

STATISTICA: esercizi svolti su

MODA, MEDIANA,
QUARTILI, DECILI e CENTILI

1 MODA, MEDIANA, QUARTILI, DECILI E CENTILI

1. Viene rilevato il tempo X (in secondi) necessario per l'esecuzione di 6 diverse procedure su uno stesso tipo di calcolatore:

Procedura	A	B	C	D	E	F
Tempo (X)	33,6	34,2	35,0	30,6	30,2	31,4

Calcolare moda, mediana e primo quartile del carattere X .

Svolgimento

La distribuzione riportata dal testo dell'esercizio è una distribuzione di unità in cui non si presenta alcuna ripetizione delle intensità, di conseguenza la moda non esiste. Per quanto riguarda il calcolo della mediana, si procede anzitutto ad ordinare in senso non decrescente gli $N = 6$ valori forniti dal testo dell'esercizio. Si ottiene:

$$x_{(1)} = 30,2; x_{(2)} = 30,6; x_{(3)} = 31,4; x_{(4)} = 33,6; x_{(5)} = 34,2; x_{(6)} = 35.$$

Dato che N è pari, si hanno le due posizioni centrali:

$$\frac{N}{2} = 3 \qquad \frac{N}{2} + 1 = 4.$$

A questo punto la mediana è data da:

$$\frac{x_{(3)} + x_{(4)}}{2} = \frac{31,4 + 33,6}{2} = 32,5.$$

Il valore assunto dalla mediana ci dice che nel 50% dei casi circa, il tempo necessario per l'esecuzione di una procedura è inferiore a 32.5 secondi. Analogamente, nel 50% dei casi circa, il tempo necessario per l'esecuzione di una procedura è superiore a 32.5 secondi.

Passiamo ora al calcolo del primo quartile.

$$\begin{aligned} Q_1 &= x_{(1, \frac{N+1}{4})} = x_{(1,75)} \\ &= x_{(1)} + 0,75(x_{(2)} - x_{(1)}) \\ &= 30,2 + 0,75(30,6 - 30,2) = 30,5 \end{aligned}$$

Il valore assunto dal primo quartile ci dice che nel 25% dei casi circa, il tempo necessario per l'esecuzione di una procedura è inferiore a 30.5 secondi. Dunque nel 75% dei casi il tempo di esecuzione è superiore a 30.5 secondi.

2. In un asilo nido si è verificata un'epidemia di influenza. La distribuzione del carattere *giorni effettivi di assenza dall'asilo per la malattia* sui 20 bambini colpiti dall'infezione è riportata nella tabella sottostante dove il carattere è stato raggruppato in classi.

Numero giorni	1 - 3	4 - 5	6 - 8
Frequenza	4	6	10

Calcolare la moda della distribuzione.

Svolgimento

Nel caso di distribuzioni con dati raggruppati in classi la determinazione della moda si riduce all'individuazione della classe modale la quale si ha in corrispondenza della frequenza specifica più elevata. E' dunque necessario calcolare le frequenze specifiche:

Numero giorni	Frequenza	ampiezza classe	frequenza specifica
1 \vdash 3	4	3	1,33
4 \vdash 5	6	2	3
6 \vdash 8	10	3	3,33

Dal confronto delle frequenze specifiche si conclude che la classe modale è 6 \vdash 8.

3. La seguente tabella riporta la distribuzione del carattere X numero di stanze di 120 abitazioni della provincia di Belluno.

numero di stanze	1	2	3	4	5	6	7	8
n_i	5	22	32	35	16	7	2	1

Calcolare moda e mediana della distribuzione. Commentare i risultati ottenuti. Quante sono le abitazioni con al più due camere? Che percentuale rappresentano? Quante sono le abitazione con almeno tre camere? Che percentuale rappresentano? Calcolare i quartili, il secondo e il settimo decile e il 59-mo centile della distribuzione. Commentare i risultati ottenuti.

Svolgimento

Riportiamo di seguito alcune calcoli che ci saranno utili nello svolgimento dell'esercizio.

Numero di stanze	n_i	$\frac{n_i}{N}$	Freq. Cumulate	Freq. Retrocum.
1	5	0,0416	5	120
2	22	0,1833	27	115
3	32	0,2666	59	93
4	35	0,2916	94	61
5	16	0,1333	110	26
6	7	0,0583	117	10
7	2	0,0166	119	3
8	1	0,0083	120	1
<i>tot</i>	120	1	—	—

Dalla tabella sopra riportata appare chiaro che la moda si ha in corrispondenza di $X = 4$. Si osservi che la moda ha una frequenza relativa pari a 0,2916. Di conseguenza essa non può ritenersi una buona sintesi in quanto rappresenta solo il 29% circa delle abitazioni.

