Esercitazione 3 del corso di Statistica (parte 1)

Dott. ssa Paola Costantini

4 Febbraio 2011

Dataset Studenti

N	*SESSO*	*ETA'*	*PESO*	*ALTEZZA*	*DIPLOMAI*	*COMPONENTI*	*OCCHIALI*	*FUMO*
1	0	20,6	65	180	Ist.Tecnico	6	0	1
2	0	20,2	75	180	Liceo	4	0	0
3	0	20,3	60	173	Ist.Tecnico	4	1	0
4	0	23,9	93	187	Liceo	8	0	1
5	0	21,4	66	164	Ist.Tecnico	5	0	0
6	0	25	84	186	Ist.Tecnico	4	0	0
7	0	20,8	67	175	Altro dipl.	4	0	1
8	0	20,6	89	170	Liceo	3	1	0
9	0	27,1	71	180	Liceo	1	0	1
10	0	23,3	63	170	Liceo	4	0	0
11	1	20,5	51	161	Ist.Tecnico	4	0	1
12	1	19,1	58	167	Ist.Tecnico	5	1	1
13	1	22,1	67	165	Altro dipl.	5	1	1
14	1	21,8	51	156	Ist.Tecnico	4	0	0
15	1	19,2	60	170	Ist.Tecnico	5	1	1
16	1	20,8	55	165	Liceo	4	1	1
17	1	21	55	158	Liceo	5	1	0
18	1	20,9	58	170	Liceo	5	1	1
19	1	22,7	76	170	Liceo	6	1	0
20	1	21	55	165	Liceo	7	0	0

Esercizio 1

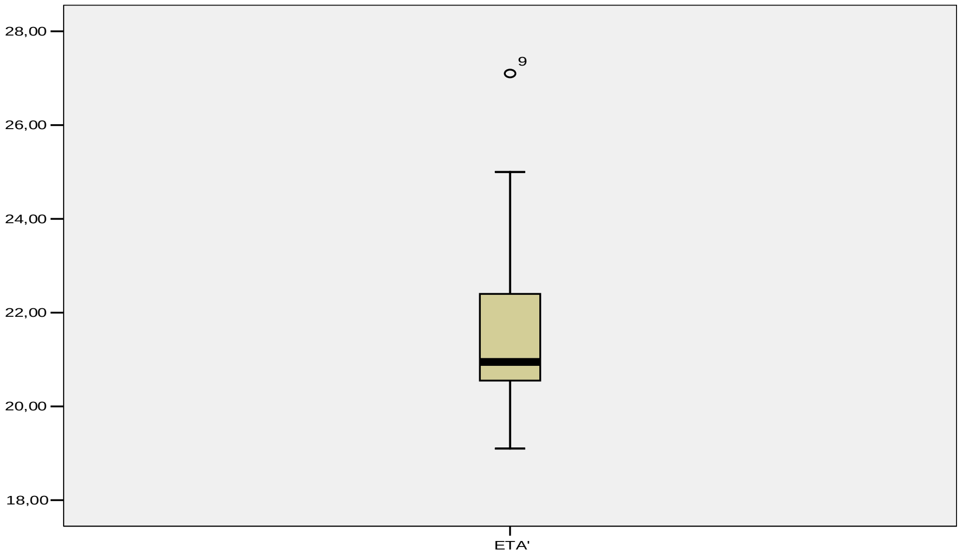
Costruire il boxplot della variabile ETA'

ETA'	n_i	f_i	N_i	F_i
19,1	1	0,05	1	0,05
19,2	1	0,05	2	0,1
20,2	1	0,05	3	0,15
20,3	1	0,05	4	0,2
20,5	1	0,05	5	0,25
20,6	2	0,1	7	0,35
20,8	2	0,1	9	0,45
20,9	1	0,05	10	0,5
21	2	0,1	12	0,6
21,4	1	0,05	13	0,65
21,8	1	0,05	14	0,7
22,1	1	0,05	15	0,75
22,7	1	0,05	16	0,8
23,3	1	0,05	17	0,85
23,9	1	0,05	18	0,9
25	1	0,05	19	0,95
27,1	1	0,05	20	1
totale	20	**1**		

Q_1 = 20,5 **Me** = 20,9 **Q_3** = 22,1

a = Q_1 – 1,5 (Q_3 –Q_1) = 20,5 -1,5 (22,1-20,5) = 18
b = Q_3 + 1,5 (Q_3 –Q_1) = 22,1 +1,5 (22,1-20,5) = 24,5

α = min = 19,1
β = max = 27,1



Esercizio 2
Calcolare media, mediana, primo e terzo quartile della variabile ETA' dopo averla ripartita in 5 classi equifrequenti.

Dati ordinati

ETA'
19,1
19,2
20,2
20,3
20,5
20,6
20,6
20,8
20,8
20,9
21
21
21,4
21,8
22,1
22,7
23,3
23,9
25
27,1

N/5= 20/5=4

C_i	n_i	f_i	N_i	F_i	c_i
[19,1; 20,3]	4	0.2	4	0.2	19,7
]20,3; 20,8]	5	0.25	9	0.45	20,55
] 20,8; 21]	3	0.15	12	0.6	20,9
[21; 22,7]	4	0.2	16	0.8	21,85
]22,7; 27,1]	4	0.2	20	1	24,9
	20	**1**			

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

Media

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

dove

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

$$\bar{x} = \frac{\sum_{i=1}^{k} (19{,}7 \times 4 + 20{,}55 \times 5 + 20{,}9 \times 3 + 21{,}85 \times 4 + 24{,}9 \times 4)}{20} = 21{,}56$$

