

Compilare in Stampatello

COGNOME:

NOME:

Matricola:

Esercizio 1

Su un campione di 220 imprese della provincia di Milano è stato rilevato il bilancio, espresso in migliaia di euro, del 2020. Qui di seguito i dati raccolti in classi e le frequenze percentuali.

- 1.a (pt4.3/31) Individuare la classe modale.
- 1.b (pt0.9/31) Quante imprese hanno un bilancio compreso tra -4 mila euro e zero.
- 1.c (pt0.6/31) La media è risultata essere $\bar{x} = 3.9865$; che relazione mi devo aspettare tra mediana e moda?
- 1.d (pt0.6/31) Siano x_1, \dots, x_n , n numeri, n dispari. Si consideri la funzione:

$$g(x) = |x_1 - x| + \dots + |x_n - x|.$$

Per quale valore di x , $g(x)$ è minima?

Esercizio 2

Una moneta perfetta viene lanciata 5 volte, se esce almeno 3 volte testa si estrae da un'urna che contiene un biglietto vincente ed uno perdente, altrimenti si estrae da un'urna che contiene due biglietti vincenti e tre perdenti.

- 2.a (pt4.3/31) Qual è la probabilità di vincere?
- 2.b (pt0.9/31) Si ripete il gioco di sopra finché non si vince due volte. Qual è la probabilità di finire alla quarta giocata?
- 2.c (pt0.6/31) Se $X \sim \text{Pois}(2)$ e $Y \sim \text{Pois}(1)$, è vero che

$$X - Y \sim \text{Pois}(1) \quad ?$$

- 2.d (pt0.6/31) Se X è una VC con supporto $\{0,1,2\}$ e Y è una VC con supporto $\{-2,-1,0\}$. Qual è il supporto di $X \times Y$?

$$\{-4, -2, -1, 0\}$$

Esercizio 3

- 3.a (pt4.3/31) Un'urna contiene 4 palline numerate da 1 a 4. Si estrae 100 volte con reinserimento e si fa la media dei 100 numeri estratti. Qual è la probabilità che la media sia compresa tra 2.5 e 2.6?

Esercizio 4

- 4.a (pt0.9/31) Sia h uno stimatore per θ , tale che

$$E(h) = \theta + \frac{\theta}{\sqrt{n}}$$

h è corretto? h è asintoticamente corretto?

- 4.b (pt0.9/31) Siano h_1 e h_2 due stimatori per θ , tali che:

$$\begin{aligned} \text{MSE}(h_1) &= \frac{\theta}{n} \\ \text{MSE}(h_2) &= \frac{2\theta}{n} \end{aligned}$$

Quale dei due stimatori è più efficiente? Perché?

4.c (pt0.9/31) Si sono osservati due gruppi di dati quantitativi e si è osservato, $\hat{\mu}_1 = 10.2$ e $\hat{\mu}_2 = 15.6$. Posto a test

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

è risultato $p_{\text{value}} = 0.0612$. La differenza tra $\hat{\mu}_1$ e $\hat{\mu}_2$ è significativa? Perché?

Esercizio 5

5.a (pt1.2/31) Su un campione di $n = 120$ startup tecnologiche italiane, è stato chiesto se abbiano implementato misure di cybersecurity avanzate. Lo studio ha riportato che 84 startup su 120 (il 70% del campione) hanno implementato queste misure.

Costruire un intervallo di confidenza al 95% per π , la quota di startup italiane che hanno implementato misure di cybersecurity avanzate.

5.b (pt3.0/31) Un'indagine molto più ampia condotta su startup europee ha mostrato che la percentuale di startup con misure di cybersecurity avanzate è del 80%. Testare l'ipotesi che in Italia la quota di startup con misure di cybersecurity avanzate sia uguale a quella europea contro l'alternativa che sia minore. Risolvere col p_{value} e confrontarlo per $\alpha = 0.1, 0.05, 0.01, 0.001$.

Esercizio 6

In uno studio sulla formazione aziendale, in un campione di $n = 30$ dipendenti, sono state analizzate le ore di formazione (in ore, X) e il punteggio di performance (in opportuna, Y).

Si osservano le seguenti statistiche: $\sum_{i=1}^{30} x_i = 1036.68$, $\sum_{i=1}^{30} y_i = 538.81$, $\sum_{i=1}^{30} x_i^2 = 39787.25$, $\sum_{i=1}^{30} y_i^2 = 10684.19$ e $\sum_{i=1}^{30} x_i y_i = 20527.76$.

6.a (pt4.3/31) Si è osservato $x_7 = 39.46$ e $y_7 = 18.26$, stimare il modello di regressione dove Y viene spiegata da X e calcolare il residuo per il punto $i = 7$.

6.b (pt0.9/31) Dare un'interpretazione dei parametri di regressione stimati.

6.c (pt0.6/31) Perché la previsione per $x = 35$ è più affidabile di quella per $x = 346$?

6.d (pt0.6/31) Cosa significa che r è un numero puro?

6.e (pt0.6/31) Se in un modello di regressione $r = 0.65$, $\hat{\sigma}_X = 1.1$ e $\hat{\sigma}_Y = 0.9$, calcolare $\hat{\beta}_1$.