

## Esercizio 1.

Si è analizzato un campione di 770 incidenti motociclistici verificatisi a Roma nel 1980 e nel 1981. Gli incidenti sono stati classificati secondo la gravità del danno alla testa sofferto dal motociclista.

0= nessun danno, 1= minore, 2= moderato,  
3= grave senza pericolo di vita, 4= grave con peric. di vita,  
5= critico, sopravvivenza incerta, 6= fatale.

In **331** incidenti il motociclista **indossava il casco**  
mentre negli altri **439 non indossava il casco.**

I dati sono stati riassunti in una tabella in cui:

- la prima colonna indica il valore,
- la seconda la frequenza assoluta con casco
- la terza la frequenza assoluta senza casco.

Valore	Freq. ass. con casco	Freq. ass. senza casco
0	248	227
1	58	135
2	11	33
3	3	14
4	2	3
5	8	21
6	1	6
<b>TOTALE</b>	<b>331</b>	<b>439</b>

**Determinare la media e la mediana campionaria** del danno alla testa sia per i motociclisti con casco che per quelli senza.

Indichiamo con

$x$  la variabile numerica corrispondente ai motociclisti **con casco**

$y$  la variabile numerica corrispondente ai motociclisti **senza casco**.

Valore	Freq. ass. con casco: $x$	Freq. ass. senza casco: $y$
0	248	227
1	58	135
2	11	33
3	3	14
4	2	3
5	8	21
6	1	6
<b>TOTALE</b>	<b>331</b>	<b>439</b>

La media campionaria con casco è

$$\bar{x} = \frac{0 \cdot 248 + 1 \cdot 58 + 2 \cdot 11 + 3 \cdot 3 + 4 \cdot 2 + 5 \cdot 8 + 6 \cdot 1}{331} = 0.43$$

La media campionaria senza casco è

$$\bar{y} = \frac{0 \cdot 227 + 1 \cdot 135 + 2 \cdot 33 + 3 \cdot 14 + 4 \cdot 3 + 5 \cdot 21 + 6 \cdot 6}{439} = 0.90$$

I dati indicano che i motociclisti con casco hanno sofferto, **in media**, danni minori di quelli senza casco.

Valore	Freq. ass. con casco	Freq. ass. senza casco
0	248	227
1	58	135
2	11	33
3	3	14
4	2	3
5	8	21
6	1	6
<b>TOTALE</b>	<b>331</b>	<b>439</b>

Le mediane campionarie con e senza casco = 0 entrambe. La mediana campionaria si calcola mettendo i dati in ordine crescente se  **$n$  è dispari** (come nel nostro caso) allora  **$m = x_{\frac{n+1}{2}}$**  **Con il**

**casco:**  $\frac{n+1}{2} = \frac{332}{2} = 166$   **$x_{166} = 0$**

**Senza casco:**  $\frac{n+1}{2} = \frac{440}{2} = 220$   **$x_{220} = 0$ .**

## Esercizio 2.

I seguenti dati forniscono il fatturato annuo in milioni di vecchie lire di **171** industrie tessili.

Fatturato	Freq. ass.
[300,500]	20
(500,800]	45
(800,1500]	56
(1500,2000]	50
<b>TOTALE</b>	<b>171</b>

- Qual è la **percentuale di industrie** con fatturato annuo **superiore a 500 milioni e non superiore a 1.5 miliardi?**
- Quale è la **classe modale** del fatturato?
- Quale è il **fatturato medio?**

Classe	Freq. ass.	Freq. rel.	Ampiezza della classe
[300,500]	20	20/171	200
(500,800]	45	45/171	300
(800,1500]	56	56/171	700
(1500,2000]	50	50/171	500
<b>TOTALE</b>	<b>171</b>	<b>1</b>	

- Quale è la percentuale di industrie con fatturato annuo **superiore a 500 milioni e non superiore a 1.5 miliardi?**

È la somma delle frequenze relative delle classi (500, 800] e (800, 1500] moltiplicata per 100.

$$\left( \frac{45}{171} + \frac{56}{171} \right) \cdot 100 = \frac{101}{171} \cdot 100 = 59.06\%$$

Classe	Freq. ass.	Freq. rel.	Ampiezza della classe
[300,500]	20	20/171	200
(500,800]	45	45/171	300
(800,1500]	56	56/171	700
(1500,2000]	50	50/171	500
<b>TOTALE</b>	<b>171</b>	<b>1</b>	

- Quale è la **classe modale** del fatturato?

