

Progetto parte in R – aa 2023-2024

- 1) Si generi un dataframe che simula un campione di $n = 200000$ studenti appartenenti a $N = 100$ scuole. Le scuole si distinguono in due gruppi: i Licei Scientifici e gli Istituti Tecnici a indirizzo economico, in rapporto 30 a 70. Sugli studenti si osservano il genere, avente distribuzione di Bernoulli con probabilità di selezionare una femmina pari al 40%, e il voto in matematica, avente una distribuzione normale con media 7.3 e deviazione standard 1.7 per i maschi del liceo scientifico e media 6.4 e deviazione standard 0.6 per le femmine del liceo scientifico. Negli istituti tecnici invece il voto in matematica è in media di 0.5 punto inferiore rispetto al liceo scientifico, per entrambi i sessi (la variabilità rimane costante).
- 2) Analizzare il dataframe simulato:
 - a) Contare quanti studenti maschi e femmine ci sono per ogni scuola e per ogni tipo di scuola.
 - b) Calcolare media, mediana, quartili, deviazione standard per il voto in matematica a livello di studente, distinguendo per genere, e a livello di scuola, distinguendo per tipo di scuola.
 - c) Rappresentare tramite un opportuno grafico il voto in matematica in modo da consentire il confronto tra maschi e femmine.
 - d) Rappresentare tramite un opportuno grafico il voto in matematica in modo da consentire il confronto tra licei scientifici e istituti tecnici.
 - e) Creare una nuova variabile che riporti il voto in matematica in classi di valori, distinguendo tra voto gravemente insufficiente (tra 0 e 5 escluso), insufficiente (tra 5 incluso e 6 escluso), sufficiente (tra 6 incluso e 7 escluso), buono (tra 7 incluso e 8.5 escluso), ottimo (almeno 8.5).
 - f) Distinguendo in base al genere, determinare la distribuzione di frequenza (frequenze assolute e relative, semplici e cumulate) per la nuova variabile di cui al punto e) e rappresentarla graficamente.
 - g) Creare una nuova variabile binaria per un voto in matematica ottimo (almeno pari a 8.5) e costruire la tabella a doppia entrata che riporta le frequenze relative condizionate degli studenti rispetto al voto in matematica (ottimo vs. non ottimo), dato il tipo di scuola.