Esercizio 1

Il corso di Statistica è frequentato da 10 studenti che presentano le seguenti caratteristiche

Studente	Sesso	Colore Occhi	Voto	Soddisfazione	Età
Stefano	M	Nero	18	Per niente	21
Francesca	F	Marrone	24	Abbastanza	22
Maria	F	Azzurro	18	Poco	22
Andrea	M	Verde	30	Molto	23
Luigi	M	Azzurro	26	Abbastanza	24
Anna	F	Marrone	18	Per niente	25
Marta	F	Marrone	30	Molto	21
Lucia	F	Nero	26	Molto	20
Matteo	M	Nero	26	Abbastanza	21
Sara	F	Nero	26	Abbastanza	21

1. stabilire la tipologia dei caratteri;
2. costruire le distribuzioni di frequenza assolute, relative, percentuali e se possibile, le corrispondenti cumulate;
3. rappresentare mediante i grafici ritenuti più idonei le distribuzioni di frequenze assolute dei diversi caratteri;
4. qual è la percentuale di studenti con un'età inferiore a 23?
5. qual è la percentuale di studenti con un'età strettamente compresa tra 21 e 24?
6. qual è la percentuale di studenti con un'età superiore a 24?

Soluzione

1. Iniziamo a classificare i diversi caratteri: sesso e colore occhi sono caratteri qualitativi sconnessi; il carattere soddisfazione è qualitativo ordinato; voto è quantitativo discreto; età è un carattere quantitativo continuo.

2. A partire dalla tabella dei dati costruiamo le distribuzioni di frequenza assolute, percentuali e le corrispondenti cumulate (se è possibile):

Sesso	n_i	f_i	p_i
M	4	0.4	40%
F	6	0.6	60%
Totale	10	1	100%

Colore Occhi	n_i	f_i	p_i
Nero	4	0.4	40%
Marrone	3	0.3	30%
Azzurro	2	0.2	20%
Verde	1	0.1	10%
Totale	10	1	100%

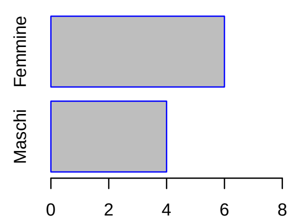
Voto	n_i	f_i	F_i	p_i	P_i
18	3	0.3	0.3	30%	30%
24	1	0.1	0.4	10%	40%
26	4	0.4	0.8	40%	80%
30	2	0.2	1	20%	100%
Totale	10	1	-	100%	-

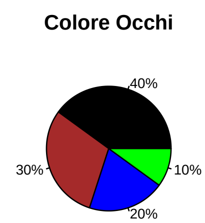
Soddisfazione	n_i	f_i	F_i	p_i	P_i
Per niente	2	0.2	0.2	20%	20%
Poco	1	0.1	0.3	10%	30%
Abbastanza	4	0.4	0.7	40%	70%
Molto	3	0.3	1	30%	100%
Totale	10	1	-	100%	-

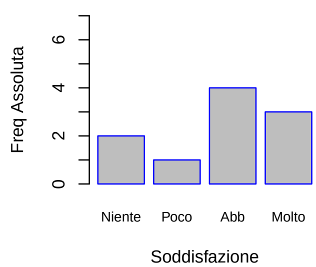
Per una lettura semplice dei dati scegliamo tre classi.

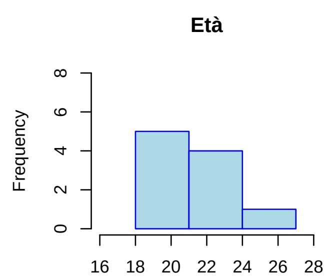
Età	n_i	f_i	F_i	p_i	P_i
$18 \dashv 21$	5	0.5	0.5	50%	50%
$21 \dashv 24$	4	0.4	0.9	40%	90%
$24 \dashv 27$	1	0.1	1	10%	100%
Totale	10	1	-	100%	-

3. Per i caratteri qualitativi non ordinati generalmente si utilizza un grafico a nastri o a torta. Un grafico a barre viene utilizzato per i caratteri qualitativi ordinati o quantitativi discreti. Per i caratteri quantitativi continui si utilizza un istogramma, avendo cura di suddividere il carattere continuo in classi.









4. $Freq(\text{Età}< 23) = \dfrac{7}{10} = 0.7$

5. $Freq(21 \leqslant \text{Età} \leqslant 24) = \dfrac{9}{10} = 0.9$

6. $Freq(\text{Età} > 24) = \frac{1}{10} = 0.1$

Esercizio 2

120 imprese vengono classificate in base al fatturato. La rappresentazione mediante istogramma produce le seguenti densità di frequenza

Fatturato	h_j
0-10	0.0333
10-15	0.0444
15-30	0.0222
30-60	0.0037

1. ricostruire la distribuzione di frequenze assolute e relative.

