CAPITOLO III

ANALISI DELLE FREQUENZE

CAPITOLO III

ANALISI DELLE FREQUENZE

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 proposto il test χ^2 (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 per classi.

Tabella 1/A. Distribuzioni fenotipiche (osservate ed attese) di *Pisum sativum* in alcuni esperimenti

Segregazione di un ibrido:	Fenotipo		
Carattere	Dominante	Recessivo	Totale
a) Colore del fiore	*Rossi 705*	*Bianchi 224*	
Distribuzione attesa (3:1)	696,75	232,25	**929**
b) Lunghezza del fusto	*Alte 787*	*Basse 277*	
Distribuzione attesa (3:1)	798	266	**1064**
c) Colore del seme	*Gialli 6022*	*Verdi 2001*	
Distribuzione attesa (3:1)	6017,25	2005,75	**8023**
d) Forma del seme	*Lisci 5474*	*Rugosi 1850*	
Distribuzione attesa (3:1)	5493	1831	**7324**

Il test χ^2 serve anche per il **confronto tra 2 o più distribuzioni osservate**. In tali situazioni sperimentali, 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 statistica, 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. 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 1/A) 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.

Segregazione di un diibrido	Frequenze	
Fenotipi	Osservate	Attese
Gialli – Lisci	***315***	9/16 = **312,75**
Gialli – Rugosi	***101***	3/16 = **104,25**
Verdi – Lisci	***108***	3/16 = **104,25**
Verdi – Rugosi	***32***	1/16 = **34,75**
Totale	*556*	556,00

Dopo aver calcolato il totale della distribuzione osservata si calcola quella attesa. Se l'ipotesi è quella espressa nella tabella precedente, il totale deve essere diviso per 16, attribuendo poi 9 sedicesimi alla prima classe (Gialli – Lisci), 3 sedicesimi alle classi seconda (Gialli – Rugosi) e terza (Verdi – Lisci) e 1 sedicesimo alla quarta classe (Verdi - Rugosi).

E' evidente (come mostra con chiarezza la tabella 1/B) che tra distribuzione osservata e distribuzione attesa non si ha mai una perfetta coincidenza, anche quando si possono constatare valori molto simili. Inoltre, 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 e quindi è sempre formata da numeri interi, una distribuzione attesa segue una legge teorica di distribuzione dell'ammontare totale e pertanto spesso è formata da classi con numeri frazionali.

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,
- se sono di dimensioni tali da fare più **ragionevolmente supporre una distribuzione realmente diversa da quella attesa**.

La prima asserzione, quella della casualità dell'evento, è chiamata **ipotesi nulla** e viene indicata con $\mathbf{H_0}$.

La seconda, quella dell'esistenza di una differenza reale anche se le cause sono ignote, è chiamata **ipotesi alternativa** e viene indicata con $\mathbf{H_1}$ oppure con H_A.

La scelta tra le due ipotesi avviene sulla base della probabilità stimata con il test. Essa è la probabilità di trovare per caso la distribuzione osservata o una distribuzione che si allontani ancor più da quella attesa, nella condizione che l'ipotesi nulla sia vera. Se la probabilità calcolata è piccola, la logica dell'inferenza statistica rifiuta l'ipotesi nulla, accettando implicitamente l'ipotesi alternativa. Tuttavia si riconosce che è possibile errare in questa scelta; ma con una probabilità definita, non superiore a quella calcolata con il test.

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 $\chi^2_{(\text{ g. d. l. })}$ (chi-quadrato), proposto da **Pearson** nel 1900 (Karl **Pearson** inglese, nato nel 1857 e morto nel 1936, docente a Cambridge, fondatore della rivista Biometrika). Con questo test, le ipotesi sono sulla distribuzione di tassi e proporzioni, ma per la stima della probabilità utilizza le **frequenze assolute**,
secondo la formula

$$\chi^2_{(g.d.l.)} = \sum_{i=1}^{n} \frac{\left(f_i^{oss} - f_i^{att}\right)^2}{f_i^{att}}$$

dove:

- f_i^{oss} = frequenza osservata i-esima,
- f_i^{att} = frequenza attesa i-esima,
- **g.d.l.** = numero di gruppi (n) meno uno (gdl = n-1),
- e la sommatoria **Σ** è estesa a tutti gli **n** gruppi.

La **distribuzione della densità di probabilità** del $\chi^2_{(g.\ d.\ l.)}$ 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** (non del totale degli individui contati).

Il numero di g.d.l. (o gdl, più rapidamente) viene riportato tra parentesi, ai piedi del simbolo: corrisponde al numero di osservazioni indipendenti. Infatti i valori attesi di ogni gruppo, che sono calcolati a partire dal totale ed attribuiti ad ogni gruppo secondo la legge di distribuzione che si vuole verificare, sono liberi di assumere qualsiasi valore. Fa eccezione il valore atteso nell'ultimo gruppo, poiché la sua 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 χ^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 $\mathbf{I^A}$, $\mathbf{I^B}$ ed **i**, mediante lo sviluppo di

$$\left(p+q+r\right)^2 = 1;$$

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 ($\mathbf{I^AI^A}$, $\mathbf{I^Ai}$, $\mathbf{I^BI^B}$, $\mathbf{I^Bi}$, $\mathbf{I^AI^B}$, **ii**) hanno 2 gdl.

Secondo uno schema valido per tutti i test statistici, **il procedimento logico che deve essere seguito** nell'applicazione del χ^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** (indicato con α, i cui criteri di scelta 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% indicato con la simbologia $\alpha = 0.05$);

6 - calcolare il **valore del test statistico** sulla base dei dati sperimentali, stimando la **probabilità** P ad esso associata;

7 - sulla base della probabilità, **trarre le conclusioni**:

- se la probabilità P calcolata risulta superiore a quella α prefissata, concludere che non è possibile rifiutare l'ipotesi nulla H_0;
- se la probabilità P calcolata risulta inferiore a quella α 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 1/B, il calcolo del χ^2 è semplice

$$\chi^2_{(3)} = \frac{(315-312,75)^2}{312,75} + \frac{(101-104,25)^2}{104,25} + \frac{(108-104,25)^2}{104,25} + \frac{(32-34,75)^2}{34,75}$$

$$\chi^2_{(3)} = \frac{(2,25)^2}{312,75} + \frac{(-3,25)^2}{104,25} + \frac{(3,75)^2}{104,25} + \frac{(-2,75)^2}{34,75} = 0,47$$

ottenendo $\chi^2_{(3)} = 0,47$.

Con l'aiuto delle tavole (riportate alla fine del capitolo), 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 χ^2

- per 3 gradi di libertà (indicato sulla riga) e
- per probabilità $\alpha = 0.05$ (indicato sulla colonna),

il valore del χ^2 (approssimato alla seconda cifra decimale) risulta uguale a 7,81.

Il valore calcolato nell'esercizio (0,47) è sensibilmente minore di quello riportato nella tabella (7,81). Si deve concludere che la probabilità di trovare per caso le differenze osservate o differenze ancora maggiori che possano essere attribuite al caso è alta (il 47%), superiore al valore prefissato del 5%. Di conseguenza, non si può rifiutare l'ipotesi nulla. Si afferma che le differenze tra distribuzione osservata e distribuzione attesa non sono significative; con alta probabilità dipendono solo dal caso.

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 χ^2 sarà elevato. Quindi, 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 χ^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 *Mixodiaptomus 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
Allele 1	**26**
Allele 2	**38**
Allele 3	**62**
Allele 4	**118**
Totale	244

E' ragionevole affermare che nella popolazione i quattro alleli abbiano le stesse frequenze e che quelle le differenze riscontrate tra essi possano essere imputate al caso (**H_0**), alle variazioni dovute al campionamento? Oppure è più logico pensare che esse sono effettivamente differenti (**H_1**) anche nella popolazione?

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

$$\chi^2_{(3)} = \frac{(26-61)^2}{61} + \frac{(38-61)^2}{61} + \frac{(62-61)^2}{61} + \frac{(118-61)^2}{61} = \frac{1225}{61} + \frac{529}{61} + \frac{1}{61} + \frac{3249}{61} = 82{,}03$$

risulta $\chi^2_{(3)}$ = 82,03.

Consultando la tabella del chi-quadrato per 3 gradi di libertà,

- alla probabilità α = 0.05 corrisponde un valore critico χ^2 = 7,82
- alla probabilità α = 0.01 corrisponde un valore critico χ^2 = 11,34
- alla probabilità α = 0.001 un valore critico χ^2 = 16,27.

Il valore del chi quadrato calcolato sui dati sperimentali (χ^2 = 82,03) 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 a 0,1%.

Si rifiuta l'ipotesi nulla e implicitamente 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 χ^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** (sulla relazione tra numero di gruppi e gradi di libertà si tornerà nel prosieguo del capitolo).

Il numero di osservazioni, in totale ed entro ogni gruppo, sono a questo riguardo ininfluenti.

Eppure, anche senza esperienza di statistica, appare logico pensare che il risultato sia tanto più attendibile quanto più elevato è il numero di osservazioni nell'esperimento.

Nel test χ^2 il numero di osservazioni, sia in totale che entro ogni classe, determina la condizione essenziale di validità.

Il χ^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 definiti in modo chiaro ed universale, con i quali si possono formare 3 classi di "credibilità" o **validità** del test.

1- **Il test è valido quando il numero totale di osservazioni è superiore a 100**; tuttavia, più recentemente, vari autori ne pretendono un numero maggiore, **circa 200**; altri, ancora più rigorosi, ne chiedono addirittura **500**.

2 - Il test è meno attendibile ed ha bisogno di **una correzione,** quando il numero di osservazioni è **tra 200** (o **100**) **e 30 circa**. E' detta **correzione di Yates**, in quanto proposta da F. **Yates** nel 1934 (con l'articolo "***Contingency tables involving small numbers and the test*** χ^2, pubblicato sulla rivista inglese **Journal of the Royal Statistical Society**, Suppl. 1, pp. 217-235). (Frank **Yates** inglese, nato nel 1902 e morto nel 1994, assistente di Fisher nel 1931 e direttore del Dipartimento di Statistica a Cambridge dal 1933 al 1968).

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, molto raramente fino a 20.

Il motivo è che, con così pochi dati, le variazioni casuali diventano così ampie da non poter mai rifiutare l'ipotesi nulla con una probabilità ragionevolmente bassa, per quanto distanti possano essere le frequenze osservate e quelle attese.

A questa condizione sul numero totale di dati è necessario aggiungerne una seconda:

- **il numero di <u>frequenze attese entro ogni classe non</u>** deve essere **<u>minore di 5</u>**.

Anche su questo secondo criterio non esiste unanimità. Alcuni testi abbassano questo limite fino a 4, molto raramente fino a 3. Altri ancora pongono questi limiti considerando anche il numero di classi: ritengono 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.

E' quindi utile ricordare che, quando ha un numero abbastanza alto di gradi di libertà, il chi quadrato è meno sensibile agli errori determinati da frequenze attese piccole.

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 attesa sarebbe di 5 e, in una distribuzione non uniforme, in alcuni gruppi le frequenze attese sarebbero necessariamente ancora inferiori a tale valore limite.

Per campioni di dimensioni inferiori a 200 (100) ma superiori a 30, si deve apportare la <u>correzione di Yates</u> o <u>correzione per la continuità</u>.

E' una procedura uguale a quella illustrata per l'uso della normale in distribuzioni binomiali con campioni di grandi dimensioni. Per passare da una distribuzione discreta, come è un conteggio, ad una distribuzione continua, come sono la normale e la distribuzione χ^2, con un numero intero (ad esempio, X = 12) si deve intendere **non la misura precisa** (come può essere una lunghezza di cm. 12,000 o un peso di g. 12,000), ma **lo spazio unitario** compreso tra **X - 0,5** e **X + 0,5**.

$$\chi^2_{(g.d.l.)} = \sum_{i=1}^{n} \frac{\left(\left|f_i^{oss} - f_i^{att}\right| - 0{,}5\right)^2}{f_i^{att}}$$

Per effettuare questa correzione richiesta dal passaggio da una misura discreta a una continua, è necessario, prima dell'elevamento al quadrato

- **sottrarre 0,5 ad ogni scarto in valore assoluto**.

Si ottiene una riduzione del valore del chi quadrato rispetto alla procedura senza correzione. E' un **risultato cautelativo**, con il quale è meno probabile rifiutare l'ipotesi nulla.

Successivamente si vedranno altre correzioni, tra le numerose proposte in letteratura. Questa rimane la più diffusa, è riportata su quasi tutti i testi di statistica elementare o di base. A livelli più sofisticati è criticata, in quanto **troppo cautelativa: ridurrebbe eccessivamente il valore del chi quadrato**, e quindi renderebbe troppo difficile rifiutare l'ipotesi nulla. Nel capitolo successivo, si dirà che **determina un test poco potente**.

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 è in accordo con la distribuzione uniforme, cioè di 20 in ogni classe.

Risposta

- Senza correzione, il valore del χ^2 è

$$\chi^2_{(2)} = \frac{(15-20)^2}{20} + \frac{(21-20)^2}{20} + \frac{(24-20)^2}{20} = \frac{(5)^2}{20} + \frac{(1)^2}{20} + \frac{(4)^2}{20} = 1{,}25 + 0{,}05 + 0{,}8 = 2{,}1$$

- Con la prima correzione di Yates diventa

$$\chi^2_{(2)} = \frac{(4{,}5)^2}{20} + \frac{(1{,}5)^2}{20} + \frac{(4)^2}{20} = 1{,}0125 + 0{,}1125 + 0{,}8 = 1{,}925$$

- Con la seconda correzione diventa

$$\chi^2_{(2)} = \frac{(4{,}5)^2}{20} + \frac{(0{,}5)^2}{20} + \frac{(3{,}5)^2}{20} = 1{,}0125 + 0{,}0125 + 0{,}6125 = 1{,}6375$$

Al posto di 2,1 il valore corretto del χ^2 diminuisce a 1,925 con un metodo e 1,6375 con l'altro.

La correzione di Yates o per la continuità ha sempre l'effetto di ridurre il valore del χ^2.

Teoricamente, dovrebbe essere applicata sempre, anche a grandi campioni. Ma quando il numero di dati diventa alto, i suoi effetti diventano nulli o almeno trascurabili.

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. 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 χ^2 non abbia validità per rifiutare l'ipotesi nulla, quando il numero di osservazioni è troppo ridotto. Con un termine più tecnico si dice che il test è troppo poco potente, poiché 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" (la variabilità casuale) 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. Per risolvere ugualmente il problema di verificare l'accordo tra una distribuzione osservata e una distribuzione attesa sono stati proposti altre metodologie, come il test di **Kolmogorov-Smirnov**, il **metodo esatto di Fisher**, il **test G** che saranno presentati nei paragrafi successivi.

3.3. LE TABELLE DI CONTINGENZA 2 X 2 (FOURFOLD TABLES)

Quando si confrontano **le frequenze di risposte binarie in due campioni indipendenti**, è utile 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 e di 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 verificare se vivere in una zona ad alto inquinamento atmosferico incide sulla frequenza di malattie polmonari. A questo scopo, in una zona con tassi elevati d'inquinamento e in una 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 con malattie	Persone senza malattie	Totale
Zona ad alto inquinamento	**32 a**	**48 b**	80 n_1
Zona a basso inquinamento	**13 c**	**57 d**	70 n_2
Totale	45 n_3	105 n_4	150 **N**

Nei testi di statistica, non esiste uniformità su come costruire la tabella. La **convenzione** qui seguita è quella proposta da H. **Zeisel** nel 1947 (nel volume ***Say it with figures***, Harper & Row, New York; tradotto in italiano nel 1968, in ***Ditelo coi numeri***, Marsilio, Padova), che riporta

- le due modalità della variabile casuale sulle righe,
- le due modalità della variabile effetto sulle colonne.

La tabella riporta i risultati dell'analisi epidemiologica. Con questi dati, si può concordare con la teoria enunciata?

Il test chi quadrato utilizza **i casi effettivamente contati, non le frequenze relative o percentuali,** anche se su di esse vengono formulate le ipotesi.

Un'altra convezione, in questo caso generalmente seguita, suggerisce di indicare le frequenze riportate in ognuna delle 4 celle con le lettere minuscole **a**, **b**, **c**, **d**, (con la disposizione utilizzata nella tabella precedente). Il totale generale dei dati è indicato con la lettera maiuscola **N**.

Per comprendere la procedura del chi quadrato in tabelle 2 x 2, 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 (a + c cioè 32 + 13 = 45) rapportate al numero totale di persone osservate:

$$(a + c)/N \quad \text{cioè} \quad 45 / 150 = 0{,}3.$$

3- Considerando che i due campioni a confronto hanno un numero differente di osservazioni, sempre nel caso che l'ipotesi nulla sia vera,

- nel primo campione (che è composto da 80 individui) dovremmo aspettarci di trovare 24 persone (0,3 x 80 = 24) con malattie polmonari e
- nel secondo campione (composto da 70 individui) di trovarne 21 (0,3 x 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 con malattie	Persone senza malattie	Totale
Zona ad alto inquinamento	24 **a**	56 **b**	80 **n_1**
Zona a basso inquinamento	21 **c**	49 **d**	70 **n_2**
Totale	45 **n_3**	105 **n_4**	150 **N**

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 in 45;
- da uno qualsiasi di questi 3 dati già scritti deriva che la **casella d** può riportare solamente **49**, affinché i vari totali marginali possano coincidere.

Il ragionamento non varia, partendo da qualsiasi casella.

Queste considerazioni aiutano a comprendere perché **una tabella attesa 2 x 2 ha solamente 1 grado di libertà**. E' possibile anche un'altra spiegazione, che assume molta importanza in statistica e che sarà costante ripresa nell'analisi della varianza, per capire come

- i gdl sono legati alla quantità di informazione raccolta (numero di dati) e
- ogni informazione ricavata da essi ne abbassa il numero.

Per stimare l'atteso di ogni casella, noi abbiamo bisogno di 3 informazioni:

- il totale di riga,
- il totale di colonna,
- il totale generale (N).

Gli altri 2 totali parziali sono derivabili da questi tre e non rappresentano una reale informazione aggiuntiva. Ad esempio, se conosco che il totale generale è 150 e il totale della prima riga è 80, so anche che il totale della seconda riga è 70. Nello stesso modo, se so che il totale della prima colonna è 45, so anche che il totale della seconda colonna è 105.

Poiché i dati sono 4, ne deriva che i gradi di libertà è uno solo (gdl = 4 – 3 = 1).

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:

$$\chi^2_{(g.d.l.)} = \sum_{i=1}^{n} \frac{\left(f_i^{oss} - f_i^{att}\right)^2}{f_i^{att}}$$

dove:

- f_i^{oss} = frequenza osservata i-esima
- f_i^{att} = frequenza attesa i-esima

ed estendendo la sommatoria ($\sum$) ai dati di tutte quattro le caselle.

Con i dati dell'esempio

$$\chi^2_{(1)} = (32 - 24)^2 / 24 + (48 - 56)^2 / 56 + (13 - 21)^2 / 21 + (57 - 49)^2 / 49 =$$

$$= 2,666 + 1,143 + 3,048 + 1,306 = 8,163$$

si ottiene un valore del chi quadrato, con 1 gdl, uguale a 8,163

La tavola sinottica del $\chi^2_{(1)}$ riporta

- il valore critico di **3,84** alla probabilità α **= 0.05** e
- il valore critico di **6,64** alla probabilità α **= 0.01**.

