STAT 516 - Spring 2001 - Homework 3
Due: Monday, February 19th, 2001

1) *(5 points)* A continuation of Problem 3, on Homework 2. For this question, you may assume the assumptions of the regression hold. **For each problem, identify which statistic you used to answer the question.**

a) Which two states' independent variables are most different from those of the other states? Most similar?

b) For which, if any, of the predictor variables is multicollinearity a concern?

c) What subset of independent variables gives the "best" model for this problem?

d) How would you respond to someone who claimed that the variables not included in the model in part c did not affect the average SAT score of the states?

2) *(5 points)* The web site contains data from the US Census Bureau concerning the populations of the 50 states in 1990 and 1999.

a) Fit a linear regression for predicting the 1999 population from the 1990 population. For computer output you need only give the regression equation and the two residual plots. Tell which plot shows you the variances are increasing along with the 1990 population, and how the plot indicates this.

b) Transform the dependent variable to correct for the problem with the variances in the regression in part a. Again, for output give the regression equation and the two residual plots.

c) Transform the independent variable to deal with the problems found in the residual plots in part b. Again, in your output give the regression equation and the two residual plots. Tell why you chose this transformation.

d) Write the estimated regression equation from part c in the form: $y = \ldots$ instead of the form: transformed $y = \ldots$ (see, for example, page 151 in the text).