Per quanto riguarda il calcolo della mediana, osserviamo che N è pari e di conseguenza abbiamo due posizioni centrali:

$$\frac{N}{2} = 60 \qquad \frac{N}{2} + 1 = 61.$$

Osservando le frequenze cumulate, si possono individuare le osservazioni che occupano le posizioni centrali. Sulla base di queste osservazioni la mediana risulta essere:

$$\frac{x_{(60)} + x_{(61)}}{2} = \frac{4 + 4}{2} = 4.$$

La mediana bi-ripartisce l'insieme delle abitazioni in due gruppi: il primo composto da abitazioni che hanno un numero di stanze minore o uguale a 4, il secondo composto da abitazioni che hanno un numero di stanze maggiore o uguale 4. Nel dettaglio, il valore assunto dalla mediana ci dice che:

- circa il 50% delle abitazioni considerate ha un numero di stanze al più pari a 4;
- circa il 50% delle abitazioni considerate ha un numero di stanze almeno pari a 4.

Il numero di abitazioni con al più due camere ci viene fornito dalla seconda frequenza cumulata che è pari a 27. Esse rappresentano una quota delle abitazioni pari a

$$\frac{27}{120} = 0,225$$

che coincide con il 22,5% delle abitazioni. Il numero di abitazioni con almeno tre camere ci viene fornito dalla terza frequenza retrocumulata che è pari a 93. Esse rappresentano una quota delle abitazioni pari a

$$\frac{93}{120} = 0,775$$

che coincide con il 77,5% delle abitazioni.

Passiamo al calcolo dei quartili.

$$\begin{aligned} Q_1 &= x_{\left(\frac{120+1}{4}\right)} = x_{(30,25)} \\ &= x_{(30)} + 0,25[x_{(31)} - x_{(30)}] \\ &= 3 + 0,25[3 - 3] = 3 \end{aligned}$$

$$\begin{aligned} Q_2 &= Me = x_{\left(\frac{120+1}{2}\right)} = x_{(60,5)} \\ &= x_{(60)} + 0,5[x_{(61)} - x_{(60)}] \\ &= 4 + 0,5[4 - 4] = 4 \end{aligned}$$

$$\begin{aligned} Q_3 &= x_{\left(3 \cdot \frac{120+1}{4}\right)} = x_{(90,75)} \\ &= x_{(90)} + 0,75[x_{(91)} - x_{(90)}] \\ &= 4 + 0,75[4 - 4] = 4 \end{aligned}$$

Il valore assunto da Q_1 ci informa che circa il 25% delle abitazioni considerate ha un numero di stanze al più pari a 3 e dunque circa il 75% delle abitazioni ha un numero

di stanze almeno pari a 3. Il valore assunto da Q_2 ci informa che circa il 50% delle abitazioni considerate ha un numero di stanze al più pari a 4 e dunque circa il 75% delle abitazioni considerate ha un numero di stanze almeno pari a 4. Tale informazione è la medesima che ci viene fornita dalla mediana. Il valore assunto da Q_3 ci informa che circa il 75% delle abitazioni considerate ha un numero di stanze al più pari a 4 e dunque circa il 25% delle abitazioni ha un numero di stanze almeno pari a 4.

Il secondo ed il settimo decile della distribuzione sono dati da:

$$\begin{aligned} D_2 &= x_{\left(\frac{120+1}{5}\right)} = x_{(24,2)} \\ &= x_{(24)} + 0,2[x_{(25)} - x_{(24)}] \\ &= 2 + 0,2[2 - 2] = 2 \end{aligned}$$

$$\begin{aligned} D_7 &= x_{\left(7 \cdot \frac{120+1}{10}\right)} = x_{(84,7)} \\ &= x_{(84)} + 0,7[x_{(85)} - x_{(84)}] \\ &= 4 + 0,7[4 - 4] = 4 \end{aligned}$$

Il valore assunto da D_2 ci informa che circa il 10% delle abitazioni considerate ha un numero di stanze al più pari a 2 e dunque circa il 90% delle abitazioni ha un numero di stanze almeno pari a 2. Il valore assunto da D_7 ci informa che circa il 70% delle abitazioni considerate ha un numero di stanze al più pari a 4 e dunque circa il 30% delle abitazioni ha un numero di stanze almeno pari a 4.

