



Facoltà di Ingegneria  
Università degli Studi di Firenze  
Dipartimento di Elettronica e Telecomunicazioni

# **Misure Elettriche**

## **- Inferenza statistica -**

### **4. Test Chi Quadro (Non parametrici)**

**Ing. Andrea Zanobini**

Dipartimento di Elettronica e Telecomunicazioni

## *Test chi-quadro*

Questa distribuzione ha numerose altre applicazioni nella statistica; in particolare ne faremo uso nella verifica di ipotesi con dati disponibili sotto forma di frequenze.

Queste procedure di verifica delle ipotesi sono note come **test di adattamento** (o anche **goodness of fit**) e **test di indipendenza**.

### **Test chi-quadro di adattamento**

I test statistici che servono a verificare se una certa distribuzione è compatibile con i dati del campione sono detti **test sulla bontà di adattamento**.

Per effettuare il test supponiamo di avere un campione di  $n$  osservazioni di una variabile, raggruppate in una tabella contenente  $k$  classi.

In altri termini, la tabella rappresenta la distribuzione di frequenza assoluta di una variabile qualitativa o di una variabile numerica discreta o continua.

Per ciascuna classe supponiamo di avere, oltre alla **frequenza osservata**  $O_i$ , una **frequenza attesa**  $A_i$ , con cui si vuole confrontare la frequenza osservata; le frequenze attese sono quelle che si osserverebbero se i dati del campione fossero distribuiti esattamente secondo la distribuzione ipotizzata.

Per valutare quantitativamente la bontà dell'adattamento delle frequenze osservate alle frequenze attese si utilizza la **statistica test**

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - A_i)^2}{A_i} \quad (9.1)$$

che viene detta il **chi-quadro** calcolato dal campione.

Si dimostra che, per  $n$  sufficientemente grande, questa statistica ha approssimativamente la distribuzione  $\chi^2$ , con grado di libertà  $v = k - 1 - m$ , dove  $m$  è il numero dei parametri della distribuzione teorica stimati servendosi dei dati del campione.

Se l'ipotesi nulla  $H_0$  è che i dati si adattino alla distribuzione teorica ipotizzata, la regola di decisione sarà: si rifiuti l'ipotesi nulla se il valore della statistica  $\chi^2$  calcolato dai dati è maggiore del valore critico  $\chi_\alpha^2$

$$\chi^2 > \chi_\alpha^2 \quad (9.2)$$

$\alpha$  è il livello di significatività stabilito e il grado di libertà della distribuzione  $\chi^2$  è  $v = k - 1 - m$ ;  $k$  indica il numero delle classi e  $m$  il numero dei parametri della distribuzione teorica stimati servendosi dei dati del campione.

Questa procedura, detta **test chi-quadro di adattamento**, è valida purché le frequenze assolute attese siano tutte maggiori o uguali a 5.

**Esempio 2**

Si effettuano 120 lanci di un dado e si osservano le seguenti uscite

|                  |    |    |    |    |    |    |
|------------------|----|----|----|----|----|----|
| <i>N° uscito</i> | 1  | 2  | 3  | 4  | 5  | 6  |
| <i>Frequenza</i> | 25 | 17 | 15 | 23 | 24 | 16 |

Provare l'ipotesi che il dado non sia truccato, usando il livello di significatività del 5%.

Se il dado non è truccato, le frequenze attese sono tutte uguali a 20 e si costruisce la tabella 3

| <i>N° uscito</i> | <i>Frequenze<br/>osservate <math>O_i</math></i> | <i>Frequenze<br/>attese <math>A_i</math></i> | $\frac{(O_i - A_i)^2}{A_i}$ |
|------------------|-------------------------------------------------|----------------------------------------------|-----------------------------|
| 1                | 25                                              | 20                                           | 1.25                        |
| 2                | 17                                              | 20                                           | 0.45                        |
| 3                | 15                                              | 20                                           | 1.25                        |
| 4                | 23                                              | 20                                           | 0.45                        |
| 5                | 24                                              | 20                                           | 0.80                        |
| 6                | 16                                              | 20                                           | 0.80                        |
| <i>Totale</i>    | 120                                             | 120                                          | 5.00                        |

