

Técnicas de Agrupación y Reducción de la Dimensión: Análisis del dataset IRIS

Elaborado por: Mayra Goicochea Neyra
15/10/2019

Introducción

El dataset Iris incluye las mediciones recopiladas por Edgar Anderson para cuantificar la variación morfológica de las flores IRIS. La variación morfológica es uno de los temas que se están desarrollando con más frecuencia debido al cambio climatológico. Se trata de medir la variación que puede encontrarse entre individuos que conforman una población. En este caso se seleccionó tres especies relacionadas, dos fueron recolectadas del mismo lugar (Península Gaspé). La información que guarda es:

- Longitud y Ancho de los sépalos (expresados en centímetros)
- Longitud y Ancho de los pétalos (expresados en centímetros)

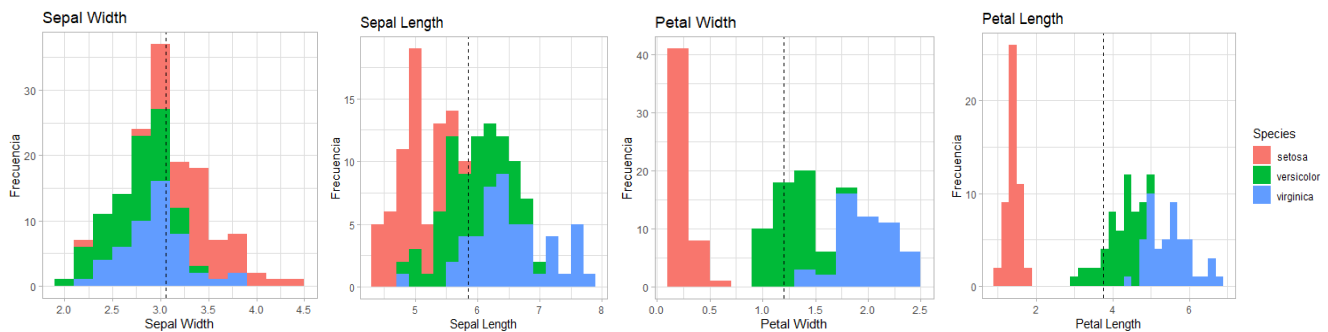
Análisis Exploratorio de Datos (AED)

Se tiene 150 casos con 5 variables: 4 variables continuas: "Sepal.Length, Sepal.Width, Petal.Length y Petal.Width"; y 1 variable categórica "Species"

Existen tres especies de flor iris: Setosa, Virginica y Versicolor

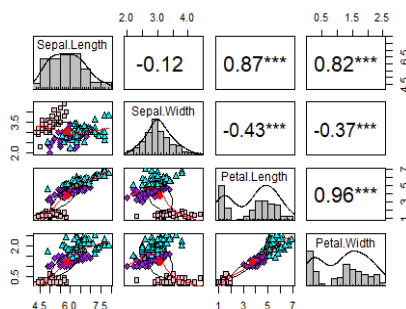
La muestra está repartida en forma homogénea de las tres especies (cada una con 50 casos). La longitud media de los sépalos es 5.843 cm, el ancho promedio de los sépalos es 3.057 cm. Mientras que los pétalos tienen 3.758 cm de longitud media y 1.199 cm de ancho medioⁱ

En cuanto al comportamiento de los datos con respecto a cada variable se observa:



Muchos casos de las variables longitud y ancho del pétalo están ubicados a la izquierda de la media. Se debe tener cuidado en estas variables si se usan las medias. En cuanto a la variable ancho del sépalo están ubicados de forma central.

Correlación



Se observa que las variables Longitud y Ancho del pétalo están altamente correlacionada, de forma positiva (con un 0.96). Se podrían agrupar para reducir variables por que actúan de forma similar.

En cuanto a la variable Longitud de sépalo tiene correlación positiva importante con las variables de pétalos.

Bibliografía

- Report on Edgar Anderson's Iris Data Analysis, [link](#)
- Variación morfológica - UNAM, [link](#)

ⁱ El archive R con las instrucciones utilizadas para el análisis exploratorio de datos se guardará en el repositorio GitHub: <https://github.com/Mayragn387/MasterDS2019/tree/master/T%C3%A9cnicas%20de%20Agrupaci%C3%B3n%20y%20de%20Reducci%C3%B3n%20de%20la%20Dimensi%C3%B3n/Source>