

Esempio di calcolo di rischio relativo

B. Cenci Goga

Università degli Studi di Perugia, Dipartimento di Scienze Biopatologiche e Igiene delle Produzioni Animali e Alimentari, Sezione di Ispezione degli Alimenti di Origine Animale

<http://www.unipg.it/facvet>

Tra i parametri di frequenza delle malattie la misura appropriata in caso di malattie alimentari è il **rischio relativo**. Il rischio relativo (RR) esprime il rischio nel gruppo dei soggetti esposti in rapporto al rischio nei soggetti non esposti.

Il tasso di attacco è, invece, un caso particolare di incidenza. Si calcola come $D/(D+N)$, dove D è il numero di casi nel lasso di tempo considerato e N i soggetti a rischio.

	Ammalati	Sani	Totale
Esposti	a	b	(a+b)
Non esposti	c	d	(c+d)
Totale	(a+c)	(b+d)	t

rischio relativo:

tasso di attacco tra le persone che hanno consumato
l'alimento

tasso di attacco tra le persone a che NON hanno consumato
l'alimento

n. di persone ammalate che hanno mangiato l'alimento/n. totale di persone che hanno mangiato l'alimento

n. di persone ammalate che non hanno mangiato l'alimento/n. totale di persone che non ha mangiato l'alimento

ossia: $a/(a+b) / c/(c+d)$

	Ammalati	Sani	Totale
Esposti	a	b	(a+b)
Non esposti	c	d	(c+d)
Totale	(a+c)	(b+d)	t

Un RR di **1,0** significa che il rischio è **simile** nel gruppo degli esposti e dei non esposti e non è associato alla malattia.

Un RR **maggiore di 1,0** significa che il rischio è **maggiore** nel gruppo esposto e che l'esposizione può essere un fattore di rischio per la malattia.

Un RR **minore di 1,0** significa che il rischio è **inferiore** nel gruppo esposto e che l'esposizione potrebbe essere un fattore di protezione.

Il calcolo del chi-square (X quadrato) o del Fisher exact test serve invece a determinare la probabilità che il RR osservato potrebbe essersi verificato casualmente, nel caso la malattia non fosse collegata all'esposizione. Questa probabilità è il valore p (p -value). Un p -value molto piccolo significa che sarebbe estremamente improbabile una simile osservazione in assenza di reale collegamento tra esposizione e malattia. Se il p -value è minore di alcuni valori soglia predeterminati (di solito 0,05 ovvero una probabilità di 5 su 100) l'associazione tra esposizione e malattia si dice "statisticamente significativa".

Per il calcolo del c quadrato far riferimento alla formula generale:

$$X^2 = \sum (\text{frequenze osservate} - \text{frequenze attese})^2 / \text{frequenza attesa} = \sum (O - E)^2 / E$$



untitled



Previous

1st

First



Data



Statistics



Choose Test



Results



Graph



Explain My Choices



InStat Guide

What kind of data do you wish to enter?

A. Specify your goal:

- Compare means (or medians)
- Regression and correlation
- Analyze a contingency table

Example:

Compare numbers of pregnant and nonpregnant women in two age groups.

B. Choose a data entry format:

- Raw data
- Averaged data: Mean, SD & N
- Averaged data: Mean, SEM & N
- X and Y (or two or more Y replicates)
- Y and 2 or more X variables (multiple reg)
- Two rows, two columns
- Larger contingency table

Based on your choices, InStat will be able to perform these tests:

- Fisher's exact test
- Chi-square test
- Chi-square test with Yates's correction
- Relative risk (with 95% CI)
- Odds ratio (with 95% CI)
- Difference of two proportions (with 95% CI)
- Sensitivity, specificity, etc. with 95% CI

Next step: Enter data (click here)



untitled

Previous Next 1st

Statistics Choose Test Results Graph Arranging Data Importing Data

?

Table Title:

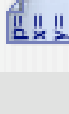
Titles	<input type="text" value="ammalati"/>	<input type="text" value="sani"/>	Total
esposti	<input type="text" value="9"/>	<input type="text" value="2"/>	11
non esposti	<input type="text" value="0"/>	<input type="text" value="10"/>	10
Total	9	12	21



untitled



1st



Previous

Next

Data

Statistics

Choose Test

Results

Graph

Help Me

Choose

Selecting Columns

1. Select a test

- Fisher's exact test (recommended)
- Chi-square test (less exact, but more widely known)

2. Use Yate's continuity correction?

- Yes (recommended)
- No

3. P value

- Two-sided (recommended)
- One-sided

4. Also calculate

- Relative risk, P1-P2, etc. (to analyze prospective, experimental and cross-sectional studies)
- Odds ratio (to analyze retrospective case-control studies)
- Sensitivity, specificity, predictive powers, etc. (to evaluate diagnostic tests)

Next step: Results (click here)





untitled

Previous Next First 1st Data Statistics Choose Test Results Graph Checklist What's Next?

Bologna Sauce

Chi-square Test

The two-sided P value is 0.8824, considered not significant. The row/column association is not statistically significant.

Note: With such small values, the chi-square P value is not accurate. Use Fisher's exact test instead.

