

STATISTICA CORSO BASE. Prova scritta dell'8-06-2018. *Tempo: 2 ore*

Cognome e Nome: **Matricola:**

Ex 1 La seguente tabella contiene la distribuzione dei minuti passati giornalmente su internet per un campione di 1000 persone

minuti	frequenze assolute
0 – 5	140
5 – 15	360
15 – 30	210
30 – 60	220
60 – 120	70
totale	1000

1. Rappresentare graficamente, in modo opportuno, la distribuzione.

2. Indicare la classe modale

3. Calcolare media, mediana e varianza della distribuzione

4. Indicare la classe dove cade il primo quartile, riportare un valore puntuale per il terzo quartile e calcolare il valore della funzione di ripartizione $F(20)$

5. Utilizzando i valori della media \bar{x} e della deviazione standard s calcolati sul campione, riportare l'intervallo di confidenza al livello del 95% per il numero medio di minuti μ passati su internet nella popolazione da cui il campione è stato estratto

6. Verificare l'ipotesi nulla che la media μ sia uguale 25 minuti o sia inferiore riportando il valore della statistica test ed il p-value

Ex 2 Referiamoci ancora ai dati dell'esercizio precedente. Riguardo alla modalità più frequente di accesso a internet, risulta che tra le persone che lo utilizzano per meno di 15 minuti, il 75% vi accede più frequentemente tramite dispositivo mobile (tablet o cellulare), mentre tale percentuale risulta pari al 25% per gli altri

1. Se scegliamo a caso una persona nel campione qual è la probabilità che navighi su internet per più di 15 minuti
2. Dopo aver scelto a caso una persona nel campione, qual è la probabilità acceda ad internete più frequentemente tramite dispositivi mobili?
3. Se, dopo aver scelto una persona a caso, rileviamo che vi accede tramite dispositivi mobili , qual è la probabilità che navighi per più di 15 minuti

EX 3 Gli alberi di villa Ada hanno un'altezza media di 8 metri e una deviazione standard di 1.5 metri. Supponendo che la distribuzione dell'altezza di questi alberi sia approssimativamente normale calcolare

1. La probabilità che un albero preso a caso sia più basso di 6 metri
2. La probabilità che un albero preso a caso sia più basso di 6 metri oppure più alto di 10 metri
3. Stabilire se gli eventi $A=$ l'albero è più basso di 6 metri e B l'albero è più alto di 9 metri sono indipendenti o incompatibili e calcolare la probabilità $p(A \cap B)$
4. Calcolare la probabilità che in un campione di 10 alberi, l'altezza media sia superiore a 9 metri

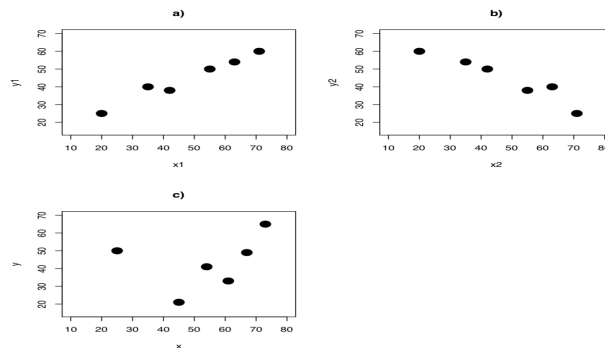
EX 4 Il presidente di un comitato di pari opportunità ha analizzato i dati provenienti da sei imprese che operano nel settore della grande distribuzione al fine di studiare la relazione tra la percentuale di dipendenti di genere femminile (indicata con x) e la percentuale di ruoli dirigenziali che, in ciascuna impresa, sono assunti da donne (indicata con y). Le sintesi ottenute sono le seguenti:

$$\bar{x} = 54.2, \bar{y} = 43.2, s_x^2 = 300.2, s_y^2 = 231.4, s_{xy} = 84.34, r = 0.32, s = 16.12$$

in cui \bar{x} e \bar{y} indicano le medie dei valori rilevati di x e y , s_x^2 e s_y^2 le rispettive varianze, r è il coefficiente di correlazione tra i valori rilevati di x e y , s_{xy} la covarianza e infine $s = \hat{\sigma}$ è l'errore standard della regressione con variabile risposta y e variabile esplicativa x , ovvero $\sqrt{\sum_{i=1}^n e_i^2 / (n - 2)}$.

A questi dati si riferiscono i quesiti che seguono.

- Ricavare l'equazione della retta dei minimi quadrati con variabile risposta y e variabile esplicativa x .
- Sulla base del modello stimato, quale percentuale di dirigenti donna dobbiamo aspettarci in un'impresa in cui il 30% dei dipendenti è di sesso femminile?
 36.42 39.4 44.98
- Sulla base del modello stimato, a quanto ci aspettiamo che ammonti l'incremento nella percentuale di donne che occupano ruoli dirigenziali se viene incrementata di uno la percentuale di dipendenti donna?
- Quale percentuale della variabilità di Y è spiegata dal modello di regressione lineare?
 5% 10% 32%
- Sulla base delle informazioni di sintesi disponibili, quale tra i seguenti ritieni che rappresenti il diagramma a dispersione dei dati raccolti?



- a) b) c)

- Alla luce della risposta al quesito precedente, ritieni che la previsione effettuata al punto ii) possa essere ritenuta affidabile? Spiega le ragioni della risposta.

7. Sulla base dei dati a disposizione verificare l'ipotesi nulla $H_0 : \beta_1 = 0$ rispetto all'ipotesi alternativa $\beta_1 \neq 0$ dove β_1 indica il coefficiente di regressione in tutta la popolazione delle imprese che operano nella grande distribuzione. Riportare il valore della statistica test e del p-value. (Ricordiamo che la stima della deviazione standard dello stimatore dei minimi quadrati B_1 è $\hat{\sigma}/\sqrt{D_x}$)

8. Riportare l'intervallo di confidenza al 95% per β_1