- Appello del 25 Giugno 2014 -

Esercizio 1)

Uno studio di analisi veterinarie vuole monitare il livello di soddifazione dei pazienti riguardo al trattamento ricevuto. Questo viene espresso mediante cinque possibili valutazioni Insufficiente (I), Scarso (S), Accettabile (A), Discreto (D) e Buono (B). Dopo due settimane si sono ottenute le seguenti valutazioni

A	D	I	S	B	A	I	I	S	A
S	D	B	S	A	B	D	A	S	I
B	I	D	A	A	D	I	S	D	A
D	B	S	I	A	I	D	B	A	D

Il candidato

a) determini la tipologia del carattere.
b) illustri la serie utilizzando una rappresentazione grafica opportuna
c) descriva cosa indica l'indice di posizione e, se possibile, ne calcoli uno adeguato alla serie di dati in esame.
d) descriva cosa indicano gli indici di curtosi e, se possibile, ne calcoli uno adeguato alla serie di dati in esame.

Esercizio 2)

I dati illustrati nel precedente esercizio sono stati organizzati tenendo conto del numero di anni di affiliazione del cliente ottenendo la seguente tabella a doppia entrata

		Y: giudizio espresso					Marginali
		I	S	A	D	B	
X: anni di affiliazione	da 0 a 4					**1**	12
	da 4 a 8	**3**	**4**	**3**	**3**		16
	da 8 a 12	**0**	**1**	**4**			12
	Marginali						

Il candidato

a) completi la tabella con i dati mancanti.
b) indichi e calcoli, se possibile, un opportuno indice di posizione per la serie bivariata
c) indichi e calcoli, se possibile, un opportuno indice di variabilità per la serie bivariata
d) se possibile, verifichi, ad un opportuno livello di significatività, se i due caratteri si possono dire indipendenti. Nel caso non fosse possibile, indichi quale ipotesi non viene verificata ed indichi il procedimento da segure nel caso non ci fossero impedimenti.

Esercizio 3)

Il candidato stimi puntualmente e per intervallo il valore atteso degli anni di affiliazione utilizzando i dati riportati come carattere X nella tabella a doppia entrata dell'Esercizio 2.

Esercizio 4)

Si considerino i seguenti eventi:

E_1: si abbia $x < 0.5$ estraendo da una $Ber(0.6)$

E_2: si abbia $y < 0$ dove y è estratto da una normale con valore atteso e varianza unitaria.

Il candidato,sapendo che la probabilità che gli eventi si verifichino contemporaneamente è del 10%,

a) calcoli le seguenti Probabilità: $P(E_1)$; $P(E_2)$; $P(E_1 \cup E_2)$; $P(E_1 | E_2)$; $P(E_2 | E_1)$;
b) indichi se gli eventi E_1 ed E_2 possono ritenersi dipendenti.

- Appello del 25 Giugno 2014 -
Svolgimento

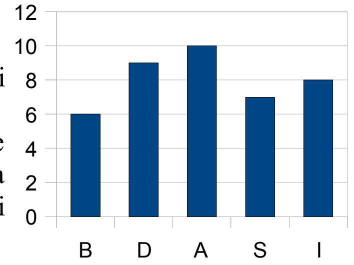
Esercizio 1)

a) Determini la tipologia del carattere.

Il carattere è di tipo qualitativo (in quanto espresso da giudizi e non numeri) ordinabile (in quanto i giudizi possono essere ordinati).

b) illustri la serie utilizzando una rappresentazione grafica opportuna.

Per caratteri qualitativi ordinabili sono possibili due tipologie di rappresentazioni grafiche: il diagramma a barre ed il diagramma a torta.
In questa soluzione si è scelto di rappresentare il primo. Un diagramma a barre è costituito da una serie di barre (orizzontali o verticali) poste in un diagramma cartesiano in cui su di un asse son riportate le modalità del carattere e sull'altro si riportano le frequenze assolute.



c*) descriva cosa indica l'indice di posizione e, se possibile, ne calcoli uno adeguato alla serie di dati in esame..*

L'indice di posizione di una serie di osservazioni indica il valore centrale che viene assunto dalla serie. Gli indici di posizione visti a lezione sono tre: moda, mediana e media. Per i caratteri in esame sono calcolabili sono i primi due. La moda (ovvero la modalità cui corrisponde la modalità maggiore) della serie è A.

d) descriva cosa indica l'indice di curtosi e, se possibile, ne calcoli uno adeguato alla serie di dati in esame.

