

Università di Cassino

Esercitazioni di Statistica 1 del 29 Gennaio 2010

Dott. Mirko Bevilacqua

DATASET STUDENTI

N	SESSO	ALTEZZA (cm)	PESO (kg)	CORSO LAUREA	NUMERO SCARPA	COLORE OCCHI	COLORE CAPELLI
1	M	179	65	INFORMAICA	43	scuri	Scuri
2	M	180	62	INFORMATICA	43	scuri	Scuri
3	F	165	50	INFORMATICA	39	scuri	Castani
4	F	160	49	INFORMATICA	37	verdi	Castani
5	F	160	47	MATEMATICA	37	azzurri	Biondi
6	F	160	48	MATEMATICA	36	scuri	Biondi
7	F	164	56	MATEMATICA	38	verdi	Castani
8	F	170	59	MATEMATICA	38	scuri	Castani
9	M	180	73	MATEMATICA	43	verdi	Castani
10	M	186	86	MATEMATICA	45	azzurri	Scuri
11	M	170	66	MATEMATICA	42	scuri	Castani
12	M	180	68	INFORMATICA	41	scuri	Scuri
13	M	180	85	INFORMATICA	43	scuri	Castani
14	F	176	56	INFORMAATICA	37	scuri	Biondi
15	M	170	72	INFORMATICA	42	verdi	Castani
16	M	180	65	INFORMATICA	42	scuri	Scuri
17	M	170	75	INFORMATICA	41	scuri	Scuri
18	M	172	70	INFORMATICA	40	scuri	Castani
19	M	178	80	INFORMATICA	45	scuri	Castani
20	F	162	49	BIOLOGIA	37	scuri	Scuri

ESERCIZIO n°1:

Costruire le distribuzioni doppie di frequenza per le seguenti coppie di caratteri:

- A) SESSO – CORSO DI LAUREA
 B) SESSO – NUMERO DI SCARPE

A) Soluzione

SESSO \ CORSO DI LAUREA	Matematica	Informatica	Biologia	Totale
	MASCHIO	3	9	0
FEMMINA	4	3	1	8
Totale	7	12	1	20

B) soluzione

N° scarpe \ Sesso	36	37	38	39	40	41	42	43	45	Totale
Maschio	0	0	0	0	1	2	3	4	2	12
Femmina	1	4	2	1	0	0	0	0	0	8
Totale	1	4	2	1	1	2	3	4	2	20

ESERCIZIO n°2:

Calcolare media aritmetica, moda e mediana e quartili per il carattere ALTEZZA a partire:

2.1) dalla distribuzione semplice del carattere;

2.2) dalla distribuzione nelle classi:

[160 – 165];]165 – 170];]170 – 175];]175 – 179];]179 – 186];

2.1) soluzione

La successione dei valori del carattere ALTEZZA è la seguente:

179 180 165 160 160 160 164 170 180 186 170 180 180 176 170 180 170 172 178 162

Media aritmetica

$$\mu = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^{20} x_i}{20} = \frac{179 + 180 + 165 + \dots + 178 + 162}{20} = 172,1$$

La **Moda** è la modalità della distribuzione che si presenta con la massima frequenza (assoluta, relativa o percentuale).

Osservando la distribuzione di frequenze:

x_i	160	162	164	165	170	172	176	178	179	180	186
n_i	3	1	1	1	4	1	1	1	1	5	1

Mo=180.

La **Mediana** (Me) di un insieme di unità ordinate (secondo un carattere ordinabile) è la modalità presentata dall'unità centrale dove per unità centrale si intende quell'unità che divide il collettivo in due parti di uguale numerosità: una parte formata dalle unità che presentano una modalità precedente o uguale a quella dell'unità centrale e una parte formata dalle unità che presentano una modalità successiva o uguale a quella dell'unità centrale.

La successione ordinata dei valori è:

160 160 160 162 164 165 170 170 170 170 | 172 176 178 179 180 180 180 180 180 186

n = 20 è pari

quindi:

$$Q_2 = Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_{10} + x_{11}}{2} = \frac{170 + 172}{2} = 171$$

Quartili

I quartili della distribuzione semplice vanno individuati determinando prima la mediana (secondo quartile) e poi le mediane delle due metà della distribuzione che si trovano a sinistra (primo quartile) e a destra (terzo quartile) della mediana.

