

# Il test del Chi-quadrato

Prof.ssa Montomoli - Univ. di Pavia

Prof.ssa Zanolin – Univ. di Verona

Test d'ipotesi

**Il rischio di contrarre epatite C è associato all'aver un tatuaggio?**



Cosa vuol dire **ASSOCIAZIONE** tra due variabili?

Due variabili sono **associate** soltanto quando sono correlate tra loro in modo **maggiore** o **minore** di quanto si verifichi per **solo effetto del caso**

I tatuaggi devono essere presenti con frequenza **maggiore** (o **minore**) nei soggetti con epatite C rispetto ai soggetti senza epatite



quindi è necessario studiare anche soggetti senza epatite (**controlli**)

poi confrontarli con i soggetti con epatite

... ed infine decidere se la differenza tra gruppi può essere dovuta al **caso**

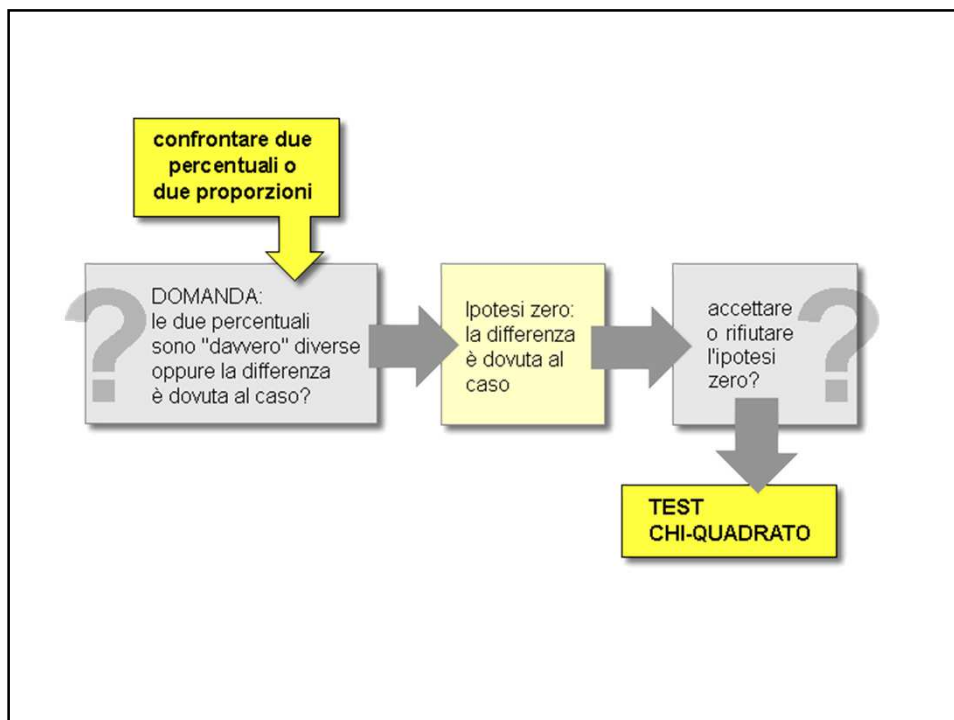
**CON UN TEST STATISTICO**

Se è vero che il caso può influenzare i risultati, come posso **dimostrare l'esistenza di associazioni** ed essere ragionevolmente sicuro che le differenze osservate non sono dovute al caso?

La **statistica** consente di escludere (con un certo grado di probabilità, ma mai con assoluta certezza) che una eventuale associazione sia dovuta appunto al caso