Calculation details:

Chi-square statistic (with Yates correction) = 0.02187
Degrees of freedom = 1

Relative Risk

Relative risk = 2.500
95% Confidence Interval: 1.461 to 4.277
(using the approximation of Katz.)

Difference between the two proportions

Top row (esposti):
Fraction in the left column: 1.000
Bottom row (non esposi):
Fraction in the left column: 0.4000
95% Confidence Interval of that fraction: 0.1911 to 0.6393
Difference:
Difference between the fractions: 0.6000

The standard error and the confidence interval of the difference between proportions can only be calculated when each cell is greater than five.

Data analyzed

esposti	annalati 1 (5%)	sani 0 (0%)	Total 1 (5%)
non esposi	8 (38%)	12 (57%)	20 (95%)
Total	9 (43%)	12 (57%)	21 (100%)

* * *

Metambre

Chi-square Test

The two-sided P value is 0.0008, considered extremely significant.
The row/column association is statistically significant.

Note: With such small values, the chi-square P value is not accurate. Use Fisher's exact test instead.

Calculation details:

Chi-square statistic (with Yates correction) = 11.172
Degrees of freedom = 1

Relative Risk

Relative risk - Infinity
95% Confidence Interval: -Infinity to Infinity
(using the approximation of Katz.)

Difference between the two proportions

Top row (esposti):
Fraction in the left column: 0.8182
Bottom row (non esposti):
Fraction in the left column: 0.000
Difference:
Difference between the fractions: 0.8182

The standard error and the confidence interval of the difference between proportions can only be calculated when each cell is greater than five.

Data analyzed

	annalati	sani	Total
esposti	9 (43%)	2 (10%)	11 (52%)
non esposti	0 (0%)	10 (48%)	10 (48%)
Total	9 (43%)	12 (57%)	21 (100%)

* * *

Statistica descrittiva

La statistica descrittiva considera numeri che riassumono i dati e non i confronti.

Tre categorie:

1. **misure di tendenza centrale** (danno un'idea del valore medio di un numero o are quantità. Logicamente il concetto di media Può assumere diversi significati).

2. **misure di variabilità** (che indicano se le misurazioni sono raggruppate in un intervallo ristretto o distribuite in un intervallo più ampio).

3. **misure di distribuzione globale** (indicate da un singolo numero).

1. Tendenza centrale - La statistica descrittiva riassume la tendenza centrale con un solo numero.

1.1 Media

La media è la somma delle osservazioni divisa per il numero delle osservazioni. La somma delle differenze tra ciascuna osservazione e la media è zero. Ogni osservazione ha la sua importanza, cosicché possono insorgere delle difficoltà nel caso in cui i dati contengano i cosiddetti "outliers", cioè osservazioni molto distanti dalla massa dei dati. Gli outlier possono essere eliminati o corretti se derivano da errori ovvi nella raccolta dei dati, ma spesso sono determinanti e non vanno ignorati.

Esempio: 5 dipendenti

1 guadagna:	24.000 euro
1 guadagna:	24.000 euro
1 guadagna:	30.000 euro
1 guadagna:	36.000 euro
1 guadagna:	200.000 euro

La media è 62.800 euro che non è una stima accurata della media degli stipendi dell'azienda.

1.2 Mediana

Una misura alternativa è la mediana, cioè il valore centrale in un gruppo di osservazioni ordinate dal valore minore al maggiore. Nel caso che le osservazioni siano in numero pari, la mediana è la media dei due valori centrali. Per definizione metà delle osservazioni sono inferiori o uguali alla mediana e l'altra metà è superiore. Nel caso degli stipendi la mediana è 30.000 euro, una stima migliore dello stipendio "medio". L'effetto degli "outliers" è così eliminato, ma anche quello delle altre misurazioni, poiché è solo l'ordine che conta.

1.3 Trimmed mean

Un'alternativa alla mediana è la trimmed mean che si calcola eliminando i valori più distanti dal centro. Prima di effettuare i calcoli si decide la percentuale di dati da eliminare. Nell'esempio dello stipendio la media trimmed al 10% è 30.000 euro.

1.4 Moda

E' il valore più frequente un set di misurazioni. I dati raccolti in maniera continua (per esempio i pesi o le altezze) potrebbero non contenere osservazioni con lo stesso valore. In questo caso i dati vanno raggruppati per classi. Un altro inconveniente è la possibilità di ottenere più mode, condizione che rende difficile la scelta della moda più appropriata. In tutte quelle condizioni in cui la variabile può assumere solo un limitato numero di valori, la moda è uno strumento utilissimo. Nell'esempio degli stipendi, la moda è di 24.000 euro.

1.5 Media geometrica.

Sono misure meno comuni che misurano la tendenza centrale di una trasformazione matematica delle osservazioni originali. In molti casi infatti, i dati trasformati no proprietà statistiche migliori di quelle dei dati originali. Nel caso delle conte batteriche, dove la variabilità aumenta all'aumentare dei valori, il logaritmo della misurazione è più adatto all'elaborazione. La media geometrica è calcolata sulla base dei logaritmi delle misurazioni e quindi trasformata di nuovo nell'unità di misura iniziale.