

# Informe Iris dataset

Isabel Afán de Ribera

31/10/2020

## 1. Presentación del dataset

Iris es una base de datos que contiene información sobre las medidas en centímetros de una serie de características del tipo de flor Iris, en concreto sobre la longitud y la anchura del sépalo y pétalo de tres especies de esta flor que son setosa, versicolor y virginica. Este dataset se compone de una muestra de 150 observaciones (50 de cada especie) y 5 variables.

## 2. Objeto

El objetivo de este análisis es conocer las diferencias morfológicas entre las distintas especies.

## 3. Presentación de variables

- `Species` : variable categórica que contiene información sobre el tipo de flor iris
- `Sepal.Length` : variable numérica que contiene información sobre la longitud del sépalo.
- `Sepal.Width` : variable numérica que contiene información sobre la anchura del sépalo.
- `Petal.Length` : variable numérica que contiene información sobre la longitud del pétalo.
- `Petal.Width` : variable numérica que contiene información sobre la anchura pétalo.

## 4. Descripción del dataset

```
data("iris")
summary(iris)
```

```
##      Sepal.Length      Sepal.Width      Petal.Length      Petal.Width
## Min.      :4.300    Min.      :2.000    Min.      :1.000    Min.      :0.100
## 1st Qu.:5.100    1st Qu.:2.800    1st Qu.:1.600    1st Qu.:0.300
## Median :5.800    Median :3.000    Median :4.350    Median :1.300
## Mean   :5.843    Mean   :3.057    Mean   :3.758    Mean   :1.199
## 3rd Qu.:6.400    3rd Qu.:3.300    3rd Qu.:5.100    3rd Qu.:1.800
## Max.   :7.900    Max.   :4.400    Max.   :6.900    Max.   :2.500
##      Species
## setosa      :50
## versicolor:50
## virginica   :50
##
##
##
```

En una primera aproximación de las variables de este dataset. Puede observarse mediante estos estadísticos descriptivos de posición que existen claras diferencias en el tamaño de los pétalos y sépalos de las distintas especies. Especialmente, en la longitud y anchura del pétalo pues la media de la longitud se sitúa en 3.758 cm mientras que el mínimo es de 1 y el máximo de 6.9. Y en el caso de la anchura la media es de 1.19 mientras que el mínimo es de 0.1 y el máximo de 2.5.

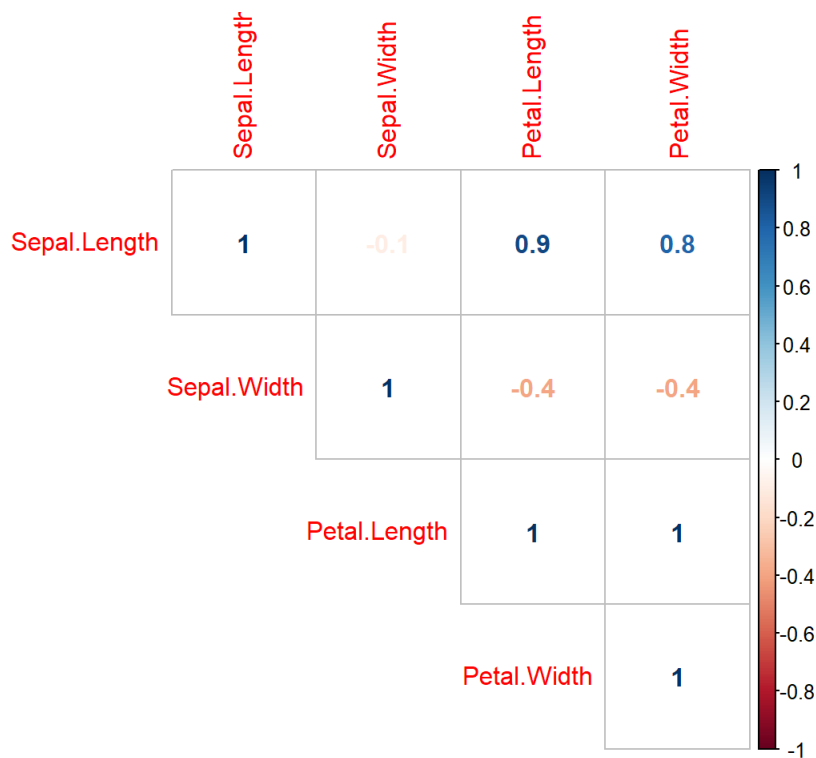
Con el fin de entender mejor las diferencias entre los atributos podemos calcular la matriz de correlaciones y graficarla.

