

## TAREA NÚMERO 3

- Las tareas son estrictamente de carácter individual, tareas idénticas se les asignará cero puntos.
- Todas las tareas tienen el mismo valor en la nota final del curso, es decir, el promedio de las notas obtenidas en la tareas será la nota final del curso.
- Todas las preguntas tienen el mismo puntaje.
- Incluir al menos 3 casos de prueba en las funciones programadas (cuando corresponda).

1. Dado  $x = (7, -5, 31, -2, -9, 10, 0, 18)$  y dado  $y = (2, 2, -3, 1, -99, -10, 10, -7)$  realice lo siguiente:

- Introduzca  $x$  y  $y$  como vectores en R.
- Calcule la media, la varianza, la raíz cuadrada y la desviación estándar de  $y$ .
- Calcule la media, la varianza, la raíz cuadrada y la desviación estándar de  $x$ .
- Calcule la correlación entre  $x$  y  $y$ .
- Escriba un comando en R para extraer las entradas 2 a la 7 de  $x$ .
- Escriba un comando en R para extraer las entradas de  $y$  excepto la 2 y la 7.
- Escriba un comando en R para extraer las entradas de  $y$  menores a -3 o mayores a 10.
- Escriba un comando en R para extraer las entradas de  $x$  mayores a 0 y que sean números pares.

2. Introduzca en R la siguiente matriz a  $4 \times 3$  usando:

```
A = matrix(c(1,2,3,4,5,6,7,8,9,10,11,12),nrow=4,"byrow"="true")
```

Luego, obtenga algunos elementos de la matriz de la siguiente manera:  $A[1,1:3]$ ,  $A[1:4,2]$ ,  $A[3,3]$ ,  $A[11]$ ,  $A[20]$ ,  $A[5,4]$ ,  $A[1,1,1]$  y explique qué pasa en cada caso.

3. Investigue para qué sirven los comandos de R `as.matrix(...)` y `as.data.frame(...)`, explique y dé un ejemplo de cada uno.
4. Introduzca usando código R (no archivos) en un `DataFrame` la siguiente tabla de datos:

Peso	Edad	Nivel Educativo
56	28	Lic
67	23	Bach
55	19	Bach
57	18	Bach
87	57	Dr
48	13	MSc

5. En muchas ocasiones nos interesa hacer referencia a determinadas partes o componentes de un vector. Defina el vector  $x = (2, -5, 4, 6, -2, 8)$ , luego a partir de este vector defina instrucciones en R para generar los siguientes vectores:

- $y = (2, 4, 6, 8)$ , así definido  $y$  es el vector formado por las componentes positivas de  $x$ .
- $z = (-5, -2)$ , así definido  $z$  es el vector formado por las componentes negativas de  $x$ .
- $v = (-5, 4, 6, -2, 8)$ , así definido  $v$  es el vector  $x$  eliminada la primera componente.
- $w = (2, 4, -2)$ , así definido  $w$  es el vector  $x$  tomando las componentes con índice impares, es decir,  $x[1] = 2, x[3] = 4$  y  $x[5] = -2$ .

6. Cargue en un `DataFrame` el archivo `EjemploAlgoritmosRecomendacion.csv` usando el siguiente comando de R:

```
Datos <- read.table('potabilidad_V2.csv',  
                    header=TRUE, sep=',',dec='.',row.names=1)
```

y haga lo siguiente:

- Calcule la dimensión de la Tabla de Datos.
- Despliegue las primeras 2 columnas de la tabla de datos.
- Ejecute un “summary” y un “str” de los datos.
- Calcule la Media y la Desviación Estándar para todas las variables cualesquiera.
- Ahora repita los ítems anteriores pero leyendo el archivo como sigue:

```
Datos <- read.table('potabilidad_V2.csv',  
                    header=TRUE, sep=';',dec=',',row.names=1)
```

Explique porqué todo da mal o genera error.

