Università di Cassino

Esercitazioni di Statistica 1 – *4 Febbraio 2011*

Dott. Mirko Bevilacqua

ESERCIZIO N° 1

Osserviamo nella seguente tabella l'altezza in cm di 12 studenti del corso di Geografia:

Studente	1	2	3	4	5	6	7	8	9	10	11	12
Altezza (cm)	158	159	165	166	170	172	173	175	185	186	190	193

Determinare media aritmetica, mediana, primo quartile, terzo quartile, campo di variazione, differenza interquartile, varianza, scarto quadratico medio, coefficiente di variazione e scarto quadratico medio relativo.

INDICI DI POSIZIONE

a) Media aritmetica

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{158+159+......+190+193}{12} = 174{,}3$$

158 159 165 | 166 170 172 | 173 175 185 | 186 190 193

b) Mediana

Essendo n pari (12) la mediana è ottenuta come:

$$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_6 + x_7}{2} = \frac{172+173}{2} = 172{,}5$$

c) Primo quartile (mediana della distribuzione formata dai primi 6 valori della successione)

$$Q_1 = \frac{x_3 + x_4}{2} = \frac{165+166}{2} = 165{,}5$$

d) Terzo quartile (mediana della distribuzione formata dagli ultimi 6 valori della successione)

$$Q_3 = \frac{x_9 + x_{10}}{2} = \frac{185+186}{2} = 185{,}5$$

INDICI DI VARIABILITÀ

La variabilità di una distribuzione esprime la tendenza delle unità di un collettivo ad assumere diverse modalità del carattere.
Per misurarla si ricorre agli indici di variabilità. Un indice di variabilità:
1) assume il suo valore minimo se e solo se tutte le unità delle distribuzione presentano uguale modalità del carattere;
2) aumenta all'aumentare della diversità tra le modalità assunte dalle varie unità.

- **INDICI DI VARIABILITÀ <u>ASSOLUTI</u>:**

e) campo di variazione

Differenza tra il più grande e il più piccolo di tali valori

$$\textbf{Range} = X_{max} - X_{min} = 193 - 158 = 35$$

f) differenza interquartile (IQR)

Differenza tra il primo e il terzo quartile della distribuzione

$$IQR = Q_3 - Q_1 = 185{,}5 - 165{,}5 = 20$$

g) varianza $\left(\sigma_x^2\right)$

Media dei quadrati degli scarti dalla media aritmetica

$$\sigma_x^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

$$\sigma^2 = [(158-174{,}3)^2 + (159-174{,}3)^2 + \ldots\ldots\ldots\ldots + (190-174{,}3)^2 + (193-174{,}3)^2]/12 = 127{,}4$$

P.S.: quando i dati sono organizzati in frequenza, la formula analitica per il calcolo della varianza è

$$\sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2 n_i}{n} = \sum_{i=1}^{k} (x_i - \mu)^2 f_i$$

h) scarto quadratico medio (σ_x)

Lo scarto quadratico medio è dato dalla radice quadrata della varianza:

$$\sigma_x = \sqrt{\sigma_x^2}$$

(L'indice di variabilità risulta ora espresso nella stessa unità di misura della variabile considerata).

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{127{,}4} = 11{,}3$$

- **INDICE DI VARIABILITÀ <u>RELATIVO</u>:**

e) coefficiente di variazione (C V)

Rapporto tra lo scarto quadratico medio e la media:

$$CV = \frac{\sigma}{|\mu|}$$

Il valore minimo di questo indice è zero, il max è $\sqrt{n-1}$

$$0 \leq CV \leq \sqrt{n-1}$$

È indice indipendente dall'unità di misura e dall'ordine di grandezza della variabile e può essere utilizzato per confrontare fenomeni espressi in unità di misura diverse o rilevati in momenti diversi.

$$CV = \frac{\sigma}{|\mu|} = \frac{10,8}{69,9} = 0,154$$

$$0 \leq CV \leq \sqrt{10-1} = 3$$

- **INDICE DI VARIABILITÀ <u>NORMALIZZATO</u>:**

f) scarto quadratico medio relativo (σ_{REL})

Rapporto tra il valore assunto dallo scarto quadratico medio ed il valore massimo che esso può assumere per la distribuzione. Gli incidici normalizzati variano da zero a uno

Il valore massimo che la varianza può assumere in un insieme di n unità statistiche con media pari a μ è:

$$\sigma^2_{MAX} = \mu^2 \cdot (n-1)$$

Di conseguenza, il valore massimo che può assumere lo scarto quadratico medio (in un insieme di n unità statistiche con media paria μ) è:

$$\sigma_{MAX} = \mu \cdot \sqrt{n-1}$$

Il valore minimo di questo indice è zero, il max è uno

$$0 \leq \sigma_{REL} \leq 1$$

$$\sigma_{REL} = \frac{\sigma_x}{MAX(\sigma_x)} = \frac{\sigma_x}{\mu_x \cdot \sqrt{n-1}} = \frac{11,3}{174,3 \cdot \sqrt{12-1}} = 0,02$$

ESERCIZIO N° 2:

Sono stati rilevati i pesi in kg di 10 maschi (M) e 10 (F) di una particolare specie di pesci.

M	1,2	3,0	5,2	4,0	3,5	4,3	3,3	4,8	3,8	3,2
F	1,3	2,2	1,5	2,3	1,8	1,7	2,1	2,0	1,9	2,1

Si calcolino i coefficienti di variazione per le due distribuzioni.

