Confrontare due proporzioni o due percentuali: il test *chi-quadrato*

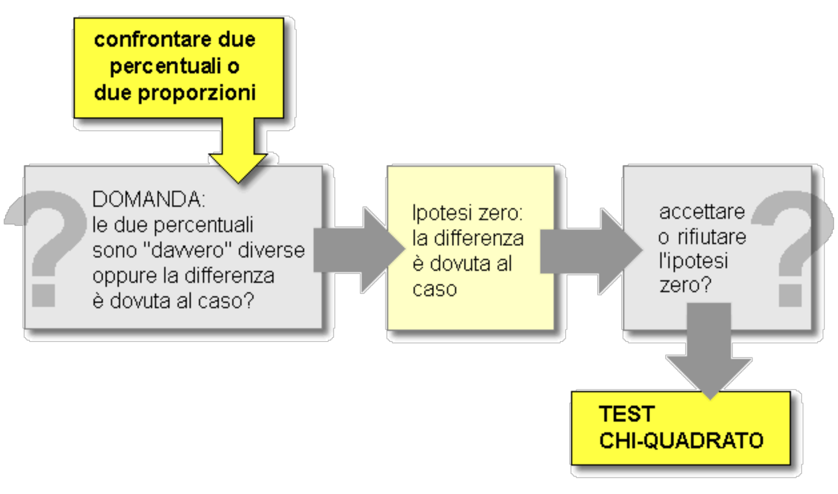
OBIETTIVO:

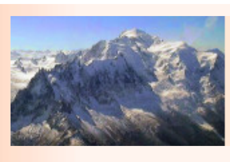
- imparare l' utilizzo del metodo del chi-quadrato attraverso un esempio
- imparare che cosa vuol dire "statisticamente significativo"

[Vai alla versione Mobile]

Nell'unità precedente, ho sottolineato che l'applicazione di un test di significatività statistica è un passo *indispensabile* nel confronto fra due gruppi o popolazioni riguardo a un parametro.

In questa unità ti presento un **esempio** di applicazione di uno dei test più comuni e più semplici, il «**chi-quadrato**».
L'esempio riguarda il confronto di due percentuali ottenute in un esperimento, allo scopo di verificare se la differenza fra tali percentuali è dovuta al caso oppure no. Se riuscirai a dimostrare che la differenza non è dovuta al caso, allora potrai affermare che essa è «statisticamente significativa».
Ti ho già spiegato lo *schema logico da seguire*. Lo schema è molto semplice, in sostanza si tratta di partire con una sorta di pregiudizio: qualsiasi sia la differenza esistente tra le due percentuali da confrontare, inizialmente devi ritenere valida l'«ipotesi zero». L'ipotesi zero (detta anche ipotesi nulla) afferma semplicemente che la differenza osservata - di qualsiasi entità essa sia - è dovuta al caso. Questa ipotesi (che può essere vera o falsa) verrà accettata oppure rifiutata sulla base del risultato di un appropriato test statistico. Nel confronto di due percentuali o di due proporzioni il test appropriato è, appunto, il test del chi-quadrato.
In sintesi:



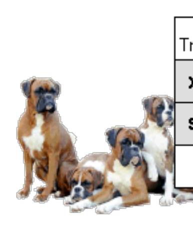


In montagna, puoi raggiungere un punto panoramico con una comoda funivia, che in dieci minuti ti porta a destinazione. Oppure puoi fare da solo, con le tue forze, ore di dura salita, una piccola sfida con te stesso. Dimenticherai facilmente la prima esperienza, mentre la seconda ti resterà, forse, nel cuore per sempre.
Analogamente, puoi fare il test chi-quadrato usando il computer (... la funivia!), e in questo caso troverai in internet tanti strumenti adatti (ce n'è uno anche alla fine di questa unità). Oppure puoi leggere qui di seguito (la sfida: lacrime e sangue!) una spiegazione passo-passo del funzionamento del test.
A te la scelta.

Un esempio: l'efficacia di un farmaco

Supponi di voler mettere a confronto l'efficacia di un nuovo antibiotico (che chiameremo con un **nome di fantasia: xmicina**) con un antibiotico già in uso (streptomicina) nella terapia di una malattia del cane (la leptospirosi), .
A questo scopo, intraprendi un test clinico su un campione di animali costituito dai cani affetti da leptospirosi che vengono presentati in alcuni ambulatori e ospedali veterinari in un determinato periodo di tempo.
Durante la sperimentazione, ogni cane viene assegnato a caso al gruppo dei trattati con il nuovo antibiotico oppure a quello dei trattati con la streptomicina (nel Cap. 9 capirai il perché di questa assegnazione a caso).
Alla fine della sperimentazione, ottieni i dati riassunti nella sottostante Tabella 1.

