NOTE DALLE LEZIONI
DI
STATISTICA MEDICA
ED ESERCIZI

VERIFICA DI IPOTESI
SUL LEGAME TRA
VARIABILI QUALITATIVE

IL PROBLEMA

Si vuole verificare l'esistenza di un legame tra il gruppo sanguigno e la gravità di una certa patologia. Si dispone del numero di individui che presentano contemporaneamente la patologia ad certo grado di gravità e un dato gruppo sanguigno.

	Gruppo sanguigno				
Patologia	A	B	AB	0	Totale
Assente	543	211	90	476	1320
Media	44	22	8	31	105
Grave	28	9	7	31	75
Totale	615	242	105	538	1500

La generalizzazione della tabella precedente è:

1° criterio

2° criterio	1	2	…	j	…	c	Tot.	
1	O_{11}	O_{12}	…	O_{1j}	…	O_{1c}	$n_1.$	
2	O_{21}	O_{22}	…	O_{2j}	…	O_{2c}	$n_2.$	
…	…	…	…	…	…	…	…	
i	O_{i1}	O_{i2}	…	O_{ij}	…	O_{ic}	$n_i.$	
…	…	…	…	…	…	…	…	
r	O_{r1}	O_{r1}	…	O_{rj}	…	O_{rc}	$n_r.$	
Tot.	$n._1$	$n._2$	…	$n._j$	…	$n._c$	N	

IL TEST PER LA VERIFICA DELL'IPOTESI

ASSUNZIONI

Le variabili di cui disponiamo sono qualitative.

Se consideriamo **una sola cella** la presenza contemporanea delle due caratteristiche è il "successo", sugli N casi possibili: si può assumere una **distribuzione binomiale**.

I dati in tabella nel loro insieme seguono una distribuzione multinomiale.

IPOTESI

H_0: $p_{ij} = p_i \times p_j$

H_1: $p_{ij} \neq p_i \times p_j$

$p_{ij} = O_{ij} / N$

$p_i = n_{i\cdot} / N$

$p_j = n_{\cdot j} / N$

Se le due variabili sono **indipendenti** la probabilità di avere la caratteristica 1 **e** la caratteristica 2 sarà data dal **prodotto** delle probabilità (legge del prodotto).

I VALORI ATTESI

Data vera l'ipotesi nulla e posta l'assunzione di distribuzione binomiale in ciascuna cella allora posso calcolare il valore atteso E_{ij} ("media") per ciascuna cella:

$E_{ij} = N\, p_{ij} = N\, p_i\, p_j = N\, (n_{.j} / N)\, (n_{i.} / N) = (n_{.j}\, n_{i.}) / N$

Si può quindi costruire una tabella di valori attesi:

2° criterio	1° criterio						
	1	2	…	j	…	c	Tot.
1	E_{11}	E_{12}	…	E_{1j}	…	E_{1c}	$n_{1}.$
2	E_{21}	E_{22}	…	E_{2j}	…	E_{2c}	$n_{2}.$
…	…	…	…	…	…	…	…
i	E_{i1}	E_{i2}	…	E_{ij}	…	E_{ic}	$n_{i}.$
…	…	…	…	…	…	…	…
r	E_{r1}	E_{r1}	…	E_{rj}	…	E_{rc}	$n_{r}.$
Tot.	$n._{1}$	$n._{2}$	…	$n._{j}$	…	$n._{c}$	N

CALCOLATE LA

TABELLA DEI VALORI ATTESI

E POI

VERIFICATE IL RISULTATO

Nel nostro esempio la tabella dei valori attesi diventa:

Gruppo sanguigno

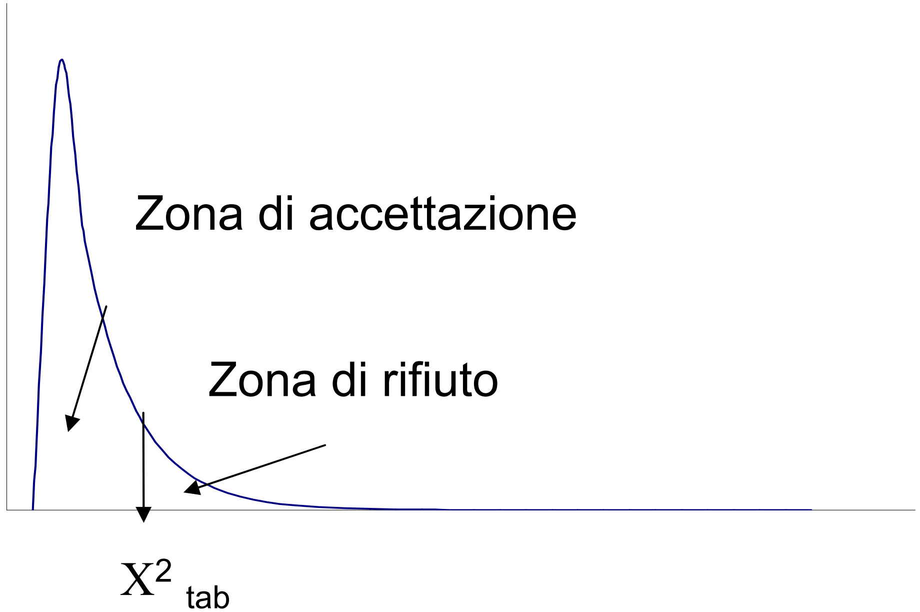
Patol.	A	B	AB	0	Totale
Assente	541,2	212,96	92,40	473,44	1320
Media	43,05	16,94	7,35	37,66	105
Grave	30,75	12,10	5,25	26,90	75
Totale	615	242	105	538	1500

STATISTICA TEST

$$X^2 = \sum_{ij} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

DISTRIBUZIONE DELLA STATISTICA TEST

La distribuzione della statistica test è una X^2 ed è caratterizzata dai gradi di libertà.



REGOLA DI DECISIONE

Fissato α accettabilmente piccolo (0,05), troverò sulle tavole X^2 un valore in corrispondenza di α prescelto e dei gradi di libertà della statistica.
Se il valore calcolato è maggiore del valore tabulato rifiuterò l'ipotesi nulla, se invece il valore calcolato è minore del tabulato accetterò l'ipotesi nulla.

CALCOLATE LA STATISTICA TEST E POI VERIFICATE IL RISULTATO

CALCOLO DELLA STATISTICA TEST

$$X^2 = \frac{(543-541{,}2)^2}{541{,}2} + \frac{(211-212{,}96)^2}{212{,}96} + \frac{(90-92{,}40)^2}{92{,}40} +$$

$$+ \ldots\ldots\ldots + \frac{(31-26{,}90)^2}{26{,}90} = 5{,}12$$

$X^2_{\alpha=0{,}05,\ gl\ 6} = 12{,}592$

DECISIONE STATISTICA

5,12<12,592 accetto l’ipotesi nulla, le due variabili sono indipendenti

DECISIONE DEL RICERCATORE

Non c’è una evidenza di associazione tra un gruppo sanguigno e l’essere affetto dalla malattia in esame.

RICHIAMI TEORICI IMPORTANTI

I GRADI DI LIBERTA'

In questo caso i gradi di libertà sono:

g.l. = (r-1) (c-1)

dove r = numero delle righe

c = numero delle colonne

$\Sigma p_{.j} = \Sigma n_{.j} / N = 1$

$\Sigma p_{i.} = \Sigma n_{i.} / N = 1$

fissato N potrò cambiare "liberamente" $n_{i.}$, totali di riga, meno 1 che mi deve garantire la somma delle probabilità di riga ($\Sigma p_i = 1$).

fissato N potrò cambiare "liberamente" $n_{.j}$, totali di colonna, meno 1 che mi deve garantire la somma delle probabilità di colonna ($\Sigma p_j = 1$).