Il cinquantanovesimo centile della distribuzione è dato da:

$$\begin{aligned} C_{59} &= x_{\left(59 \cdot \frac{120+1}{100}\right)} = x_{(71,39)} \\ &= x_{(71)} + 0,39[x_{(72)} - x_{(71)}] \\ &= 4 + 0,39[4 - 4] = 4 \end{aligned}$$

Il valore assunto da C_{59} ci informa che circa il 59% delle abitazioni considerate ha un numero di stanze al più pari a 4 e dunque circa il 41% delle abitazioni ha un numero di stanze pari almeno pari a 4.

4. La seguente tabella riporta la distribuzione del carattere X *fatturato* (in miliardi di lire) di 1240 calzaturifici italiani nell'anno 1996.

classi di fatturato	fino a 0,5	0,5 - 1	1 - 5	5 - 10	oltre 10
n_i	340	368	480	37	15

Calcolare moda e mediana della distribuzione. Commentare i risultati ottenuti. Quanti calzaturifici hanno avuto un fatturato minore o uguale a un miliardo? Che percentuale ha avuto un fatturato di almeno un miliardo? Quanti calzaturifici hanno avuto un fatturato compreso tra 3 e 6,5 miliardi?

Calcolare i quartili, i primi cinque decili e i primi 3 centili della distribuzione. Commentare i risultati ottenuti.

Svolgimento

Osserviamo innanzi tutto che la distribuzione di frequenze in considerazione riguarda un carattere quantitativo continuo con modalità raggruppate in classi.

Riportiamo di seguito alcuni calcoli che ci saranno utili nello svolgimento dell'esercizio.

Fatturato	n_i	$\frac{n_i}{N}$	Fr. Cum.	Fr. Retrocum.	Ampiezza	Fr. Spec.
fino a 0,5	340	0,2741	340	1240	0,5	680
0,5 - 1	368	0,2967	708	900	0,5	736
1 - 5	480	0,387	1188	532	4	120
5 - 10	37	0,029	1225	52	5	7,4
oltre 10	15	0,012	1240	15	10	1,5

Dalla tabella sopra riportata si osserva che la classe di fatturato in corrispondenza della quale si ha la frequenza specifica più alta è 0,5 - 1. Tale classe di fatturato coincide con la classe modale.

Per quanto riguarda il calcolo della mediana, osserviamo che $N = 1240$ è un numero pari e di conseguenza abbiamo due posizioni centrali

$$\frac{N}{2} = 620 \qquad \frac{N}{2} + 1 = 621$$

le quali, come evidenziano le frequenze cumulate, cadono entrambe nella classe 0,5 - 1. Dato che il carattere quantitativo *fatturato* è continuo con modalità raggruppate in classi, la mediana viene calcolata nel modo seguente¹:

$$\begin{aligned} Me &= l_j^- + \left[\frac{N}{2} - C_{j-1} \right] \frac{a_j}{n_j} \\ &= 0,5 + [620 - 340] \frac{0,5}{368} = 0,88 \end{aligned}$$

Il valore assunto dalla mediana ci informa che circa il 50% dei calzaturifici ha un fatturato minore o uguale a 0,88 miliardi di lire. Di conseguenza, il rimanente 50% circa ha un fatturato maggiore o uguale a 0.88 miliardi di lire.

Il numero di calzaturifici che hanno avuto un fatturato minore o uguale ad un miliardo di lire ci viene fornito dalla seconda frequenza cumulata che è pari a 708.

Il numero di calzaturifici che hanno avuto un fatturato di almeno un miliardo di lire ci viene fornito dalla terza frequenza retrocumulata che è pari a 532. Possiamo quindi concludere che la percentuale di calzaturifici che hanno avuto almeno un miliardo di fatturato è pari a

$$\frac{532}{1240} \cdot 100 = 42.8\%.$$

I calzaturifici che hanno avuto un fatturato compreso tra 3 e 6,5 miliardi sono dati da:

$$(5 - 3)120 + (6,5 - 5)7,4 = 251,1$$

¹ Zenga M. ; *Lezioni di statistica descrittiva*; pagina 99.

Tale risultato si basa sull'ipotesi che, all'interno di ogni classe di reddito, le frequenze si distribuiscano in modo uniforme. Graficamente, il numero di calzaturifici che hanno avuto un fatturato compreso tra 3 e 6,5 miliardi coincide con l'area evidenziata nel grafico:

