



Universidad Simón Bolívar
Departamento de Cómputo Científico y Estadísticas
Estadística para Ingenieros C03321 y C03322
Abril-Julio 2024

PROYECTO (15%)

En “Datos_car.xlsx”, se proporciona un conjunto de datos que contiene información sobre varios vehículos, incluidas sus especificaciones y precios de venta. El objetivo consiste en ajustar un modelo de regresión lineal múltiple que ayude a los compradores y vendedores potenciales a estimar el valor justo de mercado de carros.

El conjunto de datos contiene la siguiente información:

- ID: Identificador único para cada listado de automóviles.
- Marca: La marca o fabricante del automóvil (por ejemplo, Toyota, Honda, Ford, etc.).
- Modelo: El modelo del automóvil (por ejemplo, Camry, Civic, Mustang, etc.).
- Año: El año de fabricación del carro.
- Kilómetros_recorridos: El total de kilómetros recorridos por el automóvil.
- Tipo_combustible: Tipo de combustible utilizado por el automóvil (por ejemplo, gasolina, gasoil, eléctrico, etc.).
- Transmisión: Tipo de transmisión del automóvil (por ejemplo, manual, automática).
- Tipo_dueño: el número de propietarios anteriores del automóvil (por ejemplo, primero, segundo, tercero).
- Kilometraje: La eficiencia de combustible del automóvil en kilómetros por litro.
- Motor: La cilindrada del motor del automóvil en CC (centímetros cúbicos).
- Potencia: La potencia máxima del automóvil en CV (caballos de fuerza al freno).
- Asientos: El número de asientos disponibles en el carro.
- Precio: El precio de venta del automóvil en INR (rupias indias), que es la variable objetivo a predecir.

Para la base de datos suministrada:

1. Realice un análisis descriptivo de las variables. **(1 punto)**
2. ¿Es cierto que la potencia promedio del carro es diferente si utiliza gasoil o gasolina? (Realice previamente un análisis descriptivo y luego la prueba de hipótesis adecuada. Apóyese en gráficos para su interpretación). **(2 puntos)**

Modelo de regresión lineal múltiple

1. Calcule y grafique la matriz de correlación. Interprete los resultados **(1 punto)**
2. Ajuste un modelo de regresión lineal múltiple. Para ello: **(5 punto)**
 - a. Realice la eliminación de las variables paso a paso.
 - b. En cada paso realice las pruebas de significancia de los parámetros.
 - c. Realice transformadas de la(s) variable(s) en caso de ser necesario.
3. Verifique que el modelo encontrado en el ítem anterior es el adecuado. Muestre los gráficos necesarios. Para ello: **(3 punto)**
 - i. Verifique linealidad visualmente. Analice los resultados de acuerdo a lo observado.
 - ii. Para verificar la homocedasticidad (se puede basar en el estadístico de Durbin-Watson).
 - iii. Verifique independencia de acuerdo a lo observado en las gráficas resultantes.
4. Realice el método de comparación de modelos (AIC) para identificar cual de los modelos resultantes es mejor (en caso de haber resultado más de un candidato a mejor modelo). **(1 punto)**
5. Escriba el modelo de regresión lineal múltiple final. **(1 punto)**
6. Calcule el precio del vehículo de acuerdo a su modelo obtenido en el ítem 5, si **(1 punto)**

Precio			
Año	2019	2020	2021
Kilometraje	11	19	16
Motor	2755	2995	1968
Potencia	172	187	178
Asientos	7	5	5

Nota: Recuerde que en el modelo de regresión lineal múltiple tanto la variable dependiente, como las independientes son variables cuantitativas