Il valore calcolato (8,163) è superiore sia a quello della probabilità 0.05 che di quella 0.01; 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 $\alpha < 0.01$ 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 χ^2 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**

$$\chi^2_{(1)} = \frac{(a \cdot d - b \cdot c)^2 \cdot N}{n_1 \cdot n_2 \cdot n_3 \cdot n_4}$$

dove, con la simbologia e i valori riportati nella **tabella osservata**

	Persone con malattie	Persone senza malattie	Totale
Zona ad alto inquinamento	**32** a	**48** b	80 n_1
Zona a basso inquinamento	**13** c	**57** d	70 n_2
Totale	45 n_3	105 n_4	150 N

- **a, b, c, d** sono le frequenze osservate nei due campioni a confronto,
- $\mathbf{n_1, n_2, n_3, n_4}$ sono i totali marginali,
- N **è il totale generale di osservazioni.**

Il calcolo, con i dati sperimentali dell'esempio precedentemente utilizzato, fornisce

$$\chi^2_{(1)} = [\ (32 \text{ x } 57 - 48 \text{ x } 13)^2 \text{ x } 150\ ] / (80 \text{ x } 70 \text{ x } 45 \text{ x } 105)$$

$$\chi^2_{(1)} = (1824 - 624)^2 \text{ x } 150 / 26460000 = 1440000 \text{ x } 150 / 26460000 = 8{,}163$$

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 cavie. 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 ($\mathbf{H_0}$); oppure se i due tossici hanno probabilmente effetti letali differenti ($\mathbf{H_1}$).

Per meglio evidenziare i dati del problema, i conteggi devono essere riportati in una tabella a due entrate

	Cavie		
	Morte	Sopravv.	Totale
Tossico A	**22**	**48**	70
Tossico B	**24**	**26**	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

$$frequenza\ attesa = \frac{totale\ riga \times totale\ colonna}{totale\ generale}$$

che devono essere collocati nella tabella 2 x 2:

	Cavie		
	Morte	Sopravv.	Totale
Tossico A	**26,83**	**43,17**	70
Tossico B	**19,17**	**30,83**	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 $\mathbf{2 \times 2}$ 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 χ^2 mediante la formula generale:

$$\chi^2_{(1)} = \sum_{i=1}^{4} \frac{\left(f_i^{oss} - f_i^{att}\right)^2}{f_i^{att}}$$

Con i dati dell'esempio

$$\chi^2_{(1)} = \frac{(22-26,83)^2}{26,83} + \frac{(48-43,17)^2}{43,17} + \frac{(24-19,17)^2}{19,17} + \frac{(26-30,83)^2}{30,83}$$

$$\chi^2_{(1)} = \frac{23,33}{26,83} + \frac{23,33}{43,17} + \frac{23,33}{19,17} + \frac{23,33}{30,83} = 0,87 + 0,55 + 1,24 + 0,76 = 3,42$$

si ottiene un valore uguale a 3,42.
Oppure, ricordando la simbologia,

	X	x	Totale
Y	**a**	**b**	$\mathbf{n_1}$
y	**c**	**d**	$\mathbf{n_2}$
Totale	$\mathbf{n_3}$	$\mathbf{n_4}$	**N**

e la formula abbreviata

$$\chi^2_{(1)} = \frac{(a \cdot d - b \cdot c)^2 \cdot N}{n_1 \cdot n_2 \cdot n_3 \cdot n_4}$$

è possibile arrivare al risultato in modo molto più rapido

$$\chi^2_{(1)} = \frac{(22 \cdot 26 - 48 \cdot 24)^2 \cdot 120}{70 \cdot 50 \cdot 46 \cdot 74} = \frac{(572 - 1152)^2 \cdot 120}{11914000} = \frac{336400 \cdot 120}{11914000} = \frac{40368000}{11914000} = 3{,}39$$

ottenendo un chi quadrato uguale a 3,39.

I due risultati 3,42 e 3,39 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 χ^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.

3.4. CORREZIONI PER LA CONTINUITA' IN TABELLE 2 X 2: YATES E HABER

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 con grandi campioni (N > 100)**
- **non è attendibile per campioni eccessivamente ridotti (N < 30).**

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 proposta da F. **Yates** nel 1934 (vedi: ***Contingency tables involving small numbers and the χ^2 test*** pubblicato su **Journal of the Royal Statistical Society Suppl. 1**: 217-235), pubblicato ancora nel 1984 (con il titolo ***Tests of significance for 2 x 2 contingency tables***, su **Journal of the Royal Statistical Society**, Series A, vol. 147, pp. 426- 449 con molti articoli di commento), la formula del chi quadrato per tabelle 2 x 2 diviene:

$$\chi^2_{(1)} = \frac{\left(|a \cdot d - b \cdot c| - \frac{N}{2} \right)^2 \cdot N}{n_1 \cdot n_2 \cdot n_3 \cdot n_4}$$

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 .

Le proposte per la correzione più adeguata sono numerose, come si potrà vedere anche nel test G, dove sono fondate su principi simili. Il dibattito è stato ripreso negli ultimi decenni e nei programmi informatici viene utilizzata anche la **correzione proposta da M. Haber** nel 1980 (vedi l'articolo ***A comparison of some continuity corrections for the chi-squared test on 2 x 2 tables***, pubblicato su **Journal of the American Statistical Association**, vol. 75, pp. 510-515) e nel 1982 (***The continuity correction and statistical testing***, pubblicato su **International Statistical Review**, vol. 50, pp. 135-144), sviluppando concetti presentati già 40 anni prima, in particolare da W. G. **Cochran** nel 1942 (vedi l'articolo ***The*** χ^2 ***correction for continuity***, pubblicato su **Iowa State Coll. J. Sci.** Vol. 16, pp. 421-436) e nel 1952 (***The*** χ^2 ***test for goodness of fit***, pubblicato su **Annals of Mathematical Statistics** Vol. 23, pp. 315-345).

Il **metodo di Haber** è fondato sul **valore atteso** ($Att._{min}$) **minore**, identificato ovviamente all'incrocio tra il totale di riga ($R_{i\,min}$) minore e il totale di colonna ($C_{i\,min}$) minore,
con

$$Att._{\min} = \frac{R_{i\min} \cdot C_{i\min}}{N}$$

Lo scarto in modulo tra questa frequenza attesa e la corrispondente frequenza osservata (Oss.) permette di calcolare d

$$d = |Att._{\min} - Oss.|$$

Quando $Oss. \leq 2 \cdot Att._{\min}$, si definisce D l'arrotondamento a 0,5 inferiore a d

Quando $Oss. > 2 \cdot Att._{\min}$, si definisce D = d - 0,5.

Da D si ricava il χ^2 corretto (χ^2_c)

$$\chi_c^2 = \frac{N^3 \cdot D^2}{n_1 \cdot n_2 \cdot n_3 \cdot n_4}$$

con la solita simbologia,

ricordando

- che in una tabella 2 x 2 i quattro valori di D sono uguali, cioè gli scarti tra osservati ed attesi sono uguali in valore assoluto,
- quando $Oss. > 2 \cdot Att._{\min}$ e quindi D = d - 0,5, la correzione di Haber e quella di Yates forniscono lo stesso risultato.

ESEMPIO 1 (Correzione di Yates). 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 l'analisi dei dati e l'interpretazione dei risultati è sempre utile riportate i dati dell'esperimento in una tabella 2 x 2.

	Piante **Cresciute**	Piante **Non cresciute**	Totale
Diserbante A	**12**	**6**	18
Diserbante B	**26**	**9**	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 χ^2, ma con la correzione di Yates**. E' possibile ricorrere alla formula per il calcolo rapido, apportando la correzione per la continuità.

$$\chi^2_{(1)} = \frac{\left(\left| 12 \cdot 9 - 6 \cdot 26 \right| - \frac{53}{2} \right)^2 \cdot 53}{18 \cdot 35 \cdot 38 \cdot 15} = \frac{\left(\left| 108 - 156 \right| - 26{,}5 \right)^2 \cdot 53}{359100} = \frac{462{,}25 \cdot 53}{359100} = \frac{24499{,}25}{359100} = 0{,}0682$$

Il valore calcolato del chi quadrato con un grado di libertà (0,0682) è particolarmente basso, inferiore a quello tabulato alla probabilità $\alpha = 0.05$. Con una lettura della tavola dei valori critici più dettagliata, è possibile osservare che esso è inferiore addirittura a quello corrispondente alla probabilità $\alpha = 0.90$.
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.

ESEMPIO 2 (Confronto tra correzione di **Yates** e correzione di **Haber**). Con lo stesso esperimento dell'esempio precedente, si supponga che il risultato sia stato

	Piante **Cresciute**	Piante **Non cresciute**	Totale
Diserbante A	**15**	**9**	24
Diserbante B	**10**	**16**	26
Totale	25	25	50

I due diserbanti hanno effetti significativamente differenti sulla crescita delle pianticelle?

Risposta
Con il **metodo di Yates**

$$\chi^2_c = \frac{\left(\left|15 \cdot 16 - 9 \cdot 10\right| - \frac{50}{2}\right)^2 \cdot 50}{24 \cdot 26 \cdot 25 \cdot 25} = \frac{(150 - 25)^2 \cdot 50}{390.000} = \frac{781.250}{390.000} = 2{,}003$$

si ottiene un chi – quadrato corretto uguale a 2,003.

