Esercitazione 3 del corso di Statistica (parte 1)

Dott.ssa Paola Costantini

29 Gennaio 2009

La seguente tabella riporta le informazioni relative a 25 laureati nell'anno 2005 in Economia, ad un anno dal conseguimento del titolo.

Genere	Eta	Voto*	Durata	Tipo Contratto	Utilizzo della Laurea	Efficacia della Laurea
Uomini	47,8	83	4,4	Stabile	In misura elevata	Efficace
Uomini	26,6	113	7,4	Stabile	In misura ridotta	Poco efficace
Uomini	31,5	91	12,4	Atipico	In misura ridotta	Abb. efficace
Uomini	23,6	102	4,4	Inserimento/Form	In misura elevata	Efficace
Uomini	25,9	94	6,4	Stabile	Per niente	Per nulla efficace
Donne	23,6	108	4,7	Inserimento/Form	In misura ridotta	Abb. efficace
Donne	28,6	108	5,7	Atipico	In misura ridotta	Abb. efficace
Uomini	42,1	100	3	Stabile	Per niente	Efficace
Donne	24,3	113	3,4	Atipico	Per niente	Per nulla efficace
Donne	26,3	113	3,4	Atipico	Per niente	Per nulla efficace
Uomini	24,9	106	3,7	Inserimento/Form	In misura elevata	Molto efficace
Donne	24	95	4,5	Stabile	In misura elevata	Efficace
Uomini	34,7	92	4,4	Stabile	In misura elevata	Efficace
Donne	24,7	106	5,14	Atipico	In misura ridotta	Abb. efficace
Uomini	25,9	100	5,94	Atipico	In misura ridotta	Abb. efficace
Uomini	25,4	92	6,14	Atipico	In misura elevata	Molto efficace
Donne	27,6	113	4,14	Senza contratto	Per niente	Poco efficace
Donne	23,4	113	4,4	Inserimento/Form	In misura elevata	Molto efficace
Uomini	31,3	105	3,4	Stabile	In misura elevata	Molto efficace
Uomini	29,7	110	4,14	Atipico	In misura ridotta	Abb. efficace
Uomini	27	93	7,3	Stabile	In misura elevata	Efficace
Uomini	35,6	97	15,4	Atipico	Per niente	Poco efficace
Uomini	23,1	101	3,7	Senza contratto	Per niente	Poco efficace
Uomini	25,3	91	6,5	Atipico	Per niente	Per nulla efficace
Uomini	32,6	92	13,5	Stabile	In misura ridotta	Abb. efficace

* Il 113 indica il 110 con Lode.

Esercizio 1 A partire dalla distribuzione di frequenza del carattere Voto di Laurea, calcolare gli indici di posizione: media, moda, mediana e quartili.

Distribuzione di frequenza del carattere quantitativo *Voto di Laurea*

Durata del corso di studi
4,4
7,4
12,4
4,4
6,4
4,7
5,7
3
3,4
3,4
3,7
4,5
4,4

Dati ordinati	**n_i**	**fi**	**Ni**	**Fi**
3	1	0,077	1	0,077
3,4	2	0,154	3	0,231
3,7	1	0,077	4	0,308
4,4	3	0,23	7	0,538
4,5	1	0,077	8	0,615
4,7	1	0,077	9	0,692
5,7	1	0,077	10	0,769
6,4	1	0,077	11	0,846
7,4	1	0,077	12	0,923
12,4	1	0,077	13	1
totale	13	1		

Media aritmetica

$$\mu = \frac{\sum_{i=1}^{n} x_i \times n_i}{n} = \frac{3\times 1\times 3{,}4\times 2 + 4{,}4\times 3 + ... + 12{,}4\times 1}{13} = 5{,}2$$

Moda

Mo_1 = 4,4

Mediana

Consideriamo la successione ordinata di valori del carattere:

3 3,4 3,4 3,7 4,4 4,4 4,4 4,5 4,7 5,7 6,4 7,4 12,4

Essendo n dispari la mediana è ottenuta come:

$$\mathbf{Me} = X_{\frac{N+1}{2}} = X_{\frac{13+1}{2}} = 7 \text{ (posizione)}$$

Con i dati ordinati, alla 7° posizione c'è 4,4

Quartili

Nel nostro caso, essendo il numero delle osservazioni dispari, nel dividere in due la successione ordinata, al fine di ottenere due successioni di pari numerosità, l'osservazione mediana viene esclusa dalla procedura per il calcolo dei quartili.

Essendo 13 dati, ed avendo escluso il valore mediano, rimangono due distribuzioni di 6 osservazioni ciascuna. In tal caso, n sarà pari (6):

Calcolo del primo quartile:

3 3,4 3,4 3,7 4,4 4,4

$$Q_1 = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} = \frac{3+4}{2} = 3{,}5 \text{ (posizione)}$$

per cui il primo quartile sarà pari a $Q_1 = \frac{3{,}4+3{,}7}{2} = 3{,}55$

In ogni caso Q_1 si individua scorrendo la successione ordinata della X e corrisponde al primo valore x* tale che F(x*) ≥ 0,25

Calcolo del terzo quartile

4,5 4,7 5,7 6,4 7,4 12,4

$$Q_3 = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} = \frac{3+4}{2} = 3{,}5 \text{ (posizione)}$$

per cui il terzo quartile sarà pari a $Q_3 = \frac{5{,}7+6{,}4}{2} = 6{,}05$

In ogni caso Q_3 si individua scorrendo la successione ordinata della X e corrisponde al primo valore x* tale che F(x*) ≥ 0,75

Esercizio n. 2

Calcolare media, mediana, primo e terzo quartile della variabile Durata del corso di studio (prime 13 osservazioni), dopo averla ripartita in 3 classi equifrequenti.

N/3= 4,3 (non avrebbe senso)

Stabiliamo di avere 2 classi con frequenza pari a 4 e una con frequenza pari a 5.

Per una distribuzione in classi di frequenza, la media si calcola:

x_i	n_i	f_i	N_i	F_i	$\hat{C}_i$	a_i	di
C_1 = [3; 3,7]	4	0,31	4	0,31	3,35	0,7	0,44
C_2 =] 3,7; 4,5]	4	0,31	8	0,62	4,1	0,8	0,38
C_3 =] 4,5; 12,4]	5	0,38	13	1	8,45	7,9	0,048
Totali	**13**	**1,00**					

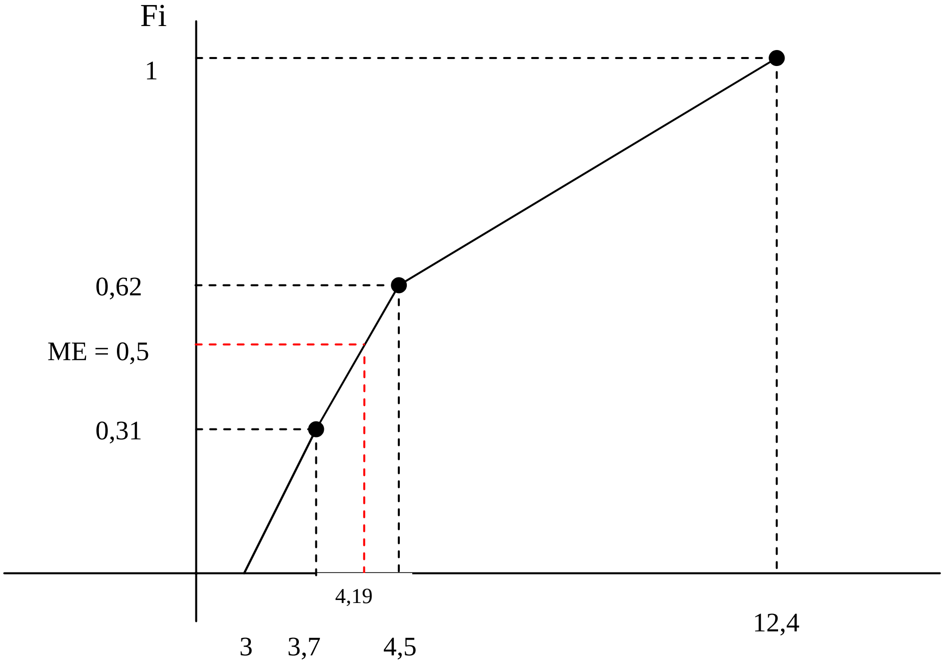
Media $\boxed{\mu_x = \frac{\sum_{i=1}^{k} \hat{c}_i \cdot n_i}{N}}$

