



Bases de Datos Masivas (11088)  
Departamento de Ciencias Básicas

## TRABAJO PRÁCTICO 00: Nociones de Estadística descriptiva

### Introducción:

Esta práctica inicial tiene como objetivo explorar y entender la información que puede proveer un dataset 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 *MPI\_subnational.csv*<sup>1</sup> (Multidimensional Poverty Measures), se solicita trabajar sobre las siguientes consignas:

1. **Exploración de datos.** Explore y explique en que consiste el dataset utilizando herramientas de exploración de datos.
  - a. Releve las características de los atributos.
  - b. Represente gráficamente la cantidad de ciudades agrupados por Región.
2. **Medidas de posición.** Calcule las medidas de posición para los atributos numéricos y agrupe los cálculos de acuerdo a la Región.
  - a. Ordene los resultados del MPI resultante y concluya al respecto. Help(order).
  - b. Grafique las variables y observe su comportamiento (graph: barplot, pie & hist).
3. **Medidas de dispersión.** Calcular el desvío estándar, la varianza y el rango para cada una de las variables.
  - a. Realice diagramas de cajas y scatterplot's. Documente las conclusiones.
  - b. ¿Qué variable es la que presenta mayor dispersión? Tenga en cuenta que cada variable puede estar expresada en diferentes unidades y magnitudes.

---

<sup>1</sup> Disponible en: <https://www.kaggle.com/ophi/mpi/data>



**Bases de Datos Masivas (11088)**  
**Departamento de Ciencias Básicas**

4. **Medidas de asociación.** Calcular el coeficiente de correlación de todas las variables y explique el resultado. ¿Qué tipo de gráficos describen mejor esta relación entre las variables?
5. Resuelva las consignas e integre en un único archivo las operaciones realizadas -se sugiere un archivo del tipo “R Markdown”- mostrando de manera sencilla el código, los resultados, las gráficas resultantes y **sus conclusiones**. Utilice como nombre del archivo **tp00-<legajo>** y envíelo al equipo docente mediante correo electrónico.

Referencias sugeridas:

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

An Introduction to R:

<https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>