STAT 210
Applied Statistics and Data Analysis: Homework 7

Due on Nov. 06/2022

Question 1

For this question use the dataset `data1`.

This dataset has information on fuel efficiency, measured in miles per gallon, and seven other variables for 80 different car models. There are two variables related to fuel efficiency, `City.Mpg` and `Highway.Mpg`. We will only consider `City.Mpg`, and we will work with the reciprocal of this variable, 1/`City.Mpg`, which we will call `City.fc` for fuel consumption. We want to explore the relation between `City.fc` and the car's weight (`Weight`). 互惠的

(i) Read the data and define a new variable called `City.fc` in the data frame equal to the reciprocal of `City.Mpg`. Draw a scatterplot of `City.fc` as a function of `Weight`. Fit a simple linear regression for `City.fc` as a function of `Weight` and add the line to the plot. Comment. Obtain a summary of the regression and comment.

(ii) Draw the diagnostic plots. Do you identify any point as an outlier? If you do, which point is this? Can you identify this point in the initial scatterplot? Can you find a reason why this point is different from the rest?

(iii) Fit a new regression model excluding the outlier(s) you identified in the previous section. Draw a scatterplot with both regression lines. Compare the summary tables. Draw the diagnostic plots and comment.

(iv) Run the Shapiro-Wilk test on the residuals for both models and compare the results.

Question 2

For this question use the data set `data2`.

The data for this question come from an experiment to determine the relationship between the volume of a gas and the pressure. The file has two variables, `Height` and `Pressure`. Height corresponds to the height of a cylindrical container with a fixed circular base and a movable top that allows changing the volume of the container. `Height` was measured in inches. `Pressure` is measured in inches of mercury as in a barometer. We want to study the relationship between these two variables.

(i) Read `data2` and plot `Pressure` as a function of `Height`. Fit a simple linear regression for `Pressure` as a function of `Height` and add the regression line to the plot. Comment. Obtain a summary for the regression and draw the diagnostic plots. Comment on the results

(ii) Use the function `boxcox` on the MASS package with the argument set to the model you fitted in (i). If the maximum value in the graph is close to an integer value, use this integer as the exponent in a power transformation for `Pressure`, i.e., if P is the `Pressure` variable in the dataset, the new variable TP is given by

$$TP = P^i$$

where i is the integer obtained from the boxcox function. Fit a new model for the transformed pressure as a function of `Height`. Obtain a summary of the new regression and compare it with the previous one. Draw the diagnostic plots and compare them with the earlier results.

(iii) If in the model you fitted in (ii), the p-value for the intercept is large, fit a model without intercept by adding `+ 0` at the end of the regression equation in the call to the `lm` function. Use this model to write down an equation for the relation between pressure and volume for a gas. What would be the predicted `Pressure` for a point with `Height` $= 32$? Draw a scatterplot of `Pressure` against `Height` and add the regression line for the first model and the curve you obtained with the second regression.