Quale genere di pesci ha maggiore variabilità assoluta?

Quale genere di pesci ha maggiore variabilità relativa?

- **Media aritmetica**

$$\mu_{PM} = \frac{\sum_{i=1}^{10} X_i \cdot n_i}{n} = \frac{36,3}{10} = 36,3$$

$$\mu_{PF} = \frac{\sum_{i=1}^{10} X_i}{n} = \frac{18,9}{10} = 1,89$$

- **Varianza**

Pesci M.	$(x_i-\mu)$	$(x_j-\mu)^2$
1,2	-2,43	5,90
3	-0,63	0,40
5,2	1,57	2,46
4	0,37	0,14
3,5	-0,13	0,02
4,3	0,67	0,45
3,3	-0,33	0,11
4,8	1,17	1,37
3,8	0,17	0,03
3,2	-0,43	0,18
		11,06

Pesci F.	$(x_j-\mu)$	$(x_j-\mu)^2$
1,3	-0,59	0,35
2,2	0,31	0,10
1,5	-0,39	0,15
2,3	0,41	0,17
1,8	-0,09	0,01
1,7	-0,19	0,04
2,1	0,21	0,04
2	0,11	0,01
1,9	0,01	0,00
2,1	0,21	0,04
		0,91

$$\sigma^2_{PM} = \frac{\sum_{i=1}^{10} (x_i - \mu)^2}{n} = \frac{11,06}{10} = 1,11$$

$$\sigma^2_{PF} = \frac{\sum_{i=1}^{10} (x_i - \mu)^2}{n} = \frac{0,91}{10} = 0,09$$

- **Scarto quadratico medio**

$$\sigma_{PM} = \sqrt{\sigma^2_{PM}} = \sqrt{1,11} = 1,05$$

$$\sigma_{PF} = \sqrt{\sigma^2_{PF}} = \sqrt{0,91} = 0,30$$

- **Coefficiente di variazione**

$$CV_{PM} = \frac{\sigma_{PM}}{|\mu_{PM}|} = \frac{1,05}{3,63} = 0,29$$

$$CV_{PF} = \frac{\sigma_{PF}}{|\mu_{PF}|} = \frac{0,30}{1,89} = 0,16$$

I pesci maschi presentano una variabilità, assoluta e relativa, più grande di quella dei pesci femmine.

ESERCIZIO N° 3

Sia data la seguente distribuzione di frequenze del carattere X:

X	1	2	4	8
n_i	30	40	70	110

Si supponga di trasformare tali valori ponderati secondo la relazione y=100+2x.

Sulla variabile trasformata y:

a) determinare la media aritmetica e la varianza;

b) verificare i risultati ottenuti in a) usando le relazioni che consentono di ricavare la media e la varianza delle trasformazioni lineari.

X_i	n_i	X_i^2
1	30	1
2	40	4
4	70	16
8	110	64
	250	

Y_i	n_i	Y_i^2
102	30	3060
104	40	4160
108	70	7560
116	110	12760
	250	

$$\mu_x = \frac{\sum_{i=1}^{4} X_i \cdot n_i}{n} = \frac{1 \cdot 30 + 2 \cdot 40 + 4 \cdot 70 + 8 \cdot 110}{250} = 5{,}08$$

<u>Trasformazione lineare della media</u>:

La media aritmetica di un carattere Y ottenuto attraverso una trasformazione lineare *Y=a+bX* di un carattere X di media μ_x è uguale a $\mu_y = a + b\mu_x$.

$$\mu_y = 100 + 2\mu_x = 100 + 2 \cdot 5{,}08 = 110{,}16$$

Calcolando la media dalla distribuzione di frequenze del carattere Y si ha:

$$\mu_y = \frac{\sum_{i=1}^{4} X_i \cdot n_i}{n} = \frac{102 \cdot 30 + 104 \cdot 40 + 108 \cdot 70 + 116 \cdot 110}{250} = 110{,}16$$

- *Media di X^2*

$$\mu_{x^2} = \frac{\sum_{i=1}^{4} X_i \cdot n_i}{n} = \frac{1 \cdot 30 + 4 \cdot 40 + 16 \cdot 70 + 64 \cdot 110}{250} = 33{,}4$$

- *Media di Y^2*

$$\mu_{y^2} = \frac{\sum_{i=1}^{4} X_i \cdot n_i}{n} = \frac{3060 \cdot 30 + 4160 \cdot 40 + 7560 \cdot 70 + 12760 \cdot 110}{250} = 8764$$

$$\sigma_x^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} = \frac{\sum_{i=1}^{K}(x_i-\mu)^2 n_i}{n} = \sum_{i=1}^{K}(x_i-\mu)^2 f_i = \mu_{x^2} - (\mu_x)^2$$

$$\sigma_x^2 = 33,40 - (5,08)^2 = 7,59$$

<u>Varianza di una trasformazione lineare:</u>

La varianza di un carattere Y, ottenuto attraverso la trasformazione Y=a+bX di un carattere X di media μ_x e varianza σ^2 è pari a: $\sigma_y^2 = b^2 \cdot \sigma_x^2$

$$\sigma_y^2 = 2^2 \cdot 7,59 = 30,37$$

Calcolando la varianza direttamente dalla distribuzione di frequenze del carattere Y, si ha:

$$\sigma_y^2 = 8764 - (110,16)^2 = 30,37$$