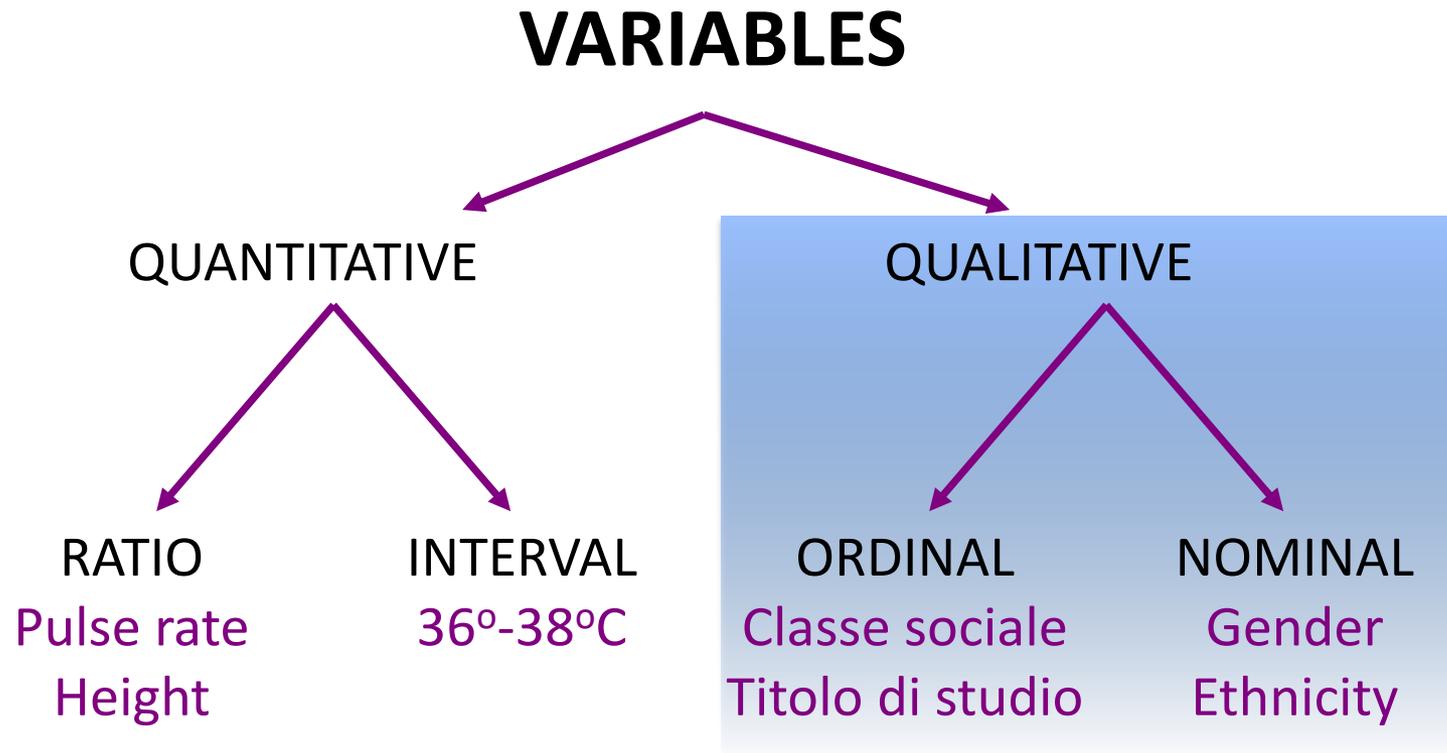


# LE TABELLE DI CONTINGENZA

# Tipi di dati in esame



# Come si rappresentano i dati per le variabili qualitative

La classificazione di un insieme di soggetti secondo **due criteri o variabili dicotomiche** (ciascuno dei quali avente due livelli di classificazione o modalità) può essere rappresentata da una tabella 2 x 2.

		Variabile 1		Totale
		1	2	
Variabile 2	1	a	b	a+b
	2	c	d	c+d
Totale		a+c	b+d	n

# Tabelle di contingenza

Una **tabella di contingenza** è una tabella di frequenza a doppia entrata in cui vengono incrociate due variabili ***qualitative***.

Variabile di  
COLONNA

FREQUENZA dei soggetti  
che rispondono alle due  
modalità che si incrociano

Variabile  
di RIGA

TRATTAMENTO	ESITO	Guariti	Non Guariti	Totale
Antibiotico 1				
Antibiotico 2				
Totale				

N= Numero di soggetti  
effettivamente osservati  
in questo esperimento

# Tabelle di contingenza 2

---

Quando si confrontano variabili categoriali non possiamo usare i parametri, *MEDIA e D.S.*, ma le percentuali (%).

