Resumenes estadisticos

Victor Lopez

2023-01-23

Funcion summary()

```
crabs = read.table("../../data/datacrab.txt", header = TRUE)
crabs = crabs[,-1]
summary(crabs)
```

```
##      color           spine           width          satell           weight
##  Min.   :2.000   Min.   :1.000   Min.   :21.0   Min.   : 0.000   Min.   :1200
##  1st Qu.:3.000   1st Qu.:2.000   1st Qu.:24.9   1st Qu.: 0.000   1st Qu.:2000
##  Median :3.000   Median :3.000   Median :26.1   Median : 2.000   Median :2350
##  Mean   :3.439   Mean   :2.486   Mean   :26.3   Mean   : 2.919   Mean   :2437
##  3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.:27.7   3rd Qu.: 5.000   3rd Qu.:2850
##  Max.   :5.000   Max.   :3.000   Max.   :33.5   Max.   :15.000   Max.   :5200
```

```
summary(subset(crabs, color == 3, c("weight", "width")))
```

```
##      weight         width
##  Min.   :1300   Min.   :22.5
##  1st Qu.:2100   1st Qu.:25.1
##  Median :2500   Median :26.5
##  Mean   :2538   Mean   :26.7
##  3rd Qu.:3000   3rd Qu.:28.2
##  Max.   :5200   Max.   :33.5
```

Funcion by()

```
# Aplica una funcion a algunas columnas de un dataframe segmentandolas segun los niveles de
# un factor

by(iris[, c(1,3)], iris$Species, FUN = summary)
```

```
## iris$Species: setosa
##   Sepal.Length    Petal.Length
##  Min.   :4.300   Min.   :1.000
##  1st Qu.:4.800   1st Qu.:1.400
##  Median :5.000   Median :1.500  
##  Mean   :5.006   Mean   :1.462  
##  3rd Qu.:5.200   3rd Qu.:1.575  
##  Max.   :5.800   Max.   :1.900  
## ------------------------------------------------------------ 
## iris$Species: versicolor
##   Sepal.Length    Petal.Length 
##  Min.   :4.900   Min.   :3.00  
##  1st Qu.:5.600   1st Qu.:4.00  
##  Median :5.900   Median :4.35  
##  Mean   :5.936   Mean   :4.26  
##  3rd Qu.:6.300   3rd Qu.:4.60  
##  Max.   :7.000   Max.   :5.10  
## ------------------------------------------------------------ 
## iris$Species: virginica
##   Sepal.Length    Petal.Length  
##  Min.   :4.900   Min.   :4.500  
##  1st Qu.:6.225   1st Qu.:5.100  
##  Median :6.500   Median :5.550  
##  Mean   :6.588   Mean   :5.552  
##  3rd Qu.:6.900   3rd Qu.:5.875  
##  Max.   :7.900   Max.   :6.900  
```

```
# VALORES NA
# datos_completos <- datos[complete.cases(datos),]
# by(datos_completos, datos_completos$grupo, function(x) mean(x$valor))

# datos_sin_NA <- na.omit(datos)
# by(datos_sin_NA, datos_sin_NA$grupo, function(x) mean(x$valor))
```

Funcion aggregate()

```
# Es equivalente a by() pero se muestra de otra forma

aggregate(cbind(Sepal.Length, Petal.Length)~Species, data = iris, FUN = summary, na.rm = TRUE)
```

```
##      Species Sepal.Length.Min. Sepal.Length.1st Qu. Sepal.Length.Median
## 1     setosa             4.300                4.800               5.000
## 2 versicolor             4.900                5.600               5.900
## 3  virginica             4.900                6.225               6.500
##   Sepal.Length.Mean Sepal.Length.3rd Qu. Sepal.Length.Max. Petal.Length.Min.
## 1             5.006                5.200             5.800             1.000
## 2             5.936                6.300             7.000             3.000
## 3             6.588                6.900             7.900             4.500
##   Petal.Length.1st Qu. Petal.Length.Median Petal.Length.Mean
## 1                1.400               1.500             1.462
## 2                4.000               4.350             4.260
## 3                5.100               5.550             5.552
##   Petal.Length.3rd Qu. Petal.Length.Max.
## 1                1.575             1.900
## 2                4.600             5.100
## 3                 5.875            6.900
```

```
# La ventaja de aggregate es que podemos evitar valores NA usando el parametro na.rm = TRUE
```