GBA 5140 Statistics Essentials for Business Analytics

Problem Set 10

Multiple Linear Regression

**Problem 1**

Load “Data-UsedCars.csv” into RStudio. This dataset contains attributes of used cars and their price. Write R code for the following tasks. ***Post your R code and RStudio output (console output) below each task.***

1. The value of a used car is measured by its price (in euros). Build a multiple linear regression model with “Price” as the dependent variable, and with the following independent variables (if necessary, please write code to adjust categorical variables in the dataset before running the regression to make sure that the correct value is used as the baseline):

“Age\_08\_04”: car age as of August 2004 in months, quantitative

“Met\_Color”: whether the car has metal color or not, categorical, use “no” as baseline

“Weight”: car weight, quantitative

“HP”: car horsepower, quantitative

“HP”: car mileage in kilometers, quantitative

“Quaterly\_Tax”: car specific quarterly road tax, quantitative

“Fuel\_Type”: car fuel type, use “Petrol” as baselineGraphical user interface, text, application

Description automatically generated

Text

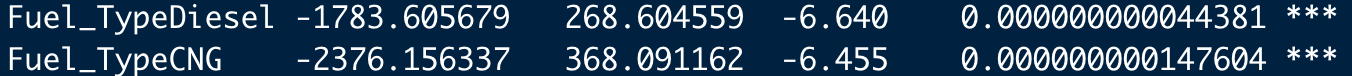
Description automatically generated with medium confidence

1. Which independent variables have significant effect on the dependent variable?

The independent variables that have a significant effect on the dependent were “Age\_08\_04”, “Weight”, “HP”, “KM”, “Quarterly\_Tax”, and “Fuel Type”

1. Explain the effect of independent variable “KM”. 

Holding all else equal one unit of increase in “KM” will decrease the Price by a factor of .017248.

1. Explain the effect of independent variable “Fuel\_Type”.

Holding all else equal if the fuel type is diesel there will be a decrease of 1783.61, in comparison to the Petrol baseline, to the Price and if the fuel type is CNG there will be a decrease of 2376.16, in comparison to the Petrol baseline, to the Price

1. Explain the effect of independent variable “Met\_Color”. 

Holding all else equal if the car has metal color there will be an increase of 1783.61 to the Price in comparison to no metal color baseline.

1. What are the model’s and adjusted ? What can you conclude?

86.77% of the variance in the dependent variable can be explained by the combination of independent variables.

1. Compute VIF of all independent variables. What can you conclude?

Text, logo

Description automatically generated with medium confidenceText

Description automatically generated

Since there is not independent variable with a VIF of greater than ten we can say that there is no multicollinearity problems with the variables.

1. Plot standardized residuals against the fitted values (i.e. estimated values) of dependent variable. What can you conclude.

Using residual analysis we can conclude that there is homoskedasticity and there are a few outliers in both the positive and negative range from the mean.

Graphical user interface, text

Description automatically generatedChart, scatter chart

Description automatically generated