Quando il risultato è una

malattia vs nessuna malattia

successo vs fallimento

guarito vs non guarito

è opportuno confrontare le percentuali di ogni categoria utilizzando **test non parametrici**.

# Confronto dell'efficacia di due antibiotici

TRATTAMENTO	ESITO			Totale
Antibiotico 1				
Antibiotico 2				
Totale				

Dei 62 bambini  
trattati  
con Antib1, 52  
risultano guariti:  
 $52/62 = 0,84$ , l'84%

Dei 61 bambini  
trattati  
con Antib2, 40  
risultano guariti:  
 $40/61 = 0,66$ , il 66%

# Confronto dell'efficacia di due antibiotici 2

TRATTAMENTO	ESITO	Guariti	Non Guariti	Totale
Antibiotico 1		a 52	b 10	62
Antibiotico 2		c 40	d 21	61
Totale		92	31	123

Dei 62 bambini trattati con Antib1, 52 risultano guariti:  $52/62 = 0,84$ , l'84%

Dei 61 bambini trattati con Antib2, 40 risultano guariti:  $40/61 = 0,66$ , il 66%

## QUESITO - Ipotesi H0 vs H1

- Esiste una relazione di **dipendenza** tra le due variabili?
- **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?



oppure



# .....Schema logico da seguire

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».

**Valori Osservati**

TRATTAMENTO	ESITO	Guariti	Non Guariti	Totale
Antibiotico 1 		52		62
Antibiotico 2 		40		61
<b>Totale</b>		<b>92</b>	<b>31</b>	<b>123</b>

**Valori Attesi**

TRATTAMENTO	ESITO	Guariti	Non Guariti	Totale
Antibiotico 1 		?		62
Antibiotico 2 		?		61
<b>Totale</b>		<b>92</b>	<b>31</b>	<b>123</b>

# Calcolo dei valori attesi

## Valori OSSERVATI

TRATTAMENTO	ESITO	Guariti	Non Guariti	Totale
Antibiotico 1		 52	 10	62
Antibiotico 2		 40	 21	61
Totale		92	31	123

**92/123=0,748 (74,8%) risultano guariti indipendentemente dall'antibiotico usato**

Applicando questa percentuale di successo (74.8%) ai valori della tabella

## Valori ATTESI

TRATTAMENTO	ESITO	Guariti	Non Guariti	Totale
Antibiotico 1		 a	 b	62
Antibiotico 2		 c	 d	61
Totale		92	31	123

  $= 62 * 74.8 / 100 = 46.37$

  $= 62 - 46.37 = 15.63$

  $= 61 * 74.8 / 100 = 45.63$

  $= 61 - 45.63 = 15.37$

## Verifica delle ipotesi e applicazione del test statistico

Il valore del chi-quadrato quantifica la differenza fra i dati **Osservati (O)** e quelli **Attesi (A)**, ed è la somma delle quattro celle a, b, c e d, per ciascuna delle quali si calcola il valore della frazione:

$$\chi^2 = \sum_{i=1}^{r \times c} \frac{(O_i - A_i)^2}{A_i}$$

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

È evidente che il **chi-quadrato aumenta con l'aumentare della differenza dei dati posti a raffronto**. Se esso supera certi valori la differenza viene ritenuta significativa; in caso contrario, non si può affermare l'esistenza di una significativa differenza tra i due eventi considerati.

Nel caso specifico, 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.

## Verifica delle ipotesi e applicazione del test statistico

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%.

Tabella dei valori di  $\chi^2$

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.	...	...

# Conclusione in base ai risultati del test

- Si può concludere che la differenza tra bambini trattati con Antibiotico1 e quelli trattati con Antibiotico2 è statisticamente significativa al livello di probabilità 5%.
- In sostanza: in base ai risultati del test del chi-quadrato, l'affermazione «**Antibiotico1 è più efficace dell'Antibiotico2**» 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, **Antibiotico1** è risultata più attiva **dell'Antibiotico2** ( $P < 0.05$ )» *dove il valore P indica la probabilità di respingere una ipotesi zero vera.*



