## Homework 7 (19 pts.) due March 7

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. (9.5 pts.) This problem is a continuation of the Commercial Properties question from Homeworks 4, 5 and 6 (Problem 6.18, CH06PR18.DAT).
  - (a) Copy down your answer from Problem 4 in Homework 6. That is, please list the 3 best regression models. There will be no credit for this part, but the information will be used in the rest of the problem.
  - (b) Find the best subset of variables according to the forward stepwise regression. How does this subset compare with the results of part a).
  - (c) Repeat part (b) for forward selection and backward elimination. How do your results compare with parts (a) and (b).
  - (d) Which model do you think is the best (remember part a)? Why? You will have to perform additional calculations to answer this question.
  - (e) Check the assumptions of the "best" model using all of the usual plots. Explain in detail whether or not each assumption appears to be substantially violated.
- 2. (9.5 pts.) This question uses the situation presented in Cosmetic sales, Problem 10.13 in the book (CH10PR13.txt: Note: this data set has been modified from the one in the textbook). For parts c, d, e, f, g ii, do not include in your answer the values for all cases. Use plots and verbal summaries instead. You may include values for a few selected cases if you wish.
  - (a) Test the regression relation between sales and the three predictor variables. State the hypotheses, test statistic and degrees of freedom, the p-value, the conclusion in words.
  - (b) Determine whether the linear regression model is appropriate by using the "usual" plots (scatterplot, residual plots, histogram/QQ plot). Explain in detail whether or not each assumption appears to be substantially violated.
  - (c) Prepare a partial regression plot for each of the predictor variables. Do your plots suggest that the regression relationship in the fitted regression function are inappropriate for any of the predictor variables? Explain.
  - (d) Are there any outlying Y observations? (Hint: studentized deleted residuals and remember to use the Bonferroni correction.)
  - (e) Are there any outlying X observations? (Hint: diagonal elements of the hat matrix)