Gli indici di curtosi indicano quanto la serie di osservazioni si discosta da una distribuzione normale avente stessa stessa media e varianza della serie di osservazioni. Gli indici di curtosi sono calcolabili solo per caratteri quantitativi.

Esercizio 2)

a) Completi la tabella con i dati mancanti.

La tabella si completa tenendo conto che la somma delle colonne deve coincidere con i dati illustrati nell'esercizio 1.

		Y: giudizio espresso					Marginali
		I	S	A	D	B	
X: anni di affiliazione	da 0 a 4	5	2	3	1	**1**	12
	da 4 a 8	**3**	**4**	**3**	**3**	3	16
	da 8 a 12	**0**	**1**	**4**	5	2	12
	Marginali	8	7	10	9	6	**40**

b) Il candidatoindichi e calcoli, se possibile, un opportuno indice di posizione per la serie bivariata

Gli indici di posizione per statistiche bivatiatre visti al lezione sono di due tipi: media e moda. Essendo presente un carattere qualitativo (il carattere Y) si può calcolare solo la moda che coincide con la modalita (della bivariata) cui corrisponde la frequenza maggiore. Nel caso in esame la frequenza assoluta maggiore è 5 rilevata per due modalità (statistica bi-modale)

(da 0 a 4; I) (da 8 a 12; D)

c) Il candidato indichi e calcoli, se possibile, un opportuno indice di variabilità per la serie bivariata

Per le statistiche bivariate è stato visto un solo indice la matrice varianza/covarianza che è calcolabile solo per serie quantitative (entrambi i caratteri quantitativi). Essendo presente un carattere qualitativo (il carattere Y) non è possibile procedere al calcolo.

c) Il candidato se possibile, verifichi, ad un opportuno livello di significatività, se i due caratteri si possono dire indipendenti. Nel caso non fosse possibile, indichi quale ipotesi non viene verificata ed indichi il procedimento da segure nel caso non ci fossero impedimenti.

Per verificare se i due caratteri sono indipendenti si può effettuare un test di ipotesi volto a verificare se le frequenze delle osservazioni rilevate nel campione sono sufficientemente vicine (ad un determinato livello di significatività) a quelle teoriche ottenute dall'ipotesi di indipendeza. Il test sfrutta la distribuzione limite dello

stimatore di Pizzetti Pearson che viene ad essere un chi quadrato avente gradi di libertà pari a quelli del numero di parametri liberi della distribuzione teorica.
Il primo punto di questa procedura consiste nel calcolo delle frequenze teoriche ricavate dalle frequenze marginali ottenute orlando la tabella delle frequenze .

$$\hat{n}_{i,j} = n\,\hat{p}_{i,j} = \frac{n_{i,+}\,n_{+,j}}{n} \quad \forall\, i,j$$

Le frequenze teoriche ottenute sono le seguenti

		Y: giudizio espresso					Marginali
		I	S	A	D	B	
X: anni di affiliazione	da 0 a 4	2.4	2.1	3	2.7	1.8	12
	da 4 a 8	3.2	3.8	4	3.6	2.4	16
	da 8 a 12	2.4	2.1	3	2.7	1.8	12
	Marginali	8	7	10	9	6	40

A questo punto è possibile valutare la convergenza dello stimatore di Pizzetti Pearson, possibile solo se tutte le frequenze teoriche sono superiori a 5. La condizione non è mai verificata pertanto non si dovrebbe procedere al calcolo.
Se l'ipotesi fosse stata verificata si avrebbe avuto la convergenza dello stimatore

$$\hat{n}_{i,j} > 5 \Rightarrow \sum_{i=1}^{M} \frac{\left(n_{i,j} - \hat{n}_{i,j}\right)^2}{\hat{n}_{i,j}} \sim \chi^2\left((M_x - 1)(M_y - 1)\right).$$

Poiché entrambi i caratteri della bi-variata hanno modalità $M_x = 3$ e $M_y = 5$, la regione di accettazione per un test al 5 % è la seguente.

$$A = \left[0\,;\, \chi^2_{0.95}(8)\right] = [0\,;\,15.507]$$

Non rimane che da calcolare il valore dello stimatore e verificare se appartiene alla regione di accettazione.
Il valore dello stimatore è

$$\sum_{i=1}^{3}\sum_{j=1}^{5} \frac{\left(n_{i,j} - \hat{n}_{i,j}\right)^2}{\hat{n}_{i,j}} = \frac{(5-2.4)^2}{2.4} + \frac{(2-2.1)^2}{2.1} + \ldots$$

se questo risulta interno ad A si accetta l'ipotesi di indipendenza altrimenti la si rifiuta.

