

## ECONOMETRIA 2024

### PRACTICA 6



**Sugerencia: lea las notas de la CLASE 7 y ejecute el “Script CLASE 7.R” Luego responda las siguientes preguntas.**

Suponga que una empresa financiera lo contrata como analista de datos para desarrollar un modelo que permita predecir el límite de crédito que se le otorgará a un cliente en función de sus ingresos. Se le proporciona la base de datos `LimiteCredito.xlsx`, la cual contiene las siguientes variables para una muestra de 400 clientes de la empresa financiera:

- ID: Identificador del cliente
- Ingreso: Ingreso anual del cliente, expresado en dólares americanos
- Limite: límite de crédito del cliente, expresado en dólares americanos

Se pide:

1. Divida la base de datos en 75% training y 25% testing. Utilice la CI de alguno de los integrantes de su equipo como semilla aleatoria: `set.seed(xxxxxxxx)`.
2. Ajuste el modelo de regresión lineal que relaciona el límite de crédito (variable de respuesta Y) en función de los ingresos anuales del cliente (variable predictora X).
  - 2.1. Reporte la ecuación de regresión lineal obtenida.
  - 2.2. Interprete el valor de la pendiente del modelo en el contexto del problema.
3. Son significativos los coeficientes del modelo? Plantee las hipótesis nulas y alternativas y el resultado de la prueba de hipótesis.
4. Grafique el modelo de regresión lineal junto con el diagrama de dispersión. Reporte la gráfica.
5. Reporte el valor de  $R^2$  del modelo en la base de training. Interprete su valor.
6. Reporte el valor del error promedio que comete el modelo en las predicciones del límite de crédito en la base de training (RCME en train).
7. Reporte el valor del error promedio que comete el modelo en las predicciones del límite de crédito en la base de testing (RCME en test).
8. Comente acerca de la presencia o no de sobreajuste (overfitting) en el modelo.
9. Utilice el modelo para predecir el límite de crédito para un cliente que tiene ingresos anuales de 25000 dólares.