

Homework1_ds

Enrique J. De La Hoz D.

UTB

Homework 1

1. Vectores

- Construir el vector u:

```
## [1]  2  4  6  8 10 12 14 16
```

- Construir el vector v:

```
## [1]  2  4  8 16 32 64 128 256 512 1024 2048 4096
```

- A partir del vector u construir la matriz A de dimensión 2×4 :

```
##      [,1] [,2] [,3] [,4]
## [1,]    2    4    6    8
## [2,]   10   12   14   16
```

- A partir del vector v construir la matriz B de dimensión 4×3 :

```
##      [,1] [,2] [,3]
## [1,]    2   32  512
## [2,]    4   64 1024
## [3,]    8  128 2048
## [4,]   16  256 4096
```

- Construir la matrix C como producto de las matrices A y B.

```
##      [,1] [,2] [,3]
## [1,]  196 3136 50176
## [2,]  436 6976 111616
```

- Obtener la matriz D que tenga la siguiente estructura:

$$D = \begin{bmatrix} A & C \\ 0 & B \end{bmatrix}$$

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,]    2    4    6    8  196 3136 50176
## [2,]   10   12   14   16  436 6976 111616
## [3,]    0    0    0    0    2   32   512
## [4,]    0    0    0    0    4   64  1024
## [5,]    0    0    0    0    8  128  2048
## [6,]    0    0    0    0   16  256  4096
```

2. Vectores

El fichero states.txt contiene datos (registrados en 1977) de los 50 estados de los Estados Unidos. Las variables son las siguientes:

*stat: Estado

*pop: Población (en miles)

*inco: Salario medio (en dólares)

*illi: Tasa de analfabetismo (en % de la población)

*lifex: Esperanza de vida (en años)

*area: Superficie (en millas²)

*reg: Región

a) Importar los datos a R.

```
## 'data.frame': 50 obs. of 7 variables:
## $ stat : Factor w/ 50 levels "Alabama","Alaska",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ pop : int 3615 365 2212 2110 21198 2541 3100 579 8277 4931 ...
## $ inco : int 3624 6315 4530 3378 5114 4884 5348 4809 4815 4091 ...
## $ illi : num 2.1 NA 1.8 1.9 1.1 0.7 1.1 0.9 NA 2 ...
## $ lifex: num NA 69.3 70.5 70.7 71.7 ...
## $ area : int 50708 566432 113417 51945 156361 103766 4862 1982 54090 58073 ...
## $ reg : Factor w/ 4 levels "North Central",...: 3 4 4 3 4 4 2 3 3 3 ...
```

b) ¿Qué variable tiene más datos perdidos (missings) y cuántos son?

```
## illi
## 4
```

c) Ordenar los datos según la región (en orden alfabético) y superficie (en orden ascendente). ¿Cuál es la población del primer y del último estado, respectivamente?

```
## [1] 5313 365
```

d) ¿Cuál es la esperanza de vida media de los estados en las regiones del oeste y del sur, respectivamente?

```
## [1] 69.72786
```

```
## [1] 71.23462
```

e) Construir la matriz que contenga los coeficientes de correlación de Pearson para todos los pares de variables numéricas descartando todos los estados con algún dato perdido.

```
##      pop    inco    illi    lifex    area
## pop    1.000  0.311  0.119 -0.106  0.228
## inco    0.311  1.000 -0.503  0.463  0.002
## illi    0.119 -0.503  1.000 -0.576  0.013
## lifex -0.106  0.463 -0.576  1.000  0.039
## area    0.228  0.002  0.013  0.039  1.000
```