www.quadernodiepidemiologia.it

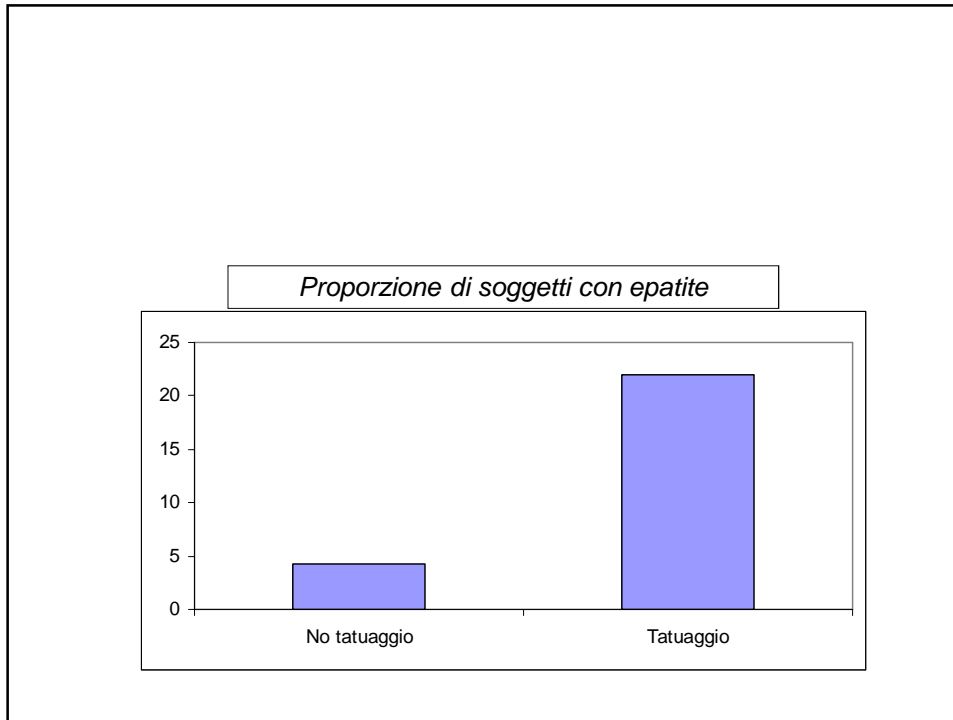


Test d'ipotesi

Il rischio di contrarre epatite C è associato all'aver un tatuaggio?

I dati sono riassunti nella *tabella di contingenza*

|               | Epatite C |            | <i>totale</i> |
|---------------|-----------|------------|---------------|
|               | Si        | No         |               |
| Tatuaggio     | 25        | 88         | <i>113</i>    |
| No tatuaggio  | 22        | 491        | <i>513</i>    |
| <i>totale</i> | <i>47</i> | <i>579</i> | <i>626</i>    |



Test d'ipotesi

Supponi che, in realtà, **NON** esistano differenze nella frequenza di epatite tra i tatuati e i non tatuati.  
 Che probabilità c'è di osservare - in uno studio di dimensioni simili a questo - differenze nella frequenza di epatite diverse da quelle che hai osservato?

La risposta a questa domanda dipende da quanto i dati ottenuti si discostano dai dati che  
**«sarebbe lecito attendersi se la frequenza di epatite C fosse influenzata soltanto dalla variazione casuale».**



## TEST D'INDIPENDENZA

Le due variabili (epatite e tatuaggio) sono **indipendenti**?



Questa è l'**IPOTESI NULLA  $H_0$**  da verificare



Se  $H_0$  è rifiutata: le due variabili  
**NON SONO INDIPENDENTI**

1. Si distribuiscono i soggetti nelle celle della tabella
2. Si calcolano le frequenze attese sotto l' $H_0$  che le due variabili siano indipendenti (per ogni cella della tabella)
3. Si paragonano le frequenze attese e le frequenze osservate
4. Si valuta la discrepanza fra frequenze ATTESE e OSSERVATE



***Come decidere se la discrepanza è piccola o grande?***

tramite il **TEST STATISTICO**

Test d'ipotesi

|              | Epatite C |       | totale |
|--------------|-----------|-------|--------|
|              | Si        | No    |        |
| Tatuaggio    | 8.5       | 104.5 | 113    |
| No tatuaggio | 38.5      | 474.5 | 513    |
| totale       | 47        | 579   | 626    |

$$P(\text{epatite}) = \frac{47}{626} = 0.075$$

$$P(\text{epatite} | \text{tatuaggio}) = 113 \times \frac{47}{626} = 8.5$$

$$P(\text{epatite} | \text{no tatoo}) = 513 \times \frac{47}{626} = 38.5$$

### PRINCIPIO del PRODOTTO

per eventi indipendenti

$$P(A \text{ e } B) = P(A \cap B) = P(A)P(B)$$

**Indipendenza:** due eventi si dicono indipendenti quando il verificarsi dell'uno non influenza il verificarsi dell'altro

es. elevati livelli di glicemia e ulcera,  
probabilità di pescare una pallina di un determinato colore da un'urna con reimpulamento

