



UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA DE
SISTEMAS INFORMÁTICOS

Máster en Software de Sistemas Distribuidos y
Empotrados

Ciencia de Datos

Ejercicio 1 - Introducción a R

Alejandro Casanova Martín

N.º de matrícula: bu0383

Madrid, 21 de abril 2024

Índice

1.	Objetos en R	3
2.	Análisis estadístico básico	6

1. Objetos en R

- a) ¿Qué contiene la constante de R *LETTERS*?

```
> LETTERS
```

```
[1] "A" "B" "C" "D" "E" "F" "G" "H" "I" "J" "K" "L" "M" "N" "O" "P" "Q" "R" "S"  
[20] "T" "U" "V" "W" "X" "Y" "Z"
```

- b) Usar la función de R *class* para determinar qué tipo de objeto es.

```
> class(LETTERS)
```

```
[1] "character"
```

- c) Construir el vector *MisLetras* con la 20 primeras letras mayúsculas del alfabeto.

```
> MisLetras <- LETTERS[1:20]
```

```
> MisLetras
```

```
[1] "A" "B" "C" "D" "E" "F" "G" "H" "I" "J" "K" "L"  
[13] "M" "N" "O" "P" "Q" "R" "S" "T"
```

- d) Extraer de *MisLetras* el vector con los elementos del 5 al 14, ambos incluidos.

```
> MisLetras[5:14]
```

```
[1] "E" "F" "G" "H" "I" "J" "K" "L" "M" "N"
```

- e) Construir un vector con los 100 primeros números positivos pares.

```
> cien_pares <- 2*(1:100)
```

```
> cien_pares
```

```
[1]  2  4  6  8 10 12 14 16 18 20 22 24  
[13] 26 28 30 32 34 36 38 40 42 44 46 48  
[25] 50 52 54 56 58 60 62 64 66 68 70 72  
[37] 74 76 78 80 82 84 86 88 90 92 94 96  
...
```

- f) Construir la matriz 10x10 *MiMatriz* con los 100 primeros números positivos pares, introducidos por filas.

```
> MiMatriz <- matrix(cien_pares, 10, 10, TRUE)
```

```
> MiMatriz
```

```
  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]  
[1,]  2  4  6  8 10 12 14 16 18 20  
[2,] 22 24 26 28 30 32 34 36 38 40  
[3,] 42 44 46 48 50 52 54 56 58 60  
[4,] 62 64 66 68 70 72 74 76 78 80  
...
```

- g) Extraer de *MiMatriz* la primera fila y la segunda columna.

```
> MiMatriz[1,] # Obtener la primera fila
```

```
[1] 2 4 6 8 10 12 14 16 18 20
```

```
> MiMatriz[,2] # Obtener la segunda columna
```

```
[1] 4 24 44 64 84 104 124 144 164 184
```

- h) Usar la función de R *diag* para construir la matriz identidad 5 x 5 *MatrizId5*.

```
> MatrizId5 <- diag(5)
```

```
> MatrizId5
      [,1] [,2] [,3] [,4] [,5]
[1,]  1   0   0   0   0
[2,]  0   1   0   0   0
[3,]  0   0   1   0   0
[4,]  0   0   0   1   0
[5,]  0   0   0   0   1
```

- i) Construir la lista *MiLista* con los objetos *MisLetras*, *MiMatriz* y *MatrizId5*.

```
> MiLista <- list(MisLetras, MiMatriz, MatrizId5)
> MiLista
[[1]]
 [1] "A" "B" "C" "D" "E" "F" "G" "H" "I" "J" "K" "L"
[13] "M" "N" "O" "P" "Q" "R" "S" "T"
```

```
[[2]]
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]   2   4   6   8  10  12  14  16  18  20
[2,]  22  24  26  28  30  32  34  36  38  40
[3,]  42  44  46  48  50  52  54  56  58  60
[4,]  62  64  66  68  70  72  74  76  78  80
...
```

```
[[3]]
      [,1] [,2] [,3] [,4] [,5]
[1,]  1   0   0   0   0
[2,]  0   1   0   0   0
[3,]  0   0   1   0   0
[4,]  0   0   0   1   0
[5,]  0   0   0   0   1
```

