**Homework 6 (50 pts)**

**due 11/9**

An assistant in the district sales office of a national cosmetics firm obtained data on advertising expenditures and sales last year in the district’s 44 territories. Data is consmetics.csv

X1: expenditures for point-of-sale displays in beauty salons and department stores (X$1000).

X2: expenditures for local media advertising.

X3: expenditures for prorated share of national media advertising.

Y: Sales (X$1000).

1. (4) Test the regression relation between sales and the three predictor variables. State the hypotheses, test statistic and degrees of freedom, the p-value, the conclusion in words.

2. (4) Determine whether the linear regression model is appropriate by using the “usual” plots (scatterplot, residual plots, histogram/QQ plot). Explain in detail whether or not each assumption appears to be substantially violated.

3. (4) Prepare a partial regression plot for each of the predictor variables. Do your plots suggest that the regression relationship in the fitted regression function are inappropriate for any of the predictor variables? Explain.

4. (4) Are there any outlying Y observations? (Do not include in your answer the values for all cases. Use plots and verbal summaries instead. You may include values for a few selected cases if you wish).

5. (4) Are there any outlying X observations? (Do not include in your answer the values for all cases. Use plots and verbal summaries instead. You may include values for a few selected cases if you wish).

6. (4) Are there any influential points?

7. Is there a serious multicollinearity problem?

1. (3) Include an appropriate scatterplot and correlation values between the explanatory variables.
2. (3) Judge by VIF, do you think there is a problem with multicollinearity? (Hint: VIP or tolerance)
3. (3) Compare your answers in parts i and ii. Are your conclusions the same or different? Please explain your answer.

8. Instead of removing variables, we are going to use the Ridge Regression to determine the parameter values.

a) (3) Make a ridge trace plot. What value of the parameter () do you believe is best? Explain your choice.

b) (3) Using the VIF factors, what value of the parameter do you believe should be used? (Hint: Look at both the graph and the printed numbers.) Explain your choice.

c) (3) Using your choice for the parameter, what are the standard regression parameters in this situation?

9. The number of defective items produced by a machine produced by a machine (Y) is known to be linearly related to the speed setting of the machine (X). Data is in machine.csv.

a) (3) 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) (3) Plot the absolute value of the residuals and the squared residuals vs. X. Which plot has a better line?

c) (4) Perform a weighted least square using the squared residuals to compute the weights. Obtain the weighted least squares estimates for the estimated parameters and their standard errors. Are these values similar to the ones produced in a)? Which results are better, the ones generated in a) or c)? Please explain your answer.

d) (4) Re-calculate the residuals for the weighted least squares and make a residual plot vs. X. Did this correct the problem that was seen in a).