Assignment #4: Problem Set for Ordinary Least Squares 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 into Blackboard. **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.

Model 1: Let's consider the regression model, which we will refer to as Model 1, given by

Y = 10,000 + 150*X1 + 25*X1^2 + 60*X2 (M1).

(1) (2 points) Is this a "linear" regression model, why or why not? Yes, because each parameter in the model enters linearly.

(2) (4 points) How do we interpret this model? Hint: how does a one unit change in X1 or X2 affect the estimated value for Y? State the interpretation for both X1 and X2.

This model is essentially a quadratic function and in its current form is not linear. A one unit change in X1 equals at least a 175 unit change for y, but every positive unit change will increase Y larger than the prior unit change, thus making it quadratic. A one unit change in X2 will change Y by 60 units, thus this variable is linear. The interpretation for X1 is such that Y increases quadratically as X increase by one unit, and y increases linearly as X2 increases.

(3) Consider the Analysis of Variance (ANOVA) table from fitting this model to a sample of 50 observations.

Analysis of Variance Table for Fitted Regression Model		
Sum of Squares from the Regression	SSR	750
Sum of Squares for the Error	SSE	250
Total Sum of Squares	SST	1000

a. (4 points) Compute the R-squared and adjusted R-squared values for this regression model. p. 46, 69

b. (2 points) Compute the estimate of the Mean Square Error (MSE). p. 303

c. (4 points) State the hypothesis and compute the test statistic for the overall F-test. p. 74

H0 = Reduced Model: Y = B0 + E

MSR/MSE

H1 = Full Model; Y = B0 + B1X1....

R-Squared = 1 - SSE/SST = = 1 - (250/1000) = .75 R-Squared || Adj R-Squ = 1 - (n-1)/ (n-p-1) * (1-R-Squ) = 1 - (50-1)/50-3-1) * (1-.75)

1 - (49/46) * (.25) = 1 - (1.0652) * (.25)

1- .2663 = .7337 = Adj R-Squ

MSE = SSE / n-p = 250/ 50-3 = 250/ 47 = 5.3191 = MSE

F = MSR/MSE || MSR= SSR/p = 750/3=250 || MSE= SSE/n-p-1= 250/46 = 5.4348|| 250/5.4348 = 45.9999 = F

H0 = Betas = 0
H1 = Betas Not 0
FCritical= K-1 Nu
n-k = Denom
Check Page 376 to verify. Fcritcal is around 4.08, reject null.

Model 2: Now let's consider an alternate regression model, which we will refer to as Model 2, given by

Y = 9,750 + 145*X1 + 75*X2 (M2).

(4) Consider the ANOVA table from fitting this model to the same sample of 50 observations that we used to fit M1.

Analysis of Variance Table for Fitted Regression Model		
Sum of Squares from the Regression	SSR	725
Sum of Squares for the Error	SSE	275
Total Sum of Squares	SST	1000

a. (4 points) Compute the R-squared and adjusted R-squared values for this regression model.

b. (2 points) Compute the estimate of the Mean Square Error (MSE).

c. (4 points) State the hypothesis and compute the test statistic for the overall F-test.

(5) Now let's consider M1 and M2 as a pair of models. We want to decide which model we should use as our final model. Here are some concepts to help us make that decision.

a. (2 points) What is the definition of a nested model? See Box below
b. (2 points) Does M1 nest M2 or does M2 nest M1?
c. (2 points) Based on any of the metrics or statistics that you have computed in Questions #3 and #4, which model should we prefer (M1 or M2) and why?
d. (10 points) Perform a F-test for nested models and determine if we should choose M1 or M2. State the hypothesis that we will be testing, compute the test statistic, and test the statistical significance using a critical value for alpha=0.05 from Table A.4 on page 358 in *Regression Analysis By Example*.

(6) In Ordinary Least Squares (OLS) Regression we assume that the response variable is normally distributed with mean XB and variance sigma^2, i.e. Y ~ N(XB, sigma^2).

a. (2 points) How do we estimate sigma^2? p.61 97
b. (6 points) What are two diagnostic checks of model goodness-of-fit that we perform in order to assess this distributional assumption? http://www.lexjansen.com/pharmasug/2004/posters/po04.pdf

(5) a. A model is nested if it can be obtained from a larger model as a special case, 71.
b. M2 Nests M1. M1 has one additional term.

Y = 10,000 + 150*X1 + 25*X1^2 + 60*X2 - M1

Y = 9,750 + 145*X1 + 75*X2 - M2

c.