- j) Construir el frame *MiFrame* con las variables '*Nombre*', '*Edad*', '*Carrera*' y '*Curso*', rellenándolo con los datos descritos a continuación:

'Juan'	20	'Informática'	3
'Ana'	19	'Matemáticas'	2
'Cristina'	22	'Periodismo'	1
'Alberto'	18	'Telecomunicaciones'	1
'Hugo'	21	'Informática'	4

```
> Nombre <- c("Juan","Ana","Cristina", "Alberto", "Hugo")
> Edad <- c(20, 19, 22, 18, 21)
> Carrera <- c('Informática', 'Matemáticas', 'Periodismo', 'Telecomunicaciones',
  'Informática')
> Curso <- c(3, 2, 1, 1, 4)
> MiFrame = data.frame(Nombre, Edad, Carrera, Curso)
> MiSubframe
```

	Edad	Curso
1	20	3
2	19	2
3	22	1
4	18	1
5	21	4

k) Extraer de *MiFrame* la variable '*Edad*':

```
> MiFrame$Edad
[1] 20 19 22 18 21
```

l) Extraer de *MiFrame* las variables '*Edad*' y '*Curso*' y asignarlas al frame *MiSubframe*:

```
> MiSubframe <-subset.data.frame(MiFrame, select=c(Edad, Curso))
> MiSubframe
```

	Edad	Curso
1	20	3
2	19	2
3	22	1
4	18	1
5	21	4

2. Análisis estadístico básico

- a) Calcular la tabla de frecuencia de la variable 'Carrera' del frame MiFrame:

```
> table(MiFrame$Carrera)
```

Informática	Matemáticas
2	1
Periodismo	Telecomunicaciones
1	1

- b) Obtener la tabla de frecuencia relativa de la variable 'Curso' del frame *MiFrame*:

```
> prop.table(table(MiFrame$Curso))
```

1	2	3	4
0.4	0.2	0.2	0.2

- c) Calcular la media de la variable 'Edad':

```
> mean(MiFrame$Edad)
```

```
[1] 20
```

- d) Calcular mínimo, 1^{er} cuartil, mediana, media, 3^{er} cuartil y máximo de la variable 'Edad':

```
> summary(MiFrame$Edad)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
18	19	20	20	21	22

- e) Calcular la desviación típica y la varianza de la variable 'Edad':

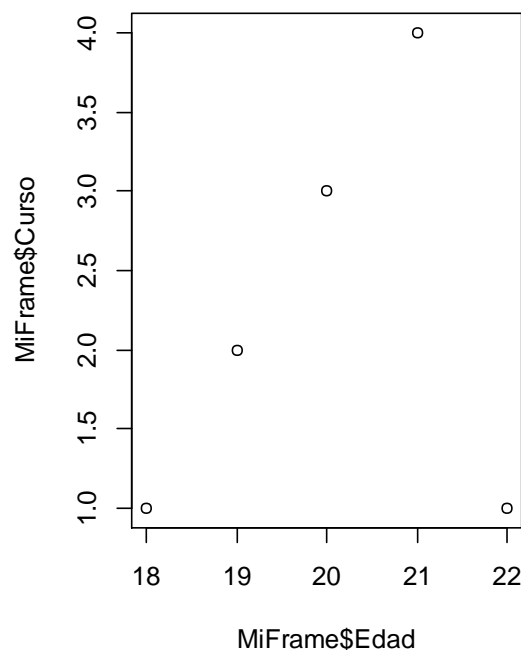
```
> var(MiFrame$Edad); sd(MiFrame$Edad)
```

```
[1] 2.5
```

```
[1] 1.581139
```

- f) Representar gráficamente las variables ('Edad', 'Curso'):