## RIASSUMENDO: **COME** UTILIZZARE IL TEST

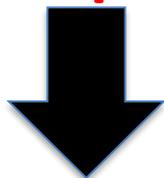
- 1.** Tabulare i dati in una tabella di contingenza, in cui compaiano i totali marginali (tabella delle frequenze osservate).
- 2.** Stabilire l'ipotesi nulla  $H_0$  e l'ipotesi alternativa  $H_1$ .
- 3.** Calcolare le frequenze attese, sotto l'ipotesi che  $H_0$  sia vera.
- 4.** Sulla base delle frequenze osservate e attese, calcolare il test  $X^2$ .
- 5.** Calcolare i gradi di libertà relativi alla tabella di contingenza.
- 6.** Cercare sulla tabella sinottica del  $X^2$  il valore critico per il valore di  $\alpha$  prefissato e per i gradi di libertà calcolati. (ora ci sono i pc!!!)
- 7.** Applicare la regola di decisione statistica e trarre la conseguente conclusione.



## RIASSUMENDO: **QUANDO** UTILIZZARE IL TEST

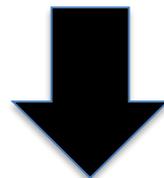
Esistono due metodi per calcolare LA PROBABILITÀ DELL'IPOTESI NULLA (valore P) da tabelle con due righe e due colonne (tabelle di contingenza):

### **test chi quadrato ( $\chi^2$ ) e test esatto di Fisher**



Il **test del chi-quadrato** è utilizzabile quando :

1. il **valore contenuto in ogni cella** (celle a, b, c, d ) è  $>5$
2. il **numero totale** di osservazioni è  $>30$ ;



Il **test esatto di Fisher** è utilizzabile quando si hanno piccoli campioni  $n < 30$



**N.B.**

**NB: Raccomandiamo di scegliere sempre il test di Fisher, in quanto calcola un valore P che è esattamente corretto**

- L'unico vantaggio del test del Chi-quadrato è che è più facile da calcolare a mano, e quindi è più noto (MA ORA ABBIAMO COMPUTER !!!)



## Correzione di Yates

Se scegli un test Chi quadrato, scegli anche **se applicare la correzione di continuità di Yates**. *Qualsiasi cella contenente un numero inferiore a 5, probabilmente stimerà il valore del Chi-quadrato.*

Questa correzione è progettata per rendere i risultati approssimativi di un test chi-quadrato più accurato con piccoli campioni

$$C^2 = \frac{\sum [(O - E) - 0.5]^2}{E}$$

**NB:**

- ***Se si seleziona sempre il test esatto di Fisher (consigliato), la correzione di Yates non è un problema***
- ***Se la tabella include numeri molto grandi (migliaia), è preferibile il test chi quadrato, poiché è molto accurato per campioni di grandi dimensioni e la correzione di continuità di Yates ha un effetto trascurabile.***

N.B.





## I dati sono stati spaiati?

In alcuni esperimenti, i soggetti sono abbinati per età e altre variabili. Un soggetto in ogni coppia riceve un trattamento mentre l'altro soggetto riceve l'altro trattamento. Questi dati dovrebbero essere analizzati con metodi speciali come il test di McNemar

- **I dati accoppiati non devono essere analizzati con il chi quadrato del test di Fisher**

## Utilizzo di alcuni rapporti comunemente utilizzati per misurare la forza di una associazione.

---

**Sep 1.** Sono state illustrate le basi razionali e qualche metodo per valutare la significatività statistica di una associazione.

**Step 2.** La **misura (o quantificazione) di una associazione** rappresenta uno fra i passi più importanti da compiere nell'indagine sulle cause delle malattie oppure nella valutazione degli effetti di un trattamento terapeutico o di una azione di prevenzione od ancora, più in generale, nell'investigazione su un qualsiasi rapporto di causa/effetto.

## Utilizzo di alcuni rapporti comunemente utilizzati per misurare la forza di una associazione.

Ora è venuto il momento di prendere in considerazione un paio di sistemi utili per quantificare una associazione.

Parlando in generale, nel procedimento di valutazione dell'esistenza di una associazione sono in gioco due variabili:



la presunta **causa o esposizione** (variabile indipendente)



**l'effetto** (variabile dipendente, così detta in quanto appunto subordinata alla variabile indipendente)/ **comparsa della malattia**

## Studi RETROSPETTIVI e studi PROSPETTICI

L'esistenza di una associazione può essere accertata attraverso studi retrospettivi (o studi caso-controllo) o studi prospettivi (o studi di coorte).

**STUDI RETROSPETTIVI** si confronta la frequenza di esposizione nei malati («casi») con quella nei non-malati («controlli»).



