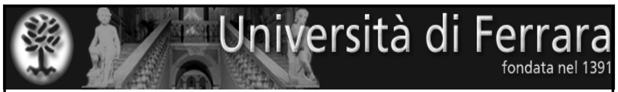


LA CAUSALITA'

LA CAUSALITA'

- **OBIETTIVO: illustrare il procedimento logico che guida la dimostrazione della causalità**

LA CAUSALITA'

- Gli studi osservazionali sono fondamentali in epidemiologia e vengono utilizzati frequentemente sia in medicina umana che in medicina veterinaria allo scopo di individuare i determinanti delle malattie.
- Occorre valutare i principi essenziali ed il flusso logico del ragionamento che viene seguito in questo tipo di studi.

LA CAUSALITA'

- Nel caso in cui si voglia tentare di verificare se un fattore è uno dei determinanti di una malattia.
- In uno studio di questo tipo le variabili in gioco sono solo due:
  - 1. la presunta causa
  - 2. la malattia
- La presunta causa viene detta variabile indipendente; la malattia viene detta variabile dipendente, in quanto *dipende* appunto dalla variabile indipendente.

LA CAUSALITA'

- Il ragionamento che conduce alla dimostrazione di un rapporto causa-effetto fra variabile indipendente e variabile dipendente può essere schematizzato così………

LA CAUSALITA'

- I stadio. È necessario accertare se la variabile indipendente (causa) è statisticamente associata con la variabile dipendente (malattia)
- II stadio. Se le variabili sono statisticamente associate, bisogna accertare che le due variabili siano causalmente associate
- III stadio. È la fase finale in cui, sulla base del rapporto causale tra le due variabili dimostrato nelle due fasi precedenti, possono essere effettuate elaborazioni sulla natura e sulle conseguenze dell'associazione, utilizzando modelli teorici, simulazioni, esperimenti in laboratorio o in campo, ecc.

LA CAUSALITA'

- Quando si può agire sulla variabile indipendente (studi sperimentali) e osservare le conseguenti modificazioni sulla variabile dipendente, si possono trarre conclusioni su quanto tali modificazioni nella variabile dipendente siano causate da quelle verificatesi nella variabile indipendente

LA CAUSALITA'

- Quando si possono soltanto osservare la contemporanea variazione delle due variabili (studi osservazionali), si può solo definire una associazione tra di esse, dimostrata dal fatto che al modificarsi dell'una si modifica anche l'altra.

LA CAUSALITA'

- In questo caso non si può escludere l'eventualità che entrambe le variabili siano, in maniera indipendente l'una dall'altra, influenzate da una terza variabile e l'apparente relazione causa-effetto sia dovuta ad un effetto detto di confondimento

ASSOCIAZIONE CAUSALITA'

- OBIETTIVO: definire il significato del termine "associazione".

ASSOCIAZIONE

- In statistica il termine associazione assume un significato diverso da quello comune
- Nel linguaggio parlato due eventi si dicono "associati" quando compaiono o si verificano in correlazione l'uno con l'altro
- In statistica per "associazione" si intende il grado di dipendenza tra due o più eventi o variabili
- Associazione = grado di dipendenza *statistica* tra due o più eventi o variabili

ASSOCIAZIONE

- Esistono *metodi* che consentono di escludere (con un certo grado di probabilità, ma non con certezza) che una eventuale associazione sia dovuta al caso.
- Anche quando la statistica afferma che "l'associazione non è casuale" (allora si dice che esiste una differenza significativa), resta ancora da dimostrare che i due fattori siano legati da un rapporto causa-effetto.
- Associazione non è sinonimo di causalità.