Per il calcolo di Q_1 e Q_3 si fa riferimento alle due semidistribuzioni di numerosità:

$$n_1 = n_2 = 10$$

160 160 160 162 164	165 170 170 170 170	172 176 178 179 180	180 180 180 180 186
---------------------	---------------------	---------------------	---------------------

$$Q_1 = \frac{x_{\frac{n_1}{2}} + x_{\frac{n_1}{2}+1}}{2} = \frac{x_5 + x_6}{2} = \frac{164 + 165}{2} = 164,5$$

Considerando la semidistribuzione di destra, ossia ricominciando a contare le posizioni dalla 11^{ma}:

$$Q_3 = \frac{x_{\frac{n_2}{2}} + x_{\frac{n_2}{2}+1}}{2} = \frac{x_5 + x_6}{2} = \frac{180 + 180}{2} = 180$$

2.2) soluzione

La distribuzione di frequenze è la seguente:

$x_{i-1} - x_i$	c_i	n_i	f_i	F_i	a_i	d_i
[160 - 165]	162,5	6	0,3	0,3	5	0,06
]165 - 170]	167,7	4	0,2	0,5	5	0,04
]170 - 175]	172,5	1	0,05	0,55	5	0,01
]175 - 179]	177	3	0,15	0,7	4	0,0375
]179 - 186]	182,5	6	0,3	1	7	0,042857
Totali		20	1			

Data la distribuzione di frequenza del carattere altezza suddivisa in 5 classi si può approssimare la **Media aritmetica** del carattere con l'espressione riportata sotto, dove k è il numero di classi della distribuzione di frequenza, C_i è il valore centrale della classe i -esima (il valore che si ottiene come semisomma degli estremi della classe) e n_i è la corrispondente frequenza assoluta:

$$\mu = \frac{\sum_{i=1}^k C_i \cdot n_i}{n} = \frac{\sum_{i=1}^5 C_i \cdot n_i}{20} =$$

$$= \frac{(162,5 \times 6) + (167,5 \times 4) + (172,5 \times 1) + (177 \times 3) + (182,5 \times 6)}{20} = 172,215$$

La **Classe modale** è definita come la classe alla quale corrisponde la frequenza più alta. Se la distribuzione possiede classi di diversa ampiezza occorre dividere le frequenze delle classi per la loro ampiezza e confrontare tali quozienti: quello più grande individuerà la classe modale.

1° classe: $6/5=1,2$; 2° classe: $4/5=0,8$; 3° classe: $1/5=0,2$; 4° classe: $3/4=0,75$

5° classe: $6/7=0,85$.

classe modale = **[160 - 165]**

Mediana

La classe mediana è **[165-170]**: prima classe avente frequenza relativa cumulata maggiore o uguale a 0,5. Individuata la classe mediana, si procede con la stima della mediana all'interno della classe, utilizzando la seguente formula:

$$Me = x_{Me-1} + \frac{(0,5 - F_{Me-1})}{F_{Me} - F_{Me-1}} \cdot (x_{Me} - x_{Me-1}) \quad \text{dove}$$

x_{Me-1} e x_{Me} individuano l'estremo inferiore e l'estremo superiore della classe dove cade la mediana

F_{Me-1} e F_{Me} indicano la frequenza relativa cumulata fino alla classe precedente a quella in cui cade la mediana e la frequenza relativa cumulata della classe mediana.

$$Q_2 = Me = 165 + \frac{(0,5 - 0,3)}{(0,5 - 0,3)} \cdot 5 = 170$$

Primo quartile

Il procedimento è analogo al precedente. Naturalmente la classe di riferimento è la prima che presenta frequenza relativa cumulata maggiore o uguale a 0,25.

Classe di riferimento: **[160-165]**.

$$Q_1 = x_{Q_{i-1}} + \frac{(0,25 - F_{Q_{i-1}})}{(F_{Q_i} - F_{Q_{i-1}})} \cdot (x_{Q_i} - x_{Q_{i-1}})$$

$$Q_1 = 160 + \frac{(25 - 0)}{0,3} \cdot 5 = 164,17$$

Terzo quartile

Stesso metodo, ma con frequenza relativa cumulata maggiore o uguale a 0,75.

Classe di riferimento: **[179-186]**.

