PROVA SCRITTA DI STATISTICA (COD 4038-5047-371-377)
4 Febbraio 2004

MODALITÀ A

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

ESERCIZIO 1 (6 punti)

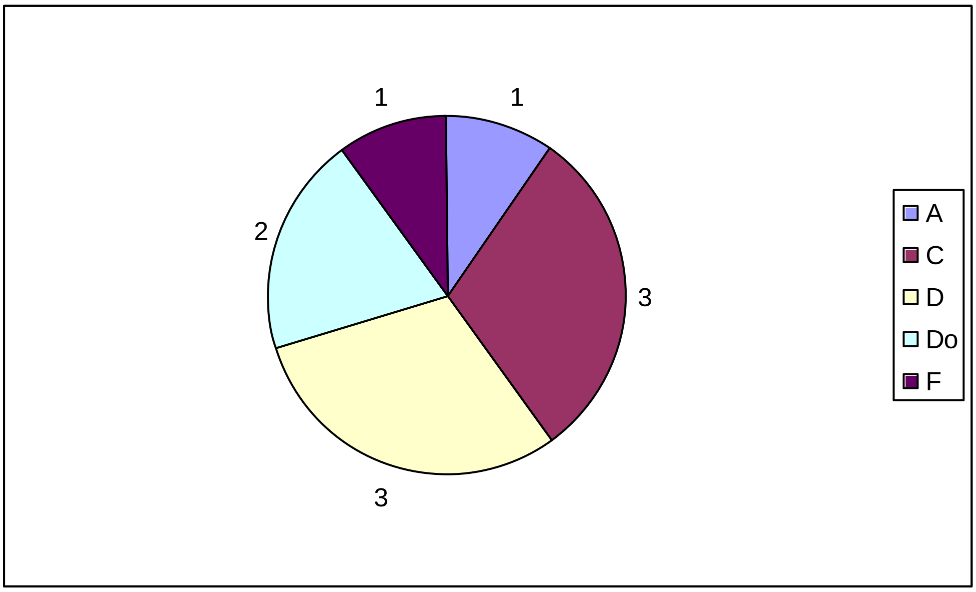
Da una classifica del sito internet IBS risulta che i 10 film più venduti in videocassetta nella settimana precedente il giorno di Natale 2003 sono quelli elencati nella tabella sottostante.

Per ciascun film è indicato:

TITOLO	
REGISTA	
PREZZO	Prezzo in euro
PREZZO INTERNET	Prezzo internet in euro
DURATA	Durata in minuti
GENERE	Genere del FILM (D=drammatico, F=fantastico, C=commedia, Do=documentario, A=animazione)

TITOLO	Regista	Prezzo	Prezzo internet	Durata	Genere	Prezzo^2	Durata^2
La meglio gioventù	M. T. Giordana	20,9	18,99	336	D	436,81	112896
Io non ho paura	G. Salvatores	11,99	10,99	109	D	143,7601	11881
Pinocchio	R. Benigni	13,39	11,39	107	F	179,2921	11449
Una settimana da Dio	T. Shadyac	11,39	9,99	101	C	129,7321	10201
Bowling a Columbine	M. Moore	9,99	6,49	120	Do	99,8001	14400
Il popolo migratore	J. Perrin, J. Cluzad, M. Debats	11,99	10,99	92	Do	143,7601	8464
La finestra di fronte	F. Ozpetek	12,1	10,99	106	D	146,41	11236
Vacanze di Natale	C. Vanzina	8	8	90	C	64	8100
Spirit, Cavallo selvaggio	K. Asbury, L. Cook	18,54	10,28	84	A	343,7316	7056
My name is Tanino	P. Verzì	9,99	8,99	100	C	99,8001	10000
SOMMA		128,28	107,1	1245		1787,0962	205683

1) Fornire un'opportuna rappresentazione grafica del carattere GENERE. (**1 punto**)



2) Calcolare la funzione di regressione di PREZZO su GENERE e dire se PREZZO è regressivamente indipendente da GENERE. Motivare la risposta. (**2 punti**)

M(Prezzo|Genere=D)=(20.9+11.99+12.1)/3=14.99667
M(Prezzo|Genere=F)=13.39
M(Prezzo|Genere=Do)=(9.99+11.99)/2=10.99
M(Prezzo|Genere=C)=(11.39+8+9.99)/3=9.7933
M(Prezzo|Genere=A)=18.54.

