

invece di 244 trattati col metodo *B* ne morirono 64. Allora $p_1 = 41/257 = 0,1596$ e $p_2 = 64/244 = 0,2623$.

La differenza tra le due mortalità si può stimare pari a $0,1595 - 0,2623 = -0,1028$

Come limiti di confidenza al 95% si ottiene

$$\begin{aligned} \text{var}(p_1 - p_2) &= \frac{0,1595 \cdot 0,8405}{257} + \frac{0,2623 \cdot 0,7377}{244} \quad (4.11) \\ &= 0,0005216 + 0,0007930 = 0,0013146 \\ \text{SE}(p_1 - p_2) &= \sqrt{0,0013146} = 0,0363 \end{aligned}$$

Pertanto i limiti di confidenza sono

$$-0,1028 \pm 1,96 \cdot 0,0363 = -0,0317 \text{ e } -0,1739$$

dove il segno meno indica la direzione della differenza.

Per ottenere il test di significatività si calcola la frequenza relativa combinata

$$\bar{p} = 105/501 = 0,2096$$

e si stima $\text{SE}(p_1 - p_2)$ pari a

$$\sqrt{0,2096 \cdot 0,7904 \left(\frac{1}{257} + \frac{1}{244} \right)} = 0,0364$$

Pertanto la deviato normale è

$$\frac{-0,1028}{0,0364} = -2,82 (p < 0,01)$$

che indica decisamente una differenza di mortalità a favore del metodo *A*.

*urlan che un
era emperso lo
o!*

Va notato che nell'Esempio 4.10 l'uso di $\bar{p}$ ha modificato di pochissimo il valore dell'errore standard, da 0,0363 a 0,0364; in effetti è più facile apportare una modifica sostanziale solo quando n_1 e n_2 sono molto diversi e p_1 e p_2 differiscono apprezzabilmente; in altre situazioni le due formule dell'errore standard possono essere considerate l'una l'approssimazione dell'altra e usate come tali.

4.8. Tabelline 2×2 e test del χ^2

La Tab. 4.5 riporta un modo alternativo per rappresentare i dati dell'esempio 4.10; si tratta di una cosiddetta *tavola* o *tabella di contingenza* 2×2 . L'angolo inferiore destro della tabella riporta il numero totale di osservazioni, in questo caso 501; il numero totale o *gran totale* è suddiviso secondo

due criteri diversi rappresentati dalle righe "orizzontali" e dalle colonne "verticali" della tabellina. Nell'esempio le righe rappresentano i trattamenti e le colonne gli esiti dei trattamenti; vi sono pertanto $2 \times 2 = 4$ combinazioni di righe e colonne; le frequenze assolute associate alle singole categorie sono riportate nel corpo della tabella (*celle interne*); le frequenze totali per le due righe e per le due colonne sono riportate rispettivamente sul margine destro e sul margine inferiore della tabella (*totali marginali*).

Tabelline 2×2 sono già state applicate per rappresentare i risultati di un confronto di proporzioni in campioni appaiati ma lo scopo era un po' diverso da quello attuale, che riguarda unicamente il caso di dati non appaiati.

Tabella 4.5. Tabella 2×2 dei risultati di un esperimento clinico.

Trattamento	Risultato		Totale
	Morte	Sopravvivenza	
<i>A</i>	41	216	257
<i>B</i>	64	180	244
Totale	105	396	501

In Tab. 4.5 interessano possibili differenze tra le mortalità dei due trattamenti. Una volta che siano stati fissati i totali marginali si possono facilmente calcolare le frequenze che si sarebbero osservate nel corpo della tabellina qualora le mortalità di *A* e di *B* fossero state identiche. Per esempio per la cella in alto a sinistra la frequenza attesa è

$$\frac{105 \times 257}{501} = 53,862$$

dato che la mortalità generale è $105/501$ e gli individui trattati con *A* sono 257. Analogamente si ottengono tutte le altre frequenze attese; in Tab. 4.6 sono riportate contemporaneamente frequenze attese e frequenze osservate distinte tra di loro dalle lettere *E* e *O*. Le frequenze attese non sono numeri interi e pertanto sono state arrotondate alla terza cifra decimale: (chiaramente è impossibile osservare 53,862 individui di una data categoria) le frequenze attese devono essere considerate come attese o valori medi calcolati su un gran numero di tabelle possibili aventi totali marginali uguali a quelli osservati e sotto la condizione che l'ipotesi zero sia vera.

Va notato che i valori di *E* danno come somma sia sulle righe che sulle colonne i totali marginali osservati. Da ciò consegue che gli scarti misurati tramite le differenze $O - E$ assommano a zero sia lungo le righe che lungo

Tabella 4.6.

Trattamento

	<i>O</i>	<i>E</i>
<i>A</i>	<i>O</i> —	(<i>O</i> —
	(<i>O</i> —	(<i>O</i> —
	<i>O</i>	<i>E</i>
<i>B</i>	<i>O</i> —	(<i>O</i> —
	(<i>O</i> —	<i>O</i> —
Totale	<i>O</i>	<i>E</i>
	<i>O</i> —	<i>O</i> —

le colonne; in altri (in questo esempio

In un certo senso contro l'ipotesi zero tali scarti; è però alle frequenze: uno = 100.

Risulta che l'ind

estendendo la somma i contributi delle

$$\chi^2 =$$

In base all'ipotesi zero e l'approssimazione Con riferimento all

Tabella 4.6. Frequenze attese e contributi all'indice X^2 per i dati di Tab. 4.5.

Trattamento		Risultato		Totale
		Morte	Sopravvivenza	
A	O	41	216	257
	E	53,862	203,138	257
	O — E	— 12,862	12,862	0
	(O — E) ²	165,431	165,431	
	(O — E) ² /E	3,071	0,814	
B	O	64	180	244
	E	51,138	192,862	244
	O — E	12,862	— 12,862	0
	(O — E) ²	165,431	165,431	
	(O — E) ² /E	3,235	0,858	
Totale	O	105	396	501
	E	105	396	
	O — E	0	0	

le colonne; in altri termini i quattro scarti hanno lo stesso valore assoluto (in questo esempio 12,862) essendo due positivi e due negativi.

In un certo senso più grande è lo scarto tanto più forti sono le indicazioni contro l'ipotesi zero; è perciò ragionevole basare il test di significatività su tali scarti; è però altrettanto ragionevole tener conto del valore effettivo delle frequenze: uno scarto di 5 è molto più importante se $E = 5$ che se $E = 100$.

Risulta che l'indice appropriato da calcolare è

$$X^2 = \sum' \frac{(O - E)^2}{E} \quad (4.12)$$

estendendo la sommatoria alle quattro celle della tabellina. La Tab. 4.6 riporta i contributi delle quattro celle a X^2 ; il totale è

$$X^2 = 3,071 + 0,814 + 3,235 + 0,858 = 7,978$$

In base all'ipotesi zero X^2 segue la distribuzione $\chi^2_{(1)}$ (vedi il paragrafo 3.4) e l'approssimazione migliora al crescere dei valori delle frequenze attese. Con riferimento alla Tab. A2 si vede che il valore osservato di 7,978 supera il

percentile dell'1% della distribuzione $\chi^2_{(1)}$ e pertanto la differenza tra le due mortalità è significativa al livello dell'1%.

Si è già visto che la distribuzione $\chi^2_{(1)}$ è la distribuzione del quadrato di una deviato normale standardizzata. Nel paragrafo 4.7 si calcolò il valore della deviato normale standardizzata in base all'errore standard della differenza di due proporzioni e si ottenne il valore numerico di 2,82. Elevando al quadrato 2,82 si ottiene 7,95 che concorda con X^2 a meno di errori di arrotondamento. In effetti si può dimostrare algebricamente che l'indice X^2 coincide sempre col quadrato della deviato normale data dal primo metodo. I livelli di probabilità dei due test sono perciò sempre in accordo, come in questo caso in cui i due test danno $P < 0,01$.

Spesso si indica l'indice X^2 con il simbolo χ^2 anche se sembra preferibile riservare quest'ultima notazione alla distribuzione e indicare con X^2 il valore calcolato sul campione.

Vi sono diverse formule alternative per calcolare X^2 ; ne possiamo ricordare una; indichiamo come segue le entrate della tabellina:

		colonne		
		1	2	
righe	1	<i>a</i>	<i>b</i>	r_1
	2	<i>c</i>	<i>d</i>	r_2
		s_1	s_2	N

Allora

$$X^2 = \frac{(ad - bc)^2 N}{r_1 r_2 s_1 s_2} \quad (4.13)$$

I prodotti ripetuti al numeratore e al denominatore di questa formula la rendono poco adatta per la calcolatrice da tavolo, a meno che non abbia il dispositivo per le divisioni successive.

Si dispone in conclusione di due test esattamente equivalenti; quale scegliere per l'applicazione è questione di preferenze personali e di opportunità. Il primo metodo, quello dell'errore standard, fornisce, come si è già visto, non solo un test di significatività ma conduce in modo naturale anche al calcolo dell'intervallo di confidenza. In generale va giudicata inadeguata qualunque metodologia statistica che dia unicamente un test di significatività e non sia in grado di stimare la grandezza delle differenze relative ai confronti di interesse tenendo conto dell'errore campionario. Questo è un forte motivo per calcolare le differenze e l'errore standard e basare il test su questi valori piuttosto che sull'indice X^2 . Il principale contro-argomento è che il metodo X^2 , come si vedrà nel Capitolo VII, può essere esteso a tabelline con più di due righe e più di due colonne.

