

STATISTICA MEDICA

Dott.ssa Marta Di Nicola
 N.P.D. 3° Blocco 2° piano
 0871-3554007
 m.dinicola@unich.it

<http://www.biostatistica.unich.it>

Dott.ssa Marta Di Nicola

STATISTICA INFERENZIALE PER VARIABILI QUALITATIVE

La presentazione dei dati per molte ricerche mediche fa comunemente riferimento a frequenze, assolute o percentuali. Osservazioni cliniche conducono sovente a risultati tipo "il 60% degli individui trattati con un farmaco è migliorato rispetto al 47% del gruppo di soggetti di controllo", implicando con ciò un confronto tra i risultati ottenuti per i due gruppi.

Risulta evidente che tali risultati non sono espressi da dati su scala quantitativa e quindi non è possibile fare riferimento alla distribuzione Gaussiana o a quelle del t di Student, ma occorre considerare metodiche specifiche che permettano, anche con tale tipo di dati, di verificare l'ipotesi zero di una differenza casuale tra le frequenze riscontrate.

Dott.ssa Marta Di Nicola

La statistica chi-quadrato (χ^2)

Esempio 1. C'è parità tra i 2 sessi nei 180 iscritti al corso di laurea in medicina? Si organizza un'indagine su un campione casuale di 80 studenti.

H_0 : %M nella popolazione=%F nella popolazione
 H_1 : %M nella popolazione \neq %F nella popolazione

I risultati osservati (O) e le attese (A) sono riportati nella tabella.

SESSO	O ₁	A	χ^2	O ₂	χ^2
M	45	40	25/40	50	100/40
F	35	40	25/40	30	100/40
TOT	80	80	50/40	80	200/40
χ^2 g.l.=1			1.25 n.s.		5*

* p < 0.05, risultato del test appena significativo

Dott.ssa Marta Di Nicola

Chi-quadro test: confronto tra campioni

Esempio 3. Si abbia un campione di 1020 soggetti diviso in Fumatori (A): $n_A=400$ \rightarrow Prevalenza BCO 30%
 Non fumatori (B): $n_B=620$ \rightarrow Prevalenza BCO 15%

Il fumo è "causa" (o fattore di rischio) per la bronchite? ossia il Δ (+15%) è statisticamente significativo?

Tabella di contingenza (2x2)

FUMO	BRONCHITE		TOT
	SI	NO	
SI	120	280	400
NO	93	527	620
TOTALE	213	807	1020

La prevalenza di bronchite risulta statisticamente \neq tra i fumatori e i non fumatori?
 H_0 : La bronchite si sviluppa **indipendentemente** dal fumo;
 H_1 : I fumatori sviluppano bronchite più dei non fumatori.

Dott.ssa Marta Di Nicola

TABELLA TETRACORICA D'INDIPENDENZA

Fattore di rischio	Malattia		TOTALE
	P (+)	NP (-)	
A (+)	a	b	$n_A(a+b)$
B (-)	c	d	$n_B(c+d)$
TOTALE	$n_1(a+c)$	$n_2(b+d)$	n

Valori delle frequenze nel caso di indipendenza

$n_1 \cdot n = a \cdot n_A \rightarrow a = (n_A n_1) / n$

$n_1 \cdot n = c \cdot n_B \rightarrow c = (n_B n_1) / n$

idem per b e d


Dott.ssa Marta Di Nicola

Tabella delle frequenze attese

FUMO	BRONCHITE CRONICA		
	SI	NO	TOT
SI	84	316	400
NO	129	491	620
TOTALE	213	807	1020

Es. $(620 \times 213) / 1020 = 129$; per differenza si calcolano le altre tre frequenze interne.

Dott.ssa Marta Di Nicola




$$\chi^2 = \frac{(120 - 84)^2}{84} + \frac{(280 - 316)^2}{316} + \frac{(93 - 129)^2}{129} + \frac{(527 - 491)^2}{491} = 32.21$$

LA FORMULA PER CALCOLARE L'INDICE-TEST CHI-QUADRATO

$$\sum_i \frac{(O_i - A_i)^2}{A_i}$$

Dott.ssa Marta Di Nicola




Nel caso di tabelle 2x2 si può determinare il valore del test chi-quadrato anche direttamente attraverso la seguente formula:

$$\chi^2 = \frac{(ad - cb)^2 n}{n_1 n_2 n_A n_B}$$

LE DUE FORMULE FORNISCONO RISULTATI EQUIVALENTI

$$\chi^2 = \frac{(120 \cdot 527 - 93 \cdot 280)^2 1020}{400 \cdot 620 \cdot 213 \cdot 807}$$

Dott.ssa Marta Di Nicola



La CORREZIONE di YATES


La correzione di Yates viene applicata nel caso di tabelle 2x2 che presentino:

- la numerosità complessiva (n) <200
- oppure una delle marginali (n_A, n_B, n₁, n₂) <40
- comunque a, b, c, d >5

la correzione si attua con la formula:

$$\chi^2 = \frac{(|ad - cb| - n/2)^2 n}{n_1 n_2 n_A n_B}$$

Dott.ssa Marta Di Nicola



Esempio Si supponga di aver rilevato, su un campione di 36 giovani, la pressione arteriosa e la pratica sportiva.


PRATICA SPORTIVA	IPERTENSIONE ARTERIOSA		
	SI	NO	TOT
SI	7	9	16
NO	14	6	20
TOT	21	15	36

Applichiamo il test del chi-quadrato con la correzione di Yates per la continuità

$$\chi^2 = \frac{((7 \cdot 6 - 14 \cdot 9) - 36/2)^2 \cdot 36}{21 \cdot 15 \cdot 20 \cdot 16} = 1.55$$

Il test risulta non significativo (1.55 < 3.84) dunque l'ipotesi nulla di indipendenza tra la pratica sportiva e l'ipertensione arteriosa viene accettata.

Dott.ssa Marta Di Nicola



TEST ESATTO di FISCHER


Viene applicato nel caso in cui in una tabella 2x2 il numero delle osservazioni è minore di 20 o una delle frequenze attese è inferiore a 5. Permette di calcolare direttamente la probabilità esatta.

$$P = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{a! b! c! d! N!}$$

PRATICA SPORT	IPERTENSIONE ARTERIOSA		
	SI	NO	TOT
SI	1	10	11
NO	15	5	20
TOT	16	15	31

$$P_1 = \frac{11! 20! 16! 15!}{1! 10! 15! 5! 31!} = 0.000567$$

Dott.ssa Marta Di Nicola



PRATICA SPORT	IPERTENSIONE ARTERIOSA		
	SI	NO	TOT
SI	0	11	11
NO	16	4	20
TOT	16	15	31

$$P_0 = \frac{11! 20! 16! 15!}{0! 11! 16! 4! 31!} = 0.000016$$

$$P = 0.000567 + 0.000016 = 0.000583$$

Altamente significativo P < 0.001

Dott.ssa Marta Di Nicola

**Generalizzazione al caso di una tabella di
dimensione rxs**

	Guariti	Migliorati	Non migliorati	Tot.
Farmaco A	21 (15)	15 (17)	7 (11)	43
Farmaco B	12 (18)	24 (22)	18 (14)	54
Tot.	33	39	25	97

Si applica la formula generale per una valutazione complessiva:

$$\chi^2 = \frac{(21-15)^2}{15} + \frac{(12-18)^2}{18} + \frac{(15-17)^2}{17} + \frac{(24-22)^2}{22} + \frac{(7-11)^2}{11} + \frac{(18-14)^2}{14} = 8.23$$

Il test risulta statisticamente
significativo

Dott.ssa Marta Di Nicola