Tabella 1. Dati ottenuti.



Trattamento ↓ / Esito →	guariti	non-guariti	*totale*
xmicina	52 (a)	10 (b)	62
streptomicina	40 (c)	21 (d)	61
totale	92	31	123

Dei 62 cani trattati con xmicina, ne sono guariti 52 (84%) 52/62 = 0.84

Dei 61 cani trattati con streptomicina, ne sono guariti 40 (66%) 40/61 = 0.66

Vale la pena di commentare in dettaglio la struttura della Tabella 1, considerato che tabelle di questo tipo verranno usate sia in questo Capitolo che, molto più estensivamente nel Cap. 11 (Test di screening e test diagnostici).
Si tratta di una tabella 2x2 (due righe per due colonne), detta anche «tabella di contingenza». Non farti trarre in inganno dal fatto che, in effetti, ci sono tre righe e tre colonne: infatti la «vera» tabella contenente i tuoi dati occupa soltanto nelle quattro celle gialle a, b, c, d. Le altre sono derivate da queste, e non sono altro che i totali di riga e di colonna.
Nella cella a sono indicati i soggetti che sono stati trattati con xmicina e che sono guariti; nella cella b i trattati con xmicina non guariti; nella cella c i trattati con streptomicina guariti; nella cella d i trattati con streptomicina non guariti. I due totali di colonna indicano rispettivamente il numero complessivo di guariti e di non guariti, mentre i totali di riga indicano rispettivamente il numero di trattati con xmicina e con streptomicina.
Nota che, con gli stessi dati, avresti potuto compilare una tabella analoga alla Tabella 1, ma ruotata di 90 gradi, ossia ponendo il trattamento sulle colonne e l'esito sulla righe, come quella a lato.

	xmicina	streptom.	totale
guariti	52 (a)	40 (b)	92
non-guariti	10 (c)	21 (d)	31
totale	62	61	123

Questa impostazione sarebbe stata ugualmente corretta. Però lo «standard» per tabelle di questo tipo è quello di disporre la variabile indipendente (nel nostro caso: il trattamento xmicina/steptomicina) sulle righe e la variabile dipendente (nel nostro caso: guarito/non guarito) sulle colonne. Noi adotteremo sempre questo standard.

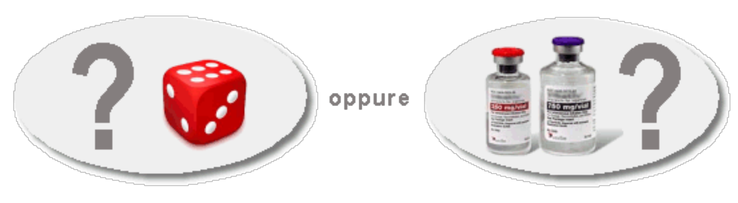
Torniamo alla tabella 1. Puoi notare che, su un totale di 123 cani, 62 sono stati sottoposti a trattamento con xmicina e, fra questi, si sono registrati 52 casi di guarigione (84%). Fra i restanti 61 animali, trattati con streptomicina, ne sono guariti 40 (66%).

È evidente che i dati grezzi indicano che la xmicina è più efficace della streptomicina. Però la superiorità di xmicina potrebbe essere dovutra al caso... Allora, prima di giungere a una conclusione affrettata, occorre rispondere alla seguente domanda:

Supponi che, in realtà, NON esistano differenze nell'efficacia dei due trattamenti. Che probabilità c'è di osservare - in uno studio di dimensioni simili a questo - differenze nell'efficacia dei due antibiotici uguali o superiori a quelle che hai osservato?

La risposta a questa domanda dipende da ***quanto i dati ottenuti si discostano dai dati che «sarebbe lecito attendersi se i trattamenti avessero la stessa efficacia e se i dati fossero influenzati soltanto dalla variazione casuale»***.

Rileggi ancora con attenzione la frase precedente: è la chiave per comprendere il background razionale di un test statistico. Riformulo la stessa frase in modo leggermente diverso, e ti pongo la seguente domanda: se xmicina = streptomicina (ipotesi zero!), è possibile ottenere risultati simili a quelli che hai osservato? La risposta è «sí», perché i tuoi dati potrebbero essere stati influenzati dal caso. Facciamo un altro passo avanti, e chiediamoci: «qual è la probabilità di osservare una differenza uguale o superiore a quella che hai ottenuto nel tuo esperimento?



Guarda ancora la Tabella 1. I tuoi dati dimostrano che **complessivamente** (cioè indipendentemente dal tipo di antibiotico) il trattamento è risultato efficace nel 74.8% dei casi. Infatti sono guariti, sempre complessivamente e indipendentemente dall'antibiotico utilizzato, 52+40=92 animali (a+c) su 123 trattati.
Applicando questa percentuale di successo (74.8%) a ciascuno dei due gruppi di cani in esame (gruppo xmicina e gruppo streptomicina), puoi ricavare i dati della sottostante Tabella 2, che illustra la situazione che ti saresti aspettato se i due antibiotici avessero avuto la stessa efficacia.

Tabella 2. Dati attesi

Trattamento ↓ / Esito →	guariti	non-guariti	*totale*
xmicina	46.37 (a)	15.63 (b)	62.00
streptomicina	45.63 (c)	15.37 (d)	61.00
totale	92.00	31.00	123.00

Nella Tabella 2, hai calcolato il valore a=46 assumendo una percentuale di guarigione del 74.8% nei 62 cani trattati con xmicina: 62*74.8/100=46.37, cioè, approssimando all'unità, a=46. Analogamente, ti saresti aspettato la guarigione del 74.8% dei 61 cani trattati con streptomicina ossia di 45.63 soggetti.
I valori delle celle b e d possono poi essere facilmente ottenuti per differenza:
b = 62-46.37 = 15.63
d = 61-45.63 = 15.37

Il valore del chi-quadrato quantifica la differenza fra i dati osservati e quelli attesi, ed è la somma delle quattro celle a, b, c e d, per ciascuna delle quali si calcola il valore della frazione:

$$\frac{(\text{dato osservato} - \text{dato atteso})^2}{\text{dato atteso}}$$

La magnitudine del chi-quadrato è determinata dalla differenza fra i numeri osservati e i numeri attesi nel caso in cui i due trattamenti avessero avuto lo stesso effetto. La differenza al numeratore della frazione viene elevata al quadrato; ciò elimina i numeri negativi che possono comparire quando il numero osservato è minore di quello atteso. Poi il quadrato della differenza viene diviso per il numero atteso; in questo modo la differenza per ogni cella viene aggiustata in rapporto al numero di individui della stessa cella.
Pertanto, calcoliamo il chi-quadrato come segue:

$$\chi^2 = \frac{(52 - 46.37)^2}{46.37} + \frac{(10 - 15.63)^2}{15.63} + \frac{(40 - 45.63)^2}{45.63} + \frac{(21 - 15.37)^2}{15.37} = 5.46$$

È evidente che il chi-quadrato aumenta con l'aumentare della differenza dei dati posti a raffronto. Se esso supera certi valori prefissati (vedi tabella «Valori di chi-quadrato»), la differenza viene ritenuta significativa; in caso contrario, non si può affermare l'esistenza di una significativa differenza tra i due eventi considerati.
Non ti resta quindi che confrontare il valore ottenuto con la Tabella dei valori di chi-quadrato (reperibile via Internet o in qualsiasi libro di statistica e di una porzione viene riportata qui sotto).

Tabella dei valori di χ2

Gradi di libertà	Probabilità	
	5%	1%
1	3.841	6.635
2	5.991	9.210
3	7.815	11.345
4	9.488	13.277
5	11.070	15.086
6	12.592	16.812
7	14.067	18.475
ecc.	...	...

Nel tuo caso, il valore ottenuto è un chi-quadrato con «**1 grado di libertà**»; infatti, per tabelle come quella che stiamo studiando, il grado di libertà è uguale a (numero di righe-1)*(numero di colonne-1). Quindi: (2-1) * (2-1) = 1 grado di libertà. ciò significa che ti interessa soltanto la prima riga della tabella (celle in verde).

Ora, confrontando il tuo valore di chi-quadrato (5.46) con quelli tabulati, noti che esso è >3.841 e <6.635. Ciò consente di ritenere che la differenza fra i due gruppi sia significativa al livello di probabilità 5% ma non al livello di probabilità 1%.

