

Università di Cassino
Corso di Statistica 1
Esercitazione del 11/02/2008
Dott. Alfonso Piscitelli

Esercizio 1

Il seguente *data set* riporta la rilevazione di alcuni caratteri su un collettivo di 20 soggetti.

| Soggetto | Età | Residenza | Reddito (Migliaia di €) | Auto Possedute | Punteggio quiz | Km giornalieri percorsi |
|----------|-----|-----------------|----------------------------|-------------------|-------------------|-------------------------------|
| 1 | 22 | Cantagallo | 0,7 | 3 | 173 | 25 |
| 2 | 18 | Cantagallo | 0,2 | 4 | 168 | 25 |
| 3 | 34 | Poggio a Caiano | 1,6 | 2 | 165 | 21 |
| 4 | 42 | Carmignano | 2,5 | 5 | 180 | 25 |
| 5 | 50 | Poggio a Caiano | 3,2 | 3 | 163 | 17 |
| 6 | 12 | Montemurlo | 0,1 | 4 | 160 | 23 |
| 7 | 46 | Carmignano | 3,8 | 4 | 177 | 26 |
| 8 | 72 | Montemurlo | 1,3 | 2 | 164 | 35 |
| 9 | 27 | Montemurlo | 1,2 | 3 | 158 | 26 |
| 10 | 48 | Carmignano | 1,7 | 5 | 170 | 30 |
| 11 | 35 | Montemurlo | 1,9 | 1 | 167 | 21 |
| 12 | 84 | Cantagallo | 0,8 | 1 | 159 | 25 |
| 13 | 21 | Montemurlo | 0,4 | 5 | 174 | 26 |
| 14 | 44 | Carmignano | 1,8 | 4 | 164 | 33 |
| 15 | 56 | Carmignano | 1,9 | 2 | 177 | 24 |
| 16 | 58 | Montemurlo | 3,2 | 3 | 172 | 29 |
| 17 | 37 | Cantagallo | 2,1 | 1 | 166 | 14 |
| 18 | 16 | Montemurlo | 0,1 | 4 | 160 | 23 |
| 19 | 73 | Carmignano | 1,6 | 2 | 170 | 21 |
| 20 | 64 | Poggio a Caiano | 2,2 | 3 | 184 | 20 |

- a. Calcolare l'indice di Asimmetria di Fisher per il carattere **Età** suddiviso in 4 classi equiampie.
- b. Rappresentare il Box Plot per il carattere **Punteggio quiz**.
- c. Calcolare l'indice di Curtosi di Pearson per il carattere **Punteggio quiz**.

Soluzioni

a) L'indice di Fisher, è un indice di forma basato sui momenti terzi standardizzati:

$$\gamma = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^3 n_i$$

$\gamma > 0 \rightarrow$ *Asimmetrica Positiva;*
 $\gamma = 0 \rightarrow$ *Simmetrica;*
 $\gamma < 0 \rightarrow$ *Asimmetrica Negativa;*

Partendo dalla distribuzione in classi del carattere **Età**

| Età | n_i | X_i^c |
|--------|-------|---------|
| 12 -30 | 6 | 21 |
| 30 -48 | 6 | 39 |
| 48 -66 | 5 | 57 |
| 66 -84 | 3 | 75 |

Tot: 20

si calcola prima la media aritmetica

$$\mu = \frac{1}{N} \sum_{i=1}^c x_i^c * n_i = 43,5$$

e poi lo scarto quadratico medio

$$\sigma = \sqrt[2]{\frac{1}{N} \sum_{i=1}^c (x_i^c - \mu)^2 * n_i} = \sqrt[2]{\sigma^2} = \sqrt{352,4} = 18,77$$

Fatto ciò, si hanno tutti gli elementi per calcolare l'indice di asimmetria di Fisher

| Età | n_i | X_i^c | $(x_i^c - \mu)$ | $Z_i = \frac{(x_i^c - \mu)}{\sigma}$ | $Z_i = \left(\frac{(x_i^c - \mu)}{\sigma} \right)^3$ | $(Z_i)^3 * n_i$ |
|---------|-------|---------|-----------------|--------------------------------------|---|-----------------|
| 12 - 30 | 6 | 21 | -22,5 | -1,19872 | -1,72248 | -10,3349 |
| 30 - 48 | 6 | 39 | -4,5 | -0,23974 | -0,01378 | -0,08268 |
| 48 - 66 | 5 | 57 | 13,5 | 0,719233 | 0,372056 | 1,860281 |
| 66 -84 | 3 | 75 | 31,5 | 1,67821 | 4,726491 | 14,17947 |

Tot: 20

5,622182

$$\gamma = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^3 n_i = \frac{5,622182}{20} = 0,2811$$

Possiamo concludere che la distribuzione è caratterizzata da un'asimmetria positiva (indice maggiore di zero).

