

Data Analytics - Homework 1

Risolvere i seguenti esercizi in uno script R. Generare dallo script un file html prodotto attraverso il comando *Knit to HTML*. Caricare lo script ed il file html generato su moodle entro il 30 aprile 2025.

- Risolvi i seguenti esercizi:
 - a. Mostra a schermo la sequenza di numeri naturali da 20 a 50.
 - b. Mostra a schermo la sequenza di numeri razionali da 20 a 50 intervallati con distanza 0.01.
 - c. Salva in un vettore chiamato `sum` la somma dei numeri trovati al punto b ed in `mean` la loro media.
- Crea una funzione che calcola i primi n numeri di Fibonacci;
 - a. E se come output volessi, oltre ai primi n numeri, anche la loro media e la loro somma?
- Crea una lista che contenga al suo interno un vettore di numeri casuali fra 1 e 10, una matrice ed una lista di 3 cantanti. Poi assegna un nome per ogni elemento della lista (ad esempio "`Numeri`", "`Matrice`", "`Cantanti`").
- Considera il dataframe `Life.csv` caricato su Moodle. Importa il dataset e svolgi i seguenti punti:
 - a. Di che tipo sono le variabili? Alcune di esse possono essere considerate come dei fattori?
 - b. Ci sono delle osservazioni mancanti? Cosa faresti per risolvere il problema?
 - c. Quante osservazioni ci sono per ogni stato? E per ogni anno? Svolgi opportune modifiche al data frame per effettuare delle analisi coerenti.
 - d. Calcola la media dell'aspettativa di vita per l'Australia. Come si potrebbe confrontarla con quella degli altri Stati?
- Costruisci una matrice 2×3 composta da multipli di 2:
 - a. Che cosa succede usando il comando `is.matrix`? E `is.array`?
 - b. Estrai la terza colonna da questa matrice, chiamando l'oggetto estratto `b`. Che tipo di oggetto è?
 - c. Converti `b` in matrice.
- Il file 'nazioni.csv' contiene informazioni su 105 Nazioni e per ognuna di esse riporta:
 - `areaGeo`: la regione geografica
 - `reddito`: il reddito pro capite in dollari
 - `infmtort`: il tasso di mortalità infantile (morti ogni 100 nascite)
 - `oil`: se il paese esporta petrolio (1: no; 2: sì)
 1. Caricare i dati nello spazio di lavoro in un data frame chiamato "nazioni", stampare il numero di righe dell'oggetto importato ed il nome delle variabili.
 2. Stampare il vettore con il numero di valori mancanti presenti in ogni variabile.
 3. Ottenere il nome dei Paesi in cui sono presenti valori mancanti.
 4. Eliminare i valori mancanti dal data set.
 5. Ottenere la distribuzione delle frequenze percentuali della variabile 'areaGeo' ed ordinarle in ordine decrescente.

6. Convertire la variabile `areaGeo` in factor ordinando i livelli secondo l'ordine ottenuto al punto precedente. Salvare il factor come nuova variabile del data frame chiamata `areaGeofact` ed eliminare la variabile `areaGeo`.
7. Convertire la variabile `oil` in factor utilizzando i livelli: "no" e "yes". Sovrascrivere la variabile `oil` già presente nel `data.frame`.
8. Quali Paesi esportano petrolio e in quali regioni si trovano? Stampare il risultato in due colonne.
9. Calcolare il tasso di mortalità infantile medio in ogni area geografica.
10. Quante nazioni hanno un tasso di mortalità infantile superiore o uguale a 300?
11. Quante delle nazioni identificate al punto 10 esportano petrolio?
12. Dividere la finestra grafica in 2 righe e 2 colonne. In ogni spazio, rappresentare con un boxplot la distribuzione della mortalità infantile condizionata alla regione geografica. Impostare lo stesso range sull'asse y ed il titolo del grafico.
13. Rappresentare con un istogramma la distribuzione del reddito. Modificare l'etichetta dell'asse x con il nome della variabile ed eliminare il titolo.
14. Aggiungere al grafico precedente le mediane del reddito per area geografica utilizzando dei punti di colore diverso.
15. Dividere la variabile `reddito` in classi utilizzando le seguenti categorie: "fino a 500", "(500, 1500]", "(1500, 4000]", "4000 e più". Salvare la nuova variabile in un oggetto chiamato `redditoCat`.
16. Quante nazioni sono nella categoria "4000 e più"? E qual è la loro distribuzione per area geografica?
17. Stampare le distribuzioni condizionate della variabile `redditoCat` rispetto all'esportazione di petrolio approssimandole a 2 cifre decimali.