SESION 9.1

Using apply(), tapply(), sapply(), lapply()

Load Required packages

Hide

library(tidyverse)

library(datasets)

Load iris dataset

Hide

data(iris)

Hide

head(iris, 3)

|  |
| --- |
|  |

|  | **Sepal.Length**  <dbl> | **Sepal.Width**  <dbl> | **Petal.Length**  <dbl> | **Petal.Width**  <dbl> | **Species**  <fctr> |
| --- | --- | --- | --- | --- | --- |
| 1 | 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| 2 | 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| 3 | 4.7 | 3.2 | 1.3 | 0.2 | setosa |

3 rows

**tapply()**

El objeto llamado "iris" estará en el espacio de trabajo. En este conjunto de datos, ¿cuál es la media de "Sepal.Width" para la especie setosa?

tapply(iris$Sepal.Width, iris$Species, mean)

setosa versicolor virginica

3.428 2.770 2.974

**Método alternativo**

iris %>%

filter(Species == 'setosa') %>%

group\_by(species = Species) %>%

summarise(avg\_sepal\_width = mean(Sepal.Width))

|  |
| --- |
|  |

| **species**  <fctr> | **avg\_sepal\_width**  <dbl> |
| --- | --- |
| setosa | 3.428 |

1 row

**lapply()**

¿Devolver una lista de las medias de las variables 'Sepal.Length', 'Sepal.Width', 'Petal.Length' y 'Petal.Width'?

lapply(iris[, -5], mean)

$Sepal.Length

[1] 5.843333

$Sepal.Width

[1] 3.057333

$Petal.Length

[1] 3.758

$Petal.Width

[1] 1.199333

**sapply()**

¿Devolver un vector de las medias de las variables 'Sepal.Length', 'Sepal.Width', 'Petal.Length' y 'Petal.Width'?

sapply(iris[, -5], mean)

Sepal.Length Sepal.Width Petal.Length Petal.Width

5.843333 3.057333 3.758000 1.199333

Hide

sapply(iris[, 1:4], mean)

Sepal.Length Sepal.Width Petal.Length Petal.Width

5.843333 3.057333 3.758000 1.199333

**apply()**

¿Utiliza apply() para encontrar las medias de las variables 'Sepal.Length', 'Sepal.Width', 'Petal.Length' y 'Petal.Width'?

apply(iris[,1:4], 2, mean) #2 indicates columns

Sepal.Length Sepal.Width Petal.Length Petal.Width

5.843333 3.057333 3.758000 1.199333

**NOTE**

RowMeans, ColMeans, RowSums, ColSums son funciones preferidas ya que han sido optimizadas para las tareas.

colMeans(iris[1:4])

Sepal.Length Sepal.Width Petal.Length Petal.Width

5.843333 3.057333 3.758000 1.199333

**Load mtcars**

Hide

data(mtcars)

Hide

head(mtcars,2)

|  |
| --- |
|  |

|  | **mpg**  <dbl> | **cyl**  <dbl> | **disp**  <dbl> | **hp**  <dbl> | **drat**  <dbl> | **wt**  <dbl> | **qsec**  <dbl> | **vs**  <dbl> | **am**  <dbl> |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mazda RX4 | 21 | 6 | 160 | 110 | 3.9 | 2.620 | 16.46 | 0 | 1 |  |
| Mazda RX4 Wag | 21 | 6 | 160 | 110 | 3.9 | 2.875 | 17.02 | 0 | 1 |  |

2 rows | 1-10 of 11 columns

**sapply()-Split() combination, tapply()**

¿Calcular el peso promedio (wt) por número de cilindros en el automóvil (cyl)?

**split()-sapply() combination**

Hide

split(mtcars$wt, mtcars$cyl)

$`4`

[1] 2.320 3.190 3.150 2.200 1.615 1.835 2.465 1.935 2.140 1.513 2.780

$`6`

[1] 2.620 2.875 3.215 3.460 3.440 3.440 2.770

$`8`

[1] 3.440 3.570 4.070 3.730 3.780 5.250 5.424 5.345 3.520 3.435 3.840 3.845 3.170 3.570

Hide

sapply(split(mtcars$wt, mtcars$cyl), mean)

4 6 8

2.285727 3.117143 3.999214

tapply produce el mismo resultado anterior con la combinación sapply-split

tapply(mtcars$wt, mtcars$cyl, mean)

4 6 8

2.285727 3.117143 3.999214

Hide

with(mtcars, tapply(wt, cyl, mean))

4 6 8

2.285727 3.117143 3.999214

Alternative Approach

Hide

mtcars %>%

group\_by(cyl) %>%

summarise(avg\_wt = mean(wt))

|  |
| --- |
|  |

| **cyl**  <dbl> | **avg\_wt**  <dbl> |
| --- | --- |
| 4 | 2.285727 |
| 6 | 3.117143 |
| 8 | 3.999214 |

3 rows