **D** e la colonna **B** e quindi è stato riportato 0 all'incrocio tra la riga **B** e la colonna **D**).

**Non sempre le scelte sono congruenti o logiche;** per esempio, posto di fronte a tre coppie d'oggetti (A,B; A,C; B,C), un individuo potrebbe fare cadere la sua rispettivamente su **A** nel primo caso, su **C** nel secondo e su **B** nel terzo, evidenziando appunto scelte contraddittorie o casuali.

Tuttavia, la tabella quadrata riassuntiva fornisce informazioni sulla concordanza dei **k** valutatori nell'attribuzione delle preferenze.

Se fosse vera l'ipotesi nulla, in ogni cella si avrebbe  $2 (k/2)$  con  $k = 4$  valutatori);

se fosse vera l'ipotesi alternativa di totale concordanza tra i valutatori, teoricamente metà delle caselle avrebbero  $4 (k)$  con  $k$  valutatori) e metà 0, escludendo la diagonale.

**Il coefficiente di concordanza  $u$  di Kendall varia tra 0, in assenza di accordo, e 1, in presenza di totale accordo.**

## **15.5. MISURE PER DATI D'INTERVALLO CON CLASSI COSTANTI: IL COEFFICIENTE DI CORRELAZIONE $r$ DI PEARSON E IL COEFFICIENTE $\eta^2$**

E' possibile costruire una **tabella di contingenza  $M \times N$**  anche con una **distribuzione delle misure d'intervallo o di rapporto**. Quando sono costruite con intervalli costanti, queste tabelle permettono di calcolare il coefficiente di correlazione  **$r$  di Pearson**, un coefficiente simmetrico che misura la relazione lineare, come già descritto dettagliatamente nel capitolo relativo.

In alcune situazioni,

la **variabile dipendente è misurata su una scala d'intervallo**, mentre

la **variabile indipendente con una scala nominale od ordinale**.

Nell'analisi delle relazioni tra le due variabili, è appropriato il **coefficiente  $\eta^2$** .

**$\eta^2$  è un indice asimmetrico che, quando elevato al quadrato, esprime la proporzione della variabilità totale della variabile dipendente, che può essere spiegata dai valori della variabile indipendente.** E' quindi analogo, in quanto svolge una funzione simile, all'indice  $R$  o  $r^2$  della regressione lineare semplice.

### 15.6. ODDS RATIO: RAPPORTO DI PROBABILITA' O TRA RISCHI.

Il test chi quadrato in tabelle 2 x 2 e il metodo esatto di Fisher per piccoli campioni servono per **confrontare le frequenze tra due campioni indipendenti**. L'ipotesi nulla  $H_0$  è che, nei due differenti gruppi, le frequenze siano uguali: sono due campioni estratti dalla medesima popolazione o da due popolazioni con la stessa frequenza relativa.

In varie situazioni, per verificare l'effetto di una situazione alterata o disturbata rispetto a quella ritenuta normale, il **confronto viene fatto con un caso di controllo**. Diventa opportuno misurare l'effetto del caso alterato, con frequenza relativa campionaria  $p_1$ , rispetto alla situazione di controllo, con frequenza campionaria  $p_2$ , attraverso il loro **rapporto R**

$$R = p_1/p_2$$

chiamato **rischio relativo, rapporto tra rischi o rapporto di probabilità**.

Con un campione di  $r_1$  individui affetti, su una numerosità  $n_1$ , e un caso di controllo con  $r_2$  individui affetti, su un campione di numerosità  $n_2$ , il rapporto  $R = p_1/p_2$  è dato da

$$R = \frac{r_1/n_1}{r_2/n_2}$$

Il suo errore standard **es**, stimato e presentato con i logaritmi naturali (**ln**), è

$$es(\ln R) = \sqrt{\frac{1}{r_1} - \frac{1}{n_1} + \frac{1}{r_2} - \frac{1}{n_2}}$$

mentre l'**intervallo di confidenza**, chiamato **limiti logit** e calcolato con l'esponenziale  $e^x$ ; alla probabilità 0.95 è

$$\exp(\ln R \pm 1,96 es(\ln R))$$

In altri **studi caso-controllo**, non è possibile calcolare direttamente il rischio relativo ma il **rapporto tra odds**, il **rapporto incrociato OR**, che fornisce una buona approssimazione del rapporto tra i rischi assoluti.

Il rapporto incrociato OR è definito come

$$OR = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

Con dati riportati in una **tabella di contingenza 2 x 2**

Campioni	Risposte		Totale
	+	-	
Caso	a	b	n <sub>3</sub>
Controllo	c	d	n <sub>4</sub>
Totale	n <sub>1</sub>	n <sub>2</sub>	N

(con la solita simbologia) **OR è dato** da

$$OR = \frac{ad}{bc}$$

ed il suo **errore standard**  $es(\ln OR)$

$$es(\ln OR) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Per stimare l'**intervallo fiduciale** di un **OR**, si deve

- trasformare il suo valore in **lnOR**,
- calcolare l'intervallo fiduciale alla probabilità  $\alpha$  prefissata, che alla probabilità 0,95 è

$$\ln OR \pm 1,96 \times es(\ln OR)$$

- riconvertire i **ln** dei due valori estremi **X<sub>1</sub>** e **X<sub>2</sub>** dell'intervallo fiduciale in **OR**, mediante elevamento a potenza di **2,7183<sup>lnX</sup>**

#### ESEMPIO

Si vuole verificare se in una popolazione anziana, che da vari anni risiede in una zona ad alto inquinamento, le malattie polmonari hanno un'incidenza maggiore rispetto a quella che risiede in una zona con inquinamento basso. La raccolta dei dati ha fornito i seguenti risultati

Inquinamento	Malattia		Totale
	Si	No	
Alto	236	524	760
Basso	104	634	738
Totale	340	1158	1498

Stimare l'intervallo fiduciale alla probabilità 0,95 e 0,99 del rischio relativo di essere affetti da malattie polmonari, per la popolazione che da anni risiede nella zona ad alto inquinamento.

Risposta

Il **rapporto incrociato o odds ratio (OR)**

tra le popolazioni che vivono nelle due zone diverse è

$$OR = \frac{236 \times 634}{524 \times 104} = \frac{149624}{54496} = 2,75$$

uguale a 2,75.

Di questa stima media è possibile **calcolare l'intervallo fiduciale**.

A questo scopo serve il suo logaritmo naturale

$$\ln OR = \ln 2,75 = 1,01$$

e l'errore standard

$$es(\ln OR) = \sqrt{\frac{1}{236} + \frac{1}{524} + \frac{1}{104} + \frac{1}{634}} = \sqrt{0,01735} = 0,1317$$

$es(\ln OR)$ , sempre espresso in logaritmo naturale, risulta uguale a 0,1317.

L'intervallo di confidenza o fiduciale deve dapprima essere calcolato per  $\ln OR$  e successivamente essere riportato al valore del rapporto.

Con i dati dell'esempio,

alla probabilità 0,95 il **ln dell'intervallo fiduciale di OR**

$$\ln OR = 1,01 \pm 1,96 \times 0,1317 = 1,01 \pm 0,258$$

risulta compreso tra 0,752 e 1,268.

Di conseguenza, alla probabilità del 95% il valore del rapporto OR, cioè il rischio relativo o rapporto tra rischi per le persone che risiedono in una zona ad alto inquinamento rispetto a quelli che risiedono in una zona a basso inquinamento,

con questi dati campionari risulta compreso tra  $2,7183^{0,752}$  e  $2,7183^{1,268}$

corrispondente ad un campo di variazione che come limiti ha 2,12 e 3,55.

## 15.7. REGRESSIONE NON PARAMETRICA

Calcolata la retta di regressione con il metodo dei minimi quadrati,

$$\hat{Y}_i = a + b \cdot X_i$$

**per la sua significatività è possibile ricorrere a test non parametrici**, quando non sono rispettate le condizioni di validità per il **test t** o per il test **F**.

E' il caso in cui i valori di **Y** non hanno una distribuzione normale (come succede con la presenza di uno o più valori anomali), i dati sono espressi in percentuale e/o come rapporti rispetto a quantità fortemente variabili (quindi non hanno lo stesso intervallo fiduciale) oppure sono un punteggio (o comunque una scala di rango).

Resta sempre il problema che, con pochi dati, diventa impossibile dimostrare la normalità della distribuzione, se non già confermata in altre ricerche.

In tale situazione di semplice incertezza o di sicurezza della non validità della regressione con il metodo parametrico dei minimi quadrati, è sempre conveniente, quando non necessario, calcolare **una retta di regressione non parametrica, con la metodologia proposta da H. Theil nel 1950.**

Successivamente, per la **significatività** sia del coefficiente angolare **b** sia dell'intercetta **a**, calcolati in qualsiasi modo, è possibile ricorrere a test non parametrici.

Tra i più diffusi, è da ricordare il **test di Theil (a distribution-free test for the slope coefficient)**, proposto con vari articoli a partire dal 1950, che può essere utilizzato per verificare:

- la significatività di una retta **b** di regressione, con ipotesi nulla

$$\mathbf{H}_0: b = 0$$

ed ipotesi alternativa **H<sub>1</sub>** sia bilaterale che unilaterale;

- la significatività della differenza tra una retta campionaria **b** ed un coefficiente angolare **b<sub>0</sub>** atteso, con ipotesi nulla

$$\mathbf{H}_0: b = b_0 \text{ (prefissato)}$$

ed ipotesi alternativa **H<sub>1</sub>** che anche in questo caso può essere sia unilaterale che bilaterale.

Questa seconda situazione è semplicemente una generalizzazione del caso precedente. Le due metodologie sono quindi identiche, con la sola eccezione di un passaggio preliminare.

Un altro test utile all'inferenza sulle rette è il **test di Hollander (a distribution-free test for the parallelism of two regression lines)**, proposto dalla seconda metà degli anni '60, per verificare

- il parallelismo tra due coefficienti angolari **b<sub>1</sub>** e **b<sub>2</sub>**, e quindi l'ipotesi nulla

$$\mathbf{H}_0: b_1 = b_2$$

e l'ipotesi alternativa sempre sia unilaterale che bilaterale.

Con altri test non parametrici sulla regressione lineare semplice, inoltre è possibile:

- con il **test di Maritz**, valutare la significatività dell'intercetta **a**

$$H_0: a = a_0$$

- con il test di **Lancaster-Quade**, verificare congiuntamente ed in modo indipendente la significatività sia di **a** sia di **b** nelle loro 3 combinazioni:  
(significatività sia di **a** che di **b**, solo di **a**, solo di **b**),
- avere stime ed intervalli di confidenza per i parametri **a** e **b**.

Nella letteratura sulla regressione non parametrica, si trovano altri test, seppure attualmente meno diffusi nei programmi informatici e nei testi, per verificare le ipotesi precedenti.

E' possibile ricordare il test di Brown e Mood, quello di Hajek, quello di Adichie e quello di Sievers, tutti sulla linearità e quindi alternativi al test di Theil;

altri test, proposti uno ancora da Brown e Mood e un altro ancora da Adichie, possono essere usati in alternativa a quello di Maritz;

Daniels, Konijn e Quade separatamente hanno proposto metodi alternativi a quello di Lancaster-Quade.

## 15.8. CALCOLO DELLA RETTA DI REGRESSIONE NON PARAMETRICA CON IL METODO DI THEIL

Nel 1950, H. **Theil** (con i tre articoli *A rank invariant method of linear and polynomial regression analysis, I, II, III*, pubblicati su *Proc. Kon. Nederl. Akad. Wetensch. A.*, 53: 386-392, 521-525, 1397-1412) ha proposto un metodo per **calcolare una retta di regressione non parametrica** (*Theil's regression method*). La procedura proposta si fonda sulla **mediana di tutte le rette, calcolate sulle possibili coppie di punti**.

Per ognuna di esse, identificate dalle coppie di variabili (**X<sub>i</sub>**, **Y<sub>i</sub>**) e (**X<sub>j</sub>**, **Y<sub>j</sub>**), si stima il coefficiente angolare **b<sub>ij</sub>**, con la relazione

$$b_{ij} = \frac{Y_j - Y_i}{X_j - X_i}$$

Poiché **b<sub>ij</sub>** è uguale a **b<sub>ji</sub>** con la sola inversione del segno, si devono **quantificare tutte le possibili combinazioni** delle **N** osservazioni in cui **j** è maggiore di **i**. Il valore di **b** è stimato dalla mediana (indicata con **b\***) di questi valori che, con **N** coppie d'osservazioni, sono **N(N-1)/2**.

Per l'inferenza si assume che

$$b = b^*$$

In assenza di un programma informatico (che pochissime librerie statistiche fino ad ora riportano), con il calcolo manuale questo metodo richiede molto tempo, quando il numero di punti diventa alto. Per esempio, già con 13 coppie di dati ( $N = 13$ ) il numero di **bij** sui quali stimare la mediana diventa 72 ( $13 \times 12 / 2$ ).

Per N maggiore di 12, lo stesso Theil ha proposto un metodo abbreviato, chiamato appunto **metodo abbreviato di Theil** (*the abbreviated Theil method*), che **richiede un numero d'operazioni nettamente minore, ma non fornisce lo stesso risultato del metodo precedente**. E' quindi da utilizzare nell'impossibilità pratica di ricorrere al primo, poiché sfrutta solo in modo parziale l'informazione contenuta nell'insieme dei dati.

Per avvalersi della procedura abbreviata, dopo aver posto i dati in ordine crescente per la variabile **X**, si conta il numero **N** di coppie di valori. Il metodo differisce leggermente, se **N** è pari o dispari.

Per N pari, dopo aver separato i dati in due metà esatte, si devono calcolare **N/2** differenze sia per la variabile **X** con

$$\mathbf{X_{ij}} = \mathbf{X_{(i + N/2)}} - \mathbf{X_i}$$

sia per la variabile **Y** con

$$\mathbf{Y_{ij}} = \mathbf{Y_{(i + N/2)}} - \mathbf{Y_i}$$

Successivamente, occorre individuare

- sia la mediana delle **N/2** differenze della **X**,
- sia la mediana delle **N/2** differenze della **Y**;

Il valore della retta **b\*** è il rapporto tra queste due mediane

$$\mathbf{b^* = mediana Y_{ij} / mediana X_{ij}}$$

Per N dispari, sempre

- dopo aver ordinato la serie dei dati originari in modo crescente per **X**,
- si elimina una coppia di valori a caso.

Alcuni autori suggeriscono la coppia di valori **X** e **Y** corrispondenti alla mediana di **X**; di conseguenza, **il numero di punti diventa pari e si ricade nel metodo precedente**.

Ottenuto il valore di **b\*** come miglior stima di **b**, è possibile calcolare il valore dell'intercetta **a** come miglior stima di **a**, con due metodi che seguono logiche diverse:

1 - in modo analogo al primo metodo descritto per **b\***, dapprima si calcolano tutte le **a<sub>i</sub>**, che in questo caso sono **N** come i punti (**X<sub>i</sub>**, **Y<sub>i</sub>**) rilevati

$$a_i = Y_i - b \cdot X_i$$

e successivamente si stima  $a^*$  come mediana delle  $N$  intercette  $a_i$  calcolate;

2 - in modo analogo alla statistica parametrica, in cui

$$a = \bar{Y} - b \cdot \bar{X}$$

si stima  $a$ , indicata appunto con  $\hat{a}$  per distinguerla dal valore parametrico, sostituendo nella formula precedente la mediana delle  $X$  e quella delle  $Y$  alle medie rispettive

$$\hat{a} = \text{mediana}(Y_i) - b^* \cdot \text{mediana}(X_i)$$

Con questo metodo, più conveniente nel caso di grandi campioni perché più rapido, si ottiene una **retta che passa non per l'incrocio delle medie ma per l'incrocio delle mediane, considerato il baricentro non parametrico del diagramma di dispersione dei punti.**

**I due metodi possono dare risultati differenti.**

La metodologia descritta per il **calcolo della retta di regressione lineare non parametrica** può essere spiegata in modo più semplice, comprensibile anche a non esperti di statistica, illustrando un esempio in tutti i suoi passaggi.

Il metodo esteso, applicabile quando si dispone di piccoli campioni, sarà presentato separatamente da quello abbreviato con un secondo esempio.

#### **Metodo per campioni piccoli (N £ 12).**

Si supponga di voler valutare gli effetti di 7 dosaggi di una sostanza tossica ( $X$ ), su vari campioni di una popolazione animale. La dose 0 (zero), detto anche campione bianco, corrisponde all'assenza del principio attivo; sovente serve come controllo. I risultati sono stati misurati come percentuale d'individui morti ( $Y$ ) su una serie di somministrazioni, ottenendo i seguenti dati:

Rango delle X	1	2	3	4	5	6	7
Valori di X	0	1	2	3	4	5	6
Valori di Y	2,9	3,1	3,4	4,0	4,6	5,1	12,4

Per calcolare il coefficiente angolare  $b$  mediante il metodo proposto da **Theil**, è utile seguire le procedure di seguito descritte:

1 – ordinare i valori di  $\mathbf{X}$  in modo crescente (operazione già effettuata nella tabella di presentazione dei dati, trattandosi di dosi crescenti);

2 – quantificare le possibili combinazioni ( $\mathbf{N} (\mathbf{N} - \mathbf{1}) / \mathbf{2}$ ),

che in questo caso sono 21 ( $7 \times 6 / 2$ )

e per ognuna calcolare il coefficiente angolare  $\mathbf{bij}$  mediante la relazione

$$b_{ij} = \frac{Y_j - Y_i}{X_j - X_i}$$

con  $\mathbf{j}$  maggiore di  $\mathbf{i}$ .

Di seguito, sono riportati tutti i risultati e i calcoli relativi nei loro passaggi:

$$b_{12} = (Y_2 - Y_1) / (X_2 - X_1) = (3,1 - 2,9) / (1 - 0) = 0,200$$

$$b_{13} = (Y_3 - Y_1) / (X_3 - X_1) = (3,4 - 2,9) / (2 - 0) = 0,250$$

$$b_{14} = (Y_4 - Y_1) / (X_4 - X_1) = (4,0 - 2,9) / (3 - 0) = 0,367$$

$$b_{15} = (Y_5 - Y_1) / (X_5 - X_1) = (4,6 - 2,9) / (4 - 0) = 0,425$$

$$b_{16} = (Y_6 - Y_1) / (X_6 - X_1) = (5,1 - 2,9) / (5 - 0) = 0,440$$

$$b_{17} = (Y_7 - Y_1) / (X_7 - X_1) = (12,4 - 2,9) / (6 - 0) = 1,583$$

$$b_{23} = (Y_3 - Y_2) / (X_3 - X_2) = (3,4 - 3,1) / (2 - 1) = 0,300$$

$$b_{24} = (Y_4 - Y_2) / (X_4 - X_2) = (4,0 - 3,1) / (3 - 1) = 0,450$$

$$b_{25} = (Y_5 - Y_2) / (X_5 - X_2) = (4,6 - 3,1) / (4 - 1) = 0,500$$

$$b_{26} = (Y_6 - Y_2) / (X_6 - X_2) = (5,1 - 3,1) / (5 - 1) = 0,500$$

$$b_{27} = (Y_7 - Y_2) / (X_7 - X_2) = (12,4 - 3,1) / (6 - 1) = 1,860$$

$$b_{34} = (Y_4 - Y_3) / (X_4 - X_3) = (4,0 - 3,4) / (3 - 2) = 0,600$$

$$b_{35} = (Y_5 - Y_3) / (X_5 - X_3) = (4,6 - 3,4) / (4 - 2) = 0,600$$

$$b_{36} = (Y_6 - Y_3) / (X_6 - X_3) = (5,1 - 3,4) / (5 - 2) = 0,567$$

$$b_{37} = (Y_7 - Y_3) / (X_7 - X_3) = (12,4 - 3,4) / (6 - 2) = 2,250$$

$$b_{45} = (Y_5 - Y_4) / (X_5 - X_4) = (4,6 - 4,0) / (4 - 3) = 0,600$$

$$b_{46} = (Y_6 - Y_4) / (X_6 - X_4) = (5,1 - 4,0) / (5 - 3) = 0,550$$

$$b_{47} = (Y_7 - Y_4) / (X_7 - X_4) = (12,4 - 4,0) / (6 - 3) = 2,800$$

$$b_{56} = (Y_6 - Y_5) / (X_6 - X_5) = (5,1 - 4,6) / (6 - 5) = 0,500$$

$$b_{57} = (Y_7 - Y_5) / (X_7 - X_5) = (12,4 - 4,6) / (7 - 5) = 3,900$$

$$b_{67} = (Y_7 - Y_6) / (X_7 - X_6) = (12,4 - 5,1) / (7 - 6) = 7,300$$

Questi risultati solitamente sono pubblicati in modo più sintetico, sotto forma di una matrice triangolare come la seguente:

X	0	1	2	3	4	5	6
Y	2,9	3,1	3,4	4,0	4,6	5,1	12,4
X = 0 ; Y = 2,9		0,200	0,250	0,367	0,425	0,440	1,583
X = 1 ; Y = 3,1			0,300	0,450	0,500	0,500	1,860
X = 2 ; Y = 3,4				0,600	0,600	0,567	2,250
X = 3 ; Y = 4,0					0,600	0,550	2,800
X = 4 ; Y = 4,6						0,500	3,900
X = 5 ; Y = 5,1							7,300
X = 6 ; Y = 12,4							

3 – stimare la mediana di questi  $N(N-1)/2$  valori **b<sub>ij</sub>**; è facilmente identificata dalla sua serie ordinata per ranghi:

Rango	1	2	3	4	5	6	7	8	9	10	<b>11</b>
<b>b<sub>ij</sub></b>	0,200	0,250	0,300	0,367	0,425	0,440	0,450	0,500	0,500	0,500	<b>0,550</b>

Rango	12	13	14	15	16	17	18	19	20	21
<b>b<sub>ij</sub></b>	0,567	0,600	0,600	0,600	1,583	1,860	2,250	2,800	3,900	7,900

In questo esempio, la mediana risulta uguale a 0,550 corrispondendo alla 11<sup>a</sup> posizione sulle 21 misure stimate; di conseguenza, si assume

$$b^* = 0,550$$

4 – con **N** uguale a 7,  
dal valore di **b\*** e mediante la relazione

$$a_i = Y_i - b^*X_i$$

si calcolano altrettanti valori delle intercette **a<sub>i</sub>**.