È la classe con la frequenza più elevata (ossia quella corrispondente alla **moda**), quindi è **(800, 1.500]**.

- Quale è il **fatturato medio**?

Il calcolo esatto non è possibile ma è sensato supporre che i dati di ogni classe siano ben approssimati dal **valore centrale della classe**

$$\bar{x} = \frac{1}{171} (400 \cdot 20 + 650 \cdot 45 + 1.150 \cdot 56 + 1.750 \cdot 50) = \mathbf{1.106,14}$$

### Esercizio 3.

Si considerino i dati

2   0   3   2   4   **x**

per quali dei seguenti valori di **x** la mediana vale 2?

☐ 0

☐ 3

☐ 4

☐ 3.5

La taglia del campione è 6, numero pari. Quindi la mediana è la media aritmetica tra il terzo ed il quarto valore  $m = \frac{x_3 + x_4}{2}$ .

Se **x** = 0 disponendo i dati in ordine crescente si ha

**0**   0   2   2   3   4

e quindi la mediana è 2.



#### Esercizio 4.

Un lavoratore neoassunto per decidere se usare la metropolitana o l'automobile per recarsi al lavoro fa la seguente indagine:  
va a lavoro per i primi 12 giorni in automobile e per i successivi 12 in metropolitana e registra il tempo impiegato in minuti.

<b>Giorno</b>	1	2	3	4	5	6	7	8	9	10	11	12
<b>Auto</b>	23	32	44	21	36	30	28	33	45	34	29	31
<b>Metro</b>	22	24	22	33	26	31	24	28	32	31	37	24

$\bar{x}_A$  **media campionaria auto:**

$$\bar{x}_A = \frac{1}{12} (23+32+44+21+36+30+28+33+45+34+29+31) = 32,167$$

$x_M$  **media campionaria metro:**

$$\bar{x}_M = \frac{1}{12} (22+24+22+33+26+31+24+28+32+31+37+24) = 27,833$$

## INDICI DI DISPERSIONE

Consideriamo le seguenti due variabili  $x$  e  $y$  entrambe di taglia 5:

$$\begin{array}{rcccccc} x : & 14 & 15 & 16 & 16 & 19 \\ y : & 0 & 1 & 16 & 17 & 46 \end{array}$$

Queste variabili hanno la stessa mediana  $m = 16$  e la stessa media:

$$\begin{aligned}\bar{x} &= \frac{14 + 15 + 2 \cdot 16 + 19}{5} = \frac{80}{5} = 16 \\ \bar{y} &= \frac{1 + 16 + 17 + 46}{5} = \frac{80}{5} = 16\end{aligned}$$

Ma i dati sono molto diversi: i valori di  $x$  sono vicini mentre quelli di  $y$  sono sparsi. Si rende necessario introdurre un **indice di dispersione**.

