- Appello del 18 Febbraio 2015 -

Esercizio 1)

Un ricercatore sta esaminando un farmaco sperimentale che dovrebbe ridurre la durata della sindrome influenzale. A tale scopo ha somministrato il nuovo farmaco sono stati selezionati 30 soggetti affetti da raffreddore ed ha somministrato loro il farmaco. Infine ha monitorato dopo quante ore sono cessati i sintomi, ottenendo i seguenti dati.

78	70	77	60	51	60	73	81	65	66
67	55	70	70	65	76	65	60	68	89
55	54	71	98	99	80	83	55	66	73

Il candidato

a) Determini la tipologia del carattere.
b) Raccolga le osservazioni in classi scegliendone un numero opportuno.
c) Determini, se possibile, un indice sintetico di posizione opportuno.
d) Determini, se possibile, un indice sintetico di variabilità.
e) Tracci l'istogramma.

Esercizio 2)

Il candidato, utilizzando i dati presentati nell'Esercizio 1 come campione, stimi puntualmente e per intervallo il valore atteso della seguente variabile casuale

P: durata della sintomatologia di una sindrome influenzare trattata con il nuovo farmaco

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

Esercizio 3)

I dati di letteratura indicano che il raffreddore in media passa dopo 3 giorni (ovvero 72 ore). Il candidato indichi, utilizzando i dati riportati nel primo esercizio, se il farmaco sia utile nella cura del raffreddore. Il candidato indichi e verifichi le ipotesi necessarie e proceda al calcolo anche se queste non fossero verificate.

Esercizio 4)

Si considerino i seguenti eventi dichiarati incompatibili.

E_1: estraendo una unità unità statistica fra quelle descritte nell'Esercizio 1, che mostri un decorso superiore a 80 ore.

E_2: si ottenga $-2<x<0$ dove x è estratto da una v.c. distribuita come $N(-6,4)$

Il candidato

a) calcoli le seguenti Probabilità $P(E_1)$; $P(E_2)$; dell'evento unione.

b) indichi cosa sarebbe successo se gli eventi fossero stati dichiarati indipendenti

- Appello del 18 Febbraio 2015 -

- svolgimento -

Esercizio 1)

a) Determinare la tipologia del carattere.

Il carattere è **quantitativo** in quanto composto da numeri **continuo** in quanto il tempo è una grandezza di sua natura continua.

b) Raccogliere il dato in classi scegliendone un numero opportuno.

Sebbene non esista una regola per decidere quali e quante classi utilizzare per raccogliere delle osservazioni continue, una indicazione può essere data dalle seguente formula empirica.

$$\log_2(N)+1\simeq 6$$

La formula indica sei classi ma per motivi di comodità (avere classi di ampiezza pari a dieci) si è deciso di raccogliere in 5 classi equispaziate fra 40 e 100, ottenendo la tabella indicata a lato.

i	inf_i	sup_i	n_i	$sup_i - inf_i$	f_i	d_i
1	50	60	8	10	0.267	0.0267
2	60	70	10	10	0.333	0.0333
3	70	80	7	10	0.233	0.0233
4	80	90	3	10	0.100	0.0100
5	90	100	2	10	0.067	0.0067
			30		1	

c)*Determinare un indice sintetico di posizione.*

Gli indici di posizione possibili sono 3: moda, media e mediana. Viste le richieste dell'enunciato (la stima per intervallo dell'esercizio seguente) si è scelto di calcolare la media. La media è data dalla somma delle osservazioni divisa per la numerosità delle stesse; essa risulta essere:

$$\bar{x}=\frac{\sum_{i=1}^{N} x_i}{N}=\frac{2100}{30}=70$$

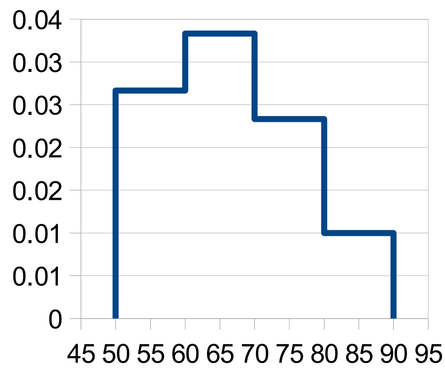
d) *Determinare, se possibile, un indice sintetico di variabilità.*

Nel corso dello svolgimento verrà richiesto il calcolo della varianza che pertanto è stata scelta come indice di variabilità. La varianza è la media del quadrato degli scarti dalla media, che risulta essere pari a.

