Nghia Lam

1001699317

1. Copy your results for model2 (or model2b): make some observations about your model2 output (Compare them to the full model (model1))
   1. How many variables are significant,
      1. There are 8 Significant figures for model2, all but highway\_mpg
      2. There are 9 Significant figures for model1
   2. What is your adjusted R-squared compared to the full model, in general, do you think your model is better etc.?
      1. Model1 has an R-squared of .849
      2. Modle2 has an R-squared of .8482
      3. I believe model2 to be better than model 1 as we have less variables
2. Copy your ANOVA output. What can you tell about the results?

A screenshot of a computer code

Description automatically generated

1. Compare the full model and reduced model for multicollinearity, what are your observations?

A close-up of a number

Description automatically generated

* + 1. Model 1 has some values with very high multicollinearity like fuel\_type, curb\_weight, comprassion, city\_mpg, highway\_mpg.
    2. Model 2 however does not have a multicollinearity problem unlike model1, as all variables are less than 10

1. A diagram of a graph

   Description automatically generated with medium confidenceA graph with a line

   Description automatically generatedA graph of a graph

   Description automatically generated with medium confidenceA graph with numbers and dots

   Description automatically generatedCopy the plots for the assumptions and compare the results for the reduced and the full model (e.g. did your plots improve?)
   * 1. A graph with a line

        Description automatically generatedA graph with numbers and lines

        Description automatically generatedA graph with dots and lines

        Description automatically generatedA graph with black dots and red line

        Description automatically generatedAbove is model1.
     2. Above is model2
     3. Model 1 and model2 look very similar, my plots did not improve.
2. Compare the stepwise selection methods. What did you notice?
   * 1. Both and Backwards are identical as Model 1 already has all the variables inputted into it so all it would need to do is go backwards, that is why I believe both and backwards are the same.
     2. Forwards wouldn’t do anything as are the variables are already inputted
     3. F statistic is overall better for backwards and both, and forwards has a very low F statistic compared to the other 2.
     4. The p values for all of them were very similar that is why I used it less in my analysis
3. In general, comparing the process for regression analysis in Python and R, which one did you like better and why?
   * 1. I liked R because it was easier to implement everything compared to python.