Homework 1

Due Date: 3:00 PM Wednesday January 18

General Instructions:

You may work on the assignment in groups of up to 3 students from the class.

Include a title page with the names and student ID numbers of all group members.

Submit your homework by the time it is due in lecture or in my mailbox in 4118 MSB.

Include all used computer code in an appendix at the end of the assignment.

Assignment:

The file *kidney.txt* contains measurements of the kidney function of 33 individuals. The variables are creatinine clearance (Y), serum creatinine concentration (X_1), age (X_2), and weight (X_3).

Creatinine clearance is an important measure of kidney function. However, it is difficult and expensive to obtain in an office setting. It is desired to see if a relatively easy and inexpensive measurement called serum creatinine concentration is effective in predicting creatinine clearance along with other patient information such as age and weight. We will fit a multiple regression model of the form
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i, i = 1, ..., n,$$
where $\{\varepsilon_i\}$ are independent $N(0, \sigma^2)$ random variables.

1. Obtain a histogram for each of the variables. Comment on any noteworthy features you see.

2. Obtain the correlation matrix, and the matrix plot of the data (make scatterplots of all variables against one another). What do the plots suggest about the relationship between Y and each of the predictor variables? Does it seem that there is any problem of multicollinearity?

3. Fit a multiple linear regression model. Obtain the parameter estimates, their standard errors, the analysis of variance table, R^2, and R^2_{adj}.

4. Obtain a plot of observed against fitted Y values. Also plot the residuals against the fitted values. Does it seem that the fitted model is reasonable? Do you suspect any nonlinearity? Is the assumption of equal variance of the of the errors (i.e., ε's) reasonable here? Explain your answers.

5. Plot the residuals against each of the independent variables (in separate plots). Do you suspect any nonlinearity? Is the assumption of equal variance of the errors reasonable here? Explain your answers.

6. Obtain a histogram of the residuals. Also obtain a normal probability plot of the residuals. Is the assumption of normality of the errors (i.e., ε's) reasonable? Explain.

7. Does it seem that all the independent variables need to be retained in the model? If you were to delete one predictor, which would be the best candidate for deletion? Provide explanations for your answers.

8. Using the AIC and BIC criteria, choose the best subset of predictors using backward elimination. For the selected model(s), obtain the parameter estimates and their standard errors.

9. Use the AIC and BIC criteria to select the best among **all possible** regression models. For the selected model(s), obtain the parameter estimates and their standard errors.