

**Curso:** Machine Learning en Agricultura.

**Tutor:** Hugo Andrés Dorado Betancourt

**Taller No 1:** Introducción a R

**1. Dado los siguientes vectores,**

`> x<-c(4,2,6)`

`> y<-c(1,0,-1)`

Intente predecir cuáles serán los resultados de las siguientes

(a) `length(x)`

(d) `x+y`

(g) `x^2`

(b) `sum(x)`

(e) `x*y`

(c) `sum(x^2)`

(f) `x-2`

**2. Crear las siguientes matrices:**

$$x = \begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix} \quad y = \begin{bmatrix} 1 & 4 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

Calcula las siguientes operaciones y chequear los resultados:

(a) `2*x`

(b) `x*x`

(c) `x%%*%x`

3. Utilizando las matrices anteriores, con x e y, calcule el efecto de las siguientes operaciones de subíndice y verifique sus respuestas en R.

(a) `x[1,]`

(c) `x[,2]`

(e) `y[,2:3]`

(b) `x[2,]`

(d) `y[1,2]`

**4. Usando un poco de matemáticas básicas podemos demostrar que:**

La media aritmética no es más que la suma de todos los elementos dividida por el número de elementos

$$\bar{x} = \frac{1}{n} \sum_{i=1}^N x_i.$$

La varianza se calcula de la siguiente manera:

$$\text{Var}(X) = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

La desviación estándar es la raíz cuadrada (sqrt ()) de la varianza

$$SD(x) = \sqrt{\text{VAR}(x)}.$$

Intente calcular la media aritmética y desviación estándar usando sum (), length () y sqrt compare el resultado con el resultado de mean (), var(), sd().

**5. Cree algunas matrices dado un vector de datos creado por replicación (consulte Replicar elementos del capítulo anterior).**

A continuación puede ver los tres vectores de datos (data\_A, data\_B, data\_C)

```
data_A <- rep(c(-1, 0, 1), 5)           # For matrix A
data_B <- rep(c(-3, -2, -1, 0, 1, 2, 3), each = 3) # For matrix B
data_C <- rep(c(1, 0, 0, 0, 0), length.out = 16) # For matrix C
```

Reproduzca a partir de código las siguientes matrices.

```
A # Matrix A
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  -1  -1  -1  -1  -1
## [2,]   0   0   0   0   0
## [3,]   1   1   1   1   1

B # Matrix B
##      [,1] [,2] [,3]
## [1,]  -3  -3  -3
## [2,]  -2  -2  -2
## [3,]  -1  -1  -1
## [4,]   0   0   0
## [5,]   1   1   1
## [6,]   2   2   2
## [7,]   3   3   3

C # Matrix C
##      [,1] [,2] [,3] [,4]
## [1,]   1   0   0   0
## [2,]   0   1   0   0
## [3,]   0   0   1   0
## [4,]   0   0   0   1
```

**6. Suponga que realiza un seguimiento de su kilometraje cada vez que aumenta. En sus últimos 6 recargas de gasolina, el kilometraje fue.**

65311 65624 65908 66219 66499 66821 67145 67447

```
> miles = c(65311, 65624, 65908, 66219, 66499, 66821, 67145, 67447)
> x = diff(miles)
```

Ingresa estos números en R. Use la función `diff` en los datos. ¿Qué da?

Debe ver el número de millas entre recargas de gasolina. Utilice el `max` para encontrar el número máximo de millas entre recargas de gasolina, la función `mean` para encontrar el número promedio de millas y el `min` para obtener el número mínimo de millas.

**7. Sea el dato  $x$  dado por**

$x = c(1, 8, 2, 6, 3, 8, 5, 5, 5, 5)$

Utilice R para calcular las siguientes funciones. Tenga en cuenta que usamos  $X_1$  para denotar el primer elemento de  $x$  (que es 0), etc.

1.  $(X_1 + X_2 + \dots + X_{10}) = 10$  (use la suma)
2. Encuentre  $\log_{10}(X_i)$  para cada  $i$ . (Utilice la función de registro que por defecto es la base  $e$ )
3. Encuentre  $(X_i - 4.4) = 2.875$  para cada  $i$ . (Hazlo todo a la vez)
4. Encuentra la diferencia entre los valores más grande y más pequeño de  $x$ . (Este es el rango. Puede usar `max` y

`min` o adivina un comando incorporado).

**8. Corrija cada uno de los siguientes errores comunes de subconjunto de marcos de datos:**

```
mtcars[mtcars$cyl = 4, ]
# use `==` (instead of `=`)

mtcars[-1:4, ]
# use `-(1:4)` (instead of `-1:4`)

mtcars[mtcars$cyl <= 5]
# `,` is missing
```

**9. ¿Por qué `mtcars[1:20]` devuelve un error? ¿En qué se diferencia de `mtcars[1:20,]`?**

**10. Crear el siguiente `data.frame`**

	Age	Height	Weight	Sex
Alex	25	177	57	F
Lilly	31	163	69	F
Mark	23	190	83	M
Oliver	52	179	75	M
Martha	76	163	70	F
Lucas	49	183	83	M
Caroline	26	164	53	F

Agregar la siguiente columna

	Working
Alex	Yes
Lilly	No
Mark	No
Oliver	Yes
Martha	Yes
Lucas	No
Caroline	Yes

- ¿Cuántas filas y columnas tiene el nuevo marco de datos?
- ¿Qué clase de datos hay en cada columna?
- Calcular en una columna el índice de masa corporal IMC utilizando la siguiente formula:

$$IMC = \frac{PESO}{ALTURA \times ALTURA}$$