

# IRIS DATASET

## 1. Análisis de los datos.

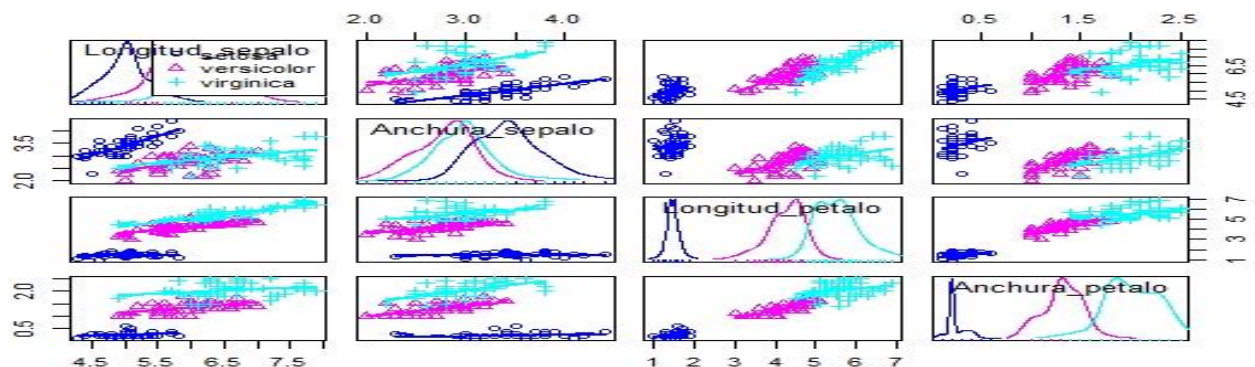
Iris es un conjunto de datos multivariante integrado por defecto en R en el cual están representados 150 datos distribuidos en cinco variables, dichos datos pertenecen a tres especies de plantas llamadas Setosa, Versicolor y Virginica que a su vez pertenecen al género Iris.

Este estudio fue desarrollado por Ronald Fisher en el año 1936 y a la finalidad de este fue crear un modelo de análisis discriminante.

Las herramientas que he utilizado para analizar dichos datos han sido las librerías de R dplyr, car y ggplot2 y el análisis de datos mediante la función summary y el calculo de la desviación estándar.

