

# Práctica Estadística Descriptiva y Análisis Exploratorio de Datos

## Introducción al análisis de datos con R A2B2C

Octubre 2020

### Análisis exploratorio de datos

Esta práctica tiene como objetivo realizar una primera aproximación a la comprensión de nuestros datos, utilizando para ello las técnicas del análisis exploratorio de datos vistas.

Para esta práctica trabajaremos con un dataset de COVID-19 de la OMS disponible en <https://data.humdata.org/dataset/>

### 1. Carga de datos y análisis general del dataset

- a) Cargar el dataset cases-covid-19-by-country.csv de una forma apropiada.
- b) ¿Cuántos datos u observaciones contiene el dataset? ¿Cuántos atributos tiene cada dato?
- c) ¿De qué tipo son los atributos? ¿Cómo se llaman?
- d) Mostrar los primeros 5 datos y los últimos 5 datos.
- e) Realizar una descripción general del dataset a partir de este análisis con sus palabras.

### 2. Medidas de resumen y visualización

- a) Calcular la media, la mediana y la moda para las variables numéricas.  
¿Son similares o muy diferentes?
- b) Calcular el rango, el desvío estandar y el IQR.  
A partir de esto indicar si los valores están muy concentrados o si por el contrario, están muy dispersos.
- c) En base a lo visto anteriormente, ¿Qué medidas usarías para resumir los valores?
- d) Realizar un histograma para cada variable y decidir si es más conveniente separar los datos en dos grupos para su análisis. En ese caso, ¿qué criterio podríamos usar?
- e) Graficar en un mismo gráfico un histograma y un boxplot para cada variable. ¿Siguen apareciendo valores extremos en los boxplot? ¿Por qué?
- f) Realizar un gráfico de scatterplot entre las variables numéricas. ¿Es necesario separar los casos más extremos para su análisis? ¿Existe relación entre la cantidad de infectados y la cantidad de fallecidos? ¿Si existe, es en todas las escalas? ¿De qué tipo?