Di seguito, sono riportati tutti i risultati con i dati dell'esempio:

$$\begin{aligned}
 a_1 &= 2,9 - 0,550 \times 0 = 2,90 \\
 a_2 &= 3,1 - 0,550 \times 1 = 2,55 \\
 a_3 &= 3,4 - 0,550 \times 2 = 2,30 \\
 a_4 &= 4,0 - 0,550 \times 3 = 2,35 \\
 a_5 &= 4,6 - 0,550 \times 4 = 2,40 \\
 a_6 &= 5,1 - 0,550 \times 5 = 2,35 \\
 a_7 &= 12,4 - 0,550 \times 6 = 9,10
 \end{aligned}$$

5 – la mediana di questi  $N$  valori  $a_i$  è identificata dalla sua serie ordinata per rango:

Rango	1	2	3	4	5	6	7
<b>ai</b>	2,30	2,35	2,35	<b>2,40</b>	2,55	2,90	9,10

Coincidendo con il quarto dei 7 valori, è uguale a 2,40.

Di conseguenza,

$$a^* = 2,40$$

6 – la retta calcolata in modo non parametrico, nella sua forma estesa, per i 7 punti rilevati è

$$\hat{Y}_i = 2,4 + 0,55 \cdot X_i$$

**Un modo alternativo per calcolare l'intercetta  $a$**  è:

1- individuare nei dati originari

Valori di X	0	1	2	<b>3</b>	4	5	6
Valori di Y	2,9	3,1	3,4	<b>4,0</b>	4,6	5,1	12,4

la mediana delle  $X_i$ , che risulta uguale a 3, e

la mediana delle  $Y_i$ , che risulta uguale a 4,0;

2 – dalla relazione

$$\hat{a} = \text{mediana}(Y_i) - b^* \cdot \text{mediana}(X_i)$$

calcolare il valore di  $\hat{a}$

$$\hat{a} = 4,0 - 0,55 \times 3 = 4,0 - 1,65 = 2,35$$

che risulta uguale a **2,35**.

**Metodo per grandi campioni** ( $N > 12$ ).

Si supponga di voler valutare la crescita media di una specie animale con l'aumentare dell'età. A questo scopo, sono stati raccolti campioni d'individui dall'età 4 all'età 20, stimando per ognuno la lunghezza media del campione:

Età X	4	5	6	7	8	9	10	11	12
Lungh. Y	40	45	51	55	60	67	68	65	71

Età X	13	14	15	16	17	18	19	20
Lungh. Y	74	76	76	78	83	82	85	89

Con 17 osservazioni, il metodo di Theil nella versione estesa richiederebbe il calcolo di 136 ( $17 \times 16 / 2$ ) coefficienti angolari **b<sub>ij</sub>**. Per effettuare l'operazione in tempi non eccessivamente lunghi, è conveniente utilizzare

il **metodo abbreviato**.

1 – Si ordinano i dati della variabile **X** per rango (spesso è un'operazione già effettuata nella tabella di presentazione, come in questo caso).

2 - Si individua la mediana delle **X**: su 17 dati è il nono valore e corrisponde al punto ( $X = 12$ ;  $Y = 71$ );

successivamente si calcolano le 8 differenze

- sia per la variabile **X** tra i valori **X(i + N/2)** e **X<sub>i</sub>**:

$$X_{10} - X_1 = 13 - 4 = 9$$

$$X_{11} - X_2 = 14 - 5 = 9$$

.....

$$X_{17} - X_8 = 20 - 11 = 9$$

(nel caso specifico sono tutte uguali a 9)

- sia per la variabile **Y** tra i valori **Y(i + N/2)** e **Y<sub>i</sub>**:

$$\begin{aligned}
Y_{10} - Y_1 &= 74 - 40 = 34 \\
Y_{11} - Y_2 &= 76 - 45 = 31 \\
Y_{12} - Y_3 &= 76 - 51 = 25 \\
Y_{13} - Y_4 &= 78 - 55 = 23 \\
Y_{14} - Y_5 &= 83 - 60 = 23 \\
Y_{15} - Y_6 &= 82 - 67 = 15 \\
Y_{16} - Y_7 &= 85 - 68 = 17 \\
Y_{17} - Y_8 &= 89 - 65 = 24
\end{aligned}$$

3 – Nelle 8 differenze di **X** e di **Y** si scelgono le 2 mediane:

- per **X<sub>ij</sub>**, essendo in questo caso tutte uguali a 9, ovviamente la mediana è 9;
- per **Y<sub>ij</sub>** conviene ordinare gli 8 valori in modo crescente

$$15 \quad 17 \quad 23 \quad 23 \quad 24 \quad 25 \quad 31 \quad 34$$

e dalla serie ordinata per rango emerge che la mediana cade tra 23 (4° valore) e 24 (5° valore) e quindi è 23,5.

4 – Utilizzando la formula

$$\mathbf{b^* = mediana Y_{ij} / mediana X_{ij}}$$

si ottiene un valore di **b\***

$$\mathbf{b^* = 23,5 / 9 = 2,611}$$

uguale a 2,611.

5 – Da **b\*** si stima  $\hat{a}$ , ovviamente con il metodo più breve: dopo aver identificato, sulla serie dei dati originari,

la mediana delle **X** che risulta uguale a 12 e

la mediana delle **Y** che risulta uguale a 71,

attraverso la relazione

$$\hat{a} = \text{mediana}(Y_i) - \mathbf{b^*} \cdot \text{mediana}(X_i)$$

si calcola il valore di  $\hat{a}$

$$\hat{a} = 71 - 2,611 \times 12 = 71 - 31,332 = 39,668$$

che risulta uguale a **39,668**

6 – **La retta di regressione lineare semplice non parametrica con il metodo abbreviato di Theil**

risulta

$$\hat{Y}_i = 39,668 + 2,661 \cdot X_i$$

### 15.9. CONFRONTO TRA LA RETTA PARAMETRICA E LA RETTA DI THEIL

Ai fini di un'interpretazione corretta dei risultati, è sempre utile costruire il diagramma di dispersione dei punti campionari, con la retta relativa. Per una scelta ragionata, è necessario comprendere esattamente sia le caratteristiche della retta di regressione lineare semplice non parametrica, calcolata con il metodo di Theil, sia le differenze rispetto a quella parametrica, calcolata con il principio dei minimi quadrati.

A questo scopo, dopo aver ripreso i dati dell'esempio precedente

Valori di X	0	1	2	<b>3</b>	4	5	6
Valori di Y	2,9	3,1	3,4	<b>4,0</b>	4,6	5,1	12,4

- sui quali è stata calcolata la regressione di Theil nella versione estesa:

$$\hat{Y}_i = 2,4 + 0,55 \cdot X_i$$

- si calcola anche la retta di regressione parametrica, che con i parametri stimati dai dati della tabella,

$$\sum X \cdot Y = 140,2 \quad \sum X = 21 \quad \sum Y = 35,5 \quad \sum X^2 = 91 \quad n = 7 \quad \bar{X} = 3 \quad \bar{Y} = 5,07$$

applicati alla formula per il coefficiente angolare

$$b = \frac{\sum X \cdot Y - \frac{\sum X \cdot \sum Y}{N}}{\sum X^2 - \frac{(\sum X)^2}{N}}$$

si ottiene un valore di **b**

$$b = \frac{140,2 - \frac{21 \times 35,5}{7}}{91 - \frac{21^2}{7}} = \frac{140,2 - 106,5}{91 - 63} = \frac{33,7}{28} = 1,20$$

uguale a 1,20 e da esso un valore di **a**

$$a = \bar{Y} - b \cdot \bar{X}$$

$$a = 5,07 - 1,20 \times 3 = 5,07 - 3,60 = 1,47$$

uguale a 1,47

per cui la retta parametrica è

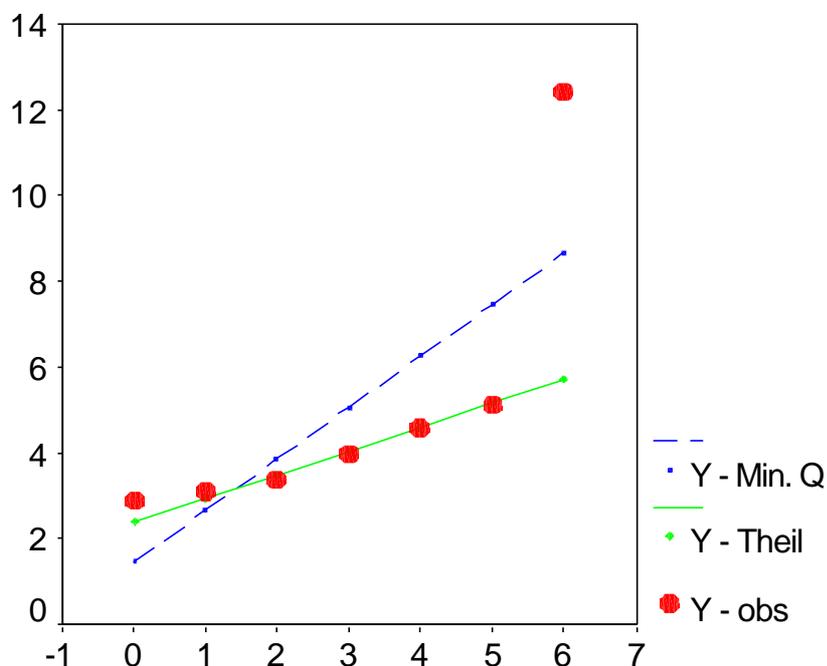
$$\hat{Y}_i = 1,47 + 1,20 \cdot X_i$$

Sia il confronto visivo tra i punti sperimentali e le due rette, sia la differenza tra i valori di Y osservati e quelli calcolati con la retta di Theil e quella dei minimi quadrati

Valori di Y osservati	2,90	3,10	3,40	4,00	4,60	5,10	12,40
Y calcolati con Theil	2,40	2,95	3,50	4,05	4,60	5,15	5,70
Y calcolati con minimi quadrati	1,47	2,67	3,87	5,07	6,27	7,47	8,67

evidenzia come l'ultima osservazione sia anomala, rispetto alle altre 5; inoltre come la retta non parametrica si avvicini molto ai primi 5 punti ignorando praticamente l'ultimo dato, mentre la retta parametrica sia da esso attratta, allontanandosi dagli altri.

E' un comportamento simile a quello della mediana e della media, rispetto ai valori anomali; in questo caso, è accentuato dal fatto che **la retta parametrica minimizza la somma dei quadrati degli scarti**. Oltre che sulle condizioni di validità, la scelta dipende quindi dal valore che si vuole attribuire all'osservazione che risulta anomala, rispetto a tutte le altre.



Nel grafico,

- i punti sono i 7 valori osservati,

- la retta che incrocia i 6 punti e si avvicina al valore anomalo è la retta parametrica calcolata con il metodo dei minimi quadrati,
- la retta che passa per i 6 punti ed ignora il valore anomalo è quella non parametrica, calcolata con il metodo di Theil.

### 15.10. TEST DI THEIL PER LA SIGNIFICATIVITA' DI **b**

Il **test di Theil** per la significatività del coefficiente **b** di regressione lineare semplice, con verifica dell'ipotesi nulla

$$H_0: b = 0$$

può avere sia un'ipotesi alternative bilaterale

$$H_1: b \neq 0$$

sia una delle due ipotesi unilaterali

$$H_1: b < 0 \quad \text{oppure} \quad H_1: b > 0$$

in funzione della conoscenza del problema e quindi della domanda formulata.

La metodologia, che si ricollega al test  $t$  (tau) di Kendall di correlazione non parametrica, può essere spiegata in modo semplice con l'elaborazione dei dati di un esempio. Al fine di evidenziare le analogie con il test parametrico e per un successivo confronto dei risultati, in quest'esempio sono riproposti gli stessi dati utilizzati per il calcolo della regressione lineare semplice.

1 - Con i dati raccolti, dopo aver individuata la variabile dipendente **Y** e la variabile indipendente o causale **X**,

Individui	A	B	C	D	E	F	G
<b>Peso (Y)</b>	52	68	75	71	63	59	57
<b>Altezza (X)</b>	160	178	183	180	166	175	162

2 - si stima il valore del coefficiente angolare **b**, che risulta

$$b = \frac{\sum (x \cdot y) - \frac{\sum x \cdot \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{76945 - \frac{1204 \cdot 445}{7}}{207598 - \frac{1204^2}{7}} = 0,796$$

uguale a 0,796

Anche se ininfluyente sul problema specifico, l'intercetta **a**

$$a = \bar{Y} - b\bar{X} = 63,571 - 0,796 \cdot 172 = -73,354$$

che risulta uguale a -73,354

è utile per scrivere in modo completo la retta di regressione

$$\hat{Y}_i = -73,354 + 0,796 \cdot X_i$$

3 - Per valutare la significatività del coefficiente angolare **b**, con ipotesi nulla

$$H_0: b = 0$$

e, in questo caso, con ipotesi alternativa bilaterale

$$H_1: b \neq 0$$

i dati devono essere ordinati secondo il rango di **X** (l'altezza),

<b>Individui</b>	A	G	E	F	B	D	C
<b>Peso (Y)</b>	<b>52</b>	<b>57</b>	<b>63</b>	<b>59</b>	<b>68</b>	<b>71</b>	<b>75</b>
<b>Altezza (X)</b>	160	162	166	175	178	180	183

riportando i dati di **Y** relativi, successivamente trasformati in ranghi

<b>Individui</b>	A	G	E	F	B	D	C
<b>Peso (Y)</b>	52	57	63	59	68	71	75
<b>Y in ranghi</b>	<b>1</b>	<b>2</b>	<b>4</b>	<b>3</b>	<b>5</b>	<b>6</b>	<b>7</b>

4 - Ponendo l'attenzione solo sui valori della Y, se all'aumentare di X

- i valori di Y tendono ad aumentare (quindi i ranghi di Y sono in ordine naturale), la regressione tende ad essere significativa, con coefficiente angolare positivo,

- il valore di Y resta approssimativamente costante (quindi i dati di Y sono in ordine casuale), la regressione è assente o non significativa,
- il valore di Y tende a diminuire ( quindi i ranghi di Y sono in ordine decrescente), la regressione tende ad essere significativa, con coefficiente angolare negativo.

Per quantificare il grado di correlazione o concordanza dei ranghi di Y con l'ordine naturale, si può utilizzare la proposta di Kendall: contare **quante sono le coppie di ranghi che sono concordanti e quante quelle discordanti dall'ordine naturale**.

Per un calcolo corretto, facilmente verificabile, è utile riportare il conteggio dettagliato e complessivo delle concordanze (+) e delle discordanze (-)

1	2	4	3	5	6	7	Totale
	+	+	+	+	+	+	<b>+6</b>
		+	+	+	+	+	<b>+5</b>
			-	+	+	+	<b>+2</b>
				+	+	+	<b>+3</b>
					+	+	<b>+2</b>
						+	<b>+1</b>
<b>Totale (concordanze meno discordanze)</b>							<b>+19</b>

**La misura della concordanza complessiva con la variabile X è dato dalla somma algebrica di tutte le concordanze e le discordanze.**

Il totale di concordanze e discordanze dei 7 valori dell'esempio è **+19**.

5 - Il numero totale di concordanze e discordanze di una serie di valori deve essere rapportato al numero **massimo totale possibile**. Poiché i confronti sono fatti a coppie, con **N** dati il numero totale di confronti concordanti o discordanti è dato dalla combinazione di **N** elementi **2 a 2**

$$C_N^2$$

Con una serie di 7 dati come nell'esempio, il numero complessivo di confronti, quindi il massimo totale possibile di concordanze o discordanze, è

$$C_7^2 = \frac{7!}{(7-2)! \cdot 2!}$$

Secondo il metodo proposto di Kendall, il grado di relazione o concordanza ( $\tau$ ) tra la variabile X e Y può essere quantificato dal rapporto

$$t = \frac{\text{totale}(\text{concordanze} - \text{discordanze})}{\text{massimo totale possibile}} = \frac{\text{totale}(\text{concordanze} - \text{discordanze})}{C_N^2}$$

Con i 7 dati riportati,

$$t = \frac{+19}{21} = +0,905$$

è uguale a +0,905.

6 - La scala dovrebbe essere continua, ma sono accettati valori discreti. In caso di due o più valori identici, il confronto tra due punteggi di Y uguali non determina né una concordanza né una discordanza: il loro confronto non contribuisce al calcolo di  $\tau$ , abbassando il valore al numeratore. Di conseguenza, deve essere diminuito anche il valore al denominatore.

**La formula, corretta per la presenza di valori identici (ties), diventa**

$$t = \frac{2 \cdot \text{totale}(\text{concordanze} - \text{discordanze})}{\sqrt{N \cdot (N - 1) \cdot T_x} \cdot \sqrt{N \cdot (N - 1) \cdot T_y}}$$

dove

$N$  è il numero totale di coppie di dati delle variabili **X** e **Y**,

$T_x$  è dato da  $T_x = \sum (t_x^2 - t_x)$  dove  $t_x$  è il numero di osservazioni identiche di ogni gruppo di valori identici della variabile X,

$T_y$  è dato da  $T_y = \sum (t_y^2 - t_y)$  dove  $t_y$  è il numero di osservazioni identiche di ogni gruppo di valori identici della variabile Y.

Per **piccoli campioni**, i valori critici sono forniti dalla tabella relativa, riportata nel paragrafo della correlazione non parametrica t di Kendall.

Con 7 dati, alla probabilità 0.005 per un test ad una coda e 0.01 per un test a 2 code, il valore critico riportato è 0.810. Il valore calcolato è superiore a quello della tabella: si rifiuta l'ipotesi nulla. In un test bilaterale la risposta dovrebbe essere: il coefficiente di regressione lineare b si discosta significativamente da 0. Se la domanda fosse stata unilaterale positiva, la risposta dovrebbe essere: all'aumentare dell'altezza il peso aumenta in modo significativo.

Per **grandi campioni** la significatività del  $\tau$  di Kendall può essere saggiata con la distribuzione normale

$$z = \frac{\tau - \mu_\tau}{\sigma_\tau} \quad (*)$$

con

$$\mu_\tau = 0$$

e

$$s_t^2 = \frac{2 \cdot (2N + 5)}{9N \cdot (N - 1)}$$

dove  $N$  è il numero di dati.

Sostituendo e semplificando, si ottiene una stima più rapida di  $z$  mediante la relazione

$$z = \frac{3t \cdot \sqrt{N \cdot (N - 1)}}{\sqrt{2 \cdot (2N + 5)}}$$

Per la **verifica dell'ipotesi nulla più generale**

$$H_0: b = b_0$$

(con ipotesi alternativa sia unilaterale che bilaterale) dove  $b_0$  è un coefficiente angolare qualsiasi, quindi anche 0, il metodo di Theil ha una leggera modifica nelle prime fasi.

Utilizziamo il medesimo esempio ed evidenziamo i metodi con i vari passaggi.

1 – Si supponga di avere già calcolato il coefficiente angolare  $b$ , (uguale a 0,796) e di volere verificare l'ipotesi se esso non si discosti dal valore  $b_0 = 0,9$  (aumento di Kg 0,9 in peso per l'aumento di 1 cm. in altezza)

$$H_0: \beta = 0,9$$

contro l'ipotesi alternativa che la crescita media in peso sia minore

$$H_1: \beta < 0,9$$

2 – Il passo successivo consiste nel calcolare i valori attesi di  $Y$ , secondo la relazione

$$\hat{Y}_i = a + 0,9 \cdot X_i$$

Tuttavia, poiché l'interesse è rivolto non ai singoli valori ma al loro rango, i valori attesi possono essere più semplicemente calcolati come

$$\hat{Y}_i = 0,9 \cdot X_i$$

con formula più generale

$$\hat{Y}_i = \beta_0 \cdot X_i$$

Con i dati dell'esempio

Individui	A	B	C	D	E	F	G
Peso (Y osservati)	52	68	75	71	63	59	57
<b>Y teorici o attesi</b>	<b>144,0</b>	<b>160,2</b>	<b>164,7</b>	<b>162,0</b>	<b>149,4</b>	<b>157,5</b>	<b>145,8</b>
Altezza (X)	160	178	183	180	166	175	162

3 – Calcolare le differenze **Di** tra **Yi attesi** e **Yi osservati**

Individui	A	B	C	D	E	F	G
<b>Di (Yi teorici – Yi attesi)</b>	<b>92,0</b>	<b>92,2</b>	<b>89,7</b>	<b>91,0</b>	<b>86,4</b>	<b>98,5</b>	<b>88,8</b>
Altezza (X)	160	178	183	180	166	175	162

4 – Ordinare i dati secondo il rango di **X** (l'altezza),

Individui	A	G	E	F	B	D	C
<b>Di</b>	<b>92,0</b>	<b>88,8</b>	<b>86,4</b>	<b>98,5</b>	<b>92,2</b>	<b>91,0</b>	<b>89,7</b>
Altezza (X)	160	162	166	175	178	180	183

riportare i dati delle differenze **Di** (nella riga centrale) e successivamente trasformarle in ranghi (come nella tabella seguente)

Individui	A	G	E	F	B	D	c
Di	92,0	88,8	86,4	98,5	92,2	91,0	89,7
<b>Di in ranghi</b>	<b>5</b>	<b>2</b>	<b>1</b>	<b>7</b>	<b>6</b>	<b>4</b>	<b>3</b>

5 – Porre l'attenzione sui soli valori delle differenze **Di (Yi attesi – Yi osservati)**:

- se il valore di **b** è statisticamente uguale a **b<sub>0</sub>**, il rango delle **Di** dipenderà solo dalle variazioni casuali in peso e sarà indipendente dal coefficiente angolare **b** (l'intercetta **a** è una costante),
- se il valore di **b** è statisticamente maggiore di **b<sub>0</sub>**, il rango delle **Di** tenderà ad essere in ordine inverso a quello delle **Xi**,
- se il valore di **b** è statisticamente minore di **b<sub>0</sub>**, il rango delle **Di** tenderà ad avere lo stesso ordine di **Xi**.

