ST 512 HW 6

All Problems: Due Tuesday, March 22^{nd} in *lecture* at the beginning of class.

Note: Remember to include your ***code*** with your typed problems submitted in class.

12.13 (10 pts)

(a) (2 pts) You can do this in SAS or by hand.

(b) (4 pts) Specifically, would you use a linear, quadratic, or cubic function? Justify your response.

(c) (4 pts) When selecting your model (2 points) be sure to give at least two statistics to support your justification (2 points).

12.30 (10 pts)

(a) (2 pts) Use R^2, not Adjusted R^2

(b) (4 pts) Ignoring the book's part (b), compute the following value (2 points).

$$\frac{R^2}{1-R^2}\frac{df_{error}}{df_{regression}}$$

How does this value relate to the output that was provided (2 points).

(c) (4 pts) Use a conceptual explanation of your computation in part (b) to ***help*** support your answer.

12.31 (5 pts) 2 points for (a) and 3 points for (b)

NBP# 1 (12 pts) Non-Book Problem: Consider the following models:

1. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$
2. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$
3. $y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \epsilon$
4. $y = \beta_0 + \beta_1 x_1 + \epsilon$
5. $y = \beta_0 + \beta_1 x_1 + \beta_2 (x_2 + x_3) + \epsilon$
6. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2 + \epsilon$

(a) (2 pts) Of the first 5 models, what would be considered the full model?

(b) (5 pts) Which of the six models are nested in model 1?

(c) (5 pts) Which of the models are nested in model 6?

Non-Book SAS Problem:

1. (43 pts) Steel.dat (see web site for data) contains data on measurements of the quality of steel. The response y is a measure of quality, x_1 and x_2 are the iron and carbon ratings - two variables used in the making of the steel. Fit the following models in SAS using `PROC GLM`.

   Model 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \epsilon$

   Model 2: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \epsilon$

   Model 3: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$

   Model 4: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$

   (a) (5 pts) Look at the Sums of Squares from each ANOVA table, what is the value of the extra regression sum of squares for adding β_3 (interaction) to the model quadratic in both x_1 and x_2 (2 points)? i.e. Find $R(\beta_3|\beta_1, \beta_2, \beta_4, \beta_5)$. Interpret this quantity. (3 points)

   (b) (5 pts) What is the value of the extra regression sum of squares for adding in β_4 and β_5 to the interaction model (2 points)? i.e. Find $R(\beta_4, \beta_5|\beta_1, \beta_2, \beta_3)$? Interpret this quantity. (3 points)

   (c) (5 pts) Are any of the models nested in one another? If so, what are the restrictions necessary to obtain the nested model in each case?

   (d) (8 pts) Test the null hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ vs the alternative that at least one is not zero - give your test statistic and p-value (2 points) as well as your decision and conclusion (3 points). What do you notice about the p-values for testing each parameter of the model individually (2 points)? What problem may be causing this (1 points)?

   (e) (5 pts) Test the null hypothesis $H_0 : \beta_4 = \beta_5 = 0$ vs the alternative that at least one is not zero.

   (f) (5 pts) Test the null hypothesis $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$ vs the alternative that at least one is not zero.

   (g) (2 pts) Use the forward selection method for model selection with an entry threshold of 0.25. Report the final model.

   (h) (2 pts) What do you conclude about the adequacy of the model assumptions for your final model from the forward selection?

   (i) (2 pts) Use the backward selection method for model selection with an exit threshold of 0.2. Report the final model.

   (j) (4 pts) The models selected by the two procedures are very different. However, the correlation between x_1 and x_2 is about -0.83. Using this fact, discuss why the models are more similar than they originally appear. Which model would you prefer for interpretation purposes?