Stat 481 - Project 1 Due Time: 5PM on Friday, March 31, 2023.

Instructions:

- Project must be typed in a word or pdf file for credit. Use COMPLETE SENTENCES and justify your findings.
- Include only necessary figures and tables, and comment on them. Attach your code at the end of the project in a separate Appendix section.
- Do not include R or SAS code in the main section of the report. Projects submitted using R-markdown or a set of code plus a sheet with brief answers will not be accepted for grading.

Dataset Location:
Use the data set provided to you on Blackboard, `MtCars.txt` (with a header).

Motor Trend Car Road Tests Description.
The data was extracted from the 1974 *Motor Trend US magazine*, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). The data considered in this project includes 32 observations on 5 (numeric) variables.

Variable Name	Description
mpg	Miles/(US) gallon
disp	Displacement (cu.in.)
hp	Gross horsepower
wt	Weight (1000 lbs)
qsec	1/4 mile time

Question:
Our interest in performing a regression analysis is almost always to answer some sort of research question. Can you think of some research questions that the researchers might want to answer here? How about the following set of questions? What procedure would you use to answer each research question? (Do the procedures that appear in parentheses seem reasonable?)

- Which, if any, predictors — displacement, gross horsepower, weight or 1/4 mile time — explain some of the variation in the fuel consumption data? (Two sets of hypotheses: primary hypotheses for all coefficient parameters, and individual hypotheses for partial test on each coefficient parameter.)
- What is the effect of gross horsepower on average mpg, after taking into account all other predictors? How about displacement and weight? (Calculate and interpret a confidence interval for the individual coefficient parameter.)

Key Items of a Statistical Report:

1. Title Page. Include project title, your name, and a brief summary of the project.
2. Section 1. Data Introduction and Description. Provide descriptive statistics such as five-number summary, mean, variance/sd, histogram, or boxplots.
3. Section 2. Multiple Linear Regression Analysis.
    - Provide a brief introduction on each model/test/method you use in the regression analysis. Write down appropriate statistical models at each step of your analysis. State your problems with null and alternative hypotheses first, based its ANOVA table and summary statistics, conclude accordingly.
    - Check the model assumptions: linearity, independence, normality and equal variance of residuals
        - Provide any applicable plots (in a reasonable size) or tests and interpret them.
        - If any of the model assumptions are not met, suggest ways to "fix" your data, for example the Box-Cox transformation.
        - Be sure to re-check all the model assumptions after any transformations and address each of the assumptions in your report.
        - Note: To simplify things, if you need to do a transformation, do *one* transformation, and proceed with analysis.
    - Check for multicollinearity. This test only needs to be done once at the beginning of the analysis.
4. Section 3. Variable Selection. Build the "best" model possible by using either backward selection or forward selection (pick one) with the criteria for inclusion as having a significance of 0.10 or lower. Make sure to check all the model assumptions on the final model.
5. Section 4. Conclusion. Draw conclusions/interpret your regression model. Include a statement about R^2 before and after creating the "best" model possible. Include statements about each variable kept in the final model. *Make sure these conclusions can be understood by a general reader.*
6. Appendix. Attach R or SAS code in this section only.

Grading:

Your project will be graded on the following items:

- Include a detailed description about data and Data Summary (descriptive statistics, scatter plot matrix, histogram/boxplot, distribution)
- Regression Analysis and Assumptions Check on each model
- Drawing Conclusions based on the "best" model possible. Includes final model statement, interpretation of parameter values, how R^2 changes before and after creating the "best" model.

Some useful SAS procedures :

- DATA or PROC IMPORT
- PROC REG
  SAS PROC REG - model options: https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_reg_sect013.htm
- PROC CORR
- PROC UNIVARIATE
- PROC TRANSREG
- SAS Procedure Help
  https://documentation.sas.com/?docsetId=proc&docsetTarget=titlepage.htm&docsetVersion=9.4&locale=en

Some useful R functions :

- read.table() or read.csv() to import data

  ```
  read.table("C:\\Stat481\\Project\\MtCars.txt", header=TRUE)
  ```

- lm() and plot(lm()) to fit linear regression model
- cor() for correlation coefficient
- boxcox() in MASS package
- vif() in car package
- Chapter 11 (R Introduction)
  https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf