|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| EXAMEN PRÁCTICO MÓDULO 1  Instrucciones   * Deben realizar los ejercicios en RMarkdown, en el que además de los códigos de R para ejecutar los ejercicios, incluirán los comentarios que consideren oportunos. Por lo tanto, ***deben entregarlo en formato HTML***. * Deben incluir en el RMarkdown todos los procesos que hayan realizado. * La resolución de sus casos prácticos deberán entregarse a través de la aplicación del curso virtual en el enlace del menú izquierdo “TAREAS” (conforme a las instrucciones dadas en el foro). * La fecha máxima de entrega del trabajo es el 15 de marzo. Pueden subirlo antes si lo tienen terminado.   En la prueba de ***Introducción a R***:   * Se deben emplear las **funciones básicas** de *R* para tratar los datos. Así pues, no está permitido usar otros paquetes, como: dplyr, data.table, etc. * Para generar los **gráficos**, se deben emplear exclusivamente las funciones propias del paquete graphics. Por lo tanto, no está permitido utilizar otros paquetes, como: ggplot2, lattice, etc.   En la prueba de ***preprocesado***, es necesario hacer comentarios aclaratorios de lo que se hace y se obtiene, así como interpretar los resultados.  Prueba de R  Vectores   1. Cree el vector *A* utilizando la función seq, de tal modo que se obtenga el siguiente resultado: 15, 12, 9, 6, 3, 0, -3, -6, -9 ¿Qué longitud tiene este vector? ¿Cuál es su suma? ¿Cuál es su valor medio? ¿Cuál es el *error típico de la media*? El *error típico de la media* se obtiene dividiendo la *desviación típica* entre la raíz cuadrada del *tamaño* de la muestra (número de elementos). 2. Utilizando las funciones rep y seq, cree el vector *B*, con los siguientes datos: 2, 2, 2, 2, 2, 5, 7.5, 10, 12.5 Calcule: la suma de *A* y *B*, la diferencia entre *A* y *B*, A2+B2−−−−−−−√A2+B2.   Matrices  Se dispone de la siguiente tabla de datos:   | **PESO** | **ESTATURA** | **EDAD** | | --- | --- | --- | | 68 | 1.73 | 22 | | 56 | 1.68 | 24 | | 81 | 1.82 | 21 | | 74 | 1.78 | 25 |  1. Cree la matriz *M* con estos datos, indicando el nombre de cada columna. Muestre la dimensión de la matriz. Utilizando la función apply, obtenga un promedio de los resultados por columna.   Dataframes   1. Con los datos de la matriz anterior, cree el dataframe *datos*. Añada dos columnas, con el *nombre* de los individuos: “Andrés”, “Lucía”, “Marcos”, “Antonio”; y el *sexo*: “Hombre” o “Mujer”. Añada un nuevo individuo al conjunto de datos: “Carmen”, peso=59, estatura=1.74 y edad=23. La variable *nombre* es de tipo carácter, *sexo* es un factor y el resto son numéricas. Muestre la estructura del dataframe *datos*. 2. Del dataframe creado, extraiga: el peso de *Carmen*, todos los *nombres*, los datos de *Marcos* y los nombres de los que tengan más de 23 años. Añada una columna al dataframe *datos*, denominada *ALTURA*. De tal modo que, si el individuo mide más 1.78, tendrá el valor “Alto”; y de lo contrario, tendrá el valor “Normal”.   Listas   1. Cree la lista *L*, con los objetos que ha creado: *A, B, M, datos*. Encuentre en la lista: los valores de *A* que cumplan la condición A+3<0A+3<0, el valor [2,3] de la matriz *M*, y los datos de *Carmen* del dataframe *datos*.   Funciones y Bucles   1. Cree la función *normalizar*, que realice la *normalización estándar* (restar la media y dividir por la desviación típica) y la *normalización min-max* (restar el mínimo y dividir por la diferencia entre el máximo y el mínimo). Dicha función devolverá una lista. Utilice esta función empleando como argumento el PESO del dataframe *datos*, y después empleando la ESTATURA. 