- Appello del 3 Settembre 2014 -

Esercizio 1)

Per verificare l'efficacia del farmaco *Y* (per trattare i tumori al fegato) si è condotto il seguente test[1]. Si sono infettate 100 cavie esponendo gli animali alla massima dose tollerata (MDT) dell'idoneo composto cancerogeno. Dopo sette giorni l'intera popolazione ha cominciato il trattamento con *Y* che è perdurato per 3 settimane. Alla fine della 4ª settimana (1 di incubazione + 3 di trattamento) tutta la popolazione è stata soppressa mediante elettroshock (di modo da preservare il fegato) e si è proceduto a misurare l'estensione della massa tumorale nel fegato; ottenendo la seguente distribuzione del volume:

Volume	0 – 200	200 – 250	250 – 300	300 – 350	350 – 400	400 – 500	500 – 700	700 – 1100
Frequenza	20	14	16	12	14	12	8	4

Il candidato

a) determini la tipologia del carattere;
b) fornisca una rappresentazione grafica dei dati;
c) indichi e descriva tutti gli indici di posizioni adeguati ai dati e ne calcoli almeno uno;
d) se possibile, determini la varianza della statistica riportata.

Esercizio 2)

Il candidato, usando come campione i dati descritti nell'Esercizio 1, stimi puntualmente e per intervallo la varianza della V.C.

R: volume di una massa tumorale dopo tre settimane di trattamento con il farmaco Y in una cavia malata

Il candidato evidenzi e valuti le ipotesi necessarie e proceda al calcolo anche quando queste non siano verificate.

Esercizio 3)

Sapendo che la dimensione di un tumore in una cavia a 4 settimane dall'inoculazione si distribuisce come una v. c. normale[2] *P* di valore atteso di 360 mm^2 e varianza di 10^6 mm^4 si verifichi se sia possibile l'ipotesi che il farmaco *Y* riduca l'aggressività del tumore (il candidato motivi il test da lui scelto). Il candidato indichi e verifichi le ipotesi richieste per l'approccio scelto e proceda al calcolo anche qualora queste non siano soddisfatte.

Esercizio 4)

Si considerino i seguenti eventi dichiarati indipendenti.

E_1 : si ottenga una palla rossa estraendo una pallina con 3 biglie verdi ed una rossa
E_2: si ottenga $x>0$ dove x è estratto da una v. c. distribuita come $Ber(0.75)$

a) Il candidato calcoli le seguenti Probabilità $P(E_1)$; $P(E_2)$; $P(E_1 \cup E_2)$ $P(E_1 | E_2)$.
b) Il candidato indichi se i due eventi E_1 ed E_2 sono complementari motivando la risposta.

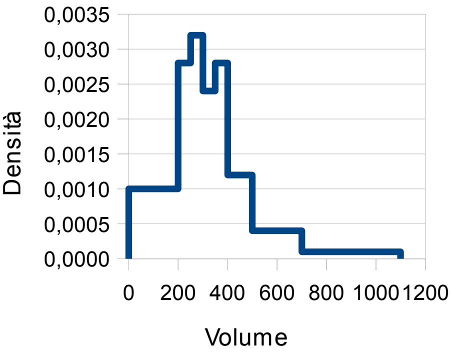
1 Il procedimento è indicativo e non rigoroso.
2 Dato fittizio.

- Appello del 3 Settembre 2014 - Svolgimento

Esercizio 1)

a) determini la tipologia del carattere;

Le osservazioni in quanto misure di volume sono rappresentabili con numeri reali, per cui il carattere è quantitativo continuo.



b) fornisca una rappresentazione grafica dei dati;

Essendo la statistica relativa ad un carattere quantitativo continuo e fornita già raggruppata in classi la scelta migliore per rappresentare il grafico in esame è l'istogramma. Per procedere alla rappresentazione dell'istogramma e necessario calcolare la densità di frequenza delle singole classi di modalità secondo la seguente formula.

$$d_i = \frac{f_i}{sup_i - inf_i}$$

Utilizzando i conti riportati in Tabella 1 si è ottenuto il grafico a lato.

c) indichi e descriva tutti gli indici di posizioni adeguati ai dati e ne calcoli almeno uno;

Gli indici di posizioni posizione indicano il valore centrale della statistica e sono:

- *Moda*: modalità avente massima frequenza.
- *Mediana*: Corrisponde al secondo quartile, ovvero l'osservazione che biparisce la statistica.
- *Media*: Somma delle osservazioni fratto la loro numerosità.

