Simulazione d'esame

Statistica I - Laurea Triennale in Marketing, Comunicazione Aziendale e Mercati Globali - Università degli Studi di Milano - Bicocca

NB: Nelle soluzioni non basta scrivere i risultati finali: riportare i passaggi principali e fornire adeguate spiegazioni. Scrivere le soluzioni a penna e con calligrafia leggibile.

Esercizio 1

Un team di ricercatori impiegato presso l'Università Bicocca ha raccolto i seguenti dati sul tempo (in minuti) necessario per completare un puzzle da 300 pezzi. La distribuzione in classi dei risultati è la seguente:

Tempo (min)	$5 \dashv 10$	$10 \dashv 15$	$15 \dashv 20$	$20 \dashv 30$	$30 \dashv 45$
Frequenza assoluta	7	18	22	10	3

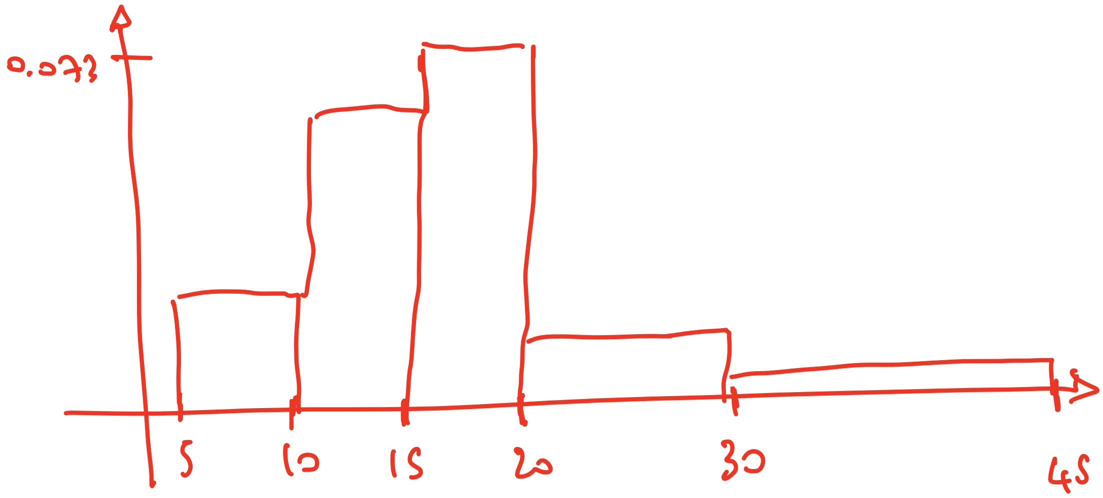
Utilizzando i dati precedenti, si risponda alle seguenti domande.

1. Dopo aver identificato la tipologia del carattere in esame, si produca una opportuna rappresentazione grafica che ne mostri l'andamento.
2. Determinare la classe modale e la classe in cui cade la mediana. Senza fare ulteriori calcoli, si risponda alla seguente domanda: è possibile che il nono decile della distribuzione sia pari a 29?
3. Si ricavi e si commenti una opportuna stima delle seguenti statistiche di sintesi: media aritmetica, mediana e terzo quartile.
4. Si dimostri empiricamente la proprietà di internalità della media aritmetica.
5. Volendo valutare quanto sono dispersi, si calcoli lo scarto quadratico medio, interpretando il valore ottenuto.

(1) IL CARATTERE È QUANTITATIVO CONTINUO, CON OSSERVAZIONI SUDDIVISE IN CLASSI.

$e_j \dashv u_j$	n_j	f_j	a_j	$d_j = f_j / a_j$	F_j	Q_j	x_j^c
$5 \dashv 10$	7	0.117	5	0.023	0.117	1.000	7.5
$10 \dashv 15$	18	0.300	5	0.060	0.417	0.883	12.5
$15 \dashv 20$	22	0.366	5	0.073	0.783	0.583	17.5
$20 \dashv 30$	10	0.167	10	0.017	0.950	0.217	25
$30 \dashv 45$	3	0.050	15	0.003	1.000	0.050	37.5
	60						

$$x_j^c = \frac{e_j + u_j}{2}$$



(2) LA CLASSE MODALE È $\{15 \dashv 20\}$

LA CLASSE IN CUI CADE LA MEDIANA
È $\{15 \dashv 20\}$

IL 9° DECILE È QUEL VALORE PER CUI $F_j \geq 0.9$, $Q_j \geq 0.1$. ESSENDO IL 9° DECILE NELLA CLASSE $\{20 \dashv 30\}$, È POSSIBILE CHE SIA UGUALE A 29.

