

N. matricola : 978797
 COGNOME - NOME: FERRARI LORENZO

<1> Scrivere una funzione f avente un unico parametro formale d e che calcoli la percentuale di tutte le studentesse ammesse nei dipartimenti A, C e D, rispetto a tutte le studentesse ammesse all'Università UC Berkeley indipendentemente dal dipartimento di ammissione (utilizzare come valore di default per il parametro d il dataset UCBAAdmissions).

<2> Quale grafico sarebbe appropriato al fine di rappresentare la distribuzione della variabile frequenza cardiaca a riposo (valori: minore di 70 battiti al minuto, maggiore o uguale a 70 battiti al minuto) in funzione della variabile ore di allenamento settimanali (valori: inferiore a quattro ore, maggiore o uguale a quattro ore) sulla base dei dati raccolti in un campione di atleti? "A") Diagramma a barre raggruppate; "B") Scatterplot.

<3> Calcolare l'unione 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 union e sample)

<4> Il più elevato dei valori non estremi (baffo superiore) del boxplot ottenuto a partire dai valori 3.3, 3.68, 4.23, 4.38, 4.39, 4.6, 4.91, 5.11, 5.84, 6.49, 12.1, 12.7 corrisponde a: "A") 6.49; "B") 5.84.

<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 media = 40.65 e deviazione standard = 3.27 nel campione 1 e da media = 43.01 e deviazione standard = 3.16 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> Creare un data frame a partire da `mtcars` estraendo le righe in cui nella colonna `cyl` è presente un valore >7 e nella colonna `mpg` è presente un valore <11. Salvare in una variabile x utilizzando un'unica istruzione R.

<8> Tra tutte le parole riportate in un testo scientifico sono presenti i termini "ecologia", "simbiosi", "biosfera", "ecosistema" con la seguente frequenza relativa: "ecologia" = 0.0975, "biosfera" = 0.1605, "ecosistema" = 0.105, "simbiosi" = 0.137. Qual è la probabilità che un termine scelto casualmente tra tutte le parole presenti nel testo sia uno tra i seguenti: "ecosistema", "ecologia", "biosfera"? "A") 0.4045; "B") 0.363; "C") 0.2237; "D") 0.0016.

<9> Utilizzando la formula disponibile al link: <https://homes.di.unimi.it/re/Corsi/TMP/tmp1/formulaPDensity.jpg> e i parametri $\mu=1$ e $\sigma=3$ calcolare la densità di probabilità dei valori contenuti in OGGETTO_002 e salvarle in una variabile x , utilizzando un'unica istruzione R.

<10> Data una variabile casuale discreta X che può assumere valori pari a 0 (probabilità = 0.32), 1 (probabilità = 0.24), 2 (probabilità = 0.1), 3 (probabilità = 0.05), 4 (probabilità = 0.29), indicare a quale delle seguenti combinazioni di valori corrisponda il valore atteso e la varianza della variabile (i valori arrotondati alla seconda cifra decimale sono riportati nell'ordine: valore atteso, varianza). "A") 1.75, 2.04; "B") 0.76, 2.67; "C") 1.75, 2.67; "D") 0.76, 1.98.

<11> Utilizzando il valore $Z = 2.58$ (livello di confidenza al 99%) calcolare in un'unica istruzione R l'errore campionario del set di dati contenuto in OGGETTO_008_b e salvare il suo valore nella variabile x , utilizzando un'unica istruzione R.

<12> Uno studio sperimentale ha l'obiettivo di verificare se il valore medio della variabile X sia uguale in tre gruppi sperimentali costituiti da unità indipendenti: si riportano di seguito i valori di numerosità campionaria, valore medio e deviazione standard della variabile nei tre gruppi. Gruppo 1: numerosità = 8, valore medio = 11.42, deviazione standard = 0.76; Gruppo 2: numerosità = 7, valore medio = 9.91, deviazione standard = 0.86; Gruppo 3: numerosità = 9, valore medio = 10.63, deviazione standard = 1.86. Assumendo di applicare il test ANOVA ad una via, quale sarebbe il valore della statistica F ? "A") 3.27; "B") 2.48; "C") 2.98; "D") 1.18.

<13> OGGETTO_013_a contiene dati di misurazione di altezze di piante prima e dopo un trattamento. Testare l'ipotesi che la differenza nelle medie delle altezze sia 0 prima e dopo il trattamento scegliendo un test statistico e salvare il risultante p value in una variabile x utilizzando un'unica istruzione R.

<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 = 0.051), se assumessi un livello di significatività $\alpha = 0.01$ incorrerei in errore nel prendere la decisione riguardo H_0 sapendo

o che il valor medio di emoglobina nei portatori della mutazione non e' di 16 g/dl (realta': H0 falsa)? "A") Si'; "B") No.

<15> OGGETTO_014_c contiene dati relativi a 18 valori distribuiti su tre gruppi. Applicare un test A NOVA ad una via. Costruite una lista x contenente un data frame contenente le colonne 2 e 3 della tabella dei risultati restituiti dal test. Attribuite a questo elemento della lista il nome SommeEMedieSq. Il tutto 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 = 0.99, studio 2: OR = 3.19, studio 3: OR = 1.5, studio 4: OR = 9.18) indicherebbe evidenza piu' forte in merito all'efficacia di una tecnica chirurgica innovativa (valori variabile tecnica chirurgica: innovativa, standard) sulla guarigione da una determinata patologia (valori variabile guarigione: guarito, non guarito), considerando come successo l'evento "guarito" (gruppo di trattamento mediante tecnica chirurgica innovativa rispetto al gruppo di trattamento mediante tecnica chirurgica standard)? "A") OR = 0.99 (studio 1); "B") OR = 3.19 (studio 2); "C") OR = 1.5 (studio 3); "D") OR = 9.18 (studio 4).