6 – Contare **quante sono le coppie di ranghi delle Di che sono concordanti e quante quelle discordanti dall'ordine naturale**:

5	2	1	7	6	4	3	Totale
	-	-	+	+	-	-	-2
		-	+	+	+	+	+3
			+	+	+	+	+4
				-	-	-	-3
					-	-	-2
						-	-1
<b>Totale (concordanze meno discordanze)</b>							<b>-1</b>

**La misura della concordanza complessiva tra Y osservati e Y attesi, quindi di b con b<sub>0</sub>, è data dalla somma totale delle concordanze e discordanze:**

**essa risulta uguale a -1.**

7 – Il valore di t, dato dal rapporto del numero totale delle concordanze meno le discordanze rispetto al numero massimo possibile, risulta

$$t = -1 / 21 = - 0,047$$

eguale a - 0,047.

Il valore calcolato non solo è inferiore a quello riportato nella tabella, ma è molto vicino a quello atteso nell'ipotesi nulla; di conseguenza, si può affermare non solo che **b non si discosta significativamente da b<sub>0</sub>**, ma anche che **b è molto vicino a b<sub>0</sub>**.

Il test di Theil per la significatività del coefficiente angolare  $b$  può essere sia unilaterale che bilaterale; è quindi analogo al test  $t$  di Student.

Per una scelta adeguata tra il test parametrico e il corrispondente non parametrico, è importante conoscere la loro efficienza. Nel caso di grandi campioni, dipende dalla forma di distribuzione dei dati, di solito schematizzata in tre situazioni: normale, rettangolare ed esponenziale doppia.

#### **L'efficienza asintotica del test di Theil rispetto al $t$ di Student**

- è uguale a  $0,95 (3/\pi)$ , quando la distribuzione è Normale;
- è uguale a 1, quando la distribuzione è Rettangolare;
- è uguale a  $1,5 (3/2)$ , quando la distribuzione è Esponenziale Doppia.

In termini elementari, quando la distribuzione è perfettamente normale, i due test hanno efficienza molto simile; ma quando la forma della distribuzione è lontana dalla normalità, il test non parametrico è più efficiente di quello parametrico.

La regressione lineare semplice non parametrica può essere utilizzata in modo appropriato, anche per **analizzare una serie di dati o comunque gli effetti della distanza da un'origine, che può essere di natura qualsiasi, da geografica a temporale.** Già H. B. Mann nel 1945 (con l'articolo *Non parametric test against trend*, sulla rivista *Econometrica* n. 13, pp. 245-259) affermava che quando  $X$  è una misura temporale, che può essere espressa in anni, mesi, giorni, ore o secondi, il test di Theil, con ipotesi  $H_0: \beta = 0$  ed ipotesi alternativa  $H_1$  sia unilaterale che bilaterale, può essere impiegato per **verificare se esiste un trend, cioè una tendenza alla diminuzione o all'aumento del carattere  $Y$ .**

#### **15.11. IL TEST DI HOLLANDER**

Quando in due campioni indipendenti bivariati  $C_1$  e  $C_2$  sono stati calcolati i due coefficienti angolari  $b_1$  e  $b_2$ , può sorgere il problema di verificare se essi siano paralleli:

$$H_0: b_1 = b_2$$

Tra i test proposti (come quello di **R. F. Potthoff** con l'articolo del 1965 *A non parametric test of whether two simple regression lines are parallel* nel volume *University of North Carolina, Institute of Statistics Mimeo series 445*, oppure quello di **P. K. Sen** per la verifica simultanea del parallelismo tra più coefficienti angolari con l'articolo del 1969 *On a class of rank order test for the parallelism of several regression lines* pubblicato in *Ann. Math. Statist*, vol. 40 pp. 1668-1683), attualmente il più diffuso nella ricerca sperimentale è quello illustrato da **M. Hollander** in modo completo nel 1970 (con l'articolo *A distribution-free test for parallelism* su *Journal Amer. Statist. Association* vol. 65, pp.387-394).

Per la significatività, il test di Hollander **utilizza il test dei ranghi con segno di Wilcoxon per due campioni dipendenti**; di conseguenza richiede le stesse condizioni di validità e come ipotesi alternativa accetta sia test bilaterali

$$H_1: b_1 \neq b_2$$

sia unilaterali nelle due differenti direzioni

$$H_1: b_1 < b_2 \quad \text{oppure} \quad H_1: b_1 > b_2$$

secondo la domanda posta dal problema che si analizza.

E' quindi analogo al test parametrico t di Student, per 2 campioni dipendenti.

Anche in questo caso, per un'illustrazione semplice e facilmente comprensibile, è utile scomporre la metodologia di **Hollander** in sette passaggi logici fondamentali.

1 – Per l'applicazione del test di Hollander, è indispensabile il rispetto di una **prima condizione preliminare**. In ognuno dei due campioni **C1** e **C2**, **il numero di osservazioni non può essere di poche unità, ma deve essere abbastanza grande, superiore almeno ad una decina**: nella fase finale entro lo stesso campione si devono formare due campioni dipendenti di dimensioni **n**, per cui il numero di osservazioni **N** si dimezza.

2 – La **seconda condizione preliminare**, insita nei concetti appena espressi sull'appaiamento dei dati, riguarda le dimensioni dei due campioni **C1** e **C2**, dei quali si vogliono confrontare i relativi coefficienti angolari: **entrambi devono avere un numero N d'osservazioni pari (N = 2n)**. Nel caso in cui il numero **N** d'osservazioni sia dispari, si elimina un'osservazione a caso.

3 – La **terza condizione preliminare** è che **i due campioni devono avere lo stesso numero d'osservazioni**; di conseguenza, se si deve eliminare un'osservazione l'operazione deve essere eseguita su entrambi i campioni.

4 – Nel campione **C1**, si stimano **n** coefficienti angolari **b<sub>1,k</sub>**, accoppiando i valori di **Y<sub>1,k</sub>** e **X<sub>1,k</sub>** rispettivamente con quelli di **Y<sub>1,k+n</sub>** e **X<sub>1,k+n</sub>**

$$b_{1,k} = \frac{Y_{1,k+n} - Y_{1,k}}{X_{1,k+n} - X_{1,k}}, \quad k = 1, 2, \dots, n.$$

5 – Nel campione C2, si stimano  $n$  coefficienti angolari  $b_{2,k}$ , accoppiando i valori di  $Y_{2,k}$  e  $X_{2,k}$  rispettivamente con quelli di  $Y_{2,k+n}$  e  $X_{2,k+n}$

$$b_{2,k} = \frac{Y_{2,k+n} - Y_{2,k}}{X_{2,k+n} - X_{2,k}}, \quad k = 1, 2, \dots, n.$$

6 – Si formano coppie a caso tra un valore  $b_{1,k}$  e di un valore  $b_{2,k}$ , calcolando  $n$  differenze  $d_k$

$$d_k = b_{1,i} - b_{2,j}$$

con il loro segno.

7 – Su queste  $n$  differenze  $d_k$  con segno si applica il test di Wilcoxon dei ranghi con segno.

Infatti dopo aver trasformato i valori ottenuti nei loro ranghi, mantenendo il segno, se è vera l'ipotesi nulla

$$H_0: b_1 = b_2$$

la somma dei ranghi positivi e di quelli negativi tenderà ad essere uguale e quindi quella minore tenderà al valore medio,

mentre se è vera l'ipotesi alternativa bilaterale

$$H_1: b_1 \neq b_2$$

oppure una delle 2 ipotesi unilaterali prefissate

$$H_1: b_1 < b_2 \quad \text{oppure} \quad H_1: b_1 > b_2$$

la somma minore tenderà a 0.

Per una spiegazione dettagliata delle diverse metodologie nel caso di campioni piccoli o grandi, si rimanda all'illustrazione del **test di Wilcoxon per 2 campioni dipendenti**.

Come esempio d'applicazione di questa metodologia, in vari testi di statistica non parametrica è riportato un esempio pubblicato da A. C. Wardlaw e G. van Belle nel 1964 (con l'articolo *Statistical aspect of the mouse diaphragm test for insulin*, sulla rivista *Diabetes* n. 13, pp. 622-633).

**Il caso è interessante anche per la ricerca ambientale, poiché illustra il confronto tra due rette calcolate solo su due punti, per i quali si abbiano misure ripetute della Y.**

E' una situazione sperimentale che ricorre con frequenza nelle ricerche di ecotossicologia o comunque nella valutazione degli effetti di un dosaggio con due soli punti, per i quali non è possibile applicare la regressione parametrica.

ESEMPIO. E' stata valutata la capacità di un ormone, somministrato a 2 dosi differenti ( $X_1 = 0,3$  e  $X_2 = 1,5$ ), di stimolare la sintesi di glicogeno, misurata in termini di densità ottica ( $Y_i$ ), in presenza di insulina di 2 tipi differenti (C1 = insulina standard; C2 = campione di insulina da saggiare). Si vuole verificare se, nelle due differenti condizioni sperimentali, le due rette dose-risposta sono parallele; in termini biologici, se si ha una risposta simile all'aumentare della dose, pure considerando gli effetti di una diversa concentrazione del principio attivo.

A questo scopo, sono state ottenute 12 misure nella condizione C1 e 12 nella condizione C2, in ognuna delle quali 6 alla dose 0,3 e 6 alla dose 1,5 (riportati nella tabella seguente). Com'è prassi in questi casi, per rendere lineare la risposta, in sostituzione del valore della dose è stato utilizzato il suo logaritmo (log dose).

Prova	C1		C2	
	$X_{1,j}$	$Y_{1,j}$	$X_{2,j}$	$Y_{2,j}$
1	log 0,3	230	log 0,3	310
2	log 0,3	290	log 0,3	265
3	log 0,3	265	log 0,3	300
4	log 0,3	225	log 0,3	295
5	log 0,3	285	log 0,3	255
6	log 0,3	280	log 0,3	280
7	log 1,5	365	log 1,5	415
8	log 1,5	325	log 1,5	375
9	log 1,5	360	log 1,5	375
10	log 1,5	300	log 1,5	275
11	log 1,5	360	log 1,5	380
12	log 1,5	385	log 1,5	380

Il metodo di Hollander di verifica del parallelismo tra le due rette può essere presentato in modo schematico nei suoi passaggi logici.

1 – Dopo aver verificato che nella programmazione dell’esperimento sono state rispettate le tre condizioni preliminari:

- numero di osservazioni per campione superiore a 10 ( $N = 12$ ),
- numero di osservazioni entro ogni campione pari e uguale per le due dosi (6 per dose),
- numero di osservazioni uguali per i due campioni (12 in ognuno dei due).

2 – Si calcolano misure ripetute del coefficiente angolare  $b_{i,k}$  in entrambi i campioni

$$b_{i,k} = \frac{Y_{i,k+n} - Y_{i,k}}{X_{i,k+n} - X_{i,k}}$$

abbinando

- la prima osservazione del dosaggio 0,3 con la prima osservazione del dosaggio 1,5
  - la seconda osservazione del dosaggio 0,3 con la seconda osservazione del dosaggio 1,5
  - e così fino a
  - la sesta osservazione del dosaggio 0,3 con la sesta osservazione del dosaggio 1,5
- ottenendo i seguenti risultati:

- per il campione C1

$$b_{1,1} = \frac{Y_{1,7} - Y_{1,1}}{X_{1,7} - X_{1,1}} = \frac{365 - 230}{\log 1,5 - \log 0,3} = \frac{135}{0,699} = 193,1$$

$$b_{1,2} = \frac{Y_{1,8} - Y_{1,2}}{X_{1,8} - X_{1,2}} = \frac{325 - 290}{\log 1,5 - \log 0,3} = \frac{35}{0,699} = 50,1$$

$$b_{1,3} = \frac{Y_{1,9} - Y_{1,3}}{X_{1,9} - X_{1,3}} = \frac{360 - 265}{\log 1,5 - \log 0,3} = \frac{95}{0,699} = 135,9$$

$$b_{1,4} = \frac{Y_{1,10} - Y_{1,4}}{X_{1,10} - X_{1,4}} = \frac{300 - 225}{\log 1,5 - \log 0,3} = \frac{75}{0,699} = 107,3$$

$$b_{1,5} = \frac{Y_{1,11} - Y_{1,5}}{X_{1,11} - X_{1,5}} = \frac{360 - 285}{\log 1,5 - \log 0,3} = \frac{75}{0,699} = 107,3$$

$$b_{1,6} = \frac{Y_{1,12} - Y_{1,6}}{X_{1,12} - X_{1,6}} = \frac{385 - 280}{\log 1,5 - \log 0,3} = \frac{105}{0,699} = 150,2$$

- per il campione C2

$$b_{2,1} = \frac{Y_{2,7} - Y_{2,1}}{X_{2,7} - X_{2,1}} = \frac{415 - 310}{\log 1,5 - \log 0,3} = \frac{105}{0,699} = 150,2$$

$$b_{2,2} = \frac{Y_{2,8} - Y_{2,2}}{X_{2,8} - X_{2,2}} = \frac{375 - 265}{\log 1,5 - \log 0,3} = \frac{110}{0,699} = 157,4$$

$$b_{2,3} = \frac{Y_{2,9} - Y_{2,3}}{X_{2,9} - X_{2,3}} = \frac{375 - 300}{\log 1,5 - \log 0,3} = \frac{75}{0,699} = 107,3$$

$$b_{2,4} = \frac{Y_{2,10} - Y_{2,4}}{X_{2,10} - X_{2,4}} = \frac{275 - 295}{\log 1,5 - \log 0,3} = -\frac{20}{0,699} = -28,6$$

$$b_{2,5} = \frac{Y_{2,11} - Y_{2,5}}{X_{2,11} - X_{2,5}} = \frac{380 - 255}{\log 1,5 - \log 0,3} = \frac{125}{0,699} = 178,8$$

$$b_{2,6} = \frac{Y_{2,12} - Y_{2,6}}{X_{2,12} - X_{2,6}} = \frac{380 - 280}{\log 1,5 - \log 0,3} = \frac{100}{0,699} = 143,1$$

3 – Si accoppiano casualmente i 6 valori  $b_{i,j}$  del campione C1 con i 6 valori del campione C2 e si calcolano le differenze; nell'esempio riportato da Wardlaw e van Belle si ottennero gli accoppiamenti e le differenze

$$d_k = b_{1,i} - b_{2,j}$$

di seguito riportate con tutti i passaggi

$$d_1 = b_{1,1} - b_{2,1} = 193,1 - 150,2 = \mathbf{42,9}$$

$$d_2 = b_{1,2} - b_{2,2} = 50,1 - 157,4 = -\mathbf{107,3}$$

$$d_3 = b_{1,3} - b_{2,4} = 135,9 - -28,6 = \mathbf{164,5}$$

$$d_4 = b_{1,4} - b_{2,5} = 107,3 - 178,8 = -\mathbf{71,5}$$

$$d_5 = b_{1,5} - b_{2,6} = 107,3 - 143,1 = -\mathbf{35,8}$$

$$d_6 = b_{1,6} - b_{2,3} = 150,2 - 107,3 = \mathbf{42,9}$$

4 – A queste 6 differenze si applica il test T di Wilcoxon per 2 campioni dipendenti: dopo averle ordinate in modo crescente, considerando il loro valore assoluto

$$- 35,8 \quad 42,9 \quad 42,9 \quad -71,5 \quad -107,3 \quad 164,5$$

si attribuisce il rango, considerando che due valori sono uguali

$$1 \quad 2,5 \quad 2,5 \quad 4 \quad 5 \quad 6$$

e si riporta il segno dei valori originari

$$-1 \quad 2,5 \quad 2,5 \quad -4 \quad -5 \quad 6$$

calcolando la somma dei ranghi positivi, che risulta uguale a 11  
e la somma dei negativi, che risulta uguale a 10.  
Il valore di T è la somma minore e quindi è uguale a 10.

5 – Per la significatività, dopo aver stimato T, occorre stabilire la direzione del confronto: se bilaterale oppure unilaterale.

Questo test sul parallelismo, come presentato dal problema, è bilaterale. Poiché sulla tabella dei valori critici per il test **T** di Wilcoxon per 2 campioni dipendenti, con **N** uguale a 6 e test a due code, il valore riportato è 2, non si può rifiutare l'ipotesi nulla. Per una conclusione più articolata, è conveniente mettere in evidenza che la somma dei negativi (10) e la somma dei positivi (11) sono tra loro molto vicine, come atteso quando l'ipotesi nulla è vera; pertanto, in termini di parallelismo tra le due rette, si deve concludere che **non solo non si può rifiutare l'ipotesi nulla ma la probabilità che le due rette siano parallele è molto elevata.**

Della metodologia presentata, un punto debole, criticato da vari autori, è l'accoppiamento casuale delle stime delle  $b_{i,k}$  nei due campioni, per calcolare le  $n$  differenze  $d_k$ . Appunto perché generato in modo casuale, può fornire risposte diverse a partire dagli stessi dati. Come in tutti questi casi, ovviamente **non è assolutamente accettabile la ripetizione fino ad ottenere la risposta desiderata** o più vicina a quanto atteso.

## CAPITOLO XVI

### IL DISEGNO SPERIMENTALE NELLA RICERCA AMBIENTALE METODI DI RICAMPIONAMENTO PER L'INFERENZA

#### 16.1. I MOTIVI DEL DISEGNO SPERIMENTALE

Per comprendere più compiutamente i motivi e le modalità del **disegno sperimentale**, (in inglese *experimental design*, tradotto in modo più significativo con **programmazione dell'esperimento**), è utile riprendere alcuni concetti che rappresentano i punti fondamentali di un corso di statistica.

Il motivo principale del ricorso all'**analisi statistica** deriva dalla **variabilità**; è essa che per la stima dei parametri di una distribuzione e per l'inferenza **impone l'uso di misure ripetute**. Se non esistesse alcuna differenza tra le singole osservazioni, raccolte in natura nelle medesime condizioni oppure sottoposte in laboratorio al medesimo trattamento, basterebbe un solo dato per ottenere indicazioni precise sull'esperimento; il confronto tra due sole misure, raccolte in condizioni diverse, permetterebbe di valutare le differenze reali determinate dai vari fattori.

**L'esistenza della variabilità impone l'estensione dell'analisi al maggior numero possibile di oggetti**, poiché l'errore nella stima dei parametri è inversamente proporzionale al numero di repliche raccolte; se si volesse ottenere la misura esatta della media, senza errore statistico, si dovrebbe giungere all'analisi di tutta la popolazione. Ma in molte discipline è un comportamento raramente possibile, quasi sempre assolutamente non conveniente nel bilancio costi-benefici, che imporrebbe gravi **problemi sia tecnici che economici, per il tempo necessario e le risorse richieste dalla ricerca**.

Per essere condotta in modo corretto, una ricerca quantitativa deve raggiungere **i due obiettivi essenziali dell'analisi statistica**:

- **le descrizioni e le inferenze tratte dallo studio di un numero limitato di casi devono essere precise ed estensibili a tutta la popolazione;**
- **i risultati devono essere pertinenti al problema che si intende affrontare.**

A questo scopo, sono indispensabili **un buon disegno sperimentale ed un campionamento corretto**, che devono **rispondere a tre quesiti essenziali**:

- **come scegliere gli individui per l'esperimento,**
- **quanti dati raccogliere,**
- **come assegnare le osservazioni o repliche ai vari fattori da analizzare.**

Nelle varie discipline ed in ogni indagine, la statistica applicata risponde in modo diverso a causa

- della differente variabilità del materiale utilizzato,
- della specificità delle domande,

- della precisione con la quale si desiderano i risultati,
- del costo di ogni singolo dato,
- del tempo richiesto dalla loro raccolta.

Tuttavia esistono alcuni criteri fondamentali, che è conveniente avere presente nella programmazione e nella conduzione dell'esperimento.

Nella ricerca ambientale, è utile ricordare i 10 principi generali che andrebbero seguiti in un'indagine statistica, come proposti da Roger Green nel suo testo del 1979 (*Sampling Design and Statistical Methods for Environmental Biologist*, John Wiley & Sons, New York).

1 - **Formulare in modo conciso e corretto la domanda** alla quale si vuole rispondere. La relazione che si deve presentare sulla ricerca condotta è sempre rivolta ad altri; pertanto, i risultati devono essere comprensibili e coerenti con la domanda.