È importante non dimenticare che l'indice X^2 si può calcolare solo su tabelline le cui entrate siano frequenze; un errore comune è quello di usarlo

per tabelline le cui entrate sono percentuali. La correzione pratica è completa.

Sia il test basato su X^2 che quello basato su χ^2 danno simulazioni valide in casi di campioni grandi. Generalmente la correzione per la continuità non è necessaria.

CORREZIONE PER LA CONTINUITÀ

Questo metodo è stato proposto da Yates. La distribuzione di X^2 , sotto l'ipotesi nulla, è una distribuzione di χ^2 con $k-1$ gradi di libertà, dove k è il numero di totali marginali. In pratica, se il numero finito di totali marginali è piccolo, si possono generare i totali marginali aggiungendo o sottraendo una unità per volta fino a che i totali raggiungano lo zero (cioè, si annullano). (Ciò è discusso nel paragrafo 4.7.) Ci si può chiedere se questa correzione era stata approssimata. In un caso attuale sarebbe stata bilità della tabella di contingenza, riducendo di $\frac{1}{2}$ il valore di X^2 . Nell'esempio 4.1, il valore di X^2 sarebbe stato uguale a 12,3, con X^2_c sarebbe stato uguale a 11,8. La correzione per la continuità non è ancora altament

La correzione per la continuità si applica quando le frequenze attese sono piccole (quasi tutti gli indici di contingenza consigliano di non usare il test di valori X^2 , se il numero di valori X^2 è piccolo).

La versione corre

Se si applica la correzione per la continuità, la formula per X^2 deve essere modificata. Se si applica la correzione per la continuità, bisogna calcolare p e q (o diminuite di $\frac{1}{2}$ le attese) senza modificare il denominatore. $P_{1(c)} = 41,5/257 =$

per tabelline le cui entrate sono valori medi di una data variabile; questa pratica è completamente erranea.

Sia il test basato sull'errore standard sia quello basato sul χ^2 sono approssimazioni valide in particolare quando le frequenze assolute sono molto grandi. Generalmente si applicano due metodi che danno risultati migliori: la correzione per la continuità e il calcolo delle probabilità esatte.

CORREZIONE PER LA CONTINUITÀ IN TABELLINE 2×2

Questo metodo fu descritto da F. Yates ed è spesso chiamato *correzione di Yates*. La distribuzione $\chi^2_{(1)}$ è stata usata come approssimazione alla distribuzione di X^2 , sotto la condizione che fosse vera l'ipotesi zero e fossero fissi i totali marginali. In base a quest'ultimo vincolo si può ottenere soltanto un numero finito di tabelline; ad esempio dati i totali marginali della Tab. 4.5 si possono generare tutte le possibili tabelline aumentando o diminuendo di una unità per volta una entrata della tabellina fintanto che o quella o altre raggiungano lo zero. (Più avanti in questo paragrafo segue una discussione più dettagliata.) Ci si trova pertanto in una situazione molto simile a quella discussa nel paragrafo 2.7 in cui una distribuzione discreta (la binomiale) era stata approssimata con una distribuzione continua (la normale). Nel caso attuale sarebbe ragionevole basare il test di significatività sulla probabilità della tabella osservata o di tabelle che dimostrino una deviazione ancora maggiore dall'ipotesi zero. La stima di tale probabilità si può migliorare riducendo di $\frac{1}{2}$ il valore assoluto della differenza $O - E$ prima di calcolare X^2 . Nell'esempio 4.10, Tab. 4.6, la differenza $|O - E|$ si sarebbe dovuta porre uguale a 12,362 invece che 12,862 e il valore corretto di X^2 , indicato con X^2_c , sarebbe stato 7,369, leggermente inferiore al valore non corretto ma ancora altamente significativo.

La correzione per la continuità ha un effetto relativamente maggiore quando le frequenze attese sono piccole, ma è certamente prudente applicarla a quasi tutti gli indici X^2 relativi a tabelline 2×2 . Un'eccezione a questo consiglio riguarda il caso in cui si debbano combinare tra loro un certo numero di valori X^2 , problema discusso nel paragrafo 12.3.

La versione corretta per la continuità della (4.13) è

$$X^2_c = \frac{(|ad - bc| - \frac{1}{2}N)^2 N}{r_1 r_2 s_1 s_2} \quad (4.14)$$

Se si applica la correzione per la continuità al test χ^2 essa andrebbe logicamente applicata anche al test basato sull'errore standard. In questo caso bisogna calcolare $p_1 - p_2$ dopo che le frequenze assolute sono state aumentate o diminuite di mezza unità (in modo che si avvicinino alle frequenze attese) senza modificare l'errore standard. Nell'esempio 4.10 si avrebbe $P_{1(c)} = 41,5/257 = 0,1615$, $p_{2(c)} = 63,5/244$ e $u_{(c)} = -0,0987/0,0364$