STAT 512

Homework 1 (9 pts.) due Jan. 17

A reminder – Please do not hand in any unlabeled or unedited SAS output. Include in your write-up only those results that are necessary to present a complete solution (what you want the grader to grade). In particular, questions must be answered in order (including graphs), and all graphs must be fully labeled (main title should include the question number, and all axes should be labeled). Don't forget to put all necessary information (see course policies) on the first page. Include the SAS input for all questions at the very end of your homework; this could be important even though it won't be graded. You will often be asked to continue problems on successive homework assignments so save all your SAS code.

- 1. (4 pts.) A regression analysis relating test scores (Y) to training hours (X) produced the following fitted equation: $\hat{Y} = 20 0.6X$. Be sure to show your work for all parts in this question.
 - a) What is the fitted value of the response variable corresponding to X = 4?
 - b) What is the residual corresponding to the data point with X = 5 and Y = 21?
 - c) If X decreases by 3 units, how does \hat{Y} change?
 - d) An additional test score is to obtained for a new observation at X = 7. Would the test score for the new observation necessarily be 15.8? Explain.
 - e) The error of sums of squares (SSE) for this model was found to be 15.3. If there were n = 20 observations, provide the best estimate for σ^2 .
 - f) Rewrite the regression equation in terms of X* where X* is training time measured in minutes. Show that your answer makes sense, i.e., gives the same predictions as the original equation (one example is sufficient).
- 2. (0.5 pts.) Explain the difference between the following two equations:

 $\hat{\mathbf{Y}} = \mathbf{b}_0 + \mathbf{b}_1 \mathbf{X}$

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- 3. (4.5 pts.) For this problem, use the "grade point average" data described in KNNL Problem 1.19. The data are on the disk that accompanies the test, online and on the homework page of the class web site (CH01PR19.DAT). make sure you understand which column is X and which is Y and read the data accordingly. See Topic 1 or nknw060.sas for an example of how to read in a data file.
 - a) Plot the data using proc gplot. Include a smoothed function on the plot by preceding the plot statement with "SYMBOL1 i = smNN" where NN is a number between 1 and 99. Note that larger numbers cause greater smoothing. make sure to indicate the smoothing number in the title of the plot. Is the relationship approximately linear?
 - b) Run a linear regression to predict GPA based on the entrance exam. Give a point estimate and a 95% confidence interval for the slope and intercept and interpret each of these in words. (*Point estimate* is another word for parameter estimate.) Clearly indicate which data is for the slope and which is for the intercept. Be sure to include the appropriate output in your assignment. Why would someone be interested if 0 is included in the interval for β_1 ?
 - c) Would it be reasonable to consider inference on the intercept for this problem? Please provide justification for your answer.
- 4. Go over Homework 0, Homework 0A and the material available on the SAS web page. (Not to be turned in.)