

N. matricola : 937618
COGNOME - NOME: CRIPPA ALESSANDRO

-
- <1> Scrivere una funzione f che riceva come argomento un dataframe a , estragga da esso tutti i valori inferiori o uguali a 3.44 e restituisca una matrice avente 12 colonne. Utilizzare come valore di default dell'argomento a il dataset `mtcars`.
- <2> Quale grafico sarebbe appropriato al fine di rappresentare la distribuzione della variabile distanza percorsa durante l'allenamento (espressa in Km) in funzione dell'età (valori: inferiore a 65 anni, superiore o uguale a 65 anni) sulla base di dati raccolti in un campione di soggetti? "A") Grafico a torta; "B") Istogramma per gruppo.
- <3> Calcolare l'intersezione di un campionamento di 7 valori da un vettore contenente gli interi da 5 a 15 e di un campionamento di 7 valori da un vettore contenente i valori da 10 a 20, e salvarla in un vettore x utilizzando un'unica istruzione R. (Leggere il manuale delle funzioni `intersect` e `sample`)
- <4> Il più elevato dei valori non estremi (baffo superiore) del boxplot ottenuto a partire dai valori 0.03, 1.62, 2.73, 3.89, 4.24, 4.28, 4.5, 4.55, 6.21, 6.96, 11.79, 12.81 corrisponde a: "A") 11.79; "B") 6.96.
- <5> Creare una matrice a partire dalla lista ottenuta mediante il comando R `list(m1=matrix(rnorm(16), nrow=4), m2=matrix(letters[1:16], ncol=4))` unendo, come righe, le prime due righe della $m1$ e le ultime due righe di $m2$, salvarla in una variabile x utilizzando un'unica istruzione R.
- <6> Una variabile è caratterizzata da mediana = 19.96 e differenza interquartile = 3.89 nel campione 1 e da mediana = 11.63 e differenza interquartile = 4.45 nel campione 2. Quale delle seguenti affermazioni è corretta? "A") La distribuzione della variabile nel campione 1 è caratterizzata da maggior dispersione e da misura di posizione più elevata rispetto alla distribuzione della variabile nel campione 2; "B") La distribuzione della variabile nel campione 1 è caratterizzata da maggior dispersione e da misura di posizione inferiore rispetto alla distribuzione della variabile nel campione 2; "C") La distribuzione della variabile nel campione 1 è caratterizzata da minor dispersione e da misura di posizione più elevata rispetto alla distribuzione della variabile nel campione 2; "D") La distribuzione della variabile nel campione 1 è caratterizzata da minor dispersione e da misura di posizione inferiore rispetto alla distribuzione della variabile nel campione 2.
- <7> Costruite la trasposizione di un data.frame composto da: una colonna x contenente i primi 4 interi, una colonna y contenente i numeri da 5 a 8 (inclusi), una colonna z ottenuta incollando la lettera "b" ed il contenuto di un vettore contenente i primi 4 elementi del vettore `LETTERS` usando un separatore vuoto (""). Salvare in una variabile x utilizzando un'unica istruzione R. Suggerimento: leggete il manuale della funzione `paste()`.
- <8> Tra tutte le parole riportate in un testo scientifico sono presenti i termini "ecologia", "simbiosi", "biosfera", "ecosistema" con la seguente frequenza relativa: "ecologia" = 0.1275, "biosfera" = 0.11, "ecosistema" = 0.0545, "simbiosi" = 0.208. Qual è la probabilità che tre parole scelte casualmente ed indipendentemente formino la sequenza "biosfera"-simbiosi"-ecologia"? "A") 0.0002521; "B") 0.0013311; "C") 0.0029172; "D") 0.0010285.
- <9> Utilizzando la formula disponibile al link: <https://homes.di.unimi.it/re/Corsi/TMP/tmp1/formulaPDensity.jpg> e i parametri $\mu=2$ e $\sigma=2$ calcolare la densità di probabilità dei valori contenuti in `OGGETTO_002` e salvarle in una variabile x , utilizzando un'unica istruzione R.
- <10> In un campione di 12 persone esposte ad un determinato fattore ambientale è stato osservato che 7 di queste si sono ammalate. Data H_0 "La proporzione di persone esposte che si ammalano è uguale alla proporzione di persone esposte che non si ammalano", H_A "La proporzione di persone esposte che si ammalano non è uguale alla proporzione di persone esposte che non si ammalano" e la distribuzione nulla della statistica test la cui distribuzione di probabilità è riportata di seguito ($\Pr[0 \text{ malati}] = 0.000244$, $\Pr[1 \text{ malato}] = 0.00293$, $\Pr[2 \text{ malati}] = 0.016113$, $\Pr[3 \text{ malati}] = 0.053711$, $\Pr[4 \text{ malati}] = 0.12085$, $\Pr[5 \text{ malati}] = 0.193359$, $\Pr[6 \text{ malati}] = 0.225586$, $\Pr[7 \text{ malati}] = 0.193359$, $\Pr[8 \text{ malati}] = 0.12085$, $\Pr[9 \text{ malati}] = 0.053711$, $\Pr[10 \text{ malati}] = 0.016113$, $\Pr[11 \text{ malati}] = 0.00293$, $\Pr[12 \text{ malati}] = 0.000244$), calcolare il p-value ed indicare a quale dei seguenti valori corrisponde il p-value corretto e se sia possibile rifiutare l'ipotesi nulla dato un livello di significatività $\alpha = 0.05$. "A") 0.006348, non rifiuto H_0 ; "B") 0.006348, rifiuto H_0 ; "C") 0.774414, non rifiuto H_0 ; "D") 0.774414, rifiuto H_0 .
- <11> Effettuate un test di normalità su un vettore di 5000 elementi campionati dalla distribuzione normale e salvate il risultante p value in una variabile x , utilizzando un'unica istruzione R.
- <12> L'altezza delle piante di una determinata varietà è caratterizzata da un certo grado di variabilità. Si suppone tuttavia che l'altezza media delle piante di tale varietà sia di 37.7 cm. Al fine di verificare tale ipotesi sono stati raccolti dati relativi all'altezza di un campione di 8 piante: l'altezza media delle piante appartenenti al campione è risultata pari a 38.55 cm con deviazione standard di 1.56 cm. Applicando il test t per un campione e facendo riferimento alla tavola statistica della distribuzione t (`tavola_statistica_distribuzione_t.jpg`), l'evidenza derivante dai dati è sufficientemente forte da poter rifiutare l'ipotesi nulla (H_0 : "l'altezza media è di 37.7 cm"; H_A : "l'altezza media non è di 37.7 cm") assumendo un livello di significatività $\alpha = 0.05$? "A") no; "B") sì.

