www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

<1> Scrivere una funzione f avente un unico parametro formale b che restituisca le medie Petal.Length suddivise per valore nella colonna Species. Utilizzate una funzione della famiglia apply. Per il calcolo delle medie dovrete utilizzare a vostra scelta almeno una delle funzioni riportate di seguito: apply, sapply, lapply, tapply. Utilizzate come valore di default per il parametro b il dataset iris.
<2> Data una variabile casuale discreta X che puo' assumere valori pari a 0 (probabilita' = 0.18), 1 (probabilita' = 0.2), 2 (probabilita' = 0.16), 3 (probabilita' = 0.29), 4 (probabilita' = 0.17), 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") 1.85, 2.45; "B") 1.85, 0.97; "C") 2.07, 1.89; "D") 1.98, 2.08; "E") 2.07, 0.76.

<3> Costruire una matrice 6x5 a partire da un vettore contenente i primi 5 numeri interi, salvatela in una variabile x utilizzando un'unica istruzione R.
<4> Data una variabile caratterizzata da distribuzione normale con media = 2.602 e deviazione standard = 1.04 in un campione di 10 unita', calcolare l'intervallo di confidenza al 95% della media utilizzando la tavola statistica della distribuzione t https://homes.di.unimi.it/re/Corsi/TMP/tmp1/t.jpg. L'intervallo di confidenza (valori arrotondati alla terza cifra decimale) corrisponde a: "A") IC 95% = 1.246 : 3.958; "B") IC 95% = 1.859 : 3.345; "C") IC 95% = 1.244 : 3.96; "D") IC 95% = 1.382 : 3.822; "E") IC 95% = 1.523 : 3.681.
<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 di m1 e le ultime due righe di m2, salvarla in una variabile x utilizzando un'unica istruzione R.
<6> In un campione di 10 piante e' stato osservato che 6 di queste hanno sviluppato un determinato tipo di malattia in seguito ad esposizione ad un parassita. Data H0: "la proporzione di piante malate in seguito ad esposizione al parassita e' uguale alla proporzione di piante non malate", HA: "la proporzione di piante malate in seguito ad esposizione al parassita non e' uguale alla proporzione di piante non malate" e la distribuzione nulla della statistica test i cui valori di probabilita' sono riportati di seguito (Pr[0 piante malate] = 0.001, Pr[1 pianta malata] = 0.01, Pr[2 piante malate] = 0.042, Pr[3 piante malate] = 0.12, Pr[4 piante malate] = 0.204, Pr[5 piante malate] = 0.248, Pr[6 piante malate] = 0.204, Pr[7 piante malate] = 0.12, Pr[8 piante malate] = 0.042, Pr[9 piante malate] = 0.01, Pr[10 piante malate] = 0.001), a quale dei seguenti valori corrisponde il p-value? "A") 0.754; "B") 0.245; "C") 0.26; "D") 0.178; "E") 0.254.