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.0.3
```

```
## corrplot 0.84 loaded
```

```
correlacion <- round(cor(iris[, -5]), 1)
corrplot(correlacion, method="number", type="upper")
```

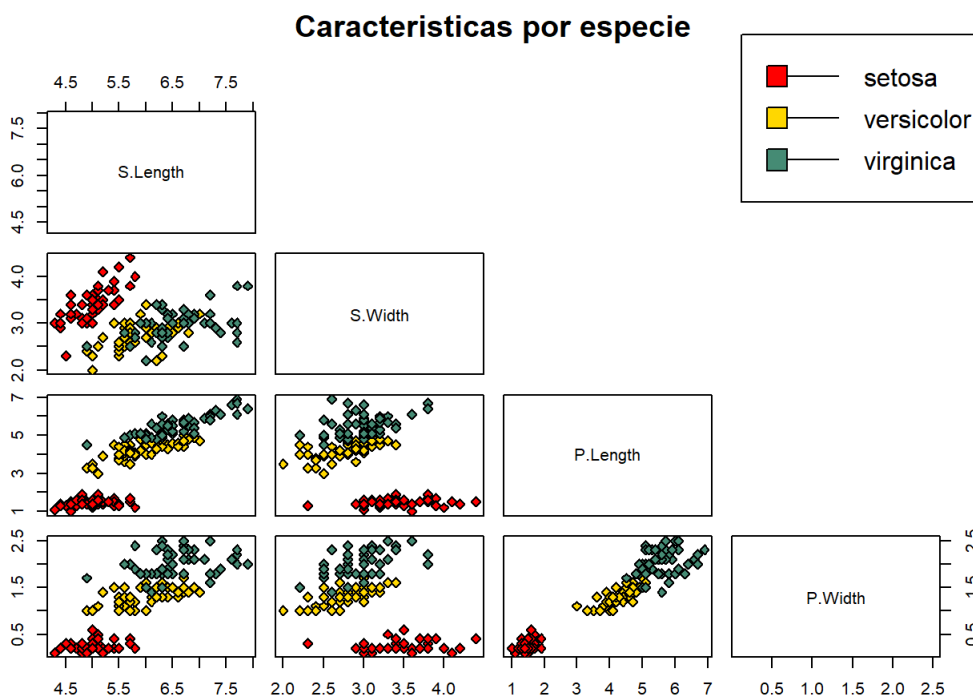


Como se observa en el gráfico de correlaciones las variable Petal Length y Petal Width tienen una correlación perfecta. Por otra parte, existe una alta correlación positiva entre Sepal Length y Petal Length y entre Sepal Length y Petal Width. Sin embargo, entre los atributos del petalo y Sepal Width la correlación es negativa al igual que entre Sepal Length y Width.

Para una mejor percepción de las diferencias entre las características de las distintas especies pasamos a representarlas graficamente a través de un gráfico de dispersión.

```
pairs(iris[1:4], main = "Características por especie", pch = 23, bg = c("red", "gold", "aquamarine4")[unclass(iris$Species)], upper.panel=NULL, labels=c("S.Length", "S.Width", "P.Length", "P.Width"), font.labels=1, cex.labels=1)
par(xpd = TRUE)

legend("topright", fill = c("red", "gold", "aquamarine4"), lty = c(1, 1, 1), legend = c( levels(iris$Species)
))
```



Como puede observarse en la gráfica las especies se dividen en dos grupos, el de las especies versicolor y virginica con características en común y el de setosas sin características en común con las otras.

## 5. Conclusiones

Tras este análisis exploratorio del dataset Iris se puede concluir sobre las relaciones entre sus atributos que existe una alta correlación entre la largura de los sépalos y la largura y anchura de los pétalos y una clara dependencia entre las características de los pétalos pues al aumentar el tamaño del largo lo hará el del ancho en la misma proporción al contrario que entre las características de los sépalos pues tienen una relación negativa, cuando uno aumenta el otro disminuye. En cuanto a la apariencia de las tres especies se concluye que existe una notable diferencia de la especie setosa con respecto a versicolor y virginica.