<13> OGGETTO_013_b contiene le misurazioni di una variabile riferita ad un test ematologico in due gruppi di soggetti portatori e non portatori di una mutazione genetica X. Testare l'ipotesi che la media nelle due popolazioni sia uguale e salvare il valore assoluto della differenza della stima dei valori medi nelle due popolazioni in una variabile x utilizzando un'unica istruzione R. Suggerimento: indagate la struttura dell'oggetto restituito dal test statistico prima di costruire la soluzione da sottomettere.

<14> Il test t per un campione è stato applicato al fine di verificare se il valore medio di emoglobina in portatori di una mutazione genetica sia di 16 g/dl (H_0 : "il valore medio di emoglobina nei portatori della mutazione è di 16 g/dl"; H_A : "il valore medio di emoglobina nei portatori della mutazione non è di 16 g/dl"). Basandosi sul p -value ottenuto (p -value = 0.03), se assumessi un livello di significatività $\alpha = 0.01$ incorrerei in errore nel prendere la decisione riguardo H_0 sapendo che il valore medio di emoglobina nei portatori della mutazione non è di 16 g/dl (realtà: H_0 falsa)? "A") Sì; "B") No.

<15> OGGETTO_014_a contiene il numero di accessi ad un pronto soccorso di un'apiccola cittadina nel corso dell'anno 2019. Effettuare un test del chi quadrato per bontà di adattamento, estrarre la stima del numero di accessi attesi nella giornata di mercoledì e salvare tale valore in una variabile x utilizzando un'unica istruzione R.

<16> Quale tra i valori di odds ratio stimati su dati raccolti nel contesto di quattro studi sperimentali indipendenti (studio 1: OR = 2.17, studio 2: OR = 1.84, studio 3: OR = 1.02, studio 4: OR = 0.99) indicherebbe evidenza più forte in merito all'efficacia di una terapia innovativa (valori variabile terapia: innovativa, standard) sulla guarigione da una determinata patologia (valori variabile guarigione: guarito, non guarito), considerando come successo l'evento "guarito" (gruppo di trattamento mediante terapia innovativa rispetto al gruppo di trattamento mediante terapia standard)? "A") OR = 2.17 (studio 1); "B") OR = 1.84 (studio 2); "C") OR = 1.02 (studio 3); "D") OR = 0.99 (studio 4).