

Esercitazione 6

ESERCIZIO 1

Sono stati rilevati i seguenti pesi, espressi in Kg, per due gruppi di atleti differenti, rispettivamente femmine e maschi.

i	x_i^F
1	51
2	56
3	58
4	59
5	62

i	x_i^M
6	75
7	77
8	82
9	84

Verificare la proprietà di scomposizione della varianza.

SOLUZIONE

Abbiamo che

$$M_1^F = \frac{1}{5} \sum_{i=1}^5 x_i^F = 57,2$$

$$M_1^M = 79,5$$

$$M_1 = 67,11$$

Le varianze quindi diventano

$$\sigma_F^2 = \frac{1}{5} \sum_{i=1}^5 (x_i^F - M_1^F)^2 = \frac{1}{5} 66,8 = 13,36$$

$$\sigma_M^2 = \frac{1}{4} \sum_{i=6}^9 (x_i^M - M_1^M)^2 = \frac{1}{4} 53 = 13,25$$

$$\sigma^2 = \frac{1}{9} \sum_{i=1}^9 (x_i - M_1)^2 = \frac{1}{9} 1224,89 = 136,10$$

applicando la formula di scomposizione della varianza abbiamo

$$\begin{aligned} \sigma^2 &= \frac{1}{n} [n^F \sigma_F^2 + n^M \sigma_M^2] + \frac{1}{n} [n^F \times (M_1^F - M_1)^2 + n^M \times (M_1^M - M_1)^2] = \\ &= \frac{1}{9} [5 \times 13,36 + 4 \times 13,25] + \frac{1}{9} [5 \times (57,2 - 67,11)^2 + 4 \times (79,5 - 67,11)^2] = \\ &= \frac{1}{9} [119,8] + \frac{1}{9} [1105,09] = 136,10 \end{aligned}$$

che coincide con il valore precedente.

ESERCIZIO 2

Calcolare e commentare la differenza media semplice e la differenza media con ripetizione per i seguenti valori, senza usare la formula di calcolo per i valori ordinati e utilizzando la formula. Confrontare i risultati ottenuti.

i	x_i
1	32
2	11
3	21
4	68
5	5

SOLUZIONE

$ x_i - x_j $	x_1	x_2	x_3	x_4	x_5	$ x_i - x_j $	x_1	x_2	x_3	x_4	x_5
x_1	$ 32 - 32 $	$ 32 - 11 $	$ 32 - 21 $	$ 32 - 68 $	$ 32 - 5 $	x_1	0	21	11	36	27
x_2	$ 11 - 32 $	$ 11 - 11 $	$ 11 - 21 $	$ 11 - 68 $	$ 11 - 5 $	x_2	21	0	10	57	6
x_3	$ 21 - 32 $	$ 21 - 11 $	$ 21 - 21 $	$ 21 - 68 $	$ 21 - 5 $	x_3	11	10	0	47	16
x_4	$ 68 - 32 $	$ 68 - 11 $	$ 68 - 21 $	$ 68 - 68 $	$ 68 - 5 $	x_4	36	57	47	0	63
x_5	$ 5 - 32 $	$ 5 - 11 $	$ 5 - 21 $	$ 5 - 68 $	$ 5 - 5 $	x_5	27	6	16	63	0

La differenza media semplice si ottiene come la media delle distanze tra i valori considerati, escludendo dal calcolo le distanze tra un valore e se stesso (quindi, escludendo la diagonale dalla tabella precedente)

$ x_i - x_j $	x_1	x_2	x_3	x_4	x_5
x_1		21	11	36	27
x_2	21		10	57	6
x_3	11	10		47	16
x_4	36	57	47		63
x_5	27	6	16	63	

Indicando con Δ la differenza media semplice, si ottiene come

$$\Delta = \frac{1}{n \times (n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n |x_i - x_j| = \frac{1}{5 \times 4} (588) = 29,4$$

La differenza media con ripetizione si ottiene considerando anche la diagonale delle matrici precedenti

$ x_i - x_j $	x_1	x_2	x_3	x_4	x_5
x_1	0	21	11	36	27
x_2	21	0	10	57	6
x_3	11	10	0	47	16
x_4	36	57	47	0	63
x_5	27	6	16	63	0

