Corso di Statistica - Esercitazione 1

Dott. Davide Buttarazzi

✉ *d.buttarazzi@unicas.it*

Esercizio 1

La seguente tabella riporta dati riguardanti la produzione di tastiere per computer risultate difettose.

Causa	**n**$_i$
Punto nero	413
Danno	1039
Jitting	258
Deformazione	834
Graffi	442
Muffa	275
Segno argentato	413
Foro	371
Alone	292
Posizione tasti	1987
Tot.	**6324**

1. Definire la tipologia della variabile *Causa*
2. Calcolare moda, media, quartili ed indice di eterogeneità di Gini della variabile *Causa*

Soluzioni esercizio 1

1. *Causa* è un carattere qualitativo nominale
2. Moda: $x_i : n_i = max = Posizione\ tasti$
   Media: non calcolabile
   Quartili: non calcolabili
   Indice di eterogeneità di Gini: $E = 1 - \sum_{i=1}^{10} f_i^2 = 0.8343$, $\quad 0 \leq E \leq (\frac{K-1}{K} = 0.9)$

Esercizio 2

La seguente tabella riporta dati relativi al giudizio espresso da alcuni clienti sulla qualità dell'ultimo modello di smartphone prodotto da una nota azienda.

Valutazione	**n**$_i$
Inaccettabile	250
Scarsa	500
Accettabile	1500
Buona	2100
Ottima	350
Tot.	**4700**

1. Definire la tipologia della variabile *Valutazione*
2. Calcolare moda, media e quartili della variabile *Valutazione*
3. Calcolare l'indice di dispersione normalizzato della variabile *Valutazione*

4. Rappresentare graficamente la variabile *Valutazione*

Soluzioni esercizio 2

1. *Valutazione* è una variabile di tipo qualitativo ordinale

   Per ottenere le soluzioni ai quesiti 2, 3 e 4 opportuno far riferimento alla seguente tabella:

Valutazione	**n**$_i$	**f**$_i$	**F**$_i$	**RF**$_i$
Inaccettabile	250	0.05	0.05	1
Scarsa	500	0.11	0.16	0.95
Accettabile	1500	0.32	0.48	0.84
Buona	2100	0.45	0.92	0.52
Ottima	350	0.07	1	0.07
Tot.	**4700**	**1**		

2. Moda: $x_i : n_i = max = Buona$
   Media: non calcolabile
   Quartili:

   $Q1 = x_{(\frac{n}{4})} = x_i : F_i \geq 0.25 = Accettabile$

   $Q2 = Me = x_{(\frac{n+1}{2})} = x_i : F_i \geq 0.5 = Buona$

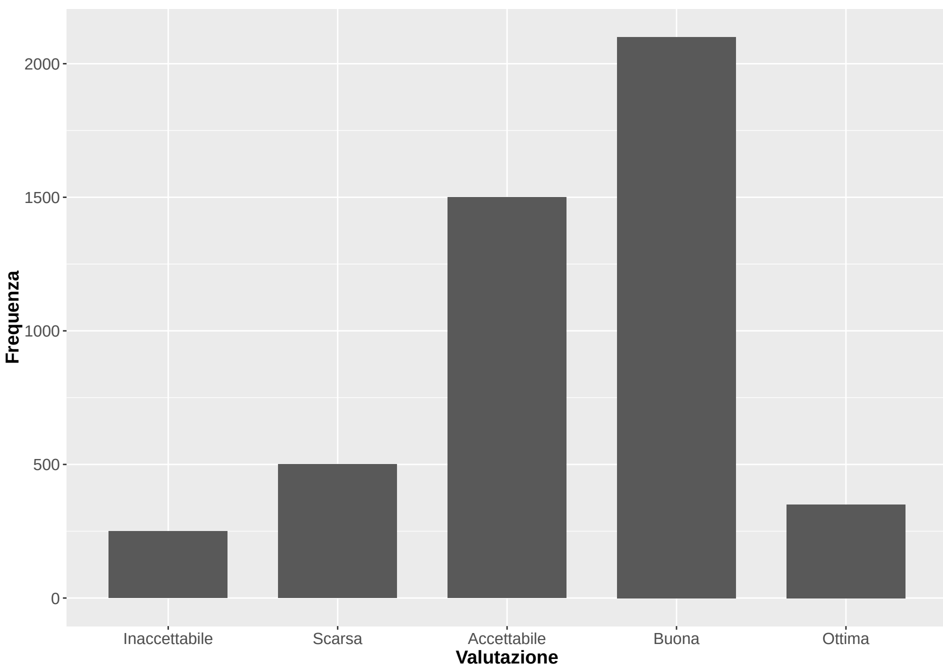
   $Q3 = x_{(\frac{3n}{4})} = x_i : F_i \geq 0.75 = Buona$

