Assignment #6: Multicollinearity and Principal Components Analysis (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 fit regression models for the gasoline consumption case, building on what we have learned from previous assignments in the course.

(1) Review Sample Data and Exploratory Data Analysis

   - Provide an executive summary of previous work from Assignment #5.

(2) Multiple Linear Regression – Full and Subset Models

   - Provide an executive summary of fitted models from Assignment #5.

(3)   Principal Components Analysis

   - Using predictor variables from the case, identify the set of principal components. Describe the process in terms of linear algebra. Determine the number of principal components needed to account for 50, 70, and 90 percent of the variance associated with the full set of predictors.

(4)   Principal Components Regression
   -   Regress the response on sets of principal components accounting for 50, 70, and 90 percent of the variance in the predictors.

(5)   Model Comparison and Recommendation

   - Compare the original full and subset models with the various principal component 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 **Assignment6_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 **Assignment6_YourLastName.zip**

Here is a reasonable section outline for this assignment report.

Section 1: Review Sample Data and Exploratory Data Analysis

Section 2: Multiple Linear Regression – Full and Subset Models

Section 3: Principal Components Analysis

Section 4: Principal Components Regression

Section 5: Model Comparison and Recommendation