7. Cargue la tabla de datos que está en el archivo `SAheartv.csv` haga lo siguiente:

- Calcule la dimensión de la Tabla de Datos.
- Despliegue las primeras 3 columnas de la tabla de datos.
- Ejecute un `summary` y un `str` de los datos.
- Usando el comando `cor` de R calcule la correlación entre las variables `tobacco` y `alcohol`.
- Calcule la suma de las columnas con variables cuantitativas (numéricas).
- Calcule para todas las variables cuantitativas presentes en el archivo `SAheart.csv`: El mínimo, el máximo, la media, la mediana y para la variables `chd` determine la cantidad de Si y de No.

8. Programe en R una función que genera 200 números al azar entre 1 y 500 y luego calcula cuántos están entre el 50 y 450, ambos inclusive.

9. Desarrolle una función que calcula el costo de una llamada telefónica que ha durado  $t$  minutos sabiendo que si  $t < 1$  el costo es de 0,4 dólares, mientras que para duraciones superiores el costo es de  $0,4 + (t - 1)/4$  dólares, la función debe recibir el valor de  $t$ .
10. Desarrolle una función que recibe una matriz cuadrada  $A$  de tamaño  $n \times n$  y calcula su traza, es decir, la suma de los elementos de la diagonal. Por ejemplo, la traza de la siguiente matriz:

$$\begin{pmatrix} 9 & 30 & 4 \\ 1 & 30 & -1 \\ 4 & 12 & -2 \end{pmatrix}$$

es 37.

11. Escribir una función que genere los  $n$  primeros términos de la serie de Fibonacci.
12. Escriba una función que retorne cuál es el mayor número entero cuyo cuadrado no excede de  $x$  donde  $x$  es un número real que se recibe como parámetro, utilizando **while**.
13. Crear un Data Frame con diez alumnos con su edad, año de nacimiento y número de teléfono. Deberá aparecer el nombre de la columna (edad, año de nacimiento, teléfono) y el nombre de la fila, que será el nombre del alumno al que corresponden los datos.
14. Desarrolle una función **R** que recibe un **DataFrame** que retorna la cantidad de entradas de este DataFrame que son divisibles entre 3.
15. Desarrolle una función **R** que recibe un **DataFrame** y dos números de columna y que retorna en una lista el nombre de las variables correspondientes a las columnas, la covarianza y la correlación entre esas dos variables.
16. Programe la siguiente función recursiva:

$$U(n) = \begin{cases} 5 & si \ n = 0 \\ -5 & si \ n = 1 \\ 2 & si \ n = 2 \\ U_{n-1} - 11U_{n-2} + 2U_{n-3} & si \ n \geq 3 \end{cases}$$

17. Programe usando recursión una función, denominada **collatz(x)**, que retorne la **cantidad** de repeticiones necesarias para que el valor llegue a 1 según la **Conjetura de Collatz**. Esta conjetura dice que, sea  $n$  un número inicial, tras un número finito de repeticiones de la operación se llega a 1. La operación se puede resumir en lo siguiente:
  - a) Toma cualquier número entero positivo  $n$ .
  - b) Si  $n$  es par, divida  $n$  entre 2.
  - c) Si  $n$  es impar, multiplique  $n$  por 3 y sume 1, es decir,  $3n + 1$ .
  - d) Repite el proceso indefinidamente hasta que el valor sea 1.

Luego calcule: **collatz(16)**, **collatz(x12)** y **collatz(31)**.

18. Un número perfecto es un número entero positivo cuya suma de todos sus factores enteros positivos es igual a el mismo. Por ejemplo, el 6 es un número perfecto ya que sus factores son (1, 2, 3) y la suma de estos es 6. Se clasifica como abundante si la suma de los factores es mayor al número y defectivo si la suma es menor al número. Escriba una función que contenga lo siguiente:

- a) `es.factor`: Determina si un valor es factor o no de otro.
- b) `suma.factor`: Obtiene la suma de los factores de un número. Debe programarla con recursión.
- c) `clasificacion`: Retorne uno de los siguientes tres símbolos: **'perfecto'**, **'defectivo'** o **'abundante'**.

**Entregables:** Debe entregar un documento autoreproducible HTML con todos los códigos y salidas.



**oldemar** **rodríguez**  
CONSULTOR en MINERÍA DE DATOS