Soluzioni

Ricordando le seguenti relazioni è possibile ricostruire la distribuzione di frequenze assolute e relative:

$$h_i = \frac{f_i}{a_i} \Rightarrow f_i = h_i \cdot a_i \quad f_i = \frac{n_i}{N} \Rightarrow n_i = N \cdot f_i$$

Fatturato	h_j	a_i	f_i	n_i
0-10	0.0333	10	0.33	40
10-15	0.0444	5	0.22	27
15-30	0.0222	15	0.33	40
30-60	0.0037	30	0.11	13

Ricorda che la densità di frequenza può essere calcolata usando le frequenze assolute, le relative o le percentuali. Qui usiamo le frequenze relative, ma il risultato grafico non cambia: l'altezza relativa dei rettangoli rimane la stessa, cambia solo la scala dell'asse verticale.

Esercizio 3

Data la seguente distribuzione di frequenze del carattere titolo di studio

Titolo di Studio	Frequenze
Lic. Elementare	12
Lic. Media	36
Diploma	36
Laurea Triennale	24
Laurea Specialistica	12
Totale	120

1. determinare i quartili.

Soluzione

Per determinare i quartili, dobbiamo calcolare la distribuzione delle frequenze relative e relative cumulate. I quartili sono dati da

Titolo di Studio	Frequenze (n_i)	N_i	f_i	F_i
Lic. Elementare	12	12	0.1	0.1
Lic. Media	36	48	0.3	0.4
Diploma	36	84	0.3	0.7
Laurea Triennale	24	108	0.2	0.9
Laurea Specialistica	12	120	0.1	1
Totale	120	1		

- $Q_1 = Lic.Media$ dato che $F_1 < 0.25 < F_2$
- $Q_2 = Diploma$ dato che $F_2 < 0.5 < F_3$
- $Q_3 = LaureaTriennale$ dato che $F_3 < 0.75 < F_4$

Esercizio 4

Data la seguente distribuzione in classi di età

Classi di età	Frequenze
18-⊣25	15
25-⊣28	69
28-⊣30	54
30-⊣35	14
Totale	152

1. calcolare la classe modale, la classe mediana, il valore mediano e la media aritmetica;

2. calcolare l'intervallo di variazione, la differenza interquartile, lo scostamento quadratico medio e il coefficiente di variazione.

Soluzione

1. Per risolvere i quesiti proposti, prima di tutto occorre ricavarsi le distribuzioni delle frequenze relative e relative cumulate

Classi di età	Frequenze (n_i)	f_i	F_i	Ampiezza (a_i)	Densità (h_i)	Val Cen (c_i)	$c_i \cdot f_i$
18-⊣25	15	0.1	0.1	7	0.01	21.5	2.12
25-⊣28	69	0.45	0.55	3	0.15	26.5	12.03
28-⊣30	54	0.36	0.91	2	0.18	29	10.30
30-⊣35	14	0.09	1	5	0.02	32.5	2.99
Totale	152	1					27.45

La classe modale corrisponde alla classe $(28 \dashv 30)$ alla quale è associata la densità di frequenza massima; la classe mediana è la classe $(25 \dashv 28)$ nella quale sono presenti le unità che si trovano nella posizione 76° e 77°, (152/2)=76 e (152/2)+1=77, rispettivamente. Il valore mediano si ottiene per interpolazione ricorrendo alla formula approssimata

$$Q_2 = Me \approx 25 + \frac{(0.5-0.1)}{(0.55-0.1)}3 = 27.67.$$

Infine la media aritmetica è pari a circa 27.45 anni.

2. Osserviamo che sia il primo che il secondo quartile cadono nella classe $25 \dashv 28$, dato che $F_1 < 0.25 < F_2$ e $F_1 < 0.5 < F_2$.

3. il terzo quartile cade nella classe $28 \dashv 30$ dato che $F_2 < 0.75 < F_3$.
   Assumendo l'uniformità della distribuzione del carattere all'interno delle classi possiamo determinare delle approssimazione per valori dell primo, dell terzo quartile e della mediana, (già calcolato al punto punto precedente). Tali valori saranno dati da:

   - $Q_1 \approx 25 + \frac{(0.25-0.1)}{(0.55-0.1)}3 = 26$
   - $Q_2 = Me \approx 25 + \frac{(0.5-0.1)}{(0.55-0.1)}3 = 27.67$
   - $Q_3 \approx 28 + \frac{(0.75-0.55)}{(0.9-0.55)}2 = 29.14$

   La differenza interquantile è data da $Q_3 - Q1 \approx 29.14 - 26 = 3.14$. La differenza interquartile rappresenta il campo di variazione per il 50% delle unità centrali.

