Esercizi di Statistica, con soluzioni

Considera 20 famiglie. Per ciascuna rileva il numero di componenti. Ecco i dati:

```
1 3 2 5 4 2 2 3 3 2 3 4 4 3 2 5 4 3 3 1
```

Costruisci la distribuzione di frequenza.

Soluzione

```
Componenti    1 2 3 4 5
Famiglie      2 5 7 4 2
```

Nell'esempio precedente dire

- qual è l'unità
- qual è la variabile
- La variabile è quantitativa o qualitativa?

Soluzione

L'unità è la famiglia, la variabile il numero di componenti, quantitativa discreta.

In un'ora una libreria fa 20 scontrini per i seguenti importi in Euro

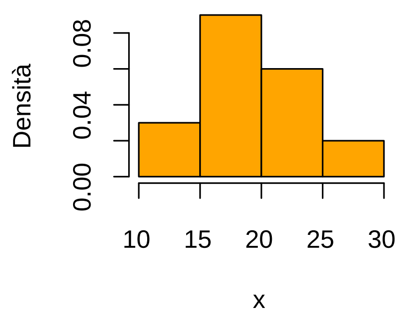
```
10 13 13 18 18 18 19 19 20 20 20 20 22 22 23 24 24 25 26 27
```

Fate un istogramma con classi

```
10 |- 15 , 15 |- 20, 20 |- 25, 25 |- 30
```

Soluzione

```
Importo    n. clienti
10 |- 15      3
15 |- 20      5
20 |- 25      9
25 |- 30      3
Totale        20
```



Ecco la distribuzione del salario mensile (in Euro) di un campione di 1000 lavoratori:

Classi di reddito	Frequenze	Altezze dei rettangoli
0 |- 500	100	0.2
500 |- 1000	200	0.4
1000 |- 2000	500	0.5
2000 |- 4000	600	0.3
4000 |- 8000	400	0.1

Disegnare l'istogramma e verificare che le aree dei rettangoli sono uguali alle frequenze. Come sono state calcolate le altezze dei rettangoli?

Soluzione

Le altezze sono = frequenze/ampiezza di classe.

Le temperature nella località X alle 12 sono state

`12.1 14.5 9.7 8.1 13.0 12.5 10.5`

Calcola la media e la mediana.

Soluzione

Media = 80.4/7 = 11.48 gradi centigradi.

Mediana = 12.1 gradi centigradi.

8.1 9.7 10.5 (12.1) 12.5 13.0 14.5

La distribuzione di 520 studenti per numero di esami superati è

Esami	Studenti
0	50
1	100
2	160
3	120
4	80
5	10

Calcolare le frequenze relative, la moda e il numero medio di esami superati. Calcolare la mediana.

Soluzione

Esami	Studenti	Freq. relative	Cumulate	Prodotto
0	50	0.10	0.10	0
1	100	0.19	0.29	100
2	160	0.31	0.60	320
3	120	0.23	0.83	360
4	80	0.15	0.98	320
5	10	0.02	1.00	50
Totale	520	1.00		1150

La moda è 2 esami superati. La media è $1150/520 = 2.21$ esami. La mediana si trova notando che le due unità centrali sono la 260 e la 261-ma nella successione ordinata e dalla distribuzione si vede che stanno entrambe nella classe di 2 esami.

Quindi la mediana è 2 esami superati.

La popolazione delle prime 10 città americane in milioni è la seguente.

New York (New York)	9.21
Los Angeles (California)	4.05
Chicago (Illinois)	2.83
Houston (Texas)	2.01
Phoenix (Arizona)	1.55
Filadelfia	1.45
Dallas (Texas)	1.31
San Diego (California)	1.30
San Antonio (Texas)	1.24
San Jose (California)	0.94

Calcolare la popolazione media e la popolazione mediana. Quale indice è migliore?

Soluzione

La popolazione totale è 25.80 milioni. Quindi la media è 2.58 milioni. La mediana è la semisomma tra 1.45 e 1.55 cioè 1.5 milioni di abitanti.

È meglio la mediana perché non risente troppo dei valori anomali (come New York).

Ecco il voto di laurea di 5 studenti di Lettere

```
110 109 108 110 110
```

Ecco il voto di laurea di 5 studenti di Economia

```
90 98 110 105 102
```

C'è maggiore variabilità di voto a Economia o a Lettere? Giustificare calcolando le varianze del voto e le deviazioni standard.