IL TEST STATISTICO

- è una “procedura di calcolo”
- permette di rifiutare un’ipotesi.
- input
  - dati (un campione)
- output
  - un consuntivo (*test statistico*)
- decisione
  - in base al consuntivo si rifiuta oppure non si rifiuta l’ipotesi

Frequenze percentuali e loro errore standard

- Frequenza assoluta

numero di volte in cui un dato avvenimento si manifesta

- Frequenza relativa

il quoziente fra il numero di prove in cui un dato evento si manifesta e il numero totale di prove

- Frequenza percentuale

frequenza relativa per 100

Frequenze percentuali
e loro errore standard

- Frequenza assoluta

numero di volte in cui un dato avvenimento si manifesta
N° nati maschi = 252

- Frequenza relativa

il quoziente fra il numero di prove in cui un dato evento si manifesta e il numero totale di prove
N° nati maschi 252 e totale nati 512
252/512 = 0,492

- Frequenza percentuale

frequenza relativa per 100
0,492 * 100 = 49,2%

Frequenze percentuali e loro errore standard

- Qualunque percentuale osservata è affetta da un errore di campionamento.
- L’intervallo entro cui può variare una percentuale osservata su un campione di n osservazioni si può valutare in base al suo errore standard.
- $s_p = \pm \sqrt{P(100-P) / n}$

ove P è la percentuale “attesa o teorica” ed n il numero di osservazioni

- L'errore standard non definisce un intervallo assoluto di errore al di là del quale non può mai cadere una percentuale osservata.

- L'errore standard definisce un intervallo entro il quale una percentuale osservata può cadere con una certa probabilità.

- La percentuale osservata potrà cadere con probabilità 2/3 entro ± 1 volte l'errore standard e con probabilità del 95% entro ± 2 volte l'errore standard.

CONFRONTARE DUE PROPORZIONI O PERCENTUALI IL TEST DEL CHI QUADRO

- Un **test di significatività statistica** è lo strumento indispensabile nel confronto fra due gruppi o popolazioni riguardo ad un parametro
- Uno dei test più comuni è il «**chi-quadrato**» e permette di confrontare due percentuali, allo scopo di verificare se la loro differenza è **probabilmente** dovuta al caso oppure no
- Se la differenza non è dovuta al caso, si dice che essa è «statisticamente significativa»

Confronto tra due percentuali osservate

Lo scopo è quello di stabilire se le due percentuali osservate differiscono soltanto per errore di campionamento casuale, oppure se è lecito ritenere che vi sia una differenza obiettiva tra i due fenomeni di cui queste due percentuali misurano la probabilità di manifestazione.

In quest'ultimo caso si dice che la differenza fra le due percentuali è statisticamente significativa.

CHI QUADRATO

- Lo schema logico da seguire:
  - Inizialmente, qualsiasi sia la differenza esistente tra le due percentuali da confrontare, si avanza l'ipotesi zero (o ipotesi nulla) che afferma semplicemente che la differenza osservata è dovuta al caso
  - Tale ipotesi può essere 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 quello del **chi-quadrato**

CHI QUADRATO

- Confrontare due proporzioni o percentuali
  - Ipotesi 0: la differenza è dovuta al caso
    - Accettare o rifiutare l'ipotesi 0 ?

CHI QUADRATO

- I dati usati nel chi quadro debbono avere scala Nominale.
- Il metodo del chi-quadrato è utilizzabile quando:
  - il valore contenuto in ogni cella (vedi tabelle successive: celle a, b, c, d) è >5
  - il numero totale di osservazioni è >30 (>40)
- In caso contrario, occorre usare il **test di Fisher**

Affinche' si possa utilizzare il chi quadrato e' indispensabile:

a) che i dati siano indipendenti, cioe' nessun soggetto puo' apparire in piu' di una cella della tabella;

b) che non piu' del 20 % delle frequenze attese nella tabella sia < 5 (altrimenti si deve usare il test esatto di Fisher);

c) nessuna cella deve avere una frequenza attesa < 1 (altrimenti si deve usare il test esatto di Fisher).

CHI QUADRATO: UN ESEMPIO

- Supponiamo di voler mettere a confronto l'efficacia, nella terapia di una malattia, un nuovo antibiotico (Farmaco A) con un antibiotico già in uso (Farmaco B).
- Intraprendiamo un test clinico su una popolazione rappresentata dai soggetti affetti dalla malattia che si presentano in alcuni ambulatori ed ospedali in un determinato periodo di tempo
- Durante la sperimentazione, ogni soggetto viene assegnato a caso al gruppo dei trattati con il Farmaco A oppure a quello dei trattati con il Farmaco B
- Alla fine della sperimentazione, abbiamo ottenuto i dati riassunti nella seguente tabella.......