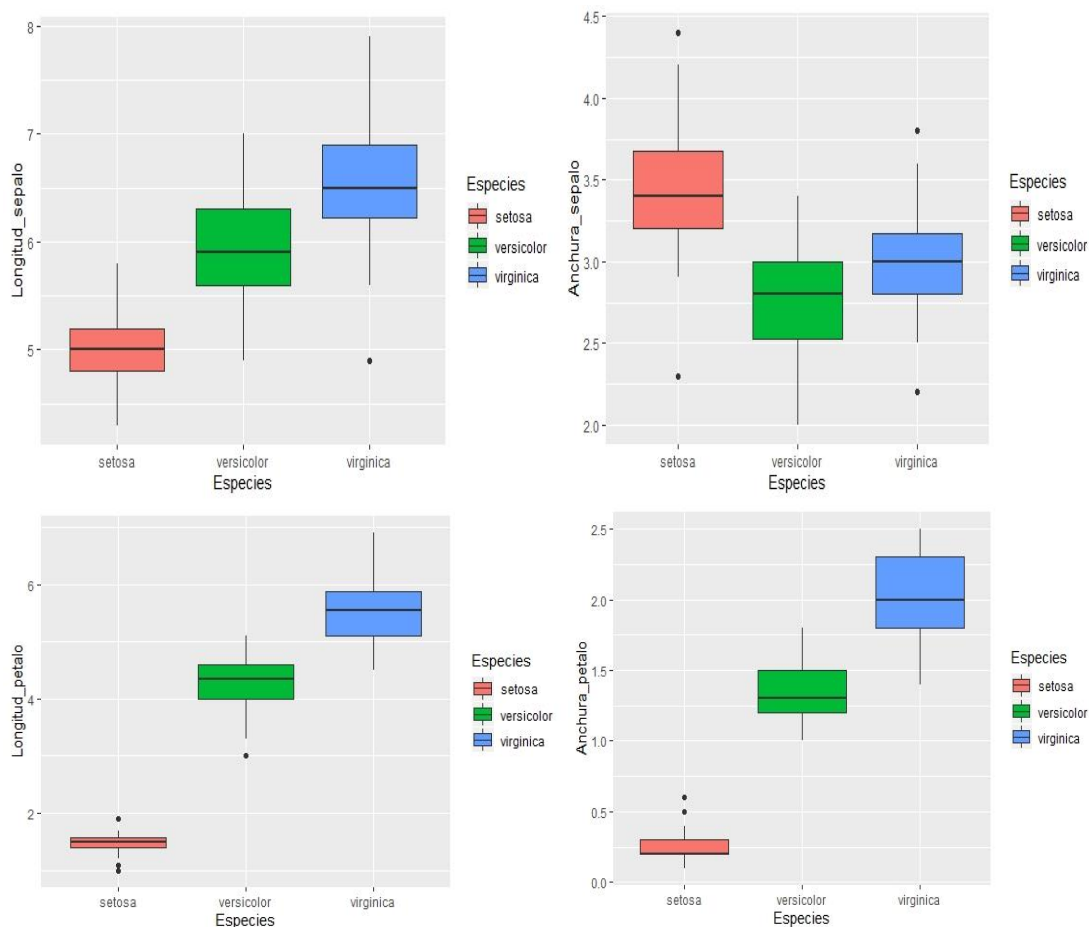
	Media longitud	Media anchura	Mediana longitud	Mediana anchura	Desviación longitud	Desviación anchura
Setosa sépalo	5.006	3.428	5	3.40	0.3525	0.3790
Virginica sepalo	6.588	2.974	6.5	3	0.6359	0.3225
Versicolor sépalo	5.936	2.77	5.90	2.8	0.5161	0.3138
Setosa pétalo	1.462	0.246	4.35	0.2	0.1737	0.1054
Virginica pétalo	5.552	2.026	5.55	2.026	0.5519	0.2746
Versicolor pétalo	4.26	1.326	4.35	1.3	0.4699	0.1977

Con estos descriptivos podemos ver que entre la media y la mediana existe mucha diferencia y si nos fijamos en la varianza la que más llama la atención es la desviación típica de la longitud del sépalo de la Virginica y nos indica que dicha magnitud se desvía en 0.6359 centímetros respecto a la media. Si las vemos una a una podemos ver que existe mas relación entre las clases Versicolor y Virginica, entre ellas, que cualquiera de ambas con la clase Setosa. Esto lo podemos comprobar con la matriz de correlación.



Respecto a lo anterior la Setosa es mas fácil de diferenciar si nos fijamos únicamente en los pétalos, pero no es tan fácil de clasificar si solo nos fijamos en los sépalos ya que todas las especies siguen una distribución normal.

Para realizar el análisis numérico he creado cuatro gráficos del tipo boxplot para analizar mas a fondo los datos de la tabla anterior.



En dichas gráficas podemos ver que existen varios datos atípicos para la clase Setosa sobre todo en la longitud del pétalo además tanto la longitud como la anchura del pétalo de dicha especie son significativamente menores que las de las medidas de las otras dos especies. En cuanto a los sépalos pasa lo mismo en cuanto a las unidades con la diferencia de que en este caso el ancho de la Setosa es mayor en términos de media y mediana que el de las otras dos especies.

En conclusión, podemos decir que la forma más sencilla para distinguir las especies Virginica y Versicolor, ya que son las que mas se parecen entre ellas, sería a través de la longitud de sus pétalos. La Setosa aun teniendo un sépalo más ancho es más fácil de diferenciar en cuanto al reducido tamaño de sus pétalos.