

Il test del CHI-QUADRATO

- Esiste una procedura più flessibile del test z che consente anche di confrontare più di due campioni o più di due esiti: il **test Chi-Quadrato** (χ^2).
- Esso rientra nei test non parametrici, in quei metodi cioè che non presuppongono alcuna condizione circa la natura o i parametri della popolazione dalla quale i campioni provengono.
- Per descrivere la procedura analizziamo le tabelle di contingenza derivate da uno studio clinico controllato in **doppio cieco** per stabilire l'efficacia di dosi leggere di aspirina nella prevenzione di trombi in pazienti dializzati (naturalmente i gruppi risultavano omogenei per quanto riguarda il tempo in dialisi ed altre presunte variabili di confondimento).

Il test del CHI-QUADRATO: tabella delle frequenze osservate

La seguente **tabella a doppia entrata** mostra i risultati ottenuti nell'esperimento con l'aspirina e con il placebo, ovvero il numero di pazienti in ogni gruppo (trattati e non trattati) che sviluppò o non sviluppò l'emorragia:

Gruppo campionario	Sviluppo emorragia	Non sviluppo emorragia	Totale righe
Placebo	18	7	25
Aspirina	6	13	19
Totale colonne	24	20	44

- La differenza riscontrata nei due gruppi è più grande di quella che ci aspetteremmo se l'aspirina agisse come un placebo (ipotesi nulla)?
- Poiché le maggiori frequenze appaiono sulla diagonale della tabella (18 pazienti non trattati hanno sviluppato emorragia, 13 pazienti trattati con aspirina non li hanno sviluppati), tenderemmo a rifiutare l'ipotesi nulla e concludere che esiste un'associazione fra presenza di emorragia e assenza del trattamento con aspirina.

Esempio tratto da "Statistica per discipline biomediche". Stanton A. Glantz. McGraw-Hill

Il test del CHI-QUADRATO: tabella delle frequenze attese

La seguente **tabella a doppia entrata** mostra come potrebbero apparire i dati se fosse vera l'ipotesi nulla, ovvero se l'aspirina non avesse alcun effetto:

Gruppo campionario	Sviluppo emorragia	Non sviluppo emorragia	Totale righe
Placebo	13,64	11,36	25
Aspirina	10,36	8,64	19
Totale colonne	24	20	44

- Essa è ricavata considerando che, a prescindere dai gruppi, una proporzione di 24/44 dei pazienti ha sviluppato *emorragia* e 20/44 non li ha sviluppati. Così, sotto l'ipotesi che il trattamento non ha alcun effetto, ci aspetteremmo che 24/44 dei 25 soggetti non trattati e 24/44 dei 19 soggetti trattati sviluppassero *emorragia* *.
- Come possiamo quantificare la differenza fra la tabella delle frequenze attese e quella delle frequenze osservate?

* Notare come, nel caso di una tabella 2X2, basti calcolare la frequenza attesa in una cella, per determinare per differenza dalle frequenze totali marginali tutte le altre.

Il calcolo del CHI-QUADRATO

- Per valutare la disomogeneità esistente fra le frequenze osservate e quelle attese sotto l'ipotesi che non ci fosse associazione fra trattamenti ed esiti, definiamo il test statistico χ^2 come:

$$\chi^2 = \sum \frac{(O - A)^2}{A}$$

dove O e A rappresentano rispettivamente le frequenze osservate e attese in una certa cella. La somma è calcolata per tutte le celle della tabella di contingenza..

- Il valore di χ^2 è tanto più elevato quanto più le frequenze osservate differiscono da quelle attese.
- Come tutti i test statistici, per effetto del campionamento casuale, χ^2 può assumere uno spettro di valori anche quando non c'è associazione fra trattamenti ed esiti e la sua distribuzione dipende dal numero di gradi di libertà che è dato da:

$$\nu = (r - 1)(c - 1) \quad \text{dove } r \text{ è il numero di righe e } c \text{ il numero di colonne della tabella.}$$

I valori critici del CHI-QUADRATO

Analogamente a quanto visto nei test precedenti, dopo aver calcolato χ^2 dai dati rifiuteremo l'ipotesi nulla con un rischio di errore $p < \alpha$ se il valore ottenuto supererà il corrispondente valore critico. Poiché, nel caso dell'esperimento sulla formazione di *emorragia* abbiamo:

