**Homework 3 – Ch. 4** – due Monday, October 27, 2014

As a reminder, you need to attach to an email your R code and output/answers for these problems. Please also bring a print out of the output/answers ready to be graded to class with your names on it.

**1.)**

a. Take a quantitative variable and turn it into a binary variable in some way. Build a model using only this new binary variable to predict what you feel is your reasonable Y variable.

a(i). Is your binary variable a significant predictor? How do you know?

b. Take a quantitative variable and turn it into a categorical variable with multiple categories in some way. Build a model using only this new categorical variable to predict what you feel is your reasonable Y variable.

b(i). Is your categorical variable a significant predictor? How do you know?

b(ii). Interpret the coefficient of the first term in your output here.

c. Produce a scatter plot matrix of at least several of your explanatory variables and your reasonable Y variable.

c(i). Which variable is most likely to need a squared term?

c(ii). Build a model with your Y variable, these explanatory variables, and your new squared term. Is your squared term significant? How do you know?

d. Build a model with your Y variable, the explanatory variables used above in part b, and an interaction term.

d(i). Is your interaction term significant? How do you know? State your null hypothesis, alternative hypothesis, t test statistic, and t critical value. Use a 5% level of significance.

e. Run a one-factor ANOVA model using an original variable present in your dataset to predict your Y variable.

e(i). Is this model significant? How do you know? State your null hypothesis, alternative hypothesis and all components of the decision-making rule. Use a 5% level of significance.

e(ii). Run Tukey to determine which groups are different from which other groups. Describe from your output how you know.

f. Evaluate the goodness of the fit of the 5 models you have built so far in parts a-e and compare them on their *s*, *R²,* *R²a,* and F p-value.Pick one model as your favorite and explain why you like it best.

g. Redo the model you built in part (c), dropping the squared X term and the original X term that you squared.

g(i). Perform a partial F-test and see whether the X and its squared term are jointly significant. State your null hypothesis, alternative hypothesis and all components of the decision-making rule. Use a 5% level of significance. Work this by hand and with code.

h. Redo part (d), with the log version of your Y variable.

h(i). Has taking the log of Y improved your model?

h(ii). Plug in reasonable numbers for each of your variables and compute a 95% confidence interval in log dollars.

h(iii). Convert the predicted y and this 95% confidence interval back to regular dollars.

h(iv). Interpret this interval on the dollar scale.