Si supponga di voler effettuare una ricerca sull'inquinamento in un tratto di fiume, al fine di valutare l'apporto specifico di un affluente. Come primo passo, l'obiettivo può essere espresso in termini di senso comune: "L'affluente causa un danno biologico?". Tale domanda deve essere formulata in modo preciso e quindi occorre studiare anticipatamente i vari aspetti del problema: "L'abbondanza della specie Y nel fiume è più ridotta dopo l'affluente rispetto alla zona precedente?". E' infatti **dal tipo di domanda che deriva il tipo di test** (unilaterale o bilaterale, ricordando che quello unilaterale è preferibile, in quanto più potente) ed è **dal tipo di problema che derivano i dati da raccogliere** (quali specie analizzare, quali indicatori di danno biologico rilevare, quali comunità studiare, ...). Inoltre è importante **valutare l'informazione contenuta nel tipo di scala**, utilizzato per misurare ogni variabile, poiché essa influenza in modo rilevante la scelta del test.

2 - **Raccogliere repliche dei campioni entro ogni combinazione di tempo, luogo ed ogni altro fattore controllato**, poiché la significatività delle differenze tra i vari livelli dello stesso fattore dipende dalla "varianza entro". **Le repliche devono essere mantenute distinte per tutti i fattori**, poiché un loro raggruppamento comporta perdita di informazioni e rende difficile, o addirittura impossibile, il ritorno successivo ad una loro distinzione. Può essere produttivo raggruppare i dati; ma tale operazione deve avvenire solo dopo analisi statistiche che dimostrino l'ininfluenza dei fattori specifici rilevati, oppure quando siano già state attuate le analisi specifiche programmate e si volesse estendere il confronto ad un'area più ampia o a un livello superiore, allo studio di alcune interazioni, evitando la raccolta di altri dati e un equivalente aumento dei costi o dei tempi.

3 - **E' conveniente che il numero di repliche, con scelta casuale** delle osservazioni entro ogni combinazione delle variabili controllate, **sia uguale in ogni sottogruppo**. E' un accorgimento che permette di ridurre al minimo gli errori standard, a parità del numero di dati raccolti. Inoltre, scegliere solo **campioni o situazioni ritenuti "rappresentativi" o "tipici" non permette un campionamento casuale** e comporta gravi distorsioni nelle conclusioni. L'assunzione di normalità della distribuzione e di indipendenza degli errori può essere violata non solo dalle caratteristiche dei dati, ma anche da un **campionamento falsamente casuale**; mentre nel primo caso l'allontanamento dalla normalità può essere sanato con una trasformazione, nel secondo caso, **con un campione non rappresentativo della popolazione, si determina una situazione che non può più essere corretta**, se non con una nuova raccolta di dati.

Nel caso di un fiume con parti rocciose ed altre sabbiose, in aree differenti possono essere presenti specie diverse o la loro densità variare moltissimo. **Differenze rilevanti tra zone limitrofe impongono un campionamento in cui siano presenti tutte le situazioni, i vari strati**; in queste condizioni, è utile passare da campionamenti completamente casuali a campionamenti stratificati.

Per alcune analisi, come nella varianza ad un solo criterio di classificazione e nella regressione lineare, si possono utilizzare campioni con un numero differente di osservazioni; ma per altre, in particolare nell'analisi fattoriale per lo studio delle interazioni, si richiedono campioni con lo stesso numero di dati, poiché trattamenti con **campioni di dimensioni diverse determinano interazioni ambigue, difficilmente interpretabili, e "varianze entro" che sono meno omogenee**. Nello studio ambientale, è quindi opportuno avere almeno due osservazioni per ogni combinazione dei fattori considerati, mentre la scelta dei siti di campionamento può essere attuata in modo corretto con il ricorso a coordinate estratte da tabelle di numeri casuali.

4 - Per verificare se una condizione particolare determina effetti differenti, occorre **raccogliere campioni sia in casi in cui la condizione analizzata è presente sia in altri in cui essa è assente**, a parità di tutti gli altri fattori (coeteris paribus). E' possibile valutare **l'effetto di un trattamento solo mediante il confronto con un controllo**, soprattutto nelle ricerche di tossicologia o di fattori che operano in condizioni non naturali, in cui la specie analizzata può non riprodursi o morire. Per verificare le conseguenze dell'affluente, che può contenere pesticidi, occorre raccogliere campioni anche in un'area a monte dell'affluente, dove non dovrebbero essere presenti o almeno avere una concentrazione nettamente inferiore; per tutti gli altri fattori non espressamente considerati, le condizioni dei due campioni dovrebbero essere analoghe.

5 - **Effettuare campionamenti preliminari**, che forniscano le informazioni di base sulle caratteristiche dei dati, per la scelta del disegno sperimentale e dei test statistici da utilizzare. In studi di campagna o in ricerche che non siano già ampiamente descritte in letteratura, **l'importanza del**

**campionamento preliminare è sovente sottostimata.** Spesso le ricerche sono condotte in tempi ristretti ed è psicologicamente difficile spenderne una parte nel campionamento, in operazioni che potrebbero non fornire dati utili per il rapporto finale. Secondo Green, la situazione è simile a quella dello scultore che inizia un'opera senza avere di fronte un modello di riferimento: è alto il rischio di errori non facilmente riparabili ed il tempo impiegato diventa in complesso maggiore. Il tempo speso nell'analisi preliminare è ampiamente recuperato successivamente; inoltre, in molti casi, anche il risultato dello studio preliminare può essere incluso nel rapporto finale, in quanto utile alla interpretazione dei risultati conclusivi.

**Il motivo fondamentale per ricorrere ad un campionamento preliminare è che non esistono altri modi per evidenziare i gravi problemi che possono insorgere in una ricerca,** in particolare se a carattere ambientale, **dove i fattori non prevedibili a priori sono numerosi.** L'efficienza dello schema di campionamento, le dimensioni del campione ed il numero di repliche per ottenere la precisione desiderata nelle stime, la possibile presenza di modelli diversi di distribuzione spaziale possono rendere necessario un campionamento stratificato a più livelli, che deve essere definito a priori anche nei particolari.

6 - **Verificare che il metodo di campionamento adottato sia appropriato per tutte le condizioni incontrate:** variazioni nell'efficienza del campionamento da un'area all'altra (ad esempio, determinate dalle diverse condizioni del letto o delle sponde del fiume) pregiudicano il confronto tra aree. Negli studi ambientali, nessuna area è immune da questo problema. Per campionare popolazioni animali, si hanno problemi sul tipo di trappole o sul mezzo di raccolta, sul modo e sull'ora della utilizzazione, sulla taglia degli individui e sulla loro densità, sul periodo di rilevazione che li vede in fasi diverse dello sviluppo o in un momento particolare di migrazione: non solo possono essere sottostimate intere comunità, ma sovente si hanno campionamenti non corretti e non confrontabili per interi gruppi tassonomici.

Purtroppo, **nella ricerca ambientale non esistono metodi validi per tutte le condizioni. Il campionamento deve quindi essere preparato in modo specifico,** finalizzato allo studio che si intende condurre.

7 - Per ogni situazione di campionamento, esistono comportamenti che devono essere stabiliti prima dell'inizio della ricerca. Se l'area da campionare presenta situazioni nettamente diversificate, è utile **suddividere l'area in sottozone relativamente omogenee ed assegnare ad esse campioni proporzionali alle loro dimensioni.** Se è richiesta una stima dell'abbondanza totale delle specie, è vantaggioso suddividere i campioni in modo proporzionale al numero di organismi presenti in ogni sottozona.

Quando il fondo di un lago è formata in prevalenza da zone rocciose e un'altra da sabbiose, oppure quando un'area è in prevalenza coltivata a prato e un'altra a bosco, un campionamento casuale entro ogni area potrebbe rappresentare un disegno sperimentale molto inefficiente, in quanto indurrebbe ad attribuire le differenze riscontrate al confronto tra esse e non all'influenza della zona prevalente. La presenza di determinate specie e la loro numerosità dipende molto più dalla zona geografica di campionamento (un fiume di montagna o presso lo sbocco al mare, una zona boscosa oltre i mille metri o vicino al litorale) che non dalle differenze tra aree confinanti (prato di montagna e bosco limitrofo). Se dalla letteratura o da un campionamento sperimentale emerge tale contrapposizione, il metodo di campionamento più appropriato è una suddivisione per zone entro ogni area mediante un disegno gerarchico. Con esso è possibile fare emergere la variabilità entro ogni area ed individuare il contributo fornito dalle varie zone.

8 - Verificare che **le dimensioni dell'unità di campionamento siano appropriate al numero di individui, alla densità e alla distribuzione spaziale dei vari organismi che si vogliono analizzare**. Il numero di repliche è una funzione della precisione desiderata nelle stime. A tale scopo è importante definire sia **l'unità di campionamento** che **l'elemento del campionamento**.

In rapporto alla distribuzione della specie che si intende studiare, **l'unità di campionamento** è la superficie o il volume (es. la pianta, i centimetri di superficie o i litri di acqua) in cui vivono gli animali che devono essere campionati. In analisi su scala molto ampia, in cui si confrontano le comunità di vari fiumi, potranno essere unità di campionamento il lago, il fiume o il bosco, ed in essi si effettuerà una serie intera di rilevazioni. Se l'oggetto di studio è un fiume, l'unità di campionamento può essere una zona indicata da parametri morfologici o idrometrici.

**L'elemento del campionamento** è il singolo animale raccolto entro l'unità di campionamento. Occorre non confondere l'unità di campionamento con l'elemento di campionamento, poiché **il processo di randomizzazione e le dimensioni del campione, o il numero di repliche, vanno riferite all'unità di campionamento non all'elemento del campionamento**.

Scelta l'unità di campionamento, la precisione con cui i parametri ecologici sono stimati dipende dal numero di unità di campionamento, non dal numero di elementi o individui contati, a parità di altre condizioni.

9 - Se l'analisi dei dati mostra che la distribuzione degli errori non è omogenea, non è normalmente distribuita o che dipende dalla media, è indispensabile **ricorrere alla loro trasformazione o all'uso di test non parametrici; per il tipo di campionamento e la verifica dell'ipotesi nulla, è utile ricorrere ad analisi sequenziali o a dati simulati**.

Il dibattito sul rispetto delle condizioni di validità dei test parametrici con dati ambientali è già stato presentato varie volte, senza una conclusione definitiva: **per un gruppo di dati reali, quasi**

**certamente le assunzioni di omogeneità e normalità non sono rigorosamente valide; ma quasi sempre sono approssimativamente valide e spesso i test per distribuzioni univariate sono estremamente robusti.** Si allontanano con probabilità maggiori dalla normalità i campioni con pochi dati, formati da gruppi di dimensioni diverse e per ipotesi unilaterali.

**I metodi non parametrici sono più frequentemente utili quando sono stati previsti nel piano sperimentale e i dati sono stati raccolti con questa finalità, piuttosto che come operazione di salvataggio di dati non attesi e non trattabili in altro modo.**

L'**analisi sequenziale** è utile quando si deve ridurre al minimo il numero di dati necessari per giungere ad una decisione, a causa del costo o del tempo richiesto per la rarità degli eventi. Nonostante il suo interesse in vari settori della ricerca ambientale, verranno di seguito ricordati solo i concetti fondamentali, rimandando a testi specifici per una conoscenza operativa, che richiede approfondimenti.

10 - **Se sono stati scelti il campionamento e i test statistici più adatti** per verificare le ipotesi formulate, **occorre accettarne i risultati**. Un risultato inatteso o non desiderato non è un motivo valido, per rifiutare il metodo seguito e ricercarne uno “migliore”. Ogni indagine statistica porta ad una conclusione, che può contenere notizie “buone o cattive”, che può essere in accordo o in disaccordo con quanto atteso; ma, se non emerge che sono stati commessi errori gravi, **tentare di cambiare la conclusione, con ripetizioni dell’esperimento** fino ad ottenere la conclusione desiderata, **non modifica la realtà e rappresenta un’alterazione della probabilità calcolata.**

## **16.2. CONCETTI SULL’ANALISI SEQUENZIALE**

Quando le osservazioni sperimentali sono fatte in successione temporale, per cui **il risultato di una prova è noto prima che sia effettuata quella successiva**, si può ricorrere **all’analisi sequenziale**. Il vantaggio fondamentale del test è **la possibilità di ridurre al minimo il numero di osservazioni necessarie per prendere la decisione** dell’esistenza o meno della significatività, per una determinata differenza. **Il test offre vantaggi, quando le osservazioni hanno costi (economici o morali) molto elevati oppure si devono analizzare gli effetti di fenomeni che succedono raramente**, con lunghi intervalli di tempo tra una manifestazione e l’altra. Per esempio, quando con misure mensili si devono analizzare molti parametri di una zona o di un lago sottoposti a risanamento, per una decisione più rapida può essere conveniente utilizzare l’analisi discriminante: si possono eliminare i parametri che forniscono risposte evidenti, per concentrare l’attenzione e le risorse su quelli più problematici.

**Nella ricerca ambientale ed in quella medica, considerazioni etiche suggeriscono di utilizzare il numero minimo di persone o di animali**, quando è logico attendersi che essi avrebbero vantaggi evidenti dalla somministrazione dell’altro tossico o dell’altro farmaco.

**Se i dati sono già stati raccolti tutti, i test sequenziali sono meno potenti dei corrispondenti metodi parametrici già presentati e pertanto la loro utilizzazione non è conveniente.**

I test sequenziali sono fondati su una regola d'arresto ed utilizzano grafici cartesiani. **A forma bidimensionale, in ascissa riportano il numero di osservazioni del campione ed in ordinata una particolare funzione  $f(x)$  dei valori osservati**, che dipende dal tipo di test. Nei casi più semplici, il grafico riporta **due linee, che dipendono sia dalle probabilità  $a$  e  $b$ , sia dall'entità della differenza dei valori del parametro espresso nell'ipotesi nulla  $H_0$  ed in quella alternativa  $H_1$** . Queste due linee individuano il confine fra **tre zone**:

- 1 - **rifiuto dell'ipotesi nulla** e quindi accettazione dell'ipotesi alternativa;
- 2 - **impossibilità di decidere** tra rifiuto ed accettazione dell'ipotesi nulla;
- 3 - non rifiuto od **accettazione dell'ipotesi nulla**.

Fino a quando il diagramma si mantiene nella zona dell'impossibilità di decidere, l'esperimento deve continuare, con la raccolta di almeno una osservazione ulteriore; l'esperimento termina, non appena è possibile decidere tra ipotesi nulla ed ipotesi alternativa.

Si supponga di disporre di due campioni dipendenti, in cui la risposta è di tipo qualitativo (+ o -) come nel test dei segni, con la seguente successione di miglioramenti e peggioramenti per 11 osservazioni.

+ + - + - + + + - + +

Ad ogni stadio dell'esperimento, dopo ogni risposta, il test più appropriato sarebbe la distribuzione binomiale, con **parametri  $n$**  (numero di dati raccolti fino a quel momento) e  **$p$**  (probabilità di un successo, nel caso fosse vera l'ipotesi nulla, uguale a 0,5). Procedendo in modo sequenziale, si potrebbe pensare di fermare l'esperimento quando la probabilità calcolata risultasse significativa alla probabilità prefissata: è un approccio che appare logico, se si considera che **vengono utilizzate le informazioni raccolte fino a quel momento**.

**In realtà, le occasioni per fermare l'esperimento potrebbero essere più di una. Anche se fosse vera l'ipotesi nulla, la probabilità di trovare per caso un risultato significativo cresce all'aumentare del numero di osservazioni; di conseguenza, con la stessa logica dei confronti multipli, si chiede una probabilità più bassa, per valutare ripetutamente i risultati dopo l'aggiunta di ogni osservazione.**

I problemi concettuali e di metodo da risolvere possono essere sintetizzati in una domanda: "se l'esperimento dimostra l'esistenza o la non-esistenza di una differenza significativa alla probabilità  $a$  dopo  $n$  osservazioni, quale deve essere il valore di  $a$  dopo ogni prova?"

Casi semplici di applicazione dell'analisi sequenziale sono il test per dati binari e il test unilaterale sulla media  $m$ .

Come esempio di presentazione della procedura d'arresto, è utile un confronto sulla media. Si supponga che gli altri parametri siano noti e che la media possa assumere solo due valori  $\mu_0$  e  $\mu_1$  (con  $\mu_1 > \mu_0$ ). Per decidere quale delle due ipotesi alternative possa essere considerata vera, si introduce il **rapporto di verosimiglianza**  $\lambda$

$$\lambda = \frac{\text{Pr obabilità..di..un..certo..campione..quando..} \mathbf{m} \mathbf{.} = \mathbf{m}_0}{\text{Pr obabilità..dello..stesso..campione..quando..} \mathbf{m} \mathbf{.} = \mathbf{m}_1}$$

Si avrà una probabilità più alta che l'ipotesi  $\mathbf{H}_0: \mathbf{m} = \mathbf{m}_0$  sia vera, quando  $\lambda$  è grande; la probabilità sarà maggiore per l'ipotesi  $\mathbf{H}_1: \mathbf{m} = \mathbf{m}_1$  quando  $\lambda$  è piccolo. In pratica, per ogni numero  $n$  di osservazioni, si individuano i limiti  $l_0$  e  $l_1$  e si calcola il valore di  $\lambda$  dopo ogni osservazione. In base a questo ultimo valore si decide di:

- 1 - accettare  $\mathbf{H}_0$  se  $\lambda \geq l_0$ ;
- 2 - accettare  $\mathbf{H}_1$  se  $\lambda \leq l_1$ ;
- 3 - raccogliere un'altra osservazione e rifare il calcolo, se  $l_1 < \lambda < l_0$ .

Il valore di  $l_0$  può essere calcolato dall'insieme dei valori campionari che portano all'accettazione di  $\mathbf{H}_0$ . Tale probabilità è  $1 - a$  se  $\mathbf{m} = \mathbf{m}_0$  e  $b$  se  $\mathbf{m} = \mathbf{m}_1$ ;

di conseguenza,  $l_0$  è uguale a

$$\lambda_0 = \frac{1 - a}{b}$$

e  $l_1$ , il rapporto tra quei valori che portano all'accettazione dell'ipotesi  $\mathbf{H}_1$ , con  $\mathbf{m} = \mathbf{m}_0$  e quella degli stessi valori con  $\mathbf{m} = \mathbf{m}_1$ , è uguale a

$$\lambda_1 = \frac{a}{1 - b}$$

Poiché  $a$  e  $b$  sono fissate a priori, è semplice determinare i valori con cui confrontare ogni singolo valore di  $\lambda$ .

Per motivi di praticità (non ripetere tutti i calcoli ad ogni osservazione) e per meglio comprendere sia l'uso che la logica dei test sequenziali, è conveniente costruire il grafico relativo al test ed utilizzare la somma ( $\mathbf{T}$ ) dei valori osservati.

Prestabiliti  $a$ ,  $b$  e  $s^2$ , dopo aver stimato le quantità  $d$ ,  $a$  e  $b$ , con

$$\delta = \mu_1 - \mu_0; \quad a = \lg\left(\frac{1 - b}{a}\right); \quad b = \lg\left(\frac{1 - a}{b}\right);$$

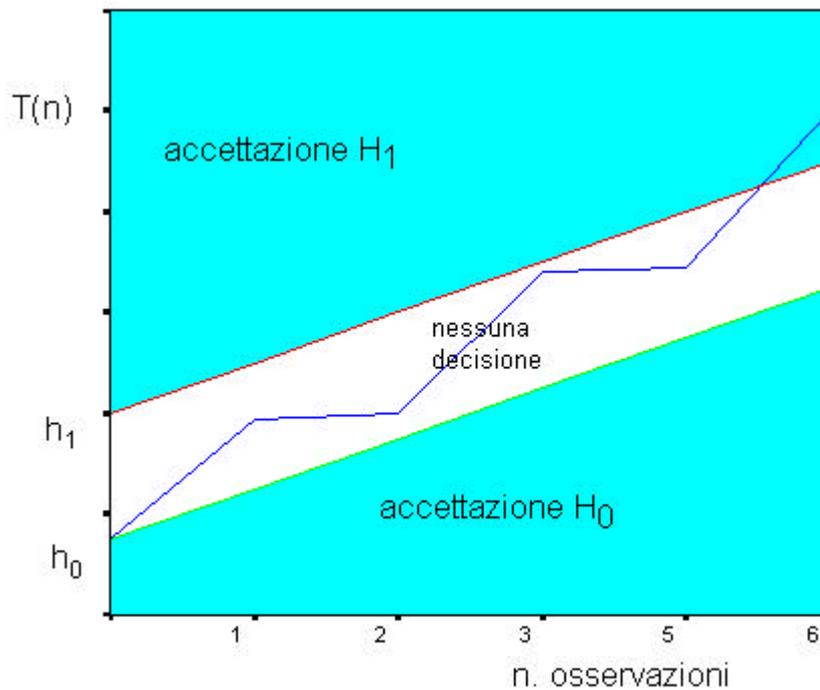
e da esse le quantità  $h_0$ ,  $h_1$  e  $c$ , con

$$h_0 = -b \cdot \sigma^2 / \delta; \quad h_1 = a \cdot \sigma^2 / \delta; \quad c = \mu_0 + 0,5 \delta$$

si calcolano i valori  $\mathbf{T}_0$  e  $\mathbf{T}_1$ ,

$$T_0 = h_0 + nc; \quad T_1 = h_1 + nc$$

che individuano le due rette di confine nel grafico, che in questo caso sarà di tipo lineare.