Puoi concludere che la differenza tra animali trattati con xmicina e quelli trattati con streptomicina è **statisticamente significativa al livello di probabilità 5%**.
In altre parole: *ammettendo che i due antibiotici abbiano pari efficacia* e ripetendo l'esperimento infinite volte, potremo osservare piuttosto raramente (ossia 5 volte su 100 o meno!) dati simili a quelli ottenuti oppure ancor più favorevoli a xmicina.
In sostanza: in base ai risultati del test del chi-quadrato, l'affermazione «xmicina è più efficace di streptomicina» ha il 95% di probabilità di essere vera (e quindi ha il 5% di probabilità di essere falsa).
Se tu dovessi stilare una relazione con i risultati del tuo lavoro, potresti concludere più o meno come segue: «In base ai risultati ottenuti, xmicina è risultata più attiva di streptomicina (P<0.05)» dove il valore P indica la *probabilità* di respingere una ipotesi zero vera.

Il metodo del chi-quadrato è utilizzabile quando il valore contenuto in ogni cella (celle a, b, c, d nella precedente Tabella 1) è >5, e il numero totale di osservazioni è >30; in caso contrario, occorre usare altri test (ad esempio, il test di Fisher, detto anche test esatto di Fisher o test delle probabilità esatte di Fisher).

Il test del chi-quadrato è uno dei tanti test di *significatività statistica* esistenti. Nella prossima Unità ne verrà trattato un altro: il test "t", che serve per confrontare due medie.
Ricordati comunque un principio generale valido sempre: qualsiasi test di significatività non può mai *provare con certezza* che una ipotesi zero è vera o falsa; esso può solo fornire una indicazione della forza con cui i dati contrastano l'ipotesi zero.

Un metodo di calcolo più semplice

Il sistema di calcolo del chi-quadrato ora fornito è piuttosto complicato, e costringe a generare - come abbiamo fatto nell'esempio - una nuova tabella con i valori «attesi». Esiste un altro tipo di calcolo, più semplice, che consente di ottenere il chi-quadrato direttamente dai valori osservati.
Tale calcolo è basato sulla seguente formula:

$$\chi^2 = \frac{(ad - bc)^2\, n}{(a+b)\,(a+c)\,(b+d)\,(c+d)}$$

Procediamo con il calcolo:

```
Calcolare il numeratore della formula:
  a*d = 52*21 = 1092
  b*c = 10*40 = 400
  1092-400 = 692
elevare al quadrato
 692^2 = 478864
e moltiplicare per il numero di osservazioni totale
 478864*123 = 58900272                      (1)

Ora calcoliamo il denominatore:
 (52+10)(52+40)(10+21)(40+21) = 10786264    (2)

Infine, dividiamo il numero trovato in (1) per quello
trovato in (2):
 58900272/10786264 = 5.46
```

Quando le frequenze attese sono basse (ma sempre >5) è consigliabile utilizzare una formula del chi-quadrato modificata secondo quanto proposto da F. Yates nel 1934:

$$\chi^2_{\text{corretto Yates}} = \frac{\left(|ad - bc| - \frac{n}{2}\right)^2 n}{(a+b)\,(a+c)\,(b+d)\,(c+d)}$$

I dati utilizzati nell'esempio sono fittizi e utilizzati esclusivamente a scopo didattico per il calcolo del chi-quadrato.
Il fatto che la differenza fra i due gruppi in studio sia risultata statisticamente significativa non implica necessariamente che, nella pratica clinica, la xmicina avrebbe sostituito la streptomicina nella terapia della leptospirosi del cane. Ad esempio, la xmicina potrebbe essere molto più tossica, oppure dotata di gravi effetti collaterali, oppure molto più costosa ecc.

Infine, ti ricordo che il test chi-quadrato si può estendere al confronto di più di due gruppi, con tabelle n x n. Però in tal caso il calcolo è diverso da quello dell'esempio.

Foglio di calcolo per Microsoft Excel® con un esempio di calcolo del chi-quadrato

Il mio consiglio: consolida quanto hai appreso risolvendo questo problema (si apre in una nuova finestra)

NELLA PROSSIMA UNITÀ:
si accenna a un caso diverso: il confronto fra due medie (anziché due percentuali). Attraverso un esempio e un foglio di calcolo, si illustra l'applicazione di uno dei test più frequentemente utilizzati per il confronto di due medie: il test t di Student.