Tyler Hudson

Regression Homework

This homework is based on the CarPrice\_Assignment file. The data dictionary is in the car price prediction zip folder as well. The dependent variable (your target variable) is Price. Alpha is equal to 0.05.

These are the continuous random variables I have identified for you.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| compressionratio | horsepower | peakrpm | citympg | highwaympg |

1. Create a table of correlations with the variables listed above and the dependent variable. Copy/paste the table and the script here.

cor(CarPrice[c(-(0:20))])

Text

Description automatically generated

1. Write the order of the correlations relative to the dependent variable.

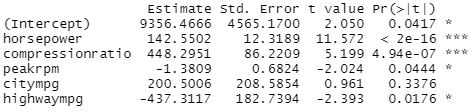
Price 0.06798351 0.8081388 -0.08526715 -0.6857513 -0.69759909 1.00000000

1. Build a simple linear regression. Explain what the p value next to the F-Statistic implies

The p value implies that the horsepower variable does not influence price because it is much smaller than the price x horsepower covariance. That p-value being 0.8081388.

F-statistic: 382.2 on 1 and 203 DF, p-value: < 2.2e-16

1. Go through the procedures described to you in class via the forward model building procedures and create a multiple linear regression with the 5 continuous random variables (not all continuous variables will necessarily make it to the final model you come up, remember how to use the adjusted r^2 and the p values next to the independent variables). Copy/paste the summary output here.



1. Add to the model you identified in 4, carbody feature. Interpret the p value of wagon.

The p value next to the F-Statistic implies that the carbodywagon variable does not influence price because it is so small compared to the price x carbody covariance (0.004941).

F-statistic: 73.84 on 9 and 195 DF, p-value: < 2.2e-16

1. What percentage of the variation in Price are you able to explain with the model you have used in (5)

Multiple R-squared: 0.7731

Adjusted R-squared: 0.7627

Approximately 77% can be explained from the variation in Price

1. Build a regression model using Carbody, Horsepower,CityMPG and Highway MPG. Predict the Price of a car that is : CarBody: Convertible, Horsepower:111,CityMPG:21,HighwayMPG:27.

$18,497.96

1. This corresponds to the car description in row number 1. What is the residual for the car? Compute it via Y-E(Y)

Residual value of car = $ -5,002.96 (overvalued)

1. Look at the summary output of your regression object. Interpret the p value of the global hypothesis test (p value next to the F-Statistic). Interpret the coefficient of wagon.

Comparing to the global hypothesis test, the p-value is again too small to influence anything. The coefficient of wagon can be interpreted by the change in price (if changing from convertible car body type to wagon).