- Appello del 4 Febbraio 2015 -

Esercizio 1)

Un ricercatore sta esaminando un farmaco sperimentale che dovrebbe ridurre gli effetti di una malattia dell'epidermide. Nello specifico dovrebbe ridurre l'estensione della micosi. A tale scopo si sono infettate 20 cavie e dopo una settimana si è cominciato il trattamento con il farmaco sperimentale. Dopo una settimana di trattamento si è misurata l'infezione nelle cavie ottenendo le seguenti misurazioni

2.30	7.10	2.30	4.20	6.60	5.90	9.00	7.80	9.40	2.90
9.20	6.10	4.90	2.60	7.90	5.70	6.70	7.90	4.30	7.20

Determinare

a) La tipologia del carattere.
b) Un indice sintetico di posizione.
c) Se possibile, un indice sintetico di variabilità.
d) Una rappresentazione grafica adeguata.
e) L'eventuale presenza di outlier.

Esercizio 2)

Il dato riportato al punto precedente è stato ulteriormente caratterizzato nella seguente tabella a doppia entrata (da completare vedi punto b) aggiungendo l'età delle cavie (all'inzio del trattamento).

		Y: estensione			Totale
		fino a 3	Fra 3 e 6	Fra 6 e 10	
X:eta in settimane	Da 2 a 4		**1**		**10**
	Da 4 a 6	**1**		**1**	**5**
	Da 6 a 8	**1**			
	Totale				

Il candidato

a) Indichi il tipo di carattere della serie bivariata.
b) Completi la tabella con i dati mancanti.
c) Se possibile, indichi e calcoli per la serie ottenuta un opportuno indice di posizione.
d) Se possibile, indichi e calcoli per la serie ottenuta un opportuno indice di variabilità.
e) Verifichi se vi è un legame fra la i due caratteri descritti.
Il candidato indichi e verifichi le ipotesi richieste per l'approccio scelto e proceda al calcolo anche qualora queste non siano soddisfatte.

Esercizio 3)

Utilizzando i dati presentati nell'esercizio 1 come campione, stimare per intervallo il valore atteso della seguente variabile casuale

P: area dell'infezione dopo una settimana di trattamento

Il candidato indichi e valuti le ipotesi richieste e proceda al calcolo della stima anche qualora non siano soddisfatte.

Esercizio 4)

Si considerino i seguenti eventi dichiarati indipendenti.

E_1 : estraendo una unità unità statistica fra quelle descritte nell'Esercizio 1, che abbia una estensione inferiore a 6

E_2: si ottenga $x<-2$ dove x è estratto da una v.c. distribuita come $N(-6,4)$

a) Il candidato 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) Il candidato indichi se i due eventi E_1 ed E_2 sono incompatibili.

- Appello del 4 Febbraio 2015 -

- svolgimento -

Esercizio 1)

a) Determinare la tipologia del carattere.

Il carattere è **quantitativo** in quanto composto da numeri **continuo** in quanto l'area è una grandezza di sua natura continua.

b)*Determinare un indice sintetico di posizione.*

Gli indici di posizione possibili sono 3: moda, media e mediana. Il primo, per la tipologia di dati in esame, non è significativo, in quanto le osservazioni raccolte non presentano elevate frequenze. Gli altri due indici sono richiesti nel prosegui dello svolgimento pertanto verranno calcolati entrambi.

La mediana è il valore che bipartisce le osservazioni: avendo 20 osservazioni la mediana sarà la media fra la decima è l'undicesima osservazione ordinata. Le osservazioni ordinate sono.

2.3	2.3	2.6	2.9	4.2	4.3	4.9	5.7	5.9	6.1
6.6	6.7	7.1	7.2	7.8	7.9	7.9	9	9.2	9.4

la mediana risulta quindi essere (6.1+6.6)/2 = 6.35.

