

**STATISTICA CORSO BASE. Prova scritta del 31-1-2012. Tempo: 2 ore - Compiti: A**

Cognome e Nome: ..... Matricola: .....

Attenzione: Prima di affrontare la prova si consiglia di leggere attentamente le note che seguono.

Indicare cognome e nome (in stampatello e senza cancellature)

Verrà corretto solo l'elaborato contenuto in questi fogli. Alla consegna, allegare comunque la brutta copia.

Non è consentita la consultazione di libri o appunti personali ma solo del formulario fornito dal docente.

Tenere spenti i telefoni cellulari. Il loro utilizzo non è consentito pena l'esclusione dalla prova.

Per ogni domanda a risposta multipla sono previste tre possibili risposte di cui una sola è corretta. Per fornire la risposta ad una domanda barrare la casella corrispondente alla risposta prescelta. Nel caso di correzioni, fare in modo che sia chiaro quale sia la risposta finale. Il punteggio assegnato ad ogni quesito è:

1 se si indica la risposta corretta, -0.5 se si indica la risposta errata, 0 se non si fornisce risposta.

Se un candidato ritiene che ci sia un errore ed in particolare che nessuna delle tre alternative disponibili risponda al quesito posto, può aggiungere la risposta corretta opportunamente motivata.

Nelle altre domande, quelle aperte, fornire, oltre al risultato, anche la spiegazione della procedura adottata.

**Esercizio 1. 8 punti**

- Nella seguente tabella, a cui si riferiscono i quesiti che seguono, sono elencate le modalità del carattere  $X$ , che rappresenta il prezzo in migliaia di Euro degli appartamenti al metro quadrato per 18 appartamenti messi in vendita nel comune di Roma:

appartamento	A	B	C	D	E	F	G	H	I	L	M	N	O	P	Q	R	S	T
$x$	5.1	3.6	1.3	4.2	4	3.6	4.1	4.7	4.4	5	2	4.8	4	3.7	4.9	4	12.1	4.7

- ★ Qual è la moda della distribuzione del carattere  $X$ ?

non esiste     4     3

- ★ Calcolare il valore della media aritmetica della distribuzione di  $X$ .

$$\bar{x} = 4.56$$

- ★ Individuare il valore della mediana ( $M$ ) dei dati raccolti.

$$M = 4.3$$

- ★ Quale, tra i seguenti, è il valore del primo quartile ( $Q_1$ )?

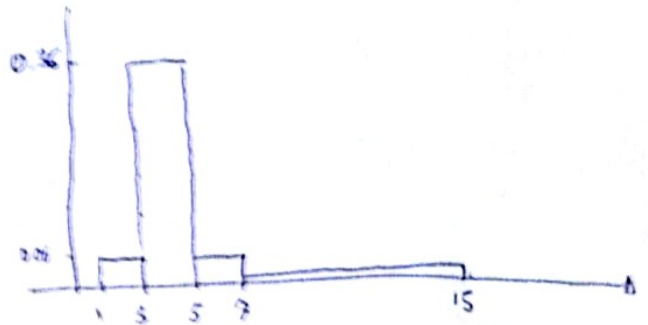
4     1.3     4.2

- ★ Immaginiamo che l'osservazione maggiore passi da 12100 Euro a 15000 Euro. Il primo quartile

Aumenta     Diminuisce     Rimane Invariato

★ Facendo riferimento alla suddivisione in classi 1 + 3, 3 + 5, 5 + 7, 7 + 15, rappresentare graficamente la corrispondente distribuzione di frequenze relative del carattere X.

classi	freq. relative	densità frequenze
[1,3]	$2/18 = 0.11$	0.06
[3,5]	$13/18 = 0.72$	0.36
[5,7]	$2/18 = 0.11$	0.06
[7,15]	$1/18 = 0.05$	0.01



★ Con riferimento alla rappresentazione grafica precedente ritieni che, eliminando il dato relativo all'appartamento "S", la figura

rimanga invariata

diventi simmetrica

diventi una curva Normale

### Esercizio 2. 7 punti

- Nel torneo 6 nazioni di rugby, l'Italia deve giocare 5 partite. Supponiamo che per l'Italia la probabilità di vincere una partita sia pari a 0.15 indipendentemente dall'avversario e che i risultati delle partite siano indipendenti tra loro. Inoltre, i bookmakers ritengono pari a 0.35 la probabilità che l'Italia sia ultima e contemporaneamente vinca almeno una partita

★ Calcolare la probabilità che l'Italia vinca almeno una partita.