### Nomenclatura di una tabella di contingenza

X (dot notation)

1.....i.....k

|   |          |  |          |  |          |
|---|----------|--|----------|--|----------|
| 1 | $n_{11}$ |  | $n_{i1}$ |  | $n_{k1}$ |
| . |          |  |          |  |          |
| j | $n_{1j}$ |  | $n_{ij}$ |  | $n_{kj}$ |
| . |          |  |          |  |          |
| t | $n_{1t}$ |  | $n_{it}$ |  | $n_{kt}$ |

Marginali di riga

$$n_{.j} = \sum_{i=1}^k n_{ij}$$

Marginali di colonna

$$n_{i.} = \sum_{j=1}^t n_{ij}$$

$$n_{..} = \sum_{i=1}^k \sum_{j=1}^t n_{ij}$$

$p_{i.}$  viene stimato da  $n_{i.}/n_{..}$   
 $p_{.j}$  viene stimato da  $n_{.j}/n_{..}$

⇒ Sotto l'ipotesi di indipendenza la probabilità di una qualsiasi combinazione delle modalità delle due variabili è data da:

$$P_{ij} = P_{i.} \cdot P_{.j}$$

Probabilità congiunta ←      Prodotto delle marginali

⇒ Gli attesi nella cella ij-ma, sotto l'ipotesi di indipendenza, saranno quindi

$$E_{ij} = P_{i.} \cdot P_{.j} \cdot N$$

Probabilità congiunta ←      Prodotto delle marginali      → N° totale di soggetti

Test d'ipotesi

## CALCOLO DEL TEST $\chi^2$

frequenza osservata

frequenza attesa

$$\chi^2 = \sum_{i=1}^N \left[ \frac{(O_i - E_i)^2}{E_i} \right]$$

Test d'ipotesi

### **FREQUENZE OSSERVATE**

|               | Epatite C |     | <i>totale</i> |
|---------------|-----------|-----|---------------|
|               | Si        | No  |               |
| Tatuaggio     | 25        | 88  | 113           |
| No tatuaggio  | 22        | 491 | 513           |
| <i>totale</i> | 47        | 579 | 626           |

### **FREQUENZE ATTESE**

|               | Epatite C |       | <i>totale</i> |
|---------------|-----------|-------|---------------|
|               | Si        | No    |               |
| Tatuaggio     | 8.5       | 104.5 | 113           |
| No tatuaggio  | 38.5      | 474.5 | 513           |
| <i>totale</i> | 47        | 579   | 626           |



Test d'ipotesi

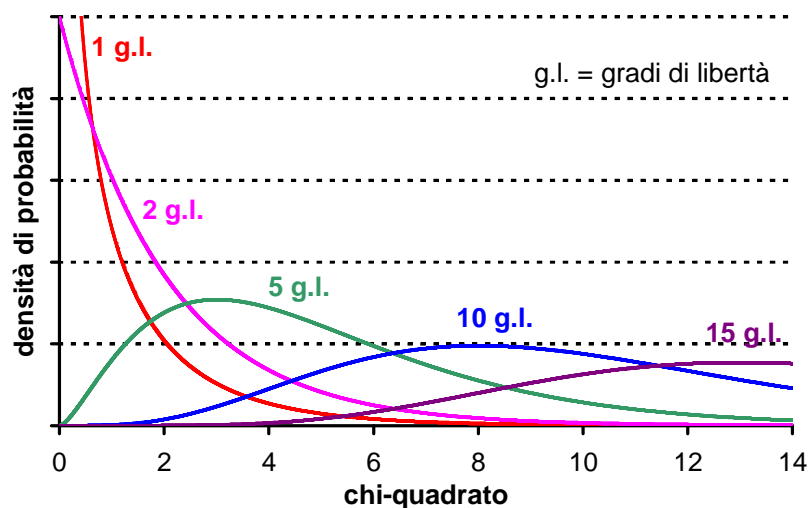
Si calcola il test sostituendo le frequenze osservate ed attese nella formula del  $\chi^2$ :

$$\chi^2 = \frac{(25-8.5)^2}{8.5} + \frac{(22-38.5)^2}{38.5} + \frac{(88-104.5)^2}{104.5} + \frac{(491-474.5)^2}{474.5} = 42.42$$

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

Se esso supera certi valori prefissati la differenza viene ritenuta **significativa**; in caso contrario, non si può affermare l'esistenza di una significativa differenza tra i due eventi considerati.

