



## TRABAJO PRÁCTICO: Nociones de Estadística descriptiva e introducción al lenguaje R

### Introducción:

Esta práctica inicial tiene como objetivo explorar y entender la información que puede ser obtenida desde un conjunto de datos, así como también repasar conceptos fundamentales de estadística descriptiva a través de la utilización de un software de análisis estadístico como R.

### Consignas:

A partir del dataset *Global Data on Sustainable Energy (2000-2020)*<sup>1</sup>, se solicita trabajar sobre las siguientes consignas:

1. **Exploración de datos.** Describa las características de los datos que contiene el dataset: tipo de los atributos, cantidad de instancias, si el mismo está completo o presenta NA, identifique cómo se agrupan los datos en tiempo y espacio, etc. Utilice herramientas gráficas y de resumen para hacer una primera aproximación.
2. **Medidas de posición.**
  - a. Calcular la media, la moda y la mediana para cada uno de los atributos y analice los resultados obtenidos.
  - b. Calcule la media por país y por año. ¿Cómo son esos resultados con respecto a la media general de esos atributos?
  - c. Documente los resultados y las conclusiones.
3. Grafique las variables cuantitativas utilizando gráficos de dispersión.
  - a. ¿Qué puede decir a partir de esos gráficos sobre la relación entre las variables?

---

<sup>1</sup> <https://www.kaggle.com/datasets/anshtanwar/global-data-on-sustainable-energy>



Bases de Datos Masivas (11088)  
Departamento de Ciencias Básicas

- b. Seleccione una relación de interés, por ejemplo: generación de electricidad versus emisiones de CO<sub>2</sub> e indique cómo es esa relación. ¿Tiene modificaciones a lo largo del tiempo? ¿Cómo se relaciona con los países?
  - c. Documente los resultados y las conclusiones.
4. **Medidas de dispersión.**
- a. Calcular el desvío estándar y la varianza para cada una de las variables.
  - b. Grafique el diagrama de cajas del producto bruto interno agrupado por año para los países del [Cono Sur](#). ¿Qué puede comentar? ¿Qué ocurre si agrega a China y Estados Unidos?
  - c. Documente las gráficas y conclusiones.
5. **Medidas de asociación.**
- a. Calcular el coeficiente de correlación de todas las variables y explique el resultado.
  - b. ¿Qué relación encuentra con los resultados del punto 3?
6. Guarde los archivos resultantes de las actividades prácticas en una carpeta denominada tp0-<legajo> que a su vez tenga un directorio por cada uno de los puntos de este trabajo, comprima la carpeta y envíelo al equipo docente.

Referencias sugeridas:

Data Mining: Concepts and techniques. Jiawei Han and Micheline Kamber.  
Chapter 2.

Introducción al Lenguaje R:

[http://www.dm.uba.ar/materias/analisis\\_expl\\_y\\_conf\\_de\\_datos\\_de\\_exp\\_de\\_marrays\\_Mae/2006/1/practicas/TP1-2006.pdf](http://www.dm.uba.ar/materias/analisis_expl_y_conf_de_datos_de_exp_de_marrays_Mae/2006/1/practicas/TP1-2006.pdf)

R: [https://cran.r-project.org/doc/contrib/rdebuts\\_es.pdf](https://cran.r-project.org/doc/contrib/rdebuts_es.pdf)