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

<7> Ordinare in modo crescente OGGETTO_005 rispetto al contenuto di colonna 2, salvare in una variabile x utilizzando un'unica istruzione R. Suggerimento: leggere il manuale della funzione order().
<8> Applicando il test di correlazione di Pearson a due variabili caratterizzate da 15 misurazioni e' stato ottenuto un valore della statistica t pari a 2.57, sulla base dei dati disponibili e delle informazioni contenute nella tavola statistica della distribuzione t https://homes.di.unimi.it/re/Corsi/TMP/tmp1/t.jpg e' possibile rifiutare H0 (la correlazione tra le due variabili e' uguale 0) in favore di HA (la correlazione tra le due variabili e' diversa da 0) assumendo una soglia di significativita' pari ad alfa = 0.05 e ad alfa = 0.01? "A") alfa = 0.05: vero, alfa = 0.01: vero; "B") alfa = 0.05: vero, alfa = 0.01: falso; "C") alfa = 0.05: falso, alfa = 0.01: falso; "D") alfa = 0.05: falso, alfa = 0.01: vero.
<9> Utilizzando la formula disponibile in https://homes.di.unimi.it/re/Corsi/TMP/tmp1/DevStrFormula.jpg calcolare la deviazione standard dei dati contenuti in OGGETTO_001_a e salvarla in una variabile x utilizzando un'unica istruzione R.
<10> Se il 68-mo percentile di una serie di misurazioni corrispondesse ad un valore pari a 77, quale percentuale delle misurazioni avrebbe valori maggiori di 77? "A") circa 32%; "B") circa 68%.
<11> Simulare 10000 lanci di un dado a 6 facce ed estrarre la frequenza osservata del numero 3, salvare in una variabile x utilizzando un'unica istruzione R.
<12> Uno studio sperimentale ha l'obiettivo di verificare se il valor medio della variabile X sia uguale in tre gruppi sperimentali costituiti da unita' indipendenti: si riportano di seguito i valori di numerosita' campionaria, valor medio e deviazione standard della variabile nei tre gruppi. Gruppo 1: numerosita' = 8, valor medio = 2.26, deviazione standard = 4.18; Gruppo 2: numerosita' = 10, valor medio = 5.26, deviazione standard = 1.91; Gruppo 3: numerosita' = 9, valor medio = 7.17, deviazione standard = 0.76. Assumendo di applicare il test di ANOVA ad una via, quale sarebbe il valore della statistica F? "A") 5.38; "B") 3.89; "C") 9.52; "D") 11.45; "E") 7.75.
<13> Utilizzando la funzione integrate calcolare l'interale di probabilita' sotto la curva della distribuzione normale tra i valori 0 e 0.5, dal risulato estrarre il valore calcolato e salvarlo in una variabile x, utilizzando un'unica istruzione R. Suggerimento: leggere il manuale della funzione integrate().
<14> Il test esatto di Fisher e' stato applicato al fine di verificare se le variabili X ed Y siano indipendenti (H0: le variabili sono indipendenti; HA: le variabili non sono indipendenti). Basandosi sul p-value ottenuto (p-value = 0.043), se applicassi una soglia di significativita' alfa = 0.10 incorrerei in errore nel prendere la decisione riguardo H0 sapendo che le due variabili sono indipendenti (verita': H0 vera)? Se si' di quale tipo? Indicare la risposta. "A") Errore tipo I; "B") Errore tipo II; "C") Decisione corretta.

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

<15> OGGETTO_004 e' un dataframe formato da due colonne. La seconda colonna contiene un carattere che indica a quale di 4 categorie appartiene il dato riportato in colonna 1. Calcolare la somma delle frequenze percentuali delle categorie A e B e salvare il risultato in una variabile x utilizzando un'unica istruzione R.
<16> In un documento scientifico sono presenti i termini "batterio", "microbiologia", "antibiotico", "resistenza", "piastra" con la seguente frequenza relativa: "batterio" = 0.312, "microbiologia" = 0.29, "antibiotico" = 0.257, "resistenza" = 0.022, "piastra" = 0.118. Qual e' la probabilita' che tre termini scelti a caso e in maniera indipendente tra quelli presenti nel documento formino la sequenza "batterio"-"piastra"-"resistenza"? "A") 0.000128; "B") 0.0021735; "C") 0.000809952; "D") 0.00032887.
<17> Utilizzando il valore Z = 1.96 (livello di confidenza al 95%) 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.
<18> Calcolare il coefficiente di correlazione di Pearson r tra le variabili X1 (valori = 8.08, 9.41, 10.52, 7.7, 10.39, 10.06, 10.17, 12.23) e Y1 (valori = 7.56, 12.53, 8.51, 7.74, 8.57, 10.51, 10.3, 9.38) e tra le variabili X2 (valori = 8.09, 8.7, 12.45, 10.4, 8.84, 8.12, 9.59, 6.67) e Y2 (valori = 9.03, 8.52, 12.32, 12.02, 9.86, 7.73, 11.8, 11.7). Quale delle seguenti affermazioni e' corretta? "A") La correlazione tra X1 ed Y1 e' piu' debole rispetto alla correlazione tra X2 ed Y2; "B") La correlazione tra X1 ed Y1 e' piu' forte rispetto alla correlazione tra X2 ed Y2.

