STAT 203 PROBLEM SET 3

(1) An economist studying the relation between household electricity consumption (Y) and number of rooms in the home (X) employed linear regression model and obtained residuals listed in file **ElectricityConsumption.txt**. Plot the residuals versus X. What problem appears to be present here? Might a transformation alleviate this problem?

(2) RABE Problem 7.4.

(3) This problem uses data from a study of computer-assisted learning by 12 students, showing the total number of responses in completing a lesson (X) and the cost of computer time (Y, in cents).
   (a) Fit a linear regression function by ordinary least squares. Obtain the residuals, and plot the residuals against X. What does the residual plot suggest?
   (b) Plot the absolute values of the residuals against X. What does this plot suggest about the relation between the standard deviation of the error term and X?
   (c) Estimate the standard deviation function by regressing the absolute values of the residuals against X, and then calculate the estimated weight for each case. Which case receives the largest weight? Which case receives the smallest weight?
   (d) Using the estimated weights, obtain the weighted least squares estimates of β_0 and β_1. Are these estimates similar to the ones obtained with ordinary least squares in (a)?
   (e) Compare the estimated standard deviations of the weighted least squares estimates b_{w0} and b_{w1} in part (d) with those for the ordinary least squares estimates in part (a). What do you find?
   (f) Iterate the steps in parts(d) and (e) one more time. Is there a substantial change in the estimated regression coefficients? If so, what should you do?

(4) Generate 200 observations of three variates X_1, X_2, X_3 according to

$$X_1 \sim Z_1$$

$$X_2 = X_1 + 0.001Z_2$$

$$X_3 = 10Z_3$$

where Z_1, Z_2, Z_3 are independent standard normal variates. Compute the leading principal component for the covariance and correlation matrices. Hence show that the leading principal component of the covariance matrix aligns itseslf in the maximal variance direction

X_3, while the leading principal component of the correlation matrix essentially ignores the uncorrelated component X_3, and picks up the correlated component $X_2 + X_1$.

(5) RABE Problem 11.7.

(6) For the presidential election data in the last problem, use the LARS package in `R` to find the best fitting model, following the following steps:

(a) Write down the criterion that is being optimized in LARS.

(b) Run LARS to find the order that variables are added to the model.

(c) Use cross-validation to select the value of the regularization parameter λ. Report the final fitted model. How does this model compare to what you fitted in the last problem?