**Module Assignment**

**Module 6 (Demonstration)**

**QMB-6304 Analytical Methods for Business**

Write a simple R script to execute the following data preprocessing and statistical analysis. Where required show analytical output and interpretations.

**Preprocessing**

1. Load the file “6304 Multiple Regression Assgt Example Data.xlsx” into R. This file contains information on 383 automobiles marketed in the United States between 1970 and 1982. The variables included miles per gallon (MPG), cylinders in the engine, engine displacement in cubic inches, horsepower, weight (in pounds), model year, country of origin, auto make, and auto model. This will be your full data set.
2. Convert the "year" and "cylinders" variables will have been converted to factors.

**Analysis**

1. Conduct a multiple regression analysis using the data in your reduced data set. Use MPG as the dependent variable and cubic inches, horsepower, and weight as independent variables.
2. Show your model output with appropriate discussion of the p value for each beta coefficient (including β0). Give proper written interpretations of the beta coefficients which explain the variable’s estimated impact on the y.
3. Report and interpret confidence intervals for each beta coefficient in your model.
4. Determine and state whether your model appears to be in conformity with the LINE assumptions of regression. Show appropriate graphics and give written interpretations where needed to justify your conclusions.
5. Using the tools presented earlier in this course determine whether any of the data points in your reduced data set have a high leverage in influencing the plot of the regression. Show appropriate graphics to support your conclusion. If you do have high-leverage data points report ONLY the year, make, and model of these autos.
6. Introduce squared terms into your model which are based on the horsepower and weight variables. Does using either or both these squared terms improve the fit of your model? Explain your reasoning on this point.
7. Using your original multiple regression model from Part 1 above, introduce the “cylinders” factor variable to your model. Do any of the factor levels appear to contribute to the fit of the original model?
8. Using your original multiple regression model from Part 1 above, introduce the “year” factor variable to your model. Do any of the factor levels appear to contribute to the fit of the original model?

Your deliverable will be a single MS-Word file created using R Markdown. Your file will show 1) the R script which executes the above instructions and 2) the results of those instructions. The first two lines of your deliverable will state this is “Assignment 5” of our course and your name as it appears in Canvas. Your code chunks and analysis results should be presented in the order in which they are listed here. Deliverable due time will be announced in class and on Canvas. This is an individual assignment to be completed before you leave the classroom. No collaboration of any sort is allowed on this assignment.