SOLUZIONE
PRIMA PROVA INTERMEDIA DI STATISTICA
CLEA (COD. 5047/4038)
6 Novembre 2002

COMPITO B1

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

Su un collettivo di 10 individui vengono rilevati i seguenti caratteri:

SESSO (SE)	sesso (Femmina = 0, Maschio = 1)
FELICITA' (FE)	livello di felicità (1=Poco felice, 2= Abbastanza felice, 3 = Molto felice)
ETA' (ET)	età dell'intervistato
ISTRUZIONE (IS)	numero di anni di studio completati (dall'intervistato)
ISTRUZIONE PADRE (IP)	numero di anni di studio completati dal padre
DISOCCUPATO (DI)	e' stato disoccupato per più di un mese? (0 = NO, 1 = SI)
STATO CIVILE (SC)	stato civile (1= nubile/celibe, 2 = sposato, 3 = vedovo, 4 = separato)
NUMERO DI FIGLI (NU)	numero di figli dell'intervistato

I dati sono contenuti nella seguente tabella:

Unità	SE	FE	ET	IS	IP	DI	SC	NU	IS^2	IP^2	IS*IP
1	1	3	26	10	8	0	1	2	100	64	80
2	0	1	35	20	17	0	1	0	400	289	340
3	0	3	64	20	12	1	3	1	400	144	240
4	1	1	56	20	12	1	2	1	400	144	240
5	1	1	48	12	17	0	1	0	144	289	204
6	0	2	37	10	8	1	2	2	100	64	80
7	1	3	29	10	8	1	4	1	100	64	80
8	0	2	32	10	2	0	2	0	100	4	20
9	1	1	48	14	2	0	2	4	196	4	28
10	0	2	72	20	8	1	2	1	400	64	160
Totali				**146**	**94**				**2340**	**1130**	**1472**

1 **(2 punti)** Si individuino tutti gli indici di sintesi che è possibile calcolare per i seguenti caratteri (relativi alla tabella precedente).

	Moda	Mediana	Media aritmetica
DISOCCUPATO (DI)	♦	ÿ	ÿ
ETA' (ET)	♦	♦	♦
FELICITA' (FE)	♦	♦	ÿ
ISTRUZIONE (IS)	♦	♦	♦

2 (2 punti) Si forniscano i tre assiomi che definiscono una funzione (misura) di probabilità. Si chiama funzione di probabilità una funzione d'insieme a valori reali definita su una famiglia di sottoinsiemi $\Re$ avente le seguenti proprietà:

1) $P(\Omega) = 1$

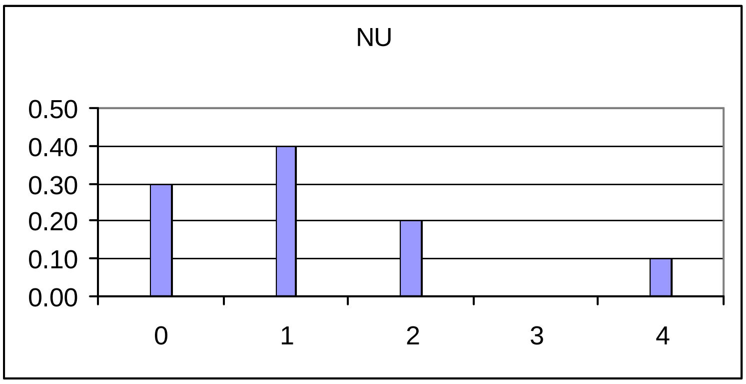
2) $P(A) \geq 0 \quad \forall A \subset \Omega$

3) $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

per ogni successione di eventi inclusi in Ω a due a due disgiunti

3 (2 punti) Si fornisca la distribuzione delle frequenze del carattere NUMERO DI FIGLI (NU). Si rappresenti poi tale distribuzione con un opportuno grafico.

NU	**Frequenze assolute**	**Frequenze relative**
0	3	0.30
1	4	0.40
2	2	0.20
4	1	0.10
Totale complessivo	10	1.00



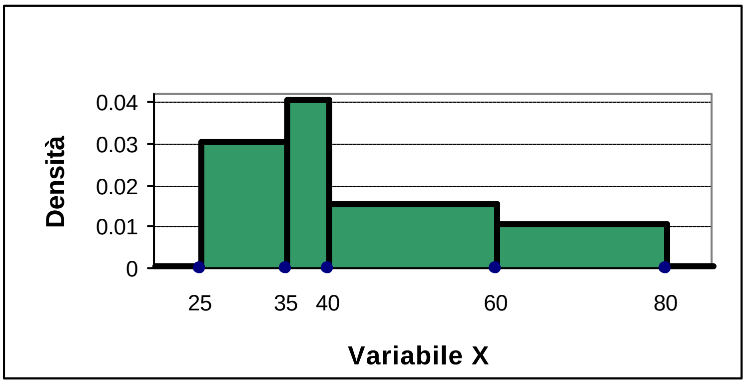
4 (2 punti) Si indichi quale tra i due caratteri ISTRUZIONE (IS) e ISTRUZIONE PADRE (IP) è maggiormente variabile. Si giustifichi la risposta.