Indicando con Δ_R la differenza media semplice, si ottiene come

$$\Delta_R = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| = \frac{1}{5^2} (588) = 23,52$$

essendo $n > n - 1$ allora $\frac{1}{n} < \frac{1}{n-1}$ quindi $\Delta_R < \Delta$, qualunque sia $n > 2.7$
 La differenza, per $n \rightarrow +\infty$, tende ad annullarsi, in quanto

$$\lim_{n \rightarrow +\infty} \frac{n-1}{n} = \lim_{n \rightarrow +\infty} \frac{n}{n-1} = 1$$

Stando alla formula per i valori ordinati, abbiamo che, considerando solo una matrice triangolare:

$ x_i - x_j $	$x(1)$	$x(2)$	$x(3)$	$x(4)$	$x(5)$
$x(1)$	$x(1) - x(1)$				
$x(2)$	$x(2) - x(1)$	$x(2) - x(2)$			
$x(3)$	$x(3) - x(1)$	$x(3) - x(2)$	$x(3) - x(3)$		
$x(4)$	$x(4) - x(1)$	$x(4) - x(2)$	$x(4) - x(3)$	$x(4) - x(4)$	
$x(5)$	$x(5) - x(1)$	$x(5) - x(2)$	$x(5) - x(3)$	$x(5) - x(4)$	$x(5) - x(5)$

Dove notiamo una certa relazione: sommando tutta la matrice triangolare, il primo valore viene aggiunto una volta e sottratto 5 (n) volte.

$ x_i - x_j $	$x(1)$	$x(2)$	$x(3)$	$x(4)$	$x(5)$
$x(1)$	$x(1) - x(1)$				
$x(2)$	$x(2) - x(1)$	$ x(2) - x(2) $			
$x(3)$	$x(3) - x(1)$	$ x(3) - x(2) $	$ x(3) - x(3) $		
$x(4)$	$x(4) - x(1)$	$ x(4) - x(2) $	$ x(4) - x(3) $	$ x(4) - x(4) $	
$x(5)$	$x(5) - x(1)$	$ x(5) - x(2) $	$ x(5) - x(3) $	$ x(5) - x(4) $	$ x(5) - x(5) $

Il secondo numero viene aggiunto due volte e sottratto 4 (n-1) volte.

$ x_i - x_j $	$x(1)$	$x(2)$	$x(3)$	$x(4)$	$x(5)$
$x(1)$	$x(1) - x(1)$				
$x(2)$	$x(2) - x(1)$	$x(2) - x(2)$			
$x(3)$	$ x(3) - x(1) $	$ x(3) - x(2) $	$ x(3) - x(3) $		
$x(4)$	$ x(4) - x(1) $	$ x(4) - x(2) $	$ x(4) - x(3) $	$ x(4) - x(4) $	
$x(5)$	$ x(5) - x(1) $	$ x(5) - x(2) $	$ x(5) - x(3) $	$ x(5) - x(4) $	$ x(5) - x(5) $

Etc. Quindi abbiamo

$$\begin{aligned} x(1) - n \times x(1) &= x(1) \times (1 - n) \\ 2 \times x(2) - (n-1) \times x(1) &= x(1) \times (2 - (n-1)) \\ &\dots \\ n \times x(n) - x(n) &= x(n) \times (n-1) \end{aligned}$$

La somma totale delle quantità precedenti è pari a

$$\sum_{i=1}^n x_i \times (2 \times i - n - 1)$$

ed essendo la matrice di partenza simmetrica (speculare rispetto alla diagonale) la somma totale è due volte la somma precedente:

$$S = 2 \times \sum_{i=1}^n x_i \times (2 \times i - n - 1)$$

La differenza media semplice e la differenza media con ripetizione possono essere ottenute a partire da questa somma

$$\Delta = \frac{S}{n \times (n-1)} = \frac{1}{n \times (n-1)} 2 \times \sum_{i=1}^n x_{(i)} \times (2 \times i - n - 1) = 29,4$$

$$\Delta_R = \frac{S}{n^2} = \frac{1}{n^2} 2 \times \sum_{i=1}^n x_{(i)} \times (2 \times i - n - 1) = 23,52$$

ESERCIZIO 3

Calcolare e commentare la differenza media semplice e la differenza media con ripetizione per la seguente distribuzione di frequenze

i	x_i	n_i
1	5	9
2	11	4
3	21	6
4	28	2
5	32	6
6	68	8
7	87	9

SOLUZIONE

i	$x_{(i)}$	n_i	C_i	$2 \times C_i - n - n_i$	$x_{(i)} n_i \times (2 \times C_i - n - n_i)$
1	5	9	9	-35	-1575
2	11	4	13	-22	-968
3	21	6	19	-12	-1512
4	28	2	21	-4	-224
5	32	6	27	4	768
6	68	8	35	18	9792
7	87	9	44	35	27405

Usiamo la formula di calcolo per i valori ordinati. Abbiamo che, la generalizzazione per il caso con le frequenze assolute è data dalla seguente

$$S = 2 \times \sum_{i=1}^k x_{(i)} n_i \times (2 \times C_i - n - n_i)$$

dove C_i sono le frequenze cumulate, ed otteniamo $\Delta = \frac{S}{n \times (n-1)}$ e $\Delta_R = \frac{S}{n^2}$. Dai dati precedenti abbiamo che

$$S = 2 \times 33686 = 67372$$

$$\Delta = \frac{S}{n \times (n-1)} = 35,61$$

$$\Delta_R = \frac{S}{n^2} = 34,80$$

ESERCIZIO 4

Un'azienda importa dagli Stati Uniti 5 differenti prodotti. Nella tabella seguente sono riportati costo del singolo prodotto, espresso in dollari, e numero di prodotti importati, per tipo, durante l'anno 2015

i	x_i	n_i
1	159,00	112
2	199,00	27
3	239,00	51
4	249,00	123
5	299,00	87

Calcolare la varianza del costo dei prodotti. Sapendo inoltre $1\$ = 0,886\text{€}$, e che per ogni prodotto l'azienda ha un costo fisso di spedizione di 8€ , calcolare la varianza del prodotto, spedizione inclusa, espresso in euro. Controllare che sia rispettata la prima proprietà della varianza, relativa alle trasformazioni lineari.

SOLUZIONE

i	x_i	n_i	$n_i \times x_i$	$n_i \times x_i^2$	y_i	$n_i \times y_i$	$n_i \times y_i^2$
1	159,00	112	17808	2831472	148,874	16673,888	2482308,402
2	199,00	27	5373	1069227	184,314	4976,478	917234,566
3	239,00	51	12189	2913171	219,754	11207,454	2462882,846
4	249,00	123	30627	7626123	228,614	28119,522	6428516,403
5	299,00	87	26013	7777887	272,914	23743,518	6479938,471

Il calcolo della varianza può essere ottenuto tramite l'applicazione della formula indiretta, ovvero

$$\sigma_{\$}^2 = \sum_{i=1}^k n_i \times (x_i - M_1^X)^2 = M_2^X - [M_1^X]^2$$

dove

$$M_1^X = \frac{1}{n} \sum_{i=1}^k n_i \times x_i = 230,025$$

$$M_2^X = \frac{1}{n} \sum_{i=1}^k n_i \times x_i^2 = 55544,70$$

e quindi la varianza diventa

$$\sigma_{\$}^2 = M_2^X - [M_1^X]^2 = 55544,70 - [230,025]^2 = 2633,199$$

Sia Y il costo espresso in euro, comprensivo di spese di spedizione, $y_i = 0,886 \times x_i + 8$, la varianza diventa

$$\sigma_{\text{€}}^2 = M_2^Y - [M_1^Y]^2$$

dove

$$M_1^Y = \frac{1}{n} \sum_{i=1}^k n_i \times y_i = 211,802$$

$$M_2^Y = \frac{1}{n} \sum_{i=1}^k n_i \times y_i^2 = 46927,202$$

ed otteniamo

$$\sigma_{\mathfrak{E}}^2 = M_2^Y - [M_1^Y]^2 = 46927,202 - [211,802]^2 = 2067,051$$

Supporta la prima proprietà della varianza, infatti

$$0,886^2 \times \sigma_{\mathfrak{S}}^2 = 0,886^2 \times 2633,199 = 2067,051$$