

Data Analytics A
23 giugno 2025 - 1 ora e 30 minuti

Nome:	N. mat.:
-------	----------

Riportare nello script R `DataAnalytics20250623.R` le risposte della sezione **R** ed i comandi necessari allo svolgimento della sezione **Analisi dei dati**. Lo script va consegnato tramite moodle.

R

- (10) Scrivere una funzione che prenda in input due variabili: una variabile numerica, diciamo Y e un fattore diciamo X . La funzione, dopo avere verificato che le due variabili siano del tipo desiderato e avere eliminato i casi per cui una delle due variabili non è osservata, restituirà una matrice che contiene in ogni riga media, mediana e scarto quadratico medio della variabile Y per ciascuna delle modalità della variabile X . Chiamare la funzione `funzione`.
- (5) Fornire un'istruzione R che consenta di contare per una variabile x quanti siano i valori che risultano mancanti.

Analisi dei dati

Ogni risposta corretta vale 2.5 punti e ogni risposta errata dà luogo ad una penalizzazione di 0.5.

Si considerino i dati contenuti nel file `fev.csv`. Si tratta di informazioni raccolte su 654 bambini e adolescenti. La variabile chiave è `FEV`, la capacità polmonare (in litri). Inoltre si dispone dell'età (`AGE`), della statura (`HEIGHT`, in pollici) e di due variabili dicotomiche, genere (`SEX`) e fumo (`SMOKE`).

1. Quanto vale la devianza della variabile `FEV`
 - a. circa 0.75
 - b. circa 0.86
 - c. circa 490.9
 - d. circa 491.7
2. Quale è la coppia di variabili quantitative per la quale è più forte la relazione lineare?
 - a. `AGE` e `FEV`
 - b. `AGE` e `HEIGHT`
 - c. `FEV` e `HEIGHT`
 - d. nessuna coppia di variabili mostra una relazione lineare
3. Si vuole usare come misura di dispersione per la variabile `FEV` la differenza fra l'80mo e il 20mo percentile. Quanto vale?
 - a. circa 3.13
 - b. circa 1.43
 - c. circa 0.87
 - d. 0

4. Si calcoli l'odds ratio per le variabili **SMOKE** e **SEX**. Quanto vale?
- a. 0.6
 - b. circa 1
 - c. circa 3.74
 - d. 0.5
5. A partire dalle variabili **SMOKE** e **SEX** si consideri la variabile che è il prodotto logico delle due (ovvero una variabile che ha 4 modalità: "M e Fumatore", "F e Fumatore", "M e Non Fumatore", "F e Non Fumatore"). Per quale gruppo identificato da questa nuova variabile risulta la media di **FEV** più elevata?
- a. "M e Fumatore"
 - b. "F e Fumatore"
 - c. "M e Non Fumatore"
 - d. "F e Non Fumatore"
6. Si determinino i parametri della funzione di regressione multipla che ha come variabile risposta **FEV** e come variabili esplicative **HEIGHT** e **SEX**.
- 6.1** La devianza spiegata dalla regressione vale circa:
- a. circa 118.4
 - b. circa 372.5
 - c. 0.76
 - d. 490.92
- 6.2** Quale tra le seguenti affermazioni è accettabile:
- a. a parità di Altezza la capacità polmonare dei maschi è inferiore in media a quella delle femmine di circa 0.125 litri
 - b. la capacità polmonare dei maschi è inferiore in media a quella delle femmine di circa 0.125 litri
 - c. a parità di Altezza la capacità polmonare dei maschi è superiore in media a quella delle femmine di circa 0.125
 - d. la capacità polmonare dei maschi è superiore in media a quella delle femmine di circa 0.125 litri
- 6.3** Quale delle seguenti affermazioni non è supportata dalle analisi. La variabile **HEIGHT**
- a. all'aumentare dell'altezza di 1 pollice la capacità polmonare aumenta in media di circa 0.13 litri
 - b. all'aumentare dell'altezza di 10 cm la capacità polmonare aumenta in media di circa 0.51 litri
 - c. è molto rilevante per spiegare la capacità polmonare
 - d. all'aumentare dell'altezza la capacità polmonare aumenta in media del 13%
- 6.4** Qual è la percentuale di devianza dei residui della funzione di regressione
- a. circa 0%
 - b. circa 10.24%
 - c. circa 24.1%
 - d. circa 42.6%