UN ESEMPIO

	Guariti	Non guariti	Totali	
Farmaco A	a=52	b=10	62	52/62=84%
Farmaco B	c=40	d=21	61	40/61= 66%
Totali	92	31	123	

UN ESEMPIO

- In particolare, si nota come, su un totale di 123 soggetti, 62 siano stati sottoposti a trattamento con il Farmaco A e, fra questi, si siano registrati 52 casi di guarigione (84%)
- Fra i restanti 61 soggetti, trattati con Farmaco B, ne sono guariti 40 (66%)
- I dati grezzi sembrano indicare che il Farmaco A sia più efficace del Farmaco B

UN ESEMPIO

- Tuttavia, prima di giungere ad una conclusione affrettata, occorre rispondere alla seguente domanda
  - *ammesso che non esistano differenze nell'efficacia dei due trattamenti, che probabilità c'è di osservare - in uno studio di dimensioni simili a questo - differenze uguali o superiori a quelle che abbiamo osservato?*

UN ESEMPIO

- 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*
- I dati dimostrano che complessivamente il trattamento è risultato efficace nel 74.8% dei casi (52+40=92 soggetti su 123 trattati)

UN ESEMPIO

- Applichiamo questa percentuale di successo (74.8%) a ciascuno dei due gruppi di soggetti in esame e avremo:

DATI OTTENUTI

	G	NG	Tot
Far. A	52	10	62
Far. B	40	21	61
Tot	92	31	123

DATI ATTESI

	G	NG	Tot
Far. A	46	16	62
Far. B	46	15	61
Tot	92	31	123

UN ESEMPIO

- Il valore 46 è stato ottenuto assumendo una percentuale di guarigione del 74.8% nei 62 soggetti trattati con Farmaco A
- Analogamente, ci si sarebbe aspettata la guarigione del 74.8% dei 61 soggetti trattati con Farmaco B ossia di 46 soggetti
- I valori delle altre celle sono stati ottenuti per differenza
- Il valore del chi-quadrato, che quantifica la differenza fra i numero osservati e quelli attesi, è la somma delle quattro celle a, b, c e d, per ciascuna delle quali si calcola il valore della frazione

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

CHI QUADRATO

- Il valore del chi-quadrato è determinato dalla differenza fra i numeri osservati ed 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

CHI QUADRATO

- **Pertanto, il chi-quadrato viene calcolato come segue:**

$$\chi^2 = \frac{(52,46,37)^2}{46,37} + \frac{(10-15,63)^2}{15,63} + \frac{(40-45,63)^2}{45,63} + \frac{(21,15,37)^2}{15,37} = 5,46$$

CHI QUADRATO

1. È evidente che il chi-quadrato aumenta con l'aumentare della differenza dei dati posti a raffronto