```
> plot(MiFrame$Edad, MiFrame$Curso)
```



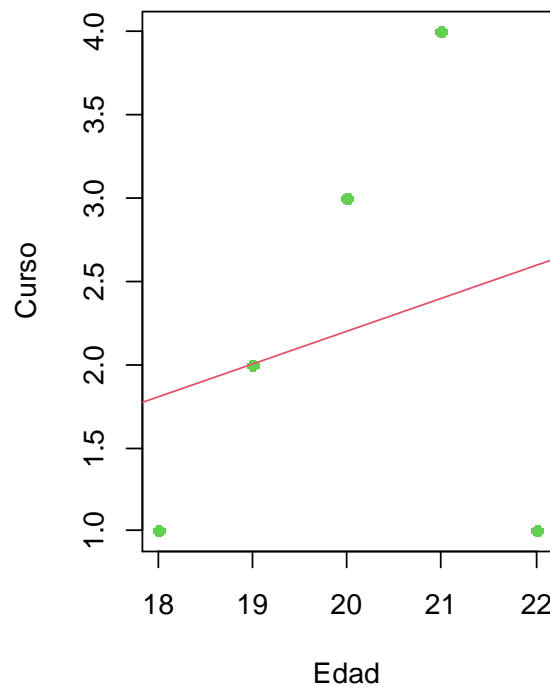
- g) Calcular (sin usar la función *lm*) los parámetros de la recta de regresión ($y = ax + b$) de las variables ('Edad','Curso'):

$$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x}) \quad \Rightarrow \quad a = \frac{\sigma_{xy}}{\sigma_x^2}, \quad b = \bar{y} - a\bar{x}$$

```
> a <- cov(MiFrame$Edad, MiFrame$Curso) / var(MiFrame$Edad)
> b <- mean(MiFrame$Curso) - a * mean(MiFrame$Edad)
> a; b
[1] 0.2
[1] -1.8
```

- h) Representar gráficamente las variables ('Edad','Curso') y la recta de regresión con un color distinto al de los marcadores:

```
> plot(MiFrame$Edad, MiFrame$Curso, xlab = 'Edad', ylab = 'Curso', col = 3, pch =
      16, cex = 1)
> abline(b, a, col=2)
```



- i) Cargar los datos del archivo *medidas_cuerpo.csv* y comprueba la estructura del *dataframe* resultante:

```
> repo <- "C:/Users/alex/Desktop/Máster Software Embebido/2 Segundo Semestre/2
      Ciencia de Datos/Ejercicios"
> setwd(repo); getwd()
[1] "C:/Users/alex/Desktop/Máster Software Embebido/2 Segundo Semestre/2
      Ciencia de Datos/Ejercicios"
> datosCSV <- as.data.frame(read.csv("Ficheros/medidas_cuerpo.csv",
      header=TRUE, sep='t', comment.char="#"))
> head(datosCSV)
  edad peso  altura  sexo muneca biceps
1  43  87.3  188.0 Hombre   12.2   35.8
```

2	65	80.0	174.0	Hombre	12.0	35.0
3	45	82.3	176.5	Hombre	11.2	38.5
4	37	73.6	180.3	Hombre	11.2	32.2
5	55	74.1	167.6	Hombre	11.8	32.9
6	33	85.9	188.0	Hombre	12.4	38.5

- j) Representar gráficamente las variables (*peso*, *altura*) y la recta de regresión, poniendo títulos a los ejes y cambiando los parámetros de forma, tamaño y color de los marcadores:

```
> plot(datosCSV$peso, datosCSV$altura, xlab = 'peso', ylab = 'altura', col = 3, pch = 16, cex = 1)
```

```
> RectaRegresion2 <- lm(datosCSV$altura~datosCSV$peso)
```

```
> RectaRegresion2
```

Call:

```
lm(formula = datosCSV$altura ~ datosCSV$peso)
```

Coefficients:

(Intercept)	datosCSV\$peso
132.6551	0.5642

```
> abline(RectaRegresion2)
```

