

# Research paper on factors affecting car prices

[Technology](#), [Cars](#)



## **MATH533 Project Part 1**

MATH533 - Managerial Statistics

- Overview of the Problem and Questions

In our research we will focus on factors which affect car prices in United States. One can easily argue that the price of a particular car depends on model of the car, the year it was manufactured, its mileage, engine capacity, horsepower, fuel consumption, etc. Our study will include only 2 main factors which are believed to influence price of the car the most - its regular age and mileage.

We realize that many other factors except age and mileage influence the price of the car, therefore, in order to reduce the variability in our measurements we were trying to analyze cars which are similar in characteristics. In our research we were looking at a particular car model - Ford Focus, Sedan, with automatic transmission and 4 cylinder engine, running on gasoline. We believe that if we analyze cars which have identical characteristics (except for age and mileage) our findings will be more adequate.

We assume that car price is negatively correlated with both, its mileage and age. In other words, we believe that the bigger the mileage - the lower the price. The same thing applies to the actual age of the car: the older the car - the lower the price is. The purpose of our study is to prove or reject our assumptions.

**We are posing the following question to be studied:**

How is a price of a car affected by its age and mileage?

- List of variables

**The following variables will be used in this study:**

- Sources of data

The data was gathered randomly from major US websites (cars. com and autotrader. com) which specialize on selling used cars. Filtering feature helped us to make other possible determinants (body type, transmission, engine capacity) fixed, so that our estimation would be more adequate. Data sample includes 50 observations, which is considered to be large enough to be able to perform analysis and form comprehensive conclusions.

- The Data

**The data about the 50 Ford Focus cars selected for this study is presented in the table below:**

- Descriptive Statistics
- Dependent Variable – Price.
- Independent Variable – Age.

Variable Age contains only integers and varies from 1 till 8, which suggests that the oldest car in the sample is 8 years old, and the newest car was manufactured last year. We may see that average age of the observed cars is 4. 5 years, with median age of 5 years. Since independent variable consists only from integer values from 1 to 8 there is not much more information which we could derive from descriptive statistics.

- Independent Variable – Mileage.
- Scatter Plots

## 6. 1 Relationship between age and price.

We may observe from the data, that there is a general negative relationship between age of the car and its price. The scatterplot shows that price of the car actually decreases as the age of the car increases. The downward sloping trend is clearly observed on the scatterplot.

## 6. 2 Relationship between Mileage and Price.

- Simple Linear Regression Analysis

### 7. 1 Regression of Age on Price.

The above table gives us the results of the linear regression of age on price. From the regression results we may infer that Age is a statistically significant variable which negatively correlates with the dependent variable price, as p-value of the age coefficient = 0. Beta coefficient equals to -0. 83. It basically says that with each additional year of age reduces the price of Ford Focus by \$1, 408. Coefficient of determination is 68%. It suggests that variation of car prices are explained by independent variable by 68%. In other words, age of the car is responsible for 68% of variation in its price.

### 7. 2 Regression of Mileage on Price.

Linear regression of mileage on price produces similar result. From the regression results we may infer that mileage and price are negatively related. Mileage coefficient is statistically significant, as p-value is close to zero. Unstandardized Beta coefficient suggests that with each additional mile traveled, Ford Focus loses \$0. 07 from its price. We may also argue that given model may not provide us with adequate results, as determination coefficient equals 56%. Therefore, 56% of price variation are explained by mileage, and the remaining 44% are explained by some other external

factors which were not included in the model.

#### - Multiple Linear Regression Analysis

Finally, multiple regression results provide us with a better picture of how car prices depend on age and mileage. From the above table we may claim, that coefficient of determination has risen dramatically compared to previous examples. Now we may conclude that 82% of variation in cars' prices is explained by variables included in the model, namely, age and mileage. All coefficients in the model are statistically significant, as p-value of all of them is very close to 0. By looking at the B table we may infer, that each additional year of a car lead to reduction in its price by \$1, 010. In addition, each additional mile traveled by the car is associated with decrease of its price by \$0. 04.

#### - Conclusions and Summary

In conclusion we may claim that our model proved our assumptions which we discussed before the start of our research. Mileage and age of the car are negatively correlated with its price. All our data point out this conclusion. Scatterplots, single linear regression and multiple regression results all produced similar results - mileage and age have a statistically significant negative relation with car's price.

However, we may not conclude that our research results apply to the whole US car market. Firstly, we were analyzing linear relationships regarding to only one model of the car. It is possible, that cars from different manufacturers or price segments have different relationship pattern between price and independent variables. In addition, we cannot be sure that our sample is truly random and independent, as it was taken from an online

source. Many used cars are sold without using ads on the internet, and pattern of relationships of that segment may also be different. Finally, sample size of 50 observations is still relatively small and results in large standard error of the data. Consequently, our findings may be not reliable. Further research may include more independent variables into the analysis in order to find out which other factors affect car prices. In addition, data collection procedures may be enhanced to produce more consistent and reliable results. It maybe also interesting for researchers to analyze how the differences in the class of the car change the influence of factors which affect its price.

## References

Brace, C. H. 2012. Understandable Statistics: Concepts and Methods

StatSoft, Inc. 2013. Electronic Statistics Textbook. Tulsa, OK: StatSoft.

Retrieved from <http://www.statsoft.com/textbook>

Berger, J. O. 1993. Statistical Decision Theory and Bayesian Analysis, 2nd Edition.

<http://www.autotrader.com>

<http://www.cars.com>