Università di Cassino

Corso di Statistica

Esercitazione del 5 febbraio 2010

Simona Balzano

ESERCIZIO 1

Data la seguente successione di valori del carattere x:

X: 3, 5, 6, 6, 7, 10, 12

calcolare:

a) la media geometrica
b) la media quadratica

Soluzione

a)
La media geometrica M_G è uguale a:

$$M_G = \sqrt[N]{\prod_{i=1}^{N} x_i} = \sqrt[7]{3 \times 5 \times 6 \times 6 \times 7 \times 10 \times 12} = \sqrt[7]{453600} = 453600^{\frac{1}{7}}$$

Passando ai logaritmi:

$$\mathit{ln}\, M_G = \frac{1}{7} \mathit{ln}\, 453600 = \frac{1}{7} 13{,}02 = 1{,}86$$

Quindi:

$$M_G = e^{1{,}86} = 6{,}43$$

Usando i logaritmi in base 10:

$$\mathit{log}\, M_G = \frac{1}{7} \mathit{log}\, 453600 = \frac{1}{7} 5{,}6567 = 0{,}8081$$

Quindi:

$$M_G = 10^{0{,}8081} = 6{,}43$$

In alternativa, poiché il logaritmo di un prodotto di n termini è uguale alla somma dei logaritmi degli n termini, si può scrivere:

$$\mathit{ln}\,M_G = \mathit{ln}\left(\prod_{i=1}^{N} x_i\right)^{\frac{1}{7}} = \frac{1}{7}\left(\mathit{ln}\,3 + \mathit{ln}\,5 + \mathit{ln}\,6 + \cdots + \mathit{ln}\,12\right) = 1,86$$

Quindi, ancora una volta:

$$M_G = e^{1,86} = 6,43$$

b)

La media quadratica M_2 è uguale a:

$$M_2 = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}} = \sqrt{\frac{3^2 + 5^2 + 6^2 + \cdots + 12^2}{7}} = \sqrt{\frac{399}{7}} = \sqrt{57} = 7,55$$

ESERCIZIO 2

Un individuo percorre la distanza in linea retta dal punto A al punto B alla velocità di 30 Km orari e la stessa distanza da B ad A alla velocità di 60 Km orari.
Determinare la velocità media su tutto il percorso A-B-A.

Soluzione

La velocità media va calcolata come media armonica delle due velocità:

$$M^{-1} = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}} = \frac{2}{\frac{1}{30} + \frac{1}{60}} = 40 Km/h$$

Se si usasse la media aritmetica $\bar{x} = (30 + 60)/2 = 45 Km/h$ si commetterebbe un errore.
Infatti la velocità media $\bar{v}$ su tutto il percorso va calcolata come rapporto tra spazio totale e tempo totale.
La distanza totale percorsa non è nota, ma si supponga che il tratto AB sia lungo 10 Km (il discorso è valido per qualsiasi distanza).
In tal caso, il tempo t_{AB} impiegato a percorrere il tratto AB sarebbe:

$$t_{AB} = \frac{10Km}{30Km/h} = \frac{1}{3} h,$$

mentre il tempo t_{BA} impiegato a percorrere il tratto BA sarebbe:

$$t_{BA} = \frac{10Km}{60Km/h} = \frac{1}{6} h,$$

per un tempo totale t pari a:

$$t = \frac{1}{3} + \frac{1}{6} = \frac{2+1}{6} = \frac{1}{2} = 0,5h.$$

La velocità media sara, quindi:

$$\bar{v} = \frac{\text{spazio tot}}{\text{tempo tot}} = \frac{AB + BA}{t} = \frac{20Km}{0,5h} = 40Km/h$$

che è il valore della media armonica delle due velocità (calcolata senza conoscere distanza e tempo).

ESERCIZIO 3

In riferimento al Dataset Emissioni atmosferiche – attività economiche (http://areadocenti.eco.unicas.it/balzano), per i 3 caratteri ANIDRIDE CARBONICA, MONOSSIDO DI CARBONIO ed EFFETTO SERRA determinare:

A. lo scostamento semplice medio dalla media
B. lo scostamento semplice medio dalla mediana;
C. il campo di variazione interquartilico;
D. disegnare il boxplot.