### ... Esempio 2 (continua)

| <i>N° uscito</i> | <i>Frequenze<br/>osservate <math>O_i</math></i> | <i>Frequenze<br/>attese <math>A_i</math></i> | $\frac{(O_i - A_i)^2}{A_i}$ |
|------------------|-------------------------------------------------|----------------------------------------------|-----------------------------|
| 1                | 25                                              | 20                                           | 1.25                        |
| 2                | 17                                              | 20                                           | 0.45                        |
| 3                | 15                                              | 20                                           | 1.25                        |
| 4                | 23                                              | 20                                           | 0.45                        |
| 5                | 24                                              | 20                                           | 0.80                        |
| 6                | 16                                              | 20                                           | 0.80                        |
| <i>Totale</i>    | 120                                             | 120                                          | 5.00                        |

Il valore della statistica chi-quadro calcolato dal campione è

$$\chi^2 = 5.00$$

Le classi sono 6 e nessun parametro è stato stimato dal campione, perciò il grado di libertà è

$$v = k - 1 - m = 6 - 1 = 5$$

Il valore critico al livello di significatività del 5% è

$$\chi_{0.05}^2 = 11.070$$

**Fare da soli  
l'esempio 1 e 3 !!**

Il test è: si rifiuta l'ipotesi che il dado sia buono se il valore del chi-quadro è maggiore del valore critico; nel nostro caso è minore, perciò in base ai dati non possiamo rifiutare l'ipotesi, e concludiamo che, al livello di significatività del 5%, non c'è una significativa evidenza che il dado sia truccato.

#### **Esempio 4**

In base a una ricerca condotta in anni precedenti, si può ritenere che il numero di incidenti stradali per settimana, in un certo tratto di autostrada, segua la distribuzione di Poisson di parametro  $\lambda = 0.4$ . Nelle ultime 90 settimane si sono rilevati i seguenti dati

|                                               |    |    |   |         |               |
|-----------------------------------------------|----|----|---|---------|---------------|
| <i>N° di incidenti per settimana</i>          | 0  | 1  | 2 | 3 o più | <i>Totale</i> |
| <i>N° di settimane in cui si è verificato</i> | 52 | 32 | 6 | 0       | 90            |

Possiamo affermare che il modello è ancora applicabile alla descrizione del fenomeno, oppure qualcosa è cambiato?

La distribuzione teorica con cui si vogliono confrontare i dati è la distribuzione di Poisson di parametro (valor medio)  $\lambda = 0.4$ . Usando questa distribuzione possiamo calcolare le seguenti probabilità<sup>1</sup>

$$P(X = 0) = e^{-0.4} = 0.6703$$

$$P(X = 1) = e^{-0.4} \cdot 0.4 = 0.2681$$

$$P(X = 2) = e^{-0.4} \cdot \frac{0.4^2}{2} = 0.0536$$

$$P(X \geq 3) = 1 - P(X < 3) = 1 - (0.6703 + 0.2681 + 0.0536) = 0.0080$$

Queste sono le probabilità con cui  $X$  appartiene alle quattro classi, ossia le frequenze relative attese;

### ... Esempio 4 (continua)

le frequenze assolute attese si ottengono moltiplicando le frequenze relative attese per il numero di osservazioni, in questo caso 90. Si ottiene la tabella 5

| <i>Classe<br/>(N° incidenti per settimana)</i> | <i>Frequenza relativa<br/>attesa</i> | <i>Frequenza<br/>assoluta attesa</i> | <i>Frequenza<br/>assoluta osservata</i> |
|------------------------------------------------|--------------------------------------|--------------------------------------|-----------------------------------------|
| $X = 0$                                        | 0.6703                               | 60.33                                | 52                                      |
| $X = 1$                                        | 0.2681                               | 24.13                                | 32                                      |
| $X = 2$                                        | 0.0536                               | 4.82                                 | 6                                       |
| $X \geq 3$                                     | 0.0080                               | 0.72                                 | 0                                       |
| <i>Totale</i>                                  | 1                                    | 90                                   | 90                                      |

Tabella 5

Si osserva che le ultime due classi hanno frequenze assolute attese minori di 5, perciò non possiamo usare questa tabella per effettuare il test; si accorpano allora le ultime due classi in un'unica classe con frequenza assoluta attesa pari a  $4.82 + 0.72 = 5.54$  e frequenza assoluta osservata pari a  $6 + 0 = 6$ .

