

LABORATORIO 1: INTRODUCCIÓN A R

1. Mirar la ayuda de la función `mean`, `range`, `length`, `sum`.
2.
  - a) Crear un vector  $x$  con los números 3.5, 2, -3, 4, 7.8 y 6.
  - b) Crear un vector  $y$  que contiene a  $x$  y que al principio tiene el número -10 y al final el 9.3.
  - c) Mirar las diferencias entre  $y * y$  e  $y \% * \%y$ .
  - d) Crear un vector  $m$  que contenga los elementos de  $y$  mayores que 3.
  - e) Crear el vector que contiene las coordenadas 2 a 5 de  $y$  y la 8.
3.
  - a) Crear una matriz  $4 \times 5$  con todas sus entradas iguales a 1.
  - b) Crear una matriz  $4 \times 5$  con todos los números del 1 al 20.
  - c) Crear un vector con 20 números (los que quieran) y a partir de él crear una matriz  $5 \times 4$  completándola por filas y otra  $2 \times 10$  completándola por columnas.
  - d) Generar dos matrices  $A$  y  $B$  que sean conformables (multiplicables). Observar el resultado de  $A \%* \%B$  y de  $A * B$
  - e) Crear una matriz  $C$   $2 \times 2$  que contenga los números 2, 3, 4 y 7 y a partir de ella crear una matriz  $D$  que contenga las entradas de  $C$  multiplicadas por 4. Concatenar por filas y por columnas las matrices  $C$  y  $D$ .
4. ¿Cuál es la diferencia entre  $1 : 30 - 1$  y  $1 : (30 - 1)$ ?  
Generar una secuencia del 20 al 1 y otra de 20 al 2 solo con números pares.
5. Dado el vector  $v = (4, 5, 9, 2, 1, 4)$  reemplazar los elementos impares por 1.
6. 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 número mayor o igual a 2.
  - c) Para  $A$  y para  $B$ : extraer la primera columna, extraer la segunda fila, extraer el elemento 2, 2, hallar el determinante, invertirla, hallar sus valores propios y subespacios propios, su descomposición en valores singulares.
  - d) Crear una matriz  $N$  de tamaño  $100 \times 50$  donde las entradas fueron generadas aleatoriamente y con iguales probabilidades en el conjunto  $\{0, 1\}$ ; en el conjunto  $[0, 1]$ . Repetir la pregunta a).
7. 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 éste 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.
8. 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.
9. Generar 1000 datos enteros uniformemente comprendidos entre 1 y 100.
- a) 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.
- b) Para cada uno de los datos, hacer aparecer el nivel al que pertenece.
- c) Calcular la cantidad de valores pertenecientes a cada uno de los niveles y la tabla de valores acumulados.
10. Se considera la función  $f(x) = \frac{1}{x^2 + 1} + \sqrt{|x + 2|}$
- a) 1) Crear un vector  $V$  que contiene los reales entre 0 y 10 con paso 0.5.  
 2) Calcular los valores de  $f(x_i)$  para todas las coordenadas  $x_i$  de  $V$ .  
 3) Calcular el valor máximo de  $f(x_i)$  y hallar en que punto se alcanza.
- b) Construir la función que calcula los valores de  $f$ .
- c) Trazar la curva que representa a la función en  $[-4, 3]$
- d) Dar un valor aproximado del mínimo y del máximo de  $f$  en  $[-4, 3]$ .
11. Sea  $X$  una variable aleatoria con distribución normal  $N(0, 2)$ .
- a) Calcular  $\mathbb{P}(X \leq 1,2)$ ,  $\mathbb{P}(X > 2,1)$ ,  $\mathbb{P}(0,4 < X < 1)$
- b) Hallar el  $x$  para el cual  $\mathbb{P}(X \leq x) = 0,75$ . Hallar el tercer quintíl.
- c) Dibujar la función de densidad y la función de distribución de  $X$ .
- d) Simular  $n = 1000$  variables aleatorias iid con la misma distribución de  $X$  y estimar las cantidades en a) y b), y las funciones en c). Comparar con los valores poblacionales.

- e) Repetir las preguntas anteriores si  $X$  tiene distribución  $\chi_3^2$ .
12. 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.
- a) ¿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?
  - b) Hacer un resumen (**summary**) de las variables de Forbes2000.
  - c) Retire los casos con datos faltantes.
  - d) Calcule el promedio y la desviación estándar de la variable **sales** con y sin datos faltantes.
  - e) Calcule la correlación entre las ventas (**sales**) y los beneficios (**profits**) y el diagrama de dispersión de ambas variables.
  - f) Encontrar los tres activos (**assets**) mayores e indicar a qué empresas pertenecen.
  - g) Realizar el histograma de la variable **marketvalue** y de su logaritmo.
  - h) Hacer el promedio de las ventas (**sales**) para cada una de las categorías (**category**).