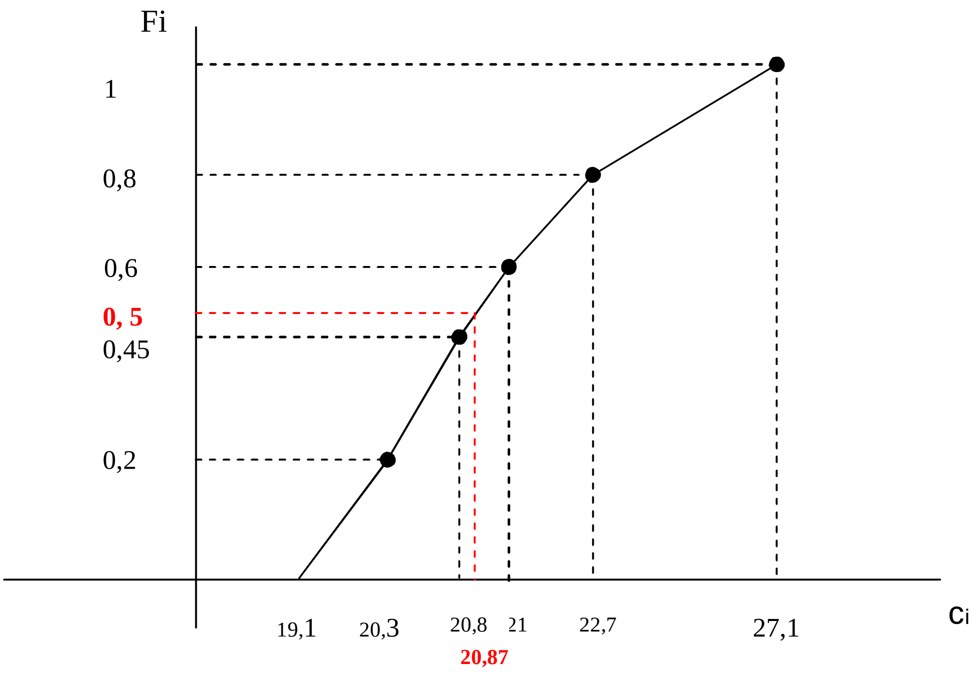
Il calcolo della mediana per la ripartizione di una distribuzione in classi avviene attraverso al formula:

Mediana

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

$$Me = 20{,}8 + (21 - 20{,}8) \cdot \frac{0{,}5 - 0{,}45}{0{,}6 - 0{,}45} = 20{,}87$$

Funzione di ripartizione



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

Primo quartile

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

$$Q_1 = 20,3 + (20,8 - 20,3) \cdot \frac{0,25 - 0,2}{0,45 - 0,2} = 20,4$$

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

Terzo quartile

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

$$Q_3 = 21 + (22,7 - 21) \cdot \frac{0,75 - 0,6}{0,8 - 0,6} = 22,27$$

Esercizio 3

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.

Scostamento semplice medio dalla media **(si vedano i calcoli nella seguente tabella)**

$$S_{\bar{x}} = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| \cdot n_i}{\sum_{i=1}^{k} n_i} = \frac{27,58}{20} = 1,379$$

Sappiamo che la media è pari a 21,32.

Scostamento semplice medio dalla mediana **(si vedano i calcoli nella seguente tabella)**

$$S_{Me} = \frac{\sum_{i=1}^{k} |x_i - Me| \cdot n_i}{\sum_{i=1}^{k} n_i} = \frac{14,9}{20} = 0,745$$

Sappiamo che la mediana è pari a 20,87.

ETA'	ni	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$	$\lvert x_i - \bar{x} \rvert \cdot n_i$	$\lvert x_i - Me \rvert \cdot n_i$
19,1	1	-2,51	6,30	6,3001	2,51	1,85
19,2	1	-2,41	5,81	5,8081	2,41	1,75
20,2	1	-1,41	1,99	1,9881	1,41	0,75
20,3	1	-1,31	1,72	1,7161	1,31	0,65
20,5	1	-1,11	1,23	1,2321	1,11	0,45
20,6	2	-1,01	1,02	2,0402	1,01	0,35
20,8	2	-0,81	0,66	1,3122	0,81	0,15
20,9	1	-0,71	0,50	0,5041	0,71	0,05
21	2	-0,61	0,37	0,7442	0,61	0,05
21,4	1	-0,21	0,04	0,0441	0,21	0,45
21,8	1	0,19	0,04	0,0361	0,19	0,85
22,1	1	0,49	0,24	0,2401	0,49	1,15
22,7	1	1,09	1,19	1,1881	1,09	1,75
23,3	1	1,69	2,86	2,8561	1,69	2,35
23,9	1	2,29	5,24	5,2441	2,29	2,95
25	1	3,39	11,49	11,4921	3,39	4,05
27,1	1	5,49	30,14	30,1401	5,49	6,15
				72,886	**25,72**	**25,75**

Varianza $\sigma_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{72,886}{20} = 3,64$

Scarto quadratico medio $\boxed{\sigma = \sqrt{\sigma^2}} = 1,9$

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:

$$\sigma_{x}^{2} = \frac{\sigma}{\bar{x}\sqrt{n-1}} = \frac{1{,}9}{21{,}61\sqrt{20-1}} = 0{,}02$$

Coefficiente di variazione

$$CV = \frac{\sigma}{\bar{x}} = \frac{1{,}9}{21{,}61} = 0{,}0879$$

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