**ISYE 4140 Statistical Analysis Fall 2012**

**Project #5: Regression Analysis**

Work in the assigned groups listed below. You may, however, consult with other groups, provided you ultimately do all the work within your own group. (Example: You can accept a tip from another group to "use the *sample()* command", but then you have to write your own R code. )

Create a report in Word by copy/pasting results from R. Hand in your hardcopy report at the start of class 23, Friday 16 November 2012. Recall that late work will be penalized per the syllabus. Start with a cover page listing all your names and signatures; your signatures signify that you have each made a significant contribution to the report.

This project asks you to make a regression model that predicts student achievement on the MCAS test from community characteristics. The data are in the file *MCAS.Rdata* on the course web site. Operationalize the notion of "student achievement" in a new variable called Score=English.score+Math.score. To better simulate a "first look" analysis, do not use Community Type or demog as predictors, since these were developed by an expert after examining the data.

Make your model using a training set that is a random selection of exactly 150 observations. Test your model on the test set comprising the remaining 75 observations.

a) [10 points] Create the training and test datasets. List the R code you used to do this. Also report the structure of both datasets using the *str()* command.

b) [5 points] For the training data, display the scatterplot matrix relating all potential predictors to each other and to the response (Score).

c) [10 points] Using the training data, build a regression model to predict Score. Report your model by including the results of the *summary()* command. Provide your assessment of the summary, indicating both positive and negative aspects.

d) [10 points] Display the residual plots for your regression model created by the *plot()* command. Assess each of the plots.

e) [10 points] Validate your model's predictive power by applying it to the testing data. Present appropriate plots and summary statistics of the prediction errors. Provide your assessment of the results, indicating both positive and negative aspects. (Remember, you are not fitting a new model to the testing data; you are applying the model developed on the training data to the testing data.)

(f) [0 points, but required] List your R code.

Teams for Project 5: Regression Analysis

