

## Desafío - Regresión desde la econometría

- Para realizar este desafío debes haber estudiado previamente todo el material disponible correspondiente a la unidad.
- Una vez terminado el desafío, comprime la carpeta que contiene el desarrollo de los requerimientos solicitados y sube el `.zip` en el LMS.
- Desarrollo desafío: Individual.
- Para la realización del desafío necesitarás apoyarte del archivo Apoyo Desafío - Regresión desde la econometría.

### Habilidades a evaluar

- Reconocer la terminología asociada a la modelación estadística.
- Identificar la regresión lineal y sus fundamentos.
- Interpretar los parámetros estimados en la regresión.
- Interpretar inicialmente estadísticas de bondad de ajuste y coeficientes.
- Reconocer los supuestos en los que la regresión tiene sustento teórico.
- Implementar un modelo de regresión con statsmodels

### Descripción

En la empresa en la cual te encuentras contratado, te solicitan trabajar con un registro de clientes de un banco e información sobre su salud financiera. La base contenida en el csv `Credit` abarca las siguientes variables:

- `Income`: Ingreso anual disponible en miles de dólares.
- `Limit`: Límite del crédito.
- `Rating`: Ranking del crédito.
- `Cards`: Cantidad de tarjetas de crédito.
- `Age`: Edad del usuario.
- `Education`: Años de educación del usuario.
- `Gender`: ¿Hombre o mujer?
- `Student`: ¿Es estudiante?
- `Married`: ¿Está casado?
- `Ethnicity`: Etnicidad del usuario (Caucásico, Afroamericano, Asiático)
- `Balance`: Promedio de deuda en tarjeta de crédito de usuario.

El objetivo que persigue el trabajo con esta base de datos es conocer los determinantes de la variación en el promedio de deuda en tarjeta de crédito. Para ello se te solicita generar

modelos econométricos para inspeccionar los efectos de distintos atributos a partir de requerimientos que serán detallados más adelante.

## Requerimientos

### 1. Preparar el ambiente de trabajo (1 puntos). Para ello debes:

- Importe las librerías básicas para el análisis de datos.
- Incluya los import base de `statsmodels`.
- Importe la base de datos `Credit.csv` y elimine la columna `Unnamed: 0`.
- Siguiendo las buenas prácticas, genere variables binarias para capturar el efecto del grupo minoritario en cada variable. Para el caso de `Ethnicity`, genere binarias para `African American`, `Caucasian` y `Asian`.



**Tip:** Utilice `value_counts()` para revisar la cantidad de casos en las variables categóricas.

### 2. Generar gráficos descriptivos para las variables continuas (1 puntos):

- Genere una función llamada `plot_hist` que genere un histograma de la variable e indique la media y mediana de la misma. Agregue una leyenda para indicar qué representa cada línea.
- ¿Cuál es el comportamiento de `Balance`, `Income`, `Cards` y `Rating`?

### 3. Regresión Lineal Simple (2 puntos). Para ello se debe:

- Generar una serie de modelos simples:

$$Balance = \beta_0 + \gamma_1 \cdot Student + \varepsilon$$

$$Balance = \beta_0 + \beta \cdot Income + \varepsilon$$

$$Balance = \beta_0 + \beta \cdot Rating + \varepsilon$$

- Utilizando `statsmodels`, comente brevemente los siguientes puntos:
  - ¿Qué variable presenta un mayor poder explicativo de la variabilidad de `Balance`?
  - ¿Qué podemos decir sobre el comportamiento de `Student`?

- ¿Qué pasa si aplicamos una prueba de hipótesis para muestras independientes de `Balance` entre aquellos que son y no son estudiantes?

4. **Generar un modelo que incluya todos los regresores (2 puntos):** A partir de la siguiente fórmula, genere el modelo que incluya dichos regresores y comente brevemente sobre el poder explicativo en la varianza de  $y$  y las principales características de los predictores.

$$Balance_i = \beta_0 + \sum_{j=1}^N \beta_j \cdot X_{ij} + \varepsilon_i$$

5. **Depurar el modelo e incluya sólo los predictores significativo (2 puntos)**

- En base al modelo anterior, refactoriza e incluya sólo las variables con un nivel de significancia  $p - value \leq .025$ . Comente brevemente sobre el poder explicativo en la varianza de  $y$  y las principales características de los predictores.
- ¿Mejoró o empeoró el modelo con la reducción de regresores?

6. **Utilizando `lmpplot` de `seaborn`, inspeccione el comportamiento de los tres principales determinantes. (2 puntos)**