### ... Esempio 4 (continua)

| <i>Classe<br/>(N° incidenti per settimana)</i> | <i>Frequenza<br/>assoluta attesa <math>A_i</math></i> | <i>Frequenza<br/>assoluta osservata<br/><math>O_i</math></i> | $\frac{(O_i - A_i)^2}{A_i}$ |
|------------------------------------------------|-------------------------------------------------------|--------------------------------------------------------------|-----------------------------|
| $X = 0$                                        | 60.33                                                 | 52                                                           | 1.15                        |
| $X = 1$                                        | 24.13                                                 | 32                                                           | 2.57                        |
| $X \geq 2$                                     | 5.54                                                  | 6                                                            | 0.04                        |
| <i>Totale</i>                                  | 90                                                    | 90                                                           | 3.76                        |

Il valore della statistica chi-quadro calcolato dal campione è

$$\chi^2 = 1.15 + 2.57 + 0.04 = 3.76$$

Le classi, dopo l'accorpamento sono 3 e nessun parametro è stato stimato dal campione, perciò il grado di libertà è

$$v = k - 1 - m = 3 - 1 = 2$$

Il valore critico al livello di significatività del 5% è

$$\chi_{0.05}^2 = 5.991$$

**Fare da soli  
l'esempio 5 e 6 !!**

Il test è: si rifiuta l'ipotesi di adattamento se il valore del chi-quadro è maggiore del valore critico; nel nostro caso è minore, perciò, in base ai dati, non possiamo rifiutare l'ipotesi nulla di adattamento, e concludiamo che, al livello di significatività del 5%, in base ai dati del campione non c'è evidenza statistica del fatto che la legge seguita dal numero settimanale di incidenti sia cambiata.



### Esempio 7

Cinque monete sono state lanciate 1000 volte, e a ciascun lancio è stato osservato il numero di teste; nella tabella è riportato il numero di lanci nei quali sono state ottenute 0, 1, ..., 5 teste.

|                            |    |     |     |     |     |    |
|----------------------------|----|-----|-----|-----|-----|----|
| <i>N° teste</i>            | 0  | 1   | 2   | 3   | 4   | 5  |
| <i>Frequenza osservata</i> | 38 | 144 | 342 | 287 | 164 | 25 |

Stabilire se le monete si possono ritenere non truccate.

Se le monete sono eque, il numero di teste su 5 monete in un singolo lancio ha una distribuzione binomiale di parametri  $p = \frac{1}{2}$ ,  $n = 5$ . Le probabilità di avere 0, 1, ..., 5 teste si possono ottenere dalla tavola della distribuzione binomiale; le corrispondenti frequenze assolute attese si ricavano moltiplicando per 1000 tali probabilità. Si ottiene così la tabella 11

| <i>Numero di teste</i> | <i>Frequenze osservate <math>O_i</math></i> | <i>Frequenze relative attese (binomiale)</i> | <i>Frequenze attese <math>A_i</math></i> | $\frac{(O_i - A_i)^2}{A_i}$ |
|------------------------|---------------------------------------------|----------------------------------------------|------------------------------------------|-----------------------------|
| 0                      | 38                                          | 0.0313                                       | 31.25                                    | 1.46                        |
| 1                      | 144                                         | 0.1562                                       | 156.25                                   | 0.96                        |
| 2                      | 342                                         | 0.3125                                       | 312.50                                   | 2.78                        |
| 3                      | 287                                         | 0.3125                                       | 312.50                                   | 2.08                        |
| 4                      | 164                                         | 0.1562                                       | 156.25                                   | 0.38                        |
| 5                      | 25                                          | 0.0313                                       | 31.25                                    | 1.25                        |
| <i>Totale</i>          | 1000                                        | 1                                            | 1000                                     | 8.91                        |