Con il metodo di **Haber**,

- dopo aver individuato il totale di riga minore (24) e il totale di colonna minore (25) si calcola il valore atteso minore ($Att._{\min}$), con i dati dell'esempio

$$Att._{\min} = \frac{24 \cdot 25}{50} = 12{,}0$$

si ottiene $Att._{\min} = 12{,}0$

- La differenza in valore assoluto dell' $Att._{\min}$ (12,0) con l'Osservato corrispondente (9)

è

$$d = \left|Att._{\min} - Oss.\right| = 12{,}0 - 9 = 3{,}0$$

uguale a 3,0.

- Poiché l'osservato è minore di 2 volte l'atteso

$$9 < 2 \cdot 12{,}0$$

con $d = 3{,}0$ si utilizza D = 2,5.

Infine con

$$\chi^2_c = \frac{N^3 \cdot D^2}{n_1 \cdot n_2 \cdot n_3 \cdot n_4} = \frac{50^3 \cdot 2{,}5^2}{24 \cdot 26 \cdot 25 \cdot 25} = \frac{125.000 \cdot 6{,}25}{390.000} = \frac{781.250}{390.000} = 2{,}003$$

si ottiene un chi–quadrato corretto uguale a 2,003.

In questa condizione il risultato è identico. Nell'altra condizione, ad alcuni autori appare più precisa la correzione di Haber.

Il metodo di Haber è più complesso e lungo di quello proposto da Yates. Ma l'uso dei programmi informatici permette ora di scegliere le misure più sofisticate, ritenute più corrette. In passato, era di estrema importanza anche la semplicità e la rapidità.

In un suo articolo 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***, Pubblicato sulla rivista **Philos. Mag.**, vol. 50 (5), pp. 157-175), Karl **Pearson** propone una formula fondata sul calcolo delle proporzioni

$$\chi^2 = N \cdot \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\left(\left| p_{ij} - p_{i.} \cdot p_{.j} \right| - \frac{1}{2N} \right)^2}{p_{i.} \cdot p_{.j}}$$

che ingloba anche la correzione per la continuità, dove

	X	X	Totale
Y	$\mathbf{p_{11}}$	$\mathbf{P_{12}}$	$\mathbf{P_{1.}}$
Y	$\mathbf{p_{21}}$	$\mathbf{P_{21}}$	$\mathbf{P_{2.}}$
Totale	$\mathbf{p_{.1}}$	$\mathbf{p_{.2}}$	**1**

- $\mathbf{p_{ij}}$ è la frequenza relativa osservata di una casella
- $\mathbf{p_i}$ e $\mathbf{p_j}$ è la somma di riga e di colonna delle frequenze relative, per cui il loro prodotto
- $\mathbf{p_i} \cdot \mathbf{p_j}$ 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	Non cresciute	Totale
Diserbante A	**0,226**	**0,113**	**0,339**
Diserbante B	**0,491**	**0,170**	**0,661**
Totale	**0,717**	**0,283**	**1**

e il calcolo del chi quadrato

$$\chi^2 = 53 \bullet \left[\frac{\left(\left| 0{,}226 - 0{,}339 \cdot 0{,}717 \right| - \frac{1}{2 \cdot 53} \right)^2}{0{,}339 \cdot 0{,}717} + \frac{\left(\left| 0{,}113 - 0{,}339 \cdot 0{,}283 \right| - \frac{1}{2 \cdot 53} \right)^2}{0{,}339 \cdot 0{,}283} + \right. $$
$$\left. + \frac{\left(\left| 0{,}491 - 0{,}661 \cdot 0{,}717 \right| - \frac{1}{2 \cdot 53} \right)^2}{0{,}661 \cdot 0{,}717} + \frac{\left(\left| 0{,}170 - 0{,}661 \cdot 0{,}283 \right| - \frac{1}{2 \cdot 53} \right)^2}{0{,}661 \cdot 0{,}283} \right]$$

$$\chi^2 = 53 \bullet \left[\frac{\left(\left| 0{,}226 - 0{,}243 \right| - 0{,}0094 \right)^2}{0{,}243} + \frac{\left(\left| 0{,}113 - 0{,}0959 \right| - 0{,}0094 \right)^2}{0{,}0959} + \frac{\left(\left| 0{,}491 - 0{,}474 \right| - 0{,}0094 \right)^2}{0{,}474} + \frac{\left(\left| 0{,}170 - 0{,}187 \right| - 0{,}0094 \right)^2}{0{,}187} \right]$$

$$\chi^2 = 53 \bullet \left(\frac{0{,}0076^2}{0{,}243} + \frac{0{,}0077^2}{0{,}096} + \frac{0{,}0076^2}{0{,}474} + \frac{0{,}0076^2}{0{,}187} \right)$$

$$\chi^2 = 53 \bullet (0{,}00024 + 0{,}00062 + 0{,}00012 + 0{,}00031) = 53 \bullet 0{,}00129 = 0{,}06837$$

stima un valore (0,06837) leggermente superiore a quello precedente (0,0682) a causa degli arrotondamenti e soprattutto dell'assenza della correzione per la continuità.

Nell'analisi statistica delle tabelle di contingenza, esistono **altre convenzioni**, oltre quelle ricordate sulla loro impostazione (la causa sulle righe, l'effetto sulle colonne) e sull'indicazione con lettere (a, b, c, d) delle frequenze riportate nelle 4 caselle.

A una **associazione** si attribuisce

- **segno positivo**, quando **le frequenze più alte** sono nelle due celle della **diagonale principale** (**a - d**),
- **segno negativo**, quando **le frequenze più alte** sono nella **diagonale secondaria** (**b – c**).

Naturalmente è una pura convenzione, non ha un valore intrinseco. Basterebbe, come è pienamente legittimo trattandosi di dicotomie, invertire la posizione delle righe o delle colonne per invertire anche il segno della associazione.

Come può essere stato dedotto in modo autonomo dallo studente osservando le caratteristiche di una tabella, si ha il massimo di equilibrio tra le modalità di una dicotomia quando a ciascuna è attribuito lo **stesso numero di dati** ($n_1 = n_2$; $n_3 = n_4$). Con un linguaggio più tecnico, si dice allora che la dicotomia ha **il massimo della varianza**; un altro modo per esprimere lo stesso concetto è che **è massima l'aspettativa di errore** nell'indovinare se un dato appartiene a una o all'altra categoria.

Per **evidenziare l'associazione**, sia visivamente nei dati, sia con l'analisi statistica, sono importanti altre due caratteristiche nella distribuzione dei dati:

- i totali marginali delle due dicotomie sono equilibrati,
- le frequenze sono maggiori lungo una diagonale.

Quando si confrontano le frequenze osservate con quelle attese, si ha

- **associazione positiva** quando la frequenza osservata è maggiore di quella attesa,
- **associazione negativa** quando quella osservata è minore di quella attesa.

Nella letteratura francese (vedi di J. P. **Benzécrì** del 1973 il testo ***L'analyse des donnés***, edito da Dunod, Paris) si parla di

- **attrazione** tra la modalità riportata nella riga e quella nella colonna che individuano la casella, quando la frequenza osservata è significativamente maggiore di quella attesa,
- **repulsione** quando la frequenza osservata è significativamente minore di quella 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 π di una popolazione è approssimativamente normale,**

- **con media campionaria p e deviazione standard della popolazione σ_π** (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 del**la 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**,

$$Z = \frac{p_1 - p_2}{\sqrt{p^*(1-p^*) \cdot (\frac{1}{n_1} + \frac{1}{n_2})}}$$

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 **π**. 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 **π**, 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 $(p_1 - p_2) - \pi$**.

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 x n_2)**, esistendo la relazione

$$\frac{1}{n_1} + \frac{1}{n_2} = \frac{n_1 + n_2}{n_1 x n_2}$$

Pertanto, **la formula per il calcolo del valore di Z nel confronto tra due proporzioni può essere scritta** pure come

$$Z = \frac{|p_1 - p_2| - \pi}{\sqrt{p^*(1-p^*) \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}}$$

dove

- $\mathbf{p_1}$ è la proporzione di casi in esame nel campione **1** con numerosità $\mathbf{n_1}$,
- $\mathbf{p_2}$ è la proporzione di casi in esame nel campione **2** con numerosità $\mathbf{n_2}$,
- $\boldsymbol{\pi}$ 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 confronto tra due proporzioni, la correzione di Yates consiste nel diminuire la differenza tra le due proporzioni campionarie di una quantità uguale a

$$\frac{1}{2}\ \mathbf{(1/n_1 + 1/n_2)}$$

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

$$\mathbf{H_0: \pi_1 = \pi_2}$$

con la correzione per la continuità di **Yates** la formula estesa del test **Z** diviene

$$Z = \frac{|p_1 - p_2| - \frac{1}{2}(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{p^*(1-p^*) \cdot (\frac{1}{n_1} + \frac{1}{n_2})}}$$

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 χ^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

$$\mathbf{H_1: \pi_1 \neq \pi_2}$$

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 χ^2:**

$$\mathbf{H_1: \pi_1 \neq \pi_2}$$

2 - se un gruppo ha una frequenza relativa significativamente maggiore (oppure minore): è un **test unilaterale o a una coda:** si confrontano

$$\mathbf{H_0: \pi_1 \leq \pi_2} \quad \text{contro} \quad \mathbf{H_1: \pi_1 > \pi_2}$$

oppure

$$\mathbf{H_0: \pi_1 \geq \pi_2} \quad \text{contro} \quad \mathbf{H_1: \pi_1 < \pi_2}$$

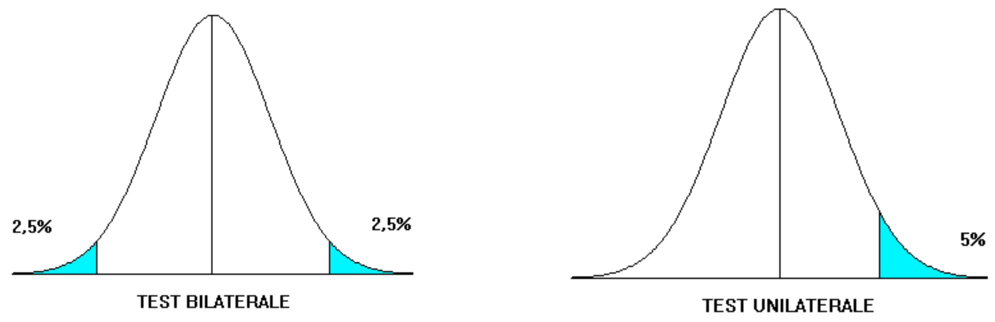
In ognuno di questi ultimi 2 casi ad una coda, viene a priori rifiutata come non accettabile od illogica la possibilità alternativa a quella proposta.

