- Appello del 8 Luglio 2015 -

Esercizio 1)
Per verificare l'efficacia di un diserbante (denominato *H-725)* si sono preparate dieci colture di un parassita e, a coltura ultimata, si sono trattati i vetrini con il diserbante. Si è studiato dopo quanto ore il parassita è stato completamente debellato dal vetrino. Si è ottenuta le la seguente statistica:

1995 1999 2005 1970 2009 2050 1996 1995 1982 2009

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 presenza di outlier.

Esercizio 2)
Il candidato, usando come campione i dati descritti nell'Esercizio 1, stimi puntualmente e per intervallo il valore atteso della V.C.

R: numero di ore in cui il diserbante H-725 rimuove la coltura parassitaria da un vetrino

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

Esercizio 3)
Studi analoghi condotti sulla precedente versione del diserbante (H-724) avevano mostrato come la variabile

T: numero di ore in cui il diserbante H-724 rimuove la coltura parassitaria da un vetrino

si distribuisse come una v. c. normale di valore atteso pari a 2000 e varianza pari a 100.
Si verifichi se sia possibile affermare che la qualità del prodotto sia migliorata passando alla nuova versione. Il candidato motivi il test da lui scelto, inoltre indichi e verifichi le ipotesi richieste per l'approccio scelto procedendo al calcolo anche qualora queste non siano soddisfatte.

Esercizio 4)
Si considerino i seguenti eventi legati allo stesso esperimento: estrarre una ed una sola biglia da un urna contente 10 palline 4 rosse, 4 nere e 2 blu.

E_1 : si ottenga una palla rossa
E_2: si ottenga una palla nera

a) Il candidato calcoli le seguenti Probabilità $P(E_1)$; $P(E_2)$; $P(E_1 \cup E_2)$ $P(E_1 | \overline{E}_2)$.
b) Il candidato indichi se i due eventi E_1 ed E_2 siano complementari motivando la risposta.

- Appello del 8 Luglio 2015 - Svolgimento

i	o_i	o_i-m	$(o_i\text{-}m)^2$
1	1970	-31	961
2	1982	-19	361
3	1995	-6	36
4	1995	-6	36
5	1996	-5	25
6	1999	-2	4
7	2005	4	16
8	2009	8	64
9	2009	8	64
10	2050	49	2401
	20010		3968

Tabella 1 – Dati esercizio 1

Esercizio 1)

a) determini la tipologia del carattere;

Le osservazioni in quanto misure di numeri di cicli sono rappresentabili con numeri naturali, per cui il carattere è quantitativo discreto.

b) fornisca una rappresentazione grafica dei dati;

d) se possibile determini la presenza di outlier;

Essendo la statistica relativa ad un carattere quantitativo discreto e dovendo successivamente recuperare gli outlier una buona scelta per la rappresentazione è il box-plot. Sfruttando le osservazioni ordinate riportate in tabella 1, diviene facile ricavare i quartili ed i valori adiacenti necessari per la visualizzazione del box-plot.

- Q_0 = osservazione minima = o_1 = 1970
- Q_1 = osservazione preceduta dal 25% delle rimanenti osservazioni= $\frac{o_3+o_4}{2}=1995$
  Il quartile viene calcolato come media in quanto non è possibile rispettare a pieno la definizione. Infatti, le osservazioni sono 10, ridotte a 9 se si esclude il quartile, l'uso della definizione richiederebbe l'utilizzo dell'osservazione $\frac{N-1}{4}+1=3.25$ come quartile. Poiché questo non è possibile si utilizza la media fra la terza e la quarta osservazione. Lo stesso ragionamento si applica a tutti i restanti quartili
- Q_2 = osservazione preceduta dalla metà delle rimanenti osservazioni= $\frac{o_5+o_6}{2}=\frac{1996+1999}{2}=1997.5$
- Q_3 = osservazione preceduta dal 75% delle rimanenti osservazioni= $\frac{o_7+o_8}{2}=\frac{2005+2009}{2}=2007$
- Q_4 = osservazione massima = o_{10} = 2050
- VAI = valore adiacente inferiore = $q_1-K\cdot(q_3-q_1)=1995-2\cdot(2007-1995)=1971$
- VAS = valore adiacente superiore = $q_3+K\cdot(q_3-q_1)=2007+2\cdot(2007-1995)=2031$

Nel calcolo si rilevano due valori esterni all'intervallo [VAI; VAS] che vengono pertanto definiti outlier. (1970 e 2050).

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 è facile ricavare la media (m)

$$m=\frac{1}{N}\sum_{i=1}^{N}o_i=\frac{1}{10}20010=2001$$

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 *E[R]* 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, in ogni caso la procedura è se non altro coerente.

