Linear Regression Project:

**Predicting the Price of a BMW**



Marketing Analytics

Professor Varki

Spring 2020 Semester

Christopher Fennimore, Qianxi Liu, Yi Lu, Natalie Quirindongo

**Table of Contents**

Descriptive Statistics……………………………………………………………………………...3

Summary Statistics (Quantitative Variable Chart)..........................................................................3

Summary Statistics (Qualitative Variable Chart)............................................................................4

Variables Chosen………………………………………………………………………………….5

Goodness of Fit Statistics…………………………………………………………………….…...5

Model Parameters…………………………………………………………………………………6

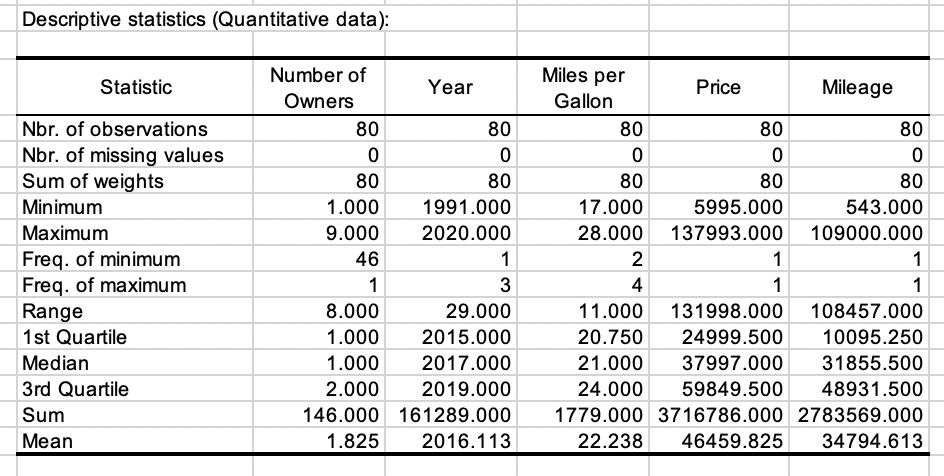
Statistically Significant Variables………………………………………………………………....7

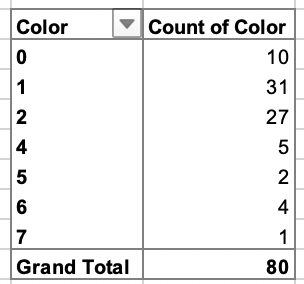
Elasticity Equations……………………………………………………………………………….9

Summary of Elasticity………………………………………………………………………...…11

**Section 1)**

In this report, we have gathered data on previously owned vehicles to predict the price of BMW’s. BMW specializes in luxury vehicles and we chose to specifically look at 5 series, 6 series, 7 series, and 8 series. These vehicles are either coupes or sedans and seat 4-5 people. Because of their similarities of size and brand, we wanted to take a closer look into what variables may affect the pricing. 80 vehicles were reported in total, with 20 observations of each car model. The information collected includes the listed price of the vehicle, year, mileage, model, the color of the vehicle, number of owners, and miles per gallon (average between city and highway).

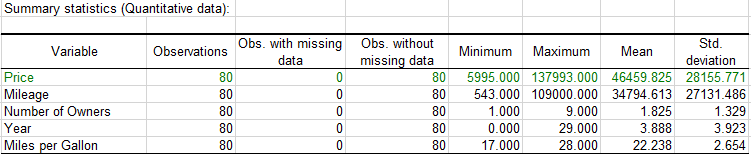


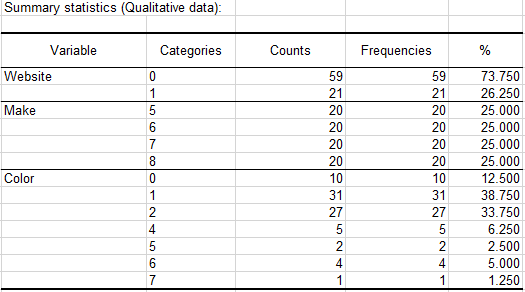
Above is the calculated descriptive statistics for the data set. Highlighting a few key observations, the range in vehicle pricing is $5,995-$137,993, with $46,459 approximately being the average. The car mileage ranges from 543-109,000, with 34,794 miles being the average. The cars have a range of 17mpg-28mpg, and the average is 22mpg. The car’s year ranged from 1991-2020, having a maximum of 9 owners within its lifespan and 1 minimum.

The most common color of car was black(31) and the least common color was orange(1).

