ESERCIZIO 3.1

In riferimento ai dati dell'esercizio 2 (esercitazione del 27-10-04), per i 3 caratteri VALORE DEL TERRENO, VALORE DELLE RISTRUTTURAZIONI e PREZZO determinare:

A. lo scostamento semplice medio dalla media
B. lo scostamento semplice medio dalla mediana;
C. la varianza e lo scarto quadratico medio;
D. il campo di variazione interquartilico.

Soluzione

Dovendo determinare mediane e quartili, conviene innanzitutto ordinare le tre successioni:

Valore del terreno	Valore delle ristrutturazioni	Prezzo
5,27	6,39	22,40
7,12	10,84	25,40
7,15	13,62	36,00
7,55	14,95	37,20
7,74	16,19	37,70
8,37	16,94	39,00
10,13	18,47	45,00
10,70	18,73	49,50
12,04	19,64	54,30
12,22	20,13	58,00
14,49	21,35	59,90
14,59	24,34	60,00
15,06	25,66	61,50
15,44	31,95	74,20
16,10	34,65	78,00
16,20	37,81	81,00
16,93	38,72	82,00
17,95	40,92	84,50
20,80	41,27	85,00
25,52	52,79	86,80
26,52	52,90	87,50
26,60	57,66	105,00
30,28	58,84	125,50
30,41	66,43	140,00
31,68	70,31	173,50
44,01	73,21	175,00
47,73	78,68	187,50
55,93	81,46	205,00
59,80	86,12	243,00
66,24	97,94	243,80

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

$$S(\mu) = \frac{\sum_{i=1}^{N} |x_i - \mu|}{N}; \qquad S(Me) = \frac{\sum_{i=1}^{N} |x_i - Me|}{N}; \qquad \sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}; \qquad \sigma = \sqrt{\sigma^2};$$

$$IQR = Q_3 - Q_1 .$$

A. Scostamento semplice medio dalla media

Valore del terreno

$$\mu_{valoredelterreno} = \frac{5{,}27 + 7{,}12 + \cdots + 66{,}24}{30} = \frac{680{,}57}{30} = 22{,}69$$

$$S(\mu)_{valoredelterreno} = \frac{|5{,}27 - 22{,}69| + |7{,}12 - 22{,}69| + \cdots + |66{,}24 - 22{,}69|}{30} = \frac{390{,}36}{30} = 13{,}01$$

Valore delle ristrutturazioni

$$\mu_{valoreristrutt.} = \frac{6{,}39 + 10{,}84 + \cdots + 97{,}94}{30} = \frac{1228{,}91}{30} = 40{,}96$$

$$S(\mu)_{valoreristrutt.} = \frac{|6{,}39 - 40{,}96| + |10{,}84 - 40{,}96| + \cdots + |97{,}94 - 40{,}96|}{30} = \frac{652{,}09}{30} = 21{,}74$$

Prezzo

$$\mu_{prezzo} = \frac{22{,}40 + 25{,}40 + \cdots + 243{,}80}{30} = \frac{2843{,}20}{30} = 94{,}77$$

$$S(\mu)_{prezzo} = \frac{|22{,}40 - 94{,}77| + |25{,}40 - 94{,}77| + \cdots + |243{,}80 - 94{,}77|}{30} = \frac{1.490{,}78}{30} = 49{,}64$$

B. Scostamento semplice medio dalla mediana

Valore del terreno

$$Me_{valoredelterreno} = \frac{x_{15} + x_{16}}{2} = \frac{16{,}10 + 16{,}20}{2} = 16{,}15$$

$$S(Me)_{valoredelterreno} = \frac{|5{,}27 - 16{,}16| + |7{,}12 - 16{,}15| + \cdots + |66{,}24 - 16{,}15|}{30} = \frac{352{,}63}{30} = 11{,}75$$

Valore delle ristrutturazioni

$$Me_{valoreristrutt.} = \frac{x_{15} + x_{16}}{2} = \frac{34{,}65 + 37{,}81}{2} = 36{,}23$$

$$S(Me)_{valoreristrutt.} = \frac{|6{,}39 - 36{,}23| + |10{,}84 - 36{,}23| + \cdots + |97{,}94 - 36{,}23|}{30} = \frac{641{,}21}{30} = 21{,}37$$

Prezzo

$$Me_{prezzo} = \frac{x_{15} + x_{16}}{2} = \frac{78 + 81}{2} = 79{,}50$$

$$S(Me)_{prezzo} = \frac{|22{,}40 - 79{,}50| + |25{,}40 - 79{,}50| + \cdots + |243{,}80 - 79{,}50|}{30} = \frac{1367}{30} = 45{,}57$$

C. Varianza e scarto quadratico medio

Valore del terreno

$$\sigma^2_{valoredelterreno} = \frac{(5{,}27 - 22{,}69)^2 + (7{,}12 - 22{,}69)^2 + \cdots + (66{,}24 - 22{,}69)^2}{30} = \frac{8.054{,}30}{30} = 268{,}48$$

$$\sigma_{valoredelterreno} = 16{,}39$$

Valore delle ristrutturazioni

$$\sigma^2_{valoreristrutt.} = \frac{(6{,}39 - 40{,}96)^2 + (10{,}84 - 40{,}96)^2 + \cdots + (97{,}94 - 40{,}96)^2}{30} = \frac{19.416{,}61}{30} = 647{,}22$$

$$\sigma_{valoreristrutt.} = 25{,}44$$

Prezzo

$$\sigma^2_{prezzo} = \frac{(22{,}40 - 94{,}77)^2 + (25{,}40 - 94{,}77)^2 + \cdots + (243{,}80 - 94{,}77)^2}{30} = \frac{116.269{,}08}{30} = 3.865{,}64$$

$$\sigma_{prezzo} = 62{,}25$$

D. Campo di variazione interquartilico

Valore del terreno

$Q_1 = x_8 = 10{,}70$

$Q_2 = Me = 16{,}15$

$Q_3 = x_{23} = 30{,}28$

$$IQR_{valoredelterreno} = Q_3 - Q_1 = 30{,}28 - 10{,}70 = 20{,}15$$

Valore delle ristrutturazioni

$Q_1 = x_8 = 18{,}47$

$Q_2 = Me = 36{,}23$

$Q_3 = x_{23} = 58{,}84$

$$IQR_{valoreristrutt.} = Q_3 - Q_1 = 58{,}84 - 18{,}47 = 40{,}37$$

Prezzo

$Q_1 = x_8 = 45$

$Q_2 = Me = 79{,}5$

$Q_3 = x_{23} = 125{,}5$

$$IQR_{prezzo} = Q_3 - Q_1 = 125{,}5 - 45 = 80{,}5$$

ESERCIZIO 3.2

In riferimento ai dati dell'esercizio 2 (esercitazione del 27-10-04), per il carattere QUARTIERE si determini:

A. l'indice di eterogeneità (o mutabilità) di Gini;
B. l'indice di diversità di Shannon (o indice di entropia).

Soluzione

Gli indici richiesti vanno entrambi calcolati a partire dalle frequenze relative:

Quartiere	$\mathbf{n_i}$	$\mathbf{f_i}$
B	6	0,20
C	5	0,17
D	4	0,13
E	4	0,13
F	3	0,10
G	8	0,27
Totale	**30**	**1,00**

A.

Indice di eterogeneità di Gini:

$$G = 1 - \sum_{i=1}^{K} f_i^2 = 1 - (0{,}04 + 0{,}03 + 0{,}02 + 0{,}02 + 0{,}01 + 0{,}07) = 1 - 0{,}18 = 0{,}82$$

$$G_{max} = \frac{K-1}{K} = \frac{5}{6} = 0{,}83$$

Indice di eterogeneità normalizzato di Gini ($0 \leq G^* \leq 1$):

$$G^* = \frac{G}{G_{max}} = \frac{0{,}82}{0{,}83} = 0{,}98$$

B.

Indice di diversità (o di entropia) di Shannon:

$$H = -\sum_{i=1}^{K} f_i \cdot \mathit{ln}(f_i) =$$

$$= -[0{,}2 \times \mathit{ln}(0{,}2) + 0{,}17 \times \mathit{ln}(0{,}17) + 0{,}13 \times \mathit{ln}(0{,}13) + 0{,}13 \times \mathit{ln}(0{,}13) + 0{,}27 \times \mathit{ln}(0{,}27)] =$$

$$= -[0{,}2 \times (-1{,}61) + 0{,}17 \times (-1{,}79) + 0{,}13 \times (-2{,}01) + 0{,}13 \times (-2{,}01) + 0{,}27 \times (-1{,}32)] = 1{,}74$$

$$H_{max} = \mathit{log}(K) = \mathit{log}(6) = 1{,}79$$

Indice normalizzato di Shannon ($0 \leq H^* \leq 1$):

$$H^* = \frac{H}{H_{max}} = \frac{1{,}74}{1{,}79} = 0{,}97$$

<u>Conclusione</u>: G^* ed H^* molto prossimi ad 1 $\rightarrow$ eterogeneità delle osservazioni molto elevata (frequenze ben distribuite fra le diverse modalità, ossia molto vicine ad 1/K).

ESERCIZIO 3.3

Si consideri la seguente distribuzione di N = 7 imprese per numero di addetti:

Impresa	Addetti
A	3
B	6
C	5
D	10
E	3
F	6
G	7

Si determini:

A. lo scarto quadratico medio del numero di addetti;
B. la differenza semplice media.

Soluzione

A.
Numero medio di addetti:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{3+6+5+10+3+6+7}{7} = \frac{40}{7} = 5{,}71$$

Scarto quadratico medio:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}} =$$

$$= \sqrt{\frac{(3-5{,}71)^2 + (6-5{,}71)^2 + (5-5{,}71)^2 + (10-5{,}71)^2 + (3-5{,}71)^2 + (6-5{,}71)^2 + (7-5{,}71)^2}{7}} =$$

$$= \sqrt{\frac{7{,}367 + 0{,}082 + 0{,}51 + 18{,}367 + 7{,}367 + 0{,}082 + 1{,}653}{7}} = \sqrt{5{,}061} = 2{,}25$$

B.

La differenza semplice media senza ripetizione si ottiene come:

$$\Delta = \frac{\sum_{i \neq j=1}^{N} |x_i - x_j|}{N(N-1)}$$

Gli scarti semplici in valore assoluto sono contenuti nella seguente tabella:

		A	B	C	D	E	F	G
	x_i	3	6	5	10	3	6	7
A	3	0	3	2	7	0	3	4
B	6	3	0	1	4	3	0	1
C	5	2	1	0	5	2	1	2
D	10	7	4	5	0	7	4	3
E	3	0	3	2	7	0	3	4
F	6	3	0	1	4	3	0	1
G	7	4	1	2	3	4	1	0

La loro media è:

$$\Delta = \frac{\sum_{i \neq j=1}^{N} |x_i - x_j|}{N(N-1)} = \frac{120}{7 \times 6} = 2{,}86 \, .$$

$$\Delta_{MAX} = 2\mu = 2 \times 5{,}71 = 11{,}43$$

Indice normalizzato di mutua variabilità $0 \leq R \leq 1$:

$$R = \frac{\Delta}{\Delta_{max}} = \frac{2{,}86}{11{,}43} = 0{,}25$$

<u>Conclusione</u>: R è più prossimo a 0 che ad 1 → le osservazioni si differenziano poco l'una dall'altra.

ESERCIZIO 3.4

Si determini la differenza semplice media per la seguente distribuzione di frequenza, relativa al carattere X = NUMERO DI AUTO relativo ad un collettivo di 20 famiglie:

auto	n_i
1	3
2	8
3	6
4	2
5	1
	20

Soluzione

La differenza semplice media per una distribuzione di frequenze è data da:

$$\Delta = \frac{\sum_{i \neq j=1}^{K} |x_i - x_j| n_i \cdot n_j}{N(N-1)}$$

dove K indica il numero di modalità che il carattere assume.

La tabella riassume le differenze in valore assoluto tra le modalità, moltiplicate per le rispettive frequenze (riportate nella riga e nella colonna di intestazione, accanto alle modalità stesse):

	n_j	3	8	6	2	1
n_i	**auto**	1	2	3	4	5
3	1	0	24	36	18	12
8	2	24	0	48	32	24
6	3	36	48	0	12	12
2	4	18	32	12	0	2
1	5	12	24	12	2	0

$$\Delta = \frac{\sum_{i \neq j=1}^{K} |x_i - x_j| n_i \cdot n_j}{N(N-1)} = \frac{(|1-2| \cdot 3 \cdot 8) + (|1-3| \cdot 3 \cdot 6) + \cdots + (|5-4| \cdot 1 \cdot 2)}{20 \cdot 19} = \frac{440}{380} = 1{,}16$$

$$\Delta_{MAX} = 2\mu = 2 \times 2{,}5 = 5$$

Indice normalizzato di mutua variabilità $0 \leq R \leq 1$:

$$R = \frac{\Delta}{\Delta_{max}} = \frac{1{,}16}{5} = 0{,}23$$

<u>Conclusione</u>: R è più prossimo a 0 che ad 1 $\rightarrow$ le osservazioni si differenziano poco l'una dall'altra.