La distribuzione di probabilità  $\chi^2$  dipende dal numero di gradi di libertà (g.l.)

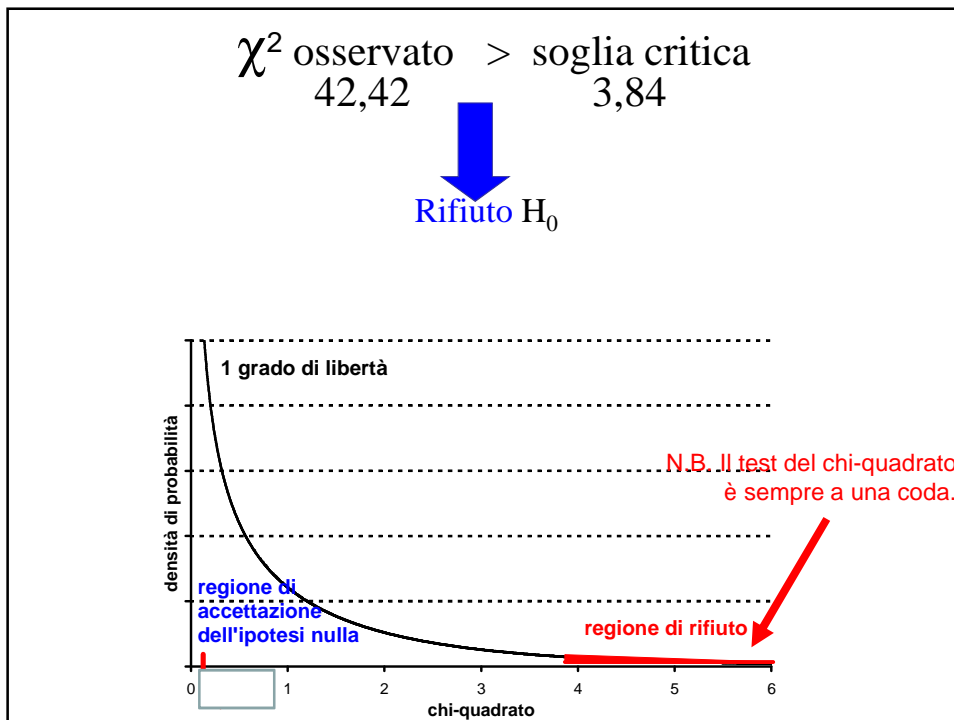


il numero di gradi di libertà di una tabella e del  $\chi^2$  calcolato su di essa è uguale a