CHI QUADRATO

2. Se esso supera certi valori (tabella valori di chi-quadrato), la differenza viene ritenuta significativa

CHI QUADRATO

3. In caso contrario, non si può affermare l'esistenza di una significativa differenza tra i due fenomeni considerati

Tabella dei valori del c^2

Gradi di libertà	Probabilità di un valore del chi-quadrato per caso								
	0,995	0,975	0,9	0,5	0,1		0,025		0,005
1	0,000	0,001	0,016	0,455	2,706		5,024		7,879
2	0,010	0,051	0,211	1,386	4,605		7,378		10,597
3	0,072	0,216	0,584	2,366	6,251		9,348		12,838
4	0,207	0,484	1,064	3,357	7,779		11,143		14,860
5	0,412	0,831	1,610	4,351	9,236		12,833		16,750
6	0,676	1,237	2,204	5,348	10,645		14,449		18,548
7	0,989	1,690	2,833	6,346	12,017		16,013		20,278
8	1,344	2,180	3,490	7,344	13,362		17,535		21,955
9	1,735	2,700	4,168	8,343	14,684		19,023		23,589
10	2,156	3,247	4,865	9,342	15,987		20,483		25,188
11	2,603	3,816	5,578	10,341	17,275		21,920		26,757
12	3,074	4,404	6,304	11,340	18,549		23,337		28,300
13	3,565	5,009	7,042	12,340	19,812		24,736		29,819
14	4,075	5,629	7,790	13,339	21,064		26,119		31,319
15	4,601	6,262	8,547	14,339	22,307		27,488		32,801
16	5,142	6,908	9,312	15,338	23,542		28,845		34,267
17	5,697	7,564	10,085	16,338	24,769		30,191		35,718
18	6,265	8,231	10,865	17,338	25,989		31,526		37,156
19	6,844	8,907	11,651	18,338	27,204		32,852		38,582
20	7,434	9,591	12,443	19,337	28,412		34,170		39,997
21	8,034	10,283	13,240	20,337	29,615		35,479		41,401
22	8,643	10,982	14,041	21,337	30,813		36,781		42,796
23	9,260	11,689	14,848	22,337	32,007		38,076		44,181
24	9,886	12,401	15,659	23,337	33,196		39,364		45,559
25	10,520	13,120	16,473	24,337	34,382		40,646		46,928
30	13,787	16,791	20,599	29,336	40,256		46,979		53,672
35	17,192	20,569	24,797	34,336	46,059		53,203		60,275
40	20,707	24,433	29,051	39,335	51,805		59,342		66,766
50	27,991	32,357	37,689	49,335	63,167		71,420		79,490

CHI QUADRATO

- Non resta che confrontare il valore ottenuto con la Tabella dei valori di chi-quadrato
- Nel nostro caso, il valore ottenuto è un chi-quadrato con 1 grado di libertà
- Il grado di libertà è uguale a (numero di righe-1)*(numero di colonne-1)

CHI QUADRATO

- Ora, confrontando il nostro valore (5.46) con quelli tabulati, notiamo 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%.

Valori di χ^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
...	...	...

CHI QUADRATO

- Si può concludere che la differenza tra soggetti trattati con Farmaco A e quelli trattati con Farmaco B è statisticamente significativa al livello di probabilità 5%

CHI QUADRATO

1. ammettendo che i due farmaci 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 al farmaco A
2. In base ai risultati del test del chi-quadrato, l'affermazione il Farmaco A è più efficace del Farmaco B ha il 95% di probabilità di essere vera (e quindi ha il 5% di probabilità di essere falsa)

TEST DI SIGNIFICATIVITA'

- Il test del chi-quadrato è uno dei tanti test di significatività statistica esistenti
- Da ricordare che un qualsiasi test di significatività non può mai provare con certezza che una ipotesi zero è vera o falsa
- Può solo fornire una indicazione della forza con cui i dati contrastano l'ipotesi zero

CHI QUADRATO (CALCOLO SEMPLICE)

- Il sistema di calcolo del chi-quadrato ora fornito è piuttosto complicato, e costringe a generare 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

LA FORMULA

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

CORREZIONE DI YATES

- Quando le frequenze attese sono basse (ma sempre >5) è consigliabile utilizzare una formula modificata
- Nel caso di tabelle 2 x 2 si deve applicare un piccolo correttivo, detto di Yates, che consiste nel sottrarre 1/2 (cioe' 0.5) ad ogni valore di O-A

CHI QUADRATO CORREZIONE DI YATES

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

TEST ESATTO DI FISHER

- Quando le dimensioni campionarie sono piccole, è possibile elencare tutte le possibili combinazioni delle osservazioni e quindi calcolare le probabilità esatte associate a ogni possibile combinazione di dati

TEST ESATTO DI FISHER

Obs	b1	b2	
a1	1	8	9
a2	10	4	14
	11	12	23

Exp	b1	b2	
a1	4,3	4,7	9
a2	6,7	7,3	14
	11	12	23

χ^2 in tabelle m x n

- Spesso occorre confrontare fra loro numerose terapie o trattamenti sperimentali

- La classificazione dei risultati, pur essendo sempre qualitativa, può avvenire secondoi più di 2 categorie

- In questo caso i dati dell'esperimento possono riassumersi in una tabella detta m x n, cioè con m trattamenti e n categorie

