

LABORATORIO 1 (REPASO DE R)

1. Crear las matrices $M = \begin{pmatrix} 4 & 3 & -2 \\ 1 & 2 & -1 \end{pmatrix}$, M' , $A = MM'$, $B = M'M$.
 - a) Para M : calcular la suma de los elementos de cada columna, los promedios por filas, los promedios por columnas. Obtener la matriz N que resulta de concatenar M con la columna $(5 \ 10)'$.
 - b) Extraer de M la submatriz cuyas columnas tienen como primer elemento un numero mayor o igual a 2.
 - c) Para A y para B : extraer la primera columna, extraer la segunda fila, extraer el elemento 22, hallar el determinante, invertirla (si se puede), hallar sus valores propios y subespacios propios, su descomposición en valores singulares, construir una lista que contiene las entradas de la matriz, el determinante y los valores propios.
2. Crear un data.frame `miejemplo` con 5 columnas (4 numéricas y 1 categórica) y 6 observaciones (o sea 6 filas).
 - a) Dar nombres a las filas y nombres a las columnas.
 - b) Añadir a este objeto una columna que sea la suma de las 3 primeras columnas numéricas y otra columna que indica con 0/1 si el valor numérico de la segunda columna es mayor que 5 o no. Dar un nombre a estas nuevas columnas.
 - c) Borrar la primera fila y la última columna. Dar un nombre a este nuevo objeto.
 - d) Hacer un resumen estadístico de los datos de este data frame cuando ésto tiene sentido.
 - e) Escribir `miejemplo` en un archivo de texto `miejemplo.txt`. Borrar el objeto de R. Cargar este archivo en el objeto `miejemplo2`.
 - f) Repetir el paso anterior con la base de datos de Iris.
3. Crear un archivo texto con 5 columnas (4 numéricas y 1 catégorica con caracteres) y 6 observaciones (o sea 6 filas) añadiendo una primera fila con el nombre que le quiera dar a cada columna.
 - a) Importar este archivo en R, en un objeto del tipo data.frame y verificar los nombres de las columnas, de las filas, las dimensiones del data.frame.
 - b) Dar un nombre a cada fila.
 - c) Añadir una columna cuyos valores indica con 0/1 si el valor numérico de la segunda columna es mayor que 6 o no.

- d) Exportar el nuevo data.frame en un nuevo archivo de texto, sin escribir los nombres de las filas ni de las columnas. Volver a importar a R.
4. Utilizar la función `runif` para generar 1000 enteros uniformemente comprendidos entre 1 y 100.
- Con la función `cut`, discretizar este vector en 4 niveles: Deficiente ("D"), Regular (R"), Bueno ("B"), Muy Bueno ("MB") si pertenece a $[0, 24]$, $[25, 49]$, $[50, 74]$, $[75, 100]$ respectivamente.
 - Para cada uno de los datos, hacer aparecer el nivel al que pertenece.
 - Calcular la cantidad de valores que pertenece a cada uno de los niveles y la distribución.
5. Se considera la función $f(x) = \frac{1}{x^2 + 1} + \sqrt{|x + 2|}$
- Crear un vector V con los reales entre 0 y 10 con paso de 0.5.
 - Calcular los valores de $f(x_i)$ para todas las coordenadas x_i de V .
 - Calcular el valor máximo de $f(x_i)$ y hallar en que punto se alcanza.
 - Construir la función que calcula los valores de f .
 - Trazar la curva que representa a la función en $[-4, 3]$
 - Dar un valor aproximado del mínimo y del máximo de f en $[-4, 3]$.
6. Sea X una variable aleatoria con distribución normal $\mathcal{N}(0, 4)$.
- Calcular $\mathbb{P}(X \leq 1,2)$, $\mathbb{P}((X > 2,1)$, $\mathbb{P}(0,4 < X < 1)$
 - Hallar el x para el cual $\mathbb{P}(X \leq x) = 0,75$. Hallar el tercer quintíl.
 - Dibujar la función de densidad y la función de distribución de X .
 - Repetir las preguntas anteriores si X tiene distribución χ_3^2 .
7. En este ejercicio nos basamos en el Capítulo 1 del libro "A Handbook of Statistical Analyses using R" de B.S. Everitt and T.Hothorn. Cargar la base de datos Forbes 2000 que se encuentra en el paquete `HSAUR2`.
- ¿De qué clase es el objeto `Forbes2000`? ¿Cuál es su dimensión? ¿Cuáles son los nombres de las variables y de qué tipo es cada una de ellas?
 - Hacer un `summary` de las variables de `Forbes2000`
 - Retire los casos con datos faltantes.
 - Obtener el promedio y la desviación estándar de la variable "Sales".
 - Obtener la matriz de correlación entre las ventas "Sales" y los beneficios "Profits" y el scatterplot de ambas variables.
 - Encontrar los tres activos ("Assets") mayores e indicar a qué empresas pertenecen.
 - Realizar el histograma de la variable "Marketvalue".
 - Hacer el promedio de las ventas "Sales" para cada una de las Categorías ("Categories").