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; $\mathbf{H_0: \pi_1 = \pi_2}$

L'ipotesi alternativa è che le due proporzioni siano diverse; $\mathbf{H_1: \pi_1 \neq \pi_2}$

I due campioni sono di dimensioni sufficientemente grandi per poter utilizzare la distribuzione **Z**, mediante

$$Z = \frac{p_1 - p_2}{\sqrt{p*(1-p*)\cdot(\frac{1}{n_1}+\frac{1}{n_2})}}$$

dove

$p_1 = 0{,}20$; $p_2 = 0{,}26$; $n_1 = 200$; $n_2 = 150$;

$m_1 = (0{,}20 \text{ x } 200) = 40$; $m_2 = (0{,}26 \text{ x } 150) = 39$; $p^* = \frac{40+39}{200+150} = 0{,}226$

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. Un gruppo di ricercatori ha apportato miglioramenti; dopo alcune prove essi 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 o minore di quello del primo aumentato del 2%; $\mathbf{H_0: \pi_2 \leq \pi_1 + 0{,}02}$

L'ipotesi alternativa è che le probabilità di ricattura del secondo metodo sia significativamente superiore a quella del primo aumentata del 2%; $\mathbf{H_1: \pi_2 > \pi_1 + 0{,}02}$

I due campioni sono di dimensioni sufficientemente grandi per poter utilizzare la distribuzione **Z**, mediante

$$Z = \frac{|p_1 - p_2| - \pi}{\sqrt{p*(1-p*) \cdot \frac{n_1 + n_2}{n_1 \times n_2}}}$$

dove

$p_1 = 0,15$; $n_1 = 400$; $m_1 = 60$; $p_2 = 0,25$; $n_2 = 100$; $m_2 = 25$;

e

$$\pi = 0,02; \quad p* = \frac{60+25}{400+100} = 0,17$$

Pertanto, il valore di **Z** è

$$Z = \frac{|0,25 - 0,15| - 0,02}{\sqrt{0,17 \cdot 0,83 \cdot \left(\frac{400+100}{400x100}\right)}} = \frac{0,08}{\sqrt{0,1411 \cdot 0,0125}} = \frac{0,08}{0,04199} = 1,90$$

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 $\mathbf{H_0}$ 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 $\mathbf{H_0 : \pi_1 \geq \pi_2}$
ed ipotesi alternativa $\mathbf{H_1 : \pi_1 < \pi_2}$
il calcolo finale per il test **Z** sarebbe stato

$$\mathbf{Z} = \frac{0,10}{0,04199} = 2,38$$

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: $\pi_1 = \pi_2$**

e **l'ipotesi alternativa è H_1: $\pi_1 \neq \pi_2$**

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

$$Z = \frac{|p_1 - p_2| - \frac{1}{2}(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{p^*(1-p^*)\cdot(\frac{1}{n_1} + \frac{1}{n_2})}}$$

dove

$p_1 = (4/80) = 0{,}05$; $n_1 = 80$; $m_1 = 4$;

$p_2 = (10/50) = 0{,}2$; $n_2 = 50$; $m_2 = 10$; $p^* = \frac{4+10}{80+50} = 0{,}108$

Sostituendo ai simboli i dati del problema, si ottiene un valore di **Z**

$$Z = \frac{|0{,}05 - 0{,}20| - \frac{1}{2}(\frac{1}{80} + \frac{1}{50})}{\sqrt{\left[0{,}108 \cdot 0{,}892 \cdot (\frac{1}{80} + \frac{1}{50})\right]}} = \frac{|0{,}15| - 0{,}01625}{\sqrt{[0{,}096 x 0{,}0325]}} = \frac{0{,}13375}{0{,}05596} = 2{,}39$$

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 prudenziale.

3.6. CONFRONTO TRA TEST χ^2 PER TABELLE 2 X 2 E TEST Z, SENZA E CON LE 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 χ^2 forniscono la stessa risposta**. Tra test **t** e test χ^2 **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 χ^2 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

$$\mathbf{Z^2 \cong \chi^2_{(1)}}$$

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 χ^2 con **n** gradi di libertà, secondo la relazione

$$\sum_{i=1}^{n} Z_i^2 = \chi^2_{(n)}$$

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 χ^2 sono possibili solo test bilaterali).
L'ipotesi nulla è **H_0: $\pi_1 = \pi_2$**
e l'ipotesi alternativa è **H_1: $\pi_1 \neq \pi_2$**
Per applicare il test χ^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	**31**	**154**	185
Lago 2	**39**	**110**	149
Totale	70	264	**334**

E' un campione di grandi dimensioni e conviene utilizzare la formula generale per il calcolo rapido,

$$\chi^2_{(1)} = \frac{(a \cdot d - b \cdot c)^2 \cdot N}{n_1 \cdot n_2 \cdot n_3 \cdot n_4}$$

senza correzione per la continuità

$$\chi^2_{(1)} = \frac{(31 \cdot 110 - 154 \cdot 39)^2 \cdot 334}{185 \cdot 149 \cdot 70 \cdot 264} = \frac{(3410 - 6006)^2 \cdot 334}{509401200} = \frac{6739216 \cdot 334}{509401200} = \frac{2250898144}{509401200} = 4{,}41$$

dal quale risulta $\chi^2_{(1)} =$ 4,41.

Nella tabella sinottica del $\chi^2_{(1)}$ ad un valore di 4,41 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

$$Z = \frac{p_1 - p_2}{\sqrt{p^*(1-p^*)\cdot(\frac{1}{n_1}+\frac{1}{n_2})}}$$

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 χ^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

$$\mathbf{Z^2 \cong \chi^2_{(1)}}$$

anche con i dati di questo esempio **è possibile osservare che** il valore del test **Z** (uguale a - 2,10) elevato al quadrato ($-2{,}10^2 = 4{,}41$) è esattamente uguale al valore calcolato con il test $\chi^2_{(1)}$ (uguale a 4,41)

- $2{,}10^2 \cong 4{,}41$

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 $\mathbf{H_0: \pi_1 = \pi_2}$

e l'ipotesi alternativa è $\mathbf{H_1: \pi_1 \neq \pi_2}$

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

$$\chi^2_{(1)} = \frac{\left(|a \cdot d - b \cdot c| - \frac{N}{2}\right)^2 \cdot N}{n_1 \cdot n_2 \cdot n_3 \cdot n_4}$$

Con i dati del problema,

$$\chi^2_{(1)} = \frac{\left(|6 \cdot 19 - 34 \cdot 10| - \frac{69}{2}\right)^2 \cdot 69}{40 \cdot 29 \cdot 16 \cdot 53} = \frac{(|114 - 340| - 34{,}5)^2 \cdot 69}{983680} = \frac{36672{,}25 \cdot 69}{983680} = \frac{2530385{,}25}{983680} = 2{,}57$$

fornisce un $\chi^2_{(1)} = 2{,}57$.

Nella tabella sinottica del $\chi^2_{(1)}$ al valore 2,57 **corrisponde una probabilità α 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**

$$Z = \frac{|p_1 - p_2| - \frac{1}{2}(\frac{1}{n_1} + \frac{1}{n_2})}{\sqrt{p^*(1-p^*) \cdot (\frac{1}{n_1} + \frac{1}{n_2})}}$$

dove:

$p_1 = 0{,}150$; $p_2 = 0{,}345$; $n_1 = 40$; $n_2 = 29$; $m_1 = 6$; $m_2 = 10$;

e

$$p^* = \frac{6+10}{40+29} = 0{,}232$$

si stima un valore di **Z**

$$\mathbf{Z} = \frac{\left|0{,}150 - 0{,}345\right| - \frac{1}{2}(\frac{1}{40} + \frac{1}{29})}{\sqrt{0{,}23 \cdot 0{,}77 \cdot \left(\frac{1}{40} + \frac{1}{29}\right)}} = \frac{0{,}195 - 0{,}03}{\sqrt{0{,}1771 \cdot 0{,}0594}} = \frac{0{,}165}{0{,}1026} = 1{,}608$$

che risulta uguale a 1,608.

Nella tabella della distribuzione normale standardizzata ad una coda, Z = 1,61 (arrotondamento di 1,608) esclude una densità di probabilità $\alpha = 0.055$; per un test bilaterale, la probabilità è $\alpha = 0.11$ (11,0%).

E' semplice osservare che 1.608 è approssimativamente uguale a 2.58: il test χ^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 E 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 μ, quando sia nota la varianza σ^2 della popolazione, attraverso la relazione**

$$Z = \frac{\bar{x} - \mu}{\sqrt{\sigma^2}}$$

poiché **la varianza di una proporzione è totalmente definita dal suo valore medio p e dal numero totale di osservazioni** essendo $\sigma^2 = n \cdot p \cdot (1-p)$

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 · 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 x 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 = ¾.
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 χ^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 n·p 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)

$$Pi = C_i^n \cdot p^i \cdot q^{n-i}$$

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

$\mathbf{P_{8+9}} = 0.00354 + 0.00026 = \mathbf{0.0038}$ (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 χ^2 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. E' lo stesso principio per cui, nel caso di una sola proporzione e un campione piccolo, si ricorre alla distribuzione binomiale.

