FIRMA DELLO STUDENTE

PRIMA PROVA INTERMEDIA DI STATISTICA
CLEA (COD. 5047/4038)
5 Novembre 2003

Cognome ______________________ *Nome* ______________________

Numero di matricola ______________________

COMPITO D2

Ai fini della valutazione si terrà conto solo ed esclusivamente di quanto riportato negli appositi spazi. Al termine della prova, è OBBLIGATORIO consegnare il presente foglio ed il foglio di brutta (DI CUI NON SI TERRÀ CONTO AI FINI DELLA VALUTAZIONE).

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

Al fine di aprire un nuovo centro sportivo, si analizzano i dati raccolti nell'ultimo anno su 10 centri sportivi dislocati in aree urbane sul territorio italiano, in relazione ai seguenti caratteri:

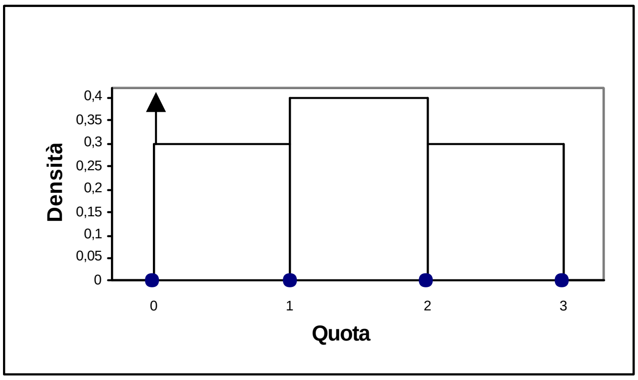
ANNI numero di anni trascorsi dalla costruzione
ISTRUTTORI numero di istruttori
QUOTA ammontare della quota annuale di iscrizione, in migliaia di euro
SOCI numero di iscritti
SPORT numero di sport praticabili
ZONA dislocazione della struttura (0= periferia, 1= centro).

Di seguito sono riportati i dati, con alcuni calcoli utili.

	Anni (A)	**Istruttori (I)**	**Quota (Q)**	**Soci (So)**	**Sport (Sp)**	**Zona (Z)**	I^2	Q^2	Sp^2	Q×I
	10	5	2.1	332	3	1	25	4.41	9	10.5
	1	5	1.5	445	2	1	25	2.25	4	7.5
	1	8	1.1	400	2	1	64	1.21	4	8.8
	5	10	0.8	521	2	0	100	0.64	4	8
	1	20	1.9	543	3	1	400	3.61	9	38
	1	8	2.8	612	5	0	64	7.84	25	22.4
	1	20	1.6	670	5	1	400	2.56	25	32
	10	5	0.7	358	2	0	25	0.49	4	3.5
	10	10	0.9	344	3	0	100	0.81	9	9
	5	8	2.5	450	3	1	64	6.25	9	20
Totali	**45**	**99**	**15.9**	**4675**	**30**		**1267**	**30.07**	**102**	**159.7**

1. (3 punti) Si scriva la distribuzione del carattere QUOTA raggruppato nelle classi [0,1), [1,2), [2,3) e si disegni il corrispondente istogramma.

QUOTA	p_i	c_i
[0,1)	0,3	0,3
[1,2)	0,4	0,4
[2,3)	0,3	0,3
	1	



2. (3 punti) Si calcoli l'indice di concentrazione del carattere ISTRUTTORI sulla base dei dati raggruppati. Si commenti il risultato.

Xi*	Ni*	Fi*	Qi*
5	3	0,3	0,1515
8	3	0,6	0,3939
10	2	0,8	0,5960
20	2	1	1
	10		

$$R^* = \frac{\sum_{i=1}^{N-1}(Fi^* - Qi^*)}{\sum_{i=1}^{N-1} Fi^*} = 1 - \frac{\sum_{i=1}^{N-1} Qi^*}{\sum_{i=1}^{N-1} Fi^*} = 1 - \frac{1{,}1414}{1{,}7} = 0{,}3286$$

Dal valore di R si deduce una bassa concentrazione per il carattere ISTRUTTORI.

3. (3 punti) Dal confronto tra media e mediana, si stabilisca se la distribuzione del carattere QUOTA è simmetrica, obliqua a destra oppure obliqua a sinistra e perché.

- Se si usano i dati grezzi: Me (Q) = 1,55; μ(Q) = 1,59 ? μ(Q) > Me (Q) (distribuzione leggermente obliqua a destra)
- Se si usano i dati riclassificati : Me (Q) = 1,5 ; μ(Q) = 1,5 ? La distribuzione è simmetrica perché Mediana e Media coincidono.

4. (2 punti) Le variabili X e Y hanno le seguenti distribuzioni marginali.

Pr{X=1}=1/4, Pr{X=2}=1/6, Pr{X=3}=1/2, Pr{X=4}=1/12

Pr{Y=-1}=1/6, Pr{Y=0}=2/3, Pr{Y =1}=1/6

Si costruisca la tabella a doppia entrata per il caso di indipendenza statistica tra X e Y e si dica quanto vale la funzione di regressione di Y su X, giustificando la risposta.

Y / X	-1	0	1	$p_X(x_i)$
1	1/24	1/6	1/24	1/4
2	1/36	1/9	1/36	1/6
3	1/12	1/3	1/12	1/2
4	1/72	1/18	1/72	1/12
$p_Y(y_j)$	1/6	2/3	1/6	1

$m_Y(x_i) = 0; \quad x_i = 1,2,3,4$

L'indipendenza statistica implica l'indipendenza regressiva.

5. (3 punti) Si costruisca la tabella a doppia entrata per i caratteri SPORT e ZONA e si dica se la percentuale di centri in cui si possono praticare almeno tre sport è più elevata per le strutture dislocate in periferia o per quelle dislocate in centro città. Si giustifichi la risposta.

Z / Sp	0	1	$p_{Sp}(s_i)$
2	0,2	0,2	0,4
3	0,1	0,3	0,4
5	0,1	0,1	0,2
$p_Z(z_j)$	0,4	0,6	1

$$\text{Fr}(Sp = 3 \mid Z = 0) = \frac{0{,}2}{0{,}4} = 0{,}5$$

$$\text{Fr}\ (Sp = 3 \mid Z = 1) = \frac{0{,}4}{0{,}6} = 0{,}6667$$

La percentuale di centri in cui si possono praticare almeno 3 sport è più elevata per le strutture dislocate in centro.

6. (3 punti) Si calcoli il coefficiente di correlazione lineare tra i caratteri QUOTA e ISTRUTTORI e si commenti la dipendenza dell'ammontare della quota d'iscrizione dal numero di istruttori.

Dalla tabella si ricavano:

$$\mu(Q) = 1{,}59 \quad ; \quad \sigma^2(Q) = 0{,}4789 \quad ; \quad \mu(I) = 9{,}9 \quad ; \quad \sigma^2(I) = 28{,}69$$
$$\text{Cov}(Q, I) = 0{,}229$$

e quindi

$$\rho = \frac{\text{Cov}(Q, I)}{\sigma_Q \sigma_I} = \frac{0{,}229}{0{,}6920 \cdot 5{,}3563} = 0{,}0618$$

La dipendenza di tipo lineare è pressoché assente .

7. (1 punto) In quale caso il coefficiente calcolato nella domanda 6 avrebbe assunto valore zero?

Nel caso di indipendenza correlativa (e, a maggior ragione, nel caso di indipendenza regressiva di Q da I o di I da Q e nel caso di indipendenza statistica trai Q e I).

8. (2 punti) Ogni mattina il signor Rossi ha una probabilità di perdere il tram pari a 1/4 e quindi di arrivare tardi in ufficio. Si scriva la funzione di probabilità della variabile aleatoria "numero di giorni in cui il signor Rossi arriva in ritardo nell'arco dei cinque giorni lavorativi".

$X \sim Bin(5,1/4)$

$$P(X=x) = \begin{cases} \binom{5}{x} \cdot \left(\frac{1}{4}\right)^{x} \cdot \left(\frac{3}{4}\right)^{(5-x)} & x = 0,1,2,3,4,5, \\ 0 & \text{altrove} \end{cases}$$

9. (3 punti) Si considerino due urne, la prima contenente 2 palline rosse e 3 blu e la seconda 4 palline rosse e 1 blu. Si consideri l'esperimento di estrazione di una pallina da una sola delle due urne, assumendo che le due urne siano equiprobabili.

a) Si calcoli la probabilità che la pallina estratta sia blu.

Siano definiti i seguenti eventi:

B: "la pallina estratta è blu"

U_1: "la pallina è estratta dalla prima urna"

U_2: "la pallina è estratta dalla seconda urna"

Applicando il teorema delle probabilità totali, di cui sono soddisfatte le ipotesi, si ottiene:

$$P(B) = P(U_1)P(B \mid U_1) + P(U_2)P(B \mid U_2) = \frac{1}{2} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{1}{5} = 0{,}4$$

b) Se la pallina estratta è blu, qual è la probabilità di aver estratto dalla prima urna?

$$P(U_1 \mid B) = \frac{P(U_1)P(B \mid U_1)}{P(B)} = \frac{3/10}{4/10} = 0{,}75$$