Esercitazione 3 del corso di Statistica (parte 1)

Dott.ssa Paola Costantini

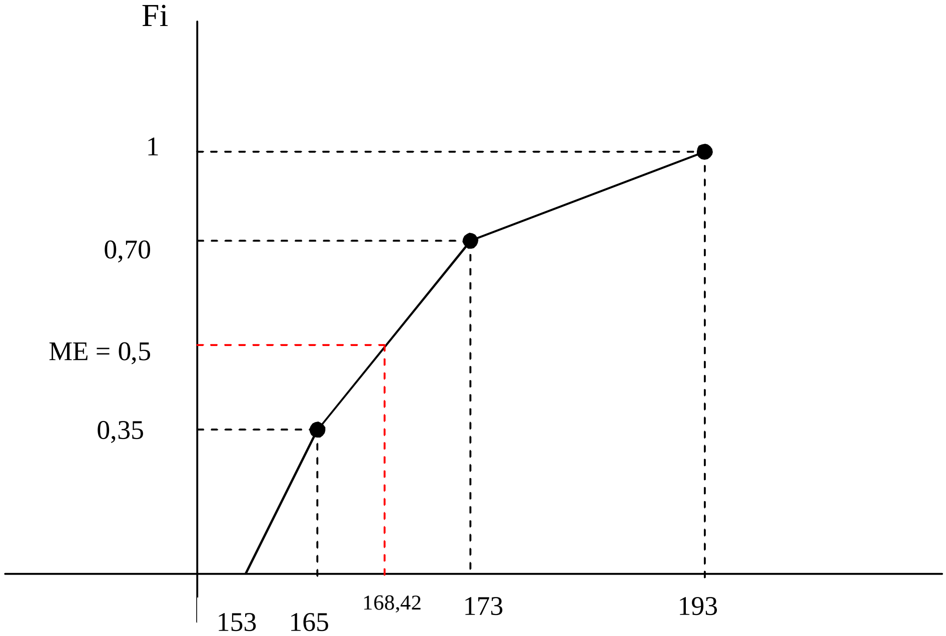
18 Ottobre 2011

Esercizio n. 1

A partire dalla distribuzione in classi della variabile Altezza ripartita in 3 classi equifrequenti, calcolare mediana, primo e terzo quartile.

C_i	n_i	f_i	N_i	F_i	$\hat{C}_i$	a_i	d_i
C_1 = [153; 165]	7	0,35	7	0,35	159	12	0,029
C_2 =] 165; 173]	7	0,35	14	0,70	169	8	0,043
C_3 =] 173; 193]	6	0,30	20	1	183	20	0,015
Totali	**20**	**1,00**					

Funzione di ripartizione



Il calcolo della mediana per la ripartizione di una distribuzione in classi avviene attraverso al formula:

Mediana

$$Me \cong x_{Me-1} + (x_{Me} - x_{Me-1}) \frac{0,5 - F_{Me-1}}{F_{Me} - F_{Me-1}}$$

$$Me = 165 + (173 - 165) \cdot \frac{0{,}5 - 0{,}35}{0{,}70 - 0{,}35} = 168{,}42$$

Il calcolo del primo quartile per la ripartizione di una distribuzione in classi avviene attraverso al formula:

Primo quartile

$$\boxed{Q_1 \cong x_{Q_1 - 1} + \left(x_{Q_1} - x_{Q_1 - 1}\right) \frac{0,25 - F_{Q_1 - 1}}{F_{Q_1} - F_{Q_1 - 1}}}$$

$$Q_1 = 153 + (165 - 153) \cdot \frac{0{,}25 - 0}{0{,}35 - 0} = 161{,}57$$

Il calcolo del terzo quartile per la ripartizione di una distribuzione in classi avviene attraverso al formula:

Terzo quartile

$$\boxed{Q_3 \cong x_{Q_3 - 1} + \left(x_{Q_3} - x_{Q_3 - 1}\right) \frac{0,75 - F_{Q_3 - 1}}{F_{Q_3} - F_{Q_3 - 1}}}$$

$$Q_3 = 173 + (193 - 173) \cdot \frac{0{,}75 - 0{,}70}{1 - 0{,}70} = 176{,}33$$

Esercizio n. 2

Si calcolino gli indici di variabilità della variabile **Componenti**: scostamento semplice medio dalla media, scostamento semplice medio dalla mediana, varianza e scarto quadratico medio della variabile Altezza, a partire dai dati grezzi.

$\mu = 4{,}65$
Me = 4,5

COMPONENTI	**n_i**	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$	$\lvert x_i - \bar{x}\rvert \cdot n_i$	$\lvert x_i - Me\rvert \cdot n_i$
1	1	-3,65	13,32	13,32	3,65	3,5
3	1	-1,65	2,72	2,72	1,65	1,5
4	8	-0,65	0,42	3,38	5,2	4
5	6	0,35	0,12	0,73	2,1	3
6	2	1,35	1,82	3,65	2,7	3
7	1	2,35	5,52	5,52	2,35	2,5
8	1	3,35	11,22	11,22	3,35	3,5
tot				**40,55**	**21**	**21**

Scostamento semplice medio dalla media

$$S_{\bar{x}} = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| \cdot n_i}{\sum_{i=1}^{k} n_i} = \frac{21}{20} = 1{,}05$$

Sappiamo che la media è pari a 4,65.

