# Esercizio sul $\chi^2$

**Esercizio.** *Si fa una rilevazione statistica sul numero di telefonate al minuto che arrivano ad un centralino. Un campione di $n = 60$ telefonate fornisce la seguente tabella:*

| *tel./min.* : | *0* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *>10* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *freq. oss.* : | *2* | *3* | *7* | *9* | *11* | *10* | *8* | *4* | *3* | *2* | *1* | *0* |

*Utilizzando il test del $\chi^2$, possiamo affermare, con un livello di confidenza del $5\%$ che il numero di telefonate al minuto che arrivano segua una legge normale?*

**Soluzione.** L'ipotesi nulla è che il numero di telefonate al minuto sia distribuito secondo una legge normale di parametri (media e varianza) incogniti $\mu$ e $\sigma^2$, che vanno quindi determinati dai dati statistici. Indichiamo con $X$ la variabile casuale che rappresenta il numero di telefonate al minuto, con $x_i$ i valori che può assumere e con $\omega_i$ ($i = 1, 2, ..., k$, con $k = 11$) la loro frequenza (osservata). Dobbiamo, quindi, calcolare preventivamente le probabilità che competono a ciascun valore $x_i$, vale a dire $p_i = \omega_i/n$. La media $\mu$ delle telefonate al minuto si calcola quindi come

$$\mu = \bar{X}_n = \sum_{i=1}^{k} p_i\, x_i = \sum_{i=1}^{k} \frac{\omega_i\, x_i}{n} = 4.433$$

mentre per la varianza usiamo come al solito lo stimatore non distorto $S^2$:

$$\sigma^2 = S^2 = \frac{n}{n-1} \sum_{i=1}^{k} p_i\, (x_i - \mu)^2 = \frac{1}{n-1} \sum_{i=1}^{k} \omega_i\, (x_i - \mu)^2 = 5.03$$

e quindi $\sigma = 2.24$. Da notare che, nel calcolo della varianza, la media non è nota, ma dobbiamo usare la sua stima $\bar{X}_n$ ottenuta dai dati. Quindi, si deve usare lo stimatore $S^2$ per la varianza, con il fattore $n - 1$ al denominatore. Con la determinazione di $\mu$ e $\sigma^2$ siamo così in possesso della distribuzione normale ipotizzata, della quale ci serviremo per calcolare le frequenze attese.

Prestiamo ora attenzione agli intervalli in cui si deve suddividere il dominio della variabile $X$. Siccome $X$ è una variabile discreta, la scelta più conveniente è quella di introdurre intervalli centrati attorno ai valori $x_i$, aggiungendo poi due intervalli illimitati di coda. In realtà, la variabile $X$ non può assumere valori negativi (un numero negativo di telefonate al minuto non avrebbe significato); siccome, però, approssimiamo la distribuzione vera con una normale, e questa si estende su tutto $\mathbb{R}$, è necessario introdurre anche un intervallo di coda illimitato verso $-\infty$ (altrimenti la densità di probabilità non sarebbe normalizzata). La suddivisione in intervalli è raffigurata in Figura 1: $A_i = [a_i, b_i] \equiv [x_{i-1} - 0.5, x_{i-1} + 0.5]$ per $i = 2, 3, ..., 12$, mentre $A_1 = (-\infty, b_1] \equiv (-\infty, x_1 - 0.5]$ ed $A_{k+2} = [a_{k+2}, \infty) \equiv [x_k + 0.5, \infty)$ (con $k + 2 = 13$). Per calcolare le frequenze attese, dobbiamo prima calcolare la probabilità teorica che compete

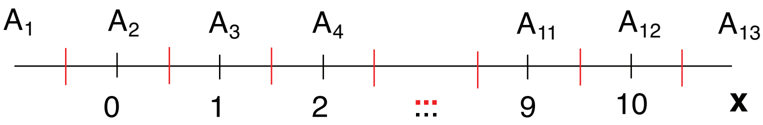


Figura 1: La suddivisione in intervalli centrati.

