

Università di Cassino
Corso di Statistica 1
Esercitazione del 21/01/2008
Dott. Alfonso Piscitelli

Esercizio 1

Il seguente *data set* riporta la rilevazione di alcuni caratteri su un collettivo di 20 soggetti.

Soggetto	Età	Residenza	Reddito (Migliaia di €)	Auto Possedute	Punteggio quiz
1	22	Cantagallo	0,7	3	173
2	18	Cantagallo	0,2	4	168
3	34	Poggio a Caiano	1,6	2	165
4	42	Carmignano	2,5	5	180
5	50	Poggio a Caiano	3,2	3	163
6	12	Montemurlo	0,1	4	160
7	46	Carmignano	3,8	4	177
8	72	Montemurlo	1,3	2	164
9	27	Montemurlo	1,2	3	158
10	48	Carmignano	1,7	5	170
11	35	Montemurlo	1,9	1	167
12	84	Cantagallo	0,8	1	159
13	21	Montemurlo	0,4	5	174
14	44	Carmignano	1,8	4	164
15	56	Carmignano	1,9	2	177
16	58	Montemurlo	3,2	3	172
17	37	Cantagallo	2,1	1	166
18	16	Montemurlo	0,1	4	160
19	73	Carmignano	1,6	2	170
20	64	Poggio a Caiano	2,2	3	184

- a. Costruire la distribuzione di frequenza per il carattere **Reddito** suddividendo la distribuzione in 5 classi equifrequenti.
- b. Rappresentare graficamente la distribuzione equifrequente.
- c. Determinare la mediana ed i quartili per il carattere **Auto Possedute** a partire sia dalla successione di valori sia dalla distribuzione di frequenze.
- d. Esprimere il carattere **Reddito** in 3 modalità ordinali (basso, medio e alto) secondo la seguente corrispondenza:

basso (fino a 1,3 Migliaia di €)
medio (da 1,4 a 2,6 Migliaia di €)
Alto (da 2,7 a 3,8 Migliaia di €)

costruire per il nuovo carattere ottenuto la distribuzione di frequenza e determinarne la mediana.

e. Costruire la distribuzione di frequenza per il carattere **Età** suddividendo la distribuzione in 4 classi equiampie e determinare il valore della mediana e dei quartili.

f. Determinare il valore del secondo decile, del settantesimo e del novantesimo percentile per il carattere **Età** utilizzando la suddivisione in classi operata precedentemente.

Soluzioni

a) Nella tabella sono riportati i valori ordinati della variabile **Reddito**.

Osservazione	Reddito
6	0,1
18	0,1
2	0,2
13	0,4
1	0,7
12	0,8
9	1,2
8	1,3
3	1,6
19	1,6
10	1,7
14	1,8
11	1,9
15	1,9
17	2,1
20	2,2
4	2,5
5	3,2
16	3,2
7	3,8

Avendo a disposizione 20 unità e dovendo costruire 5 classi equifrequenti (cioè con la stessa frequenza) viene da se che ogni classe dovrà avere una frequenza pari a $20 / 5 = 4$. In altre parole ogni classe deve contenere 4 unità statistiche.

