**Assignment #4: Statistical Inference in Linear Regression (50 points)**

This assignment will be made available in both pdf and Microsoft docx format. Answers should be typed into the docx file, saved, and converted into pdf format for submission. **Color your answers in green so that they can be easily distinguished from the questions themselves.**

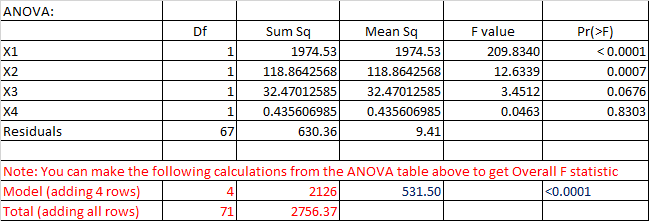
**Throughout this assignment keep all decimals to four places, i.e. X.xxxx.**

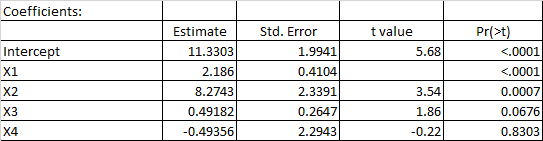
**Any computations that involve “the log function”, denoted by log(x), are always meant to mean the natural log function (which will show as ln() on a calculator). The only time that you should ever use a log function other than the natural logarithm is if you are given a specific base.**

In this assignment we will review model output from R and perform the computations related to statistical inference for linear regression. By performing this computations we are ensuring that we understand how the numbers in this R output are computed. **Students are expected to show all work in their computations. A good practice is to write down the generic formula for any computation and then fill in the values need for the computation from the problem statement.**

**Model 1:** Let’s consider the following R output for a regression model which we will refer to as Model 1. (Note 1: In the ANOVA table, I have added 2 rows – (1) Model DF and Model SS - which is the sum of the rows corresponding to all the 4 variables (2) Total DF and Total SS - which is the sum of all the rows;

Note 2: The F test corresponding to the Model denotes the overall significance test. In R output, you will see that at the bottom of the Coefficients table)









1. (5 points) How many observations are in the sample data?

Since the F-statistic is calculated “on 4 and 67 DF” there are n-1 = 4+67 observations.

n = 72

1. (5 points) Write out the null and alternate hypotheses for the t-test for Beta1.

Null Hypothesis: β1 = 0

Alternative Hypothesis: β1 != 0

1. (5 points) Compute the t- statistic for Beta1.

MSE = SSE/ N-t , MSTR = SSTR/t-1, T-statistic = t-value = MSE/MSTR = 2.186.0.4104=5.3265

1. (5 points) Compute the R-Squared value for Model 1, using ANOVA.

R-Sq= 1-(SSE/SST)=1-(630.36/2756.37)=0.7713

1. (5 points) Compute the Adjusted R-Squared value for Model 1.

R^2 adjusted = 1 – (1-R^2)(N-1)/ (N – p – 1) = 1 – (1 - .7713)(72-1)/ (72-4-1) =.7576

N = Total sample size

P = Number of predictors

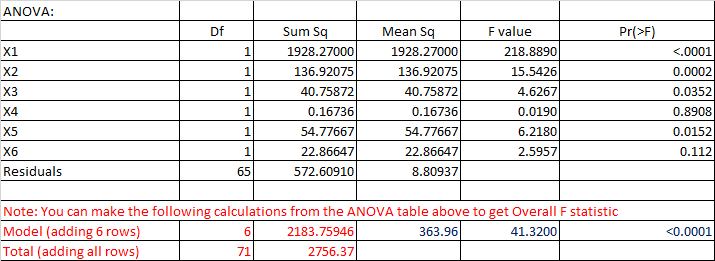
1. (5 points) Write out the null and alternate hypotheses for the Overall F-test.

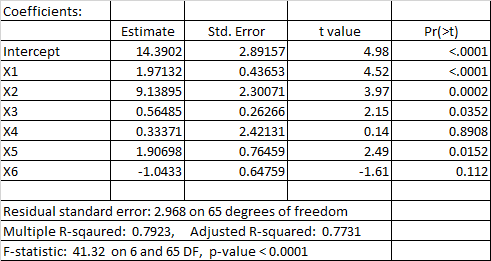
Null Hypothesis: β1 = β2 = β3 = β4 = 0

Alternative Hypothesis: At least one β != 0

1. (5 points) Compute the F-statistic for the Overall F-test.

**Model 2:** Now let’s consider the following R output for an alternate regression model which we will refer to as Model 2.







1. (5 points) Now let’s consider Model 1 and Model 2 as a pair of models. Does Model 1 nest Model 2 or does Model 2 nest Model 1? Explain.
2. (5 points) Write out the null and alternate hypotheses for a nested F-test using Model 1 and Model 2.
3. (5 points) Compute the F-statistic for a nested F-test using Model 1 and Model 2.

**Here are some additional questions to help you understand other parts of inference.**

1. (0 points) Compute the AIC values for both Model 1 and Model 2.
2. (0 points) Compute the BIC values for both Model 1 and Model 2.

1. (0 points) Compute the Mallow’s Cp values for both Model 1 and Model 2.
2. (0 points) Verify the t-statistics for the remaining coefficients in Model 1.
3. (0 points) Verify the Mean Square values for Model 1 and Model 2.
4. (0 points) Verify the Root MSE values for Model 1 and Model 2.