Per passare da indicazioni di principio a raccomandazioni pratiche, per la 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 χ^2 quando il valore di $\mathbf{n \cdot p}$ oppure quello di $\mathbf{n \cdot (1-p)}$ 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***, sulla rivista inglese **Journal of the Royal Statistical Society**, vol. 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	**a**	**b**	$\mathbf{n_1}$ = a + b
Campione y	**c**	**d**	$\mathbf{n_2}$ = c + d
Totale	$\mathbf{n_3}$ = a + c	$\mathbf{n_4}$ = a + d	**N =** a+b+c+d

con la **distribuzione ipergeometrica** la probabilità $\mathbf{P_i}$ è calcolata con la formula

$$P_i = \frac{C^a_{a+c} \cdot C^b_{b+d}}{C^{a+b}_N} = \frac{\frac{(a+c)!}{a! \cdot c!} \cdot \frac{(b+d)!}{b! \cdot d!}}{\frac{N!}{(a+b)! \cdot (c+d)!}} = \frac{(a+b)! \cdot (c+d)! \cdot (a+c)! \cdot (b+d)!}{a! \cdot b! \cdot c! \cdot d! \cdot N!}$$

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	Cavie		
	Sopravvissute	Morte	Totale
Pesticida A	**7**	**1**	8
Pesticida B	**3**	**6**	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

$$Pi = \frac{8! \cdot 9! \cdot 10! \cdot 7!}{7! \cdot 1! \cdot 3! \cdot 6! \cdot 17!} = 0.03455$$

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	Cavie		
	Sopravvissute	Morte	Totale
Pesticida A	**8**	**0**	8
Pesticida B	**2**	**7**	9
Totale	10	7	17

La probabilità di avere per caso questa possibile risposta, mantenendo fissi i totali, è data da

$$Pi = \frac{8! \cdot 9! \cdot 10! \cdot 7!}{8! \cdot 0! \cdot 2! \cdot 7! \cdot 17!} = 0.00185$$

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)

8	0
2	7

(2)

7	1
3	6

(3)

6	2
4	5

(4)

5	3
5	4

(5)

4	4
6	3

(6)

3	5
7	2

(7)

2	6
8	1

(8)

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 χ^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 di K. D. Tocher** del 1950 (vedi l'articolo ***Extension of the Neyman-Pearson theory of tests to discontinuous variates*****,** pubblicato da **Biometrika** Vol. 37, pp. 130-144). Queste modifiche, citate in alcuni testi specialistici, in questo manuale sono tralasciate, perché nella ricerca applicata 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 χ^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

$$H_0: \pi_1 = \pi_2 = \pi_3 = \cdots = \pi_M$$

e l'ipotesi alternativa è

$$H_1 = \text{almeno una delle } \pi \text{ è diversa dalle altre.}$$

Il caso più semplice di tabelle M x N è la tabella di contingenza $\mathbf{2 \times N}$, per **risposte dicotomiche di N gruppi a confronto. Essa ha N - 1 gradi di libertà,** derivati dalla formula generale

$$(N\text{-}1) \times (2\text{-}1)$$

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 χ^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

Animali	Pesticida					
	A	B	C	D	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 π di animali morti $H_0: \pi_1 = \pi_2 = \pi_3 = \cdots = \pi_M$

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

Animali	Pesticida					
	A	B	C	D	E	Totale
Morti	**8,33**	**6,67**	**14,17**	**13,75**	**7,08**	50,0
Sopravvissuti	**11,67**	**9,33**	**19,83**	**19,25**	**9,92**	70,0
Totale	20,0	16,0	34,0	33,0	17,0	120,0

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

$$\chi^2_{(g.d.l.)} = \sum_{i=1}^{M \cdot N} \frac{\left(f_i^{oss} - f_i^{att}\right)^2}{f_i^{att}}$$

si calcola il valore del $\chi^2_{(4)}$, estendendo la sommatoria a tutte 10 le caselle

$$\chi^2_{(4)} = \frac{(8-833)^2}{8,33} + \frac{(10-6,67)^2}{6,67} + \cdots + \frac{(10-9,92)^2}{9,92} = 3,9266$$

Il valore del $\chi^2_{(4)}$ 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 χ^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**

$$\chi^2_{g.d.l.} = \frac{C \cdot 100}{\overline{p} \cdot (1-\overline{p})}$$

con C uguale a

$$C = \sum_{i=1}^{k} p_i \cdot n_i - \overline{p} \cdot \sum_{i=1}^{k} n_i$$

e dove

- **k** = numeri di gruppi a confronto,
- $\mathbf{p_i}$ = frequenza percentuale del carattere in esame nel gruppo i,
- $\mathbf{n_i}$ = frequenza assoluta del carattere in esame nel gruppo o campione i,
- **N** = numero totale di osservazioni,
- $\overline{p}$ = 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					
Animali	A	B	C	D	E	Totale
Morti n_i	8	10	14	11	7	50
P_i in %	**40,0**	**62,5**	**41,2**	**33,3**	**41,2**	**41,66**
Sopravvissuti	12	6	20	22	10	70
Totale	20	16	34	33	17	120

e in riferimento a questi dati sono calcolati i parametri richiesti dalla formula

$$\sum_{i=1}^{5} p_i \cdot n_i = 40{,}0x8 + 62{,}5x10 + 41{,}2x14 + 33{,}3x11 + 41{,}2x7 = 2176{,}5$$

$$\overline{p} = 41{,}66; \quad \text{N} = 120$$

$$\overline{p} \cdot \sum_{I=1}^{5} n_i = 41{,}7x50 = 2085$$

C = 2176,5 - 2085 = 91,5

Da essi si stima un valore del $\chi^2_{(4)}$

$$\chi^2_{(4)} = \frac{91{,}5x100}{41{,}66x58{,}34} = 3{,}765$$

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.

<u>Nel caso più generale di una tabella di contingenza $\mathbf{M \times N}$,</u> il χ^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 x N, i gradi di libertà sono

[illegible]

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 $\mathbf{M \times N}$ 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 χ^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	**12**	**8**	**21**	**5**	**4**	50
Coltivazione II	**15**	**10**	**5**	**20**	**8**	58
Coltivazione III	**9**	**6**	**10**	**17**	**11**	53
Coltivazione IV	**23**	**12**	**12**	**31**	**17**	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	**11,5**	**7,0**	**9,4**	**14,3**	**7,8**	50
Coltivazione II	**13,4**	**8,2**	**10,9**	**16,5**	**9,1**	58
Coltivazione III	**12,2**	**7,5**	**9,9**	**15,1**	**8,3**	53
Coltivazione IV	**21,9**	**13,3**	**17,8**	**27,1**	**14,8**	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

$$\chi^2_{(12)} = \frac{(12-11,5)^2}{11,5} + \frac{(8-7)^2}{7} + \cdots + \frac{(17-14,8)^2}{14,8} = 32,251$$

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)^2 / ATTESO

	Specie A	Specie B	Specie C	Specie D	Specie E	Totale
Coltivazione I	**0,022**	**0,143**	**14,315**	**6,048**	**1,851**	22,379
Coltivazione II	**0,191**	**0,395**	**3,194**	**0,742**	**0,133**	4,655
Coltivazione III	**0,839**	**0,300**	**0,001**	**0,239**	**0,878**	2,257
Coltivazione IV	**0,055**	**0,127**	**1,890**	**0,561**	**0,327**	2,960
Totale	1,107	0,965	19,400	7,590	3,189	**32,251**

Alla probabilità $\alpha = 0.05$ con 12 g.d.l., la tabella del χ^2 dà un valore critico uguale a 21.03.

Il valore del $\chi^2_{(12)}$ calcolato (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.

<u>**La scomposizione dei gradi di libertà di queste tabelle complesse**</u> è un altro modo che permette di avere informazioni più dettagliate, sugli effetti di ogni particolare gruppo di dati.

La proprietà additiva del χ^2 e dei relativi gradi di libertà consente la scomposizione di una tabella $\mathbf{M \times N}$ in tanti test $\mathbf{2 \times 2}$, 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 $\mathbf{3 \times 3}$ 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 χ^2 che ha 4 gradi di libertà. Se risulta significativo, è utile scomporre questa valutazione globale, per conoscere quali confronti singoli $\mathbf{2 \times 2}$ 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 x 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 χ^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 χ^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 $\mathbf{G^2}$, che utilizza i logaritmi naturali.

Rispetto al test χ^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 χ^2 e a volte più potente, nel caso di frequenze piccole**.

Nel 1978 K. **Larntz** (nell'articolo ***Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics***, pubblicato su **Journal of the American Statistical Association** 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 χ^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** nel 1976 da D. A. **Williams** (con l'articolo ***Improved likelihood ratio test for complete contigency 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 χ^2corrispondente.

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

$$\mathbf{G = 2 \ln L}$$

(dove ln indica il logaritmo naturale).