<1> Scrivere una funzione f avente un unico parametro formale c e che estragga da esso il numero di passeggeri di eta' "Child" e di classe "3rd", trasformi i dati ottenuti in un vettore numerico, lo normalizzi rispetto a media e deviazione standard e restituisca il vettore normalizzato. Utilizzare come parametri di default per l'argomento c il dataset Titanic.
<2> Data una variabile casuale discreta X che puo' assumere valori pari a 0 (probabilita' = 0.17), 1 (probabilita' = 0.16), 2 (probabilita' = 0.23), 3 (probabilita' = 0.21), 4 (probabilita' = 0.23), indicare a quale delle seguenti combinazioni di valori corrisponda il valore atteso e la deviazione standard della variabile (i valori arrotondati alla seconda cifra decimale sono riportati nell'ordine: valore atteso, deviazione standard). "A") 2.17, 0.98; "B") 1.98, 0.87; "C") 2.17, 2.37; "D") 1.98, 0.50; "E") 2.17, 1.39.
<3> Utilizzando nel comando R i vettori c(1,3,5,7,9,11,12) e c(1,2,4,6,8,10,12) costruire il vettore che rappresenti la loro intersezione, salvarlo nella variabile x utilizzando un'unica istruzione R. (leggere il manuale della funzione intersect)
<4> Data una variabile caratterizzata da distribuzione normale con media = 3.24 e deviazione standard = 0.86 in un campione di 7 unita', calcolare l'intervallo di confidenza al 95% della media utilizzando la tavola statistica della distribuzione t

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

https://homes.di.unimi.it/re/Corsi/TMP/tmp1/t.jpg. L'intervallo di confidenza (valori arrotondati alla terza cifra decimale) corrisponde a: "A") IC 95% = 1.872 : 4.608; "B") IC 95% = 2.481 : 3.999; "C") IC 95% = 2.444 : 4.036; "D") IC 95% = 1.959 : 4.521; "E") IC 95% = 2.863 : 3.617.

<5> Creare una lista contenente un vettore vlog1 ottenuto applicando il test ogico >-1.5 al contenuto di OGGETTO_001_a ed un vettore vlog2 ottenuto applicando il test logico >7 al contenuto di OGGETTO_002, salvarla in una variabile x utilizzando un'unica istruzine R.

<6> In un campione di 10 piante e' stato osservato che 6 di queste hanno sviluppato un determinato tipo di malattia in seguito ad esposizione ad un parassita. Data H0: "la proporzione di piante malate in seguito ad esposizione al parassita e' uguale alla proporzione di piante non malate", HA: "la proporzione di piante malate in seguito ad esposizione al parassita non e' uguale alla proporzione di piante non malate" e la distribuzione nulla della statistica test i cui valori di probabilita' sono riportati di seguito (Pr[0 piante malate] = 0.001, Pr[1 pianta malata] = 0.01, Pr[2 piante malate] = 0.042, Pr[3 piante malate] = 0.12, Pr[4 piante malate] = 0.204, Pr[5 piante malate] = 0.248, Pr[6 piante malate] = 0.204, Pr[7 piante malate] = 0.12, Pr[8 piante malate] = 0.042, Pr[9 piante malate] = 0.01, Pr[10 piante malate] = 0.001), a quale dei seguenti valori corrisponde il p-value? "A") 0.754; "B") 0.245; "C") 0.26; "D") 0.178; "E") 0.254.

<7> Ordinare in modo decrescente il contenuto di OGGETTO_005 rispetto al contenuto di colonna 1, salvare in una variabile x utilizzando un'unica istruzione R. Suggerimento: leggere il manuale della funzione order().

<8> I risultati di uno studio finalizzato a confrontare l'efficacia di due farmaci hanno dimostrato che delle 839 persone trattate con il farmaco X 825 sono guarite (14 non sono guarite) e che delle 1283 persone trattate con il farmaco Y 945 sono guarite (338 non sono guarite). Se applicassimo il test G per indipendenza delle variabili farmaco e guarigione il valore della statistica G sarebbe pari a: "A") 341.9; "B") 284.8; "C") 91.4; "D") 63.6; "E") 82.4.

<9> Utilizzando la formula disponibile in https://homes.di.unimi.it/re/Corsi/TMP/tmp1/DevStrFormula.jpg calcolare la deviazione standard dei dati contenuti in OGGETTO_002 e salvarla in una variabile x utilizzando un'unica istruzione R.

<10> Se il 19-mo percentile di una serie di misurazioni corrispondesse ad un valore pari a 76, quale percentuale delle misurazioni avrebbe valori maggiori di 76? "A") circa 81%; "B") circa 19%.

<11> Simulare 100000 lanci di un dado a 6 facce ed estrarre la frequenza osservata del numero 1, salvare in una variabile x utilizzando un'unica istruzione R.

<12> Dato il coefficiente di correlazione di Pearson r = -0.37 tra due variabili, si calcoli l'intervallo di confidenza al 95% sapendo che il coefficiente e' stato stimato su una numerosita' campionaria di 11 osservazioni. A quale dei seguenti valori corrisponde

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

l'intervallo di confidenza al 95%? "A") IC 95% = -0.75:0.2; "B") IC 95% = -0.79:0.3; "C") IC 95% = -0.64:-0.02; "D") IC 95% = -0.71:0.1.
<13> Effettuare un t test a due code per valutare l'ipotesi che i dati contenuti in OGGETTO_002 abbiano una media pari a 6.8, estrarre il risultante p-value e salvarlo nella variabile x utilizzando un'unica istruzione R.
<14> L'altezza delle piante di una determinata varieta' e' caratterizzata da un certo grado di variabilita'. Si suppone tuttavia che l'altezza media delle piante di tale varieta' sia di 38.83 cm. Al fine di verificare tale ipotesi sono stati raccolti dati relativi all'altezza di un campione di 10 piante: l'altezza media delle piante appartenenti al campione e' risultata pari a 41.805 cm con deviazione standard di 3.121 cm. Applicando il test t per un campione e facendo riferimento alla tavola statistica della distribuzione t https://homes.di.unimi.it/re/Corsi/TMP/tmp1/t.jpg, l'evidenza derivante dai dati e' sufficientemente forte da poter rifiutare l'ipotesi nulla (H0: l'altezza media e' uguale a 38.83 cm; HA: l'altezza media non e' uguale a 38.83 cm) assumendo una soglia di significativita' alfa = 0.05? "A") no; "B") si'.
<15> OGGETTO_004 e' un dataframe formato da due colonne. La seconda colonna contiene un carattere che indica a quale di 4 categorie appartiene il dato riportato in colonna 1. Calcolare la somma del numero di elementi appartenenti alle categorie A e B e salvare il risultato in una variabile x utilizzando un'unica istruzione R.
<16> Nel testo di un articolo scientifico sono presenti, tra tutte le parole, i termini "gene", "polimorfico", "patogeno", "mutato" con la seguente frequenza relativa: "gene" = 0.0265, "polimorfico" = 0.005, "patogeno" = 0.0225, "mutato" = 0.039. Qual e' la probabilita' che un termine scelto a caso tra tutte le parole presenti nel testo sia uno tra i seguenti: "mutato", "patogeno", "gene"? "A") 0.088; "B") 0.023; "C") 0.062; "D") 0.012.
<17> Effettuare un t test su un set di 100 valori campionati dalla normale con media=7.2 e deviazione standard=1 testando l'ipotesi che la media del campione sia pari a 6.8, salvare l'intervallo di confidenza al 95% per mu in una variabile x utilizzando un'unica istruzione R.
<18> Calcolare il coefficiente di correlazione di Pearson r tra le variabili X1 (valori = 9.83, 11.68, 9.07, 8.9, 11.47, 9.78, 9.66, 7.82) e Y1 (valori = 3.98, 8.81, 8.48, 10.58, 10.84, 7.41, 10.14, 8.37) e tra le variabili X2 (valori = 13.02, 9.46, 13.12, 9.53, 12.57, 9.98, 9.2, 10.04) e Y2 (valori = 13.49, 7.79, 7.88, 13.9, 11.21, 5.96, 13.01, 11.93). Quale delle seguenti affermazioni e' corretta? "A") La correlazione tra X1 ed Y1 e' piu' forte rispetto alla correlazione tra X2 ed Y2; "B") La correlazione tra X1 ed Y1 e' piu' debole rispetto alla correlazione tra X2 ed Y2.