3 PUNTI

Sia  $X =$  numero partite vinte dall'Italia,  $X \sim \text{BINOMIALE}(0.15, 5)$

$A = \{ \text{Italia vinca almeno una partita} \}$   $P(A) = P(X \geq 1) = 1 - P(X=0) = 1 - 0.15^0 \cdot 0.85^5 = 0.556$

★ Calcolare la probabilità che l'Italia vinca almeno una partita e non sia ultima.

2 PUNTI

$B = \{ \text{Italia è ultima} \}$

poiché  $A = (A \cap B) \cup (A \cap \bar{B})$  e  $P(A) = P(A \cap B) + P(A \cap \bar{B})$

dove  $P(A) = 0.556$  e  $P(A \cap B) = 0.35$

Viene richiesto  $P(A \cap \bar{B})$

Abbiamo  $0.556 = 0.35 + P(A \cap \bar{B})$ . Quindi  $P(A \cap \bar{B}) = 0.556 - 0.35 = 0.206$

★ Calcolare la probabilità che l'Italia non sia ultima sapendo che ha vinto almeno una partita

$P(\bar{B} | A) = \frac{P(\bar{B} \cap A)}{P(A)} = \frac{0.206}{0.556} = 0.37$

★ Gli eventi {L'Italia vinca almeno una partita} e {L'Italia non è ultima} sono

Compatibili

Incompatibili (disgiunti)

Indipendenti

gli eventi sono compatibili Infatti  $P(A \cap \bar{B}) > 0$

Inoltre se l'assegnazione delle probabilità fatta dai bookmakers è corretta gli eventi ~~devono~~ essere dipendenti. Infatti se l'Italia perdere tutte le partite dovrebbe ~~avere~~ aumentare la probabilità di finire ultima



Esercizio 3. 7 punti

- Una società telefonica, all'interno di una ricerca di mercato, ha chiesto a 81 studenti maschi possessori di un cellulare con connessione internet, quante ore hanno navigato su internet tramite il proprio telefono nel corso dell'ultimo mese. Il numero medio di ore rilevato su questo campione è risulta pari a 5 mentre la deviazione standard campionaria è risultata pari a 2

- ★ Ricavare un intervallo di confidenza, al livello 99%, per il livello medio  $\mu$  di ore di navigazione su internet per gli studenti maschi con età tra i 18 e i 23 anni.

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = \left[ 5 \pm t^* \frac{2}{9} \right] = \left[ 5 \pm 2.639 \cdot \frac{2}{9} \right] = [4.11, 5.59]$$

- ★ Il P-value relativo al sistema di ipotesi  $H_0: \mu = 4$  contro l'alternativa  $\mu \neq 4$  risulta

> 1                       < 0.05                       > 0.05

$$t = \frac{5-4}{2/9} = 4.2 \quad P\text{-value} = 2 \cdot P(t_{80} > 4.2) < 2 \cdot 0.0005 = 0.001$$

Supponiamo ora che per un campione di 20 studentesse si sia osservata una media campionaria relativa al tempo passato su internet tramite cellulare pari a 3 ore al mese con una deviazione standard campionaria pari a 4 ore. La società telefonica è convinta che i ragazzi passino in media più ore su internet rispetto alle ragazze e vuole verificare tale ipotesi sulla base dei dati a disposizione

- ★ Qual è il sistema di ipotesi statistiche che la società telefonica deve verificare?

$$\begin{array}{ll}
 \mu_1 = \text{media ragazzi} & H_0: \mu_1 = \mu_2 \\
 \mu_2 = \text{media ragazze} & H_a: \mu_1 > \mu_2
 \end{array}$$

- ★ Calcolare la statistica test relativamente a tale sistema di ipotesi 2 PUNTI

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{5-3}{\sqrt{\frac{4}{81} + \frac{16}{20}}} = 2.17$$

- ★ Al livello di significatività del 5%, ritieni che l'ipotesi nulla formulata in precedenza possa essere rifiutata? (Motivare la risposta) 2 PUNTI

$$P\text{-value} = P(t_{\min(n_1-1, n_2-1)} > 2.17) \in [0.02, 0.025] \Rightarrow P\text{-value} < 0.05$$