a ciascun valore $x_i$, che identifichiamo con la probabilità $\bar{p}_i$ da attribuire a ciascun intervallo, sulla base della distribuzione normale ipotizzata. Introducendo la variabile standardizzata

$$Z = \frac{X - \mu}{\sigma},$$

con i valori calcolati sopra per $\mu$ e $\sigma$, abbiamo per l'$i-$esimo intervallo:

$$\bar{p}_i = P(\{X \in A_i\}), \qquad i = 1, ..., k + 2$$

Per gli intervalli limitati abbiamo:

$$\bar{p}_i = P(\{X \in A_i\}) = P\left(\left\{\frac{a_i - \mu}{\sigma} < Z < \frac{b_i - \mu}{\sigma}\right\}\right) = \Phi\left(\frac{b_i - \mu}{\sigma}\right) - \Phi\left(\frac{a_i - \mu}{\sigma}\right)$$

dove $\Phi$ è la funzione di ripartizione della distribuzione normale standard, da calcolare usando le tavole o al computer. Per gli intervalli di coda invece abbiamo:

$$\bar{p}_1 = P(\{X \in A_1\}) = P\left(\left\{Z < \frac{b_1 - \mu}{\sigma}\right\}\right) = \Phi\left(\frac{b_1 - \mu}{\sigma}\right)$$
$$\bar{p}_{k+2} = P(\{X \in A_{k+2}\}) = P\left(\left\{\frac{a_{k+2} - \mu}{\sigma} < Z\right\}\right) = 1 - \Phi\left(\frac{a_{k+2} - \mu}{\sigma}\right)$$

Le frequenze attese $\Omega_i$ sono quindi date da

$$\Omega_i = \bar{p}_i\, n$$

Riassumiamo i risultati ottenuti nella seguente tabella:

| Intervallo | $x_i$ | freq. oss. ($\omega_i$) | $\bar{p}_i$ | freq. attese ($\Omega_i$) |
|---|---|---|---|---|
| $A_1$ | $< 0$ | 0 | 0.014 | 0.835 |
| $A_2$ | 0 | 2 | 0.026 | 1.55 |
| $A_3$ | 1 | 3 | 0.056 | 3.34 |
| $A_4$ | 2 | 7 | 0.099 | 5.93 |
| $A_5$ | 3 | 9 | 0.14 | 8.66 |
| $A_6$ | 4 | 11 | 0.17 | 10.4 |
| $A_7$ | 5 | 10 | 0.17 | 10.26 |
| $A_8$ | 6 | 8 | 0.14 | 8.33 |
| $A_9$ | 7 | 4 | 0.093 | 5.56 |
| $A_{10}$ | 8 | 3 | 0.051 | 3.05 |
| $A_{11}$ | 9 | 2 | 0.023 | 1.38 |
| $A_{12}$ | 10 | 1 | 0.085 | 0.511 |
| $A_{13}$ | $> 10$ | 0 | 0.0034 | 0.205 |

Prima di procedere con il calcolo della grandezza

$$w = \sum_i \frac{(\omega_i - \Omega_i)^2}{\Omega_i}$$

dobbiamo raggruppare le classi in modo che le frequenze attese siano $\geq 5$. Raggruppando gli intervalli $A_1$, $A_2$ ed $A_3$ e gli intervalli $A_{10}$, $A_{11}$, $A_{12}$ ed $A_{13}$ otteniamo la seguente tabella, in cui omettiamo l'indicazione degli intervalli:

| $x_i$ | freq. oss. ($\omega_i$) | $\bar{p}_i$ | freq. attese ($\Omega_i$) |
|---|---|---|---|
| $< 2$ | 5 | 0.096 | 5.726 |
| 2 | 7 | 0.099 | 5.93 |
| 3 | 9 | 0.14 | 8.66 |
| 4 | 11 | 0.17 | 10.4 |
| 5 | 10 | 0.17 | 10.26 |
| 6 | 8 | 0.14 | 8.33 |
| 7 | 4 | 0.093 | 5.56 |
| $> 7$ | 6 | 0.162 | 5.145 |

Il numero delle classi è ora uguale ad 8. Il calcolo esplicito sui dati dell'ultima tabella dà $w = 0.93$, mentre $\chi^2_{0.05}(5) = 11.07$ (dalle tavole). L'ipotesi nulla è pertanto accettata.