As we know, taking data is susceptible to random measurement error, and the line and its equation are an approximation of the data, so here, we are going to develop a model from the trend of the data accounting for the variability. We used mileage, number of owners, year, miles per gallon, website, make series, and color as our independent variables to predict the price of the BMW car.

Mileage, number of owners, year, and miles per gallon are our quantitative data; website, make series, and color are our qualitative data. Variables are shown as the following:



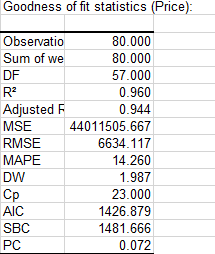


|  |  |  |
| --- | --- | --- |
| **Make** | **Color** | **Website** |
| 5 = 5 series  6 = 6 series  7 = 7 series  8 = 8 series | 0 = gray  1 = black  2 = white  3 = red  4 = silver  5 = brown  6 = blue  7 = orange | Cargurus.com = 0  Cars.com = 1 |

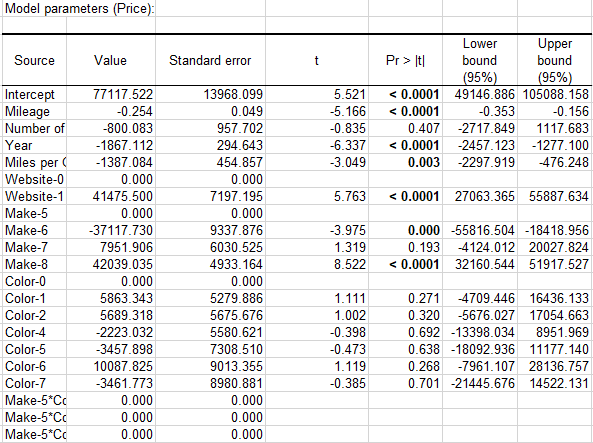
**Section 2)**

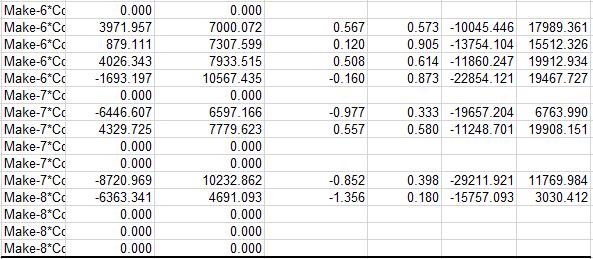
The percentage of the variation r^2 in the dependent variable that is explained by the variation in the independent variable is a measurement of how successful the regression was in explaining the response variable. When r^2 = 1, we can say all of the variation in one variable is accounted for by the linear relationship with the other variable. As our data analysis shows, r^2 = 0.944 which is very close to 1. As a result, we justify our model for those data is linear regression.

The only non-obvious variable among all of the variables is website. Other variables such as mileage, number of owners, year, miles per gallon, make series, and color are all related to the car itself, which means these factors have an effect on the price of the car obviously, but the website is not. Website is just a platform provided for consumers to look for and buy cars, and its effect on the price may not be very obvious. However, this variable is still useful for predicting the price of the cars. Cars.com and Cargurus are competitors, so they may have different dispositioning, including price difference, among the competition. Also, different websites have their own brand reputation and sales volume, and all of these might be the consideration of buying a used car for customers, then sellers can adjust their price in different websites according to the condition of websites.



**Section 3)**





During this project, we used a variety of different variables to determine the price of a BMW car. We ran a linear regression analysis on the 80 different BMW cars and came up with interesting results. With the variables of mileage, make, the number of owners, color, mileage, website, and miles per gallon the results came back unexpected.

The first independent variable that was statistically significant was the year. As an analyst, I figured this was going to be one of the most important variables. The results came back proving that year had a p-value of <0.0001. This was significantly lower than the alpha value of 0.05 showing that year has an impact on predicting the price of a BMW. From the Modeling Parameters chart, this statistic can be read as follows:

On average with a one-unit increase in year, all else constant equals an $1,867,11 decrease in price in BMW cars.

The second independent variable that was statistically significant was the type of website. I threw in this variable because I was curious to see if the type of website had a factor in price premiums. We used two different websites Cargurus.com and Cars.com to find our pricing and variables for each BMW. The results show that the type of website had a p-value of <0.0001 significantly lower than the alpha of 0.05. From the Parameters chart, this statistic can be read as follows:

On average, all else constant, when using Cars.com a BMW will increase in price by $41,475.50 dollars.

The third independent variable that was statistically significant for determining the price of a BMW was mileage. This resulted in a p-value of <0.0001 lower than the alpha of 0.05. This statistic could be read as the following.

