Stat 401B Project 2 (Multiple Regression) Fall 2008

The second project (worth 40 points, 8% of your final grade) has three objectives:

1. Demonstrating you can use the JMP Stepwise platform.
2. Building a model that is "best"
   - the model is statistically useful at the 5% level.
   - each variable in the model is statistically significant at the 5% level, given the other variables in the model.
   - among those models that meet the two criteria above, the one with the highest R^2.
3. Analyzing residuals including identifying outliers, high leverage and influential values.

There are four data sets that you can choose from. A brief description of each of the data sets along with response and explanatory variables can be found in the handout, Data Sets for Model Building Project. This can be an individual or group project. The result of the project is a thorough, but concise, professional quality technical report of no more than 5 typed pages. The report is to be handed in, or emailed, to your instructor by 5 pm on December 10, 2008. You can turn in your report sooner if you wish. Your report should contain:

1. Demonstrating you can use the JMP Stepwise platform.
   - Provide a description of the "best" and second "best" single variable models. Give the variable, value of R^2, RMSE, and the test that the variable is statistically significant.
   - Provide a description of the final models fit by JMP using Forward, Backward and Mixed (Forward) selection procedures (using the default settings for the Prob to Enter and Prob to Leave). Give the variables in the final model, value of R^2, RMSE, and indicate whether each variable in the model is statistically significant or not by reporting the appropriate test statistic and P-value.
   - Discussing how the final models fit by JMP using Forward, Backward and Mixed (Forward) selection procedures change when the default settings are set so that both Prob to Enter and Prob to Leave are 0.05.
2. Building a model that is "best."
   - Provide a description of your final fitted model with supporting evidence that it is the "best" model. This evidence should include a description of the process you used to find your "best" model.
3. Analyzing residuals including identifying outliers, high leverage and influential values.
   - Provide an analysis of residuals for your "best" model. This analysis should include appropriate plots and tests for outliers, high leverage points and influential points. You should also comment on whether the conditions for regression analysis are satisfied.

Simply attaching a ream of computer printout in the appendix and expecting the instructor to find what is important is not acceptable. I will deduct points for including

JMP output that does not address the questions asked. The main body of the report should include only the end products of any statistical calculations. It is appropriate to include plots and important summary values within the body of the report.

The following due dates will be used to assure satisfactory progress on the project, you may complete each progress step before the date listed.

- You will have time to work on the project during lab on November 18, December 2 and December 9.
- November 18 – Individuals or groups indicate which of the four data sets they have chosen for their project. This should be done by emailing me at wrstephe@iastate.edu
- December 10 – The final reports are due. These are to be typed (or word processed) on plain white paper and should not exceed 5 pages (12 pt font or larger) in length. A title page, JMP output, etc. all count toward the 5 page limit.