χ^2 in tabelle m x n

		Esito		
terapia	Buono	Mediocre	Cattivo	Totale
A	3	2	2	7
B	4	8	6	18
C	15	5	5	25
Totale	22	15	13	50

χ^2 in tabelle m x n

		Esito		
terapia	Buono	Mediocre	Cattivo	Totale
a	b	c	d	n1
e	f	g	h	n2
i	j	k	l	n3
m1	m2	m3	m4	N

χ^2 in tabelle m x n

		Esito		
terapia	Buono	Mediocre	Cattivo	Totale
A	3	2	2	7
B	4	8	6	18
C	15	5	5	25
Totale	22	15	13	50

In questo esempio i trattamenti sono 3 ed i tipi di esiti sono 3
Perciò, i gradi di libertà sono = (3-1) * (3-1) = 4
c^2 = 6,22

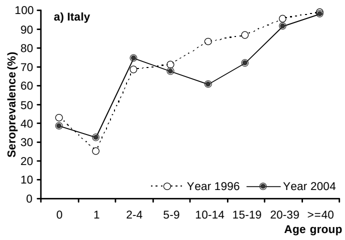
CHI QUADRATO

- Ora, confrontando il nostro valore (6,22) con quelli tabulati, notiamo che esso è < 9,488
- Ciò consente di dire che il chi quadrato non è significativo e perciò i dati non sono sufficienti per concludere che vi è differenza tra le terapie in esame.

Valori di χ^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
...	...	...

Morbillo in Italia: sieroprevalenza per classi di età, 1996 e 2004



Measles serological survey in the Italian population: interpretation of results by qualitative methods
MC Rota, M Massari, G Gabutti, M Guido, A De Donno, MLCiofi degli Atti on behalf of the Serological Study Group (Vaccine, 2008)

Morbillo in Italia: sieroprevalenza per classi di età, 2004

- Vogliamo verificare se la differenza nel tasso di sieroprevalenza rilevato nelle classi di età 1aa e 2-4 aa è significativo
- Per prima cosa dobbiamo impostare una tabella 2 x 2

ESEMPIO

	Con Ab	Senza Ab	Totale	
Classe 1 aa	31	62	94	33,3 %
Classe 2-4 aa	194	59	253	76,7 %
Totale			347	

CHI QUADRATO

- Ora, confrontando il nostro valore (56,18) con quelli tabulati, notiamo che esso è > 6,635
- Ciò consente di dire che il chi quadrato è significativo (p<0.01)

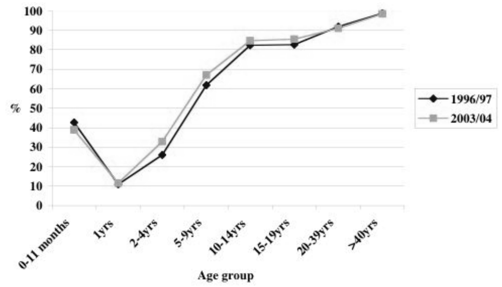
Valori di χ^2

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
...	...	...

Varicella in Italia

Figure 4

Comparison between VZV seroprevalence by age-group in Italy, 1996 and 2004



The Epidemiology of Varicella Zoster Virus Infection in Italy. Gabutti G, Rota MC, Guido M, De Donno A, Bella A, Ciofi degli Atti ML, Crovari P, and the Seroepidemiology Group (BMC Public Health 2008)

Varicella in Italia: sieroprevalenza per classi di età, 1996 e 2004

- Vogliamo verificare se la differenza nel tasso di sieroprevalenza rilevato nelle classi di età 2-4 aa è significativo
- Per prima cosa dobbiamo impostare una tabella 2 x 2

ESEMPIO

	Con Ab	Senza Ab	Totale	
1996	96	196	292	32,9 %
2004	87	248	335	26 %
Totale			627	

CHI QUADRATO

- Ora, confrontando il nostro valore (3,601) con quelli tabulati, notiamo che esso è < 3,841
- Ciò consente di dire che il chi quadrato non è significativo (p = 0,0578)
- I dati non sono sufficienti per concludere che vi è differenza tra le sieroprevalenze in esame.

Valori di χ^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
...	...	...