<u>**Nel caso di grandi campioni è bene approssimata dalla distribuzione χ^2 con gli stessi gdl**</u> (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

$$L = \frac{C_N^{n1} \cdot p_o^{n1} \cdot q_o^{n2}}{C_N^{n1} \cdot p_a^{n1} \cdot q_a^{n2}} = \left(\frac{p_o}{p_a} \right)^{n1} \cdot \left(\frac{q_o}{q_a} \right)^{n2}$$

Poiché

n1 osservato = Np_o e n2 osservato = Nq_o

e similmente

n1 atteso = Np_a e n2 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

$$\ln L = n_o^1 \ln\left(\frac{n_o^1}{n_a^1} \right) + n_a^2 \ln\left(\frac{n_o^2}{n_a^2} \right)$$

Per la significatività, **il valore di G non ha una sua tabella, ma utilizza la stessa distribuzione dei valori critici del χ^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 χ^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

$$\textbf{G o likelihood ratio} = 2\sum_{i=1}^{k}\left[Oss.\ln(Oss./Att.)\right] = 2\left[\sum_{i=1}^{k}Oss.(\ln Oss.) - \sum_{i=1}^{k}Oss.(\ln Att.)\right]$$

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**

$$q = 1 + \frac{k^2 - 1}{6N\nu}$$

dove

k = il numero di gruppi;

ν = 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 **ν = 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 χ^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 χ^2 che ha 3 gdl

$$\chi^2_{(3)} = \frac{(55-40{,}75)^2}{40{,}75} + \frac{(28-40{,}75)^2}{40{,}75} + \frac{(37-40{,}75)^2}{40{,}75} + \frac{(43-40{,}75)^2}{40{,}75}$$

$$\chi^2_{(3)} = \frac{14{,}25^2}{40{,}75} + \frac{12{,}75^2}{40{,}75} + \frac{3{,}75^2}{40{,}75} + \frac{2{,}25^2}{40{,}75} = 4{,}983 + 3{,}989 + 0{,}345 + 0{,}124 = 9{,}441$$

e risulta uguale a 9,441; con la correzione di Yates il valore diminuisce a

$$\chi^2_{(3)} = \frac{13{,}75^2}{40{,}75} + \frac{12{,}75^2}{40{,}75} + \frac{3{,}75^2}{40{,}75} + \frac{2{,}75^2}{40{,}75} = 4{,}640 + 3{,}989 + 0{,}345 + 0{,}186 = 9{,}160$$

9,160.

Usando il metodo G

$$G = 2\left(55 \ln\frac{55}{40{,}75} + 28 \ln\frac{28}{40{,}75} + 37 \ln\frac{37}{40{,}75} + 43 \ln\frac{43}{40{,}75} \right)$$

$$G = 2(16,493 - 10,507 - 3,572 + 2,311) = 2x4,725 = 9,450$$

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

$$q = 1 + \frac{4+1}{6x163} = 1 + \frac{5}{978} = 1,0051$$

risulta uguale a 1,0051

e quindi il valore di $\mathbf{G_{adj}}$

$$G_{adj} = \frac{9,450}{1,0051} = 9,402$$

risulta uguale a 9,402.

In questo esempio, il valore di **G** risulta leggermente superiore a quello del corrispondente χ^2. Poiché il valore critico del χ^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 si 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 χ^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 Tabelle 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	$\mathbf{n_1}$
Camp. y	c	d	$\mathbf{n_2}$
Totale	$\mathbf{n_3}$	$\mathbf{n_4}$	N

il valore di **G o Log-likelihood ratio** è dato da

$$G = 2\left[(a\ln a + b\ln b + c\ln c + d\ln d) - (n_1\ln n_1 + n_2\ln n_2 + n_3\ln n_3 + n_4\ln n_4) + (N\ln N)\right]$$

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 $\mathbf{G_{adj}}$ dato dal rapporto

$$\mathbf{G_{adj} = G/q}$$

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 χ^2**.

Consiste nell'aggiungere o togliere 0,5 ad ognuna delle 4 frequenze osservate, sulla base del confronto dei prodotti delle due diagonali: ***a*** **x** ***d*** contro ***b*** **x** ***c***.

Se il prodotto ***a*** **x** ***d*** è maggiore di ***b*** **x** ***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* x *d*** è minore di ***b* x *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 χ^2 senza (D) e con la correzione (E) per campioni non grandi.

Risposta.

A) **<u>Con il metodo G</u>**, 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

$$G = 2 \cdot (333,33 - 803,14 + 471,75) = 2x1,94 = 3,88$$

un valore di G uguale a 3,88.

B) Con la **<u>correzione di Williams</u>**

$$q = 1 + \frac{\left(\frac{102}{53} + \frac{102}{49} - 1\right) \cdot \left(\frac{102}{58} + \frac{102}{44} - 1\right)}{6x102} = 1 + \frac{(1{,}925 + 2{,}082 - 1) \cdot (1{,}759 + 2{,}318 - 1)}{612} = 1{,}015$$

$$G_{adj} = \frac{3{,}88}{1{,}015} = 3{,}82$$

si ottiene un valore di **G** aggiustato uguale a 3,82.

C) Con la **<u>correzione di Mantel-Haenszel</u>**,

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

$$G_{adj} = 2 \cdot (332{,}92 - 803{,}14 + 471{,}75) = 2x153 = 3{,}06$$

si ottiene un valore di **G** aggiustato uguale a 3,06.

D) **<u>Il χ^2 con la formula abbreviata per grandi campioni</u>**

$$\chi^2_{(1)} = \frac{(35 \cdot 26 - 18 \cdot 23)^2 \cdot 102}{53 \cdot 49 \cdot 58 \cdot 44} = \frac{(910 - 414)^2 \cdot 102}{6627544} = \frac{25093632}{6627544} = 3{,}78$$

da un valore uguale a 3,78 con 1 gdl.

E) **<u>Il χ^2 con la correzione di Yates</u>** per campioni non grandi

$$\chi^2_{(1)} = \frac{\left(|35 \cdot 26 - 18 \cdot 23| - \frac{102}{2}\right)^2 \cdot 102}{53 \cdot 49 \cdot 58 \cdot 44} = \frac{(|910 - 414| - 51)^2 \cdot 102}{6627544} = \frac{20198550}{6627544} = 3{,}05$$

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 = ∑ f ln f di ogni casella,

II = ∑ f ln f di ogni totale marginale sia di riga che di colonna,

III = ∑ N ln N, con **N** uguale al totale generale del numero di osservazioni,

si ottiene il valore di **G** con

$$G = 2 \cdot (I - II + III)$$

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	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

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			
Zona	Bassi	Medi	Alti	Totale
A	**5**	**2**	**1**	8
B	**3**	**10**	**1**	14
C	**5**	**2**	**4**	11
Totale	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 χ^2 ed il valore del **likelihood ratio**, confrontando i risultati.)

Risposta.

L'ipotesi nulla $\mathbf{H_0}$ 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 $\mathbf{H_1}$ 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

Zona	Punteggi			
	Bassi	Medi	Alti	Totale
A	**3,15**	**3,39**	**1,46**	8
B	**5,52**	**5,94**	**2,54**	14
C	**4,33**	**4,67**	**2,00**	11
Totale	13	14	6	33

Per il calcolo del χ^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

Frequenza attesa in ogni casella = Totale di riga x Totale di colonna / Totale generale

FREQUENZE OSSERVATE - FREQUENZE ATTESE

Zona	Punteggi			
	Bassi	Medi	Alti	Totale
A	**+1,85**	**-1,39**	**-0,46**	0
B	**-2,52**	**+4,06**	**-1,54**	0
C	**+0,67**	**-2,67**	**+2,00**	0
Totale	0	0	0	0

Il passo successivo è la stima del contributo di ogni casella al valore complessivo del χ^2. Ogni casella indica la differenza tra frequenza osservata e frequenza attesa, in rapporto alla frequenza attesa.

(FREQUENZE OSSERVATE - FREQUENZE ATTESE)^2 / FREQUENZE ATTESE

Zona	Punteggi			
	Bassi	Medi	Alti	Totale
A	**1,087**	**0,570**	**0,145**	1,802
B	**1,150**	**2,775**	**0,934**	4,859
C	**0,104**	**1,527**	**2,000**	3,631
Totale	2,341	4,872	3,079	**10,292**

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 χ^2deriva 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

$$G = 2\begin{bmatrix} (5\ln 5 + 2\ln 2 + 1\ln 1 + 3\ln 3 + 10\ln 10 + 1\ln 1 + 5\ln 5 + 2\ln 2 + 4\ln 4) - \\ -(8\ln 8 + 14\ln 14 + 11\ln 11 + 13\ln 13 + 14\ln 14 + 6\ln 6) + \\ +(33\ln 33) \end{bmatrix}$$

G = 2 x{[(5 x 1,609) + (2 x 0,693) + (1 x 0) + (3 x 1,099) + (10 x 2,303) + (1 x 0) + (5 x 1,609) +

+ (2 x 0,693) + (4 x 1,386)] -

- [(8 x 2,079) + (14 x 2,639) + (11 x 2,398) + (13 x 2,565) + (14 x 2,639) + 6 x 1,792)]+

+ [(33 x 3,497)]}

G = 2 x [(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 x (50,733 - 160,999 +115,409) = 2 x 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 χ^2, per gli stessi gradi di libertà.

