



Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística

**ANÁLISIS DE REGRESIÓN PARA LA PREDICCIÓN DEL PRECIO DE UN
AUTOMÓVIL**

Docente:
Mario Enrique Arrieta Prieto

Grupo 7

Autores:

Andrés Mauricio Rico Parada	Estadística	aricop@unal.edu.co
Ander Steven Cristancho Sanchez	Estadística	acristanchos@unal.edu.co
John Anderson Guarín López	Estadística	jaguarinl@unal.edu.co
Juan David Carrascal Ibañez	Matemáticas	jdcarascali@unal.edu.co

Septiembre 2022

1. Descripción

Este trabajo, de carácter académico, busca afianzar los conocimientos adquiridos en el curso Análisis de Regresión, aplicando algunas de las temáticas vistas en clase en un caso práctico. Es por ello que, para esta primera parte se realizará una pequeña introducción y marco general del trabajo que se va a abordar a lo largo del semestre académico. La temática en cuestión consiste en el análisis y predicción de los precios de los automóviles en Estados Unidos. Para abordar esta temática, en esta primera parte se indicará el alcance del trabajo, estableciendo la población objetivo, unidad poblacional, unidad de observación y unidad muestral. Luego, se establecerá un objetivo general y tres (3) objetivos específicos del trabajo. Después se muestra una lista de variables que utilizarán para este trabajo. Finalmente, se describirá la metodología, la cual indica el conjunto de datos a utilizar, y un plan tentativo de trabajo con el fin de llevar a cabo el logro de los objetivos propuestos.

2. Alcance

Población objetivo

Mercado de automóviles estadounidense.

Unidad poblacional

Modelos de automóviles en el mercado estadounidense.

Unidad de observación

Registro de producción del automóvil en el “1985 Ward’s Automotive Yearbooks”.

Constitución de la muestra

Está conformada por las unidades muestrales, que son los modelos de automóviles registrados en el “1985 Ward’s Automotive Yearbook”.

3. Objetivos

General

Estudiar los factores que influyen en el precio de un modelo de automóvil en el mercado estadounidense.

Específicos

- Estudiar las posibles diferencias en el precio de un automóvil de acuerdo a las variables categóricas, como el tipo de combustible, tipo de motor, seguridad, aspiración o forma del auto.
- Proponer futuros planes de negocio con base en la información relevante, para producir automóviles con un precio competitivo en el mercado.
- Analizar si existe relación de algún tipo entre los caballos de fuerza y el tamaño del motor, además, observar si estas variables se ven afectadas por el tipo de motor del automóvil.

4. Variables

A continuación se muestran las variables a considerar en el estudio del mercado de automóviles. Para efectos de este trabajo, sólo se tendrán en cuenta diez (10) variables (5 Cuantitativas y 5 Cualitativas) de las ventiséis (26) disponibles en el conjunto de datos.

Nombre variable	Descripción	Clasificación
Price	Precio del automóvil en dólares	Variable cuantitativa, escala de razón
Highwaympg	Cantidad de millas por galón que puede recorrer en autopista	Variable cuantitativa, escala de razón
Horse Power	Caballos de fuerza del automóvil (Hp)	Variable cuantitativa, escala de razón
Engine Size	Tamaño del motor del automóvil en cm ³	Variable cuantitativa, escala de razón
Aspiration	Sistema de aspiración: puede ser std o turbo.	Variable cualitativa, escala nominal
Curbweight	Peso del automóvil en Kg	Variable cuantitativa, escala de razón
Engine Type	Tipo de motor: dohc, dohcvt, l, ohc, ohcvt, ohcvt, rotor	Variable cualitativa escala nominal
Fueltype	Tipo de combustible: Gas o Diesel	Variable cualitativa, escala nominal
Carbody	Forma del automóvil: sedán, convertible, hatchback, wagon, hardtop	Variable cualitativa, escala nominal
Symboling	Seguridad del automóvil, escala de -3 a 3. Con 3: inseguro y -3: bastante seguro	Variable cualitativa, escala ordinal

5. Metodología de recolección

Para llevar a cabo el proyecto de análisis, se utilizará una fuente secundaria, “imports-85.data”. Un dataset creado originalmente por Jeffrey C. Schlimmer, que cuenta con 205 observaciones y 26 variables, el cual extrae información de diferentes modelos de automóviles en “1985 Ward’s Automotive Yearbook”, correspondiente a sus respectivas especificaciones en “1985 Model Import Car and Truck Specifications”, su rating de riesgo en “Personal Auto Manuals, Insurance Services Office”, y sus pérdidas normalizadas en “Insurance Collision Report, Insurance Institute for Highway Safety”. El dataset original se puede encontrar en el sitio web: <https://archive.ics.uci.edu/ml/datasets/Automobile>.

6. Descripción del plan tentativo del análisis

- Realizar un análisis descriptivo sobre las variables a considerar, graficando la información relevante de los datos e identificando datos atípicos en la muestra.
- Estudiar si las variables a considerar en el estudio tienen alguna relación entre sí.
- Analizar la media del precio de un automóvil en las diferentes categorías de autos.
- Estimar las distribuciones de las variables cuantitativas analizadas en el estudio.
- Identificar un modelo que posiblemente permita predecir el precio de un automóvil, con base en las demás variables del conjunto de datos.
- Estimar los parámetros del modelo seleccionado.

- Identificar los outliers, si los hay, y manejarlos, para después realizar una mejor estimación de los parámetros.
- Determinar cuáles variables tienen mayor y menor incidencia para el modelo.
- Validar el modelo, verificando si se cumplen los supuestos y observando su rendimiento, determinando si los valores del modelo se ajustan bien al verdadero precio de los automóviles.
- Así mismo, aplicar otros modelos para el precio de los automóviles y contrastar su rendimiento.