STAT 700
Homework 2 Problems
due Wed. Sept. 14

2 Problems. Show all work.
Please follow the Lab report directions off the homework web page for R Problems.
Please work in Groups 2 (or 3)!

1. Consider the linear model from class,

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}.$$

Assume that the ε_i are independent $N(0, \sigma^2)$ random variables or equivalently

$$\boldsymbol{Y} \sim N_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}_n).$$

Also, assume that $\boldsymbol{X'X}$ is invertible.

The prediction of a future observation, $Y_0 = \boldsymbol{x}_0'\boldsymbol{\beta} + \varepsilon_0$ at a given vector of independent variables $\boldsymbol{x}_0'$, is given by $\widehat{Y}_0 = \boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}$. Find the expected value, variance, and the distribution of $\widehat{Y}_0$.

2. **Forbes' data**. James D. Forbes a Scottish physicist in 1857 collected data to see if the simpler measurement of the boiling point of water could be substituted for a direct reading of barometric pressure. The data are measurements in the Alps and Scotland from a barometer and a thermometer. Boiling point measurements were adjusted for the difference between ambient air and a standard temperature. The data are for $n = 17$ locations and measurements on $Temp$=boiling point (degrees Fahrenheit) and $Pressure$=corrected barometric pressure (inches of mercury).

The data is available off the class web page:
    http://www.rohan.sdsu.edu/~babailey/stat700/forbes.txt

Use the R read.table command with the header=T option. You do not need to make your own labels!

(a) Make a scatter plot of $Pressure$ (Y) versus $Temp$ (X). Make an appropriate title for the plot.

(b) Would a straight line closely match the data? To answer this question fit the model $Pressure = \beta_0 + \beta_1 Temp + \varepsilon$. Include the R summary of the model fit from lm. Also, include the four R diagnostic plots. Do you detect any pattern in the residuals when you examine the plot of the residuals versus fitted values? Does the Normal Q-Q plot of the residuals indicate any departures from normality?

(c) Fit the model $\log(Pressure) = \beta_0 + \beta_1 Temp + \varepsilon$. Include summary and diagnostic plots from `lm`. Do you detect any pattern in the residuals when you examine the plot of the residuals versus fitted values?

(d) Construct 90% CIs for β_0 and β_1 for the model in part (c) using R commands. (Hint: there is a `confint` function in R.) Using your 90% CI for β_1, expain if the slope of the regression line is significant or not significant.