Stat 704, Homework 3

1. Consider regression of Y on X_1, centered at $\bar{X}_1$ (See equation 1.6):

$$Y_i = \beta_0^* + \beta_1(X_{1i} - \bar{X}_1) + \epsilon_i,\ i = 1, \ldots, n$$

   (a) Confirm that the LS estimates for β_0^* and β_1 are $\bar{Y}$ and the familiar $b_1 = \frac{\sum(X_{1i}-\bar{X}_1)(Y_i-\bar{Y})}{\sum(X_{1i}-\bar{X}_1)^2}$, respectively. What are the implications for interpretation of β_0^*?

   (b) We can write the design matrix for this problem in the following form:

$$X = \begin{bmatrix} 1 & X_{11} - \bar{X}_1 \\ 1 & X_{12} - \bar{X}_1 \\ \vdots & \vdots \\ 1 & X_{1n} - \bar{X}_1 \end{bmatrix}$$

   Use the form above to confirm that $X'X = \begin{bmatrix} n & 0 \\ 0 & X_1'(I - \frac{1}{n}J)X_1 \end{bmatrix}$, where $J = 11'$. What is $(X'X)^{-1}$? What insights does this provide on inference for the LS estimators for this model?

2. **Patient Satisfaction** (Use SAS): 6.15a-e (The stem-and-leaf plot can be a histogram incorporated into the scatterplot matrix in part b, just as we did in class. For other parts of this problem, you do not have to generate graphs that are automatically generated by SAS). 6.16a, b (obtain separate 90% CI's for β_1, β_2 and β_3). 6.17a.

3. 6.30a-d. Use R (the `pairs()` function can be used to generate scatterplot matrices).