**STAT 501 – Homework 6 – Spring 2015 – due Feb 22**

**Instructions**: Use Word to type your answers within this document. Then, submit your answers in the appropriate dropbox in ANGEL by the due date. The point distribution is located next to each question. If there are multiple parts, then the points are divided equally over the subparts.

1. (**10 points**) Suppose Y has 4 possible predictors: X1, X2, X3 and X4.
   1. If SSTO = 11489, SSE(X1, X2, X3) = 335, SSE(X1, X2, X4) = 990, and SSR(X1, X2) = 10493, what is the value of SSR(X4| X1, X2)?
   2. If SSR(X1) = 9140 and SSR(X2) = 2177, under what condition can you find SSR(X1, X2)? What would be the value? *[Hint: Take a look at Example 1 on Pastry Sweetness in Section 5.5 Further Examples and pay particular attention to point number 2 about R2. What does it say about the sums of squares if the SLR R2 values sum to the MLR R2 value? And what is special about the predictors in this example?]*
2. **(10 + 6x5 = 40 points)**

a. Fill in the blanks in the following tables. The column labeled “Seq SS” represents Minitab “sequential sums of squares” (measures the reduction in the SS when a term is added to a model that contains only the terms before it), while the column labeled “Adj SS” represents Minitab “adjusted sums of squares” (measures the reduction in the SS for each term relative to a model that contains all of the remaining terms).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | df | Seq SS | Adj SS | F value  based on Adj SS | p-value  based on Adj SS |
| Regression | 3 | 100.866 |  | 35.14 | 0.000 |
| X1 | 1 | 67.444 | 33.031 | 34.52 | 0.000 |
| X2 | 1 | 3.883 |  |  |  |
| X3 | 1 |  |  | 30.88 | 0.000 |
| Error | 93 |  |  | ---- | ------- |
| Total | 96 | 189.842 | 189.842 | ---- | -------- |

Coefficients

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Term | coef | SE coef | T-value | p-value |
| Constant | 0.58 | 1.24 | 0.45 | 0.652 |
| X1 | 0.34 | 0.058 | 5.88 | 0.000 |
| X2 | -0.01 | 0.0245 |  |  |
| X3 | 0.06 | 0.0103 | 5.56 | 0.000 |

1. Calculate SSR(X3|X1), that is the sequential sum of squares obtained by adding X3 to a model already containing only the predictor X1. Show your work.
2. Explain in words what is measured by the quantity calculated in the previous part.
3. Discuss the conceptual difference between the sequential sum of squares (Seq SS) and adjusted sum of squares (Adj SS) in terms of the predictor X2. For this data, what are the numerical values of these sums of squares for the predictor X2?
4. Calculate the value of an F-statistic for testing H0: β2= β3 = 0 within the model Yi = β0 + β1 Xi,1 + β2 Xi,2 + β3 Xi,3 + εi. It is not necessary to carry out the test – just calculate the value of F. Show your work.
5. Calculate the value of the coefficient of partial determination R2Y,2|1.
6. Write a sentence that interprets the value calculated in the previous part.
7. **(25 points)**

**Regression Analysis: weight versus trunk, width, length**

Analysis of Variance

Source DF Seq SS Seq MS F-Value P-Value

Regression 3 8208.9 2736.31 39.38 0.000

trunk 1 5453.4 5453.42 78.48 0.000

width 1 2551.7 2551.70 36.72 0.000

length 1 203.8 203.82 2.93 0.090

Error 93 6462.6 69.49

Total 96 14671.5

Coefficients

Term Coef SE Coef T-Value P-Value

Constant -15.71 4.60 - 3.42 0.001

trunk 2.638 0.522 5.05 0.000

width 0.5108 0.0842 6.07 0.000

length 0.0106 0.00620 1.71 0.090

Regression Equation

weight = -15.71 + 2.638 trunk + 0.5108 width + 0.01062 length

**Regression Analysis: weight versus length**

Analysis of Variance

Source DF Seq SS Seq MS F-Value P-Value

Regression 1 2103 2103.5 15.90 0.000

length 1 2103 2103.5 15.90 0.000

Error 95 12568 132.3

Lack-of-Fit 84 11363 135.3 1.23 0.370

Pure Error 11 1205 109.6

Total 96 14672

Coefficients

Term Coef SE Coef T-Value P-Value VIF

Constant 14.73 1.93 7.61 0.000

length 0.03046 0.00764 3.99 0.000 1.00

Regression Equation

weight = 14.73 + 0.03046 length

Let Y = weight, X1 = trunk, X2 = width, X3 = length

* 1. Test H0: β2 = 0 vs. H0: β2 ≠ 0 in the model E(Y) = β0 + β1 X1 + β2 X2 + β3 X3? What is the value of the test statistic you will use? What is the p-value and conclusion?
  2. Calculate the value of an F-statistic for testing H0: β1 = β2 = 0 in the model E(Y) = β0 + β1 X1 + β2 X2 + β3 X3, where the X variables are defined in the order given above.
  3. Propose a model to test H0 in part b by using a partial F-statistic. Remember to specify the order of the predictors in your model.
  4. Write an interpretation of the significance of the test results given for the predictor variable *length* within the three-predictor model (the first output). Be careful – you might also look at what happens in the one-predictor model in which *length* is the only predictor.

1. (**25 points**) Use the “Infection Risk” data set in the Lesson 6 folder. Preliminary data analyses have revealed that the variable *Y* = *InfctRsk* could be related to the variables *X1* = *Stay*, *X2* = *Culture*, *X3* = *Xray*, *X4* = *Beds*, *X5* = *Census*, and *X6* = *Nurses.*
   1. Fit a multiple linear regression model that relates *InfctRsk* to the predictor variables *X1-X6*. Perform a hypothesis test at significance level 0.05 to determine if at least one of the predictors in this model is useful in predicting *Y*. State your null and alternative hypotheses in terms of the slope parameters, the test statistic value with calculations shown, the decision rule and the conclusion.
   2. Perform a hypothesis test to determine if the predictor variables X5 = Census and X6 = Nurses can be deleted from the model while retaining the four remaining variables X1 = Stay, X2 = Culture, X3 = Xray, and X4 = Beds. Again, state your null and alternative hypotheses in terms of slope parameters, show your work in calculating the test statistic, state the decision rule and the conclusion.
   3. Perform a hypothesis test to determine if *X4* = *Beds* can be dropped from a model with the four predictors, *X1* = *Stay*, *X2* = *Culture*, *X3* = *Xray*, and *X4* = *Beds,* by using:
      * 1. a t-statistic
        2. an F-statistic

(State your null and alternative hypotheses in terms of the slope parameters, the test statistic values, the decision rules and the conclusions.)

Is there any relationship between the two test statistics in (i) and (ii) above?

* 1. Write down the fitted regression equation based on your conclusion in part (c).