Classi di età	Frequenze (n_i)	f_i	Val Cen (c_i)	$c_i \cdot f_i$	$c_i^2 * f_i$
18-⊣25	15	0.1	21.5	2.12	45.62
25-⊣28	69	0.45	26.5	12.03	318.78
28-⊣30	54	0.36	29	10.30	298.78
30-⊣35	14	0.09	32.5	2.99	97.29
Totale	152	1		27.45	760.46

Lo scarto quadratico medio può essere calcolato come segue

$$\sigma \approx \sqrt{\sum_{i=1}^{n} c_i^2 f_i - \bar{x}^2} = \sqrt{760.46 - (27.45)^2} = \sqrt{6.96} = 2.64.$$

Il coefficiente di variazione sarà pari a

$$CV = \frac{\sigma}{\bar{x}}100 = \frac{2.64}{27.45}100 = 9.62.$$

Esercizio 5

Immaginiamo di aver osservato il numero di spot pubblicitari trasmessi tra le 20 e le 21 da alcune reti televisive

Rete TV	Numero di Spot
Rai 1	10
Rai 2	10
Rai 3	10
Canale 5	20
Italia 1	30
La 7	20

1. calcolare sia il numero medio di spot pubblicitari che la mediana.

Soluzione

In questo caso abbiamo una distribuzione unitaria e non una distribuzione di frequenze, quindi per determinare la mediana occorre ordinare in senso crescente i dati (10 10 10 20 20 30) e da questo collettivo ordinato calcolare f_i e F_i. La media

Modalità	n_i	f_i	F_i
10	3	0.5	0.5
20	2	0.33	0.83
30	1	0.17	1
Totale	6	1	-

è data da $\bar{x} = (10 * 3 + 20 * 2 + 30 * 1)/6 = 100/6 = 16.67$.
Per il calcolo della mediana, notiamo che si ha un numero pari di osservazioni, quindi le posizioni mediane sono date da $N/2 = 3$ e $N/2 + 1 = 4$, e la mediana sarà la media della terza e della quarta osservazione, ossia $(10 + 20)/2 = 15$.
Inolre possiamo notare che in corrispondenza della prima modalità si ha una frequenza cumulata $F_1 = 0.5$. Quindi essendo nel caso $F_1 = 0.5 < F_2$ e visto che il carattere che stiamo analizzando è quantitativo possiamo scegliere come mediana la media aritmetica delle modalità corrispondendi alle frequenze relative cumulate $F_1 = 0.5$ e $F_2 = 0.83$. Quindi avremo che $Me \approx (10 + 20)/2 = 15$.

Esercizio 6

Dato il seguente collettivo di valori

$$4, 9, 5, 10, 2$$

1. calcolare la media aritmetica e la mediana;
2. verificare le prime tre proprietà della media aritmetica;
3. verificare la proprietà della mediana;
4. determinare la devianza, varianza e scarto quadratico medio.

Soluzione

1. la media aritmetica sarà pari a

   $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}(4+9+5+10+2) = 6.$$

   L'esercizio fornisce una distribuzione unitaria non ordinata, quindi il primo passo per poter calcolare la mediana è ordinare la distribuzione,

   $$2,4,5,9,10. \Rightarrow Me = 5.$$

2. Verifichiamo le prime tre proprietà della media aritmetica

   (a) $\sum_{i=1}^{n} x_i = n\bar{x} \Rightarrow \quad n\bar{x} = 6 \cdot 5 = 30 \quad \sum_{i=1}^{n} x_i = 4+9+5+10+2 = 30$

   (b) $\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \Rightarrow$
   $(4-6)+(9-6)+(5-6)+(10-6)+(2-6) = -2+3-1+4-4 = 0$

   (c) $\sum_{i=1}^{n}(x_i - c)^2$ è minimo per $c = \bar{x} \Rightarrow (4-6)^2 + (9-6)^2 + (5-6)^2 + (10-6)^2 + (2-6)^2 = 4+9+1+16+16 = 46.$
   Calcoliamo gli scarti quadratici dalla mediana
   $(4-5)^2 + (9-5)^2 + (5-5)^2 + (10-5)^2 + (2-5)^2 = 1+16+0+25+9 = 51.$
   Notiamo che $46 < 51$ e questo è vero per ogni $c \neq \bar{x}$.

3. Verifichiamo la proprietà della mediana

    $\sum_{i=1}^{n} \mid x_i - c \mid$ è minimo per $c = Me \Rightarrow$
    calcoliamo gli scarti in valore assoluto dalla mediana
    $\mid 4-5 \mid + \mid 9-5 \mid + \mid 5-5 \mid + \mid 10-5v+ \mid 2-5 \mid = 1+4+0+5+3 = 13$;
    calcoliamo gli scarti in valore assoluto dalla media aritmetica $\bar{x} = 6$
    $\mid 4-6 \mid + \mid 9-6 \mid + \mid 5-6 \mid + \mid 10-6 \mid + \mid 2-6 \mid = 2+3+1+4+4 = 14$.
    Notiamo che $13 < 14$ e questo è vero per ogni $c \neq Me$.