Sono tutti calcolabili con la statistica in esame. Utilizzando i conti in tabella si è calcolata la media ($\bar{x}$) che vale 325 mm^2.

d) se possibile, determini la varianza della statistica riportata.

Come accade per la media, anche il calcolo della varianza può venir esteso alle statistiche raccolte in classi. Basta introurre il valore di centro classe (ovvero il punto medio dell'intervallo caratterizzante la classe). Utilizzando i dati riportati in Tabella 1 si ha che:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{M} n_i(\bar{x}-\bar{c}_i)^2 = \frac{1}{100}3342500 = 33425$$

Classe (i)	Inf_i	Sup_i	Sup_i - Inf_i	$\bar{c}_i$	n_i	f_i	d_i	$\bar{c}_i*f_i$	$\bar{x}$ - $\bar{c}_i$	$(\bar{x}-\bar{c}_i)^2$	$n_i(\bar{x}-\bar{c}_i)^2$
1	0	200	200	100	20	0.20	0.00100	2000.00	-225	50625	1012500
2	200	250	50	225	14	0.14	0.00280	3150.00	-100	10000	140000
3	250	300	50	275	16	0.16	0.00320	4400.00	-50	2500	40000
4	300	350	50	325	12	0.12	0.00240	3900.00	0	0	0
5	350	400	50	375	14	0.14	0.00280	5250.00	50	2500	35000
6	400	500	100	450	12	0.12	0.00120	5400.00	125	15625	187500
7	500	700	200	600	8	0.08	0.00040	4800.00	275	75625	605000
8	700	1100	400	900	4	0.04	0.00010	3600.00	575	330625	1322500
				Totale	100	1		32500.00			3342500

Esercizio 2)

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 *Var[X]* il cui stimatore è la varianza campionaria la quale converge in legge per campioni avente numerosità superiore a 30 (ipotesi confermata).
- L'ipotesi di prove i.i.d. è difficilmente valutabile con le informazioni che si hanno a disposizione.

La stima puntuale può essere effettuata ricordando che la varianza viene stimata correttamente mediante la varianza campionaria. Parte del calcolo è già stata effettuata nella tabella riportata nello svolgimento del primo esercizio, utilizzando quei conti si ottiene:

$$\hat{Var}[P] = s^2 = \frac{n}{n-1}\sigma^2 = \frac{100}{99}\sigma^2 = 33763$$

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 = 10%*. Determinato il livello di confidenza la stima *I* è data dalla seguente formula:

$$I=\left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)};\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right]=\left[\frac{99\cdot 33763}{\chi^2_{0.95}(100)};\frac{99\cdot 33763}{\chi^2_{0.05}(100)}\right]=\left[\frac{3342500}{126.58};\frac{3342500}{71.42}\right]=[26428;46800]$$

Si noti come le tavole non consentano il calcolo del chi-quadro a 99 gradi di libertà. In questo caso si può utilizzare la convergenza in legge della distribuzione chi-quadro a quella normale. Si ha infatti che per un numero di gradi libertà sufficientemente elevato

$$\chi^2(\nu)\sim N(\nu,2\nu)\Rightarrow\chi^2(99)\sim N(99,198)$$

Pertanto i valore richiesti posson essere recuperati usando la distribuzione normale. Ovviamente le tavole consento l'uso della sola normale standard. Pertanto si recupera il valore che lascia nella coda della normale la probabilità richiesta e lo si destandardizza. Si ottengono quindi i seguenti valori

$$\chi^2_{0.95}(99)=N_{0.95}(99;198)=Z_{0.95}*Var[N(99;198)]+E[N(99;198)]=1.96*14.07+99=126.58$$

$$\chi^2_{0.05}(99)=N_{0.05}(99;198)=Z_{0.05}*Var[N(99;198)]+E[N(99;198)]=-1.96*14.07+99=71.42$$

Esercizio 3)

Uno modo semplice per motivare l'asserzione con i dati in esame è quello di determinare se il valore atteso del volume della massa neoplastica sia minore assumendo il farmaco rispetto al normale decorso del tumore. Pertanto, utilizzando le vv. cc.