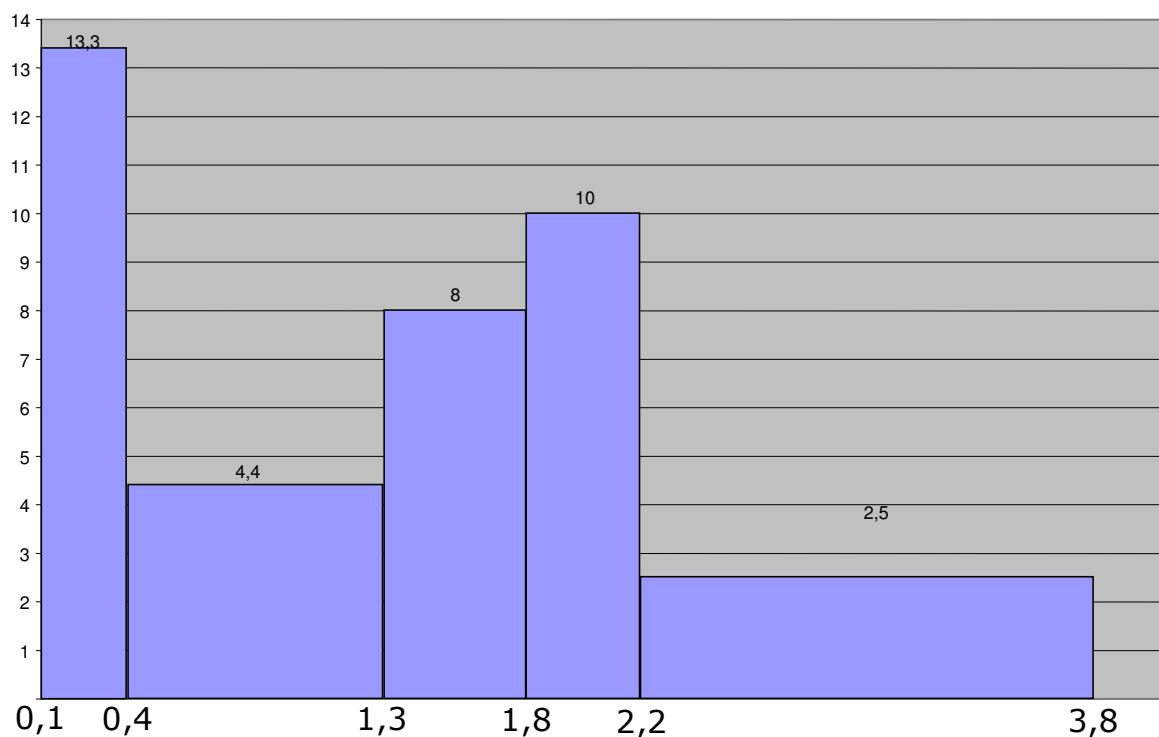
La prima classe sarà quindi ottenuta prendendo i primi 4 valori della variabile in esame, la seconda i successivi 4 e così via.

Nel nostro caso:

Reddito	d_i	n_i	f_i	F_i	h_i
0,1 - 0,4	0,3	4	0,02	0,02	13,3
0,4 - 1,3	0,9	4	0,02	0,04	4,4
1,3 - 1,8	0,5	4	0,02	0,06	8,0
1,8 - 2,2	0,4	4	0,02	0,08	10,0
2,2 - 3,8	1,6	4	0,02	1	2,5

Tot: 20 1

b) Sulla base dei risultati ottenuti nella precedente tabella è possibile costruire l'**istogramma** associato alla nostra distribuzione di frequenze. Essendo le classi di ampiezza diversa le basi dei rettangoli associati alle diverse classi avranno lunghezza diversa, in particolare maggiore è l'ampiezza della classe più deve essere lunga la base del rettangolo associato a quella classe.



- c) La successione dei valori ordinati in senso non decrescente e la corrispondente distribuzione di frequenza della variabile **Auto Possedute** sono le seguenti:

Soggetto	Posizione	Auto Possedute
11	1	1
12	2	1
17	3	1
3	4	2
8	5	2
15	6	2
19	7	2
1	8	3
5	9	3
9	10	3
16	11	3
20	12	3
2	13	4
6	14	4
7	15	4
14	16	4
18	17	4
4	18	5
10	19	5
13	20	5

Auto Possedute	n_i	f_i	F_i
1	3	0,15	0,15
2	4	0,2	0,35
3	5	0,25	0,60
4	5	0,25	0,85
5	3	0,15	1

Tot: 20 1

Il primo quartile corrisponde a quel valore del carattere X che lascia alla sua sinistra il 25% delle osservazioni e alla sua destra il 75%.

La mediana è quel valore che suddivide la distribuzione, ordinata in modo decrescente o non decrescente, in due parti uguali lasciando quindi il 50% delle osservazioni alla sua destra e il restante 50% alla sua sinistra.