(3) $$M_1(T) = \frac{1}{N} \sum_{j=1}^{K} n_j \times x_j^c$$

$$= \frac{1}{60} \left[7 \times 7.5 + 18 \times 12.5 + 22 \times 17.5 + 10 \times 25 + 3 \times 37.5 \right]$$

$$= \frac{1}{60} [1025] = 17.083$$

$$M_e(T) = l_j + (0.5 - F_{j-1}) \cdot \frac{1}{d_j} =$$

$$= 15 + (0.5 - 0.417) \cdot \frac{1}{0.073} = 16.137$$

$$q_{0.75} = l_i + (0.75 - F_{i-1}) \cdot \frac{1}{d_i}$$

$$= 15 + (0.75 - 0.417)\frac{1}{0.073} = 19.562$$

IN MEDIA, IL TEMPO DI RISOLUZIONE È PARI A 17.083, CON MEDIANA 16.137, CON PIÙ TEMPI BASSI CHE ALTI. IL 75% DEI TEMPI OSSERVATI È MINORE O UGUALE A 19.562.

(4) LA MEDIA È COMPRESA TRA IL VALORE MINIMO E IL MASSIMO OSSERVATI.

$$\min(T) = 5, \quad \max(T) = 45$$

$$M_1(T) = 17.083,$$

È COMPRESA TRA I DUE VALORI PRECEDENTI.

(5) $\sigma_T = \sqrt{\sigma_T^2}$

$$\sigma_T^2 = M_1(T^2) - (M_1(T))^2$$

$$M_1(T^2) = \frac{1}{N}\sum_{j=1}^{k} n_j \cdot (x_j^c)^2$$

$$= \frac{1}{60}\left[7 \times 7.5^2 + 18 \times 12.5^2 + 22 \times 17.5^2 + 10 \times 25^2 + 3 \times 37.5^2\right]$$

$$= \frac{1}{60}(20412.5) = 340.208$$

$$\sigma_T^2 = 340.208 - 17.083^2 = 43.379$$

$$\sigma_T = \sqrt{\sigma_T^2} = 6.956$$

I TEMPI IN MEDIA SONO 17.083, CONCENTRATI IN $\pm$ 6.956 INTORNO ALLA MEDIA ARITMETICA.

Esercizio 2

Un reparto di produzione automotive sta monitorando i tempi (in minuti) impiegati da dieci robot di saldatura per completare una singola giunzione in una fase di collaudo. L'obiettivo è analizzare la variabilità e le caratteristiche principali di questi tempi di ciclo per migliorare l'efficienza del processo. I dati di partenza sono i seguenti:

$$12, 15, 14, 17, 20, 22, 19, 18, 16, 210$$

Utilizzando i dati precedenti, si risponda alle seguenti domande.

1. Dopo aver calcolato il range e la distanza interquartile, si disegni in maniera approssimata la distribuzione dei dati utilizzando un *boxplot*. Quale fra i due indicatori vi sembra il più opportuno per riassumere la distribuzione dei dati in presenza di osservazioni anomale?

2. Si dimostri empiricamente che lo scostamento medio dalla mediana è inferiore allo scostamento medio dalla media aritmetica.

Solo dopo aver completato i calcoli, venite avvisati che l'ultimo valore era errato: non 210, ma 21. Dopo aver corretto i dati, si risponda alle seguenti domande.

3. Si calcolino la varianza ed il coefficiente di variazione per i tempi di saldatura corretti.

4. Usando le proprietà di trasformazioni lineari della media aritmetica e la varianza, si ricavi la media aritmetica e la varianza dei tempi di produzione misurati in secondi. Si calcoli, inoltre, il coefficiente di variazione, confrontandolo con il coefficiente ottenuto al tempo precedente.

(1) Range = $\max(T) - \min(T) = 210 - 12 = 198$

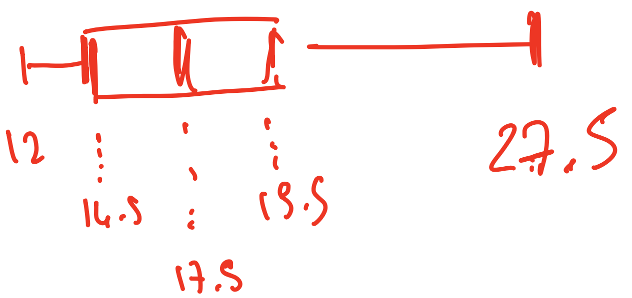
$T_{(i)} = \{12, 14, 15, 16, 17, 18, 19, 20, 21, 210\}$

$Me(T) = \frac{17+18}{2} = 17.5$

$q_{0.25} = \frac{14+15}{2} = 14.5$

$q_{0.75} = \frac{19+20}{2} = 19.5$

$\min(210, 20 + 1.5 \cdot 5) = 27.5$



(2) $Me(T) = 17.5$

$$M_1(T) = \frac{1}{N}\sum_{i=1}^{N} x_i =$$

$$= \frac{1}{10}[12+14+15+16+17+18+19+20+22+210]$$

$$= \frac{1}{10}(363) = 36.3$$

$$S_{Me}(T) = \frac{1}{N}\sum_{i=1}^{N}|x_i - Me|$$

$$= \frac{1}{10}\Big[|12-17.5| + |14-17.5| + |15-17.5| + |16-17.5| + |17-17.5| + |18-17.5| + |19-17.5| + |20-17.5| + |22-17.5| + |210-17.5|\Big]$$

$$= \frac{1}{10}[5.5+3.5+2.5+1.5+0.5+0.5+1.5+2.5+4.5+192.5]$$

$$= \frac{1}{10}(215) = 21.5$$

$$S_{M_1}(T) = \frac{1}{N}\sum_{i=1}^{N}|x_i - \bar{x}| = \frac{1}{10}\Big[|12-36.3| + |14-36.3| + |15-36.3| + |16-36.3| + |17-36.3| + |18-36.3| + |19-36.3| + |20-36.3| + |22-36.3| + |210-36.3|\Big]$$

$$= \frac{1}{10}(347.4) = 34.74$$

EMPIRICAMENTE ABBIAMO MOSTRATO CHE LO SCOSTAMENTO MEDIO DALLA MEDIANA È INFERIORE ALLO SCOSTAMENTO MEDIO DALLA MEDIA ARITMETICA.

(3) $\sigma_T^2 = M_1(T^2) - (M_1(T))^2$

$$M_1(T^2) = \frac{1}{N}\sum_{i=1}^{N} x_i^2$$

$$= \frac{1}{10}\left[12^2 + 14^2 + \dots + 22^2 + 21^2\right]$$

$$= \frac{1}{10}\left[3120\right] = 312$$

$$M_1(T) = \frac{1}{10}\left(12 + 14 + \dots + 22 + 21\right) = 17.4$$

$$\sigma_T^2 = 312 - 17.4^2 = 9.24$$

$$CV = \frac{\sigma_T}{M_1(T)} = \frac{\sqrt{9.24}}{17.4} = 0.175$$

(4) $Y = 60 \times T$ TEMPO IN SECONDI, ALLORA

$$M_1(Y) = M_1(60 \times T) = 60 \times M_1(T) = 60 \times 17.4 = 1044$$

$$\sigma_Y^2 = Var(Y) = Var(60 \times T) = 60^2 \times Var(T) = 60^2 \times 9.24 = 33264$$

$$CV(Y) = \frac{\sigma_Y}{M_1(Y)} = \frac{\sqrt{\sigma_Y^2}}{M_1(Y)} = \frac{\sqrt{33264}}{1044} = 0.175$$

È UGUALE AL VALORE PRECEDENTE.

Esercizio 3

Un reparto di produzione automotive oggi a prodotto le seguenti automobili, suddivise per colore e per tipologia di automobile.

	rosso	blu	grigio	bianco	$n_{i.}$
suv	12	10	27	3	52
utilitaria	2	14	3	11	30
berlina	12	6	0	0	18
$n_{.j}$	26	30	30	14	100

Considerare la distribuzione di frequenze marginale del colore e rispondere alle seguenti domande.

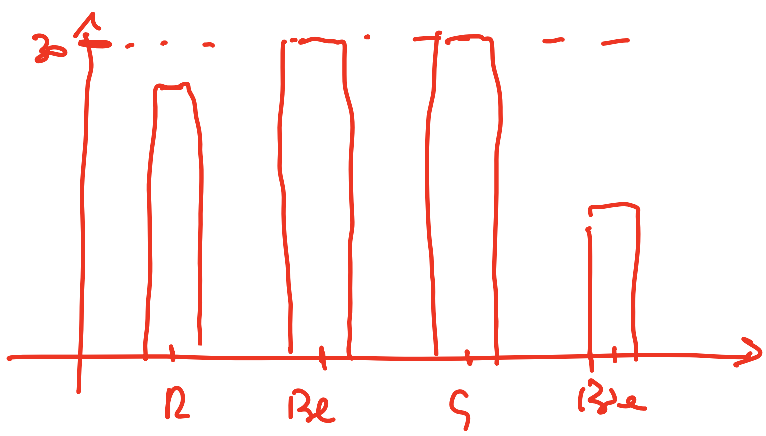
1. Dopo aver identificato la tipologia del carattere in esame, si produca una opportuna rappresentazione grafica che ne mostri l'andamento.
2. Si identifichi la classe modale, commentando la sua rappresentatività.
3. Si calcoli l'indice di mutabilità di Gini, commentando il risultato ottenuto.

Considerando la distribuzione di frequenze congiunta del colore e della tipologia di automobile, riportata nella tabella a doppia entrata precedente, si risponda alle seguenti domande.

4. I due caratteri osservati sono indipendenti in distribuzione? Motivare la risposta data.
5. Volendo interpretare la connessione presente tra le due variabile osservate, si calcoli l'indice di connessione di Pearson normalizzato e si commenti il risultato ottenuto.

(1) CARATTERE QUALITATIVO SU SCALA NOMINALE

X	n_j	f_j
R	26	0.26
Be	30	0.30
G	30	0.30
Bia	14	0.14
	100	



(2) LA CLASSE MODALE È INDETERMINATA, QUINDI NON POSSIAMO DIRE NULLA SULLA RAPPRESENTATIVITÀ.

(3) $G = 1 - \sum_{j=1}^{k} f_j^2 = 1 - \left[0.26^2 + 0.3^2 + 0.3^2 + 0.14^2\right]$

$= 1 - 0.267 = 0.733$

$$G_{Norm} = \left(\frac{N}{N-1}\right) G = \left(\frac{100}{99}\right) 0.733 = 0.740$$

MOLTO VICINO A 1, ALLA MUTABILITÀ.

(4) NON SONO INDIPENDENTI IN DISTRIBUZIONE, ESSENDO PRESENTI DEGLI ZERI NELLA TABELLA A DOPPIA ENTRATA, E IN CASO DI INDIPENDENZA LE FREQUENZE CONGIUNTE SONO UGUALI AL PRODOTTO DELLE MARGINALI CORRISPONDENTI.

(5) $$M_2(\beta) = \sqrt{\frac{1}{N}\sum_{j=1}^{c}\sum_{i=1}^{r}\frac{(n_{ij}-\hat{n}_{ij})^2}{\hat{n}_{ij}}}$$

$\hat{n}_{ij}$	R	Be	G	Bie	
Suv	13.52	15.6	15.6	7.28	52
Uti	7.8	9	9.0	4.20	30
Ber	4.68	5.4	5.4	2.52	18
	26	30	30	14	100

$$\hat{n}_{ij} = \frac{n_{i\cdot} \times n_{\cdot j}}{N}$$

$$\hat{n}_{11} = \frac{52 \times 26}{100} = 13.52$$

$$\hat{n}_{21} = \frac{30 \times 26}{100} = 7.8$$

$(n_{ij}-\hat{n}_{ij})$	R	Be	G	Bie
Suv	1.52	-5.6	11.4	-4.28
Uti	-5.8	5.0	-6.0	6.30

$n_{11} - \hat{n}_{11} = 12 - 13.5 = 1.52$

$n_{21} - \hat{n}_{21} = 2 - 7.8$

Bel	7.32	0.6	-5.4	-2.52

$M_{31} - \hat{M}_{31} = 12 - 4.68$

$$M_2(\rho) = \sqrt{\frac{1}{100} \sum_{j=1}^{4} \sum_{i=1}^{3} \frac{(M_{ij} - \hat{M}_{ij})^2}{\hat{M}_{ij}}}$$

$$= \sqrt{\frac{1}{100}\left[\frac{1.52^2}{13.52} + \frac{-5.6^2}{15.6} + \frac{11.4^2}{15.6} + \frac{-4.18^2}{7.28} + \frac{-5.8^2}{7.8} + \frac{5.0^2}{9} + \frac{-6.0^2}{9.0} + \frac{6.30^2}{4.20} + \frac{7.32^2}{4.68} + \frac{0.6^2}{5.4} + \frac{-5.4^2}{5.4} + \frac{-2.52^2}{2.52} \right]}$$

$$= \sqrt{\frac{1}{100}\left[13.023 + 22.100 + 19.436\right]}$$

$$= \sqrt{\frac{1}{100} 54.564} = 0.739$$

$$\tilde{M}_1(\rho) = \frac{M_2(\rho)}{\sqrt{\min(r-1, c-1)}} = \frac{M_2(\rho)}{\sqrt{2}}$$

$$= 0.522$$

È PRESENTE UNA CONNESSIONE

MO DSMMTTANGNTE ACΓA -