### ... Esempio 7 (continua)

| <i>Numero di teste</i> | <i>Frequenze osservate <math>O_i</math></i> | <i>Frequenze relative attese (binomiale)</i> | <i>Frequenze attese <math>A_i</math></i> | $\frac{(O_i - A_i)^2}{A_i}$ |
|------------------------|---------------------------------------------|----------------------------------------------|------------------------------------------|-----------------------------|
| 0                      | 38                                          | 0.0313                                       | 31.25                                    | 1.46                        |
| 1                      | 144                                         | 0.1562                                       | 156.25                                   | 0.96                        |
| 2                      | 342                                         | 0.3125                                       | 312.50                                   | 2.78                        |
| 3                      | 287                                         | 0.3125                                       | 312.50                                   | 2.08                        |
| 4                      | 164                                         | 0.1562                                       | 156.25                                   | 0.38                        |
| 5                      | 25                                          | 0.0313                                       | 31.25                                    | 1.25                        |
| <i>Totale</i>          | 1000                                        | 1                                            | 1000                                     | 8.91                        |

Il valore della statistica chi-quadro calcolato dal campione è

$$\chi^2 = 8.91$$

Le classi sono 5 e nessun parametro è stato stimato dal campione, perciò il grado di libertà è

$$v = k - 1 - m = 5$$

Il valore critico al livello di significatività del 5% è

$$\chi_{0.05}^2 = 11.070$$

**Fare da soli  
l'esempio 8 !!**

Il valore del chi-quadro è minore del valore critico, perciò in base ai dati non possiamo rifiutare l'ipotesi nulla di adattamento, e concludiamo che, al livello di significatività del 5%, c'è un buon adattamento dei dati alla distribuzione binomiale; in altri termini non possiamo rifiutare l'ipotesi che la moneta sia equa.

### **Esempio 9**

Sono state misurate le lunghezze di 150 sbarrette di metallo simili, e i dati sono stati raggruppati nella tabella seguente

| <i>Classi<br/>(lunghezza in mm)</i> | <i>Frequenze<br/>osservate</i> |
|-------------------------------------|--------------------------------|
| $27 < x \leq 28$                    | 3                              |
| $28 < x \leq 29$                    | 23                             |
| $29 < x \leq 30$                    | 53                             |
| $30 < x \leq 31$                    | 50                             |
| $31 < x \leq 32$                    | 21                             |

Stabilire se in base a questi dati si può affermare che la lunghezza delle sbarrette segue una distribuzione normale.

L'ipotesi nulla è che la lunghezza abbia una distribuzione normale. Per sottoporre a test questa ipotesi occorre calcolare dai dati la stima per la media e la varianza della popolazione.

Usando i dati raggruppati si ottiene

$$\bar{x} = 29.92$$

$$s^2 = 0.9566 \quad s = 0.978$$

### ... Esempio 9 (continua)

Calcoliamo ora le frequenze relative attese delle classi. Oltre alle classi indicate, nelle quali cadono le osservazioni, occorre considerare anche la classe  $x \leq 27$  e la classe  $x > 32$ .