La media è data dalla somma delle osservazioni divisa per la numerosità delle stesse; essa risulta essere:

$$\bar{x}=\frac{\sum_{i=1}^{N} x_i}{N}=\frac{120}{20}=6.0$$

c) *Determinare, se possibile, un indice sintetico di variabilità.*

Nel corso dello svolgimento verranno richiesti due indici di variabilità: la varianza e la distanza interquartile. Pertanto verranno calcolati entrambi in questa sede.

La distanza interquartile è la differenze fra il terzo ed il primo quartile (l'osservazione, rispettivamente preceduta e seguita, da un quarto delle restanti osservazioni ordinate): avendo 20 osservazioni si avrà.

$$q_1=\frac{x_5+x_6}{2}=\frac{4.2+4.3}{2}=4.25 \qquad q_3=\frac{x_{15}+x_{16}}{2}=\frac{7.8+7.9}{2}=7.85$$

Da cui *D= 7.85-4.25 = 3.6*

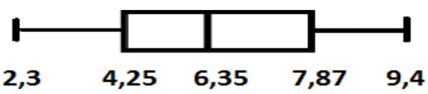
La varianza è la media del quadrato degli scarti dalla media, che risulta essere pari a.

$$\sigma^2=\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N}=\frac{100.76}{20}=5.038$$

c) *Determinare una rappresentazione grafica adeguata.*

Il box plot risulta essere una rappresentazione adeguata ad i dati in esame. Per poterlo calcolare serve riconoscere il valore adiacente inferiore (VAI) e quello superiore (VAS). Nel calcolo di questi indici è stato posto *k = 1*.

$$VAI=max(q_{0,}q_1-kD)=max(2.3\,,4.25-3.6)=max(2.3\,,0.65)=2.3$$
$$VAS=min(q_{4,}q_3+kD)=max(9.4\,,7.85+3.6)=min(9.4\,,11.45)=9.4$$



Il boxplot risulta essere quello diseganto a lato

d) *Determinare l'eventuale presenza di outlier.*

Non essendoci esterni all'interni all'intervallo *[VAI, VAS]* possiamo considerare la statistica priva di outlier.

Esercizio 2)

a) *Indichi il tipo di carattere della serie bivariata*

Essendo composta da due caratteri quantitativo continui, risulta essere una serie quantitativa continua.

b) Completi la tabella con i dati mancanti.

I totali di colonna si sono ricavati dai dati precedenti, le restanti frequenze si hanno dai totali di riga e di colonna.

		Y: estensione			Totale
		fino a 3	Fra 3 e 6	Fra 6 e 10	
X:eta in settimane	Da 2 a 4	2 (2)	**1** (2.5)	7 (5.5)	**10**
	Da 4 a 6	**1** (1)	3 (1.25)	**1** (2.75)	**5**
	Da 6 a 8	**1** (1)	1 (1.25)	3 (2.75)	5
	Totale	4	5	11	20

c) Se possibile, indichi e calcoli per la serie ottenuta un opportuno indice di posizione.

Essendo il carattere della bivariata quantitativo è possibile calcolare sia la moda che la media. Procediamo al calcolo di quest'ultima. La media si può calcolare componente per componente. Ricordando che per dati raccolti in classe la classe viene sostituita con il solo valore di centro classe si ottengono i seguenti valori:

$$\bar{x}=\frac{\sum_{i=1}^{Mx} cx_i * n_i}{N}=\frac{3*10+5*5+7*5}{20}=\frac{30+25+35}{20}=\frac{90}{20}=4.5$$

$$\bar{y}=\frac{\sum_{i=1}^{My} cy_i * n_i}{N}=\frac{1.5*4+4.5*5+8*11}{20}=\frac{6+22.5+88}{20}=\frac{116.5}{20}=5.825\sim 5.8$$

Pertanto la media risulta essere *(4.5 , 5.825)*

d)Se possibile, indichi e calcoli per la serie ottenuta un opportuno indice di variabilità.

Per i dati in esame si procede al calcolo della matrice varianza-covarianza. Si ricorda che anche in questo caso nel calcolo la classe viene sostituita con il valore di centro classe. Onde facilitare il compito si è annotata la precedente tabella a doppia entrata in modo da riportare i valori di centro classe e gli scarti delle modalità dei singoli caratteri ed inserendo nel corpo della tabella i prodotti degli scarti .