$$\text{(numero righe - 1) x (numero colonne - 1)}$$

**Tavola distribuzione CHI-QUADRATO**

| Gradi di libertà | Livello di Probabilità $\alpha$ |       |       |       |       |       |       |       |       |       |
|------------------|---------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                  | 1.00                            | 0.99  | 0.95  | 0.90  | 0.25  | 0.10  | 0.05  | 0.025 | 0.01  | 0.005 |
| 1                |                                 |       |       | 0.02  | 1.32  | 2.71  | 3.84  | 5.02  | 6.64  | 7.88  |
| 2                | 0.01                            | 0.02  | 0.10  | 0.21  | 2.77  | 4.61  | 5.99  | 7.38  | 9.21  | 10.60 |
| 3                | 0.07                            | 0.12  | 0.35  | 0.58  | 4.11  | 6.25  | 7.82  | 9.35  | 11.35 | 12.84 |
| 4                | 0.21                            | 0.30  | 0.71  | 1.06  | 5.39  | 7.78  | 9.49  | 11.14 | 13.28 | 14.86 |
| 5                | 0.41                            | 0.55  | 1.15  | 1.61  | 6.63  | 9.24  | 11.07 | 12.83 | 15.09 | 16.75 |
| 6                | 0.68                            | 0.87  | 1.64  | 2.20  | 7.84  | 10.65 | 12.59 | 14.45 | 16.81 | 18.55 |
| 7                | 0.99                            | 1.24  | 2.17  | 2.83  | 9.04  | 12.02 | 14.07 | 16.01 | 18.48 | 20.28 |
| 8                | 1.34                            | 1.65  | 2.73  | 3.49  | 10.22 | 13.36 | 15.51 | 17.54 | 20.09 | 21.96 |
| 9                | 1.74                            | 2.09  | 3.33  | 4.17  | 11.39 | 14.68 | 16.92 | 19.02 | 21.67 | 23.59 |
| 10               | 2.16                            | 2.56  | 3.94  | 4.87  | 12.55 | 15.99 | 18.31 | 20.48 | 23.21 | 25.19 |
| 11               | 2.60                            | 3.05  | 4.58  | 5.58  | 13.70 | 17.28 | 19.68 | 21.92 | 24.73 | 26.76 |
| 12               | 3.07                            | 3.57  | 5.23  | 6.30  | 14.85 | 18.55 | 21.03 | 23.34 | 26.22 | 28.30 |
| 13               | 3.57                            | 4.11  | 5.89  | 7.04  | 15.98 | 19.81 | 22.36 | 24.74 | 27.69 | 29.82 |
| 14               | 4.08                            | 4.66  | 6.57  | 7.79  | 17.12 | 21.06 | 23.69 | 26.12 | 29.14 | 31.32 |
| 15               | 4.60                            | 5.23  | 7.26  | 8.55  | 18.25 | 22.31 | 25.00 | 27.49 | 30.58 | 32.80 |
| 16               | 5.14                            | 5.81  | 7.96  | 9.31  | 19.37 | 23.54 | 26.30 | 28.85 | 32.00 | 34.27 |
| 17               | 5.70                            | 6.41  | 8.67  | 10.09 | 20.49 | 24.77 | 27.59 | 30.19 | 33.41 | 35.72 |
| 18               | 6.27                            | 7.02  | 9.39  | 10.87 | 21.61 | 25.99 | 28.87 | 31.53 | 34.81 | 37.16 |
| 19               | 6.84                            | 7.63  | 10.12 | 11.65 | 22.72 | 27.20 | 30.14 | 32.85 | 36.19 | 38.58 |
| 20               | 7.43                            | 8.26  | 10.85 | 12.44 | 23.83 | 28.41 | 31.41 | 34.17 | 37.57 | 40.00 |
| 21               | 8.03                            | 8.90  | 11.59 | 13.24 | 24.94 | 29.62 | 32.67 | 35.48 | 38.93 | 41.40 |
| 22               | 8.64                            | 9.54  | 12.34 | 14.04 | 26.04 | 30.81 | 33.92 | 36.78 | 40.29 | 42.80 |
| 23               | 9.26                            | 10.20 | 13.09 | 14.85 | 27.14 | 32.01 | 35.17 | 38.08 | 41.64 | 44.18 |
| 24               | 9.89                            | 10.86 | 13.85 | 15.66 | 28.24 | 33.20 | 36.42 | 39.36 | 42.98 | 45.56 |
| 25               | 10.52                           | 11.52 | 14.61 | 16.47 | 29.34 | 34.38 | 37.65 | 40.65 | 44.31 | 46.93 |



Test d'ipotesi

## Output di un programma statistico

| row   | col |     | Total |
|-------|-----|-----|-------|
|       | 1   | 2   |       |
| 1     | 25  | 88  | 113   |
| 2     | 22  | 491 | 513   |
| Total | 47  | 579 | 626   |

