

Università di Cassino

Esercitazioni di Statistica 1 del 23 ottobre 2006

Dott.ssa Simona Balzano

Considerando il DATASET STUDENTI:

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

SESSO – CORSO LAUREA

SESSO – ATTIVITÀ SPORTIVA

CORSO LAUREA – ATTIVITÀ SPORTIVA

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];

2.3) a partire dalla distribuzione in classi determinare la proporzione di studenti:

a) con altezza compresa tra 170 e 173 cm;

b) più alti di 173 cm;

c) con altezza compresa tra 173 e 178 cm;

d) alti precisamente 173cm.

Soluzione

1) Tabelle di contingenza

SESSO – CORSO LAUREA

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

ATTIVITÀ SPORTIVA – SESSO

ATTIVITÀ SPORTIVA \ SESSO	SESSO		Totale
	Femmine	Maschi	
Nulla	1	5	6
Media	7	6	13
Alta	0	1	1
Totale	8	12	20

ATTIVITÀ SPORTIVA – CORSO LAUREA

ATTIVITÀ SPORTIVA \ CORSO LAUREA	CORSO LAUREA			Totale
	Biologia	Informatica	Matematica	
Nulla	0	4	2	6
Media	1	7	5	13
Alta	0	1	0	1
Totale	1	12	7	20

2.1)

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

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

Moda

Osservando la distribuzione di frequenze:

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

$Mo = 180$

Mediana

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:

$$n_1 = n_2 = 10$$

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

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

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

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

In alternativa, ma obbligatoriamente quando n è dispari, i quartili possono essere individuati sia a partire dalle frequenze relative cumulate:

x_i	n_i	f_i	F_i
160	3	0,15	0,15
162	1	0,05	0,2
164	1	0,05	0,25
165	1	0,05	0,3
170	4	0,2	0,5
172	1	0,05	0,55
176	1	0,05	0,6
178	1	0,05	0,65
179	1	0,05	0,7
180	5	0,25	0,95
186	1	0,05	1
	20	1	

sia come media aritmetica delle osservazioni che si trovano a destra e a sinistra dei valori $n/4$ (per Q_1) e $3/4n$ (per Q_3).

Non è detto che i risultati delle diverse procedure coincidano.

2.2) La distribuzione di frequenze è:

$x_{i-1} - x_i$		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			

Media aritmetica

$$\begin{aligned}\mu &= \frac{\sum_{i=1}^k x_i^c n_i}{n} = \frac{\sum_{i=1}^5 x_i^c n_i}{20} = \\ &= \frac{(162,5 \times 6) + (167,5 \times 4) + (172,5 \times 1) + (177 \times 3) + (182,5 \times 6)}{20} = 172,215\end{aligned}$$

Classe modale

Avendo le classi ampiezza non costante, la classe modale è identificata dalla densità di frequenza più elevata:

$$\text{classe modale} = [160 - 165]$$

come risulta anche osservando l'istogramma normalizzato.

Mediana e quartili

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

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

equivalente a:

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

per la quale si procede in questo modo:

- 1) si individua la classe di riferimento;
- 2) si sostituisce ad $F_{\text{desiderata}}$ il valore 0,25 per Q_1 , 0,5 per Q_2 e 0,75 per Q_3 .

In base alla frequenza relativa cumulata, il **primo quartile** si trova nella classe **[160 - 165]**, quindi:

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

Il secondo quartile, ossia la **mediana**, si trova nella classe **[165 - 170]**, anzi, avendo questa una frequenza relativa cumulata pari a 0,5, si vede immediatamente che coincide con l'estremo superiore della classe. Infatti anche applicando la formula avremmo:

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

Infine, il **terzo quartile** si trova nella classe **]179 - 186]**, quindi:

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

2.3)

Le quantità necessarie per determinare proporzioni di unità comprese in un dato intervallo rispetto al totale delle frequenze sono le seguenti:

$x_{i-1} - x_i$	f_i	d_i
[160 - 165]	0,3	0,06
]165 - 170]	0,2	0,04
]170 - 175]	0,05	0,01
]175 - 179]	0,15	0,0375
]179 - 186]	0,3	0,042857
totali	1	

a)

la proporzione di studenti con altezza compresa tra 170 e 173 cm si determina come segue:

$$f_{[170-173]} = (173 - 170) \times d_{[170-175]} = 3 \times 0,01 = 0,03$$

b)

Il totale degli studenti più alti di 173 cm è dato dalla somma di quelli compresi tra 173 e 175 più quelli più alti di 175 (ossia compresi nelle classi]175 - 179] e]179 - 186])

La proporzione di studenti più alti di 173 cm si determina, dunque, come:

$$\begin{aligned} f_{[>173]} &= f_{[173-175]} + f_{[175-179]} + f_{[179-186]} = \\ &= [(175 - 173) \times d_{[170-175]}] + 0,15 + 0,3 = \\ &= (2 \times 0,01) + 0,15 + 0,3 = 0,47 \end{aligned}$$

c)

La proporzione di studenti con altezza compresa tra 173 e 178 cm è uguale alla somma di due frazioni: di quelli compresi tra 173 e 175 più quelli compresi tra 175 e 178:

$$\begin{aligned}f_{[173-178]} &= f_{[173-175]} + f_{[175-178]} = \\&= \left[(175 - 173) \times d_{[170-175]} \right] + \left[(178 - 175) \times d_{[175-179]} \right] = \\&= (2 \times 0,01) + (3 \times 0,0375) = 0,1325\end{aligned}$$

d)

La proporzione di studenti alti **precisamente 173 cm non è determinabile.**