

Laboratorio # 11 – R (Análisis Exploratorio de Datos)

17/18 de junio de 2019

En parejas, suban un único archivo de extensión `.r` con las respuestas a cada ejercicio. **Recuerden escribir los nombres de ambos/as integrantes como comentario en la primera línea.**

En este laboratorio trabajaremos con el dataset `iris`, disponible en la instalación base de R. Este conjunto de datos contiene 150 muestras de tres especies de flores del iris, con los valores de cuatro variables: largo y ancho del sépalo y pétalos. Puede consultar la documentación del dataset con el comando `help(iris)`.

1. Usando todos los datos disponibles, calcule la distribución de frecuencias y frecuencias relativas según **especie** de flor.
2. Calcule el promedio del **largo del pétalo** por **especie** de flor.
3. Represente visualmente la distribución de frecuencias del **largo del sépalo**. ¿Qué puede interpretar al respecto?
4. Genere un gráfico de dispersión para estudiar la relación entre:
 - a) largo del sépalo vs. ancho del sépalo
 - b) largo del pétalo vs. ancho del pétalo¿Qué puede concluir en cada caso?
5. Seleccione los registros correspondientes a la especie “virginica”, guardando su resultado en una variable llamada `virginica`. Sobre estos datos, calcule el promedio, mediana, varianza y desviación estándar para todas las columnas numéricas (i.e., largo y ancho de pétalos y sépalos). ¿Qué puede interpretar al respecto?