

UNIVERSIDAD AUSTRAL

Maestría: Maestría en Ciencia de Datos

Materia: Regresiones Avanzadas

Trabajo Práctico Grupal

Se adjunta un archivo llamado CarPrices.csv con las siguientes columnas:

- # Car_ID: Identificación única de cada observación
- # Symboling: Su clasificación de riesgo de seguro asignada, Un valor de +3 indica que el auto es riesgoso, -3 que probablemente es bastante seguro.(Numerica)
- # carCompany: Nombre de la empresa de automóviles (Categórico)
- # fueltype: Tipo de combustible de coche, es decir, gas o diesel (Categórico)
- # aspiration: Aspiración utilizada en un coche (Categórico)
- # doornumber: Número de puertas de un coche (Categórico)
- # carbody: Carrocería del coche (Categórico)
- # drivewheel: Tipo de rueda motriz (Categórico)
- # enginelocation: Ubicación del motor del coche (Categórico)
- # wheelbase: Base del coche (Numérico)
- # carlength: Longitud del coche (Numérico)
- # carwidth: Anchura del coche (numérica)
- # carheight: Altura del coche (Numérico)
- # curbweight: El peso de un coche sin ocupantes ni equipaje. (Numérico)
- # enginetype: Tipo de motor. (Categórico)
- # cylindernumber: Cilindro colocado en el coche (Categórico)
- # enginesize: Tamaño del coche (numérico)
- # fuelsystem: Sistema de combustible del coche (Categórico)
- # boreratio: Boreratio de coche (Numérico)
- # stroke: Carrera o volumen dentro del motor (Numérico)
- # compressionratio: Relación de compresión del coche (Numérico)
- # horsepower: Caballos de fuerza (numérico)
- # peakrpm: RPM máximo del coche (Numérico)
- # citympg: Kilometraje en la ciudad (Numérico)
- # highwaympg: Kilometraje en carretera (Numérico)
- # price: Precio del coche (Numérico)

1. ¿Cuál podría ser a su criterio la variable dependiente para pensar en un modelo de regresión lineal múltiple? Justifique
2. Generen información exploratoria de la variable dependiente elegida, y algunas relaciones con al menos tres columnas.
3. Prueben y generen al menos 3 modelos de regresión lineal múltiple con la variable dependiente elegida. (con variables cuantitativas y cualitativas).

4. Chequeen supuestos para los 3 modelos pedidos
5. Chequeen puntos de influencia para los 3 modelos.
6. Si encuentran puntos de influencia, Que podría plantear? Realice lo planteado.
7. Seleccione el mejor modelo y Justifique.
8. Escriban todo en un script de R , con las justificaciones de lo que están haciendo, colocando el nombre de los integrantes dentro del script, luego de terminado, vamos a poner en común todo lo realizado y pondrán el script en el campus, carpeta TP1 hasta el 24 de Noviembre a las 23:59.