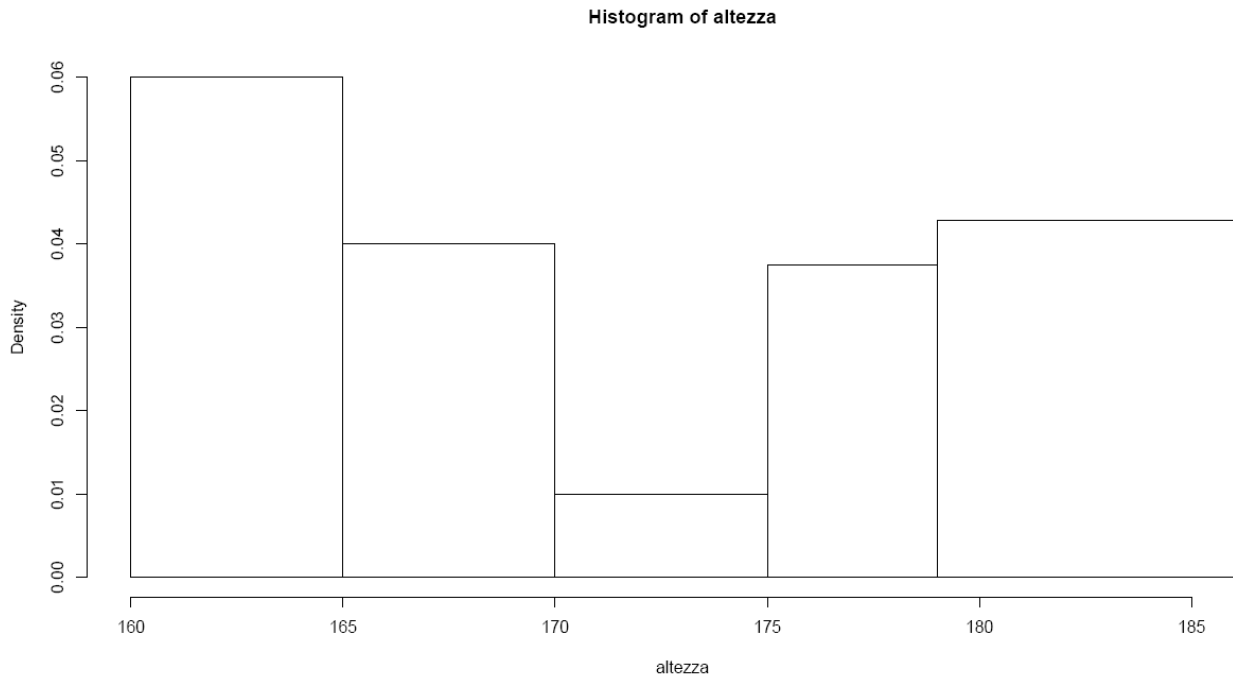
$$Q_3 = x_{Q_{3-1}} + \frac{(0,75 - F_{Q_{3-1}})}{(F_{Q_3} - F_{Q_{3-1}})} \cdot (x_{Q_3} - x_{Q_{3-1}})$$

$$Q_3 = 179 + \frac{(0,75 - 0,7)}{1 - 0,7} \cdot 7 = 180,17$$

REGOLA GENERALE

I quantili di una distribuzione in classi possono essere determinati in base alla formula generica:

$$x_{P_X} = x_{P_X-1} + \frac{(F_{\text{desiderata}} - F_{P_X-1})}{F_{P_X} - F_{P_X-1}} \cdot (x_{P_X} - x_{P_X-1})$$



ESERCIZIO n°3:

Rappresentare con un istogramma la distribuzione in classi del carattere peso dopo averla ripartita in 5 classi equifrequenti. Calcolare la media aritmetica e gli altri indici di posizione.

SOLUZIONE

Dividendo la frequenza assoluta totale ($n=20$) per il numero di classi richieste (5), si ottiene la frequenza (costante) di ogni classe (4). Dalla distribuzione semplice, ordinata in modo crescente, del carattere peso è possibile costruire 5 classi aventi frequenza assoluta costante e pari a 4.

peso: { 47 48 49 49 50 56 56 59 62 65 65 66 68 70 72 73 75 80 85 86 }

$x_{i-1} - x_i$	c_i	n_i	f_i	F_i	a_i	d_i
[47 - 49]	48	4	0,2	0,2	2	0,1
]49 - 59]	54	4	0,2	0,4	10	0,02
]59 - 66]	62,5	4	0,2	0,6	7	0,029
]66 - 73]	69,5	4	0,2	0,8	7	0,029
]73 -86]	79,5	4	0,2	1	13	0,015
Totali		20	1			

Media aritmetica

$$\mu = \frac{\sum_{i=1}^k C_i n_i}{n} = \frac{\sum_{i=1}^5 C_i n_i}{20} =$$

$$= \frac{(162,5 \times 6) + (167,5 \times 4) + (172,5 \times 1) + (177 \times 3) + (182,5 \times 6)}{20} = 172,215$$

Classe modale

Classe con estremi [**47-49**], ossia quella con densità di frequenza più elevata.

Mediana

La classe mediana è] **59-66** [. $F_{]59-66]} = 0,6$

$$Q_2 = Me = 59 + \frac{(0,5 - 0,4)}{(0,6 - 0,4)} \cdot 7 = 62,5$$

Primo quartile

Classe di riferimento:] **49-59** [. $F_{]49-59]} = 0,4$

$$Q_1 = 49 + \frac{(0,25 - 0,2)}{0,4 - 0,2} \cdot 10 = 51,5$$

Terzo quartile

Classe di riferimento:] **66-73**].

$$F_{]66-73]} = 0,8$$

$$Q_3 = 66 + \frac{(0,75 - 0,6)}{0,8 - 0,6} \cdot 7 = 71,25$$

Histogram of peso