Poiché la funzione di regressione non è costante PREZZO non è regressivamente indipendente da GENERE.

3) Quale tra i due caratteri PREZZO e DURATA presenta maggiore variabilità? (**2 punti**)

Bisogna confrontare i coefficienti di variazione di PREZZO e DURATA.

M(P)=12.828 $\quad$ M(P^2)=178.70962 $\quad$ V(P)=M(P^2)-M(P)^2=178.70962-164.557584=14.152036

$$CV(P) = \frac{\sqrt{V(P)}}{M(P)} = \frac{3.7619}{12.828} \approx 0.2932$$

M(D)=124.5 $\quad$ M(D^2)=20568.3 $\quad$ V(D)=M(D^2)-M(D)^2=20568.3-15500.25=5068.05

$$CV(D) = \frac{\sqrt{V(D)}}{M(D)} = \frac{71.1902}{124.5} \approx 0.5718$$

Poiché CV(D)>CV(P) DURATA presenta una variabilità maggiore di PREZZO.

4) Calcolare il risparmio medio che si ottiene acquistando una videocassetta su internet. (**1 punto**)

(PREZZO TOTALE - PREZZO INTERNET TOTALE)/10=2.118

ESERCIZIO 2 (3 punti)

Un'urna contiene 3 palline nere e 7 bianche. Si estrae a caso *senza reimmissione* una pallina e

a) se è nera nell'urna vengo aggiunte due palline bianche ed una nera;

In questo caso la composizione dell'urna diventa 3 nere e 9 bianche.

b) se è bianca non viene aggiunta nessuna pallina all'urna.

In questo caso la composizione dell'urna diventa 3 nere e 6 bianche.

Viene estratta successivamente una seconda pallina dall'urna.

1) Calcolare la probabilità che entrambe le palline siano bianche. (**1 punto**)

$$P(B_1 \cap B_2) = \frac{7}{10} \times \frac{6}{9} = \frac{7}{15} = 0.4\overline{6}$$

2) Calcolare la probabilità che le palline estratte siano di due colori differenti. (**2 punti**)

$$P\left(\underbrace{(B_1 \cap N_2) \cup (N_1 \cap B_2)}_{\text{sono incompatibili}}\right) = P(B_1 \cap N_2) + P(N_1 \cap B_2) = \frac{7}{10} \times \frac{3}{9} + \frac{3}{10} \times \frac{9}{12} \approx 0.4583$$

ESERCIZIO 3 (2 punti)

Si consideri la seguente tabella a doppia entrata in cui sono indicate unicamente le distribuzioni marginali di X e Y. Completare la tabella in modo tale che X sia regressivamente indipendente da Y, ma le due variabili non siano statisticamente indipendenti.

Una possibile tabella è la seguente:

X\Y	5	6	10	p_X
1		0,3		0,3
2	0,2		0,2	0,4
3		0,3		0,3
p_Y	0,2	0,6	0,2	

ESERCIZIO 4 (5 punti)

Sia X una variabile aleatoria avente funzione di densità

$$f_x(x) = \begin{cases} \frac{1}{k} x^2 & -2 < x < 2 \\ 0 & altrove \end{cases}$$

1) Determinare il valore di k. (**2 punti**)