Nel caso di N pari, cioè di un numero di osservazioni pari, la mediana è data dalla media aritmetica del valore che occupa la $(N/2)$ -esima posizione e il successivo, cioè quello che occupa la $(N/2)+1$ -esima. Nel nostro caso $20/2=10 \Rightarrow$ media tra la 10° e l'11°posizione.

Il terzo quartile corrisponde a quel valore del carattere X che lascia alla sua sinistra il 75% delle osservazioni e alla sua destra il rimanente 25%.

$$Q_1 = \frac{X_{\frac{N}{4}} + X_{\frac{N}{4}+1}}{2} = \frac{X_5 + X_6}{2} = \frac{2+2}{2} = 2$$

$$Q_2 = M_e = 3$$

$$Q_3 = \frac{X_{\frac{3 \cdot N}{4}} + X_{\frac{3 \cdot N}{4}+1}}{2} = \frac{X_{15} + X_{16}}{2} = \frac{4+4}{2} = 4$$

Nel caso delle distribuzioni di frequenza semplice, invece, i quartili vengono individuati facendo riferimento alle frequenze cumulate o alle frequenze relative cumulate. In questo caso:

- . \Rightarrow il primo quartile è quel valore della x associato alla prima frequenza relativa cumulata maggiore di 0,25. [$Q_1=2$]
- . \Rightarrow il secondo quartile o mediana è quel valore della x associato alla prima frequenza relativa cumulata maggiore di 0,50. [$Q_2=Me=3$]
- . \Rightarrow il terzo quartile è quel valore della x associato alla prima frequenza relativa cumulata maggiore di 0,75. [$Q_3=4$].

d) La distribuzione di frequenza dalla variabile **Reddito** suddivisa nelle 3 modalità ordinali indicate dal testo dell'esercizio è la seguente:

Reddito	n_i	f_i	F_i
Basso	8	0,4	0,4
Medio	9	0,45	0,85
Alto	3	0,15	1

Tot 20

Per individuare la mediana si fa riferimento alla distribuzione di frequenze cumulate relative. Individueremo il valore mediano come quello associato alla prima frequenza cumulata relativa che supera lo 0,50. Nel nostro caso la mediana è la classe di reddito [1,4 e 2,6 Migliaia di €], in altre parole, la modalità "Medio".

e) La distribuzione in classi di frequenza del carattere **Età**, è:

Età	n_i	N_i	f_i	F_i	d_i
12 - 30	6	6	0,3	0,3	18
30 - 48	6	12	0,3	0,6	18
48 - 66	5	17	0,25	0,85	18
66 - 84	3	20	0,15	1	18
Tot:	20		1		

Prima di passare al calcolo della mediana, per dati in classi, bisogna evidenziare la classe mediana. La classe mediana è quella associata alla prima frequenza cumulata relativa che supera lo 0,50.

. ⇒ Classe Me=30-|48 [in cui il valore della x associato alla prima frequenza cumulata è maggiore di 0,50].

Quindi, la mediana sarà:

$$M_e = l_e + \frac{\sum n_i - N_{e-1}}{n_e} d_e$$

dove:

l_e =limite inferiore della classe Mediana;

N_{e-1} =frequenza cumulata associata alla classe precedente a quella Mediana;

n_e =frequenza assoluta della classe Mediana;

d_e =ampiezza della classe Mediana;

$$M_e = 30 + \frac{10-6}{6} 18 = 42$$

Le classi del primo e del terzo quartile, si individuano in corrispondenza delle rispettive frequenze cumulate relative. Avremo quindi che:

. ⇒ Classe $Q_1=12-|30$ [valore della x associato alla prima frequenza cumulata maggiore di 0,25].

. ⇒ Classe $Q_3=48-|66$ [valore della x associato alla prima frequenza cumulata maggiore di 0,75].