					Y: estensione			
					classe	fino a 3	Fra 3 e 6	Fra 6 e 10
					Centro classe	1.5	-1.3	8.0
	classe	Centro classe	Scarto	Scarto al quadrato	scarto	-4.30	1.69	2.20
					Scarto al quadrato	18.49	2.86	4.84
X:età in settimane	Da 2 a 4	3	-1.5	2.25		*6.45*	*1.95*	*-3.3*
	Da 4 a 6	5	0.5	0.25		*-2.15*	*0.85*	*1.1*
	Da 6 a 8	7	2.5	6.25		*-10.75*	*4.23*	*5.5*

Procedendo al calcolo delle varianze

$$\sigma_x^2=\frac{\sum_{i=1}^{Mx} n_i(cx_i-\bar{x})^2}{N}=\frac{10(2.25)+5(0.25)+5(6.25)}{20}=\frac{22.5+1.25+31.25}{20}=2.75$$

$$\sigma_y^2=\frac{\sum_{i=1}^{My} n_i(cy_i-\bar{y})^2}{N}=\frac{4(18.49)+5(2.86)+11(4.48)}{20}=\frac{73.96+14.3+49.28}{20}=6.87$$

e della covarianza $\sigma_{xy}=\frac{\sum_{i=1}^{Mx}\sum_{j=1}^{My} n_{i,j}(cx_i-\bar{x})(cy_j-\bar{y})}{N}$

$$\sigma_{xy}=\frac{2(6.45)+1(1.95)+7(-3.3)+1(-2.15)+3(-0.85)+1(1.1)+1(-10.75)+1(4.23)+3(5.5)}{20}=-\frac{1.87}{20}=-0.0935$$

La matrice richiesta risulta essere $\Sigma=\begin{bmatrix} 2.75 & -0.0935 \\ -0.0935 & 6.87 \end{bmatrix}$

e) Verifichi se vi è un legame fra la i due caratteri descritti.

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 sufficiente mente vicine (ad un determinato livello di significatività) a quelle teoriche ottenute dall'ipotesi di indipendeza. Il test viene fatto sfruttando la distribuzione limite dello stimatore di Pizzetti Pearson che viene ad essere un chi quadranto avente gradi di libertà paria 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 marginali sono state inserite fra parentesi nella tabella al punto b)

A questo punto è possibile valutare la convergenza dell stimatore di Pizzetti Pearson, possibile solo se tutte le frequenze teoriche sono superiori a 5. Constatato che la condizione non è verificata il testo richiede di procedere. Si valuta quindi al calcolo della regione di accettazione fissato il livello di significatività al 5%.

$$A=\left[0\,;\chi^2_{1-\alpha}\left((M_x-1)(M_y-1)\right)\right]=\left[0\,;\chi^2_{1-0.05}\left((3-1)(3-1)\right)\right]=\left[0\,;\chi^2_{0.95}(4)\right]=[0\,;9.49]$$

Si può ora procedere al calcolo dello stimatore vero e proprio

$$\frac{\sum_{i=1}^{3}\sum_{j=1}^{3}\left(n_{i,j}-\hat{n}_{i,j}\right)^2}{\hat{n}_{i,j}}=\frac{(2-2)^2}{2}+\frac{(1-2.5)^2}{2.5}+\frac{(7-5.5)^2}{5.5}+\frac{(1-1)^2}{1}+\frac{(3-1.25)^2}{1.25}+\frac{(1-2.75)^2}{2.75}+\frac{(1-1)^2}{1}+\frac{(1-1.25)^2}{1.25}+\frac{3.0625}{2.75}=$$

$$0+\frac{2.25}{2.5}+\frac{2.25}{5.5}+0+\frac{3.0625}{1.25}+\frac{3.0625}{2.75}+0+\frac{0.0625}{1.25}+\frac{0.0625}{2.75}=0.9+0.41+2.5+1.17=4.98$$

Poichè il valore dello stimatore è interno all'intervallo di accettazione potrei dire, se fossero rispettate le ipotesi del test, che i due caratteri sono indipendenti ad un livello di significatività del 5 per cento e che quindi l'età non influisce sul efficacia del trattamento.