**STUDI PROSPETTICI** si confronta la frequenza di malattia negli esposti con quella nei non-esposti:



**Esposizione = presenza di un fattore che può (ipoteticamente) causare un certo effetto**

## Studi RETROSPETTIVI e studi PROSPETTIVI



Come si fa a *quantificare* l'associazione eventualmente riscontrata, ossia a **quantificare il rischio** cui sono soggetti , ad esempio, i bambini esposti?

Dipende dallo studio:

Studi Retrospettivi → **ODDS RATIO !**

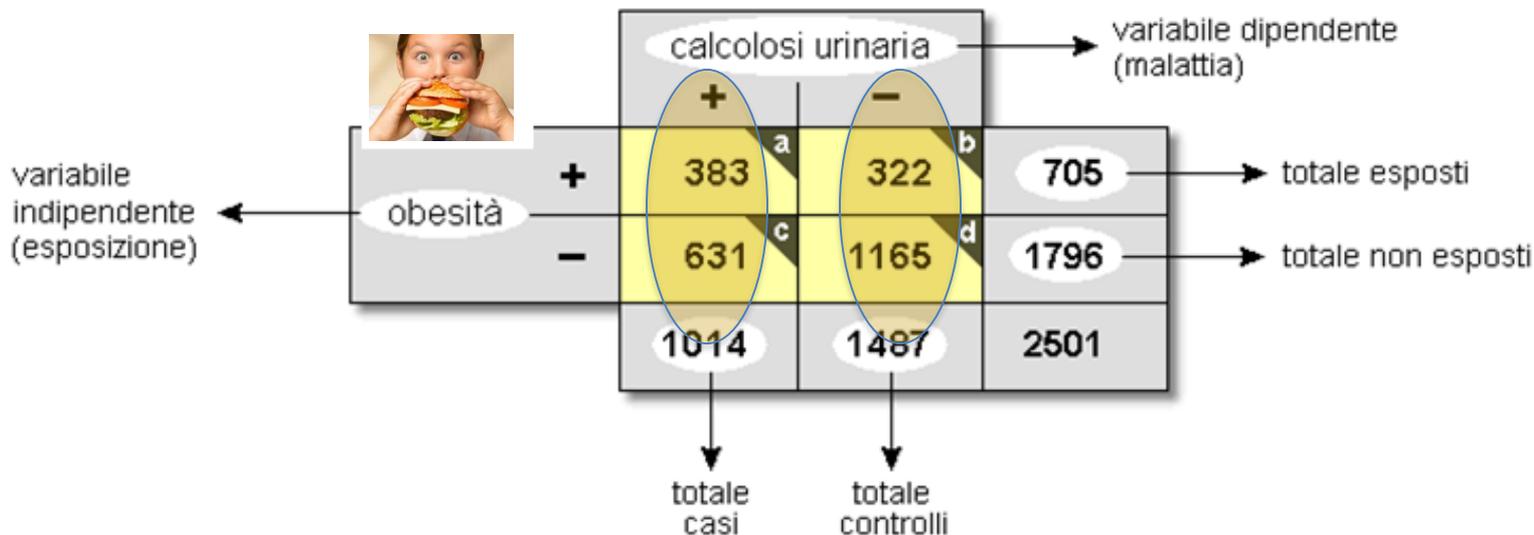
Studi Prospettivi → **RISCHIO RELATIVO !**

## Studio retrospettivo: l'*odds ratio*

Prendiamo come esempio l'associazione fra obesità e urolitiasi, ricordando che in uno studio retrospettivo si inizia selezionando i casi e i controlli e poi andando ad accertare quanti fra i casi (e quanti fra i controlli) sono stati esposti alla presunta causa.

Lo studio ha fornito i seguenti risultati

Risultati dello studio retrospettivo



The diagram illustrates a retrospective study contingency table. It features a 2x2 grid of cells representing the relationship between obesity (independent variable) and urinary calculi (dependent variable). The cells are labeled with letters a, b, c, and d. Marginal totals are provided for both rows and columns. An arrow points from a small image of a child eating a burger to the 'obesità +' row. Another arrow points from the text 'variabile dipendente (malattia)' to the 'calcolosi urinaria' header.

		calcolosi urinaria		
		+	-	
obesità	+	383 (a)	322 (b)	705 (totale esposti)
	-	631 (c)	1165 (d)	1796 (totale non esposti)
		1014 (totale casi)	1487 (totale controlli)	2501

## Studio retrospettivo: l'*odds ratio*

L'*odds ratio* si calcola attraverso i semplici rapporti (*odds*) fra le frequenze osservate e non attraverso le *proporzioni*.