## VARIANZA CAMPIONARIA

Data una variabile  $x$  di taglia  $n$  e di valori  $x_1, \dots, x_n$ , i seguenti numeri

$$x_i - \bar{x}, \quad i = 1, 2, \dots, n$$

si chiamano **scarti dalla media**. La somma degli scarti è nulla:

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) = x_1 + x_2 + \dots + x_n - n\bar{x} = 0$$

Per avere un indice di dispersione significativo si introduce la **varianza campionaria**

$$\begin{aligned} s^2 &= \frac{1}{n-1} \left[ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \right] \\ &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

Calcoliamo le varianze dei dati dell'esempio considerato prima:

$$\begin{array}{rcccccc} x : & 14 & 15 & 16 & 16 & 19 \\ y : & 0 & 1 & 16 & 17 & 46 \end{array}$$

Avevamo calcolato  $\bar{x} = 16 = \bar{y}$ .

$$\begin{aligned} s_x^2 &= \frac{1}{4} \left[ (14 - 16)^2 + (15 - 16)^2 + 2 \cdot (16 - 16)^2 + (19 - 16)^2 \right] \\ &= \frac{1}{4} [4 + 1 + 9] = \frac{14}{4} = 3,5 \end{aligned}$$

$$\begin{aligned} s_y^2 &= \frac{1}{4} \left[ (0 - 16)^2 + (1 - 16)^2 + (16 - 16)^2 + (17 - 16)^2 + (46 - 16)^2 \right] \\ &= \frac{1}{4} [256 + 225 + 1 + 900] = \frac{1382}{4} = 345,5 \end{aligned}$$

La varianza 345,5 è un numero grande come ci aspettavamo, inoltre è un numero molto maggiore di 3,5 che è la varianza di  $x$ .

## UNA PROPRIETA' DELLA VARIANZA CAMPIONARIA

Vale la seguente identità:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

La dimostrazione è una semplice verifica, infatti:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

dove si è usata la definizione di media:

$$\sum_{i=1}^n x_i = n\bar{x}$$

## SCARTO QUADRATICO MEDIO

Si definisce scarto quadratico medio o **deviazione standard** la **radice quadrata della varianza**

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)}$$

Nota che lo scarto quadratico medio è misurato nelle stesse unità delle variabili  $x_i$  e della loro media  $\bar{x}$ .

## Esercizio

Calcolare la varianza campionaria  $s^2$  del seguente campione di taglia  $n = 5$ :

1   2   5   6   6

La **media campionaria** è :  $\bar{x} = \frac{1}{5}(1 + 2 + 5 + 2 \cdot 6) = \frac{20}{5} = 4$

Usando la formula  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  otteniamo:

$$s^2 = \frac{1}{4} [(-3)^2 + (-2)^2 + 1 + 2 \cdot 2^2] = \frac{1}{4} (9 + 4 + 1 + 2 \cdot 4) = \frac{22}{4}$$

Usando la formula  $s^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$

$$s^2 = \frac{1}{4} [1 + 2^2 + 5^2 + 2 \cdot 6^2 - 5 \cdot 4^2] = \frac{1}{4} (1 + 4 + 25 + 2 \cdot 36 - 80) = \frac{22}{4}$$

## CALCOLO DELLA VARIANZA CON LE FREQUENZE

**Esempio.** Abbiamo i seguenti dati disposti in una tabella delle frequenze assolute

Valore	Freq. assoluta
4	1
6	4
7	2
<b>TOTALE</b>	<b>7</b>

Quindi l'insieme dei dati originali è composto da 7 valori che disposti in modo crescente sono:

4 6 6 6 6 7 7.



Abbiamo visto che la media campionaria è  $\bar{x} = 6$ . La varianza campionaria è

$$s^2 = \frac{1 \cdot (4 - 6)^2 + 4 \cdot (6 - 6)^2 + 2 \cdot (7 - 6)^2}{7 - 1} = 6$$

In generale sia  $x$  una variabile numerica di taglia  $n$  con  $k$  valori distinti ordinati  $x_1 < x_2 < \dots < x_k$  con frequenze assolute  $n_1, n_2, n_3, \dots, n_k$  ( $n_1 + n_2 + \dots + n_k = n$ ). La varianza campionaria è data da

$$s^2 = \frac{n_1 \cdot (x_1 - \bar{x})^2 + n_2 \cdot (x_2 - \bar{x})^2 + \dots + n_k \cdot (x_k - \bar{x})^2}{n - 1}$$