Data and Decisions

In Class Assignment



Hedonic regression is commonly used in business analytics. It is a revealed preference method for estimating the monetary value of the characteristics of a particular good. To estimate a hedonic regression, we regress the price of a good (e.g., the price of a car) on several of its attributes. Under certain assumptions, the coefficients on the regression will provide us with an estimate of the monetary value contribution of each characteristic. For this assignment, we use the data in the excel file titled “CarPrices.xlsx” and it can be found on *Canvas*. These data describe 205 vehicle models sold in the United States along their various attributes, including price. For this assignment, we will focus on the variables: price and highway mpg.

1. *Computing summary statistics*: What is the mean miles per gallon on the highway in these data? What is the standard deviation? Which car(s) get the lowest miles per gallon on the highway? Which vehicle has the largest highway mpg?
2. *Linear prediction line*: Estimate the best linear prediction line with price as the response variable and highway mpg as the explanatory variable. Display the regression output.
   1. Interpret the both the slope and the intercept. Be sure to use the correct units.
   2. Create a scatterplot of price on highway mpg with your estimated prediction line. This is what EXCEL calls the “Line Fit Plot”. Please label the units on both axes. How is the fit of this regression model? Does the model tend to underpredict or overpredict the price of vehicles with respect to mpg?
   3. Create a scatterplot of the residuals on highway mpg. This is what EXCEL calls the “Residual Plot”. Please label the units on both axes. Do we notice any similarities between the residual plot and line-fit plot? What does a positive residual represent?
3. *Curved prediction line*: Now estimate a regression with the natural log of price as the response variable and highway mpg as the explanatory variable. You will need to create the new response variable yourself. Display the output table.
   1. How do you interpret the slope coefficient?
   2. Create a residual plot. How does it compare to the previous residual plot. Explain briefly.
   3. The prediction line we have estimated can be written as follows: . We can transform this equation by taking the exponential function of both sides: . Compute this transform prediction and add to the scatterplot of price and highway mpg. What do we notice?