Per calcolare queste frequenze usiamo la funzione di ripartizione della variabile aleatoria normale  $X$  di media  $\mu = 29.92$  e scarto quadratico medio  $\sigma = 0.978$ ; passando alla variabile aleatoria standardizzata

$$Z = \frac{X - 29.92}{0.978}$$

Ad esempio per la classe  $27 < x \leq 28$  la frequenza relativa attesa (probabilità) si calcola nel modo seguente

$$X = 27 \quad \Rightarrow \quad Z = \frac{27 - 29.92}{0.978} = -2.99$$

$$X = 28 \quad \Rightarrow \quad Z = \frac{28 - 29.92}{0.978} = -1.96$$

$$\begin{aligned} P(27 < X < 28) &= P(-2.99 < Z < -1.96) = P(1.96 < Z < 2.99) = \\ &= 0.9986 - 0.9750 = 0.0236 \end{aligned}$$

La corrispondente frequenza attesa è quindi  $150 \cdot 0.0236 = 3.54$

**... Esempio 9  
(continua)**

| <i>Classi</i>    | <i>Frequenza<br/>osservata <math>f_i</math></i> | <i>frequenze<br/>relative attese<br/>(probabilità)</i> | <i>frequenze<br/>attese</i> |
|------------------|-------------------------------------------------|--------------------------------------------------------|-----------------------------|
| $x \leq 27$      | 0                                               | 0.0014                                                 | 0.21                        |
| $27 < x \leq 28$ | 3                                               | 0.0236                                                 | 3.54                        |
| $28 < x \leq 29$ | 23                                              | 0.1486                                                 | 22.29                       |
| $29 < x \leq 30$ | 53                                              | 0.3583                                                 | 53.74                       |
| $30 < x \leq 31$ | 50                                              | 0.3324                                                 | 49.86                       |
| $31 < x \leq 32$ | 21                                              | 0.1191                                                 | 17.87                       |
| $x > 32$         | 0                                               | 0.0166                                                 | 2.49                        |
| <i>Totale</i>    | 150                                             | 1                                                      | 150                         |

Poiché ci sono delle classi che hanno frequenze assolute attese minori di 5, procediamo ad accorpare le prime tre classi e le ultime due e otteniamo la tabella 17

| <i>Classi</i>    | <i>Frequenza<br/>osservata <math>f_i</math></i> | <i>frequenze<br/>attese <math>A_i</math></i> | $\frac{(O_i - A_i)^2}{A_i}$ |
|------------------|-------------------------------------------------|----------------------------------------------|-----------------------------|
| $x \leq 29$      | 26                                              | 26.04                                        | 0.0001                      |
| $29 < x \leq 30$ | 53                                              | 53.74                                        | 0.0102                      |
| $30 < x \leq 31$ | 50                                              | 49.86                                        | 0.0004                      |
| $x > 31$         | 21                                              | 20.36                                        | 0.0201                      |
| <i>Totale</i>    | 150                                             | 150                                          | 0.0308                      |

**... Esempio  
9 (continua)**

| <i>Classi</i>    | <i>Frequenza<br/>osservata <math>f_i</math></i> | <i>frequenze<br/>attese <math>A_i</math></i> | $\frac{(O_i - A_i)^2}{A_i}$ |
|------------------|-------------------------------------------------|----------------------------------------------|-----------------------------|
| $x \leq 29$      | 26                                              | 26.04                                        | 0.0001                      |
| $29 < x \leq 30$ | 53                                              | 53.74                                        | 0.0102                      |
| $30 < x \leq 31$ | 50                                              | 49.86                                        | 0.0004                      |
| $x > 31$         | 21                                              | 20.36                                        | 0.0201                      |
| <i>Totale</i>    | 150                                             | 150                                          | 0.0308                      |

Il valore della statistica chi-quadro calcolato dal campione è

$$\chi^2 = 0.0308$$

Le classi sono 4 e due parametri, valor medio e varianza della popolazione, sono stati stimati dal campione, perciò il grado di libertà è

$$v = k - 1 - m = 4 - 1 - 2 = 1$$

Il valore critico al livello di significatività del 5% è

$$\chi_{0.05}^2 = 3.841$$

**Fare da soli  
l'esempio 10 !!**

Il valore del chi-quadro è minore del valore critico, perciò in base ai dati non possiamo rifiutare l'ipotesi nulla di adattamento, e concludiamo che, al livello di significatività del 5%, c'è un buon adattamento dei dati alla distribuzione normale.

Esistono altri test più indicati per distribuzioni continue, ad esempio il test di Kolmogorov-Smirnov, che non sarà trattato in queste lezioni.