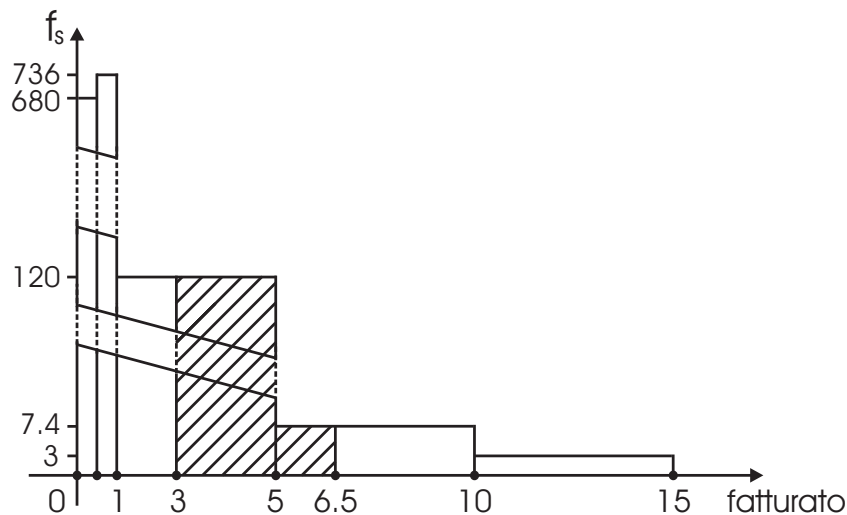


Fig. 1: Grafico delle frequenze specifiche del carattere "Fatturato".

Per quanto riguarda il calcolo dei quartili, determiniamo innanzi tutto le loro posizioni:

$$Pos(Q_1) = \frac{1241}{4} = 310,25 \quad Pos(Q_2) = 2 \cdot \frac{1241}{4} = 620,5$$

$$Pos(Q_3) = 3 \cdot \frac{1241}{4} = 930,75.$$

In base alle frequenze cumulate, le classi del primo, secondo e terzo quartile risultano essere rispettivamente: fino a 0,5; 0,5 + 1; 1 + 5.

Nel caso della prima classe, sembra opportuno fissare a 0 il suo estremo inferiore. I quartili risultano dunque essere:

$$Q_1 = 0 + [310,25 - 0 - 0,5] \cdot \frac{0,5}{340} = 0,46$$

$$Q_2 = Me = 0,88$$

$$Q_3 = 1 + [930,75 - 708 - 0,5] \cdot \frac{4}{480} = 2,852$$

Il valore assunto da Q_1 ci informa che all'incirca il 25% dei calzaturifici ha avuto nel 1996 un fatturato minore o uguale a 0,46 miliardi di lire. Il rimanente 75% circa ha avuto un fatturato maggiore o uguale a 0,46 miliardi di lire. Q_2 coincide con la mediana ed ha dunque il suo stesso significato. Il valore assunto da Q_3 ci informa che il 75% dei calzaturifici ha avuto nel 1996 un fatturato minore o uguale a 2,852 miliardi di lire. Il rimanente 25% circa ha avuto un fatturato maggiore o uguale a 2,852 miliardi di lire.

Per quanto riguarda il calcolo dei primi cinque decili, determiniamo innanzi tutto le loro posizioni:

$$\begin{aligned} Pos(D_1) &= \frac{1241}{10} = 124,1 & Pos(D_2) &= 2 \cdot \frac{1241}{10} = 248,2 \\ Pos(D_3) &= 3 \cdot \frac{1241}{10} = 372,3 & Pos(D_4) &= 4 \cdot \frac{1241}{10} = 496,4 \\ Pos(D_5) &= 5 \cdot \frac{1241}{10} = 620,5 \end{aligned}$$

Osservando le frequenze cumulate, le classi associate a tali posizioni sono rispettivamente: fino a 0,5; fino a 0,5; 0,5 + 1; 0,5 + 1; 0,5 + 1. Utilizzando 0 come estremo inferiore della prima classe, i primi cinque decili risultano essere:

$$\begin{aligned} D_1 &= 0 + [124,1 - 0 - 0,5] \frac{0,5}{340} = 0,182 \\ D_2 &= 0 + [248,2 - 0 - 0,5] \frac{0,5}{340} = 0,36 \\ D_3 &= 0,5 + [372,3 - 340 - 0,5] \frac{0,5}{368} = 0,543 \\ D_4 &= 0,5 + [496,4 - 340 - 0,5] \frac{0,5}{368} = 0,712 \\ D_5 &= Me = 0,88 \end{aligned}$$

