Lab 2: Correlation and Regression

Overview

In this lab, we will examine the Motor Trend Cars data set.

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).

Variable	Description
mpg	Miles/(US) Gallon
cyl	Number of Cylinders
disp	Displacement (cu.in.)
hp	Gross horsepower
wt	Weight (lb/1000)
qsec	1/4 mile time
vs	V or Straight Configuration Engine
am	Transmission Type (auto/manual)

Load the Data in StatCrunch

A Tab Separated File (.tsv) containing the data can be found here:

http://stat.wvu.edu/~draffle/111/week1/lab1/mtcars2.tsv

1. In StatCrunch, use the menus `Data` → `Load` → `From File` → `On the web`.
2. Copy and Paste the address given above into `WWW Address`
3. Click `Load File` at the bottom of the page.

Complete the following problems on a separate piece of paper.

Hand in your answers at the end of lab – make sure to include you name and the lab number. Instructions are included for making graphs and finding the necessary statistics for each problem.

Each question will be worth two points, and you can receive partial credit for incorrect answers if your process was correct.

Write your final answer as a sentence and include all steps you used to get there, otherwise you will receive partial credit.

When sketching scatterplots, you do not need to include every point, just make sure you show the overall pattern and any outliers, if they exist. Keep in mind that standard notation for scatterplots is Y vs. X.

For all problems, you can find your answers following the steps:

Correlation:

- `Stat` → `Summary Stats` → `Correlation`

Regression:

- `Stat` → `Regression` → `Simple Linear`

- Note that this output also includes correlation information and a scatterplot with fitted line
- The scatterplot with fitted line can be found by clicking the arrow at the bottom of the output window.

1. Make a scatterplot of quarter mile time (`qsec`) vs. weight (`wt`). Sketch the graph.
2. Does there appear to be a strong linear relationship between `qsec` and `wt`? Include any relevant statistics.
3. Fit the regression line that that predicts engine size (`disp`) with weight (`wt`). Report the line equation and interpret the slope and intercept in the context of the data set; make sure to include units in your description. (keep in mind that weight is measured in *thousands* of pounds).
4. Report the R^2 statistic for this model. What does it tell us about the relationship?
5. Make a scatterplot of `mpg` vs. `hp`. Describe the relationship, making sure to include any relevant statistics.
6. Fit the Least Squares Regression line for predicting fuel efficiency `mpg` from horsepower `hp`. Report the line equation and interpret the slope and intercept in the context of the data set; make sure to include units in your description.
7. Find the expected fuel efficiency for the Maserati Bora based just on its horsepower (show your work). Describe how well the prediction matches with the actual value using the residual.
8. Fit a linear model which predicts quarter mile time (`qsec`) from horsepower (`hp`). Report the line equation and interpret the slope and intercept in the context of the data set; make sure to include units in your description.
9. What percent of the variability in quarter mile time is accounted for by horsepower?
10. Which car has the highest residual for this data set (you can get information about a point by hovering over it with your mouse)? What is the residual?