3. Indice di dispersione normalizzato:

   $D = \sum_{i=1}^{5} [F_i(1-F_i) + RF_i(1-RF_i)] = 2\sum_{i=1}^{4} F_i(1-F_i) = 1 \quad , \qquad 0 \leq D \leq 2$

   $D^* = \frac{2}{K-1} D = 0.5 \quad , \qquad 0 \leq D^* \leq 1$

4. La variabile *Valutazione* può essere rappresentata graficamente tramite grafico a barre



Esercizio 3

La seguente tabella riporta il prezzo di mercato di un campione di smartphone considerati da una nota azienda come principali competitor.

Prezzo	200	99	180	450	20	130	100	100	100	360	150	130	200	50	100	195	140	140

1. Definire la tipologia della variabile *Prezzo*
2. Calcolare moda, media e quartili della variabile *Prezzo*
3. Calcolare le seguenti misure di variabilità della variabile *Prezzo*:
    - differenza semplice media (senza ripetizione)
    - scostamento semplice medio
    - scostamento quadratico medio
    - varianza
    - devianza
    - coefficiente di variazione
    - campo di variazione interquartile
    - campo di variazione assoluto

Soluzioni esercizio 3

1. *Prezzo* è una variabile di tipo quntitativo discreta.
2. Moda: $x_i : n_i = max = 100$

   Media: $\mu = \sum_{i=1}^{N} \frac{x_i}{N} = \sum_{i=1}^{18} \frac{Prezzo_i}{18} = 158$

   Quartili:

   Per calcolare i quartili occorre riorganizzare la serie grezza in ordine non-decrescente:

   20 50 99 100 100 100 100 130 130 140 140 150 180 195 200 200 360 450

   $Q1 = x_{(\frac{n}{4})} = x_i : F_i \geq 0.25 = 100$

   $Q2 = Me = x_{(\frac{n+1}{2})} = x_i : F_i \geq 0.5 = 140$

   $Q3 = x_{(\frac{3n}{4})} = x_i : F_i \geq 0.75 = 195$

3. - Differenza semplice media:

     Per calcolare l'indice $\Delta_{SR} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} |x_i - x_j|}{n(n-1)}$ è opportuno costruire la seguente matrice di distanze in valore assoluto:

	20	50	99	100	100	100	100	130	130	140	140	150	180	195	200	200	360	450
20	0	30	79	80	80	80	80	110	110	120	120	130	160	175	180	180	340	430
50		0	49	50	50	50	50	80	80	90	90	100	130	145	150	150	310	400
99			0	1	1	1	1	31	31	41	41	51	81	96	101	101	261	351
100				0	0	0	0	30	30	40	40	50	80	95	100	100	260	350
100					0	0	0	30	30	40	40	50	80	95	100	100	260	350
100						0	0	30	30	40	40	50	80	95	100	100	260	350
100							0	30	30	40	40	50	80	95	100	100	260	350
130								0	0	10	10	20	50	65	70	70	230	320
130									0	10	10	20	50	65	70	70	230	320
140										0	0	10	40	55	60	60	220	310
140											0	10	40	55	60	60	220	310
150												0	30	45	50	50	210	300
180													0	15	20	20	180	270
195														0	5	5	165	255
200															0	0	160	250
200																0	160	250
360																	0	90
450																		0

Il numeratore di Δ_{SR} corrisponde alla somma di tutti gli elementi della matrice (SR=senza ripetizione, quindi vengono esclusi gli elementi presenti sulla diagonale principale). Poiché la matrice è simmetrica, ciò corrisponde a 2 volte la somma degli elementi della parte superiore (o inferiore) della matrice.

Quindi
$\Delta_{SR} = \frac{2(30+79+80+...+160+250+90)}{18(18-1)} = \frac{32156}{306} = 105.08$
L'indice puó essere normalizzato (ovvero $0 \leq \Delta_{SR} \leq 1$) dividendo per 2μ

- scostamento semplice medio
  Per calcolare gli indici di variabilità rispetto alla media ($\mu = 158$), é opportuno costruire la seguente tabella:

x_i	$\lvert x_i - \mu \rvert$	$(x_i - \mu)^2$
200	42	1764
99	59	3481
180	22	484
450	292	85264
20	138	19044
130	28	784
100	58	3364
100	58	3364
100	58	3364
360	202	40804
150	8	64
130	28	784
200	42	1764
50	108	11664
100	58	3364
195	37	1369
140	18	324
140	18	324
Totale	1274	181374

  Quindi si avrà che lo scostamento semplice medio sarà:

  $S_\mu^1 = \frac{\sum_{i=1}^{n} |x_i - \mu|}{n} = \frac{1274}{18} = 70.7$
- scostamento quadratico medio:
  $S_\mu^2 = \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}} = \frac{181374}{18} = \sqrt{10076} = 100.3$
- varianza:
  $\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} = \frac{181374}{18} = 10076$
- devianza:
  $dev(x) = \sum_{i=1}^{n} (x_i - \mu)^2 = 181374$
- coefficiente di variazione:
  $\frac{\sigma}{\mu} = \frac{100.3}{158} = 0.63$
- campo di variazione interquartile:
  $q_3 - q_1 = 195 - 100 = 95$
- campo di variazione assoluto:
  $max - min = 450 - 20 = 430$

Esercizio 4

Dopo aver organizzato in 5 classi di uguale ampiezza i dati in serie grezza dell'esercizio 3:

1. calcolare la mediana della variabile *Prezzo* tramite il metodo dell'interpolazione lineare utilizzando le frequenze relative (e relative cumulate)
2. calcolare la media
3. rappresentare graficamente i dati

Soluzioni esercizio 4

1. L'ampiezza delle classi (uguale per tutte le classi) può essere così calcolata:

   $\Delta = \frac{max-min}{K} = \frac{430}{5} = 86$

   Quindi la serie grezza può essere riorganizzata nella seguenta tabella per classi:

Prezzo	$\mathbf{n}_i$	$\mathbf{f_i}$	$\mathbf{F_i}$
20 $\vert-$ 106	7	0.39	0.39
106 $\vert-$ 192	6	0.33	0.72
192 $\vert-$ 278	3	0.16	0.88
278 $\vert-$ 364	1	0.06	0.94
364 $\vert-\vert$ 450	1	0.06	1

   $q_x = x_{i-1} + \frac{x_i - x_{i-1}}{F_i - F_{i-1}}[F(q_x) - F_{i-1}] = 106 + \frac{86}{0.33}(0.5 - 0.39) = 135$

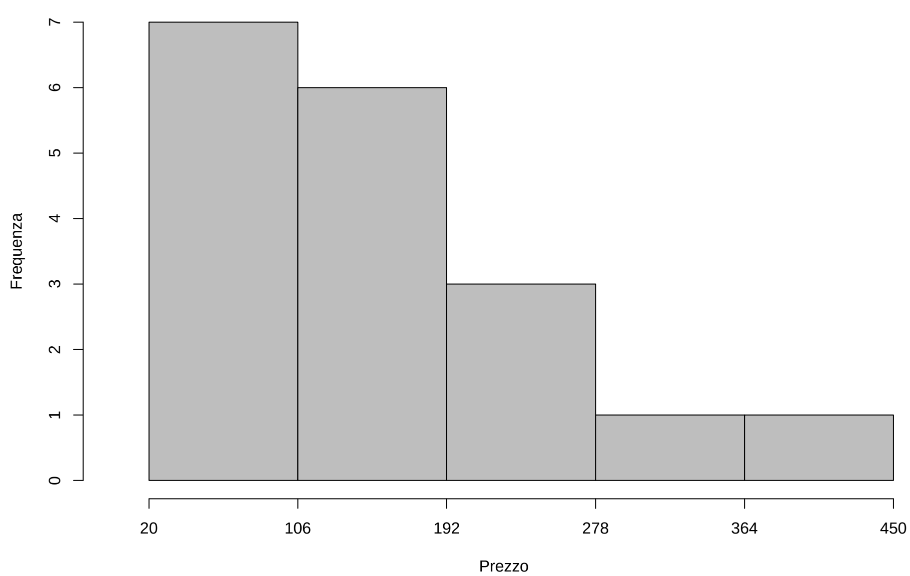
2. Per calcolare la media occorre individuare i centri delle classi utilizzando la formula $\hat{x}_i^c = \frac{x_{i-1}+x_i}{2}$
   Si avrà quindi:

Prezzo	$\mathbf{n}_i$	$\mathbf{\hat{x}_i^c}$
20 $\vert-$ 106	7	63
106 $\vert-$ 192	6	149
192 $\vert-$ 278	3	235
278 $\vert-$ 364	1	321
364 $\vert-\vert$ 450	1	397

   La media sarà quindi:

   $\mu = \sum_{i=1}^{K} \frac{\hat{x}_i^c n_i}{N} = \frac{7\times63+6\times149+3\times235+1\times321+1\times397}{18} = \frac{2734}{18} = 151.8$

3. La variabile *Prezzo* può essere rappresentata graficamente tramite istogramma



   In questo caso, poiché le classi hanno uguale ampiezza, ponendo sull'asse delle ordinate la densità si ottiene lo stesso risultato grafico.