Il valore assunto da D_1 ci informa che il 10% dei calzaturifici ha avuto nel 1996 un fatturato minore o uguale a 0,182 miliardi di lire. Il restante 90% circa ha avuto un fatturato maggiore o uguale a 0,182 miliardi di lire. Il valore assunto da D_2 ci informa che il 20% dei calzaturifici ha avuto nel 1996 un fatturato minore o uguale a 0,36 miliardi di lire. Il restante 80% circa ha avuto un fatturato maggiore o uguale a 0,36 miliardi di lire. Il valore assunto da D_3 ci informa che il 30% dei calzaturifici ha avuto nel 1996 un fatturato minore o uguale a 0,543 miliardi di lire. Il restante 70% circa ha avuto un fatturato maggiore o uguale a 0,543 miliardi di lire. Il valore assunto da D_4 ci informa che il 40% dei calzaturifici ha avuto nel 1996 un fatturato minore o uguale a 0,712 miliardi di lire. Il restante 90% circa ha avuto un fatturato maggiore o uguale a 0,712 miliardi di lire. D_5 coincide con la mediana ed ha dunque il suo stesso significato.

Per quanto riguarda il calcolo dei primi tre centili, determiniamo innanzitutto le loro posizioni:

$$\begin{aligned} Pos(C_1) &= \frac{1241}{100} = 12,41 & Pos(C_2) &= 2 \cdot \frac{1241}{100} = 24,82 \\ Pos(C_3) &= 3 \cdot \frac{1241}{100} = 37,23 \end{aligned}$$

Osservando le frequenze cumulate, la classe associata a tali posizioni è: fino a 0,5. Utilizzando 0 come estremo inferiore della prima classe, i primi tre centili risultano essere:

$$C_1 = 0 + [12,41 - 0 - 0,5] \frac{0,5}{340} = 0,0175$$

$$C_2 = 0 + [24,82 - 0 - 0,5] \frac{0,5}{340} = 0,0358$$

$$C_3 = 0,5 + [37,23 - 0 - 0,5] \frac{0,5}{340} = 0,0540$$

Il valore assunto da C_1 ci informa che l'1% dei calzaturifici ha avuto nel 1996 un fatturato minore o uguale a 0,0175 miliardi di lire. Il rimanente 99% circa ha avuto un fatturato maggiore o uguale a 0,0175. Il valore assunto da C_2 ci informa che il 2% dei calzaturifici ha avuto nel 1996 un fatturato minore o uguale a 0,0358 miliardi di lire. Il rimanente 98% circa ha avuto un fatturato maggiore o uguale a 0,0358. Il valore assunto da C_3 ci informa che il 3% dei calzaturifici ha avuto nel 1996 un fatturato minore o uguale a 0,0540 miliardi di lire. Il rimanente 97% circa ha avuto un fatturato maggiore o uguale a 0,0540.

5. Gli esercizi alberghieri della provincia di Belluno sono stati classificati in base al numero di camere per esercizio. I risultati dell'indagine sono riportati nella seguente tabella:

numero di camere	fino a 5	6 † 9	10 † 13	14 † 17	18 † 22	23 † 30	totale
numero esercizi	20	20	20	24	25	16	125

Calcolare la moda della distribuzione e commentare.

Svolgimento

La tabella fornita dal testo dell'esercizio riporta la distribuzione di frequenze di un carattere quantitativo discreto con dati raggruppati in classi. In tal caso il calcolo della moda si riduce all'individuazione della classe modale che si identifica in quella con frequenza specifica maggiore. Nel caso della prima classe, sembra opportuno fissare ad 1 il suo estremo inferiore. Alla luce di ciò calcoliamo le frequenze specifiche.

numero di camere	fino a 5	6 † 9	10 † 13	14 † 17	18 † 22	23 † 30	totale
numero esercizi	20	20	20	24	25	16	125
ampiezza classe	5	4	4	4	5	8	
freq. specifiche	4	5	5	6	5	2	

La classe modale risulta quindi essere: 14 † 17.

6. La seguente tabella riporta le distribuzioni degli impiegati civili dello stato secondo la qualifica funzionale:

Qualifica	II	III	IV	V	VI	VII	VIII	Totale
n_i	58038	308249	287707	71974	52232	28081	12259	818540

Calcolare i quartili, i decili, il 96-esimo centile della distribuzione. Commentare i risultati.

Svolgimento

La distribuzione di frequenze in considerazione riguarda un carattere qualitativo rilevato su scala ordinale. Al fine di calcolare i percentili richiesti dal testo dell'esercizio, ricaviamo le frequenze cumulate.