Le due rette sono parallele e la loro distanza, l'ampiezza del corridoio di nessuna decisione, è proporzionale in modo diretto alla varianza della popolazione e in modo inverso al valore di  $\delta$ , la differenza tra le due medie.

Come nel test t, per evidenziare la significatività di una differenza piccola occorre un numero elevato di dati, mentre con una varianza piccola ne sono sufficienti pochi.

Nel grafico precedente è riportato il totale cumulato dei valori osservati: esce dal corridoio nella parte superiore e quindi permette di rifiutare l'ipotesi nulla, accettando l'ipotesi alternativa

**In test bilaterali sulla media, la probabilità prefissata diventa  $\alpha/2$  ed occorre costruire sia il corridoio ascendente, per scarti mediamente positivi, che quello discendente, per scarti negativi.** Il metodo è analogo a quello precedentemente descritto, con i valori  $h_0$ ,  $h_1$  e  $c$  che possono essere sia positivi che negativi.

Per il corridoio superiore si calcolano le due linee di confine con

$$T_0 = h_0 + nc; \quad T_1 = h_1 + nc$$

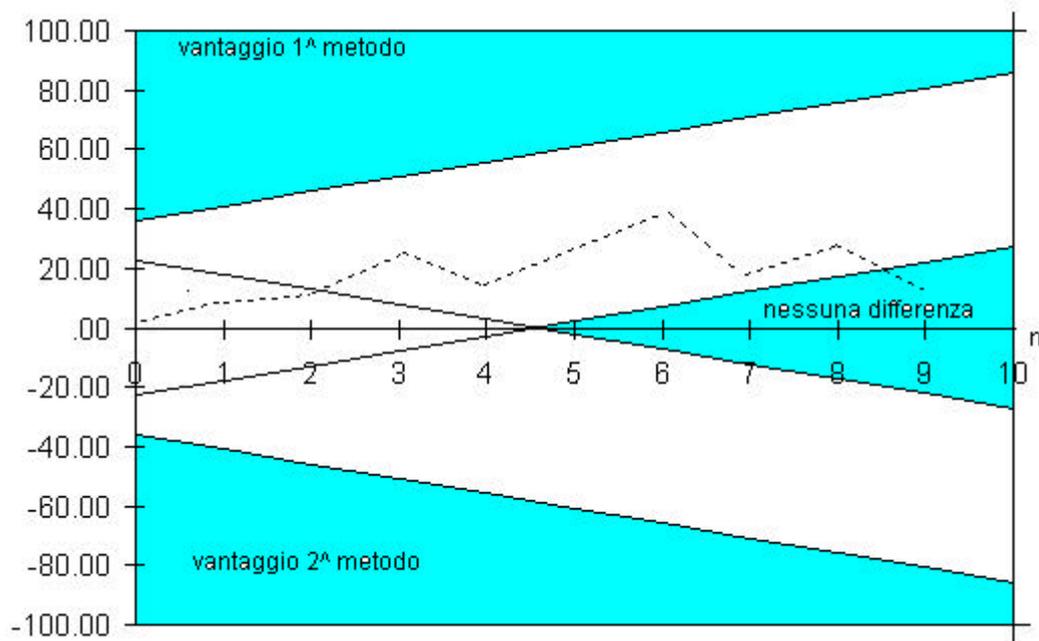
e per il corridoio inferiore con

$$T'_0 = -h_0 - nc; \quad T'_1 = -h_1 - nc$$

Dopo ogni prova si stima la somma ( $T$ ) delle differenze tra i due metodi, come nel caso di 2 campioni dipendenti; il valore viene riportato nel grafico. Se resta in uno dei due corridoi, è necessario raccogliere un altro dato; se ne esce, è possibile decidere tra vantaggio significativo del 1° metodo,

nessuna differenza significativa, vantaggio significativo del 2° metodo, secondo l'area in cui viene a trovarsi il valore di T.

Un esempio può essere fornito dal confronto tra due metodi di riduzione dei livelli d'inquinamento, che sono applicati in situazioni d'emergenza e quindi raramente, per verificare il più rapidamente possibile quale sia il migliore.



Nella figura precedente, il totale cumulato delle differenze esce dalla zona d'incertezza ed entra in quella di nessuna differenza: non esiste una differenza significativa e si deve accettare l'ipotesi nulla.

**Quando la varianza della popolazione è ignota non si ricorre alla distribuzione normale, ma si applica il test sequenziale di Barnard, equivalente al test t di Student.**

### 16.3. CAMPIONE E SUPERPOPOLAZIONE

Tutti i ragionamenti inferenziali hanno come premessa la natura campionaria dei dati. Quando la loro numerosità è molto piccola, la variabile-risposta ha un campo di variazione così ampio da rendere impraticabile o comunque inutile l'uso di molti test statistici, per la scarsa potenza; ma **quando la loro numerosità è molto grande, con elevata probabilità si rifiuta l'ipotesi nulla.**

Per esempio, con migliaia di dati il test chi quadrato evidenzia che nessun modello si adatta ai dati; il test t di Student e il test F portano quasi sempre alla conclusione che tra due o più medie esistono differenze significative, anche quando lo scarto è minimo.

**Per chi non ha sufficiente familiarità con la statistica, questa elevata probabilità di rifiutare l'ipotesi nulla sembra quindi rendere inutile i test, in quanto forniscono quasi sempre risultati significativi.** In questi casi, è conveniente fornire una **misura precisa delle differenze tra i gruppi a confronto e valutare se esse hanno un significato biologico, ecologico, medico** oppure sono oggettivamente trascurabili, pur essendo statisticamente significative.

Il **campione** è ritenuto **grande o piccolo** in considerazione del **numero assoluto di unità** che lo compongono, in totale ed entro ogni incrocio di modalità. **I suoi rapporti relativi alla popolazione totale, se ne rappresenta una quota piccola oppure molto grande, non sono mai presi in considerazione.**

Spesso gli studi statistici hanno per oggetto **non dati campionari ma tutta la popolazione**, come succede in studi demografici di comunità umane od animali, in analisi di ambienti come laghi, fiumi o città, dopo un censimento completo, soprattutto in ambito ristretto. In queste situazioni, cade non solo il concetto di "significatività statistica", già indebolito come nel caso di alta numerosità, ma anche il concetto di distribuzione di probabilità.

**L'inferenza diviene teoricamente superflua, poiché i valori e le differenze riscontrate sui dati dell'universo non hanno bisogno d'inferenza**, essendo quelli reali esistenti nella popolazione: sono significativi per definizione, per quanto essi siano piccoli.

Tuttavia, in particolare quando la popolazione non è molto grande, composta al massimo da poche centinaia di individui, ai fini dell'analisi statistica spesso risulta utile ed opportuno **considerare i dati della popolazione come frutto di un campionamento casuale semplice di una superpopolazione.** Il gruppo può cambiare continuamente nel tempo, come una popolazione d'individui per nuove nascite, decessi e migrazioni. **Il contenuto di ogni gruppo di osservazioni può essere considerato come una stima corretta dei dati di una superpopolazione di numerosità ignota;** in tal modo, tutti gli strumenti inferenziali ritrovano il loro significato corretto. La scelta è giustificata dal concetto di variabilità statistica di un campione rispetto alla popolazione: a causa della dinamica demografica, la popolazione cambia casualmente in tempi ristretti, non diversamente da quanto avviene per un campione rispetto alla sua popolazione.

**L'idea di superpopolazione è un artificio che non appare in contrasto con considerazioni statistiche sostanziali; è adottato con frequenza, tanto che quando si utilizzano i dati di tutta la popolazione il ricorso a test statistici è comunemente accettato, senza necessità di evidenziarne l'anomalia concettuale.**

#### **16.4. STIMATORI E LORO PROPRIETA'**

**Scopo della statistica è anche quello di descrivere le caratteristiche costanti di una popolazione, i suoi parametri, come la media, la varianza, la simmetria, la curtosi e la**

**correlazione, basandosi sulle informazioni che possono essere ottenute da un campione estratto dalla medesima.**

Durante il corso, la scelta è avvenuta in modo intuitivo, senza una scelta ragionata. Ad esempio, in modo aprioristico si è ritenuto che la media del campione sia una stima soddisfacente della media della popolazione. Nella logica statistica, occorre invece **esaminare quali condizioni deve soddisfare una stima per essere buona**, la migliore tra quelle proponibili.

Il problema si presenta solamente quando si dispone dei dati di uno o più campioni estratti dalla popolazione in modo casuale. Se nella formazione del campione è intervenuta qualche distorsione, non è possibile alcuna inferenza sulla popolazione originale.

Per meglio comprendere i concetti, è utile distinguere tra **stima e stimatore**:

**la prima indica il valore calcolato,**

**il secondo la particolare funzione dei valori campionari che permettono di calcolare la stima.**

Per scegliere gli stimatori più adeguati, occorre utilizzare 4 criteri: **correttezza, efficienza, consistenza, sufficienza.**

**Uno stimatore è detto corretto quando il suo valore atteso nel campione, per qualunque sua dimensione, coincide con quello della popolazione.** Quando non gode della proprietà della correttezza è detto **distorto** o **deviato**, mentre la differenza tra il valore corretto e quello distorto è chiamata **distorsione** o **deviazione**. E' ovvio che tra uno stimatore senza deviazioni (*unbiased estimator*) ed uno distorto la preferenza vada al primo, in assenza di altre informazioni. Ma quando il valore corretto ha una varianza grande, spesso è preferibile quello distorto se, appunto per una varianza minore, complessivamente fornisce una stima dell'intervallo più vicina alla realtà.

La **quantità d'informazione** di uno stimatore è determinata dalla sua **varianza**; essa rappresenta il criterio fondamentale per scegliere quello più adeguato. **Uno stimatore è più efficiente di un altro quando la sua varianza è minore; l'efficienza relativa è misurata dal rapporto tra le due varianze.**

I concetti di correttezza ed efficienza spesso sono illustrati con un disegno, in cui la media della popolazione è rappresentata come un bersaglio, formato da un punto centrale e da centri concentrici di raggio crescente, mentre i dati del campione sono i lanci o tiri verso di esso.

Le varie possibilità sono schematizzate in quattro situazioni.

- 1- Se i lanci risultano lontani dal centro e la loro media è diversa dal centro del bersaglio, lo stimatore è **non corretto e non efficiente** ( di norma, lo spostamento sistematico dal centro è dovuto allo strumento, la variabilità dei tiri all'incapacità dell'individuo a ripetere con precisione la stessa operazione).
- 2- Se i lanci risultano distanti dal centro, ma la loro media è collocata sul centro, lo stimatore è **corretto ma non efficiente** (lo strumento è centrato, ma l'individuo non sa usarlo).

- 3- Se i lanci sono tutti molto vicini tra loro, ma lontani dalla media, lo stimatore è **non corretto ma efficiente** (lo strumento è tarato male, ma l'individuo sa usarlo).
- 4- Se i lanci colpiscono tutti il centro del bersaglio o sono ad esso molto vicini, **lo stimatore è corretto ed efficiente** (lo strumento è tarato bene e l'individuo sa usarlo).

Un'altra proprietà è la **consistenza: uno stimatore è consistente quando all'aumentare delle osservazioni il valore del campione è sempre più vicino a quello reale della popolazione**. E' il caso della media e della varianza che, all'aumentare del numero di repliche, si avvicinano sempre più a quelle della popolazione. Se il campione comprende tutti gli individui, ovviamente deve trattarsi di una popolazione finita, i parametri coincidono con quelli della popolazione; in tale situazione è implicito il concetto di **stimatore asintoticamente corretto**.

Uno **stimatore è detto sufficiente quando non è influenzato in modo sistematico dal numero di dati** che formano il campione.

Per ottenere uno stimatore che goda di queste quattro proprietà è necessaria una strategia di campionamento adeguata; di conseguenza, le quattro proprietà sono riferite anche al **disegno sperimentale, che dovrebbe essere corretto, efficiente e consistente**.

#### **16.5. TEST DI NORMALITA', SIMMETRIA E CURTOSI, IL TEST DI LILLIEFORS E IL $\chi^2$ PER L'ADATTAMENTO DI UN CAMPIONE AD UNA DISTRIBUZIONE NORMALE**

Tra le condizioni di validità, i test statistici parametrici richiedono **la normalità e la simmetria della distribuzione dei dati**, insieme con l'omogeneità della varianza per i gruppi a confronto. In altri casi, si pone il problema di testare se un dato campione si allontana significativamente da una distribuzione normale. Dopo la raccolta dei dati, sia la distribuzione di frequenza in forma tabellare che la rappresentazione grafica con istogrammi forniscono l'impressione di quanto i valori del campione siano vicini ad una distribuzione normale. Ma è indispensabile verificare questa interpretazione soggettiva, con metodi quantitativi che conducano tutti alle stesse conclusioni sull'approssimazione della distribuzione osservata a quella normale.

L'ipotesi nulla è che la popolazione dalla quale è stato estratto il campione non è troppo lontana dalla famiglia di distribuzioni che seguono la legge di Gauss, quindi che sia  $N(\mu, \sigma^2)$  con  $\mu$  e  $\sigma$  qualsiasi, contro l'ipotesi alternativa che sia diversa dalla normale con media e varianza stimate.

Dopo aver calcolato la funzione di ripartizione della legge normale ridotta  $N(0, 1)$

- si calcola la cumulata delle probabilità,
- poi la cumulata delle frequenze relative del campione
- e lo scarto massimo tra le due distribuzioni.

**Quantili della statistica di Lilliefors**  
**per verificare la normalità della distribuzione campionaria**

| <b>n</b>      | <b>a</b>         |                  |                  |                  |                  |
|---------------|------------------|------------------|------------------|------------------|------------------|
|               | <b>0,20</b>      | <b>0,15</b>      | <b>0,10</b>      | <b>0,05</b>      | <b>0,01</b>      |
| <b>4</b>      | 0,300            | 0,319            | 0,352            | 0,381            | 0,417            |
| <b>5</b>      | 0,285            | 0,299            | 0,315            | 0,337            | 0,405            |
| <b>6</b>      | 0,265            | 0,277            | 0,294            | 0,319            | 0,364            |
| <b>7</b>      | 0,247            | 0,258            | 0,276            | 0,300            | 0,348            |
| <b>8</b>      | 0,233            | 0,244            | 0,261            | 0,285            | 0,331            |
| <b>9</b>      | 0,223            | 0,233            | 0,249            | 0,271            | 0,311            |
| <b>10</b>     | 0,215            | 0,224            | 0,239            | 0,258            | 0,294            |
| <b>11</b>     | 0,206            | 0,217            | 0,230            | 0,249            | 0,284            |
| <b>12</b>     | 0,199            | 0,212            | 0,223            | 0,242            | 0,275            |
| <b>13</b>     | 0,190            | 0,202            | 0,214            | 0,234            | 0,268            |
| <b>14</b>     | 0,183            | 0,194            | 0,207            | 0,227            | 0,261            |
| <b>15</b>     | 0,177            | 0,187            | 0,201            | 0,220            | 0,257            |
| <b>16</b>     | 0,173            | 0,182            | 0,195            | 0,213            | 0,250            |
| <b>17</b>     | 0,169            | 0,177            | 0,189            | 0,206            | 0,245            |
| <b>18</b>     | 0,166            | 0,173            | 0,184            | 0,200            | 0,239            |
| <b>19</b>     | 0,163            | 0,169            | 0,179            | 0,195            | 0,235            |
| <b>20</b>     | 0,160            | 0,166            | 0,174            | 0,190            | 0,231            |
| <b>25</b>     | 0,142            | 0,147            | 0,158            | 0,173            | 0,200            |
| <b>30</b>     | 0,131            | 0,136            | 0,144            | 0,161            | 0,187            |
| <b>&gt;30</b> | $0,736/\sqrt{n}$ | $0,768/\sqrt{n}$ | $0,805/\sqrt{n}$ | $0,886/\sqrt{n}$ | $1,031/\sqrt{n}$ |

La procedura è analoga a quella di Kolmogorov-Smirnov per un test bilaterale. Ma la distribuzione delle probabilità è differente, poiché la distribuzione normale ridotta è calcolata a partire dalla media e dalla varianza campionarie, che introducono una variabilità ulteriore.

Nella tavola dei quantili di Lilliefors ai vari livelli di probabilità sono riportati i valori critici. Se lo scarto massimo calcolato è superiore a quello della tabella, si rifiuta l'ipotesi nulla: il campione non è stato estratto da una popolazione distribuita secondo la legge di Gauss.

ESEMPIO. Nello studio di una specie rara, di cui sono stati misurati 10 individui, si ha il sospetto di essere in presenza di 2 varietà, che hanno dimensioni diverse.

Nella tabella sottostante sono state riportate le 10 misure, in millimetri, già ordinate per rango

|                  |    |    |    |    |    |    |    |    |    |    |
|------------------|----|----|----|----|----|----|----|----|----|----|
| Individui        | a  | b  | c  | d  | e  | f  | g  | H  | i  | L  |
| Dimensioni $x_i$ | 10 | 11 | 12 | 12 | 13 | 15 | 15 | 16 | 17 | 19 |

E' possibile dimostrare che le dimensioni degli individui che formano il campione non hanno una distribuzione normale?

Sarebbe una indicazione importante a favore dell'ipotesi che è formato da individui estratti da due popolazioni diverse.

Risposta

Dopo aver calcolato la media del campione e la deviazione standard  $s$ , con

- media = 14
- $s = 2,87$

per ogni  $x_i$  si stimano i valori di  $z$  corrispondenti (riportati nella seconda colonna)

$$z_i = \frac{x_i - \bar{x}}{s}$$

e la ripartizione delle probabilità della normale ridotta corrispondente

(vedi: S att. riportati nella terza colonna)

| $X_i$  | $Z_i$ | S att. | S oss. | $D_i$        |
|--------|-------|--------|--------|--------------|
| 10     | -1,39 | 0,083  | 0,100  | 0,083        |
| 11     | -1,05 | 0,147  | 0,200  | 0,047        |
| 12, 12 | -0,70 | 0,242  | 0,400  | 0,042        |
| 13     | -0,35 | 0,363  | 0,500  | -0,037       |
| 15, 15 | 0,35  | 0,637  | 0,700  | <b>0,137</b> |
| 16     | 0,70  | 0,758  | 0,800  | 0,058        |
| 17     | 1,05  | 0,853  | 0,900  | 0,053        |
| 19     | 1,74  | 0,959  | 1,000  | 0,059        |

Dopo aver effettuato la cumulata osservata per ogni  $x_i$  (**S oss.** riportata nella 4 colonna), si calcolano le differenze  **$D_i$**  (quinta colonna), sottraendo ad ogni valore della cumulata delle frequenze attese (**S att.**) quello della cumulata delle frequenze osservate (**S oss.**) riportato sulla riga antecedente al valore  $x_i$ .

Per esempio, la prima  **$D$**  (0,083) è data da  $0,083 - 0,000$ ; la seconda  **$D$**  (0,047) da  $0,147 - 0,100$ ; la quarta  **$D$**  (-0,037) da  $0,363 - 0,400$  e la decima  **$D$**  (0,059) da  $0,959 - 0,900$ .

La differenza massima è  **$D = 0,137$** .

Nella tabella dei valori critici di Lilliefors,

- per  $n = 10$  alla probabilità  $\alpha = 0.05$
- il valore riportato è 0,258 e alla probabilità 0.20 è 0,215.

Il valore  **$D$**  calcolato è inferiore anche a questo ultimo. Non solo non è possibile rifiutare l'ipotesi nulla, ma è possibile affermare che lo scostamento della distribuzione del campione dalla normale è trascurabile, perché molto più piccolo. Facilmente, il campione è formato da individui della stessa varietà, appartenenti alla stessa popolazione

Esistono numerosi test per la normalità. Possono essere ricordati **il test di Shapiro e Wilk** e i test fondati sui coefficienti di simmetria e curtosi.

In letteratura con frequenza si ricorre all'uso del test  $\chi^2$  per la bontà dell'adattamento (*goodness of fit test*).

| Frequenze |           |        |  |
|-----------|-----------|--------|--|
| Classe    | Osservate | Attese | $(\text{Oss}-\text{Att})^2/\text{Att}$ |
| 10-19     | 18        | 40,6   | 12,58                                  |
| 20-29     | 70        | 61,6   | 1,15                                   |
| 30-39     | 136       | 111,4  | 5,43                                   |
| 40-49     | 188       | 161,3  | 4,42                                   |
| 50-59     | 180       | 187,1  | 0,27                                   |
| 60-69     | 152       | 174,0  | 2,78                                   |
| 70-79     | 124       | 129,6  | 0,24                                   |
| 80-89     | 56        | 77,4   | 5,92                                   |
| 90-99     | 54        | 37,1   | 7,70                                   |
| 100-109   | 8         | 14,2   | 2,71                                   |
| 110-119   | 10        | 4,4    |  |
| 120-129   | 2         | 1,0    | 12,09                                  |
| 130-140   | 2         | 0,3    |  |
| TOTALE    | 1000      | 1000,0 | $\chi^2 = 5,29$                        |

I dati devono essere raggruppati in classi, per formare una distribuzione di frequenza, della quale devono essere calcolate la media  $\bar{X}$  e la deviazione standard  $s$ . Mediante esse si stimano le frequenze attese in ogni classe, in accordo con la distribuzione normale equivalente.

Da alcuni autori il metodo è criticato, per i motivi addotti nel test di Lilliefors sulla stima della normale ridotta corrispondente. Per i vincoli imposti al calcolo delle frequenze attese nella distribuzione normale, **i gdl sono n-3**, dove **n** è il numero di gruppi. E' un approccio cautelativo, che riduce la probabilità di rifiutare l'ipotesi nulla e che tenta di superare le obiezioni precedenti.

Nella tabella i gruppi sono 11, a causa dell'aggregazione delle tre classi all'estremo superiore; con 8 gdl, il valore critico riportato nella tabella del  $\chi^2$  è uguale a 20,09 alla probabilità 0.01 e uguale a 26,12 alla probabilità 0.001.

Poiché il valore del  $\chi^2_{(8)}$  calcolato dai dati è 55,29 si rifiuta l'ipotesi nulla: esiste una differenza significativa tra la distribuzione osservata e l'equivalente distribuzione normale.

Nel caso di campioni con un numero ridotto d'osservazioni, è possibile ricorrere al test di Kolmogorov-Smirnov per un campione, poiché le classi sono formate sui valori della distribuzione, che sono almeno di tipo ordinale.

Il chi quadrato ed il metodo di Kolmogorov-Smirnov sono **due test generalisti**: nel caso di confronto con la normale, risultano significativi per qualunque deviazione dalla normalità, determinata dalla simmetria come alla curtosi.

## 16.6. CAMPIONI NON PROBABILISTICI E CAMPIONI PROBABILISTICI

Nel **campionamento probabilistico**, ogni unità dell'universo ha una probabilità prefissata e non nulla di essere inclusa nel campione.

Esistono **metodi non probabilistici** di formazione del campione, detti **campionamenti a scelta ragionata**, che prescindono dai criteri di scelta casuale delle unità campionarie. E' una tecnica adatta a piccoli campioni, in cui le unità sono scelte sulla base di informazioni preliminari.

Si supponga, a causa di ridotte disponibilità economiche, di poter attrezzare una sola stazione di rilevamento dell'inquinamento dell'aria in una città. Facilmente verrà collocata in una zona ritenuta rappresentativa dell'inquinamento medio o di quello massimo della città. Nello stesso modo, se è possibile studiare l'evoluzione dell'inquinamento solo in alcuni laghi di una provincia, la scelta non sarà casuale ma fondata sulla loro rappresentatività della situazione generale oppure sui laghi che per la vicinanza o la loro storia attraggono maggiormente l'attenzione della cittadinanza.

Un campionamento non probabilistico è il **campionamento bilanciato**: il criterio di scelta è che la media delle rilevazioni sia uguale, o comunque molto vicina, a quella della popolazione.

Tra i metodi non probabilistici rientra anche il **campionamento per quote**: la popolazione è divisa in gruppi omogenei, sulla base di caratteristiche discriminanti rilevate con un censimento generale; la scelta della unità campionarie rispecchia le proporzioni esistenti nella popolazione, ma l'elemento caratteristico è che essa non è fissata da metodi generali in quanto demandata al ricercatore. Per un'analisi statistica corretta, il rischio deriva dalla possibilità che il ricercatore preferisca, anche inconsciamente, unità di rilevazione di più facile accesso, di costi più contenuti, di maggiore visibilità. Diventa difficile, se non impossibile, valutare la precisione delle stime: sarebbe necessario avere campionamenti replicati, ma sono poco compatibili con la scelta del campionamento per quote, che offre vantaggi soprattutto dal punto di vista del risparmio.

Per indagini sulla popolazione, sono campionamenti non probabilistici anche quelli definiti **di convenienza**, come i **campioni volontari**, utilizzati soprattutto nelle indagini sociologiche o a carattere medico ed epidemiologico, oltre al **campione a valanga** o a **palla di neve**. Questo ultimo metodo viene applicato nel caso di popolazioni rare e non registrate in modo completo: da un gruppo iniziale noto, si possono ottenere informazioni per risalire agli altri individui appartenenti alla stessa categoria. E' applicato in indagini su stranieri illegali: da un gruppo piccolo conosciuto, con domande si possono ricavare informazioni per giungere agli altri individui non ufficialmente registrati, che spesso vivono in clandestinità.

Nonostante il ricorso a questi modelli, nella ricerca ambientale con frequenza maggiore vengono utilizzati i **campionamenti probabilistici**. Il metodo fondamentale è il **campionamento casuale semplice senza ripetizione** (*simple random sampling* o *random sampling without replacement*), in cui ogni individuo della popolazione ha le stesse probabilità di essere inserito nel campione. Sebbene non sia quello più utilizzato, questo metodo riveste grande importanza teorica poiché rappresenta il termine di confronto di tutti gli altri piani, quando si intenda misurarne l'efficienza relativa.

Le unità sono estratte una alla volta, mentre quelle rimanenti hanno la stessa probabilità di essere estratte successivamente. Si utilizzano numeri casuali, che fino ad alcuni anni fa erano presi da tabelle ed ora spesso sono prodotti mediante computer. Si utilizzano estrazioni caratterizzate dall'assenza di una legge di ordinamento o di successione. Due campioni casuali semplici senza ripetizione sono considerati distinti, se contengono almeno un elemento differente; il loro numero è dato dalle **combinazioni di N elementi n a n**

$$C_N^n$$

dove **N** è il numero di individui che formano la popolazione ed **n** quello che forma il campione, mentre la probabilità di un campione specifico è  $1/C_N^n$ .

Il **campionamento sistematico** è un metodo semplice: da un elenco numerato degli individui che formano la popolazione, dopo l'estrazione casuale della prima unità, eventualmente con un numero random, si estraggono gli individui successivi a distanza costante. Per esempio, se in un elenco si opta per un intervallo di dieci unità, dopo aver estratto per primo il numero 6 le unità campionate saranno 6, 16, 26, ecc. Oltre all'estrema semplicità, tale metodo, di uso molto popolare quando si disponga di un elenco nominativo completo, ha due vantaggi: richiede l'estrazione di un solo numero casuale e fornisce un risultato più accurato della semplice estrazione casuale. Presenta anche due svantaggi potenziali: se, per motivi di distribuzione geografica o temporale, la popolazione è soggetta ad una variazione periodica di entità simile, il campione è errato in quanto lo elimina; inoltre, non esiste un metodo per stimare l'errore standard della media. Il metodo del campionamento sistematico supera questi inconvenienti e non determina errori, quando è parte di un disegno sperimentale più complesso.

Nel **campionamento casuale semplice con ripetizione**, le  $n$  unità del campione vengono estratte con ripetizione e con probabilità costante, uguale a  $1/N$ . Due campioni sono considerati distinti se contengono almeno una unità differente oppure le stesse unità ma in un ordine differente; il numero di campioni possibili è determinato dalle disposizioni con ripetizione di  $N$  elementi presi  $n$  a  $n$ , equivalente a  $N^n$  ed il singolo campione ha probabilità  $1/N^n$ .

Anche in questo caso si possono utilizzare i numeri casuali, con l'avvertenza che se un numero compare più volte l'unità corrispondente deve essere inserita nel campione altrettante volte.

Il **campionamento casuale stratificato** rappresenta un raffinamento, sulla base di conoscenze delle caratteristiche della popolazione, per aumentare l'efficienza del metodo. La popolazione deve essere riunita in gruppi tra loro omogenei e l'estrazione casuale è esercitata all'interno di essi, come se si trattasse di tanti campioni casuali semplici. La stratificazione può essere fatta sulla base di due o più caratteri.

Nella ricerca ambientale, spesso la stratificazione è intesa in senso geografico: dopo aver prefissato il numero di unità da rilevare per ogni zona, l'estrazione dalla popolazione avviene per caso. Offre l'opportunità di ottenere risultati distinti per ogni area e quindi un miglioramento delle stime. Si supponga di avere una popolazione in cui i maschi siano il 60 % e le femmine il 40% e di voler esaminare l'altezza media con i maschi che sono più alti delle femmine. Un campionamento totalmente casuale, senza distinzione tra i sessi, non permetterebbe un'analisi separata e potrebbe determinare una media totale distorta, se la proporzione tra i sessi nel campione risultasse differente da quella della popolazione. Insieme con la stratificazione in sessi, se importanti per il parametro altezza si potrebbero considerare anche altre stratificazioni, come quella per età.

Con  $n_1$  modalità del primo fattore (sesso) e  $n_2$  modalità del secondo (classe d'età), il numero di strati è  $n_1 \times n_2$ .

Il **campionamento casuale a grappoli** è utilizzato quando gli individui sono suddivisi, in modo naturale od artificiale, in gruppi legati da vincoli di contiguità. Caratteristica distintiva del metodo è che le unità non sono scelte in modo diretto, ma estratte in quanto appartenenti ad un certo gruppo. Ad esempio, per rispondere alle domande di un questionario sul traffico si immagini di interrogare tutti gli abitanti di alcune vie, scelte in modo casuale o ragionato. Le domande sono rivolte agli individui, ma la scelta è avvenuta per la strada. Per analisi dell'inquinamento idrico di una regione, si immagini di rilevare tutti i laghi o i fiumi di alcune aree prestabilite, eventualmente scelte a caso.

Spesso il motivo principale di tale procedimento è la mancanza di un censimento completo delle unità da rilevare. Il metodo ha lo svantaggio di contenere un numero variabile di unità entro ogni gruppo e, in genere, di essere meno efficiente del campionamento casuale semplice.

Il **campionamento a due stadi**, detto anche **campionamento a grappoli con sottocampionamento**, è analogo a quello a grappoli in quanto le aree da campionare sono scelte come i grappoli. Questo metodo si differenzia dal precedente, in quanto solo una parte delle unità elementari contenute nei grappoli fanno parte del campione. Al primo stadio, o livello, si estraggono i grappoli, chiamati unità primarie; al secondo, si estraggono le unità secondarie o elementari. Utilizzando ancora l'esempio precedente, si scelgono dapprima le aree entro le quali misurare l'inquinamento dei laghi e successivamente i laghi da campionare.

Il metodo può essere generalizzato con facilità ed essere esteso al caso di 3 o più stadi. Per analizzare i laghi di una regione si scelgono dapprima due o tre province, poi alcune zone entro le province, infine un certo numero di laghi entro le zone.

Il **campionamento con probabilità variabili** si differenzia dai precedenti, in quanto le unità sono scelte con probabilità differenti. Le modalità sono numerose; ma per illustrare il metodo può essere utile il concetto di scelta dei laghi, dopo aver fissato le aree, con probabilità differenti sulla base delle loro dimensioni. Il vantaggio di tale procedimento è quello di una rappresentatività migliore: i laghi maggiori, e quindi presumibilmente quelli più importanti, hanno probabilità maggiori di essere estratti.

## 16.7. TIPO DI CAMPIONAMENTO E VARIANZA

I **vari tipi di campionamento**, dei quali sono stati riportati alcuni tra quelli più semplici e di uso più frequente, permettono una **stima più corretta della media generale e di quelle dei gruppi**. Ma è importante ricordare che **la scelta del tipo di campionamento incide anche sulla varianza** e soprattutto **permette una riduzione della varianza d'errore**, quando la popolazione sia realmente suddivisa in strati con caratteristiche differenti. E' possibile fornire una **dimostrazione matematica** di quanto ripetutamente evidenziato sia nell'analisi della varianza a uno sia in quella a più criteri di classificazione: **la riduzione della varianza d'errore aumenta la significatività della differenza tra le medie a confronto**.

Si supponga di voler studiare i valori e le caratteristiche dell'inquinamento dell'aria in una città. Poiché le fonti che lo possono determinare sono molteplici (riscaldamento, traffico, insediamenti industriali, forni inceneritori, ...) e quasi sempre hanno una distribuzione territoriale che non è casuale, è opportuno distinguere l'intera area in quartieri, ciascuno con alta omogeneità al suo interno e con il massimo di differenza dagli altri.

Con  $r$  strati, ciascuno con numerosità  $n_i$ , media  $\bar{x}_i$  e varianza  $S_i^2$ , è possibile costruire lo schema

$$n_1 \quad n_2 \quad \dots \quad n_i \quad \dots \quad n_r$$

$$\begin{array}{ccccccc} \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_i & \dots & \bar{x}_r & \\ \mathbf{s}_1^2 & \mathbf{s}_2^2 & \dots & \mathbf{s}_i^2 & \dots & \mathbf{s}_r^2 & \end{array}$$

in cui

$N$  è il numero totale di individui della popolazione,

$\bar{x}$  la media e  $\mathbf{s}^2$  la sua varianza.

Se  $\mathbf{p}_i$  è la proporzione del numero di osservazioni in ogni strato rispetto al totale  $N$ , cioè

$$\mathbf{p}_i = n_i / N$$

è possibile definire  $\bar{\mathbf{s}}^2$  la media delle varianze come

$$\bar{\mathbf{s}}^2 = \sum \mathbf{p}_i \sigma_i^2$$

della quale è semplice stimare le relazioni con la varianza totale  $\mathbf{s}^2$  della popolazione.

In una popolazione suddivisa in vari strati, le modalità di estrazione degli elementi che formano il campione possono essere diversi, come riportato nel paragrafo sulle modalità di campionamento. Si supponga di formare tanti campioni di  $n_i$  ripetizioni e che ogni campione contenga elementi scelti a caso entro ciascun strato, con  $\mathbf{p}_i$  frequenza proporzionale alla sua numerosità, cioè

$$n_i = N\mathbf{p}_i$$

In questo modo di scelta delle osservazioni, chiamato anche campionamento secondo lo schema di Poisson, ogni strato è rappresentato in ogni campione di numerosità  $n_i$  con un peso proporzionale alla sua numerosità relativa  $\mathbf{p}_i$ , come non succede con l'estrazione di un solo campione completamente casuale di numerosità  $n$  dalla popolazione, chiamato anche schema di Bernoulli.

Con lo schema di Poisson, la varianza delle medie campionarie  $\bar{\mathbf{s}}^2$  è inferiore a quella dell'intera popolazione  $\mathbf{s}^2$ , ottenuta con lo schema di Bernoulli, in accordo con la relazione

$$\mathbf{s}^2 = \bar{\mathbf{s}}^2 + \mathbf{s}_a^2$$

dove  $\mathbf{s}_a^2$  è la varianza delle medie degli strati, calcolata con la solita formula della varianza tra

$$\mathbf{s}_a^2 = \sum_{i=1}^r \mathbf{p}_i (\bar{x}_i - \bar{x})^2$$

La varianza delle medie campionarie è sempre minore della varianza generale, di una quantità che dipende dalla varianza tra le medie degli strati. Di conseguenza, per ottenere la significatività più alta è vantaggioso che la scelta degli strati sia effettuata con il criterio della massima diversificazione per il fattore preso in esame.

Se il fine è quello di ottenere una riduzione ancora maggiore del valore medio delle varianze  $\bar{\mathbf{s}}^2$ , il criterio della proporzionalità rispetto al numero di individui che compongono i vari strati può essere perfezionato, considerando anche la variabilità interna ai singoli strati. Il numero di individui da

estrarre da ogni popolazione parziale deve essere proporzionale sia alla numerosità dello strato sia alla sua variabilità interna, misurata come deviazione standard  $\sigma_i$ .

Di conseguenza, il numero di individui  $n'_i$  da estrarre da ogni strato deve essere

$$n'_i = n_i \sigma_i / \bar{S}$$

dove

$\sigma_i$  è la deviazione standard dello strato e

$\bar{S}$  è la deviazione standard media di tutti gli strati, misurata con  $\bar{S} = \sum p_i \sigma_i$ .

Se la deviazione standard  $\sigma_i$  di uno strato è maggiore di quella media ( $\bar{S}$ ), la **numerosità ottimale** di quel gruppo è maggiore di quella proporzionale  $n'_i$ ; viceversa in caso contrario.

Con il **campionamento Poissoniano ottimale**, effettuata in rapporto alla variabilità degli strati, si realizza un ulteriore guadagno rispetto al **campionamento proporzionale**.

## 16.8. LA SCELTA DEL TEST

Insieme con la programmazione dell'esperimento e la raccolta dei dati campionari, anche la scelta del test è strettamente collegata all'obiettivo del lavoro. Per una utilizzazione completa dei dati raccolti, è necessario che il tipo di test sia deciso a priori; infatti da esso dipende il tipo di scala, la qualità dell'informazione ed il numero di osservazioni necessari per avere una potenza sufficiente.

Quando il test è scelto a posteriori, a causa di dati che rivelano caratteristiche non desiderate e quindi non è possibile ricorrere alla statistica parametrica, si ha perdita d'informazione con utilizzazione parziale dei dati raccolti. Al contrario, nel caso molto meno frequente in cui si ricorra a un test parametrico dopo che i dati sono stati raccolti per un test non parametrico, si dispone di un numero molto grande di dati, che hanno richiesto tempi più lunghi del necessario e costi molto maggiori.

Utilizzando solo gli argomenti di statistica univariata e bivariata che sono stati affrontati nel corso, i test possono essere raggruppati in tre categorie principali, sulla base della finalità della ricerca:

- **analisi della casualità degli eventi** in una serie di dati;
- **valutazione dei parametri** di un gruppo e **confronti** tra quelli di due o più gruppi;
- **indipendenza od associazione** tra due o più gruppi.

Per ogni obiettivo è possibile scegliere tra una serie di test specifici, in rapporto alle caratteristiche dei dati. Per presentare le analisi che ricorrono con maggiore frequenza nella ricerca scientifica a carattere ambientale ed ecologico, è utile un elenco, anche se parziale.

L'**analisi della casualità** in una serie di dati può riguardare il **numero di eventi** oppure **la loro periodicità**.

La verifica del numero di eventi è condotta mediante le distribuzioni delle probabilità:

- la binomiale, la multinomiale, la ipergeometrica e la poissoniana servono per categorie nominali o qualitative;
- la distribuzione normale è utile sia nel caso di variabili continue o distribuite in modo approssimativamente continuo, sia in quello di variabili binarie quando è possibile avere campioni con un numero elevato di osservazioni.

**L'identità o accordo tra una distribuzione osservata e una distribuzione attesa** possono essere verificati con il ricorso al test  $\chi^2$  o a quello di Kolmogorov-Smirnov; si utilizza il test di Lilliefors per verificare l'accordo con la distribuzione normale. Sono test che permettono l'accertamento di un accordo complessivo, che riguarda contemporaneamente, senza possibilità di distinzione, i quattro parametri di una distribuzione: tendenza centrale, dispersione, simmetria e curtosi.

La periodicità di eventi alternativi può essere indagata mediante il test delle successioni.

**L'analisi dell'indipendenza o dell'associazione tra due o più variabili** nel caso di variabili nominali o qualitative richiede il ricorso al test  $\chi^2$ , al test di Kolmogorov-Smirnov, al metodo esatto di Fisher, al coefficiente C di Cramer o al phi, al test di Mc Nemar. La scelta tra essi dipende da vari fattori: le differenti modalità con le quali possono essere condotti gli esperimenti relativi (campioni dipendenti o indipendenti), il tipo di scala, il numero di osservazioni.

Nel caso di una scala di rango, l'analisi della correlazione richiede il ricorso a test non parametrici come tau di Kendall o il rho di Spearman, mentre analisi che verificano ipotesi analoghe sia alla regressione che all'analisi della varianza possono essere condotte con il test di Cox e Stuart, quello di Jonckheere o quello di Page. Con scale d'intervallo o di rapporto, la correlazione è condotta mediante il prodotto-momento r di Pearson e la regressione lineare semplice con il coefficiente b. La correlazione netta permette di eliminare gli effetti dovute ad una terza variabile.

**L'identità della tendenza centrale in due o più serie di dati** con misure di rango può essere verificata mediante il test dei segni, il test della mediana, il test di Wilcoxon, il test U di Mann Whitney, i test di casualizzazione, quando si utilizzano due campioni dipendenti od indipendenti. Nel caso di k campioni, si ricorre al test di Kruskal-Wallis, a quello di Friedman, all'estensione del test della mediana. E' sempre indispensabile ricordare che questi test si caratterizzano per potenze nettamente differenti, dipendenti dalla metodologia e dal tipo di misura o scala utilizzate; pertanto, per rifiutare l'ipotesi nulla quando essa è falsa, richiedono un numero minimo di osservazioni nettamente differente.

Quando siano soddisfatte le condizioni di validità della statistica parametrica, si utilizzano le procedure del test t o del test F, seguito eventualmente da confronti multipli.

#### **16.9. PERDITA DI SOGGETTI NEL CORSO DELL'INDAGINE**

Nelle ricerche su cavie o persone che durano nel tempo, sovente si perde una certa quota di individui scelti per formare il campione. Di conseguenza, già all'inizio dell'esperimento occorre assicurarsi una nuova dimensione  $n'$  maggiore di quella  $n$  stimata con i metodi presentati.

Se, con l'esperienza derivata da esperimenti precedenti o dall'analisi della letteratura, la quota di perdita di risultati finali è stimata uguale a  $k$ , il campione deve essere di dimensioni  $n'$ , con

$$n' = \frac{n}{1 - k}$$

Nel corso dell'esperimento, gli individui possono essere persi per motivi casuali, che potrebbero colpire senza distinzione tutti quelli che formano il campione oppure per fattori selettivi, che colpiscono solamente una parte, non rappresentativa dell'intero collettivo. Con una perdita selettiva, si determina un campione distorto. L'analisi del campione rimanente non può essere ritenuta corretta, anche se si è aumentata la dimensione del campione.

I metodi per stimare dati perduti dopo la misurazione sono concetti già discussi nell'analisi della varianza a più criteri di classificazione. Le possibilità di ovviare all'inconveniente dipendono dal disegno sperimentale utilizzato. Quando il disegno sperimentale è analogo all'analisi della varianza ad un solo criterio di classificazione non è possibile né utile sostituire i dati perduti; i risultati sono tanto più precisi quanto più alto è il numero di criteri utilizzati per la loro classificazione.

## 16.10. METODI DI RICAMPIONAMENTO PER L'INFERENZA:

### IL JACKKNIFE E IL BOOTSTRAP

Quando non è possibile ricorrere ai test della statistica parametrica (per l'assenza delle condizioni di validità) e a quelli di statistica non parametrica (per ora adatti solo a problemi semplici), per l'inferenza e la stima dei parametri è possibile ricorrere ai **metodi nuovi di ricampionamento**: i due più diffusi, il **jackknife** prima e il **bootstrap** più recentemente, hanno avuto origine da una logica nuova e sono adatti anche per risolvere problemi complessi.

Sono procedure concettualmente elementari, ma che richiedono un numero elevato di calcoli ripetitivi anche quando il campione è formato da poche repliche, soprattutto nel caso del bootstrap. Di conseguenza, il loro uso è divenuto frequente solo dall'inizio degli anni '80, con la diffusione dei computer; a parere di molti, hanno **aperto un nuovo settore della ricerca statistica inferenziale e delle sue applicazioni**.

Secondo alcuni autori (come Robert Sokal e James Rohlf, nella terza edizione del volume *Biometry* del 1995) sono metodi di **statistica parametrica** in quanto, come il test delle permutazioni fondato sul calcolo combinatorio, permettono il confronto tra medie, ne stimano l'errore standard e l'intervallo fiduciale con la distribuzione  $z$  oppure  $t$ .

Per la maggioranza degli statistici, sono da classificare tra i **metodi non parametrici**, perché non richiedono né la normalità della distribuzione della variabile casuale, né la sua conoscenza; inoltre, possono essere applicati a dati ordinali, quali gli indici di similarità, di distanza, di eterogeneità, per citare solo le situazioni nelle quali anche nella ricerca ambientale si ricorre a queste tecniche con frequenza maggiore.

Testi di statistica più recenti (come D. Cook et al. 1999, *Beyond Traditional Statistical Methods*, nel Capitolo 4) parlano sia di un metodo **bootstrap parametrico**, che di un metodo **bootstrap non parametrico**. In questo capitolo seguiremo questa impostazione, relativamente al bootstrap.

Sono chiamati **metodi di ricampionamento** (*resampling methods*), perché sono fondati sull'utilizzazione ripetuta dell'unico campione osservato.

Il **principio** sul quale sono costruiti, detto **della sostituzione**, è elementare: **sussiste una buona validità statistica, se alla funzione di ripartizione della popolazione** (ad esempio, la tabella della distribuzione normale, quella del test  $t$  o di quella  $F$  oppure l'estrazione casuale dai dati di una popolazione) è **sostituita la funzione di ripartizione del campione**, costruendo una distribuzione simile in modo empirico.

Nei test di statistica univariata e bivariata, le **procedure jackknife e bootstrap** sono applicate al confronto tra varianze, all'analisi della varianza ad un criterio di classificazione, al coefficiente di correlazione. Per il confronto tra due varianze campionarie, sono utilizzati in sostituzione del test  $F$ ;

per il confronto tra più varianze, possono essere usati in modo alternativo ai test di Hartley, di Cochran e di Bartlett. Nel caso dell'inferenza su medie, sono applicati al fine di verificare sia la significatività dello scostamento di una misura tratta da un campione rispetto ad un valore atteso e la differenza tra due misure campionarie (in sostituzione del test t), sia per il confronto simultaneo tra più medie (in sostituzione del test F).

Rispetto ai test parametrici, **i due metodi non offrono vantaggi quando sono presenti dati anomali**, in quanto il valore ottenuto risulta ugualmente alterato. E' tuttavia possibile valutare l'effetto dei dati anomali ed individuarli.

Uno degli scopi fondamentali della statistica è la stima di un parametro, già presentata nei paragrafi dedicati agli **intervalli fiduciali o di confidenza**. Di essi vengono ricordati i concetti fondamentali, utili per la comprensione dei metodi qui discussi.

1 - Quando **i dati sono distribuiti in modo normale e la deviazione standard della popolazione (s) è nota**, a partire dalla formula dell'inferenza sulla media

$$Z = \frac{\bar{X} - m}{\frac{S}{\sqrt{n}}}$$

è possibile conoscere la media della popolazione (m) stimandone l'intervallo fiduciale intorno alla media di un campione ( $\bar{x}$ ) di n dati; alla probabilità 1-a, con la distribuzione z si stima l'intervallo fiduciale mediante la relazione:

$$1-\alpha = \Pr\left(\bar{x} - z_{\alpha/2} \frac{S}{\sqrt{n}} < m < \bar{x} + z_{\alpha/2} \frac{S}{\sqrt{n}}\right)$$

Inversamente, quando sono noti la media della popolazione (m) e la sua deviazione standard (s), ogni media campionaria ( $\bar{x}$ ) di n dati con probabilità 1-a è compresa nell'intervallo

$$1-\alpha = \Pr\left(m - z_{\alpha/2} \frac{S}{\sqrt{n}} < \bar{x} < m + z_{\alpha/2} \frac{S}{\sqrt{n}}\right)$$

sempre stimato con la distribuzione normale z.

2- Quando **la distribuzione dei dati della popolazione è normale e la deviazione standard della popolazione è ignota**, a partire da

$$t_{(n-1)} = \frac{\bar{X} - m}{\frac{s}{\sqrt{n}}}$$

è possibile conoscere la media della popolazione ( $m$ ) stimandone l'intervallo fiduciale intorno alla media di un campione ( $\bar{x}$ ) di  $n$  dati, ricorrendo alla sua deviazione standard ( $s$ ); alla probabilità  $1-\alpha$ , è compresa nell'intervallo determinato con la distribuzione  $t$  mediante la relazione:

$$1-\alpha = \Pr\left(\bar{x} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} < m < \bar{x} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}\right)$$

Inversamente, quando è nota la media della popolazione ( $m$ ) e la sua deviazione standard è ignota, ogni media campionaria ( $\bar{x}$ ) di  $n$  dati, di cui sia calcolata la deviazione standard ( $s$ ), con probabilità  $1-\alpha$  è compresa nell'intervallo

$$1-\alpha = \Pr\left(m - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} < \bar{x} < m + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}\right)$$

stimato con la distribuzione  $t$  di Student.

Per determinare l'intervallo di confidenza di un qualsiasi parametro della popolazione è richiesta la conoscenza della sua variabilità associata ad uno stimatore del parametro.

In modo più generale, analogo al caso della media e quindi con  $q = m$  e  $\hat{q} = \bar{X}$ , è possibile pervenire alla stima del parametro della popolazione quando è noto l'errore standard della popolazione oppure del campione o almeno la distribuzione

$$\frac{\hat{q} - q}{\hat{e}s(\hat{q})}$$

Ma quando  $q$  non è la media della popolazione e  $\hat{q}$  non è la media del campione, ma rispettivamente un parametro e una statistica che non godono delle stesse proprietà, può essere difficile o addirittura impossibile ottenere la stima dell'errore standard e la sua distribuzione. In queste situazioni, per trovare una soluzione è possibile ricorrere alla **simulazione**, quando si disponga, caso più teorico che reale, dei dati di una popolazione di forma qualsiasi od ignota.

La tecnica e le sue potenzialità sono meglio illustrate con un esempio.

Si supponga di avere una popolazione di dati e di voler stimare il suo 75° percentile, per campioni formati da 20 dati (non è nota la sua distribuzione, che ovviamente si discosta da quella della media).

I passaggi richiesti possono essere schematizzati in 4 punti:

1 - estrarre dalla popolazione un campione delle dimensioni desiderate (20 dati);

- 2 - calcolare  $\hat{q}$ , identificando il 75° percentile (cioè il 15° valore nella serie dei 20 dati, ordinati in modo crescente);
- 3 - estrarre un altro campione e calcolare il suo  $\hat{q}$ , ottenendo  $\hat{q}_1, \hat{q}_2, \dots, \hat{q}_{1000}$ , fino al numero desiderato di repliche che deve essere alto, per esempio 1000;
- 4 - calcolare la media dei 1000  $\hat{q}$  (che sarà l'indicatore migliore di  $q$ , il 75° percentile della popolazione) e la sua deviazione standard; l'intervallo fiduciale è facilmente costruito dalla distribuzione di frequenza dei 1000  $\hat{q}$ : per la probabilità  $\alpha = 0.05$  è sufficiente escludere il 2,5% dei valori nei due estremi (in altri termini, tra  $\alpha/2$  e  $1-\alpha/2$ ).

**Le tecniche di ricampionamento seguono una logica simile, ma disponendo solo dei dati di un campione.**

**A - Il JACKKNIFE serve per ridurre le distorsioni sistematiche, che dipendono dai dati campionari, nella stima delle statistiche di una popolazione di cui fornisce l'errore standard.** Permette quindi, con l'uso del  $t$  di Student, di calcolare l'intervallo di confidenza per la statistica in esame, assumendo che essa sia distribuita normalmente, almeno in modo approssimato. E' essenziale comprendere che **l'assunzione di normalità riguarda la statistica elaborata con il metodo jackknife, non la distribuzione delle misure campionarie.**

Il termine jackknife in inglese indica il coltello a serramanico; per estensione, il coltello degli esploratori che contiene varie lame e molti altri strumenti, come apribottiglie, lime, forbici, cacciavite. E' funzionale in situazioni di emergenza; è inutile quando si disponga degli strumenti specifici, più solidi e professionali. Secondo Garhwaite et al. (*Statistical inference*, London, Prentice-Hall, 1995), il termine è stato scelto opportunamente, poiché il metodo ha una applicazione appropriata quando non è possibile utilizzare uno strumento analitico, che sia accurato e specifico per l'inferenza e la stima del parametro della popolazione.

L'idea di base del metodo jackknife, **come proposta da Tukey nel 1958 sviluppando l'idea proposta da Quenouille nel 1949, serve anche per costruire intervalli di confidenza intorno alla media.**

**La metodologia è bene evidenziata dalla serie di operazioni richieste, che possono essere schematizzate in 7 passaggi:**

- 1 - calcolare la statistica  $St$  desiderata (per esempio la varianza  $s^2$  o la correlazione  $r$ ) utilizzando le  $N$  osservazioni del campione raccolto;

- 2 – dividere il campione in sottogruppi; se il campione è di grandi dimensioni, i sottogruppi sono formati da  $k$  unità; se il campione è di piccole dimensioni, come spesso succede, i sottogruppi possono essere formati da una sola unità;
- 3 – calcolare il valore della statistica desiderata senza un sottogruppo, ignorando ogni volta un sottogruppo diverso  $St_i$ ; si ottengono  $N/k$  differenti stime della statistica;
- 4 – calcolare i cosiddetti **pseudovalori**  $q_i$  (chiamati in questo modo perché cercano mediamente di stimare il parametro  $q$  riproducendo le variabili originarie) per ogni stima di  $St_i$  mediante la differenza

$$q_i = N \cdot St - (N - 1) \cdot St_i$$

- 5 - la stima con il jackknife della statistica in oggetto  $\hat{S}_t$  è semplicemente la media aritmetica  $\bar{q}$  di questi valori  $q_i$

$$\hat{S}_t = \frac{\sum q_i}{N} = \bar{q}$$

- 6 – mentre l'errore standard  $es$  di  $\hat{S}_t$  è

$$es_{\hat{S}_t} = \sqrt{\frac{\sum (q_i - \bar{q})^2}{N(N - 1)}}$$

in cui

$$s_{\hat{S}_t} = \sqrt{\frac{\sum (q_i - \bar{q})^2}{(N - 1)}}$$

è la deviazione standard;

- 7 – con il valore del  $t$  di Student alla probabilità  $\alpha$  prescelta e per gdl  $N-1$ , si stimano i limiti di confidenza

$$\hat{S}_t \pm t_{(\alpha/2, N-1)} \cdot es_{(St)}$$

entro il quale alla probabilità  $\alpha$  prefissata si troverà il parametro della popolazione.

Quando è applicato al coefficiente di correlazione,

- per confrontare un  $r$  sperimentale con un valore teorico o per verificare l'uguaglianza di due  $r$  campionari ( $H_0: \rho_1 = \rho_2$ ),
- al fine di eliminare l'asimmetria di un valore diverso da 0 (vedi le trasformazioni dei dati)

si ricorre alla trasformazione  $z$ , sviluppata da R. A. Fisher

$$z = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right)$$

(E' la formula per l'iperbolica inversa tangente di  $r$ , simboleggiato con  $\tanh^{-1}$ , discussa nella trasformazione dei dati).

Essa rende ogni valore di  $r$  distribuito in modo approssimativamente normale.

La sua varianza  $s_z^2$  è

$$s_z^2 = \frac{1}{N-3}$$

dove  $N$  è il numero di coppie di valori.

Per la **verifica dell'omogeneità della varianza** tra due o più gruppi con distribuzione non normale, con il jackknife è possibile stimare  $N/k$  varianze in ogni gruppo e quindi ottenere altrettante misure diverse dai loro rapporti.

Nell'**analisi della varianza ad un criterio di classificazione** con  $p$  gruppi, si hanno altrettanti rapporti

$$s_{tra}^2 / s_{entro}^2$$

con l'esclusione di un gruppo alla volta.

Se sono presenti dati anomali (**outliers**), l'errore standard calcolato con il metodo jackknife è fortemente sottostimato. L'esame dei pseudo-valori permette di evidenziare gli effetti dei dati di volta in volta esclusi e quindi di vedere se ognuno di essi, e quale, è un dato anomalo. La quantità  $\bar{q} - q_i$ , definita da Devlin nel 1975 "la **funzione di influenza del campione**", è una misura conveniente dell'effetto che un dato specifico ha sulla statistica in studio.

ESEMPIO.

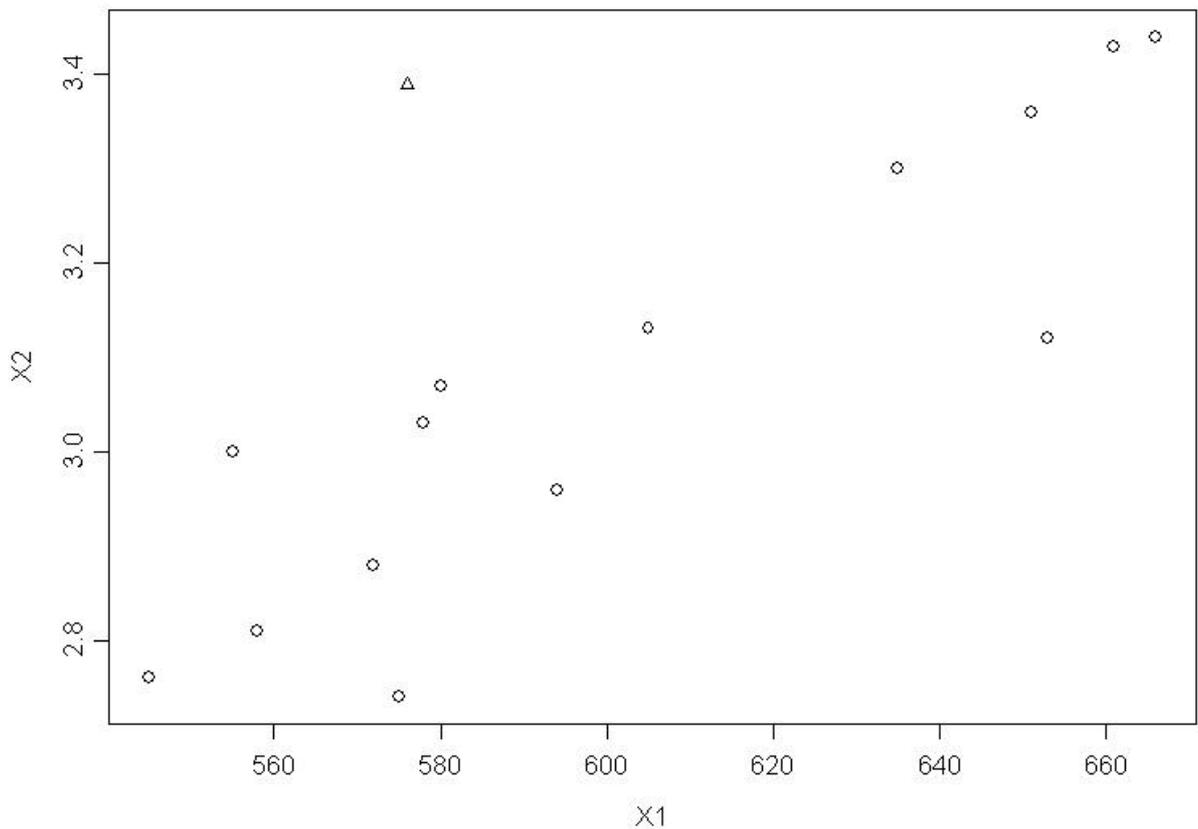
Applicazione del metodo Jackknife per l'analisi di un coefficiente di correlazione, con presentazione dei risultati ottenuti con un programma informatico a grande diffusione.

Consideriamo le due variabili della tabella che segue

| X1  | X2   |
|-----|------|
| 576 | 3.39 |
| 635 | 3.30 |
| 558 | 2.81 |
| 578 | 3.03 |
| 666 | 3.44 |
| 580 | 3.07 |
| 555 | 3.00 |
| 661 | 3.43 |
| 651 | 3.36 |
| 605 | 3.13 |
| 653 | 3.12 |
| 575 | 2.74 |
| 545 | 2.76 |
| 572 | 2.88 |
| 594 | 2.96 |

con 15 osservazioni campionarie.

Per meglio comprendere le caratteristiche dei dati, il primo passo è la rappresentazione grafica



Il diagramma di dispersione già a prima vista indica la presenza di un possibile outlier, rappresentato con un triangolo nell'angolo in alto a sinistra.

Il calcolo del coefficiente di correlazione con i 15 punti fornisce un valore di **r (Observed)** dei valori osservati uguale a 0,7764.

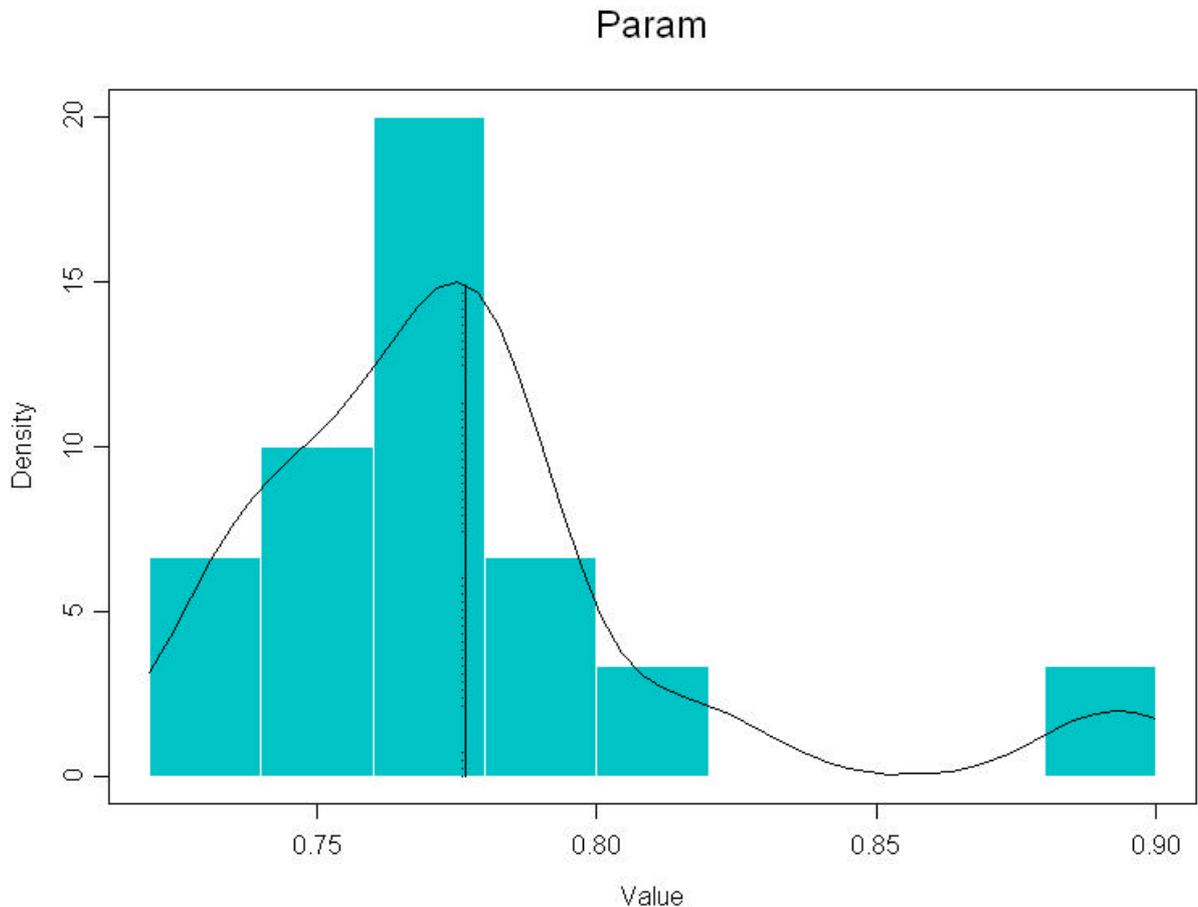
Ma è possibile calcolare altri 15 valori di **r**, togliendo ogni volta una coppia di dati e quindi utilizzare ogni volta solo 14 valori.