## Test chi-quadro di indipendenza

Il test chi-quadro può essere utilizzato anche per verificare l'indipendenza o meno di due variabili: questa è forse la più frequente fra le applicazioni della distribuzione  $\chi^2$ .

In questo test si vuole sottoporre a test l'ipotesi nulla che due criteri di classificazione, quando applicati al medesimo insieme di dati, siano indipendenti.

Si dice che due criteri di classificazione sono indipendenti se la distribuzione rispetto a un criterio non viene influenzata dalla classificazione rispetto all'altro criterio. Se l'ipotesi nulla viene rifiutata, concludiamo che i due criteri di classificazione sono indipendenti.

### *Esempio 11*

Un corso universitario è impartito dallo stesso insegnante a studenti del secondo anno di tre indirizzi di laurea diversi; gli esami superati e non superati sono registrati nella seguente tabella

|                           | <i>Laurea A</i> | <i>Laurea B</i> | <i>Laurea C</i> |
|---------------------------|-----------------|-----------------|-----------------|
| <i>esame superato</i>     | 80              | 40              | 110             |
| <i>esame non superato</i> | 50              | 20              | 40              |

Il rendimento degli studenti dei tre corsi, rispetto a questo esame, si può ritenere sostanzialmente equivalente, oppure le differenze sono statisticamente significative?

Questo equivale a chiedersi se le due variabili (qualitative) “indirizzo di laurea” e “superamento dell'esame” sono indipendenti.

### **Esempio 12**

Per stabilire l'efficacia di un vaccino anti-influenzale è stata condotta una ricerca, somministrando il vaccino a 500 persone e controllando il loro stato di salute in un anno; lo stesso controllo è stato fatto per un gruppo di altre 500 persone non vaccinate; in base ai risultati dell'esperimento si è ottenuta la seguente tabella

|                      | <i>nessuna influenza</i> | <i>una influenza</i> | <i>più di una influenza</i> |
|----------------------|--------------------------|----------------------|-----------------------------|
| <i>vaccinati</i>     | 252                      | 145                  | 103                         |
| <i>non vaccinati</i> | 224                      | 136                  | 140                         |

Si può ritenere che il vaccino sia efficace, ossia sottoponendosi alla vaccinazione si ha un minor rischio di contrarre la malattia, oppure il vaccino non è efficace?

Questo equivale a chiedersi se le due variabili (qualitative) “vaccinazione” e “minor numero di influenze” sono indipendenti oppure no.

### **Esempio 13**

Per verificare la qualità della produzione in una fabbrica, un ingegnere controlla il numero di pezzi difettosi prodotti da tre macchine diverse, e ottiene la seguente tabella di dati

|                  | <i>macchina 1</i> | <i>macchina 2</i> | <i>macchina 3</i> |
|------------------|-------------------|-------------------|-------------------|
| <i>buoni</i>     | 150               | 140               | 200               |
| <i>difettosi</i> | 25                | 40                | 20                |

Si può ritenere che la quantità di pezzi difettosi non dipenda dalla macchina che si utilizza?



In tutti gli esempi considerati disponiamo di  $n$  osservazioni congiunte di due variabili e ci chiediamo se esiste una forma di dipendenza fra le due variabili.

L'ipotesi nulla sarà che le due variabili siano indipendenti; se si rifiuta l'ipotesi nulla, la conclusione sarà che vi sia qualche interazione fra i due criteri di classificazione.