<1> Calcolare la media di un vettore contenente 100 numeri campionati dalla distribuzione uniforme in [0,1] (vedere manuale della funzione runif) e salvare il suo valore nella variabile x utilizzando un'unica istruzione R.

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

<2> La quantità di pioggia annua (in mm) è un esempio di variabile: "A") numerica continua; "B") numerica discreta; "C") categorica ordinale; "D") categorica nominale
<3> Calcolare la differenza tra un vettore ottenuto campionando 7 elementi da un vettore contenente gli interi da 5 a 15 e un secondo vettore ottenuto campionando 7 elementi da un vettore contenente gli interi da 10 a 20, salvare il suo contenuto in un vettore di nome x utilizzando un'unica istruzione R. (Leggere il manuale delle funzioni setdiff e sample)
<4> Uno studio che consiste nel verificare l'inquinamento dell'aria in due città differenti è uno studio di tipo: "A") sperimentale; "B") osservazionale
<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 di m1 e le ultime due righe di m2, salvarla in una variabile x utilizzando un'unica istruzione R.
<6> Sono stati raccolti dati relativi ad una misurazione di pressione arteriosa sistolica in due gruppi di pazienti. Al primo gruppo è stato somministrato il farmaco X, al secondo gruppo è stato somministrato il farmaco Y; le misurazioni di pressione arteriosa sono state condotte in entrambe i gruppi solo dopo il trattamento con farmaci. La distribuzione di frequenza della variabile pressione arteriosa sistolica nei due gruppi non risulta normale nonostante opportune trasformazioni. Quale tipologia di test dovrei applicare al fine di verificare se la pressione arteriosa sistolica in pazienti trattati con farmaco X ed in pazienti trattati con farmaco Y sia uguale? "A") Test U di Mann-Whitney; "B") Test dei segni per ranghi di Wilcoxon; "C") Test dei segni; "D") ANOVA a due vie
<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> E stato condotto uno studio al fine di verificare lefficacia di un trattamento sulla guarigione da una malattia fungina (condizione di successo: guarigione) in una determinata specie vegetale. A tale scopo 20 piante sono state sottoposte a trattamento e 20 piante non sono state trattate. Al termine dello studio è stato contato il numero di piante guarite e non guarite in entrambi i gruppi. Se lodds ratio per la condizione "guarito" (gruppo di trattamento rispetto al gruppo di non trattamento) fosse pari a 4 quale delle seguenti affermazioni risulterebbe falsa? "A") Lodds di successo nel gruppo di trattamento è maggiore rispetto allodds di successo nel gruppo di piante non trattate; "B") Lodds di successo nel gruppo di trattamento è minore rispetto allodds di successo nel gruppo di piante non trattate; "C") Il rapporto tra la proporzione di successi e la proporzione di insuccessi nel gruppo di trattamento è maggiore rispetto al rapporto tra la proporzione di successi e la p
<9> Dato il vettore numerico contenuto nella seconda colonna di OGGETTO_005 calcolare la sua varianza e salvarla in una variabile x, utilizzando un'unica istruzione R.