Soluzione

- A Lettere il voto medio di laurea è 109.4.
- A Economia il voto medio è 101.
- La varianza del voto a Lettere è 0.8
- La varianza del voto a Economia è 57. (Ho usato il denominatore $n = 4$.

Le deviazioni standard sono perciò

- Lettere punti 0.8944
- Economia punti 7.5498.

Evidentemente la variabilità è minore a Lettere. Volendo usare il *coefficiente di variazione* per ottenere una misura relativa di variabilità si ottiene

- Lettere CV = 0.0082
- Economia CV = 0.0748

Quindi il coefficiente di variazione è minore a Lettere.

Considera la distribuzione di 1000 studenti secondo il voto di laurea

```
Voto  Frequenza
98         10
99         40
100       250
101       400
102       250
103        40
104        10
```

- Mostrate che la media la moda e la mediana sono uguali a 101. Verificate che la deviazione standard è di 1 punto.
- Calcolate la frequenza relativa di studenti che hanno preso un voto compreso tra 101 - 2 = 99 e 101 + 2 = 103. Secondo la disuguaglianza di Chebyshev questa frequenza relativa quanto dovrebbe essere?

Soluzione

La distribuzione è simmetrica e quindi media moda e mediana sono uguali. La moda e la mediana sono evidentemente uguali a 101 punti. La media è 101000/1000 = 101.

Per calcolare la varianza compiliamo la tabella seguente

Voto	Freq.	Voto^2	Voto^2 * freq.
98	10	9604	96040
99	40	9801	392040
100	250	10000	2500000
101	400	10201	4080400
102	250	10404	2601000
103	40	10609	424360
104	10	10816	108160
Totale	1000		10202000

Quindi la media dei voti al quadrato è

$$MQ = 10202000/1000 = 10202$$

e la varianza è uguale alla MQ meno la media al quadrato :

$$MQ - (media^2) = 10202 - 101^2 = 1.$$

Quindi anche la deviazione standard è 1.

La proporzione di studenti che hanno preso un voto tra 99 e 103 è 980/1000 = 98%.

Come vedremo la disuguaglianza di Chebychev asserisce che per forza questa proporzione deve essere maggiore di 1 - 1/4 = 75%. E infatti così avviene.

Su 4 famiglie di 2 componenti misuriamo il reddito di Febbraio X e le relative spese per l'alimentazione Y.

```
X:     1500 1700  1400 1600
Y:      200  350   150  300
```

Calcolare la covarianza e il coefficiente di correlazione e interpretarli.

Soluzione

media(X) = 1550, media(Y) = 250, var(X) = 12500, var(Y) = 6250.

(X - media(X))(Y - media(Y)) : $2500, 15000, 15000, 2500$

cov(X, Y) = 35000/4 = 8750.

cor$(X, Y) = 8750/\sqrt{(12500)(6250)} = 0.9899$.

Provate a calcolare la covarianza con la formula equivalente

media(XY) - media(X) media(Y)

Soluzione

Poiché media$(XY) = (1500 \times 200 + 1700 \times 350 + 1400 \times 150 + 1600 \times 300)/4$ abbiamo

$\text{cov}(X,Y) = \text{media }(XY) - \text{ media }(X) \text{ media}(Y) = 396250 - (1550)(250) = 8750.$

Rappresentate la relazione tra spesa e reddito con la retta dei minimi quadrati e verificate che questa è: Spesa $= -835 + 0.7$ Reddito. Provate a interpretare il risultato.

Soluzione

Coefficiente angolare:

$$\text{cor}(X,Y)\sqrt{\text{var}(Y)/\text{var}(X)} = \text{cov}(X,Y)/\text{var}(X) = 8750/12500 = 0.7.$$

La retta deve passare per il punto (media(X), media(Y)) $= (1550, 250)$ e quindi ha equazione

$$y = 250 + 0.7(x - 1550) = -835 + 0.7x.$$

La pendenza ha l'interpretazione: per ogni Euro in più di reddito la spesa aumenta di 70 centesimi.

Vero o Falso? Se X e Y sono due variabili con $\text{var}(X) = 3.25$, $\text{var}(Y) = 5.8$, $\text{cov}(X,Y) = 14.703$

allora il coefficiente di correlazione è 0.78. Giustificare.

Soluzione

Falso. Perché $0.78 = \text{cov}(X,Y)/(\text{var}(X)\text{var}(Y))$ mentre il coefficiente di correlazione è

$$\text{cor}(X,Y) = \text{cov}(X,Y)/\sqrt{\text{var}(X)\text{var}(Y)}.$$