**Ch. 3 homework** – due Tuesday, October 1, 2013

As a reminder, you need to provide a word document with your answers, along with appropriate output and graphs from the R program that helps to answer each question, as well as your code. Please let me know if you have any questions.

**1.) Use data file Insolvency2.**

a. Produce a table of summary statistics for all variables. Comment on what you feel is most noticeable.

b. Produce a table of correlations. What two variables are most highly correlated with PREM\_LEV?

c. Produce a scatter plot matrix of the explanatory variables (ASSET87, ASSET88, ANN\_PREM, CUR\_LIAB, SURPLUS) and PREM\_LEV. Comment appropriately.

d. Fit a linear model of PREM\_LEV using the above explanatory variables. Summarize the fit of this model by citing the F p-value, the *R²* and the adjusted *R²a* and commenting on them appropriately*.*

e. Based on your part (d) model fit, is ASSET87 a statistically significant variable? To respond to this question, use a formal test of hypothesis. State your null and alternative hypotheses, use the t score as your decision making criterion and your decision-making rule at the significance level of alpha=0.05.

f. Based on your part (d) model fit, is ANN\_PREM a statistically significant variable? To respond to this question, use a formal test of hypothesis. State your null and alternative hypotheses, use the p-value as your decision making criterion and the significance level of alpha=0.05.

g. Using the model in part (d), interpret the y-intercept.

h. Note that the sign of the regression coefficient associated with ANN\_PREM while the correlation is negative. Make a scatterplot of PREM\_LEV and ANN\_PREM and comment on whether the slope of this plot corresponds to the regression coefficient or the correlation. Is this expected? Explain why or why not.

i. To get further insights into the relationship between PREM\_LEV and ANN\_PREM, produce an added variable plot and compute the corresponding partial correlation coefficient. What is the interpretation of this coefficient? Do this plot and this coefficient coincide? Explain.

**2.) Use data file HospitalCosts2.**

a. Produce summary statistics of the variables. Comment on TOTCHG.

b. Run a regression predicting TOTCHG using AGE and LOS. Using alpha = 0.01, comment on whether these variables are significant.

c. Using the model in part (b), interpret the coefficient of LOS.

d. Provide a boxplot of TOTCHG by the variable FEMALE. Which gender has higher expenses? Comment on this plot.

e. Provide a boxplot of TOTCHG by the variable RACE. Comment on this plot appropriately.

f. Take the log version of TOTCHG. Then fit a linear model of this logged variable using AGE and LOS. Compare the fit of this model to the model in part (b) using multiple goodness of fit statistics.