Nel nostro esempio sulla urolitiasi, calcoliamo:

- gli *odds* di esposizione nel gruppo dei casi
  - gli *odds* di esposizione nel gruppo dei controlli
- e poi ne facciamo il rapporto.

(ricordati:  $\text{odds} = p \text{ a favore} / p \text{ contro}$ )

		urolitiasi	
		SI	NO
obesità	SI	383	322
	NO	631	1165

**ODDS di esposizione  
nei casi:**  
 $383 / 631 = 0.607$

**ODDS di esposizione  
nei controlli:**  
 $322 / 1165 = 0.277$

**ODDS RATIO =  $0.607 / 0.277 = 2.19$**

Calcolo alternativo:

$(383 \cdot 1165) / (322 \cdot 631) = 2.19$

## Studio prospettivo: il rischio relativo

Diversamente da uno studio retrospettivo, uno studio prospettivo inizia suddividendo la **popolazione in esposti e non esposti** e poi osservando nel tempo quanti fra gli esposti (e quanti fra i non-esposti) si ammalano.

Risultati dello studio prospettivo

		polmonite		
		+	-	
variabile indipendente (esposizione)	chiuso	240 (a)	230 (b)	470 (totale esposti)
	aperto	160 (c)	1070 (d)	1230 (totale non esposti)
		400 (totale casi)	1300 (totale controlli)	1700

## Studio prospettivo: il rischio relativo

Il rischio relativo (RR) (anche detto relative risk o risk ratio) è il rapporto tra la probabilità che si verifichi un evento in un gruppo esposto (ad un trattamento, ad un fattore di rischio, ad un fattore protettivo), e la probabilità che si verifichi lo stesso evento in un gruppo di non esposti (allo stesso trattamento, fattore di rischio, fattore protettivo)

		Polmonite		
		SI	NO	
Ambiente	chiuso	240	230	470
	aperto	160	1070	1230
		400	1300	1700

**INCIDENZA  
negli esposti:**  
 $240 / (240+230) = 0.511$

**INCIDENZA  
nei non esposti:**  
 $160 / (160 + 1070) = 0.130$

**RISCHIO RELATIVO =  $0.511 / 0.130 = 3.93$**

# Odds Ratio e Rischio Relativo: Interpretazione dei risultati

Se **RR=0**: non c'è associazione tra la presenza o meno del fattore di rischio e la malattia

Se **RR = 1** il rischio di contrarre la malattia è uguale per i soggetti esposti e per i soggetti non esposti al fattore di rischio.

Se **RR>1** il rischio di contrarre la malattia è maggiore tra i soggetti esposti.  
In questo caso **l'esposizione è dannosa**.

Se **RR<1** il rischio di contrarre la malattia è minore tra i soggetti esposti.  
In questo caso **l'esposizione è protettiva**.

Schema di interpretazione dei valori di Rischio Relativo e Odds Ratio



# Odds Ratio e Rischio Relativo: Interpretazione dei risultati

---

## IC, proporzioni e significatività statistica

Se invece di misurare una media in un campione si compara l'associazione fra una esposizione e un esito, la stima del rischio nel campione selezionato non coinciderà con il rischio reale (e sconosciuto) dell'intera popolazione, quindi si ricorrerà anche in questo caso all'IC.

In uno studio condotto su 1700 bambini si registra un rischio di polmonite nei bambini che vivono prevalentemente in ambiente chiuso pari a RR: 3.93 (IC 95%: 3.31, 4.65). Questo significa che i bambini che vivono in ambiente chiuso hanno 3 volte più probabilità di ammalarsi di polmonite rispetto a quelli che vivono prevalentemente in ambiente aperto. Nel campione studiato, la stima puntuale del rischio 3.93; è giusto sostenere, con una fiducia del 95%, che il valore di questo rischio, trasferito nella popolazione generale, è compreso fra 3.31 e 4.65.

## MANTEL-HAENSZEL - Variabile di confounding

---

Nello studio della relazione tra una data malattia e un presunto fattore di rischio, può capitare che vi sia **un'altra variabile** (associata alla malattia, al fattore di rischio o ad entrambi), **che può falsare la vera relazione** tra le due variabili.

La tecnica di Mantel-Haenszel consente di controllare la variabile di confounding, in modo da ottenere una valutazione non ambigua della relazione tra malattia e fattore di rischio.

Come procedere? I soggetti, casi o controlli, vengono assegnati a strati, che corrispondono alle diverse modalità della variabile di confounding. La variabile di confounding può essere categoriale o continua; se è continua deve essere categorizzata.