Dove $\boxed{\hat{c}_i = \frac{x_{i-1} + x_i}{2}}$

$$\mu_x = \frac{\sum_{i=1}^{k} (3{,}35 \times 4 + 4{,}1 \times 4 + 8{,}45 \times 5)}{13} = 5{,}54$$

Classe Modale = C_3 =]4,5;12,4] = la classe modale è quella cui corrisponde la frequenza più elevata.

Funzione di ripartizione



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

Mediana

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

$$Me = 3{,}7 + (4{,}5 - 3{,}7) \cdot \frac{0{,}5 - 0{,}31}{0{,}62 - 0{,}31} = 4{,}19$$

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

Primo quartile

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

$$Q_1 = 3 + (3{,}7 - 3) \cdot \frac{0{,}25 - 0}{0{,}31 - 0} = 3{,}56$$

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

Terzo quartile

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

$$Q_3 = 4,5 + (12,4 - 4,5) \cdot \frac{0,75 - 0,62}{1 - 0,62} = 7,2$$

Si calcolino gli indici di variabilità: scostamento semplice medio dalla media, scostamento semplice medio dalla mediana, varianza e scarto quadratico medio della variabile Durata del corso di studi, a partire dai dati grezzi.

Durata del corso di studi	n_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$	$\lvert x_i - \bar{x}\rvert \cdot n_i$	$\lvert x_i - Me\rvert \cdot n_i$
3	1	-2,2	4,84	4,84	2,2	1,4
3,4	2	-1,8	3,24	6,48	3,6	2
3,7	1	-1,5	2,25	2,25	1,5	0,7
4,4	3	-0,8	0,64	1,92	2,4	0
4,5	1	-0,7	0,49	0,49	0,7	0,1
4,7	1	-0,5	0,25	0,25	0,5	0,3
5,7	1	0,5	0,25	0,25	0,5	1,3
6,4	1	1,2	1,44	1,44	1,2	2
7,4	1	2,2	4,84	4,84	2,2	4
12,4	1	7,2	51,84	51,84	9,4	8
totale	**13**			**74,6**	**24,2**	**19,8**

Scostamento semplice medio dalla media

$$S_{\bar{x}} = \frac{\sum_{i=1}^{k} \lvert x_i - \bar{x}\rvert \cdot n_i}{\sum_{i=1}^{k} n_i} = \frac{24,2}{13} = 1,86$$

Sappiamo che la media è pari a 5,2.

Scostamento semplice medio dalla mediana

$$S_{Me} = \frac{\sum_{i=1}^{k} \lvert x_i - Me\rvert \cdot n_i}{\sum_{i=1}^{k} n_i} = \frac{19,8}{13} = 1,5$$

Sappiamo che la mediana è pari a 4,4

Varianza $\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{74,6}{13} = 5,74$

Scarto quadratico medio $\boxed{\sigma = \sqrt{\sigma^2}} = 2,4$

CALCOLARE INOLTRE I SEGUENTI INDICI DI VARIABILITÀ RELATIVA:

Scarto quadratico medio relativo

Si ottiene come rapporto tra il valore assunto dallo scarto ed il valore massimo che esso può assumere per la distribuzione:

$$\sigma_x^2 = \frac{\sigma}{\bar{x}\sqrt{n-1}} = \frac{2,4}{5,2\sqrt{13-1}} = 0,1333$$

Coefficiente di variazione

$$CV = \frac{\sigma}{\bar{x}} = \frac{2,4}{5,52} = 0,46$$

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