Statistics 4590
Assignment #5: March 23, 2010
Due in class: April 6, 2010

All problem numbers are from the textbook *The Statistical Sleuth, 2nd edition* by Ramsey and Schafer.

Make sure to submit the relevant output with your assignment, and clearly label material on this output. Remember the rules on assignment #1 and the course syllabus for what affects your grade.

1. Consider the multiple linear regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + e$$

   Suppose we want to see if x_1 and x_3 can be dropped from the model. Explain why it is appropriate to use the partial F-test for $H_o : \beta_1 = \beta_3 = 0$ rather than conducting two t-tests for $H_o : \beta_1 = 0$ and $H_o : \beta_3 = 0$.

2. Refer to #15, page 261–262 to complete the following. You used this data in Assignment #4.

   (a) Fit the linear regression model of corn yield on rain. Comment on the residual plots provided by R for this model.

   (b) Fit the multiple regression model of corn yield on rain, rain^2 and year. Use this model to estimate the corn yield if 10.8 inches of rain fell in 1928.

   (c) For your model in (b), can the terms for year and rain^2 be dropped?

3. Problem #12, page 604. Complete all parts of the question, but you are not required to give a description of the data, or address issues about model assumptions or scope of inference.

   The data is on the course website.

4. What is the difference between a log-linear regression model and a linear regression model after a log transformation of the response variable?

5. Problem #23, p. 666. Here you will look at another way to analyze the *Challenger* data that we have studied in class. The data is available on the course website.

   You do not have to give any description of the data, or use any plots to support your analysis. Fit an appropriate model, and conduct an appropriate test, to answer the question of interest. You can exclude any discussion of scope of inference.