Qualifica	n_i	Freq. Cumulate
II	58038	58038
III	308249	366287
IV	287707	653994
V	71974	725968
VI	52232	778200
VII	28081	806281
VIII	12259	818540

Le posizioni dei quartili sono date da:

$$Pos(Q_1) = \left(\frac{N+1}{4} \right) = 204635,25$$

$$Pos(Q_2) = 2 \cdot \left(\frac{N+1}{4} \right) = 409270,5$$

$$Pos(Q_3) = 3 \cdot \left(\frac{N+1}{4} \right) = 613905,75$$

Sulla base delle frequenze cumulate calcolate in precedenza possiamo concludere che:

$$Q_1 = III; \quad Q_2 = Me = IV; \quad Q_3 = IV.$$

Tali valori ci dicono che:

- All'incirca il 25% degli impiegati civili dello stato ha una qualifica al più pari a *III*. Pertanto il rimanente 75% circa ha una qualifica almeno pari a *III*.
- All'incirca il 50% degli impiegati civili dello stato ha una qualifica al più pari a *IV*. Pertanto il rimanente 50% circa ha una qualifica almeno pari a *IV*.
- All'incirca il 75% degli impiegati civili dello stato ha una qualifica al più pari a *IV*. Pertanto il rimanente 25% circa ha una qualifica almeno pari a *IV*.

Le posizioni dei decili sono date da:

$$Pos(D_1) = \left(\frac{N+1}{10} \right) = 81854 \quad Pos(D_2) = 2 \cdot \left(\frac{N+1}{10} \right) = 163708$$

$$Pos(D_3) = 3 \cdot \left(\frac{N+1}{10} \right) = 245562 \quad Pos(D_4) = 4 \cdot \left(\frac{N+1}{10} \right) = 327416$$

$$Pos(D_5) = 5 \cdot 81854 = 409270 \quad Pos(D_6) = 6 \cdot 81854 = 491124$$

$$Pos(D_7) = 7 \cdot 81854 = 572978 \quad Pos(D_8) = 8 \cdot 81854 = 654832$$

$$Pos(D_9) = 9 \cdot 81854 = 736686$$

Sulla base delle frequenze cumulate calcolate in precedenza possiamo concludere che:

$$D_1 = III \quad D_2 = III \quad D_3 = III \quad D_4 = III \quad D_5 = Me = IV$$

$$D_6 = IV \quad D_7 = IV \quad D_8 = V \quad D_9 = VI$$

Tali valori ci dicono che:

- All'incirca il 10% degli impiegati civili dello stato ha una qualifica al più pari a *III*. Il rimanente 90% circa ha una qualifica almeno pari a *III*.
- All'incirca il 20% degli impiegati civili dello stato ha una qualifica al più pari a *III*. Il rimanente 80% circa ha una qualifica almeno pari a *III*.
- All'incirca il 30% degli impiegati civili dello stato ha una qualifica al più pari a *III*. Il rimanente 70% circa ha una qualifica almeno pari a *III*.
- All'incirca il 40% degli impiegati civili dello stato ha una qualifica al più pari a *III*. Il rimanente 60% circa ha una qualifica almeno pari a *III*.
- All'incirca il 50% degli impiegati civili dello stato ha una qualifica al più pari a *IV*. Il rimanente 50% circa ha una qualifica almeno pari a *IV*.
- All'incirca il 60% degli impiegati civili dello stato ha una qualifica al più pari a *IV*. Il rimanente 40% circa ha una qualifica almeno pari a *IV*.
- All'incirca il 70% degli impiegati civili dello stato ha una qualifica al più pari a *IV*. Il rimanente 30% circa ha una qualifica almeno pari a *IV*.
- All'incirca il 80% degli impiegati civili dello stato ha una qualifica al più pari a *V*. Il rimanente 20% circa ha una qualifica almeno pari a *V*.
- All'incirca il 90% degli impiegati civili dello stato ha una qualifica al più pari a *VI*. Il rimanente 10% circa ha una qualifica almeno pari a *VI*.

La posizione del 96° percentile è data da:

$$Pos(C_{96}) = 96 \cdot \frac{N+1}{100} = 96 \cdot 8185,4 = 785798,4$$

In base alle frequenze cumulate abbiamo che $C_{96} = VII$. Questo risultato ci dice che circa il 96% degli impiegati civili dello stato ha una qualifica al più pari a *VII*. Il rimanente 4% circa ha una qualifica almeno pari a *VII*.

7. La seguente tabella riporta la distribuzione del carattere X numero di stanze di 129 abitazioni di un quartiere milanese.

Numero di stanze	1	2	3	4	5	6	7	8	9
n_i	10	35	40	21	10	6	4	2	1

Calcolare la moda, i quartili, il primo e il nono decile, il 73-esimo centile della distribuzione. Commentare i risultati ottenuti.

Svolgimento

La distribuzione in considerazione riguarda un carattere quantitativo discreto rilevato su scala di rapporti. Al fine del calcolo dei percentili richiesti dall'esercizio, calcoliamo le frequenze cumulate.

