Problem 1

The data can be found at

http://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/KutnerData/Chapter%20%206%20Data%20Sets/CH06PR05.txt

Kutner 10.9

Refer to **Brand preference** Problem 6.5.

a. Obtain the studentized deleted residuals and identify any outlying Y observations. Are any of the Y observations outlying according to the rule of thumb stated in the lecture?

b. Are any of the observations outlying with regard to their X values according to the rule of thumb stated in the chapter?

c. The largest absolute studentized deleted residual is for case 14. Obtain the DFFITS and Cook's distance values for this case to assess the influence of this case. What do you conclude?

d. Fit two regressions with and without influential points (if any) and compare the coefficients/standard errors of all variables. Any major changes in directionality and/or magnitude?

Problem 2

Kutner 7.38

Refer to the "hospital.csv" data set. For predicting the average length of stay of patients in a hospital (Y), it has been decided to include age (X_1) and infection risk (X_2) as predictor variables. The question now is whether an additional predictor variable would be helpful in the model and, if so, which variable would be most helpful. Assume that a first-order multiple regression model is appropriate.

a. Create the correlation matrix for all 6 predictors and comment.

b. For each of the following variables, calculate the coefficient of partial determination given that X_1 and X_2 are included in the model: routine culturing ratio (X_3), average daily census (X_4), number of nurses (X_5), and available facilities and services (X_6).

c. On the basis of the results in part (b), which of the four additional predictor variables is best? Is the extra sum of squares associated with this variable larger than those for the other three variables?

d. Use ANOVA to compare the models with X_1+X_2 vs model in (c). Use VIF to check for multicollinearity among these three variables and comment.

e. Model building: use automatic procedures (backward, forward, stepwise) and criterion based procedures (Cp, adj R, AIC) to find the 'best' predictive model. What is the final result? How do the assumptions hold?