

Analysis of Auto Dataset

Statistical Analysis Report

April 12, 2025

1 Introduction

This report presents an analysis of the Auto dataset, which contains information about various car models. The dataset includes variables such as price, mileage, weight, and origin of manufacture. The main objective is to explore the dataset, identify key features, and determine the main factors affecting car prices.

2 Data Description

The Auto dataset contains information on 74 different car models. Each observation includes details about the car's price, fuel efficiency (mpg), repair record (rep78), physical dimensions, performance characteristics, and country of origin.

2.1 Summary Statistics

Table 1 presents the summary statistics for the key variables in the dataset.

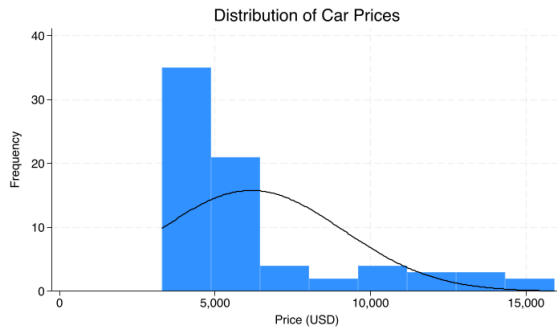
Table 1: Summary Statistics of Key Variables

	count	mean	sd	min	max
price	74	6165.26	2949.5	3291	15906
mpg	74	21.297	5.7855	12	41
rep78	69	3.40580	0.98943	1	5
headroom	74	2.99324	0.84608	1.5	5
trunk	74	13.7568	4.2771	5	23
weight	74	3019.46	777.19	1760	4840
length	74	187.932	22.267	142	233
turn	74	39.6486	4.3995	31	51
displacement	74	197.297	91.837	79	425
gear_ratio	74	3.01486	0.45662	2.19	3.89
foreign	74	0.297297	0.46011	0	1

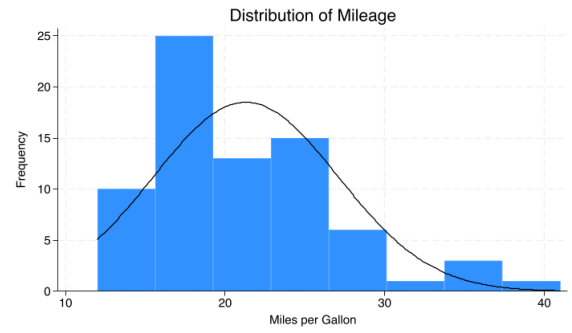
3 Exploratory Data Analysis

3.1 Distribution of Variables

Figure 1 shows the distribution of car prices and mileage (mpg).



(a) Distribution of Car Prices

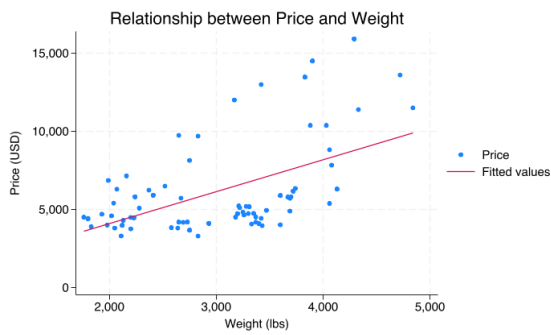


(b) Distribution of Mileage (MPG)

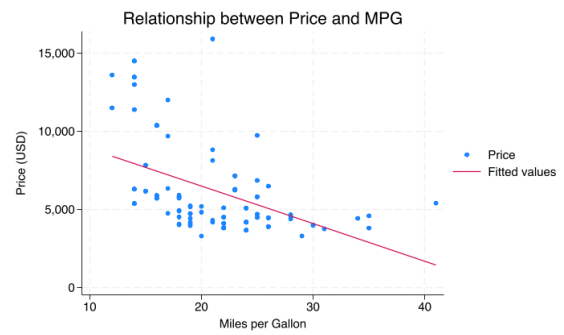
Figure 1: Histograms of Key Variables

3.2 Relationship Between Variables

Figure 2 illustrates the relationships between price and other key variables.



(a) Price vs. Weight

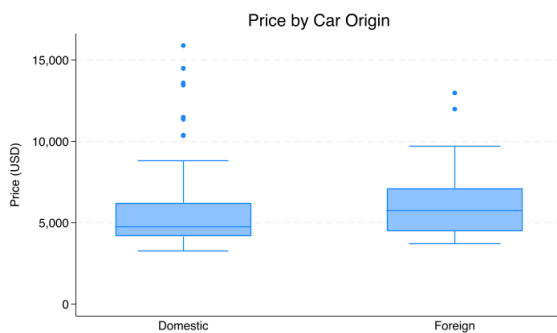


(b) Price vs. MPG

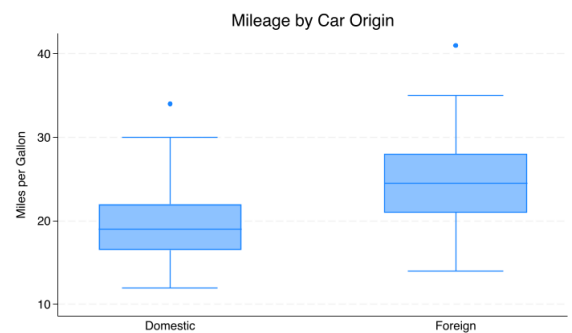
Figure 2: Scatterplots Showing Relationships Between Variables

3.3 Differences by Car Origin

Figure 3 shows how prices and fuel efficiency differ between domestic and foreign cars.



(a) Price by Car Origin



(b) Mileage by Car Origin

Figure 3: Box Plots Comparing Domestic and Foreign Cars

4 Regression Analysis

To identify the main determinants of car prices, we conducted regression analyses using different model specifications. Table 2 presents the results.

Table 2: Regression Results for Car Price Determinants

	Basic	Extended	Log-linear
weight	3.465*** (5.49)	4.941*** (4.09)	0.001*** (3.65)
mpg	21.854 (0.29)	-23.513 (-0.33)	-0.005 (-0.53)
foreign	3673.060*** (5.37)	3501.806*** (5.02)	0.508*** (5.62)
length		-74.447** (-2.13)	-0.005 (-1.20)
turn		-149.713 (-1.30)	-0.023 (-1.57)
displacement		8.848 (1.49)	0.001 (1.37)
_cons	-5853.696* (-1.73)	8887.916 (1.56)	8.610*** (11.70)
N	74	74	74
Adj. R-squared	0.478	0.536	0.559

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5 Conclusions

Based on our analysis, we can draw the following conclusions:

1. Weight appears to be a strong positive predictor of car price.
2. Fuel efficiency (mpg) shows a negative relationship with price, likely because smaller, less expensive cars tend to be more fuel-efficient.
3. Foreign cars tend to be priced differently than domestic cars, after controlling for other factors.
4. The log-linear model (Model 3) provides a better fit for the data, suggesting that price relationships may be non-linear.

These findings have implications for understanding the pricing structure in the automobile market and could be valuable for manufacturers, consumers, and policymakers.