



## UNIVERSIDAD AUTÓNOMA DE OCCIDENTE

**Asignatura:** Estadística y Probabilidad 2

**Periodo académico:** 2023

**Profesor:** Johann A. Ospina

**Correo:** jaospina@uao.edu.co

### Condiciones:

- Realizar la tarea en grupo.
- Subir el informe con el análisis de resultados e interpretaciones de gráficas y tablas en formato pdf y script en R.
- Interpretar todos los resultados encontrados.
- Referenciar tablas y gráficas de forma adecuada.
- Investigar de otras fuentes bibliográficas y referencias.
- Realizar presentación del informe en ingles.
- Se bonificará a los grupos que realicen la presentación y el informe en LaTeX. Adjuntar código .tex.
- Todo lo que investiguen nuevo por favor deben ponerlo en la metodología

## Contexto

El conjunto de datos `Datos_Paises.xlsx` contiene indicadores socioeconómicos de 96 países. Las variables observadas son:

- $X_1$  = Tasa anual de crecimiento de la población.
- $X_2$  = Tasa de mortalidad infantil por cada 1000 nacidos vivos.
- $X_3$  = Porcentaje de mujeres en la población activa.
- $X_4$  = PNB en 1995 (en millones de dólares).
- $X_5$  = Producción de electricidad (en millones kW/h).
- $X_6$  = Líneas telefónicas por cada 1000 habitantes.
- $X_7$  = Consumo de agua per cápita.
- $X_8$  = Proporción de la superficie del país cubierta por bosques.
- $X_9$  = Proporción de forestación anual.
- $X_{10}$  = Consumo de energía per cápita.
- $X_{11}$  = Emisión de CO2 per cápita.

El conjunto de datos `BASE.HURTOS.csv` corresponde a datos de hurtos en la ciudad de Cali. El objetivo es utilizar el ACM para describir de forma multivariada el hurto en la ciudad. Para esto los expertos en crimen sugieren utilizar las variables: dinámica, hora del hecho y agruparla en franjas (mañana, tarde y noche), día de la semana, mes del hecho, agrupar la ciudad en zonas utilizando la variable comuna (investigar) y el estrato moda.

## Mini proyecto 2

### Aplicación del ACP y ACM para la Ciencia de Datos

[1] Aplicación del algoritmo de Análisis de Componentes Principales (ACP):

- a) Examine la correlación de las variables usando un gráfico de la matriz de correlación. Interprete sus resultados.
- b) Realice un análisis de componentes principales a partir de la matriz de correlaciones. Interprete sus resultados.
- c) Realice el gráfico de sedimentación y un gráfico que muestre la varianza acumulada por cada componente. Interprete sus resultados.
- d) Interprete las dos primeras componentes. Intente darle nombre a los componentes.
- e) Encuentre la función con la que se describen los 2 primeros componentes y las variables que más peso tienen. Interprete sus resultados.
- f) Construya un índice e interprete de acuerdo al contexto del problema.

[2] Aplicación del algoritmo de Análisis Correspondencias Múltiples (ACM)

- a) Recodifique las variables para análisis.
- b) Realice una análisis exploratorio de datos.
- c) Contruya el ACM con las variables mencionadas al inicio del documento.
- d) Interprete el primer plano factorial.
- e) Realice un análisis clúster con al ACM y caracterice cada clúster.