

## 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.**

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.

When stating the null and alternate hypotheses in any statistical test in PREDICT 410, we should always state these hypotheses in terms of the model parameters, i.e. the model coefficients denoted by the betas.

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

$$Y = 10,000 + 150 \cdot X_1 + 25 \cdot X_1^2 + 60 \cdot X_2 \quad (M1).$$

- (1) (2 points) Is this a “linear” regression model, why or why not?
- (2) (4 points) How do we interpret this model? Hint: how does a one unit change in  $X_1$  or  $X_2$  affect the estimated value for  $Y$ ? State the interpretation for both  $X_1$  and  $X_2$ .
- (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.
- b. (2 points) Compute the estimate of the Mean Square Error (MSE).
- c. (4 points) Perform the overall F-test for this model, i.e. state the null and alternate hypothesis, compute the test statistic for the overall F-test, and make a decision to “reject “ or “fail to reject” the null hypothesis. Test the statistical significance of the overall F-test using a critical value for  $\alpha=0.05$  from Table A.4 on page 376 in *Regression Analysis By Example*.

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

$$Y = 9,750 + 145 \cdot X_1 + 75 \cdot X_2 \quad (M2).$$

- (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

- (4 points) Compute the R-squared and adjusted R-squared values for this regression model.
  - (2 points) Compute the estimate of the Mean Square Error (MSE).
  - (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.
- (2 points) What is the definition of a nested model?
  - (2 points) Does M1 nest M2 or does M2 nest M1?
  - (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?
  - (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 376 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.  $Y \sim N(XB, \sigma^2)$ .
- (2 points) How do we estimate  $\sigma^2$ ?
  - (6 points) What are two diagnostic checks of model goodness-of-fit that we perform in order to assess this distributional assumption?