

STATISTICA CORSO BASE. Prova scritta del 3-07-2018. *Tempo: 2 ore*

Cognome e Nome: **Matricola:**

Ex 1 La seguente tabella contiene le distribuzioni dei voti all'esame di statistica condizionate al sesso per 30 maschi e 50 femmine

voto	frequenze assolute M	freq. assolute F
18 – 20	9	15
21 – 24	12	10
25 – 27	6	10
28 – 30	3	15
totale	30	50

1. Rappresentare graficamente, in modo opportuno, la distribuzione marginale del voto per gli 80 studenti
2. Calcolare il voto medio e la deviazione standard per gli 80 studenti
3. Calcolare primo quartile, mediana e terzo quartile per il voto degli studenti maschi
4. Utilizzando i valori della media \bar{x} e della deviazione standard s calcolati sull'intero campione di 80 studenti, riportare un intervallo di confidenza al livello del 95% per il voto medio all'esame di statistica nella popolazione da cui il campione è stato estratto
5. L'intervallo trovato al punto precedente viene detto *intervallo per grandi campioni*. Enunciare il teorema che permette di costruire tale intervallo anche se la popolazione di riferimento non è chiaramente Normale

6. Sempre con riferimento al campione di 80 studenti verificare l'ipotesi nulla che la media nella popolazione sia di 26 rispetto all'ipotesi che sia diversa riportando il livello di significatività osservato o P-value

Ex 2 Referiamoci ancora ai dati dell'esercizio precedente.

1. Calcolare il voto medio dei maschi e quello delle femmine e stabilire sulla base di questi valori se vi è dipendenza in media del carattere Y voto a statistica rispetto al carattere X dato dal sesso.
2. Calcolare il rapporto di correlazione η_Y^2 e stabilire se il livello di dipendenza in media è elevato
3. Indicare il valore delle devianza residua, ovvero della somma delle devianze delle due distribuzioni condizionate di Y dato X

EX 3 Alcune palme di un noto parco nazionale sono malate per via della presenza di un parassita noto come *punteruolo rosso*. In particolare il parassita è presente nel 6% delle palme.

1. Prendendo a caso un campione di 10 palme qual è la probabilità che nessuna sia malata?
2. Sempre in un campione in un campione di 10 palme, qual è la probabilità che almeno una sia malata

3. Considerando un campione di 100 palme, qual è la probabilità di osservare al massimo 6 palme malate

4. Alcuni ricercatori ritengono che la presenza del parassita sia aumentata. Per verificare la loro ipotesi prendono un campione di 100 palme. Su queste 100 palme, 10 risultano avere il parassita. Cosa possono concludere i ricercatori ad un livello di significatività del 5%?

EX 4 Su un campione di 8 studenti si vuole analizzare la relazione tra esami sostenuti in un anno (variabile Y) e ore settimanali dedicate allo studio (variabile X). Gli 8 studenti hanno fornito i seguenti valori

Y	6	3	0	2	3	1	5	6
X	55	40	7	25	42	15	40	50

1. Determinare l'equazione della retta di regressione di Y rispetto a X con i dati sopra riportati

2. Sulla base della retta di regressione quale sarebbe il numero di esami sostenuti per uno studente che studia 6 ore al giorno?

3. Aumentando lo studio di circa 16 ore a settimana quanti esami in più si potrebbero superare?

4. Calcolare il valore dell'indice R^2 per misurare la bontà dell'adattamento della retta di regressione ai dati osservati. Sulla base di tale indice ritieni che la retta sia uno strumento utile per rappresentare i dati osservati.

5. Disegnare il diagramma a dispersione per le 8 osservazioni riportando la retta di regressione sopra calcolata.