

Tabelle di Contingenza, Bivariata, Indice chi quadro (quesiti da 1 a 2)

La statistica bidimensionale o bivariata si occupa dello studio del grado di dipendenza di due caratteri distinti della stessa unità statistica. E' possibile, ad esempio, studiare il legame che esiste tra il peso e l'altezza di un gruppo di persone oppure, considerando 100 famiglie, si può studiare il legame che esiste tra il numero di membri di una famiglia e il numero di automobili di proprietà di ciascuna famiglia ecc..

Il tipo di dipendenza che si studia è solo la "dipendenza statistica", e non è detto che ogni volta che si nota una dipendenza statistica c'è anche una dipendenza causa-effetto tra le due variabili in esame. Ecco un esempio significativo: statisticamente si può dimostrare che gli studenti con i piedi piccoli fanno più errori di ortografia di quelli con i piedi grandi. Questo non significa che l'avere i piedi piccoli sia la causa del fare errori di ortografia, piuttosto i ragazzi con i piedi piccoli sono i più giovani e per questo fanno più di frequente errori di ortografia. In questo caso c'è una terza variabile, l'età degli studenti, che genera la relazione di causa-effetto.

Si consideri ad esempio questa tabella 2x2 contenente la sintesi dei risultati riguardanti uno studio clinico riguardante uno specifico farmaco

Tabulazione dei dati osservati

	Aspirina	Placebo	Totale
Esito Positivo	33	14	47
Esito Negativo	17	41	58
Totale	50	55	105

I dati di questa tabella rappresentano le frequenze assolute, cioè è riportato il numero di persone che assumendo il farmaco o il placebo hanno ottenuto esito positivo o negativo in relazione ad una specifica patologia.

I totali riportati sono definiti totali marginali di riga e colonna.

Il numero 105, è il totale generale e rappresenta la numerosità del campione.

L'osservazione di questi dati assoluti mostra un deciso legame tra l'uso dell'aspirina e l'esito positivo del farmaco.

Per una valutazione oggettiva e sintetica di questo legame di dipendenza o indipendenza è possibile calcolare le percentuali di riga e di colonna e le percentuali generali

Percentuali di riga

	Aspirina	Placebo	Totale
Esito Positivo	33/47=70,2%	14/47=29,8%	47/47=100%
Esito Negativo	17/58=29,3%	41/58=70,7%	58/58=100%
Totale	50	55	105

Percentuali di colonna

	Aspirina	Placebo	Totale
Esito Positivo	33/50=66%	14/55=25,5%	47
Esito Negativo	17/50=34%	41/55=74,5%	58
Totale	50/50=100%	55/55=100%	105

Percentuali rispetto al totale generale

	Aspirina	Placebo	Totale
Esito Positivo	33/105=31,4%	14/105=13,3%	47/105=44,8%
Esito Negativo	17/105=16,2%	41/105=39%	58/105=55,2%
Totale	50/105=47,6%	55/105=52,4%	105/105=100%

L'analisi di queste tabelle percentuali permette una prima valutazione del fenomeno di dipendenza statistica. Si possono anche utilizzare i risultati di queste tabelle per realizzare grafici a torta che rappresentano il fenomeno in modo significativo.

Un indice sintetico che esprime in modo oggettivo il fenomeno di dipendenza è l'indice chi quadro. Se tale indice è 0 significa che non è presente alcuna relazione di dipendenza. L'intensità di questo indice è direttamente proporzionale all'intensità del legame tra i due caratteri.

Per il calcolo dell'indice chi quadro si esegue la seguente procedura

- 1) Calcolo dei dati attesi in caso di relazione completamente assente

dati attesi

	Aspirina	Placebo	Totale
Esito Positivo	$47 \cdot 50 / 105 = 22,4$	$47 \cdot 55 / 105 = 24,6$	47
Esito Negativo	$50 \cdot 58 / 105 = 27,6$	$55 \cdot 58 / 105 = 30,4$	58
Totale	50	55	105

questo tipo di calcolo permette di determinare gli attesi se non c'è relazione per il motivo che di seguito espongo utilizzando ad esempio il primo calcolo (prima riga e prima colonna) di questa tabella.

Il dato atteso 22,4 è stato determinato utilizzando questa uguaglianza di rapporti

$$47/105 = X/50$$

questa equazione esprime il seguente concetto:

Se non c'è relazione fra i due caratteri la percentuale di esiti positivi sul totale sarà uguale alla percentuale di esiti positivi rispetto alla totalità delle persone che assumono l'aspirina.

Risolviendo l'equazione rispetto ad X si ottiene:

$$X = 47 \cdot 50 / 100.$$

Si può ragionare analogamente per tutte le altre celle della tabella.

2)

$$\text{chi_quadro} = \sum_{i=1}^4 (\text{osservati}_i - \text{attesi}_i)^2 / \text{attesi}_i$$

La somma sarà estesa a tutte le celle della tabella

$$= (33 - 22,4)^2 / 22,4 + (14 - 24,6)^2 / 24,6 + (17 - 27,6)^2 / 27,6 + (41 - 30,4)^2 / 30,4 = 17,41 \quad (\text{forte dipendenza infatti l'indice è molto maggiore di 1})$$

La significatività di questo indice potrà essere anche confermata utilizzando il test statistico chi-quadro con i gradi di libertà opportunamente calcolati.

Quesito 1

Colore occhi/capelli	guariti	Non guariti	totale
xmicina	52	10	62
streptomicina	40	21	61
totale	92	31	100

Utilizzando la tabella dei dati osservati qui sopra proposta, riguardante l'effetto di due diversi farmaci per il trattamento della leptospirosi, malattia che colpisce i cani.
determinare

- 1) la tabella delle percentuali di riga
- 2) La tabella delle percentuali di colonna
- 3) la tabella delle percentuali rispetto al totale generale
- 4) L'indice chi quadro su questi dati (risultato atteso 5,46)
- 5) Un sintetico commento dei risultati ottenuti

Quesito 2

classe /promozione	promossi	respinti	con debito
classe prima	193	34	20
classe seconda	156	18	32
classe terza	148	22	56
classe quarta	147	11	23

Utilizzando la tabella dei dati osservati qui sopra proposta, riguardante il risultato scolastico nelle diverse classi di un istituto tecnico
determinare

- 1) la tabella delle percentuali di riga
- 2) La tabella delle percentuali di colonna
- 3) la tabella delle percentuali rispetto al totale generale
- 4) L'indice chi quadro su questi dati
- 5) Un sintetico commento dei risultati ottenuti

Utilizzare preferibilmente il foglio elettronico per effettuare i calcoli.