4. per determinare la devianza, la varianza e lo scarto quadratico medio, utilizziamo la seguente tabella

x_i	$(x_i - \bar{x})^2$
4	4
9	9
5	1
10	16
2	16
Totale	46

Quindi

$$devianza = \sum_{i=1}^{n}(x_i - \bar{x})^2 = 46 \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 9.2 \quad \sigma = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)} = 3.03$$

Calcolo semplificato, ma **NON** la **definizione**, della varianza:
$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{226}{5} - 6^2 = 45.2 - 36 = 9.2$;
Calcolo semplificato, ma **NON** la **definizione**, della devianza:
$devianza = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = 226 - 5 \cdot (6^2) = 226 - 5 \cdot 36 = 46$.
I risultati si ottengono facilmente utilizzando la tabella seguente.

x_i	x_i^2
4	16
9	81
5	25
10	100
2	4
Totale	226

Esercizio 7

Le seguenti distribuzioni sono state rilevate

Peso Madri (in kg)
60.0, 70.8, 61.8, 63.7, 60.5, 62.9, 68.1, 57.5, 62.4, 65.8
Peso Neonati (in kg)
3.2, 3.5, 3.2, 2.7, 3.8, 4.0, 3.1, 2.4, 3.4, 3.0
Altezza Neonati (in cm)
49.5, 46.4, 50.2, 49.6, 52.1, 47.8, 51.3, 49.0, 48.6, 47.1

Sapendo che:

Peso Neonati: media=3.23 kg, sd=0.45 kg;
Altezza Neonati: media=49.16 cm, sd=1.69 cm.

1. calcolare la media e la deviazione standard del Peso delle madri;
2. calcolare il coefficiente di variazione delle tre distribuzioni;
3. nei neonati, è più variabile la distribuzione del peso o dell'altezza?
4. la distribuzione del peso è più variabile nei neonati o nelle madri?

Soluzione

1. Utilizziamo la seguente tabella per calcolare la media e la deviazione standard del peso delle madri.

x_i	x_i^2
60.0	3600.00
70.8	5012.64
61.8	3819.24
63.7	4057.69
60.5	3660.25
62.9	3956.41
68.1	4637.61
57.5	3306.25
62.4	3893.76
65.8	4329.64
$\sum_i x_i = 633.5$	$\sum_i x_i^2 = 40273.49$

$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} = \frac{633.5}{10} = 63.35$ e $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2} = \sqrt{14.13} = 3.76$.

2. I coefficienti di variazione saranno pari a:

$$\textbf{Peso Madri}: CV = \frac{3.76}{63.35}100 = 5.93$$
$$\textbf{Peso Neonati}: CV = \frac{0.45}{3.23}100 = 14.05$$
$$\textbf{Altezza Neonati}: CV = \frac{1.69}{49.16}100 = 3.44$$

3. I neonati sono più variabili rispetto al peso che all'altezza.

4. Il peso dei neonati è più variabile rispetto a quello delle madri (circa il doppio).

Notiamo che dai valori delle deviazioni standard si potrebbe concludere che la distribuzione relativa al peso delle madri sia più dispersa rispetto a quella del peso dei neonati. Tuttavia questa affermazione non tiene conto delle possibili differenze di dimensione tra i valori delle due distribuzioni. Per un corretto confronto dobbiamo determinare i corrispondenti coefficienti di variazione.

Esercizio 8

La temperatura a Roma durante la scorsa settimana ha avuto una media di $25°C$ e varianza $4°C^2$, sapendo che la relazione tra gradi Celsius e Faharenheit è la seguente

$$F = \frac{9}{5}C + 32$$

determinare la media e la varianza della temperatura espressa in Fahrenheit.

Soluzione

Ricordando le proprietà della media e della varianza per trasformazioni lineari, sappiamo che conoscendo la media $\bar{x}$ e la varianza σ_x^2, la media e la varianza di $y = a + bx$ saranno date rispettivamente da

$$\bar{y} = a + b\bar{x}, \quad \sigma_y^2 = b^2\sigma_x^2.$$

Nell'esercizio la trasformazione lineare è data da $y = \frac{9}{5}x + 32$; quindi la media e la varianza di y sono pari a

$$\bar{y} = \frac{9}{5}25 + 32 = 77, \quad \sigma_y^2 = \left(\frac{9}{5}\right)^2 \sigma_x^2 = \frac{81}{25}4 = 12.96,$$

rispettivamente.