Valori critici del χ^2

v	α			
	0,10	0,05	0,025	0,01
1	2,706	3,841	5,024	6,635
2	4,605	5,991	7,387	9,210
3	6,251	7,815	9,348	11,345
4	7,779	9,488	11,143	13,277
5	9,236	11,070	12,833	15,086
6	10,645	12,592	14,449	16,812
7	12,017	14,067	16,013	18,475
8	13,362	15,507	17,535	20,090
9	14,684	16,919	19,023	21,666
10	15,987	18,307	20,483	23,209

Probabilità di valori maggiori, p

* in tabelle più grandi di 2X2 le frequenze attese in ogni casella non dovrebbero mai essere inferiori a 1 e non più del 20% inferiori a 5.

$$\chi^2 = \sum \frac{(O-A)^2}{A} = \frac{(18-13,64)^2}{13,64} + \frac{(7-11,36)^2}{11,36} + \frac{(6-10,36)^2}{10,36} + \frac{(13-8,64)^2}{8,64} = 7,10$$

e il valore ottenuto supera il valore critico 6,635 associato ad $\alpha=0,01$, possiamo affermare che l'aspirina è associata ad un tasso più basso di emorragia e possiamo rifiutare l'ipotesi di assenza di associazione con un rischio di errore $p < 0,01$.

• Nelle tabelle 2X2 tutte le frequenze attese devono essere > 5 affinché il test sia accurato*. Inoltre, per una migliore compatibilità con la distribuzione teorica, sarebbe opportuno applicare la correzione di Yates per la continuità:

$$\chi^2 = \sum \frac{(|O-A| - 0,5)^2}{A}$$

Applicazione del CHI-QUADRATO a tabelle con dimensioni $>$ di 2x2

• La tabella mostra i risultati di un'indagine sulle comorbidità in un campione di 98 pazienti ospedalizzati con diagnosi di Schizofrenia, Disturbi Schizoaffettivi e Disturbi Bipolari. Fra parentesi sono indicate le frequenze attese sotto l'ipotesi nulla che non ci sia una prevalenza di comorbidità significativamente più alta in qualcuno dei tre gruppi.

Comorbidità	Schizofrenia	Dist. Schizoaff.	Dist. Bipolari	Totale righe
SI	17 (10,79)	6 (8,91)	23 (26,30)	46
NO	6 (12,21)	13 (10,09)	33 (29,70)	52
Totale colonne	23	19	56	98

$$\chi^2 = \frac{(17-10,79)^2}{10,79} + \frac{(6-8,91)^2}{8,91} + \frac{(23-26,30)^2}{26,30} + \frac{(6-12,21)^2}{12,21} + \frac{(13-10,09)^2}{10,09} + \frac{(33-29,70)^2}{29,70} = 9,293 \quad v = (3-1)(2-1) = 2$$

• La tabella dei valori critici mostra che, se la differenza fra le frequenze attese e osservate è dovuta solo al caso, meno del 1% delle volte il valore del χ^2 supera 9,21. Dunque concludiamo che c'è una relazione fra la classificazione diagnostica e la presenza di comorbidità ($p < 0,01$). Tuttavia non sappiamo quale o quali gruppi si differenziano maggiormente.

La scomposizione delle tabelle di contingenza

- Per verificare quali coppie di gruppi presentano una differenza significativa sulla prevalenza di comorbidità dobbiamo effettuare i confronti multipli applicando la disuguaglianza di Bonferroni.
- Le percentuali di comorbidità mostrano valori più alti (73,9% vs 31,6% e 41,1%) negli Schizofrenici rispetto agli Schiozaffettivi ed ai Bipolari; perciò decidiamo dapprima di confrontare fra loro questi due gruppi e successivamente, in caso di differenza non significativa, di accorparli e confrontarli con gli Schizofrenici .

Comorb.	Schizoaff.	Bipol.	Totale
SI	6 (7,35)	23 (21,67)	29
No	13 (11,65)	33 (34,33)	46
Totale	19	56	75

Comorb.	Schizofren.	Schizaff.+Bip.	Totale
SI	17 (10,79)	29 (35,18)	46
No	6 (12,21)	46 (39,82)	52
Totale	23	75	98

- Il valore del χ^2 nel 1° confronto (0,21) ci conferma che gli Schizoaffettivi ed i Bipolari hanno una prevalenza simile, mentre il valore del 2° confronto (7,422), dimostra una prevalenza significativamente più alta negli Schizofrenici ($p < 0,01$) rispetto agli altri due gruppi accorpati. Tenuto conto della disuguaglianza di Bonferroni, il rischio totale di commettere almeno un errore di tipo I è $p < 2\alpha$, ovvero $p < 0,02$.

Calcolo del rischio relativo in una tabella di contingenza

La tabella illustra lo schema generale per il calcolo del rischio relativo che è semplicemente il rapporto fra la probabilità (incidenza) che si verifichi un evento (ad es. ammalarsi) nel gruppo degli esposti (o sottoposto a trattamento) e la probabilità dell' evento corrispondente nel gruppo dei non esposti. Ovvero:

	Non malati	
	malati	malati
Esposti	a	c
non esposti	b	d

$$RR = \frac{a/(a+c)}{b/(b+d)} = \frac{a}{a+c} \times \frac{b+d}{b}$$

L' ipotesi nulla su RR che generalmente si vuol saggiare è che questo sia pari a 1 (incidenza di esposti e non esposti uguale). Questo viene fatto utilizzando l' errore standard di RR.

Esempio:

	Sviluppo emorragia	Non svil. emorragia
aspirina	6	13
Placebo	18	7

$$RR = \frac{6/(6+13)}{18/(18+7)} = \frac{6}{6+13} \times \frac{18+7}{18} = 0,32 \times 1,38 = 0,44$$

Nota: ricordo che, per le tabelle di contingenza, si applica generalmente il test del Chi-Quadrato come già illustrato precedentemente proprio in questo studio sulla prevenzione di trombi in soggetti dializzati.

Intervallo di confidenza del rischio relativo

Si può dimostrare che la distribuzione campionaria del logaritmo naturale dei rischi relativi è approssimativamente normale. Inoltre, l'errore standard del logaritmo di RR è:

$$\sigma_{\log_e RR} = \sqrt{\frac{1 - a/(a+b)}{a} - \frac{1 - c/(c+d)}{c}}$$

L'intervallo di confidenza al livello del 95% del vero logaritmo naturale di RR è:

$$\log_e RR - 1,96\sigma_{\log_e RR} < \log_e RR_{vero} < \log_e RR + 1,96\sigma_{\log_e RR}$$

Da cui, applicando la funzione esponenziale agli elementi della disequazione, abbiamo:

$$e^{\log_e RR - 1,96\sigma_{\log_e RR}} < RR_{vero} < e^{\log_e RR + 1,96\sigma_{\log_e RR}}$$

➤ **Pertanto, è possibile verificare l'ipotesi nulla che RR sia pari a 1, ovvero che il fattore di rischio (o trattamento) non abbia effetto, controllando che l'intervallo di confidenza non includa il valore 1.**

Intervallo di confidenza del rischio relativo, esempio

	Sviluppo trombi	Non svil. trombi
aspirina	6	13
Placebo	18	7

$$RR = \frac{6/(6+13)}{18/(18+7)} = \frac{6}{6+13} \times \frac{18+7}{18} = 0,32 \times 1,38 = 0,44$$

$$\sigma_{\log_e RR} = \sqrt{\frac{1 - a/(a+b)}{a} - \frac{1 - c/(c+d)}{c}} = \sqrt{\frac{1 - 6/(6+13)}{6} - \frac{1 - 13/(13+7)}{13}} = 0,390$$

$$e^{\log_e RR - 1,96\sigma_{\log_e RR}} < RR_{vero} < e^{\log_e RR + 1,96\sigma_{\log_e RR}} \quad \text{ovvero:}$$

$$e^{\log_e 0,44 - 1,96 \times 0,390} < RR_{vero} < e^{\log_e 0,44 + 1,96 \times 0,390}$$

da cui:

$$0,20 < RR_{vero} < 0,94 \quad \text{➤ Poiché questo intervallo non comprende il valore 1, possiamo essere confidenti al 95% che l'aspirina previene lo sviluppo di una trombosi.}$$

Calcolo dell'Odds-Ratio in una tabella di contingenza

Anche in presenza di eventi rari (incidenza molto bassa), il rischio relativo può consentire di misurare la forza dell'associazione fra il fattore di esposizione e la malattia. In questo caso, l'indice ottenuto come rapporto tra i prodotti dei valori in diagonale della tabella di contingenza, detto **Odds Ratio (OR)**, rappresenta un'ottima approssimazione del rischio relativo.

	malati	Non malati
Esposti	a	c
non esposti	b	d

$$RR = \frac{a/(a+c)}{b/(b+d)} \approx OR = \frac{a/c}{b/d} = \frac{a}{c} \times \frac{d}{b}$$

Infatti se l'evento è raro, abbiamo:

$$a + c \approx c$$

$$b + d \approx d$$

Anche in questo caso l'ipotesi nulla su OR che generalmente si vuol saggiare è che questo sia pari a 1 (incidenza di esposti e non esposti uguale). Questo viene fatto utilizzando l'errore standard di OR.

Intervallo di confidenza dell'Odds-Ratio

Si può dimostrare che la distribuzione campionaria del logaritmo naturale dell'odds ratio è approssimativamente normale. Inoltre, l'errore standard del logaritmo di OR è:

$$\sigma_{\log_e OR} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

L'intervallo di confidenza al livello del 95% del vero logaritmo naturale di OR è:

$$\log_e OR - 1,96\sigma_{\log_e OR} < \log_e OR_{vero} < \log_e OR + 1,96\sigma_{\log_e OR}$$

Da cui, applicando la funzione esponenziale agli elementi della disequazione, abbiamo:

$$e^{\log_e OR - 1,96\sigma_{\log_e OR}} < OR_{vero} < e^{\log_e OR + 1,96\sigma_{\log_e OR}}$$

➤ Pertanto, è possibile verificare l'ipotesi nulla che OR sia pari a 1, ovvero che il fattore di rischio (o trattamento) non abbia effetto, controllando che l'intervallo di confidenza non includa il valore 1.

- Va ricordato che il test chi-quadrato va usato con tabelle le cui entrate siano frequenze. E' un errore usarlo con valori medi oppure percentuali.
- Il test chi-quadrato è un metodo approssimato valido quando le frequenze sono grandi.
- Una regola perchè sia valido è che il valore atteso di ogni cella sia maggiore o uguale a 5.
- Quando le frequenze attese sono basse (ma sempre >5) si applica la correzione di Yates che riduce di $\frac{1}{2}$ la grandezza assoluta di (O-E) per ciascuna cella:

I dati appaiati e il test di McNemar

• Il Test di McNemar è utilizzato quando la variabile in studio è qualitativa nominale ed è osservata in situazioni in cui le unità statistiche dei due campioni sono appaiate (un esempio tipico è quando confrontiamo i risultati di due diversi test diagnostici che, negli stessi individui, sono classificati come positivi o negativi).

– *Illustriamo il procedimento utilizzando come esempio i dati tratti da uno studio sull'attendibilità di un test diagnostico basato sull'applicazione sulla cute del dinitroclorobenzene (DNCB) per verificare le condizioni del sistema immunitario (il DNCB è un irritante chimico che, se il sistema è attivo, provoca una reazione).*

– *Poiché molti ricercatori attribuivano la reazione cutanea prevalentemente al danneggiamento in situ del tessuto indipendentemente dal sistema immunitario, fu organizzato un esperimento in cui il DNCB e l'olio di Croton (un noto irritante estratto da semi di un arbusto tropicale) vennero applicati alla cute di 173 soggetti affetti da cancro. – L'idea era che se il DNCB e l'olio di Croton avessero prodotto effetti simili, si sarebbe dovuto respingere la tesi che la reazione dei pazienti al DNCB fosse legata all'attivazione del sistema immunitario.*

Il Test di McNemar (esempio)

Reazione cutanea al DNCB e all' olio di Croton nei pazienti affetti da cancro

Olio di Croton	DNCB	
	positiva	negativa
positiva	O ₁₁ =81	O ₁₂ =48
negativa	O ₂₁ =23	O ₂₂ =21

Fonte: J.A.Roth et al. "Lack of correlation between skin reactivity to Dinitrochlorobenzene and Croton Oil in patient with cancer". N. Engl. J. of Med., 293: 388-389, 1975

- Nella tabella le righe e le colonne non sono indipendenti, in quanto rappresentano le risposte degli stessi soggetti a due trattamenti diversi.
- Gli 81 soggetti che avevano risposto ad ambedue i trattamenti ed i 21 che non avevano risposto a nessuno dei due non ci danno informazioni sul fatto che i soggetti rispondano in modo diverso ai due irritanti.
- Il Test di McNemar focalizza allora l'attenzione sui 71 soggetti che hanno risposto ad uno degli irritanti, ma non all'altro.

Il Test di McNemar: calcolo del χ^2

- Se non ci fosse differenza negli effetti del DNCB e dell' olio di Croton, dovremmo attenderci che metà dei soggetti che hanno risposto ad uno solo degli agenti irritanti abbia risposto al DNCB e non all'olio di Croton e l'altra metà all'olio di Croton e non al DNCB.
- Perciò le frequenze attese nelle celle relative ai soggetti che hanno risposto ad uno solo dei due trattamenti sono pari a $71/2=35,5$.
- Per paragonare queste frequenze con quelle osservate possiamo usare il test χ^2 ottenendo:

$$\chi^2 = \sum \frac{(O - A)^2}{A} = \frac{(23 - 35,5)^2}{35,5} + \frac{(48 - 35,5)^2}{35,5} = 8,113$$

- Il valore del test* è maggiore di 6,635 (il valore critico della distribuzione χ^2 associato, per un grado di libertà, al livello di significatività $\alpha=0,01$).
- Il test porta dunque a concludere che c'è una differenza significativa fra la risposta al DNCB e quella all'olio di Croton con una probabilità di errore $p < 0,01$.

*Il valore χ^2 può essere calcolato anche come:
$$\chi^2 = \frac{(O_{12} - O_{21})^2}{O_{12} + O_{21}} = \frac{(48 - 23)^2}{48 + 23} = 8,113$$

Il Test di McNemar: riepilogo

In breve ecco i passi necessari al calcolo del Test di McNemar:

- Ignorare i soggetti che hanno risposto in modo uguale ad entrambi i trattamenti.
- Calcolare il numero totale dei soggetti che hanno risposto in modo diverso ai due trattamenti.
- Calcolare il numero atteso dei soggetti che avrebbero dovuto rispondere positivamente a un trattamento ma non all'altro: questo numero è la metà del numero totale dei soggetti che hanno risposto in modo diverso ai due trattamenti.
- Confrontare il numero osservato di soggetti che hanno risposto ad uno solo dei trattamenti col numero atteso calcolando il χ^2 con la correzione di Yates per la continuità.
- Confrontare il valore del χ^2 ottenuto, con i valori critici della distribuzione del χ^2 associati a 1 grado di libertà.

* Esercizio *

* Esempio tratto da: " Biostatistica" di M. Pagano e K. Gauvreau. Edizioni Gnocchi Napoli.

In uno studio sulla variabilità intra osservatore nella valutazione degli strisci cervicali, sono stati esaminati 3325 vetrini per individuare l'eventuale presenza di cellule squamose anormali. Ogni vetrino è stato esaminato da un osservatore e riesaminato dopo 6 mesi dallo stesso osservatore. I risultati di questo studio sono riportati in tabella.

Primo screening	Secondo screening		Totale
	positivo	negativo	
positivo	1763	489	2252
negativo	403	670	1073
Totale	2166	1159	3325

Questi dati confermano l'ipotesi nulla che non esiste alcuna associazione tra momento dello screening e diagnosi?

* **Esercizi 10.12 (soluzione)** *

Il totale delle valutazioni discordanti è:

$$Tot_{v_disc} = 489 + 403 = 892$$

Il numero atteso di valutazioni con esito discordante è:

$$Tot_{a_disc} = 892 / 2 = 446$$

$$\chi^2 = \frac{(489 - 446 - 0,5)^2}{446} + \frac{(403 - 446 - 0,5)^2}{446} = 8,10$$

$$v = 1$$

$P < 0,01$ rifiutiamo H_0

La diagnosi ha maggiore probabilità di essere presente al 1° screening