

N. matricola : 05298A  
 COGNOME - NOME: DE CASTRO NEGREIROS HELEM VITORIA

- <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 temperatura dell'acqua dei laghi alpini (valori: inferiore a 10 gradi centigradi, superiore o uguale a 10 gradi centigradi) in funzione dell'altitudine (valori: inferiore a 1500 metri sul livello del mare, superiore o uguale a 1500 metri sul livello del mare) sulla base di misure effettuate in laghi alpini differenti? "A") Diagramma a barre raggruppate; "B") Scatterplot.
- <3> Costruire una matrice  $5 \times 10$  contenente numeri campionati dalla distribuzione normale con  $\text{media}=0$  e deviazione standard=1, salvarla in una variabile  $x$  utilizzando un'unica istruzione R.
- <4> Il più elevato dei valori non estremi (baffo superiore) del boxplot ottenuto a partire dai valori 0.29, 1.88, 2.53, 2.69, 3.06, 3.15, 3.3, 4.35, 4.6, 5.22, 10.95, 11.53 corrisponde a: "A") 5.22; "B") 10.95.
- <5> Creare un vettore di valori logici a partire dal secondo elemento della lista `list(m=matrix(rnorm(64),nrow=8),s=c(rep(c("A","T"),3),rep(c("C","G"),5)))` utilizzando un test logico che identifichi il carattere "T", salvarlo in una variabile  $x$ , utilizzando un'unica istruzione R.
- <6> Una variabile è caratterizzata da mediana = 46.86 e differenza interquartile = 5.17 nel campione 1 e da mediana = 11.32 e differenza interquartile = 3.86 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> Costruire un data.frame contenente i primi 120 numeri interi utilizzando un vettore  $v$  contenente i primi 20 interi ed una matrice  $m$ , salvarlo 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" = 0.1615, "biosfera" = 0.1285, "ecosistema" = 0.145, "simbiosi" = 0.065. Qual è la probabilità che un termine scelto casualmente tra tutte le parole presenti nel testo sia uno tra i seguenti: "ecosistema", "simbiosi", "ecologia"? "A") 0.4045; "B") 0.0015; "C") 0.2224; "D") 0.3715.
- <9> Effettuare un t test per dati appaiati testando l'ipotesi che la differenza tra le medie dei campioni sia pari a 0 confrontando: un vettore di 100 valori campionati dalla normale con  $\text{media}=0$  e  $\text{sd}=1$ , un vettore di 100 valori campionati dalla normale con  $\text{media}=1$  e  $\text{sd}=1$ , estrarre il p-value risultante e salvarlo in una variabile  $x$ , utilizzando un'unica istruzione R. Suggerimento: leggere il manuale della funzione `t.test()`
- <10> In un campione di 12 persone esposte ad un determinato fattore ambientale è stato osservato che 9 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 ( $\text{Pr}[0 \text{ malati}] = 0.000244$ ,  $\text{Pr}[1 \text{ malato}] = 0.00293$ ,  $\text{Pr}[2 \text{ malati}] = 0.016113$ ,  $\text{Pr}[3 \text{ malati}] = 0.053711$ ,  $\text{Pr}[4 \text{ malati}] = 0.12085$ ,  $\text{Pr}[5 \text{ malati}] = 0.193359$ ,  $\text{Pr}[6 \text{ malati}] = 0.225586$ ,  $\text{Pr}[7 \text{ malati}] = 0.193359$ ,  $\text{Pr}[8 \text{ malati}] = 0.12085$ ,  $\text{Pr}[9 \text{ malati}] = 0.053711$ ,  $\text{Pr}[10 \text{ malati}] = 0.016113$ ,  $\text{Pr}[11 \text{ malati}] = 0.00293$ ,  $\text{Pr}[12 \text{ malati}] = 0.000244$ ), 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") 0.145996, non rifiuto  $H_0$ ; "B") 0.145996, rifiuto  $H_0$ ; "C") 0.006348, non rifiuto  $H_0$ ; "D") 0.006348, rifiuto  $H_0$ .
- <11> Ipotizzando che il valore medio di una popolazione sia pari a  $\mu=8$  (ipotesi nulla) effettuare un test t, calcolare il p value (a 2 code) a partire dal set di dati (campione singolo) contenuto in `OGGETTO_007_b`, e salvare il p value in una 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 = 10.69, deviazione standard = 2.17; Gruppo 2: numerosità = 7, valore medio = 11.2, deviazione standard = 1.03; Gruppo 3: numerosità = 9, valore medio = 10.55, deviazione standard = 0.85. Assumendo di applicare il test ANOVA ad una via, quale sarebbe il valore della statistica  $F$ ? "A") 0.11; "B") 2.1; "C") 0.41; "D") 0.07.
- <13> `OGGETTO_013_a` contiene dati di misurazione di altezze di piante prima e dopo un trattamento. T

estare 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 esatto di Fisher e' stato applicato al fine di verificare se le variabili X ed Y siano indipendenti ( $H_0$ : "le variabili sono indipendenti";  $H_A$ : "le variabili non sono indipendenti"). Basandosi sul p-value ottenuto (p-value = 0.043), se assumessi un livello di significativita' alfa = 0.10 incorrerei in errore nel prendere la decisione riguardo  $H_0$  sapendo che le due variabili sono indipendenti (realta':  $H_0$  vera)? "A") Si'; "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 bonta' di adattamento, estrarre la stima del numero di accessi attesi nella giornata di mercoledi' 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 piu' 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).