$$\sigma^2=\frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}{N}=\frac{426}{30}=141.87$$

e) *Se possibile tracciare l'istogramma*

L'istogramma è una rappresentazione comunemente utilizzata quando si tratta un dato quantitativo continuo che, per diverse esigenze, viene rappresentato in classi di modalità. Il grafico riporta le modalità sull'asse delle ascisse e sulle ordinate la densità di frequenza di ogni classe. Il grafico si compone di rettangoli fra di loro adiacenti la cui base si ricava dagli estremi della classe mentre l'altezza, l'altezza coincide con la densità di frequenza, pertanto l'area di ogni rettangolo sarà uguale alla frequenza relativa della classe. A lato si riporta l'istogramma richiesto. I conti per ricavare il suddetto istogramma sono riportati nella tabella creata al punto b)



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[P]* il cui stimatore è la media 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 consiste nel calcolo dello stimatore utilizzando i dati in esame. Il calcolo è già stato effettuato nel primo esercizio ottenendo:

$$\hat{E[P]}=\bar{x}=70$$

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:

$$\hat{Var}[P]=s^2=\sigma^2\frac{n}{n-1}=\frac{4256}{29}=146.76$$

La stima risulta quindi essere:

$$I=\left[70-1.96\sqrt{\frac{146.76}{30}}\ ;\ 70+1.96\sqrt{\frac{146.76}{30}}\right]=[70-2.21\,;70+2.21]=[67.79;72.21]$$

Esercizio 3)

Possiamo definire che un farmaco sia utile se mediamente riduce la durata della malattia. Pertanto l'informazione richiesta può essere ottenuta mediante un test di ipotesi in cui si controllano le seguenti ipotesi

$$H_0:E[P]=72 \quad \text{e} \quad H_1:E[P]<72$$

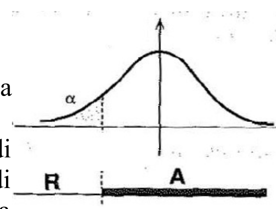
Nel caso si accetti l'ipotesi alternativa si conclude che il farmaco è utile. Il test verte sullo stimatore del valore atteso che è la media campionaria; le tecniche di stima viste nel corso prevedono che:

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

$$\bar{x}\sim N\left(E[P],\frac{Var[P]}{n}\right)\sim N\left(E[P],\frac{s^2}{n}\right)\sim N\left(72,\frac{146.76}{30}\right)\sim N(72,4.892)$$



Si noti come nella formula precedente la varianza della v.c. P che risulta ignota sia stata stimata utilizzando la varianza campionaria.
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 si ha che il valore limite è:

$$N_{0.10}(72,4.892)=Z_{0.10}*\sqrt{Var[N(72,4.892)]}+E[N(72,4.892)]=-1.28*\sqrt{4.892}+72=-2.83+70=69.17$$

cui corrisponde la regione di accettazione

$$A=[69.17\,;+\infty]$$

Il valore dello stimatore (la media campionaria) è stato calcolato nel primo esercizio e corrisponde a 70 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 il valore atteso della statistica è compatibile con quello del decorso normali dell'influenza. Potremmo quindi affermare che il farmaco non è utile.

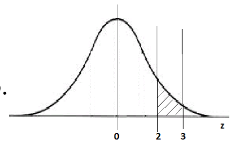
Esercizio 4)

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

$$P(E_1)=\frac{\textit{esiti favorevoli}}{\textit{esiti possibil}}=\frac{|E_1|}{|U|}=\frac{5}{30}=\frac{1}{6}$$

Per calcolare $P(E_I)$ si deve rincondurre i due limiti $x_1 = -2$ e $x_2 = 0$ dalla normale in esame $y\sim N(-6,4)$ a quella standardizzata $z\sim N(0,1)$.

$$z_1=\frac{x_1-E[X]}{\sqrt{Var[X]}}=\frac{-2+6}{\sqrt{4}}=2 \qquad z_2=\frac{x_2-E[X]}{\sqrt{Var[X]}}=\frac{0+6}{\sqrt{4}}=3$$



da cui si ricava che il punto $x=-2$ corrisponde al punto $z=1$, come indicato nella figura a lato. Pertanto la probabilità richiesta diviene

$$P(E_2) = P(-2<X<0) = P(2<Z<3) = 0.0215$$

Utilizzando la probabilità assiomatica e ricordando che per eventi incompatibili la probabilità dell'evento intersezione è pari a zero, si possono ricavare la probabilità dell'evento unione:

$$P(E_1\cup E_2)=P(E_1)+P(E_2)-P(E_1\cap E_2)=0.1667+0.0215-0=0.1882$$

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

Nel caso gli eventi fossero stati dichiarati indipendenti nulla sarebbe cambiato nel calcolo della probabilità dei singoli eventi elementari, sarebbe cambiata la probabilità dell'evento unione in qunato la probabilità dell'evento intersezione per eventi indipendenti è data dal prodotto delle probabilità dei singoli eventi. Si otterrebbe quindi:

$$P(E_1\cup E_2)=P(E_1)+P(E_2)-P(E_1\cap E_2)=0.1667+0.0215-0.1667*0.0215=0.1846$$