

Terza esercitazione Data Analytics

Filippo Piccotto
filippo.piccotto@phd.units.it

18/04/2023

1. Scrivi una funzione `threshold` per valutare se un numero è maggiore di una certa soglia fissata dall'utente.
 - a. Estrai un vettore di 30 interi (da 1 a 100). Usa la funzione sopra costruita per valutare quanti elementi del vettore considerato superano la soglia di 50.
 - b. Importa il dataset `chicago.air` e usa la funzione `threshold` per valutare in quanti giorni il valore dell'ozono ha superato 0.075 ppm.
 - c. Usa la funzione `mapply()` per vedere quanti giorni hanno avuto un valore dell'ozono superiore a 0.075 ppm, quanti una temperatura superiore a 90 F, e quanti un valore di radiazioni solari superiore a 1.25.
2. Il dataset `Ceo.csv` considera i dati relativi ai pagamenti nei confronti degli amministratori delegati di alcune imprese corporative per l'anno 2008.
 - a. Ottieni delle prime statistiche descrittive per la variabile `Tot.Comp` e commenta quanto hai trovato. Produci un istogramma della variabile in questione sovrapponendo le informazioni che ritieni più opportune.
 - b. Grafica la funzione di ripartizione empirica della variabile in questione. Plotta sopra quella di una appropriata normale. Cosa puoi dire a proposito?
 - c. Produci un boxplot (cercando di identificare gli outlier) della compensazione. Commenta il grafico e se ritieni fai le modifiche che ritieni più opportune.
3. Scrivi una funzione che prenda in input la numerosità del campione (n), la media (μ) e la deviazione standard (s) e:
 - a. Simuli un campione casuale di dimensione n da una normale con media μ e deviazione standard s ;
 - b. Restituisca l'istogramma del campione simulato sovrapponendo la densità della distribuzione teorica.