Esercizio 3) *Il candidato stimi puntualmente e per intervallo il valore atteso degli anni di affiliazione uluzzando i dati riportati come carattere X nella tabella a doppia entrata dell'Esercizio 2.*
I dati coinvolti nell'esercizio in esame sono riportati nella seguente tabella ad entrata semplice.

i	*classe$_i$*	*c_i*	*n_i*	*c_i*n_i*	*c_i - μ*	*$(c_i - \mu)^2$*	*n_i*$(c_i - \mu)^2$*
1	da 0 a 4	2	12	24	-4	16	192
2	da 4 a 8	6	16	96	0	0	0
3	da 8 a 12	10	12	120	4	16	192
		Totale	40	240			384

Lo stimatore puntuale del valore atteso è la media campionaria (indicata con μ) che è data dalla somma delle osservazioni fratto la dimensione della statistica. Poichè i dati sono raccolti in m classi di modalità per procedere al calcolo della media è opportuno, per ogni classe i, calcolare il rispettivo valore di centro classe c_i. Utilizzando i conti riportati in tabella produce la seguente stima:

$$\hat{E[P]} = \mu = \frac{\sum_{i=1}^{m} c_i * n_i}{n} = \frac{240}{40} = 6$$

Per ottenere la stima per intervallo occorre fissare un livello di confidenza $1\text{-}\alpha = 95\%$; da cui si ricava $\alpha = 5\%$. La stima per intervallo è in caso che la varianza della popolazione sia ignota è data dalla

$$E[P] \in \left[\mu - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \;\; ; \;\; \mu + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$$

Ricordando la formula del calcolo della varianza campionaria ed utilizzando i conti riportati in tabella si ha che:

$$s^2=\sigma^2\frac{n}{n-1}=\left(\frac{\sum_{i=1}^{m} n_i*(c_i-\mu)^2}{n}\right)\frac{n}{n-1}=\frac{\sum_{i=1}^{m} n_i*(c_i-\mu)^2}{n-1}=\frac{384}{39} \qquad z_{1-\frac{\alpha}{2}}=1.96$$

Pertanto l'intervallo richiesto è: $E[P]\in\left[6-1.96\sqrt{\frac{384}{40*39}}\ ;\ 6+1.96\sqrt{\frac{384}{40*39}}\right]=[5.028\ ;\ 6.972]$

Esercizio 4)

a) calcoli le seguenti Probabilità: $P(E_1)$; $P(E_2)$; $P(E_1 \cup E_2)$; $P(E_1 | E_2)$; $P(E_2 | E_1)$;

L'evento E_1: prevede l'estrazione di un numero minore di *0.5* estraendo da una *Ber(0.6).* Poiche da una bernulliana sono estraibili solo i valori 0 o 1 la probabilità dell'evento richiesto è pari a quella di estrarre uno zero. Pertano si ha che

$$P(E_1) = q = 1-p = 1- 0.6 = 0.4$$

Per calcolare la probablità richiesta si deve prima rincondurre la normale in esame $y\sim N(1,1)$ a quella standardizzata $z\sim N(0,1)$.

$$z=\frac{y-E[Y]}{\sqrt{Var[Y]}}=y-1$$

da cui si ricava che il punto $y=0$ corrisponde al punto $z=-1$. Pertanto la probabilità richiesta diviene

$$P(E_2) = P(Y<0) = P(Z<-1) = 0.1587$$

Utilizzando la probabilità assiomatica si possono ricavare le altre probabilità richieste

$$P(E_1\cup E_2)=P(E_1)+P(E_2)-P(E_1\cap E_2)=0.4+0.1587-0.1=0.4587$$

$$P(E_1|E_2)=\frac{P(E_1\cap E_2)}{P(E_2)}=\frac{0.1}{0.1587}=0.6301$$

$$P(E_2|E_1)=\frac{P(E_1\cap E_2)}{P(E_1)}=\frac{0.1}{0.4}=0.25$$

b) indichi se gli eventi E_1 ed E_2 possono ritenersi dipendenti.

L'indipendenza statistica si ha quando le pobabilità condizionate corrispondono con le probabilità non condizionate.

$$P(E_1|E_2)=P(E_1) \quad \text{e} \quad P(E_2|E_1)=P(E_2)$$

Poichè questa condizione non è verificata si può asserire che gli eventi sono dipendenti.