**CORREZIONI 1**

**ESERCIZIO 1**

Si vuole studiare la durata di vita media µ della popolazione di malati di tumore alla prostata con 3 diverse diagnosi di gravità (livello di gravità da 1 a 3) dopo la prima visita in ospedale. Per poter estrarre un campione di 150 individui con tale diagnosi, vengono considerati i database di 5 ospedali a Roma negli anni 2013-2014.

i) dite che tipo di campionamento usereste **2 PT**

A STRATI

**ii) spiegate** come fareste, **in pratica**, per estrarre il campione. **2.0 PT**

DIVISIONE DI TUTTI I PAZIENTI IN 3 STRATI (livelli di gravità)

ESTRAZIONE DI 1 CAMPIONE CASUALE SEMPLICE DI 50 individui per ciascuno dei LIVELLI DI GRAVITA’. POI UNIONE DEI 3 CAMPIONI

iii) µ indica una statistica campionaria **V F 0.5 PT**

**ESERCIZIO 2**

In una ricerca genetica si è scoperto che 400 soggetti, su 1200 esaminati, sono portatori di una mutazione che può indurre una malattia rara. Il campione esaminato è stato estratto casualmente.

i) Quale modello (distribuzione) probabilistico può descrivere questo esperimento aleatorio?

modello binomiale **1.5 PT**

ii) Quali **condizioni** sono necessarie per poter applicare questa distribuzione? **2.5 PT**

* Esperimento casuale costituito da n prove ( risultati dell’esp.) ripetute e indipendenti in ciascuna delle quali sono possibili 2 risultati che indicheremo con 1 e 0.

Per ogni prova è costante la probabilità di successo.

ii) dite **quali sono e quali valori** assumono i parametri di questo modello (distribuzione) **2.0 PT**

**n = 1200 p = 400/1200**

ESERCIZIO 3

Considerate queste coppie di dati.

**x y**

80.7 2.00

71.8 3.64

75.6 3.42

72.7 2.84

75.3 1.83

73.6 2.82

75.1 4.04

74.0 2.56

73.8 0.80

70.4 4.19

Dopo aver osservato il grafico di dispersione e calcolato l’equazione della retta di regressione



i) dite a **quale valore x** corrisponde il residuo maggiore in valore assoluto **1+1 PT**

ii) quanto vale tale residuo

i) x=73.8 ii) -2.14

iii) il punto corrispondente a tale valore x può essere considerato un outlier o un’osservazione influente rispetto alla retta di regressione? **Spiegate la vostra scelta**

Outlier nella direzione delle y **2 PT**

iv) quale percentuale di variabilità delle y è spiegata dal modello di regressione dei minimi quadrati?

16.6% **1.5 PT**

v) **spiegate** come avete calcolato tale percentuale

r2 = 16.6% = variabilità spiegata/ variab. tot **1.0 PT**

vi) spiegate come si interpreta questo risultato circa il modello di regressione

Il modello spiega solo il 16.6% della variabilità delle y , non è buono **1.0 PT**

**ESERCIZIO 4**

Osservate il seguente box-plot

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |  | | | | |

i) dite che forma ha la distribuzione dei dati rappresentati dalla figura **1 PT**

asimmetrica a sinistra

ii) riportate tutti gli indici di centralità e di variabilità che si possono usare per questi dati

mediana, moda, range, quartili, range interquartile **2 PT**

iii) riportate la formula della differenza interquartile e calcolatela per questi dati

Q3 - Q1 = 18-10 = 8 **1 PT**

iv) in un box-plot cosa rappresenta la scatola che compare nella fig.?

Il 75% centrale dei dati **1 PT**

iv) dite quanto può valere approssimativamente la mediana **0.5 PT**

M = 15

v) la media è più grande della mediana V **F 0.5 PT**

vi) quale percentuale di osservazioni è approssimativamente inferiore o uguale a 12?

a- 20%

b- 40%

c- 25%

**d- nessuna delle precedenti 1 PT**

viii) il box-plot può essere usato per variabili qualitative ordinate V **F 1 PT**

ESERCIZIO 5

Se una distribuzione normale ha media pari a 200 e deviazione standard pari a 50, trovare quel valore K tale che la proporzione di valori minori di K sia pari a 0.975.

**a b c d e 1.5 PT**

a) 239

b) 268

c) **298**

d) 300

e) 296

ii) per la stessa distribuzione calcolate il percentile corrispondente al valore 157 **1.5 PT**

0.194895

**ESERCIZIO 6**

Il numero di anelli sulle scaglie indica l’età di un pesce. In un allevamento di trote è stato osservato un campione grande, ed è stata determinata l’età (anni) di ogni pesce. I dati sono riportati nella tabella.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Età in anni | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| N° di pesci catturati | 16 | 19 | 13 | 12 | 10 | 6 | 1 |

i) Calcolate media, mediana e moda dell’età per questo campione.  **1 PT**

Variable N N\* Mean SE Mean StDev Minimum Q1 Median Q3 Maximum

C13 77 0 **4,039** 0,187 1,642 2,000 3,000 **4,000** 5,000 8,000

MODA = 3

ii) Disegnate, qui sotto, il grafico che ritenete più opportuno per rappresentare l’età considerata come variabile discreta. **2 PT**



iv) Dite quale o quali altri grafici sarebbe possibile usare (senza disegnarli) per rappresentare questo campione di dati.

ramo-foglia **PT 2**