Soluzione

Ai fini del calcolo dei quartili, conviene innanzitutto ordinare le tre successioni:

Anidride carbonica
0,49
3,50
3,59
3,89
5,94
5,97
6,80
8,62
10,31
11,34
16,72
19,26
22,43
23,72
28,09
29,87
32,41
38,36
55,46
56,36

Monossido di carbonio
1,76
8,40
12,78
12,92
16,05
19,38
22,51
26,03
26,15
48,68
56,77
56,95
57,80
60,74
68,18
68,67
74,05
84,62
102,93
433,04

Effetto serra
0,68
4,09
4,84
5,01
7,23
7,39
8,42
10,41
11,94
14,99
17,83
25,78
26,86
31,71
33,04
38,36
38,85
43,34
59,69
68,71

Gli indici cercati per i tre caratteri possono essere determinati in base alle seguenti relazioni:

Scostamento semplice medio dalla media:

$$S(\mu) = \frac{\sum_{i=1}^{N} |x_i - \mu|}{N}$$

Scostamento semplice medio dalla mediana:

$$S(Me) = \frac{\sum_{i=1}^{N} |x_i - Me|}{N}$$

Differenza interquartile:

$$IQR = Q_3 - Q_1$$

A. Scostamento semplice medio dalla media

Anidride carbonica (AC)

$$\mu_{AC} = \frac{0,49+3,50+3,59+\cdots+56,36}{20} = 19,16$$

$$S(\mu)_{AC} = \frac{|0,49-19,16| + |3,5-19,16| + \cdots + |56,36-19,16|}{20} =$$
$$= \frac{18,67+15,66+\cdots+37,2}{20} = 13,36$$

Monossido di carbonio (MC)

$$\mu_{MC} = \frac{1,76+8,4+12,78+\cdots+433,04}{20} = 62,92$$

$$S(\mu)_{MC} = \frac{|1,76-62,92| + |8,4-62,92| + \cdots + |433,04-62,92|}{20} =$$
$$= \frac{61,16+54,52+\cdots+370,12}{20} = 45,4$$

Effetto serra (ES)

$$\mu_{ES} = \frac{0,68+4,09+4,84+\cdots+68,71}{20} = 22,96$$

$$S(\mu)_{ES} = \frac{|0,68-22,96| + |4,09-22,96| + \cdots + |68,71-22,96|}{20} =$$
$$= \frac{22,28+18,87+\cdots+45,76}{20} = 15,97$$

B. Scostamento semplice medio dalla mediana

Anidride carbonica (AC)

$$Me_{AC} = \frac{x_{10}+x_{11}}{2} = \frac{11,34+16,72}{2} = 14,03$$

$$S(Me)_{AC} = \frac{|0,49-14,03| + |3,5-14,03| + \cdots + |56,36-14,03|}{20} =$$
$$= \frac{13,54+10,53+\cdots+42,32}{20} = 13,11$$

Monossido di carbonio (MC)

$$Me_{MC} = \frac{x_{10} + x_{11}}{2} = \frac{48,68 + 56,77}{2} = 52,73$$

$$S(Me)_{MC} = \frac{|1,76 - 52,73| + |8,4 - 52,73| + \cdots + |433,04 - 52,73|}{20} =$$
$$= \frac{50,97 + 44,32 + \cdots + 380,31}{20} = 43,45$$

Effetto serra (ES)

$$Me_{ES} = \frac{x_{10} + x_{11}}{2} = \frac{14,99 + 17,83}{2} = 16,41$$
$$S(Me)_{ES} = \frac{|0,68 - 16,41| + |4,09 - 16,41| + \cdots + |68,71 - 16,41|}{20} =$$
$$= \frac{15,73 + 12,32 + \cdots + 52,3}{20} = 15,46$$

D. Campo di variazione interquartilico

Anidride carbonica

$$Q_1 = \frac{x_5 + x_6}{2} = 5,95$$
$$Q_2 = Me = 14,03$$
$$Q_3 = \frac{x_{15} + x_{16}}{2} = 28,98$$
$$\Delta Q_{AC} = Q_3 - Q_1 = 28,98 - 5,95 = 23,03$$

Monossido di carbonio

$$Q_1 = \frac{x_5 + x_6}{2} = \frac{16,05 + 19,38}{2} = 17,72$$
$$Q_2 = Me = 52,73$$
$$Q_3 = \frac{x_{15} + x_{16}}{2} = \frac{68,18 + 68,67}{2} = 68,42$$

$$\Delta Q_{MC} = Q_3 - Q_1 = 68,42 - 17,72 = 50,7$$

Effetto serra

$$Q_1 = \frac{x_5 + x_6}{2} = \frac{7,23 + 7,39}{2} = 7,31$$
$$Q_2 = Me = 16,41$$
$$Q_3 = \frac{x_{15} + x_{16}}{2} = \frac{33,04 + 38,36}{2} = 35,7$$

$$\Delta_{ES} = Q_3 - Q_1 = 35,7 - 7,31 = 28,39$$

D)

<u>Boxplot</u>

L'intervallo [α; β] rappresenta il massimo intervallo "teorico" tollerabile, oltre il quale eventuali valori presenti nella distribuzione devono essere considerati anomali.
L'intervallo [a; b] è il più ampio intervallo "reale" interno ad [α; β].
Eventuali valori presenti nella distribuzione ed esterni ad [a; b] (e ad [α; β]) sono da considerare anomali.

<u>Anidride carbonica</u>

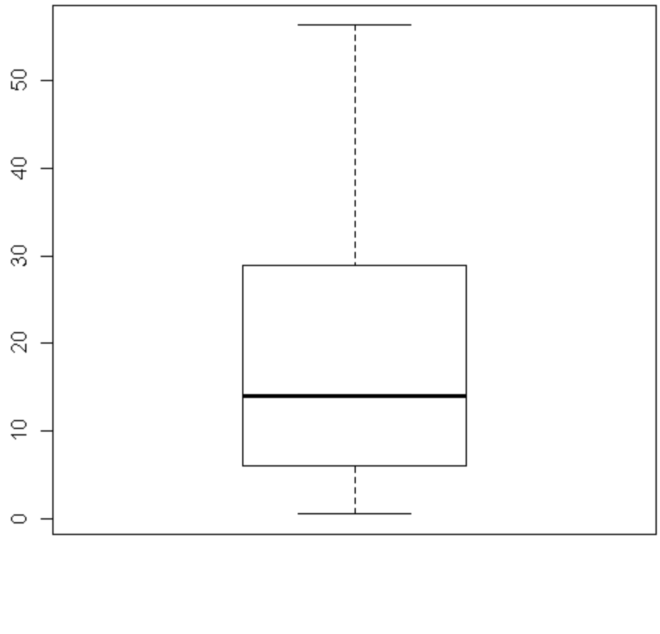
Q_1 = 5,95

Q_2 = Me = 14,03

Q_3 = 29,98

$$\alpha = Q_1 - 1,5(Q_3 - Q_1) = 5,95 - 1,5 \times 23,03 = -28,59 \quad \rightarrow \quad a = \min\{x_i \geq \alpha\} = 0,49$$

$$\beta = Q_3 + 1,5(Q_3 - Q_1) = 29,98 + 1,5 \times 23,03 = 63,53 \quad \rightarrow \quad b = \max\{x_i \leq \beta\} = 53,36$$



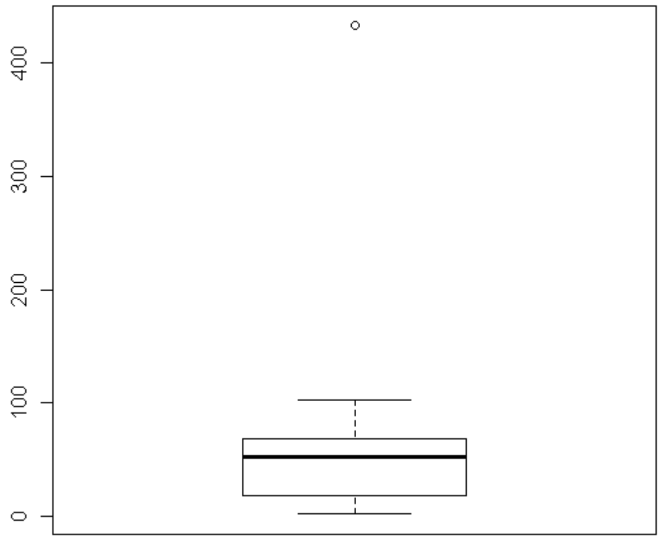
Non sono presenti "valori anomali".

Monossido di carbonio

$$Q_1 = 17,72$$
$$Q_2 = Me = 52,73$$
$$Q_3 = 68,42$$

$$\alpha = Q_1 - 1,5(Q_3 - Q_1) = 17,72 - 1,5 \times 50.7 = -58,35 \quad \rightarrow \quad a = \min\{x_i \geq \alpha\} = 1,76$$

$$\beta = Q_3 + 1,5(Q_3 - Q_1) = 68,42 + 1,5 \times 50,7 = 144,49 \quad \rightarrow \quad b = \max\{x_i \leq \beta\} = 102,93$$



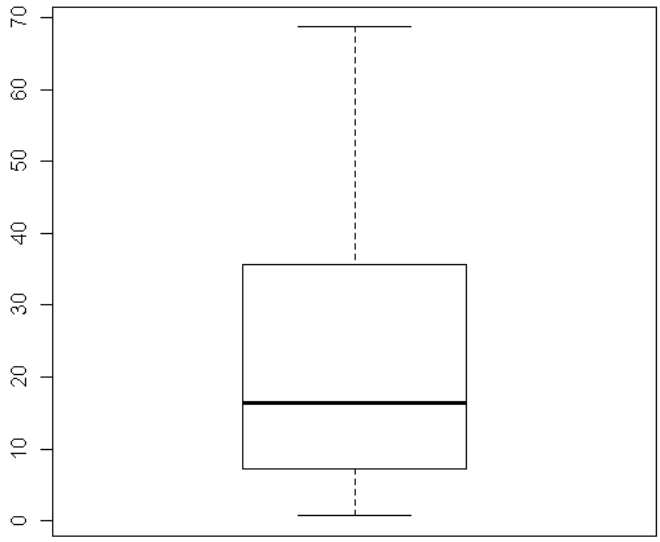
Il valore 433,04 è un “valore anomalo”.

Effetto serra

$$Q_1 = 7,31$$
$$Q_2 = Me = 16,41$$
$$Q_3 = 35,7$$

$$\alpha = Q_1 - 1,5(Q_3 - Q_1) = 7,31 - 1,5 \times 28,39 = -35,28 \quad \rightarrow \quad a = \min\{x_i \geq \alpha\} = 0,68$$

$$\beta = Q_3 + 1,5(Q_3 - Q_1) = 35,7 + 1,5 \times 28,39 = 78,28 \quad \rightarrow \quad b = \max\{x_i \leq \beta\} = 68,21$$



Non sono presenti valori anomali.

Esercizio 4

Data la seguente distribuzione di frequenza del carattere MONOSSIDO DI CARBONIO:

$x_{i-1} - x_i$	n_i	f_i	F_i
0 |-| 35	3	0.15	0.15
35 -| 90	6	0.30	0.45
90 -| 200	7	0.35	0.80
200 -| 350	4	0.20	1
	20	**1**	

determinarne lo scostamento semplice medio dalla media e dalla mediana.

Soluzione

La mediana è:

$$Me = x_{Me-1} + \left(0,5 - F_{Me-1}\right)\frac{b_{Me}}{f_{Me}} =$$

$$= 90 + \left(0,5 - 0,45\right)\frac{200-90}{0,35} = 105,7$$

La media è:

$$\bar{x} = \sum_{i=1}^{k} c_i f_i = 17,5 \times 0,15 + 62,5 \times 0,3 + 145 \times 0,35 + 275 \times 0,2 = 127,13$$

Scostamento semplice medio dalla media:

$$S(\mu) = \frac{\sum_{i=1}^{k} \left|c_i - \mu\right| n_i}{N}$$

Scostamento semplice medio dalla mediana:

$$S(Me) = \frac{\sum_{i=1}^{k} \left|c_i - Me\right| n_i}{N}$$

È necessario calcolare i valori centrali delle classi:

$$c_1 = \frac{x_0 + x_1}{2} = \frac{35}{2} = 17,5\,; \qquad c_2 = \frac{x_1 + x_2}{2} = \frac{35+90}{2} = 62,5\,;$$

$$c_3 = \frac{x_2 + x_3}{2} = \frac{90+200}{2} = 145\,; \qquad c_4 = \frac{x_3 + x_4}{2} = \frac{200+350}{2} = 275\,.$$

Scostamento semplice medio dalla media

$$S(\mu) = \frac{\sum_{i=1}^{k} |c_i - \mu| n_i}{N} = \sum_{i=1}^{k} |c_i - \mu| f_i =$$

$$= |17,5 - 127,13| \times 0,15 + |62,5 - 127,13| \times 0,30 +$$

$$+ |145 - 127,13| \times 0,35 + |275 - 127,13| \times 0,20 = 71,66$$

Scostamento semplice medio dalla mediana

$$S(Me) = \frac{\sum_{i=1}^{k} |c_i - Me| n_i}{N} \sum_{i=1}^{k} |c_i - Me| f_i =$$

$$= |17,5 - 105,7| \times 0,15 + |62,5 - 105,7| \times 0,30 +$$

$$+ |145 - 105,7| \times 0,35 + |275 - 105,7| \times 0,20 = 73,81$$