Quindi, il primo quartile sarà:

$$Q_1 = l_{Q_1} + \frac{\sum n_i - N_{Q_1-1}}{n_{Q_1}} d_{Q_1}$$

dove:

l_{Q_1} = limite inferiore della classe Q_1 ;

N_{Q_1-1} = frequenza cumulata associata alla classe precedente a quella di Q_1 ;

n_{Q_1} = frequenza assoluta della classe Q_1 ;

d_{Q_1} = ampiezza della classe Q_1 ;

$$Q_1 = 12 + \frac{5-0}{6} 18 = 27$$

Quindi, il terzo quartile sarà:

$$Q_3 = l_{Q_3} + \frac{3 \sum n_i - N_{Q_3-1}}{n_{Q_3}} d_{Q_3}$$

dove:

l_{Q_3} = limite inferiore della classe Q_3 ;

N_{Q_3-1} = frequenza cumulata associata alla classe precedente a quella di Q_3 ;

n_{Q_3} = frequenza assoluta della classe Q_3 ;

d_{Q_3} = ampiezza della classe Q_3 ;

$$Q_3 = 48 + \frac{15-12}{5} 18 = 58,8$$

f) Prima di passare al calcolo del secondo decile, bisogna evidenziare la classe in cui è presente il secondo decile.

Età	n_i	N_i	f_i	F_i	d_i
12 - 30	6	6	0,3	0,3	18
30 - 48	6	12	0,3	0,6	18
48 - 66	5	17	0,25	0,85	18
66 - 84	3	20	0,15	1	18
Tot:	20		1		

La classe del secondo decile è quella associata alla prima frequenza cumulata relativa che supera il valore di 0,20.

. ⇒ Classe $D_2=12|-30$ [in cui il valore della x associato alla prima frequenza cumulata è maggiore di 0,20].

Quindi, il secondo decile sarà:

$$D_2 = l_d + \frac{\sum n_i (2) - N_{d-1}}{n_d} d_d$$

dove:

l_d = limite inferiore della classe del secondo decile;

N_{d-1} = frequenza cumulata associata alla classe precedente a quella del secondo decile;

n_d = frequenza assoluta della classe del secondo decile;

d_d = ampiezza della classe del secondo decile;

$$D_2 = 12 + \frac{4-0}{6} 18 = 12 + 12 = 24$$

Le classi del settantesimo e del novantesimo percentile, si individuano in corrispondenza delle rispettive frequenze cumulate relative. Avremo quindi che:

. ⇒ Classe $C_{70}=48|-66$ [valore della x associato alla prima frequenza relativa cumulata maggiore di 0,70].

. ⇒ Classe $C_{90}=66|-184$ [valore della x associato alla prima frequenza relativa cumulata maggiore di 0,90].

Quindi, il settantesimo percentile sarà:

$$C_{70} = l_{C_{70}} + \frac{\sum n_i (70) - N_{C_{70}-1}}{n_{C_{70}}} d_{C_{70}}$$

dove:

$l_{C_{70}}$ = limite inferiore della classe C_{70} ;

$N_{C_{70}-1}$ = frequenza cumulata associata alla classe precedente a quella di C_{70} ;

$n_{C_{70}}$ = frequenza assoluta della classe C_{70} ;

$d_{C_{70}}$ = ampiezza della classe C_{70} ;

$$C_{70} = 48 + \frac{14-12}{5} 18 = 48 + 7,2 = 55,2$$

Quindi, il novantesimo percentile sarà:

$$C_{90} = l_{C_{90}} + \frac{\sum n_i (90) - N_{C_{90-1}}}{n_{C_{90}}} d_{C_{90}}$$

dove:

$l_{C_{90}}$ =limite inferiore della classe C_{90} ;

$N_{C_{90-1}}$ =frequenza cumulata associata alla classe precedente a quella di C_{90} ;

$n_{C_{90}}$ =frequenza assoluta della classe C_{90} ;

$d_{C_{90}}$ =ampiezza della classe C_{90} ;

$$C_{90} = 66 + \frac{18-17}{3} 18 = 66 + 6 = 72$$