2. Dadas las siguientes instrucciones:   df <- as.data.frame(HairEyeColor, stringsAsFactors = FALSE)  str(df)  ## 'data.frame': 32 obs. of 4 variables:  ## $ Hair: chr "Black" "Brown" "Red" "Blond" ...  ## $ Eye : chr "Brown" "Brown" "Brown" "Brown" ...  ## $ Sex : chr "Male" "Male" "Male" "Male" ...  ## $ Freq: num 32 53 10 3 11 50 10 30 10 25 ...  Cree un bucle for que recorra cada columna del dataframe df, si la columna es de tipo *carácter*, convierta la variable en un *factor*. Utilice la función is.character para comprobar si la columna es de tipo *carácter*.  Dataframes y Gráficos  En los siguientes ejercicios trabajará con el dataset mtcars, que consta de 32 coches y 11 variables. Descripción de las variables que se van a utilizar:   * mpg: Millas recorridas por galón de combustible. * cyl: Número de cilindros. * hp: Caballos de fuerza. * wt: Peso (1000 libras). * qsec: Tiempo empleado en recorrer 1/4 de milla (segundos).  1. ¿Cuáles son los 5 coches más pesados? ¿Qué coches tienen 8 cilindros y 175 o más caballos de fuerza? (ordene el resultado por caballos de fuerza) ¿Cuál es el coche más rápido? ¿Cuál es el coche que menos combustible consume? ¿Cuál es el consumo medio de los coches de 4 cilindros? 2. Cree un *histograma* del peso de los coches, y mediante la función abline dibuje: una línea vertical (roja) en la *media* de la distribución y otra línea vertical (verde) en la *mediana* de la distribución. 3. Cree un gráfico que muestre la relación entre el *peso* y las *millas por galón*: mpg en función de wt. Interprete el gráfico. 4. Convierta la variable cyl en un factor. A continuación, cree un *diagrama de caja* que represente los *caballos de fuerza* en función del *número de cilindros*. Interprete el gráfico.   Prueba de Preprocesado  La base de datos contiene 1.500 registros que representan a los clientes de un banco que demandaron un crédito personal en el año 2019. Este conjunto de datos contiene seis variables numéricas y cuatro variables categóricas además de la variable CLASE que muestra la información de si se ha devuelto o no el crédito solicitado. La descripción de la base de datos es la siguiente:  **Variables numéricas**:   * *NUM\_FAMILIA*: número de miembros de la unidad familiar. * *INVERSION*: importe de la inversión. * *IMPCUOTA*: importe de la cuota del préstamo. * *VALVIV*: valor neto de las viviendas propias. * *INGRESOS*: es la suma de los ingresos anuales fijos, los variables y otros ingresos provenientes de otras fuentes. * *SALDO*: consta del saldo medio a la vista y el saldo medio no vista.   **Variables categóricas**:   * *CIVIL*: estado civil de la persona que pide el préstamo: 1. Casado; 2. Separado; 3. Soltero. * *TIPO\_TRABAJO*: El tipo de trabajo se ha recodificado en las siguientes categorías: 1. Fijo; 2. Temporal; 3. Autónomo; 4. Pensionista; 5. Otros. * *VIVIENDA*: modalidades de la vivienda: 1. Propiedad libre de cargas; 2. Propiedad hipotecada; 3. Alquiler; 4. Domicilio con la familia; 5. Otros. * *NACIONALIDAD*: nacionalidad del que solicita el préstamo: 1. Español; 2. Extranjero. * *CLASE*: Toma el valor SÍ, si se ha devuelto el crédito y NO, si no se ha devuelto.   Se solicita que se realicen los siguientes análisis:   1. Análisis descriptivo de las variables. 2. Tablas de datos. 3. Análisis gráfico. 4. Tabla de correlaciones. 5. Imputación de datos ausentes. 6. Equilibrado de la muestra.   https://ssl.gstatic.com/ui/v1/icons/mail/images/cleardot.gif   |  |  | | --- | --- | | https://ssl.gstatic.com/ui/v1/icons/mail/no_photo.png | ResponderReenviar | |  |  |