La media di questi 15 valori (**Mean**) è uguale a 0,7759 con una distorsione (**Bias**) di 0.0005 rispetto al valore calcolato con il metodo classico; l'errore standard (**SE**) del valore **r** medio è 0,1425.

| Observed | Bias    | Mean   | SE     |
|----------|---------|--------|--------|
| 0.7764   | -0.0005 | 0.7759 | 0.1425 |

E' sempre conveniente, per una lettura più agevole, fare una distribuzione grafica di questi 15 valori calcolati, come riportato nella figura sottostante. In essa emerge con chiarezza come il valore di **r** calcolato senza la presenza di quel punto anomalo, sia diverso da quelli ottenuti quando esso è

compreso. E' una dimostrazione dell'anomalia di quel punto, che a prima vista né sulla tabella dei dati né sul diagramma di dispersione, risultava con altrettanta evidenza.



Distribuzione dei 15 pseudo-valori di  $r$  calcolati dalle 15 coppie di dati campionari.

(Le ordinate sono state moltiplicate per 3 per motivi grafici. Il programma utilizzato e' S-plus ver. 5.0. L'esempio e' stato tratto dal manuale d'istruzione di S-plus di pag. 970).

Il metodo jackknife permette quindi una stima del parametro e del suo intervallo fiduciale quando non sono rispettate le condizioni di validità per la statistica parametrica classica.

Mediante l'intervallo fiduciale è possibile passare all'inferenza rispetto alla sua media, verificando se un certo valore atteso oppure un altro valore sperimentale è compreso nell'intervallo calcolato.

Ma per questi aspetti inferenziali attualmente si preferisce ricorrere all'uso del bootstrap.

**B – Il BOOTSTRAP**, introdotto da **B. Efron** nel 1979 (per una sua presentazione più aggiornata, vedere dello stesso autore: *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM, 1982; *An introduction to the Bootstrap*, Chapman & Hall, 1993), è una tecnica che permette di ricavare

**gli errori standard e i limiti di confidenza di varie misure statistiche;** è un metodo generale per ottenere informazioni circa la variabilità e la distribuzione di statistiche campionarie  $\hat{q}$  e quindi da esse stimare i limiti di confidenza del parametro  $q$  della popolazione, quando non si possiedono informazioni sulla sua distribuzione.

Il metodo bootstrap ha il grande vantaggio di fornire le stime statistiche necessarie all'inferenza anche per funzioni molto complesse; oltre a quelli già indicati nella presentazione generale, esempi più specifici d'applicazione sono il calcolo dei momenti, dei coefficienti di variazione, dei rapporti tra valori medi e fra varianze, dei coefficienti di correlazione, degli autovalori delle matrici di varianze e covarianze. Usi frequenti nella ricerca ambientale sono l'inferenza su indici di similarità o distanza calcolati mediante la presenza-assenza di specie o l'intensità di alcuni parametri fisici e chimici, la costruzione e l'interpretazione di dendrogrammi; ugualmente diffusa è l'applicazione nella tassonomia numerica e nella costruzione di alberi filogenetici, fondati sulle frequenze geniche o di più caratteri fenotipici.

**L'uso del bootstrap non è possibile con i quantili** (quanto si disponga solo di essi non dei valori reali), **con dati incompleti, non indipendenti o alterati da errori grossolani.**

**Se il campione è formato da  $k$  dati, l'idea di base è simulare il processo di selezione di molti campioni di  $k$  osservazioni, allo scopo di trovare la probabilità che i valori cadano all'interno di intervalli predeterminati.** Il campione bootstrap è nient'altro che il campione originario nel quale, per effetto dell'estrazione con ripetizione, alcuni dati sono ripetuti ed altri, per mantenere lo stesso numero d'osservazioni, sono assenti.

E' proprio la **modalità di estrazione, fondata sulla ripetizione**, a generare la variabilità nelle stime; poiché è richiesto che i campioni abbiano tutti lo stesso numero d'osservazioni, se si estraesse senza ripetizione sarebbero tutti identici. Il nome di bootstrap (letteralmente stringhe o lacci da scarpe), derivato dall'espressione inglese "to pull yourself up by your own bootstrap" (*tirarsi su attaccandosi ai lacci delle proprie scarpe*), evidenzia il fatto paradossale che l'unico campione disponibile serve per generarne molti altri.

La metodologia è fondata sulla **generazione casuale di un numero grandissimo di copie**, di norma da un migliaio a un milione, **del campione di  $k$  osservazioni**. Il migliaio o milione di copie viene casualmente mescolato; con la generazione di numeri casuali vengono estratti migliaia o milioni di stringhe di  $k$  osservazioni.

Ognuna di queste stringhe di  $k$  osservazioni può contenere due o più valori identici, con l'ovvia esclusione d'altri valori che sono contenuti nel campione originale. Sono chiamati campioni di bootstrap, ognuno dei quali permette di ottenere una stima della statistica desiderata.

La distribuzione della misura statistica calcolata è trattata come una distribuzione costruita a partire da dati reali e fornisce una stima dell'accuratezza statistica.

Per esempio, se con 15 coppie d'osservazioni è stata calcolato un  $r$  uguale a 0,782, con 10 mila stringhe è possibile stimare che il 90% (quindi con esclusione dei 500 valori minori e i 500 maggiori) variano tra 0,679 e 0,906. L'ampiezza dell'intervallo (0,227) è la misura dell'accuratezza fornita dal bootstrap, per il valore  $r$  calcolato sul campione delle 15 osservazioni raccolte.

Vari autori hanno dimostrato che, molto spesso, l'ampiezza dell'intervallo stimato in questo modo corrisponde all'intervallo calcolato su campioni reali dello stesso numero d'osservazioni  $k$ .

E' un risultato che può apparire paradossale. Tuttavia, come nelle misure statistiche, il risultato non è garantito: anche il bootstrap può fornire risposte fuorvianti, in una percentuale ridotta di campioni possibili, senza che sia possibile sapere in anticipo quali siano.

**Il bootstrap è considerato un miglioramento del jackknife e può essere applicato anche a stime di jackknife (jackknife bootstraped).**

Per meglio comprendere i concetti in modo operativo, si possono definire i passaggi fondamentali richiesti dalla metodologia:

- 1 – a partire dal campione osservato ( $x_1, x_2, \dots, x_k$ ) si costruisce una popolazione fittizia, ripetendo  $n$  volte ognuno dei  $k$  dati;
- 2 – si estrae un **campione casuale, definito campione bootstrap**, estraendo  $k$  dati; nella striscia possono essere presenti una o più repliche dello stesso dato e quindi mancare un numero corrispondente di valori presenti nel campione originale
- 2 – per ciascuno di tali campioni bootstrap si calcola lo stimatore  $q$  desiderato;
- 3 – con  $n$  estrazioni si ottiene la successione di stime, che sono la realizzazione della variabile casuale “stimatore bootstrap  $T$ ”;
- 4 – la funzione di ripartizione empirica dei  $k$  valori  $q$  ottenuti fornisce una stima accurata delle caratteristiche della variabile casuale  $T$ ,
- 5 – l'approssimazione è tanto più precisa quanto più  $n$  è elevato.

Dalla serie dei valori  $q$ , è possibile ottenere

la **stima della distorsione**,

la **stima della varianza** e

gli **intervalli di confidenza**.

La **stima bootstrap della distorsione**  $b_k$  è la distanza tra il valore calcolato con i  $k$  dati originari e quello medio calcolato con le  $n$  estrazioni del metodo bootstrap.

Se  $\theta$  è il valore calcolato con il campione osservato

e  $\hat{q}^*$  è la media dei  $k$  valori  $\hat{q}_i^*$  calcolati con  $n$  replicazioni bootstrap, cioè

$$\hat{q}^* = \frac{\sum_{i=1}^k \hat{q}_i^*}{n}$$

la stima bootstrap  $b_k$  della distorsione è data da

$$b_k = \hat{q}^* - \theta$$

In modo analogo e con la stessa simbologia, una **stima dell'errore standard della distribuzione** è data da

$$\hat{e}_s(\hat{q}) = \sqrt{\frac{\sum_{i=1}^k (\hat{q}_i^* - \hat{q}^*)^2}{n-1}}$$

Per la stima dell'**intervallo di confidenza**, che è l'applicazione più importante del bootstrap, è utile distinguere tra

- **bootstrap parametrico** quando è nota la forma funzionale della popolazione e
- **bootstrap non parametrico** quando la forma della distribuzione è completamente sconosciuta; è il caso più frequente di utilizzazione del bootstrap e la stima dei parametri richiede la metodologia precedentemente descritta; con il solo termine "bootstrap", di norma si intende quello non parametrico.

Nel caso di **bootstrap parametrico**, rispetto alla metodologia appena presentata la differenza fondamentale consiste nell'estrazione del campione di  $k$  dati, che avviene direttamente dalla popolazione, invece che da una popolazione fittizia formata appunto da  $n$  copie dei  $k$  valori osservati. L'**intervallo di confidenza** può essere determinato, in modo facilmente intuitivo, sulla base della distribuzione dei  $k$  valori  $\hat{q}_i^*$  calcolati con  $n$  replicazioni bootstrap, con il **metodo percentile**: con 1000 repliche, se  $\alpha = 0.05$ , è sufficiente considerare l'intervallo che esclude il 2,5% dei valori nei due estremi.

Il **metodo del percentile** ordina per rango i valori ottenuti con le  $k$  replicazioni bootstrap, individuando i valori estremi entro i quali si collocano,  $1-\alpha$  (in percentuale) valori della distribuzione, con  $\alpha/2$  in ognuno dei due estremi.

**Fondato sui ranghi, il metodo ovviamente non varia anche con trasformazione dei dati. Richiede che la distribuzione bootstrap non sia distorta e possiede un'accuratezza limitata per piccoli campioni.**

Può essere applicato nel caso dell'esempio precedente, per individuare il 75° percentile di un gruppo di dati, se la popolazione d'origine è distribuita in modo normale.

Il metodo del percentile non sempre è corretto: l'intervallo di confidenza che comprende il 95% dei  $k$  valori  $\hat{q}_i^*$  calcolati non sempre contiene  $\theta$  il 95% delle volte.

Una stima più prudente, anche se valida quando i dati sono distribuiti in modo normale, è dato dall'**intervallo standard z**, compreso tra

$$\hat{q} - z_{1-\alpha/2} \cdot \hat{es}(\hat{q}) \quad \text{e} \quad \hat{q} + z_{\alpha/2} \cdot \hat{es}(\hat{q})$$

considerati in valore assoluto.

Nel caso di **bootstrap non parametrico**, per calcolare l'**intervallo di confidenza** sono state proposte diversi metodi alternativi a quello del percentile (ugualmente indicativo anche in questo caso, ma con gli stessi limiti) con numerose varianti, di cui le 4 ora più diffuse sono:

- 1) il metodo bootstrap-t,
- 2) il metodo della varianza stabilizzata bootstrap-t,
- 3) il metodo BCa (**B**ias-**C**orrected and **a**ccelerated),
- 4) il metodo ABC (**A**pproximate **B**ootstrap **C**onfidence).

Il metodo bootstrap-t e il metodo della varianza stabilizzata bootstrap-t (the variance-stabilized bootstrap-t) hanno in comune l'idea di correggere la distribuzione ottenuta mediante la distribuzione z; come già visto nel caso delle medie, sono da applicare soprattutto quando il campione è di piccole dimensioni.

L'intervallo è compreso tra

$$\hat{q} - t_{(n-1; 1-\alpha/2)} \cdot \hat{es}(\hat{q}) \quad \text{e} \quad \hat{q} + t_{(n-1; \alpha/2)} \cdot \hat{es}(\hat{q})$$

E' importante ricordare che i **2 metodi t non necessariamente utilizzano la distribuzione t di Student**, ma una distribuzione ottenuta con una seconda serie di valori "bootstrappati" derivati da

$$t^* = \frac{\hat{q} - q}{\hat{es}(\hat{q})}$$

che può coincidere con la distribuzione t di Student solo quando  $q = m$  e  $\hat{q} = \bar{X}$ .

In generale, se l'errore standard della statistica campionaria  $\hat{es}(\hat{q})$  **non è noto**, si procede ad una sua stima con un secondo livello di bootstrap, mediante i seguenti passaggi:

- 1) dopo una parte iniziale identica a quella precedente,

- a) dalla popolazione fittizia di  $n$  repliche dei  $k$  dati campionari, estrarre un campione di  $k$  dati e calcolare  $\hat{q}$  (la stima di  $q$  ottenuta dal campione);
- b) generare una intera striscia con  $n$   $\hat{q}$ , dalla quale stimare  $\hat{q}^*$
- 2) generare una serie di queste stime  $\hat{q}^*$ 
  - a) dalla quale calcolare  $\hat{q}^{**}$ ;
  - b) ripetere queste ultime operazioni in modo da ottenere  $\hat{q}_1^{**}, \hat{q}_2^{**}, \dots, \hat{q}_n^{**}$
  - c) calcolare la deviazione standard di questi  $\hat{q}_i^{**}$  valori, cioè  $\hat{e}s_{n^*}(\hat{q})$  che sarà una stima dell'errore standard della statistica campionaria  $\hat{e}s(\hat{q})$
  - d) calcolare

$$t^* = \frac{\hat{q} - q}{\hat{e}s_{n^*}(\hat{q})}$$

- 3) ripetere l'operazione per ottenere  $t_1^*, t_2^*, \dots, t_n^*$ ;
- 4) dalla distribuzione di questi valori, utilizzando i percentili  $t_{\alpha/2}^*$  e  $t_{1-\alpha/2}^*$  si ottiene l'intervallo di confidenza per  $\theta$ , che alla probabilità  $\alpha$  sarà compresa nell'intervallo tra

$$\hat{q} - t_{1-\alpha/2}^* \cdot \hat{e}s_i(\hat{q}) \quad \text{e} \quad \hat{q} - t_{\alpha/2}^* \cdot \hat{e}s_i(\hat{q})$$

preso in valore assoluto.

Questo metodo è generalmente valido per stimare parametri di posizione (come la mediana), in modo particolare se si ricorre al bootstrap parametrico; ma non è ugualmente attendibile nella stima di altri parametri, specialmente quando i campioni sono piccoli-

- Il metodo **variance-stabilized bootstrap-t** corregge il metodo bootstrap -t precedente.
- Il **metodo BC** (Bias-Corrected) **indicato ora più frequentemente BCa** (Bias-Corrected and accelerated) è un altro miglioramento del metodo percentile, da utilizzare soprattutto nel caso di bootstrap non parametrico: "centra" la distribuzione del bootstrap sul valore del parametro calcolato con i dati del campione osservato e per essa stima i quantili asintotici. Non tiene conto della eventuale asimmetria ed implica un'assunzione parametrica, come avviene per la distribuzione normale e la distribuzione  $t$ .
- Il **metodo ABC** (Approximate Bootstrap Confidence) usa alcuni risultati analitici per ridurre il numero di campioni bootstrap necessari nel metodo **BCa**

ESEMPIO. Bootstrap di un coefficiente di correlazione.

Vengono utilizzate le variabili X1 e X2 dell'esempio precedente sul jackknife, con lo stesso diagramma di dispersione per rappresentare i punti. Di ognuna delle 15 coppie ( $k$ ) di dati sono state fatte 1000 copie ( $n$ ) e dall'insieme di 15000 dati sono state fatte 1000 estrazioni casuali di 15 dati.

Di ogni serie di 15 dati è stata calcolato il coefficiente di correlazione  $r$ .

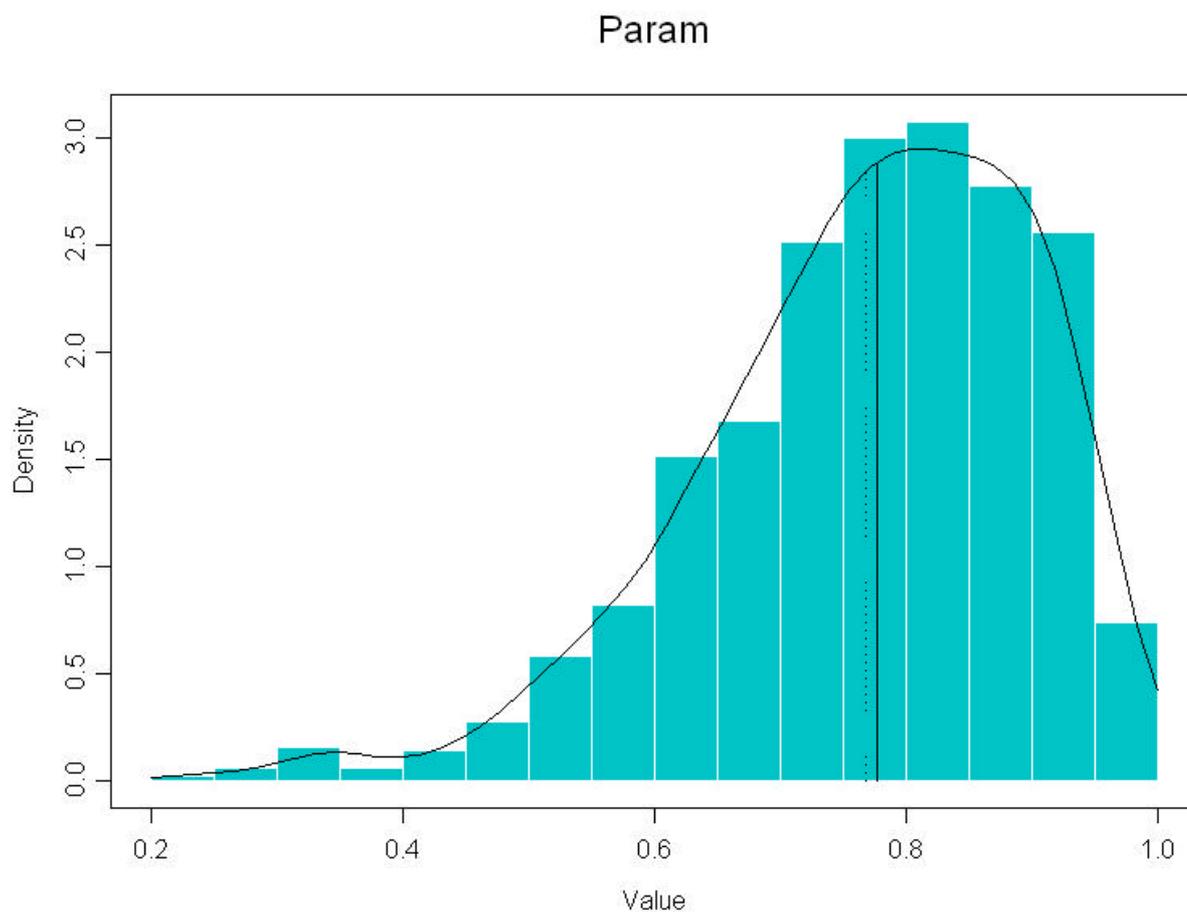


Figura. Distribuzione di frequenza dei 1000 valori di  $r$  calcolati con il metodo bootstrap, a partire dalle 5 coppie di osservazioni riportate nella tabella precedente.

Insieme con il grafico, i programmi informatici riportano le statistiche relative.

Summary Statistics:

| Observed | Bias    | Mean   | SE     |
|----------|---------|--------|--------|
| 0.7764   | -0.0088 | 0.7676 | 0.1322 |

Empirical Percentiles:

| 2.5%   | 5%    | 95%    | 97.5%  |
|--------|-------|--------|--------|
| 0.4673 | 0.523 | 0.9432 | 0.9593 |

BCa Percentiles:

| 2.5%  | 5%     | 95%    | 97.5%  |
|-------|--------|--------|--------|
| 0.344 | 30.453 | 0.9255 | 0.9384 |

Il valore di  $r$  calcolato sui 15 dati osservati è uguale a 0.7764 (ovviamente la risposta è uguale a quella data dal jackknife). La media delle 1000 repliche campionarie ottenute per estrazione casuale di 15 numeri è uguale a 0,7676 e il suo errore standard è uguale a 0,1322 (sono leggermente diversi dalla stima jackknife; anche un nuovo calcolo bootstrap, fondato sull'estrazione casuale, non darebbe risultati esattamente coincidenti, anche se molto simili per le dimensioni delle repliche).

Il bias, la differenza tra la stima di  $r$  con i 15 dati campionari e stima della media degli  $r$  ottenuti con il metodo bootstrap, è uguale a  $-0,0088$ .

Quando il campione è piccolo, un aumento del numero di ricampionamenti non determina un miglioramento reale, poiché i cambiamenti sono piccoli rispetto alla deviazione standard.

Le due strisce precedenti danno le stime dell'intervallo di variazione, con il metodo dei percentili empirici e quella dei percentili BCa.

Il jackknife ed il bootstrap rientrano nell'area dei metodi statistici chiamata **data analysis**, insieme con i test di casualizzazione, con alcune tecniche di rappresentazione grafica (che mostrano le caratteristiche di una distribuzione e gli effetti delle trasformazioni) e con la **cross-validation**, che consiste nel dividere casualmente i campioni in due parti uguali per eseguire alcune analisi alternative (come una regressione con numero differente di parametri alla quale sono applicate trasformazioni differenti), finché con il secondo campione si ottiene una risposta simile a quella del primo). Definita come **la ricerca sistematica e approfondita di una serie di dati per evidenziarne le informazioni e le relazioni**, rappresenta la **parte nuova della statistica applicata alla ricerca biologica ed ambientale**. Per alcuni autori, è anche quella che promette gli sviluppi interessanti per i prossimi anni.