Affinché f(x) sia una funzione di densità di probabilità devono essere verificate le seguenti due condizioni:

a) $f(x) \geq 0$

b) $\int_{-\infty}^{+\infty} f(x)dx = 1$

Dalla condizione a) si ottiene k>0

Dalla condizione b) si ottiene

$$\int_{-2}^{2} \frac{1}{k} x^2 dx = \frac{1}{3k} x^3 \Big|_{-2}^{2} = \frac{8}{3k} + \frac{8}{3k} = \frac{16}{3k}$$

$\frac{16}{3k} = 1$ k=16/3

$$f_x(x) = \begin{cases} 3/16\, x^2 & -2 < x < 2 \\ 0 & altrove \end{cases}$$

2) Data la variabile aleatoria Y=3X+2, calcolare E(Y). (**3 punti**)

E(Y)=3E(X)+2

$$E(X) = \int_{-2}^{2} x \frac{3}{16} x^2 dx = \frac{3}{16} \int_{-2}^{2} x^3 dx = \frac{3}{16} \frac{x^4}{4} \Big|_{-2}^{2} = \frac{3}{16}\left(\frac{16}{4} - \frac{16}{4}\right) = 0$$

E(Y)=2

ESERCIZIO 5 (punti 4)

Sia consideri un campione bernoulliano di ampiezza 5 estratto da una popolazione X con E(X)=(μ+3)/2 e V(X)=3.

1) Proporre uno stimatore non distorto per μ. (**2 punti**)

$E(\overline{X}) = \frac{m+3}{2}$ di conseguenza $T = 2\overline{X} - 3$ è uno stimatore non distorto per μ.

2) Calcolare l'errore quadratico medio dello stimatore proposto al punto precedente. (**2 punti**)

Poiché lo stimatore considerato è non distorto

$$EQM(T) = V(T) = V(2\overline{X} - 3) = 4V(\overline{X}) = 4 \times \frac{3}{5} = \frac{12}{5}$$

ESERCIZIO 6 (2 punti)

Sia $X_1,\ldots X_5$ un campione casuale estratto da una variabile aleatoria normale con media pari a θ e varianza pari a 1.
Al fine di verificare le seguenti ipotesi

$$H_0 : q = \tfrac{1}{4} \qquad H_1 : q > \tfrac{1}{4}$$

è stata proposta la seguente regione critica

$$R = \{(x_1, \cdots x_5) : \bar{x} \geq 1.5\}.$$

Calcolare la probabilità di errore di prima specie.

$$P(R \mid H_0) = P(\bar{X} \geq 1.5 \mid q = 0.25) = P\left(Z \geq \frac{1.5 - 0.25}{1/\sqrt{5}} \right) = 1 - f(2.79) = 1 - 0.9974 = 0.0026$$

ESERCIZIO 7 (5 punti)

Da un'indagine effettuata tra gli stranieri residenti in Grecia risulta che su 1200 intervistati 48 sono italiani.

1) Proporre uno stimatore non distorto per la proporzione di italiani residenti in Grecia e calcolarne la stima. (**2 punti**)

Stimatore : $\bar{X} = \dfrac{\sum_i X_i}{n}$

Stima: $\bar{x} = \dfrac{\sum_i x_i}{n} = \dfrac{48}{1200} = 0.04$

2) Calcolare l'intervallo di confidenza al 90% per la proporzione di italiani residenti in Grecia. (**2 punti**)

$$\left(\bar{x} - z_{1-\alpha/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} ; \bar{x} + z_{1-\alpha/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} \right)$$

$$\left(0.04 - 1.645 \sqrt{\frac{0.04(1-0.04)}{1200}} ; 0.04 + 1.645 \sqrt{\frac{0.04(1-0.04)}{1200}} \right)$$

$$(0.0307; 0.0493)$$

3) Ad un livello di significatività del 10% accettereste o rifiutereste l'ipotesi nulla che la proporzione di italiani residenti in Grecia sia uguale al 4% contro l'ipotesi che sia diversa dal 4%? Motivare la risposta. (**1 punto**)

Poiché 0.04 appartiene all'intervallo di confidenza al 90% accettiamo l'ipotesi H_0.