Numero di stanze	n_i	Freq. Cumulate
1	10	10
2	35	45
3	40	85
4	21	106
5	10	116
6	6	122
7	4	126
8	2	128
9	1	129
tot	129	–

Dato che $N = 129$ è dispari, la posizione mediana è data da:

$$Pos(Me) = \frac{129 + 1}{2} = 65.$$

Basandoci sulle frequenze cumulate calcolate in precedenza possiamo concludere che $Me = 3$. Tale valore ci dice che circa il 50% delle abitazioni del quartiere milanese in considerazione, hanno un numero di stanze minore o uguale a 3. Analogamente, il valore della mediana ci dice che circa il 50% delle abitazioni del quartiere milanese in considerazione, hanno un numero di stanze maggiore o uguale a 3.

Le posizioni dei quartili sono date da:

$$Pos(Q_1) = \left(\frac{N + 1}{4} \right) = 32,5$$

$$Pos(Q_2) = 2 \cdot \left(\frac{N + 1}{4} \right) = 65$$

$$Pos(Q_3) = 3 \cdot \left(\frac{N + 1}{4} \right) = 97,5$$

Sulla base delle frequenze cumulate calcolate in precedenza possiamo concludere che:

$$Q_1 = 2; \quad Q_2 = Me = 3; \quad Q_3 = 4.$$

Tali valori ci dicono che:

- All'incirca il 25% delle abitazioni del quartiere milanese in considerazione ha un numero di stanze minore o uguale a 2. Il restante 75% circa ha un numero di stanze maggiore o uguale a 2.

- All'incirca il 50% delle abitazioni del quartiere milanese in considerazione ha un numero di stanze minore o uguale a 3. Il restante 50% circa ha un numero di stanze maggiore o uguale a 3.
- All'incirca il 75% delle abitazioni del quartiere milanese in considerazione ha un numero di stanze minore o uguale a 4. Il restante 25% circa ha un numero di stanze maggiore o uguale a 4.

Le posizioni del primo e nono decile sono date da:

$$Pos(D_1) = \frac{N+1}{10} = 13 \quad Pos(D_9) = 9 \cdot \frac{N+1}{10} = 117$$

Sulla base delle frequenze cumulate calcolate in precedenza possiamo concludere che $D_1 = 2$ e $D_9 = 6$. Tali risultati ci dicono che:

- All'incirca il 10% delle abitazioni del quartiere milanese in considerazione ha un numero di stanze minore o uguale a 2. Il restante 90% circa ha un numero di stanze maggiore o uguale a 2.
- All'incirca il 90% delle abitazioni del quartiere milanese in considerazione ha un numero di stanze minore o uguale a 6. Il restante 10% circa ha un numero di stanze maggiore o uguale a 6.

La posizione del 73-esimo centile è data da:

$$Pos(C_{73}) = 73 \cdot \frac{N+1}{100} = 73 \cdot 1,3 = 94,9$$

Sulla base delle frequenze cumulate abbiamo che $C_{73} = 4$. Tale valore ci dice che circa il 73% delle abitazioni del quartiere milanese in considerazione ha un numero di stanze minore o uguale a 4. Il restante 27% circa ha un numero di stanze maggiore o uguale a 4.

8. La seguente tabella riporta il numero di incidenti stradali subiti dai possessori di patente di un comune negli anni 1983 e 1984.

Numero incidenti	fino a 2	3-4	5-8	9-11	12-16	17-19	20-24	totale
Frequenze	380	305	285	84	67	24	5	1150

Calcolare la moda e commentare.

Svolgimento

La tabella fornita dal testo dell'esercizio riporta la distribuzione di frequenze di un carattere quantitativo discreto con dati raggruppati in classi. In tal caso il calcolo della moda si riduce all'individuazione della classe modale che si identifica in quella con frequenza specifica maggiore. Calcoliamo dunque le frequenze specifiche dopo

aver evidentemente fissato in 0 l'estremo inferiore della prima classe.

Num. Incidenti fino a 2	n_i	Ampiezza	Freq. Specifiche
3-4	380	3	126,67
5-8	305	2	152,5
9-11	285	4	71,25
12-16	84	3	28
17-19	67	5	13,4
20-24	24	3	8
	5	5	1

La Classe modale risulta quindi essere 3-4. Si può quindi affermare che il numero di incidenti che, più comunemente, subiscono i possessori di patente nel comune in considerazione è pari a 3 o a 4. Si osservi che la classe modale rappresenta solo il 26,5% circa dell'intera popolazione ed è di conseguenza scarsamente rappresentativa della stessa.