La stima puntuale consiste nel calcolo dello stimatore utilizzando i dati in esame. Il calcolo è già stata effettuata nella tabella riportata nello svolgimento del primo esercizio, utilizzando quei conti si ottiene:

$$\hat{E[R]}=m=2001$$

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:

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

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

$$\hat{Var}[P]=s^2=\frac{\sum_{i=1}^{n}(o_i-m)^2}{n-1}=\frac{3968}{9}=440.89$$

La stima risulta quindi essere:

$$I=\left[2001-1.96\sqrt{\frac{440.89}{10}}\ ;\ 2001+1.96\sqrt{\frac{440.89}{10}}\right]=[2001-13.01\,;2001+13.01]=[1987.99\,;2014.01]$$

Esercizio 3)

Uno modo semplice per motivare l'asserzione con i dati in esame è quello di determinare se il valore atteso del tempo richiesto dal fertilizzante per "disinfestare" un vetrino sia diminuito o meno. Pertanto, utilizzando la v. c. R introdotta al punto precedente si potrebbe fare un semplice test sul valore atteso contrapponendo le due ipotesi

$$H_0: E[R]=2000 \quad \text{e} \quad H_1: E[R]<2000$$

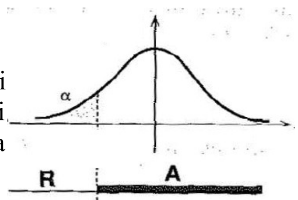
ad un adeguato livello di significatività *α*, ad esempio il *10%.* Il test verte sullo stimatore del valore atteso che è la media campionaria; le tecniche di stima viste nel corso prevedono che:

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

$$\bar{x}\sim N\left(E[P],\frac{Var[P]}{n}\right)\sim N(2000,10)$$



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.

Il valore dello stimatore (la media campionaria) è stato calcolato nel primo esercizio e corrisponde a 2001 che esendo superiore al valore atteso di riferimento (2000) non può cadere nella regione di rifiuto, indipendentemente dal valore livello di significatività. Pertanto, se le ipotesi (a e b) fossero valide, potremmo accettare l'ipotesi nulla (H_0) ovvero che il valore atteso della statistica è compatibile con quello dichiarato e quindi che la qualità del prodotto non risulta migliorata.

Esercizio 4)

a) calcoli le seguenti Probabilità: $P(E_1)$; $P(E_2)$; $P(E_1 \cup E_2)$ $P(E_1 | \bar{E}_2)$;

$P(E_1)$, $P(E_2)$, $P(E_1 \cup E_2)$: Le probabilità richieste sono ottenibile usando la definizione di probabilità classica:

$$P(E_1)=\frac{esiti\ favorevoli}{esiti\ possibil}=\frac{|E_1|}{|U|}=\frac{4}{10}=0.4 \qquad P(E_2)=\frac{esiti\ favorevoli}{esiti\ possibil}=\frac{|E_2|}{|U|}=\frac{4}{10}=0.4$$

L'evento $E_1 \cup E_2$ indica l'estrazione di una palla rossa o nera (ricordo che l'esperimento su cui insistono E_1 ed E_2 è l'estrazione di una ed una sola pallina). Pertanto si ha che:

$$P(E_1\cup E_2)=\frac{esiti\ favorevoli}{esiti\ possibil}=\frac{|E_1\cup E_2|}{|U|}=\frac{8}{10}=0.8$$

$P(E_1 | \bar{E}_2)$: come prima cosa si deve vedere a cosa corrisponda l'evento $\bar{E}_2$ che indica il fatto che non si sia verificato l'evento E_2 pertanto è la probabilità che non si sia estratta una pallina nera. In sostanza l'evento da considerare è

$E_1 | \bar{E}_2$: si è estragga una pallina rossa, sapendo che non è stata estratta una pallina nera.

La probabilità condizionante ci dice che delle 10 palline originali solo 6 sono valide (le 4 rosse e le due blu) di queste i casi favorevoli sono 4 (le palline rosse). Pertanto si ha

$$P(E_1|\bar{E}_2)=\frac{esiti\ favorevoli}{esiti\ possibil}=\frac{4}{6}=0.6667$$

Lo stesso risultato si poteva ottenere applicando la probabilità assiomatica.

$$P(\bar{E}_2)=1-P(E_2)=1-0.4=0.6 \qquad P(E_1\cap\bar{E}_2)=\frac{esiti\ favorevoli}{esiti\ possibil}=\frac{4}{10}$$

$$P(E_1|\bar{E}_2)=\frac{P(E_1\cap\bar{E}_2)}{P(\bar{E}_2)}=\frac{0.4}{0.6}=\frac{4}{6}=0.6667$$

b) Il candidato indichi se i due eventi E_1 ed E_2 sono complementari motivando la risposta.

Due eventi sono complementari quando il non verificarsi di uno implica il verificarsi dell'altro e viceversa. Essendo le biglie nell'urna di tre colori il sapere che non è stata estratta una biglia rossa, non implica che sia stata estratta una biglia nera, infatti potrebbe essere stata estratta una biglia blu. Pertanto i due eventi non sono complementari.