**Last Team Assignment, DAPT 621 (20 points)**

**Submit by 9 AM Monday, December 8, 2014 in an Excel file sent by e-mail to dapt.stat@gmail.com**

This is a group assignment submitted electronically in an Excel spreadsheet by 9 AM, Monday, December 8. There are two parts to this assignment.

1. Evaluate the model/procedure your team recommended for Old Faithful assignment one.

2. Build the best model for the data found in the Excel file

[**SENIC Data (pg. 1348 of *Applied Linear Stat. Models* by Kutner, Nachtsheim, Neter & Li 5th edition**](http://www.people.vcu.edu/~randrews/632/KNNL5th_hw/Senic_C01.xls)

on the Lecture Materials webpage at <http://www.people.vcu.edu/~randrews/632/KKNR5th_lect/KKNR5th_lect.html>.

1. To evaluate your recommended procedure for DAPT tours obtain the data for Old Faithful eruptions for the month of October. On tab labelled 1, list each day for the month and indicate the eruption time that is in (could have 2 and list both if that is the case) or closest to your recommended 90-minute interval. Then as a summary record the percentage of the days when your procedure chose a 90 minute time period that allowed the guests to see Old Faithful erupt.

2. Create a tab in Excel for each part below for the SENIC data. For this modeling process use Length\_stay as the response or dependent variable.

2A. Indicate if you found any data points that you think are not indicative for building a regression model for this dependent variable. If you choose to exclude any points give a reason for excluding each of the points you excluded.

2B Give the regression output for the model that you think is the best model. Make sure to include R-square, p-value for the overall significance of the model, p-values for testing the significance of each variable in your model and the VIF values for each variable in your model.

2C Tell which data point (use the ID number in the original data set to identify the point) that has the most influence on the model as measured by Cook’s D measure of influence.

2D Find the mean of each of the variables in the data set you used to create your model then use your model to

1. Predict the value for the length of stay (Length\_ stay) for a hospital that has variable values equal to the mean for each variable.
2. Find the 95% confidence interval for a new individual value of Y (Length\_ stay) for a hospital **in the west with no medical school affiliation that has variable values equal to the mean for each quantitative variable**.

2E The above confidence interval assumes that all assumptions were met, including Homoscedasticity and Normality. If the assumptions are not met then the confidence interval could be off the mark. Tell whether you observed anything that would cause you to believe that this interval would be off the mark in some way or not. Provide a reason or reasons for your conclusion.