Tabelle come quelle riprodotte negli esempi si chiamano **tabelle di contingenza**. In una tabella di questo tipo  $n$  osservazioni sono classificate secondo un certo criterio  $X$ , ossia secondo il valore di una certa variabile, in  $r$  classi e, contemporaneamente, sono classificate secondo un altro criterio  $Y$ , ossia secondo i valori assunti da un'altra variabile, in  $c$  classi; la tabella riporta all'incrocio di ogni riga con ogni colonna la frequenza assoluta osservata  $O_{ij}$

|               |     | <i>Classi</i> |          |          |     |          |
|---------------|-----|---------------|----------|----------|-----|----------|
|               |     | 1             | 2        | 3        | ... | $c$      |
| <i>Classi</i> | 1   | $O_{11}$      | $O_{12}$ | $O_{13}$ | ... | $O_{1c}$ |
|               | 2   | $O_{21}$      | $O_{22}$ | $O_{23}$ | ... | $O_{2c}$ |
|               | 3   | $O_{31}$      | $O_{32}$ | $O_{33}$ | ... | $O_{3c}$ |
|               | ... | ...           | ...      | ...      | ... | ...      |
|               | $r$ | $O_{r1}$      | $O_{r2}$ | $O_{r3}$ |     | $O_{rc}$ |

Partendo da questa tabella si costruisce la **tabella delle frequenze attese**, ossia delle frequenze che si avrebbero nell'ipotesi di indipendenza; ogni frequenza attesa  $A_{ij}$  si ottiene con la seguente formula

$$A_{ij} = \frac{(\text{totale riga } i) \cdot (\text{totale colonna } j)}{\text{totale generale}}$$

Si dimostra che, per  $n$  sufficientemente grande, la statistica

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - A_{ij})^2}{A_{ij}} \quad (9.3)$$

detta il **chi-quadro** calcolato dal campione, ha approssimativamente la distribuzione  $\chi^2$  con grado di libertà  $\nu = (r-1) \cdot (c-1)$ .

Se l'ipotesi nulla  $H_0$  è che le due variabili siano indipendenti, la regola di decisione sarà: si rifiuti l'ipotesi nulla, se il valore della statistica  $\chi^2$  calcolato dai dati è maggiore del valore critico  $\chi_{\alpha}^2$

$$\chi^2 > \chi_{\alpha}^2 \quad (9.4)$$

$\alpha$  è il livello di significatività stabilito e il grado di libertà della distribuzione  $\chi^2$  è  $\nu = (r-1) \cdot (c-1)$ .

Questa procedura, detta **test chi-quadro di indipendenza**, è valida purché le frequenze assolute attese siano tutte maggiori o uguali a 5.

### **Esempio 11 – parte 2**

Riprendiamo l'esempio 11.

Un corso universitario è impartito dallo stesso insegnante a studenti del secondo anno di tre indirizzi di laurea diversi; gli esami superati e non superati sono registrati nella tabella 21

| <i>Tabella di contingenza – Frequenze osservate</i> |                 |                 |                 |                     |
|-----------------------------------------------------|-----------------|-----------------|-----------------|---------------------|
|                                                     | <i>Laurea A</i> | <i>Laurea B</i> | <i>Laurea C</i> | <i>Totale esami</i> |
| <i>esame superato</i>                               | 80              | 40              | 110             | 230                 |
| <i>esame non superato</i>                           | 50              | 20              | 40              | 110                 |
| <i>Totale studenti iscritti</i>                     | 130             | 60              | 150             | 340                 |

Il rendimento degli studenti dei tre corsi, rispetto a questo esame, si può ritenere sostanzialmente equivalente, oppure le differenze sono statisticamente significative?

Costruiamo la tabella delle frequenze attese, ricordando che ogni casella contiene il prodotto del totale di riga per il totale di colonna, diviso per il totale generale. Nella tabella 21 l'ultima colonna e l'ultima riga contengono i totali parziali delle righe e delle colonne, che servono per calcolare le frequenze attese, l'ultima casella in basso a destra contiene il totale generale.

| <i>Frequenze attese</i>   |                 |                 |                 |
|---------------------------|-----------------|-----------------|-----------------|
|                           | <i>Laurea A</i> | <i>Laurea B</i> | <i>Laurea C</i> |
| <i>esame superato</i>     | 87.94           | 40.59           | 101.47          |
| <i>esame non superato</i> | 42.06           | 19.41           | 48.53           |

**... Esempio 11  
(continua)**

| <i>Frequenze attese</i>   |                 |                 |                 |
|---------------------------|-----------------|-----------------|-----------------|
|                           | <i>Laurea A</i> | <i>Laurea B</i> | <i>Laurea C</i> |
| <i>esame superato</i>     | 87.94           | 40.59           | 101.47          |
| <i>esame non superato</i> | 42.06           | 19.41           | 48.53           |

Servendosi delle tabelle delle frequenze osservate e delle frequenze attese si calcola il valore della statistica chi-quadro con la formula (9.3)

$$\chi^2 = \frac{(80 - 87.94)^2}{87.94} + \frac{(40 - 40.59)^2}{40.59} + \frac{(110 - 101.47)^2}{101.47} + \\ + \frac{(50 - 42.06)^2}{42.06} + \frac{(20 - 19.41)^2}{19.41} + \frac{(40 - 48.53)^2}{48.53} = 4.46$$

Il grado di libertà è  $v = (2 - 1) \cdot (3 - 1) = 2$ ; il valore critico al livello di significatività del 5% è  $\chi_{0.05}^2 = 5.991$ . Poiché il valore della statistica chi-quadro è minore del valore critico, i dati non consentono di rifiutare l'ipotesi nulla e si conclude che il risultato dell'esame è indipendente dall'indirizzo di laurea, ossia il rendimento è equivalente.

**Fare da soli  
l'esempio 12  
parte 2 !!**

### *Esempio 13 – parte 2*

Per verificare la qualità della produzione in una fabbrica, un ingegnere controlla il numero di pezzi difettosi prodotti da tre macchine diverse, e ottiene la seguente tabella di dati

| <i>Frequenze osservate</i> |                   |                   |                   |                     |
|----------------------------|-------------------|-------------------|-------------------|---------------------|
|                            | <i>macchina 1</i> | <i>macchina 2</i> | <i>macchina 3</i> | <i>Totale buoni</i> |
| <i>buoni</i>               | 150               | 140               | 200               | 490                 |
| <i>difettosi</i>           | 25                | 40                | 20                | 85                  |
| <i>Totale macchina</i>     | 175               | 180               | 220               | 575                 |

Tabella 25

Si può ritenere che la quantità di pezzi difettosi non dipenda dalla macchina che si utilizza?

Costruiamo la tabella delle frequenze attese

| <i>Frequenze attese</i> |                   |                   |                   |
|-------------------------|-------------------|-------------------|-------------------|
|                         | <i>macchina 1</i> | <i>macchina 2</i> | <i>macchina 3</i> |
| <i>buoni</i>            | 149.13            | 153.39            | 187.48            |
| <i>difettosi</i>        | 25.87             | 26.61             | 32.52             |

### ... Esempio 13 (continua)

| <i>Frequenze attese</i> |                   |                   |                   |
|-------------------------|-------------------|-------------------|-------------------|
|                         | <i>macchina 1</i> | <i>macchina 2</i> | <i>macchina 3</i> |
| <i>buoni</i>            | 149.13            | 153.39            | 187.48            |
| <i>difettosi</i>        | 25.87             | 26.61             | 32.52             |

Servendosi delle tabelle delle frequenze osservate e delle frequenze attese, si calcola il valore della statistica chi-quadro

$$\chi^2 = \frac{(150 - 149.13)^2}{149.13} + \frac{(140 - 153.39)^2}{153.39} + \frac{(200 - 187.48)^2}{187.48} + \frac{(25 - 25.87)^2}{25.87} + \frac{(40 - 26.61)^2}{26.61} + \frac{(20 - 32.52)^2}{32.52} = 13.60$$

Il grado di libertà è  $\nu = (2 - 1) \cdot (3 - 1) = 2$ ; il valore critico al livello di significatività del 5% è  $\chi_{0.05}^2 = 5.991$ . Poiché il valore della statistica chi-quadro è maggiore del valore critico, i dati consentono di rifiutare l'ipotesi nulla e si conclude che c'è evidenza statistica di una dipendenza del numero dei pezzi difettosi dalla macchina che si utilizza.

**Fare da soli  
l'esempio 14 !!**