Esercizio 3)

Le tecniche di stima viste nel corso prevedono che:

- la popolazione sia descrivibile mediante una variabile casuale,
- che il campione abbia una numerosità tale da far convergere lo stimatore e
- che le prove siano indipendenti ed identicamente distribuite (i.i.d.).

Nel caso in esame

- il testo fornisce la variabile da utilizzare.
- la grandezza da stimare risulta *E[P]* il cui stimatore è la media campionaria la quale converge in legge per campioni avente numerosità superiore a 30 (ipotesi non confermata).
- L'ipotesi di prove i.i.d. è difficilmente valutabile con le informazioni che si hanno a disposizione.

La stima puntuale consiste nel calcolo dello stimatore utilizzando i dati in esame. Il calcolo è già stato effettuato nel primo esercizio ottenendo:

$$\hat{E[P]}=\bar{x}=6$$

Le stime per intervallo sono regolate dal livello di confidenza (*1-α*) che solitamente è fissato da chi svolge l'analisi dei dati. Nel caso in esame una scelta valida è porre *α = 5%*. Determinato il livello di confidenza la stima *I* è data dalla seguente formula:

$$I=\left[m-z_{1-\frac{\alpha}{2}}\sqrt{\frac{Var[P]}{n}}\ \ ;\ \ m+z_{1-\frac{\alpha}{2}}\sqrt{\frac{Var[P]}{n}}\right]$$

Si noti come non sia nota *Var[P]* che deve essere stimata mediante il suo stimatore corretto: la varianza campionaria, utilizzando i conti riportati nel primo esercizio si ha che:

$$\hat{Var[P]}=s^2=\sigma^2\frac{n}{n-1}=5.038\frac{20}{19}=5.303$$

La stima risulta quindi essere:

$$I=\left[6-1.96\sqrt{\frac{5.303}{20}}\ \ ;\ \ 6+1.96\sqrt{\frac{5.303}{20}}\right]=[6-1.009\,;6+1.009]=[4.991\,;7.009]$$

Esercizio 4)

La probabilità delle evento E_1 è ottenibile usando la definizione di probabilità classica:

$$P(E_1)=\frac{esiti\ favorevoli}{esiti\ possibil}=\frac{|E_1|}{|U|}=\frac{9}{20}=0.45$$

Per calcolare $P(E_1)$ si deve rincondurre la normale in esame *y~N(-6,4)* a quella standardizzata *z~N(0,1)*.

$$z=\frac{x-E[X]}{\sqrt{Var[X]}}=\frac{x+6}{\sqrt{4}}=\frac{x+6}{2}$$

da cui si ricava che il punto *x=-2* corrisponde al punto *z=1*. Pertanto la probabilità richiesta diviene

$$P(E_2) = P(X<-2) = P(Z<2) = 0.9772$$

Utilizzando la probabilità assiomatica e ricordando che per eventi indipendenti la probabilità dell'evento intersezione è pari al prodotto delle probabilità, si possono ricavare le altre probabilità richieste

$$P(E_1\cup E_2)=P(E_1)+P(E_2)-P(E_1\cap E_2)=0.45+0.9772-0.45*0.9772=0.98746$$

$$P(E_1|E_2)=\frac{P(E_1\cap E_2)}{P(E_2)}=\frac{P(E_1)P(E_2)}{P(E_2)}=P(E_1)=0.45 \qquad P(E_2|E_1)=\frac{P(E_1\cap E_2)}{P(E_1)}=\frac{P(E_1)P(E_2)}{P(E_1)}=P(E_2)=0.9772$$

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

Due eventi incompatibili non possono verificarsi contemporaneamente, pertanto la probabilità dell'evento intersezione è nulla. Poichè nel caso in esame questa è positiva, si può asserire che gli eventi sono compatibili.