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

<10> E' stato condotto uno studio finalizzato a verificare la possibile influenza dell'età sul risultato della compilazione di un questionario. Quale variabile è la variabile esplicativa? "A") Risultato del questionario; "B") Età

<11> Sia df= (frequenza teorica del numero 5 - frequenxa osservata del numero 5 in una simulazione di 100 lanci di un dado a 6 facce). Calcolare il valore di df e salvarlo in una variabile x, utilizzando un'unica istruzione R.

<12> Una variabile è caratterizzata da media = 49 e deviazione standard = 26.5 nel campione X e da media = 49 e deviazione standard = 11.5 nel campione Y, quale delle seguenti affermazioni è vera? "A") La distribuzione della variabile nel campione Y è caratterizzata da maggior dispersione rispetto alla distribuzione della variabile nel campione X; "B") La distribuzione della variabile nel campione X e nel campione Y sono caratterizzate dalla stessa dispersione; "C") La distribuzione della variabile nel campione X è caratterizzata da maggior dispersione rispetto alla distribuzione della variabile nel campione Y; "D") Sulla base delle informazioni disponibili non si può stabilire in quale campione la variabile sia caratterizzata da maggior dispersione

<13> Effettuare un t test a due code per valutare l'ipotesi che i dati contenuti in OGGETTO_002 abbiano una media pari a 6.8, estrarre il risultante p-value e salvarlo nella variabile x utilizzando un'unica istruzione R.

<14> Supponiamo di essere interessati a prevedere il numero di volte in cui esce "croce" dal lancio di 4 monete a due facce. Calcolare la probabilità che esca "croce" 3 volte dato lo spazio campionario riportato di seguito, in cui la lettera T indica testa e la lettera C croce. S = (T,T,T,T),(T,T,T,C),(T,T,C,T),(T,T,C,C),(T,C,T,T),(T,C,T,C),(T,C,C,T),(T,C,C,C),(C,T,T,T),(C,T,T,C),(C,T,C,T),(C,T,C,C),(C,C,T,T),(C,C,T,C),(C,C,C,T),(C,C,C,C). Il risultato è: "A") 0.0625; "B") 0.375; "C") 0.25; "D") 0

<15> OGGETTO_004 e' un dataframe formato da due colonne. La seconda colonna contiene un carattere che indica a quale di 4 categorie appartiene il dato riportato in colonna 1. Calcolare la somma del numero di elementi appartenenti alle categorie B e D e salvare il risultato in una variabile x utilizzando un'unica istruzione R.

<16> E' stato applicato il test t per un campione al fine di verificare se la media della variabile V1 sia uguale a 0 (H0: il valor medio della variabile V1 è pari a 0; HA: il valor medio della variabile V1 non è uguale a 0). Il valore della statistica test (t) calcolato è pari a 0.7381324 (9 df) ed il p-value = 0.4792378. Assumendo una soglia di significatività alfa = 0.05 è possibile rifiutare H0? "A") vero; "B") falso

<17> Effettuare un t test su un set di 100 valori campionati dalla normale con media=7.2 e deviazione standard=1 testando l'ipotesi che la media del campione sia pari a 6.8, salvare l'intervallo di confidenza al 95% per mu in una variabile x utilizzando un'unica istruzione R.

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

<18> In riferimento al grafico a torta in figura "OGGETTO_010_g.jpg" quale sarebbe la frequenza relativa della categoria C3 se venisse esclusa la categoria C2? "A") 0.08860759; "B") 0.06086957; "C") 0.34259259; "D") 0.1666667

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari