

# Tarea 2: Modelos de corte transversal

DPDC 102 : Modelos Estadísticos

07.05.2022

---

Considere la base de datos 'crime\_usa.xls', disponible en el apartado **Tarea 2** del aula virtual del curso. Esta base de datos toma en consideración las variables ( $X_1, X_2, X_3, X_4, X_5, X_6, X_7$ ) que describen el crimen en distintas ciudades de Estados Unidos, donde:

- $X_1$  = Tasa de crímenes reportados totales por cada 1 millón de habitantes.
- $X_2$  = Tasa de crímenes violentos reportados por cada 100 mil habitantes.
- $X_3$  = Financiamiento anual para la policía en \$/habitante.
- $X_4$  = % de personas de más de 25 años con 4 años de enseñanza media.
- $X_5$  = % de personas entre 16 a 19 años que no estén estudiando en enseñanza media ni tampoco sean hayan terminado dicha enseñanza.
- $X_6$  = % de personas entre 18 a 24 años que estudian en la universidad.
- $X_7$  = % de personas mayores de 25 años con al menos 4 años de estudios universitarios. *Referencia: Life In America's Small Cities, By G.S. Thomas*

---

## Pregunta 1: (10 puntos)

Describa los datos, entregando gráficos y tablas que muestren las principales características de la base de datos.

*Ayuda:* puede emplear el método `.hist(bins=n_bins, figsize = (ancho, alto))` para graficar un resumen de histogramas de los datos.

## Pregunta 2: (30 puntos)

Efectúe un análisis de regresión lineal múltiple que explique la tasa de crímenes reportados totales por cada 1 millón de habitantes. Para esto, implemente la metodología de **stepwise** hacia atrás o hacia adelante, para encontrar la mejor combinación de variables explicatorias. Puede emplear cualquier métrica de desempeño de modelo:  $RMSE$ ,  $R^2$ ,  $AIC$ , o  $BIC$ .

*Ayuda:* Todas las métricas, parámetros, y otros resultados relevantes pueden ser revisados a partir del objeto `'model'`, en donde quedan almacenados al crear algún modelo de regresión.

## Pregunta 3: (10 puntos)

Muestre el resumen (`summary()`) del modelo escogido y describa sus principales características: -  $R^2$  - Estadístico  $F$  - Estadístico  $t$

## Pregunta 4: (20 puntos)

Comente sus resultados en función del problema. ¿Cuáles son las variables que mejor explican el problema de la tasa de crimen en las ciudades de Estados Unidos?

---

## Condiciones:

- Fecha de entrega viernes 13 de mayo del 2022.
- Los grupos de trabajo son los mismos grupos ya establecidos en el módulo anterior.
- Formato de entrega: Un documento en archivo `.pdf` que contenga gráficos y comentarios, que describan los análisis efectuados y los resultados obtenidos, y un archivo `.py` con los códigos implementados de Python para replicar los análisis.