Assignment #5: Automated Variable Selection and Predictive Models (50 points)

Data: The data for this assignment are associated with the Exercise 9.3 in Chatterjee and Hadi (2012, pages 255–256), which we refer to as the "gasoline consumption case."

Assignment Instructions:

In this assignment we build regression models for the gasoline consumption case, building on what we have learned from previous assignments in the course.

(1) Define the Sample Population

- Define the appropriate sample population for your statistical problem.

(2) Exploratory Data Analysis

- Utilize statistical graphics to explore the data from the case.

- Ensure that you employ individual scatter plots or a scatter plot matrix showing relationships between pairs of variables.

(3) Multiple Linear Regression – The Full Model

- Fit a model with the complete set of predictors in this case. Interpret the results.

(4) Multiple Linear Regression – Subset Model

- Employ a systematic method for selecting a subset of predictors. Fit the subset model and interpret the results.

(5) Model Comparison and Recommendation

- Compare the full and subset models. Which model would you recommend to management, and how can management use this model in making predictions?

Assignment Document:

All assignment reports should conform to the standards and style of the report template provided to you. Results should be presented and discussed in an organized manner with the discussion in close proximity of the results. The report should not contain unnecessary results or information. The document should be submitted in pdf format. Name your file **Assignment5_YourLastName.pdf**

The assignment pdf file and accompanying plain text files for Python programs should be included in a zip archive using standard zip compression with the name **Assignment5_YourLastName.zip**

Here is a reasonable section outline for this assignment report.

Section 1: Define the Sample Population

Section 2: Exploratory Data Analysis and Simple Linear Regression

Section 3: Multiple Linear Regression – Full Model

Section 4: Multiple Linear Regression – Subset Model

Section 5: Model Comparison and Recommendation