FIRMA DELLO STUDENTE

PRIMA PROVA INTERMEDIA DI STATISTICA

CLEA (COD. 5047/4038)

5 Novembre 2003

Cognome ***Nome***

Numero di matricola

COMPITO A1

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

Sono stati raccolti alcuni dati relativi a dei test sulle 6 automobili disponibili presso un concessionario. In particolare, si conosce:

FRENI bontà del dispositivo di frenata (ottimo, medio, pessimo)
CONS consumo medio di benzina (litri per 100 km)
VEL velocità massima (km/h)
TRAZ trazione (ant= anteriore, pos= posteriore, int= integrale)

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

AUTO	FRENI	CONS	VEL	TRAZ	$CONS^2$	VEL^2	CONS*VEL
Panda	Pessimo	7	90	Ant	49	8100	630
Tigra	Medio	8	110	Int	64	12100	880
Puma	Ottimo	7	100	Pos	49	10000	700
Pantera	Medio	9	120	Ant	81	14400	1080
Jaguar	Medio	11	140	Int	121	19600	1540
Furetto	Ottimo	10	130	Int	100	16900	1300
Totali		52	690		464	81100	6130

1. **(2 punti)** Quali indicatori si possono calcolare per i seguenti caratteri? Si indichi con ✓ dove si possono calcolare.

	FRENI	CONS	TRAZ	VEL
Coefficiente di variazione		✓		✓
Media		✓		✓
Mediana	✓	✓		✓
Moda	✓	✓	✓	✓

2. (4 punti)

a) Date l'equazione della retta di regressione lineare del consumo (CONS) sulla velocità massima (VEL).

(2 punti) Dalla tabella si ricavano $\mu(\mathrm{VEL}) = 115$, $\mu(\mathrm{CONS}) = 8{,}6667$, Cov(CONS,VEL) = 24,9962, $\sigma^2(\mathrm{VEL}) = 291{,}6667$ da cui $b = 0{,}0857$ e $a = -1{,}1888$. La retta di regressione cercata è quindi:

$$r_{\mathrm{CONS}}(\mathrm{vel}) = -1{,}1888 + 0{,}0857 \cdot \mathrm{vel}$$

b) Calcolate con un indice opportuno la bontà della retta di regressione di cui al punto precedente e commentate il risultato.

(1 punto) Dalla tabella risulta $\sigma^2(\mathrm{CONS}) = 2{,}2216$ e quindi

$$R^2 = \frac{[\mathrm{Cov}(\mathrm{CONS},\mathrm{VEL})]^2}{\sigma^2(\mathrm{CONS}) \cdot \sigma^2(\mathrm{VEL})} = \frac{24{,}9962^2}{2{,}2216 \cdot 291{,}6667} = 0{,}9643$$

La retta di regressione risulta pertanto un modello molto buono.

c) Dal concessionario è stata testata un'altra automobile, la cui velocità massima è 160 km/h. Sulla base dei parametri stimati precedentemente prevedere il consumo dell'automobile.

(1 punto) $r_{\mathrm{CONS}}(160) = -1{,}1888 + 0{,}0857 \cdot 160 = 12{,}5232$

3. (2 punti) Nella seguente tabella è riportato, per 6 grandi aeroporti americani, il numero di passeggeri registrati nel corso dell'ultimo anno (i dati sono espressi in milioni). Definite un indice che consenta di valutare il livello di concentrazione del carattere analizzato e calcolatelo sui dati riportati nella tabella.

Aeroporto	***Numero di passeggeri***
Atlanta	62
Los Angeles	57
Miami	33
Chicago	69
Dallas/Ft W.	58
San Francisco	38

x_i	F_i	Q_i
33	1/6	0,1041
38	2/6	0,2240
57	3/6	0,4038
58	4/6	0,5867
62	5/6	0,7823
69	1	1
317		

$$R = \frac{\sum_{i=1}^{N-1}(Fi - Qi)}{\sum_{i=1}^{N-1} Fi} = 1 - \frac{\sum_{i=1}^{N-1} Qi}{\sum_{i=1}^{N-1} Fi}$$

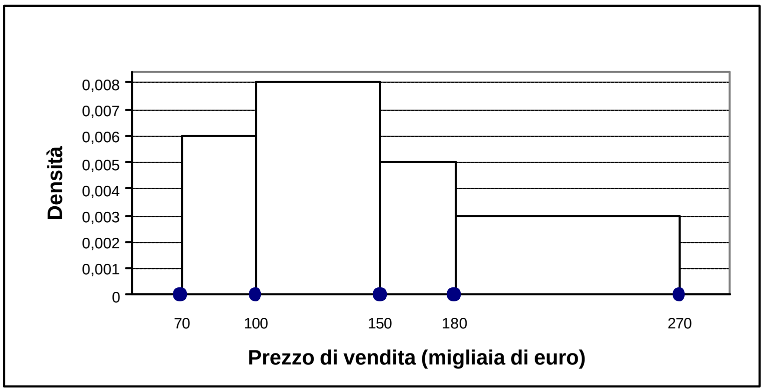
$$R = 1 - \frac{2{,}1009}{(15/6)} = 0{,}1596$$

Il carattere risulta poco concentrato.

4. (4 punti) Nella tabella è riportata la distribuzione del carattere X, che rappresenta i prezzi di vendita di 100 appartamenti (i dati sono espressi in migliaia di euro).

a) Rappresentate il carattere X utilizzando un opportuno grafico (specificate chiaramente i valori sugli assi).
b) Calcolate la percentuale di appartamenti venduti ad un prezzo compreso fra 120 e 180 mila euro.
c) Determinate la mediana del carattere X.

$[x_i, x_{i+1})$	n_i	p_i	c_i
[70-100)	18	0,18	0,006
[100-150)	40	0,40	0,008
[150-180)	15	0,15	0,005
[180-270)	27	0,27	0,003
TOTALE	**100**	**1**	



(1 punto)

$$\text{Fr}(120 \le X \le 180) = 30 \cdot 0{,}008 + 0{,}15 = 0{,}39 = 39\%$$

(1 punto)

$$\text{Me}(X) = 100 + \frac{(0{,}5 - 0{,}18)}{0{,}008} = 140$$

5. (3 punti) Per i 200 alunni delle scuole superiori di un istituto tecnico sono stati rilevati i caratteri Casco (ovvero se generalmente portano il casco in motorino) e Numero incidenti avuti.

		Numero incidenti avuti			
		0	1	4	
Casco	Sì	0,16	0,2	0,04	0,40
	No	0,24	0,3	0,06	0,60
		0,40	0,5	0,10	1

a) Tali caratteri sono tra loro statisticamente indipendenti? (Giustificare la risposta)

(1 punto)

I due caratteri sono statisticamente indipendenti perché è verificata la condizione

$$p_{XY}(x_i^*, y_j^*) = p_X(x_i^*) \cdot p_Y(y_j^*) \qquad \forall i, \forall j$$

b) Sulla base di quanto ottenuto nel punto precedente, cosa si può affermare sulla dipendenza in media (dipendenza regressiva) di Numero di incidenti rispetto a Casco? Perché?

(2 punti)
Il Numero di incidenti è regressivamente indipendente da Casco perché l'indipendenza statistica implica l'indipendenza regressiva.

6. (2 punti) Nella seguente tabella, gli impiegati di una grande società sono stati classificati in base alle loro abitudini legate al fumo, e in base alla posizione occupata nella società (amministrativi o venditori). Si ipotizzi che le due variabili siano *indipendenti*. Si sa inoltre che il 40% degli impiegati fa il venditore. Il 30% degli impiegati non fuma *e* occupa una posizione da amministratore. Infine, *fra i venditori*, il 25% è fumatore leggero. Sulla base delle informazioni fornite, completate la tabella a doppia entrata inserendo le frequenze relative.

Posizione / ***Fumo***	Amministrativo	Venditore	
Non fumatore	**0,30**	**0,2**	0,5
Fumatore leggero	**0,15**	**0,1**	0,25
Fumatore accanito	**0,15**	**0,1**	0,25
	0,6	0,4	1

7. (2 punti) Fate un esempio di una tabella delle frequenze congiunte in cui ci sia indipendenza regressiva (ovvero indipendenza in media) di Y rispetto a X ma non indipendenza statistica.

X \ Y	0	1	2	$p_X(x_i)$
10	**0,15**	**0,50**	**0,15**	0,80
20	**0,05**		**0,05**	0,10
30		**0,10**		0,10
$p_Y(y_j)$	0,20	0,60	0,20	1

8. (3 punti) Il sindaco di New York ha deciso di intraprendere un programma per i giovani, al fine di ridurre il tasso di criminalità giovanile. La probabilità che il suo programma abbia successo è valutata a 0.9, se il programma è supportato da un finanziamento integrativo da parte della Law Enforcement Assistance Administration. Se invece la LEAA non sostiene il programma, la probabilità di successo scende a 0.3. Sapendo che la probabilità che il programma ottenga il finanziamento è pari a 0.6, qual è la probabilità che il programma abbia successo?

Siano gli eventi S: “il programma ha successo” ed F: “il programma è finanziato dalla LEAA”
Applicando il teorema delle probabilità totali, di cui sono soddisfatte le ipotesi, si ottiene:

$$P(S) = P(S \mid F) \cdot P(F) + P(S \mid \overline{F}) \cdot P(\overline{F}) = 0{,}9 \cdot 0{,}6 + 0{,}3 \cdot 0{,}4 = 0{,}66$$

9. (2 punti) La probabilità per ciascun studente di riuscire a entrare nella discoteca “Ripa90” è pari a un terzo. Si presentano indipendentemente l’uno dall’altro 5 studenti. Qual è la probabilità che entrino esattamente tre studenti?

$X \sim Bin(5,1/3)$

$$P(X=3) = \binom{5}{3} \cdot \left(\frac{1}{3}\right)^3 \cdot \left(\frac{2}{3}\right)^2 = 10 \cdot \frac{4}{3^5} = 0{,}1646$$