

Clase de programación en R - N⁰.5

Victor Augusto Lizcano Sandoval

November 24, 2021

Funciones apply

Las funciones apply son usadas para aplicar funciones a elementos de una estructura de datos (matrices, data frames, arrays y listas).

Función apply

Para matrices, arreglos y dataframes

```
df <- data.frame(x = 1:4, y = 5:8, z = 10:13)
```

```
apply(X = df, MARGIN = 1, FUN = sum)
```

MARGIN = 1 aplica para filas

MARGIN = 2 aplica para columnas

Función lapply

Para listas. Recibe una lista y devuelve una lista

```
A <- matrix(1:9,nrow = 3, ncol = 3)
B <- matrix(11:19,nrow = 3, ncol = 3)
C <- matrix(21:29,nrow = 3, ncol = 3)
miLista <- list(A,B,C)
lapply(miLista,"[,1,1)
```

Función sapply

Para listas. Recibe una lista y devuelve un vector

```
sapply(miLista,"[,1,1)
```

Función tapply

Para factores

```
x <- 1:20  
y <- factor(rep(letters[1:5], each = 4))  
tapply(x, y, sum)
```

Función mapply

Para matrices. Devuelve una lista o un vector

```
mapply(sum, 1:5, 1:5, 1:5) # Suma el primer elemento de cada  
vector, después el segundo y luego el tercero  
mapply(rep, 1:4, 4:1) # Repite cada elemento del primer vector el  
número de veces que indique el segundo vector.
```

Función vapply

Para listas. Devuelve un vector

```
x <- list(A = 1, B = 1:3, C = 1:7)
vapply(x, FUN = length, FUN.VALUE = 0L)
```


Librería reshape

Para transformar la estructura de los datos

```
library(reshape)
str(mtcars)
mtcars2 = melt(mtcars,id=c("mpg","cyl"))
cast(mtcars2, mpg ~ variable, mean)
```

Librería readxl

Para leer datos de Excel

```
library(readxl)
misDatos <- read_excel("mi_archivo.xlsx", sheet = "Hoja 1")
```

Texto plano

Para leer datos de texto

```
misDatos <- read.csv("mi_archivo.csv", header = TRUE, na.strings  
= "NA", dec=".",sep=";",stringsAsFactors = TRUE)  
misDatos <- read.csv2(file.choose(), header = TRUE, na.strings  
="NA", dec=".",sep=";",stringsAsFactors = TRUE)  
misDatos <- read.table(mi_archivo.csv, header = TRUE, na.strings  
="NA", dec=".",sep="\t",stringsAsFactors = TRUE)
```

Valores ausentes

NA (not available)-NaN (not a number)

```
y <- c(1,2,3,NA)
```

```
is.na(y) # Existencia de datos ausentes
```

```
mean(y, na.rm=TRUE) # Saca el promedio sin tener en cuenta el  
dato ausente
```

```
MisDatos[!complete.cases(MisDatos),] # Lista las filas de datos que  
contienen valores ausentes
```

```
DatosSiNA <- na.omit(MisDatos) # Crea un conjunto de datos sin  
valores ausentes
```

```
colSums(is.na(MiDataFrame)) # Suma datos ausentes por columnas
```

Función which

Devuelve la posición o el índice del valor que satisface la condición dada

```
UnVector<- c(5,4,3,2,1) which(UnVector==2)
```

Ordenar datos

Funciones `order()` y `sort()`

```
str(iris)
sort(iris$Sepal.Length)
sort(iris$Sepal.Length, decreasing=TRUE)
order(iris$Sepal.Length)
order(iris$Sepal.Length, decreasing=TRUE)
```

Data cheat sheets

Resumen para el manejo de algunas librerías

Las *sata cheat sheets* facilitan el uso de algunos paquetes. Buscar en <https://www.rstudio.com/resources/cheatsheets/> y descargar la *cheat sheet* de la librería *dplyr*.