Scostamento semplice medio dalla mediana

$$S_{Me} = \frac{\sum_{i=1}^{k} |x_i - Me| \cdot n_i}{\sum_{i=1}^{k} n_i} = \frac{21}{20} = 1{,}05$$

Sappiamo che la mediana è pari a 4,5.

Varianza $\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{40{,}55}{20} = 2{,}2275$

Scarto quadratico medio $\boxed{\sigma = \sqrt{\sigma^2}} = 1{,}49$

Coefficiente di variazione

$$CV = \frac{\sigma}{\bar{x}} = \frac{1{,}49}{4{,}65} = 0{,}32$$

E' un indice indipendente dall'unità di misura (è un numero puro) e può essere utilizzato per confrontare distribuzioni diverse

Esercizio n. 3

1) Calcolare l'indice di eterogeneità di Gini per i caratteri "Qualifica Funzionale" e "Regime di Impiego".

Condizione di massima omogeneità:
le *n* unità statistiche presentano tutte la stessa modalità

Condizione di massima eterogeneità:
nella distribuzione di frequenza appaiono tutte le *k* modalità, e ad ognuna di esse è associata la medesima frequenza

Calcoliamo l'indice del Gini per la seguente variabile ordinata *Esito dell'esame finale.*

Esito esame	n_i	f_i
Insufficiente	6	0,21
Sufficiente	10	0,345
Buono	8	0,275
Ottimo	5	0,17
Totale	29	

$$G = 1 - \sum f_i^2 = 1 - (0{,}21^2 + 0{,}345^2 + 0{,}275^2 + 0{,}17^2) = 1 - (0{,}044 + 0{,}119 + 0{,}0756 + 0{,}0289) = 0{,}73$$

In caso di minima eterogeneità G=0

In caso di massima eterogeneità l'indice assume valore $G = 1 - \frac{1}{k}$

$$G\max = 1 - \frac{1}{k} = 1 - \frac{1}{4} = 0{,}75$$

Avendo definito il valore massimo dell'indice è possibile ottenerne la sua versione normalizzata

$$G^* = \frac{G}{G\max} = \frac{0{,}73}{0{,}75} = 0{,}9733$$

Conclusione

G^* prossimo ad 1, la distribuzione è molto eterogenea: tutte le modalità sono presenti e con frequenze molto simili tra loro.

Esercizio n. 4

Calcolare l'indice di dispersione D per la seguente variabile ordinata *esito dell'esame finale.*

Esito esame	n_i	$N_i (cumulate\ n_i)$	$F_i (cumulate\ f_i)$	$Fr_i (retro\ cumulate\ f_i)$
Insufficiente	6	6	0,21	1
Sufficiente	10	16	0,55	0,79
Buono	8	24	0,82	0,45
Ottimo	5	29	1	0,18
Totale	29			

$$D = \sum_{j=1}^{k} \left[F_j (1 - F_j) + RF_j (1 - RF_j) \right] =$$

$$= [0,21(1-0,21)+1(1-1)] + [0,55(1-0,55)+0,79(1-0,79)] + [0,82(1-0,82)+0,45(1-0,45)] +$$
$$+ [1(1-1)+0,18(1-0,18)] =$$
$$= 0,166+0+0,2475+0,166+0,1476+0,2475+0+0,1476 = 1,12$$

Essendo una somma di termini tutti non negativi, l'indice D è non negativo.
La distribuzione della variabile esito dell'esame finale è una distribuzione caratterizzata da un'alta variabilità (massima eterogeneità o minima omogeneità).

Esercizio5

Si ricalcolino l'indice di eterogeneità di Gini e l'indice di dispersione D cambiando l'ordine delle modalità della distribuzione.

Indice di eterogeneità di Gini

Esito esame	n_i	f_i	$N_i\,(cumulate\ n_i)$	$F_i\,(cumulate\ f_i)$	$Fr_i\,(retro\ cumulate\ f_i)$
Insufficiente	10	0,34	10	0,34	1
Sufficiente	5	0,17	15	0,51	0,66
Buono	8	0,28	23	0,79	0,49
Ottimo	6	0,21	29	1	0,21
Totale	29	1			

$$G = 1 - \sum f_i^2 = 1 - (0,34^2 + 0,17^2 + 0,28^2 + 0,21^2) = 1 - (0,115 + 0,029 + 0,078 + 0,044) = 0,73$$

$$G\max = 1 - \frac{1}{k} = 1 - \frac{1}{4} = 0,75$$

$$G^* = \frac{G}{G\max} = \frac{0,73}{0,75} = 0,9733$$

Il risultato è identico rispetto all'esercizio 3

Indice di dispersione D

$$= [0,34(1-0,34)+1(1-1)] + [0,51(1-0,51)+0,66(1-0,66)] + [0,79(1-0,79)+0,49(1-0,49)] +$$
$$+ [1(1-1)+0,21(1-0,11)] =$$
$$= 0,2244+0+0,2499+0,2244+0,1659+0,2499+0+0,1659 = 1,28$$

In questo caso il risultato cambia, in quanto l'indice di dispersione D tiene conto dell'ordine delle modalità.