RIFIUTO  $H_0$

Esercizio 4. 8 punti

- Il sito web dell'Eurostat riporta le previsioni per la crescita del prodotto interno lordo nel 2012 (variabile  $y$ ) e i tassi di disoccupazione aggiornati al Novembre 2011 (variabile  $x$ ) per 26 paesi dell'area Euro). I dati osservati sono sintetizzati mediante le seguenti statistiche riassuntive, relative alle due variabili (medie, varianze e coefficiente di correlazione):

$$\bar{x} = 8.765; \quad \bar{y} = 1.053; \quad s_x^2 = 6.894; \quad s_y^2 = 1.958; \quad r = -0.104.$$

- ★ Ricavare, con il metodo dei minimi quadrati, il valore dei coefficienti della retta di regressione con variabile risposta la crescita del prodotto interno lordo

$$\begin{aligned}
 \hat{y} &= a + bx \quad \text{dove} \quad b = r \frac{s_y}{s_x} = -0.104 \cdot \sqrt{\frac{1.958}{6.894}} = -0.0554 \\
 a &= \bar{y} - b\bar{x} = 1.053 + 0.055 \cdot 8.765 = 1.53
 \end{aligned}$$



- ★ Misurare la bontà di adattamento del modello di regressione ai dati attraverso un adeguato indice, spiegando anche il significato di quest'ultimo. 2 PUNTI

$$r^2 = 0.104^2 = 0.01 = \frac{\text{variabilità di } y \text{ spiegata dalla retta}}{\text{variabilità totale di } y}$$

L'indice  $r^2$  è molto basso. Solo l'1% della var. di  $y$  è spiegata dalla regressione.

- ★ Sulla base del modello stimato, quando il tasso di disoccupazione aumenta di un'unità di quanto varia la crescita del prodotto interno lordo?

~~Il~~ La crescita del prodotto interno lordo diminuisce di 0.055 punti.

- ★ Sulla base dei dati disponibili, vogliamo confrontare le seguenti ipotesi riguardanti il valore del coefficiente angolare del modello di regressione

$$H_0: \beta = 0$$

$$H_a: \beta < 0$$

Sapendo che l'errore standard di  $b$ , la stima del coefficiente angolare, è pari a  $ES_b = 0.108$ , qual è il valore della statistica test?

-0.513

-5.13

0

- ★ Qual è il valore  $P$  corrispondente al valore della statistica test calcolata al punto precedente?

 compreso tra 0.0025 e 0.005

 compreso tra 0.005 e 0.01

 maggiore di 0.01

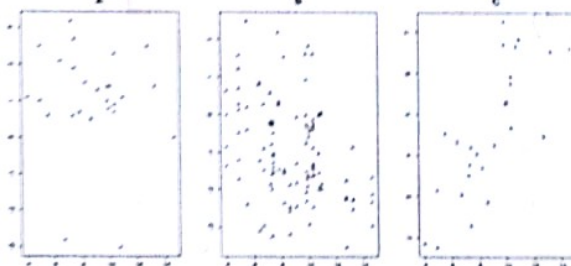
- ★ Con riferimento al problema di verifica delle ipotesi ed alla luce del risultato ottenuto nel punto precedente, cosa concludiamo ad un livello di significatività dell'1%?

 non c'è evidenza sufficiente per rifiutare l'ipotesi nulla

 i dati forniscono evidenza sufficiente per il rifiuto dell'ipotesi nulla

 le informazioni a disposizione non sono sufficienti a prendere una decisione

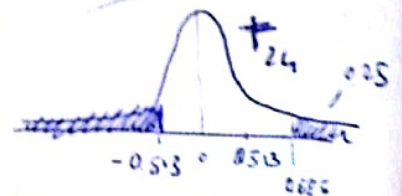
- ★ Quale dei seguenti tre diagrammi a dispersioni ritieni sia quello relativo ai dati appena analizzati


 A

 B

 C

Infatti: nel grafico C c'è una chiara crescita mentre i dati hanno  $b < 0$  mentre il grafico B ha troppi punti. Il grafico A ha esattamente 26 punti e mostra un andamento leggermente decrescente.



$$P\text{-value} = P(t_{24} < -0.513) \\ = P(t_{24} > 0.513) > 0.25$$