Pearson chi2(1) = 42.4189 Pr = 0.000

Test d'ipotesi

$P < 0.00001$



**Rifiuto l'ipotesi nulla di indipendenza delle due variabili**

**CONCLUSIONI**

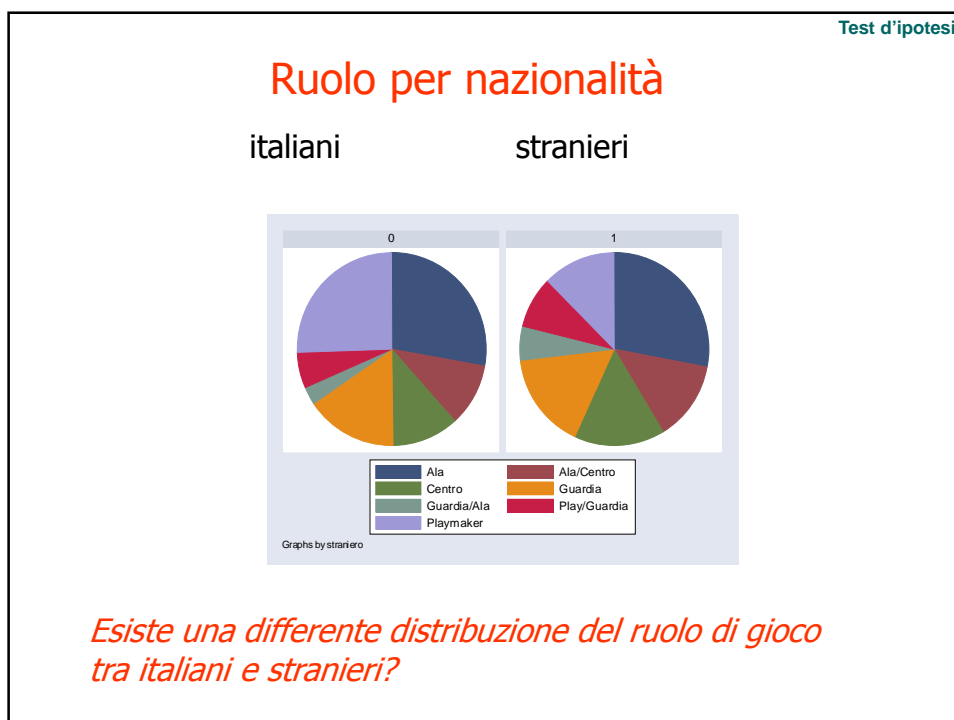
i dati della tabella sono improbabili, se è vera l'ipotesi che epatite C e tatuaggi sono indipendenti



**esiste una relazione tra epatite C e tatuaggi**

Test d'ipotesi

|  | Situazione reale  |  |
|--|---|--|
| <b>Conclusioni</b>   | Epatite e tatuaggi sono indipendenti<br><b>H<sub>0</sub> VERA</b> | Epatite e tatuaggi NON sono indipendenti<br><b>H<sub>0</sub> FALSA</b> |
| Epatite e tatuaggi sono indipendenti<br><b>RIFIUTO H<sub>0</sub></b>         | $\alpha$<br><b>errore di I tipo</b>                               | <b>OK</b>  |
| Epatite e tatuaggi NON sono indipendenti<br><b>NON RIFIUTO H<sub>0</sub></b> | <b>OK</b>   | $\beta$<br><b>errore di II tipo</b>                                    |



Test d'ipotesi

## Ruolo per nazionalità

| ruolo        | straniero |     | Total |
|--------------|-----------|-----|-------|
|              | 0         | 1   |       |
| Ala          | 37        | 29  | 66    |
| Ala/Centro   | 14        | 14  | 28    |
| Centro       | 15        | 16  | 31    |
| Guardia      | 21        | 17  | 38    |
| Guardia/Ala  | 4         | 6   | 10    |
| Play/Guardia | 8         | 9   | 17    |
| Playmaker    | 34        | 13  | 47    |
| Total        | 133       | 104 | 237   |

Pearson chi2(6) = 7.8336 Pr = 0.251

*Le due variabili (ruolo e nazionalità) sono indipendenti.*

*La distribuzione del ruolo non è diversa per nazionalità*

Test d'ipotesi

**Ipotesi: i giocatori stranieri vengono acquistati per ricoprire il ruolo di centro**

| ruolo     | straniero |    | Total |
|-----------|-----------|----|-------|
|           | 0         | 1  |       |
| Centro    | 15        | 16 | 31    |
| Playmaker | 34        | 13 | 47    |
| Total     | 49        | 29 | 78    |

Pearson chi2(1) = 4.5887 Pr = 0.032

*Le due variabili (ruolo e nazionalità) non sono indipendenti.*