La **correzione q di Williams** è

$$q = 1 + \frac{(3+1)\cdot(3+1)}{6x33} = 1 + \frac{16}{198} = 1{,}081$$

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 α = 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 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 (nell'articolo ***The use of transformations and maximum likelihood in the analysis of quantal experiments involving two tratments***, 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**

$$d = \frac{p_1 - p_2}{\overline{p} \cdot (1 - \overline{p})}$$

dove

p_1 e **p_2** sono rispettivamente le frequenze relative dei due campioni a confronto,

$\overline{p}$ 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 x 2**, con misure di associazione la **differenza standardizzata** diventa

$$d_k = \frac{p_{k1} - p_{k2}}{\overline{p}_k \cdot (1 - \overline{p}_k)}$$

e **il suo errore standard e.s.(d_k)** è

$$\text{e.s.}(d_k) = \sqrt{\frac{1}{\overline{p}_k \cdot (1 - \overline{p}_k)} \cdot \frac{n_k}{n_{k1} \cdot n_{k2}}}$$

dal quale si ricava

$$w_k = \frac{\overline{p}_k \cdot (1 - \overline{p}_k) \cdot n_{k1} \cdot n_{k2}}{n_k}$$

Derivato dal metodo che, per grandi campioni, utilizza la distribuzione **z** e sulla base della relazione

$$\chi^2_{(n)} = \sum_{i=1}^{n} z_i^2$$

il **metodo del χ^2 proposto da W. G. Cochran** nel 1954 (nell'articolo ***Some methods of strengthening the common χ^2 tests***, pubblicato da **Biometrics**, vol. 10, pp. 417-451)

nel caso di **due campioni** è

$$\chi^2_{(1)} = \frac{(p_1 - p_2)^2}{\overline{p} \cdot (1-\overline{p}) \cdot (\frac{1}{n_1} + \frac{1}{n_2})}$$

e nel caso di **<u>k campioni</u>** diventa

$$\chi^2_{(k)} = \sum \frac{(p_{k1} - p_{k2})^2}{\overline{p}_k \cdot (1-\overline{p}_k) \cdot (\frac{1}{n_{k1}} + \frac{1}{n_{k2}})}$$

Lo **stesso risultato** è ottenuto con la formula abbreviata

$$\chi^2_{(k)} = w_k \cdot d_k^2$$

dove $\mathbf{w_k}$ e $\mathbf{d_k}$ 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

	Zona A		Zona B	
Rilevazione	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	$p_{kA} - p_{kB}$	$\overline{p}_k$	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) $$p_{kA} - p_{kB} = 0{,}289 - 0{,}351 = -0{,}062$$

2) $$\overline{p}_k = \frac{(0{,}289 \cdot 152) + (0{,}351 \cdot 174)}{152 + 174} = \frac{54 + 61}{326} = 0{,}322$$

3) $$d_k = \frac{0{,}289 - 0{,}351}{0{,}322 \cdot 0{,}678} = \frac{-0{,}062}{0{,}218} = -0{,}284$$

4) $$n_k = 326$$

5) $$w_k = \frac{0{,}322 \cdot 0{,}678 \cdot 152 \cdot 174}{326} = \frac{0{,}218 \cdot 26448}{326} = 17{,}69$$

6) $$w_k d_k = 17{,}69 \cdot (-0{,}284) = -5{,}02$$

7) $$w_k d_k^2 = 17{,}69 \cdot (-0{,}284)^2 = 1{,}43$$

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

$$\chi^2_{\text{om}og.} = \sum_{k=i}^{3} w_k d_k^2 - \frac{\left(\sum_{k=1}^{3} w_k d_k\right)^2}{\sum_{k=1}^{3} w_k}$$

IV - Con i dati dell'esempio

$$\chi^2_{omog.} = 3{,}75 - \frac{2{,}51^2}{43{,}5} = 3{,}75 - 0{,}145 = 3{,}605$$

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 χ^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

$$\chi^2_{assoc.} = \left(\frac{d}{e.s.(d)}\right)^2 = \left(\frac{0{,}058}{0{,}152}\right)^2 = 0{,}146$$

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 χ^2 proposto da N. Mantel e W. Haenszel** nel 1959 (con l'articolo ***Statistical aspects of the analysis of data from retrospective studies of disease*** pubblicato su **J. Matl. 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 **Journal of the American Statistical Association**, vol. 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

$$\chi^2_{(k)} = w_k \cdot d_k^2$$

apporta una correzione

- sia a **w$_k$**

$$w_k = \frac{\overline{p}_k \cdot (1-\overline{p}_k) \cdot n_{k1} \cdot n_{k2}}{n_k - 1}$$

- sia a **d$_k$**

$$d_k = \frac{p_{k1} - p_{k2}}{\overline{p}_k \cdot (1-\overline{p}_k)} \cdot \frac{n_k - 1}{n_k}$$

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	$p_{kA} - p_{kB}$	$\overline{p}_k$	d_k	n_k	w_k	$w_k d_k$	$w_k d_k^2$
I	0,048	0,338	**0,214**	210	**11,80**	2,53	**0,54**
II	-0,062	0,322	**-0,283**	326	**17,72**	-5,02	**1,43**
III	0,069	0,263	**0,355**	290	**14,09**	5,00	**1,78**
Totale	-----	-----	-----	-----	**43,61**	2,51	**3,75**

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.

3.12. 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	**7**	**244**	**49**	300
Pisanino	**8**	**156**	**24**	188
Pradarena	**22**	**231**	**31**	284
Pradaccio	**143**	**185**	**116**	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	**44,4**	**201,3**	**54,3**	300
Pisanino	**27,8**	**126,2**	**34,0**	188
Pradarena	**42,0**	**190,6**	**51,4**	284
Pradaccio	**65,8**	**297,9**	**80,3**	444
Totale	180	816	220	1216

Con la formula generale (Oss. - Att)^2/ Att. si calcola il valore del χ^2, che avrà 6 gradi di libertà.

$$\chi^2_{(6)} = \frac{(7-44,4)^2}{44,4} + \frac{(244-201,3)^2}{201,3} + \frac{(49-54,3)^2}{54,3} + \frac{(8-27,8)^2}{27,8} + \frac{(156-126,2)^2}{126,2} +$$

$$+ \frac{(24-34)^2}{34} + \frac{(22-42)^2}{42} + \frac{(231-190,6)^2}{190,6} + \frac{(31-51,4)^2}{51,4} + \frac{(143-65,8)^2}{65,8} +$$

$$+\frac{(185-297,9)^2}{297,9}+\frac{(116-80,3)^2}{80,3}=240,571$$

Si ottiene un chi quadrato con 6 gradi $\chi^2_{(6)}$ di libertà uguale a 240,571.

Il valore risulta molto alto; la probabilità α 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	**6**	**23**	29
Trattamento II	**2**	**26**	28
Trattamento III	**8**	**22**	30
Trattamento IV	**3**	**20**	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	**5,0**	**24,0**	29
Trattamento II	**4,8**	**23,2**	28
Trattamento III	**5,2**	**24,8**	30
Trattamento IV	**4,0**	**19,0**	23
Totale	19	91	110

e successivamente, attraverso la formula generale, il valore del chi quadrato con 3 gradi di libertà

$$\chi^2_{(3)} = \frac{(6-5)^2}{5} + \frac{(23-24)^2}{24} + \frac{(2-4,8)^2}{4,8} + \frac{(26-23,2)^2}{23,2} +$$
$$+ \frac{(8-5,2)^2}{5,2} + \frac{(22-24,8)^2}{24,8} + \frac{(3-4)^2}{4} + \frac{(20-19)^2}{19} = 4,02325$$

Si ottiene un valore $\chi^2_{(3)}$ 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	**131**	**32**	163
24°c	**100**	**64**	164
28°c	**90**	**91**	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	**102,7**	**60,3**	163
24°c	**103,3**	**60,7**	164
28°c	**114,0**	**67,0**	181
Totale	320	188	508

e attraverso la formula generale si può computare il valore complessivo del chi quadrato,

$$\chi^2_{(2)} = \frac{(131-102{,}7)^2}{102{,}7} + \frac{(32-60{,}3)^2}{60{,}3} + \frac{(100-103{,}3)^2}{103{,}3} +$$

$$+\frac{(64-60{,}7)^2}{60{,}7} + \frac{(90-114)^2}{114} + \frac{(91-67)^2}{67} = 35{,}004$$

con il contributo fornito da ogni casella

(OSS. - ATT.)2 / ATT   PER OGNI CASELLA

Temperatura	Uova schiuse	Uova non-schiuse	Totale
16° C	**7,798**	**13,282**	**---**
24° C	**0,105**	**0,179**	**---**
28° C	**5,053**	**8,587**	**---**
Totale	**---**	**---**	**35,004**

Il valore del $\chi^2_{(2)}$ (uguale a 35,004) è molto elevato; per 2 gradi di libertà, è associato ad una probabilità α 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
Euplotes p.	**11**	**37**	48
Paramecium c.	**15**	**33**	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

$$\chi^2_{(1)} = \frac{(11 \cdot 33 - 37 \cdot 15)^2 \cdot 96}{48 \cdot 48 \cdot 26 \cdot 70} = 0{,}843$$

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

$$\chi^2_{(1)} = \frac{\left(|11 \cdot 33 - 37 \cdot 15| - \frac{96}{2}\right)^2 \cdot 96}{48 \cdot 48 \cdot 26 \cdot 70} = 0{,}4747$$

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 $\mathbf{G^2}$:

$$G^2 = 2 \cdot \{[\, 11 \cdot \ln(11) + 37 \cdot \ln(37) + 15 \cdot \ln(15) + 33 \cdot \ln(33) + 96 \cdot \ln(96)] \; -$$

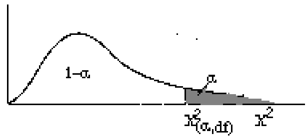
$$- [\ 48 \cdot \ln(48) + 48 \cdot \ln(48) + 26 \cdot \ln(26) + 70 \cdot \ln(70)]\} =$$

$$G^2 = 2 \cdot \{[(11x2,398)+(37x3,611)+(15x2,708)+(33x3,497)+(96x4,564)]-$$
$$- [(48x3,871)+(48x3,871)+(26x3,258)+(70x4,248)]\}=$$

$$G^2 = 2 \cdot (754,104 - 753,742) = 2 \text{ x } 0,362 = 0,724$$

Il valore del χ^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.

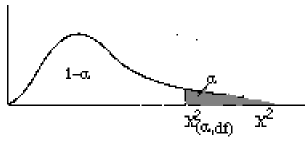
VALORI CRITICI DELLA DISTRIBUZIONE χ^2 (con gdl da 1 a 30)



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	49.5B8	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 χ^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	32.4á0	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