9. I dati nella tabella riportano la distribuzione di frequenza delle entrate provenienti dai redditi soggetti ad imposta di una certa categoria di lavoratori del Canada (dati espressi in migliaia di dollari \$):

entrate	0-2	2-4	4-6	6-8	8-10	10-15	15-20	20-50	50-100	100-200	totale
frequenze	250	1589	1768	1473	1172	1298	306	200	21	3	8080

Calcolare:

- i) la percentuale di entrate minori di 5000\$;
- ii) la percentuale delle entrate maggiori di 12000\$;
- iii) i quartili;
- iv) l'entrata di reddito che è superata solo dal 10% delle entrate.

Svolgimento

Al fine dello svolgimento dell'esercizio saranno utili i risultati riportati nella seguente tabella.

Entrate	Frequenze	Freq. Cumulate	Freq. Retrocum.	Freq. Spec.
0-2	250	250	8080	125
2-4	1589	1839	7830	794.5
4-6	1768	3607	6241	884
6-8	1473	5080	4473	736.5
8-10	1172	6252	3000	586
10-15	1298	7550	1828	259.6
15-20	306	7856	530	61.2
20-50	200	8056	224	6.6
50-100	21	8077	24	0.42
100-200	3	8080	3	0.03
totale	8080	-	-	

- i) Il numero di entrate minori di 5000\$ è dato dalla somma delle:

- entrate minori o uguali a 4000\$;
- entrate maggiori di 4000\$ e minori o di 5000\$.

Il valore del primo addendo ci viene fornito dalla seconda frequenza cumulata: 1839. Il valore del secondo addendo viene individuato ipotizzando che le frequenze si distribuiscano uniformemente all'interno di ogni classe. Con questa ipotesi la frequenza x associata alla sottoclasse 4-5 è sostanzialmente ottenuta ricorrendo alla frequenza specifica della classe 4-6. La frequenza specifica associata alla classe 4-6 è pari a 884 e ci indica la frequenza che corrisponde ad ogni intervallo unitario della classe 4-6 (e dunque anche all'intervallo 4-5). Il numero di entrate maggiori di 4000 e minori o di 5000 è dunque pari a 884 e di conseguenza il numero di entrate minori di 5000 è 2723. La percentuale di entrate minori di 5000\$ è quindi pari a

$$\frac{2723}{8080} \cdot 100 = 33,7\%.$$

ii) il numero di entrate maggiori di 12000\$ è dato dalla somma delle:

- entrate maggiori di 15000\$;
- entrate maggiori di 12000\$ e minori o uguali a 15000\$.

Il valore del primo addendo ci viene fornito dalla settima frequenza retrocumulata: 530. Il valore del secondo addendo viene individuato ipotizzando che le frequenze si distribuiscano uniformemente all'interno di ogni classe. Con questa ipotesi la frequenza x associata alla sottoclasse 12-15 è sostanzialmente ottenuta ricorrendo alla frequenza specifica della classe 10-15. tale frequenza specifica è data da 259,6 così che la frequenza associata alla sottoclasse (di ampiezza 3) è data da:

$$3 \cdot 259,6 = 778,8.$$

Il numero di entrate maggiori di 12000\$ è quindi pari a $530 + 778,8 = 1308,8$. La percentuale di entrate maggiori a 12000\$ è pari a

$$\frac{1308,8}{8080} \cdot 100 = 16,2\%.$$

iii) Le posizioni delle tre entrate quartili sono date da:

$$Pos(Q_1) = \frac{8081}{4} = 2020,25$$

$$Pos(Q_2) = 2 \cdot \frac{8081}{4} = 4040,50$$

$$Pos(Q_3) = 3 \cdot \frac{8081}{4} = 6060,75$$

Osservando le frequenze cumulate si può dunque stabilire che:

$$Q_1 = x_{(2020,25)} = 4 + (2020,25 - 1839 - 0,5) \cdot \frac{2}{1768} = 4,2045.$$

$$Q_2 = x_{(4040,5)} = 6 + (4040,5 - 3607 - 0,5) \cdot \frac{2}{1473} = 6,5879.$$

$$Q_3 = x_{(6060,75)} = 8 + (6060,75 - 5080 - 0,5) \cdot \frac{2}{1172} = 9,6728.$$

iv) L'entrata del reddito che è superata solo dal 10% delle entrate coincide con D_9 . La posizione del nono decile è data da:

$$Pos(D_9) = 9 \cdot \frac{8081}{10} = 7272,9.$$

$$D_9 = x_{(7272,9)} = 10 + (7272,9 - 6252 - 0,5) \frac{5}{1298} = 13,9366.$$