R: volume di una massa tumorale dopo tre settimane di trattamento con il farmaco Y in una cavia malata

P: volume di una massa tumorale dopo un mese dall'inoculazione del tumore

si potrebbe fare un semplice test sul valore atteso contrapponendo le due ipotesi

$$H_0: E[R]=E[P] \quad \text{e} \quad H_1: E[R]<E[P]$$

ad un adeguato livello di significatività *α*, ad esempio il *10%.*

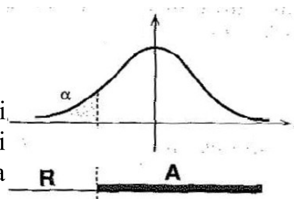
Il test verte sullo stimatore del valore atteso che è la media campionaria; le tecniche di stima viste nel corso prevedono che:

a) il campione abbia una numerosità tale da far convergere lo stimatore (30) e

b) le prove siano indipendenti ed identicamente distribuite (i.i.d.).

Entrambe le ipotesi sono state discusse nell'esercizio precedente, pertanto non si ripeteranno le stesse considerazioni. Nel caso lo stimatore converga questo si distribuirebbe come

$$\bar{x}\sim N\left(E[P],\frac{Var[P]}{n}\right)\sim N(360,10^4)$$



Determinata la distribuzione limite dello stimatore, è possibile determinare la regione di accettazione *A*. Il test in esame è di tipo unilaterale sinistro. Fissato un livello di significatività al 10 % si ha che la regione di accettazione non comprende il 10% della coda di sinistra. Utilizzando il processo di standardizzazione di ha che il valore limite è:

$$N_{0.10}(360;10^4)=Z_{0.10}*Var[N(360;10^4)]+E[N(360;10^4)]=-1.28*100+360=-128+360=232$$

cui corrisponde la regione di accettazione

$$A=[232;+\infty]$$

Il valore dello stimatore (la media campionaria) è stato calcolato nel primo esercizio e corrisponde a 325 che ricade nella regione di accettazione dell'ipotesi nulla. Pertanto, se le ipotesi (a e b) fossero valide, potremmo accettare l'ipotesi nulla (H_0) ovvero che la distribuzione empirica ha lo stesso valore atteso di quella data ad un livello si significatività del 5% e quindi che il farmaco non riduce l'incidenza tumorale in maniera significativa..

Esercizio 4)

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

$P(E_1)$: La probabilità richiesta è ottenibile usando la definizione di probabilità classica:

$$P(E_1)=\frac{esiti\ favorevoli}{esiti\ possibil}=\frac{|E_1|}{|U|}=\frac{1}{4}=0.25$$

$P(E_2)$: L'evento E_2 prevede l'estrazione di numero strettamente positivo da una *Ber(0.8).* Poiché in una bernoulliana solo due esiti sono possibili (0 e 1) tale evento coincide con il verificarsi dell'esito 1. In una bernoulliana questa probabilità è chiamata p e corrisponde al parametro della bernoulliana p . Pertanto si ha che

$$P(E_2) = p = 0.75$$

Ricordando che per due eventi indipendenti la probabilità dell'evento intersezione è dato dal prodotto delle probabilità e che le probabilità condizionate coincidono con quelle non condizionate si ha che

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) = \frac{1}{4} + \frac{3}{4} - \frac{1}{4} * \frac{3}{4} = 1 - \frac{3}{16} = \frac{13}{16} = 0.8125$$

$$P(E_1 | E_2) = P(E_1) = 0.25$$

$$P(E_2 | E_1) = P(E_2) = 0.75$$

b) Il candidato indichi se i due eventi E_1 ed E_2 sono complementari motivando la risposta.
Due eventi sono complementari se quando il non verificarsi di uno implica il verificarsi dell'altro. Pertanto due eventi indipendenti non possono essere complementari.