On average, all else constant, a one unit increase in mileage results in a decrease of $0.25 in price for a BMW.

The fourth independent variable that was statistically significant was miles per gallon. This came back with a p-value of 0.003 which was lower than the alpha value of 0.005. Miles per gallon affected the price of a BMW by this:

On average, all else constant a one unit increase in miles per gallon per resulted in a decrease of $1,387.08 in price of a BMW.

The two different makes that were statistically significant were the model 6 and model 8.

Model 6 had a p-value of 0.000 and Model 8 had a p-value of <0.0001. For determining the price of a BMW for these two models is as follows:

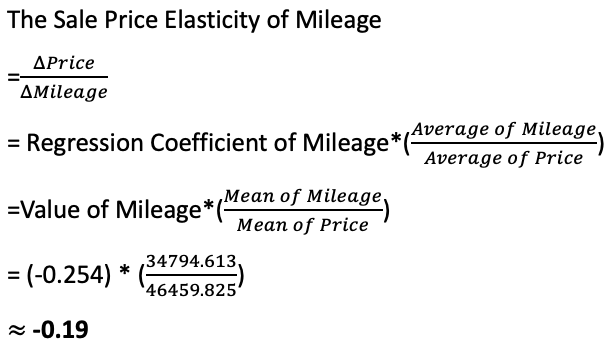
Model 6: On average, with all else constant, a one unit increase in model of BMW to a model 6 results in a $37,117.73 decrease in price of a BMW

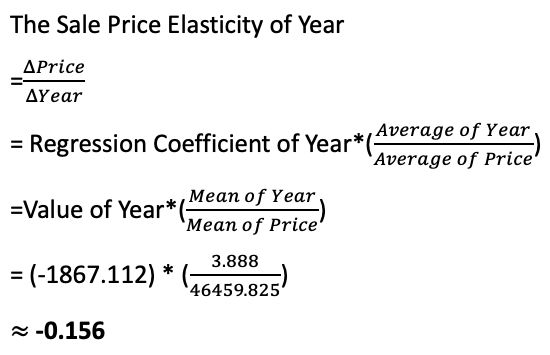
Model 8: On average, with all else constant, a unit increase in BMW to a model 8 results in a $42,039.03 increase in price of a BMW.

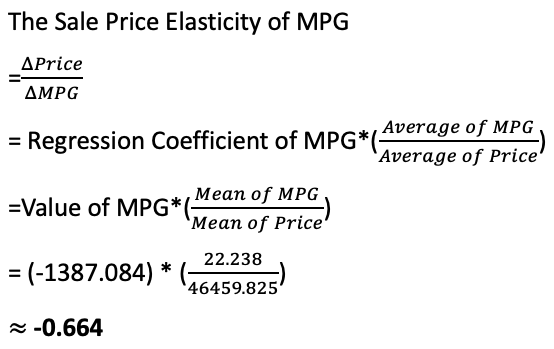
**Section 4)**

In terms of continuous and significant variables in this case, we conclude that year, mileage, miles per gallon,website, and makeare the target according to P-Values of variables.

|  |  |  |
| --- | --- | --- |
| **Variables** | **Value** | **Mean** |
| **Mileage** | -0.254 | 34794.613 |
| **Year** | -1867.112 | 3.888 |
| **Mile per Gallon** | -1387.084 | 22.238 |
| **Price** | - | 46459.825 |







|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | **Type of variable** | **P-Value** | **Significant** | **Elasticity** |
| **Mileage** | **Quantitative** | **<0.0001** | **Yes** | **Inelasticity**  **(-0.19)** |
| **Number of Owners** | Quantitative | 0.407 | Non | - |
| **Year** | **Quantitative** | **<0.0001** | **Significant** | **Inelasticity**  **(-0.156)** |
| **Miles Per Gallon** | **Quantitative** | **0.003** | **Yes** | **Inelasticity**  **(-0.664)** |

The result of the sale price elasticity of mileage in this regression model is -0.19, the absolute of elasticity is -0.19 which is less than 1, therefore we can report that the Sale Price Elasticity of Year is ***Inelasticity.***

The result of the sale price elasticity of year in this regression model is -0.156, the absolute of elasticity is -0.156 which is less than 1, therefore we can report that the Sale Price Elasticity of Year is ***Inelasticity.***

The result of the sale price elasticity of miles per gallon in this regression model is -0.664, the absolute of elasticity is -0.664 which is less than 1, therefore we can report that the Sale Price Elasticity of Year is ***Inelasticity.***