Homework one Due Feb. 7th.

Please submit a hard copy that includes the R scripts, abbreviated outputs and answers/explanations to the questions.

1. In this problem we will perform multiple regression on the Boston housing data. The data contains 506 records with 14 variables. The variable medv is the response variable. To assess the data use

   library(MASS)

   data(Boston)

   (a) First perform a multiple regression with all the variables, what can you say about the significance of the variables based on only the p-values. Next use the "step" function to perform backward selection using (1) the AIC criteria and (2) the BIC criteria then compare the results. (By default the step function in R performs variable selection based on AIC criteria. Read the documentation to find out how to do the selection using BIC criteria.)

   (b) Now make a histogram of the response variable (use hist()) to see if it is skewed. Using log(medv) as the response variable, perform the stepwise selection as previously using both AIC and BIC criteria. Compare with the previous results in terms of selected variables and adjusted R^2.

2. The data set fancy (you need to library the fpp package to get the dataset) concern the monthly sales figures of a shop which opened in January 1987 and sells gifts, souvenirs, and novelties. The sales volume varies with the seasonal population of tourists.

   (a) Produce a time plot of the data and describe the patterns in the graph. Identify any unusual or unexpected fluctuations in the time series.

   (b) Use R function tslm to fit a regression model to the logarithms of these sales data with a linear trend and seasonal component.

   (c) Use multiple regression with trend variable and seasonal dummy variables to redo the regression as shown in the lecture example. Check to see that you obtain the same results as tslm.