Tale risultato è confermato dal confronto tra la mediana e la media aritmetica.

$$\mu = 43,5 > M_e = 42$$

b) Il box-plot è una particolare rappresentazione di una distribuzione, dove è possibile evidenziare: la simmetria della distribuzione; la variabilità e la presenza di eventuali valori anomali.

| Soggetto | Punteggio quiz |
|----------|----------------|
| 1 | 158 |
| 2 | 159 |
| 3 | 160 |
| 4 | 160 |
| 5 | 163 |
| 6 | 164 |
| 7 | 164 |
| 8 | 165 |
| 9 | 166 |
| 10 | 167 |
| 11 | 168 |
| 12 | 170 |
| 13 | 170 |
| 14 | 172 |
| 15 | 173 |
| 16 | 174 |
| 17 | 177 |
| 18 | 177 |
| 19 | 180 |
| 20 | 184 |

A partire dalla distribuzione ordinata del carattere **Punteggio quiz**, si calcolano i valori della Mediana, del primo e del terzo Quartile.

$$M_e = 167,5; Q_1 = 163,5 \text{ e } Q_3 = 173,5.$$

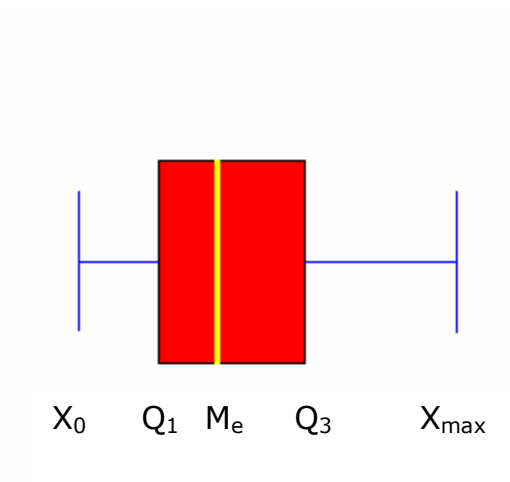
La Mediana, il primo e il terzo Quartile ci consentono di disegnare la scatola, mentre per disegnare i baffi dobbiamo prima calcolarci i Limiti Superiori ed Inferiori del box plot.

$$L_I = Q_1 - 1,5 * (Q_3 - Q_1)$$

Il baffo di sinistra (o di sotto) sarà il valore più grande tra L_I e X_{\min} . Nel nostro caso $L_I = 148,5$ mentre $X_{\min} = 158$. Il più grande è X_{\min} , quindi il primo baffo coinciderà con il valore di X_{\min} .

$$L_S = Q_3 + 1,5 * (Q_3 - Q_1)$$

Il baffo di destra (o quello di sopra) sarà il valore più piccolo tra L_S e X_{\max} . Nel nostro caso $L_S = 188,5$ mentre $X_{\max} = 184$. Il più piccolo è X_{\max} , quindi il secondo baffo coinciderà con il valore di X_{\max} .



- c)** La Curtosi riguarda un maggiore o minore appiattimento della forma della distribuzione. L'indice γ_C è un indice di forma basato sui momenti quarti standardizzati.

$$\gamma_C = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^4 - 3$$

$\gamma_C > 0 \rightarrow$ *Leptocurtica*;

$\gamma_C = 0 \rightarrow$ *Normocurtica*;

$\gamma_C < 0 \rightarrow$ *Platicurtica*;

Partendo dalla distribuzione del carattere **Punteggio quiz** si calcola la media aritmetica e lo scarto quadratico medio:

$$\mu = 168,55$$

$$\sigma = 7,145$$

| Soggetto | Punteggio quiz | $(x_i - \mu)$ | $Z_i = \frac{(x_i - \mu)}{\sigma}$ | $Z_i = \left(\frac{(x_i - \mu)}{\sigma}\right)^4$ |
|------------|----------------|---------------|------------------------------------|---|
| 1 | 158 | -10,55 | -1,47661 | 4,754019 |
| 2 | 159 | -9,55 | -1,33665 | 3,192012 |
| 3 | 160 | -8,55 | -1,19668 | 2,050763 |
| 4 | 160 | -8,55 | -1,19668 | 2,050763 |
| 5 | 163 | -5,55 | -0,77679 | 0,364102 |
| 6 | 164 | -4,55 | -0,63683 | 0,164474 |
| 7 | 164 | -4,55 | -0,63683 | 0,164474 |
| 8 | 165 | -3,55 | -0,49687 | 0,060949 |
| 9 | 166 | -2,55 | -0,35691 | 0,016226 |
| 10 | 167 | -1,55 | -0,21694 | 0,002215 |
| 11 | 168 | -0,55 | -0,07698 | 3,51E-05 |
| 12 | 170 | 1,45 | 0,202946 | 0,001696 |
| 13 | 170 | 1,45 | 0,202946 | 0,001696 |
| 14 | 172 | 3,45 | 0,482872 | 0,054366 |
| 15 | 173 | 4,45 | 0,622835 | 0,150484 |
| 16 | 174 | 5,45 | 0,762798 | 0,338561 |
| 17 | 177 | 8,45 | 1,182686 | 1,956491 |
| 18 | 177 | 8,45 | 1,182686 | 1,956491 |
| 19 | 180 | 11,45 | 1,602575 | 6,595884 |
| 20 | 184 | 15,45 | 2,162426 | 21,86578 |
| Tot | | | | 45,74148 |

$$\gamma_c = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^4 - 3 = \left(\frac{45,74148}{20} \right) - 3 = -0,713$$

La distribuzione è Platicurtica.