

Máster universitari en Estadística i Investigació Operativa (MESIO)
Computación en Estadística y en Optimización
Test 1 con R (Grupo B)

Instrucciones:

- Bajar los ficheros GrBTestR1.R y wine.csv de ATENEA y guardarlos en un disco local o una memoria USB.
- Cambiar el nombre del script GrBTestR1.R a GrBTestR1_ApellidoNombre.R.
- Incluir en este script todas las instrucciones necesarias para resolver los ejercicios. Posibles comentarios (se valorarán) se pueden incluir detrás de una almohadilla (#).
- Entregar el script vía ATENEA o por correo electrónico a amonleong@ub.edu antes de las 15h.

Ejercicio 1 (4,5 puntos)

- a) Cread el siguiente vector **a** de longitud 49. Contad el número de elementos del vector para comprobar que es correcto:

```
## [1] 1 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
## [24] 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
## [47] 48 49 50
```

- b) Utilizad el vector **a** para construir la matriz cuadrada **A** con los últimos 25 elementos:

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  26  31  36  41  46
## [2,]  27  32  37  42  47
## [3,]  28  33  38  43  48
## [4,]  29  34  39  44  49
## [5,]  30  35  40  45  50
```

- c) Obtened 3 matrices aleatorias (**A1**, **A2** y **A3**) con remplazamiento a partir de los elementos de la matriz **A** con dimensión 50x50. Comprobad que la dimensión de la matriz obtenida es realmente 50X50 y que el máximo y el mínimo (de **A1**, **A2**, **A3**) están dentro del máximo y el mínimo de los elementos de la matriz **A**.
- d) Cread una matriz (**B**) de dimensión 1500 × 1500 con valores aleatorios de una distribución Normal con media $\mu = 40$, $\sigma = 14$. Representad el histograma del vector generado para comprobar que es realmente normal.
- e) Cread un vector **b** con números pares de longitud 48, consecutivos, y tomados de 4 en 4, como por ejemplo:

```
## [1] 2 6 10 14 18 22 26 30 34 38 42 46 50 54 58 62 66
## [18] 70 74 78 82 86 90 94 98 102 106 110 114 118 122 126 130 134
## [35] 138 142 146 150 154 158 162 166 170 174 178 182 186 190
```

- f) Extraed a partir de la matriz (B) una submatriz cuadrada a partir de las 48 primeras filas y 48 primeras columnas y nombradla como (C). Comprobad su dimensión y multiplicad todos los valores por 2 y renombradla como C.
- g) Sumad el vector **b** a la tercera fila de la matriz (C) y nombrad el resultado como matriz (C1). Comprobad el resultado obtenido.
- h) Extraed a partir de la matriz (B) una submatriz cuadrada con las 5 primeras filas y 5 primeras columnas (redondead a 2 decimales) y nombradla como matriz (D).Obtened la matriz traspuesta y la inversa de D. Comprobad que el producto de D y de su inversa es la matriz identidad, como en el siguiente ejemplo:

```
## [1] "Matriz D"
##      [,1] [,2] [,3] [,4] [,5]
## [1,] 33.45 28.21 27.07 31.08 48.57
## [2,] 42.13 33.31 15.94 56.34 45.65
## [3,] 60.42 24.92 48.46 55.53 56.87
## [4,] 31.71 27.73 39.88 64.56 29.05
## [5,] 46.79 44.82 67.70 40.35 64.25
## [1] "Matriz traspuesta de D"
##      [,1] [,2] [,3] [,4] [,5]
## [1,] 33.45 42.13 60.42 31.71 46.79
## [2,] 28.21 33.31 24.92 27.73 44.82
## [3,] 27.07 15.94 48.46 39.88 67.70
## [4,] 31.08 56.34 55.53 64.56 40.35
## [5,] 48.57 45.65 56.87 29.05 64.25
```

Ejercicio 2 (5,5 puntos)

- a) Leed el fichero `wine.csv` de manera correcta. Este fichero contiene una serie de análisis químicos a tres tipos de vinos, con las características químicas que los caracterizan. Importad correctamente el fichero a un dataframe con el nombre `wine`. Describid qué variables contiene y comprobad de alguna manera que la importación es correcta.
- b) Contad el número de valores missing (NA) total y por columna. Indicad en qué fila y columna se encuentran los valores missing (NA). Eliminad estos valores NA, renombrado el nuevo dataframe sin NA como `wine.sinNA`; comprobad que no queda ningún valor NA. A partir de ahora trabajad siempre con `wine.sinNA` sin NA (evitará errores).
- c) Utilizando el dataframe `wine.sinNA`. Indicad la cantidad de tipos de vino que aparecen analizadas: frecuencia y porcentaje por tipo de vino (variable `Wine`).
- d) Calculad el valor medio y la desviación estandar de la cantidad de alcohol por tipo de vino. Opcionalmente, puede listarse conjuntamente en una sóla línea como lo que se muestra a continuación para cada tipo de vino, como en el ejemplo que aparece a continuación.

```
## [1] "Vino tipo 1: Media alcohol=13.75 ,sd alcohol=0.44"  
## [1] "Vino tipo 2: Media alcohol=12.28 ,sd alcohol=0.54"  
## [1] "Vino tipo 3: Media alcohol=13.15 ,sd alcohol=0.53"
```

- e) Utilizando el *data frame* `wine.sinNA`, indicad qué vino tiene más contenido en cenizas (Ash) y listar todos sus valores, así como su tipo (1,2,3).