Il carattere ISTRUZIONE PADRE è più variabile, in quanto il coefficiente di variazione e' maggiore
m_{IS}**=14.6** s_{IS}^2**=234-(14.6)^2=234-213.16= 20.84** m_{IP}**=9.4** s_{IP}^2**=113-(9.4)^2=113-88.36= 24.64** $CV_{IS} = \frac{s_{IS}}{\mid m_{IS} \mid} = \frac{\sqrt{20.84}}{14.6} = \frac{4.5651}{14.6} = 0.3127 \quad CV_{IP} = \frac{s_{IP}}{\mid m_{IP} \mid} = \frac{\sqrt{24.64}}{9.4} = \frac{4.9639}{9.4} = 0.5281$

5 **(2 punti)** In una contea del Michigan (USA) si è ipotizzato che è più facile cadere nella dipendenza da alcol se ci si iscrive al college. Si considera un campione di residenti nella contea.
Supponiamo di conoscere le seguenti probabilità:
a) la probabilità che la persona non sia alcolizzata è 0.7.
b) la probabilità che, nell'ipotesi che la persona sia alcolizzata, essa sia stata iscritta al college è 0.8.
c) la probabilità che, nell'ipotesi che la persona non sia alcolizzata, essa sia stata iscritta al college è 0.7
Si calcoli la probabilità che la persona sia alcolizzata, nell'ipotesi che sia stata iscritta al college.

$P(C) = P(C \mid A) \times P(A) + P(C \mid \overline{A}) \times P(\overline{A}) = 0.8 \times 0.3 + 0.7 \times 0.7 = 0.24 + 0.49 = 0.73$ $P(A \mid C) = \frac{P(C \mid A) \times P(A)}{P(C)} = \frac{0.8 \times 0.3}{0.73} = \mathbf{0.3288}$

6 **(2 punti)** Si consideri il carattere ETA' (ET) come carattere continuo per intervalli, utilizzando le seguenti classi [25; 35), [35; 40), [40; 60), [60; 80)**.** Si rappresenti la distribuzione di tale carattere per mezzo di un istogramma, (si riportino in un'opportuna tabella le quantità necessarie alla realizzazione del grafico).



Estremi inferiori delle classi	Estremi superiori delle classi	Frequenze assolute	Frequenze relative	Densità
25	35	3	0.3	0.03
35	40	2	0.2	0.04
40	60	3	0.3	0.015
60	80	2	0.2	0.01

7 **(2 punti)** Si supponga che in un'urna U vi siano 2 palline bianche, 5 rosse e 3 nere.

a) Si estraggono due palline SENZA reimmissione. Si calcoli la probabilità di ottenere due palline bianche.

b) Si supponga invece che l'estrazione avvenga CON reimmisisone, si calcoli la probabilità di ottenere due palline bianche.

a)	b)
A_1={pallina bianca alla prima estrazione} A_2={pallina bianca alla secondaestrazione} $P(A_1 \cap A_2) = P(A_2 \mid A_1)P(A_1) = \frac{1}{9}\frac{2}{10} = \frac{1}{45}$	$P(A_1 \cap A_2) = P(A_2)P(A_1) = \frac{2}{10}\frac{2}{10} = \frac{1}{25}$

8 **(2 punti)** La seguente tabella a doppia entrata contiene le frequenze relative congiunte delle variabili statistiche X e Y. Si valuti con un opportuno indice il grado di connessione fra le variabili X e Y.

X/Y	**0**	**1**	**2**
-1	0.2		
2			0.4
3		0.4	

$j^2 = 2 \quad \tilde{j}^2 = 1$

Poiché vi è una sola frequenza non nulla su ogni riga e su ogni colonna, ci troviamo in una situazione di massima connessione (bilaterale). L'indice relativo di connessione $\tilde{j}^2$ assumerà quindi il valore 1.

9 **(2 punti)** Si scriva l'espressione analitica dell'interpolante lineare (retta dei minimi quadrati) del carattere ISTRUZIONE (IS) sul carattere ISTRUZIONE PADRE (IP).

COV(IS,IP)=147.2-14.6*9.4=147.2-137.24=9.96 s_{IP}^2**=24.64** **b=COV(IS,IP)/VAR(IP)=9.96/24.64=0.4042** **a=**m_{IS}-m_{IP}***b=14.6-9.4*0.4042=14.6-3.7995=10.8005**
IS=10.8005+0.4042*IP

10 **(2 punti)** Si determini la funzione di regressione della variabile ISTRUZIONE (IS) rispetto alla variabile SESSO (SE). E' possibile affermare che in media il grado di istruzione del soggetto non dipende dal sesso del soggetto medesimo?

μ(IS|SE=0)=16 μ(IS|SE=1)=13.2
NO, ISTRUZIONE non è regressivamente indipendente da SESSO Il grado di istruzione medio delle donne (condizionato a SESSO=0) è più elevato del grado di istruzione medio degli uomini (condizionato a SESSO=1)

11 **(2 punti)** Si consideri la seguente distribuzione delle frequenze assolute di una variabile statistica X. Si fornisca la definizione di primo quartile per una variabile statistica e si calcoli il valore per la v.s. X.

x_i^*	n_i
0	10
1	40
2	30
3	20

Il primo quartile di una distribuzione statistica X è quel valore q_1 in corrispondenza del quale valgono entrambe le relazioni: $Fr\{X \leq q_1\} \geq 0,25$ $Fr\{X \geq q_1\} \geq 0,75$
q_1=1 oppure Nel caso della v.s. X, la posizione occupata dal primo quartile nella serie ordinata delle osservazioni è $\frac{N+1}{4} = 25.25$. Tale valore si approssima all'intero più vicino (25). Quindi si avrà: $q_1 = x_{(25)} = 1$.

12 (2 punti) 20 aziende hanno vinto *complessivamente* 40 gare di appalto. Si scriva la distribuzione delle frequenze della v.s. “numero di gare di appalto vinte” nel caso di equiripartizione e nel caso di massima concentrazione.

Equiripartizione

x_i^*	n_i	p_i
2	20	1

Massima concentrazione

x_i^*	n_i	p_i
0	19	0.95
40	1	0.05