Stat 401, Section F Homework 10
Due Date: Wednesday, November 7

1. The file `meatproc.txt` contains the 12 data points described in problem 16 of Chapter 8 regarding the pH level in steer muscle at various hours after slaughter. Test for a lack of fit in the regression of pH on $\log(Hour)$. Give the test statistic, degrees of freedom, a p-value, and a conclusion. If you want, you could use the SAS program `meatproc.sas` to do some of the necessary calculations, although you are not required to.

2. Animal ecologists wish to track the health of bear populations in the wild. By obtaining physical measurements of bears, researchers can gauge the effects of changes taking place in and around bear habitat (e.g., nearby housing and business development, recreational activities, introduction of non-native plant and animal species, severe storms, harsh winters, etc.). Researchers have studied bears by anesthetizing them in order to obtain vital measurements, such as age, gender, length, and width. A bear's weight is another important variable that is quite difficult to obtain in the wild because most bears are heavy and difficult to lift. The scientific problem is to develop a method for predicting the weight of a bear, given other more easily obtained measurements. A good method might alleviate the need to weigh bears in the wild and greatly simplify the data collection process.

   In one detailed study, researchers were able to obtain weight measurements along with several other variables for each of 54 bears. The file *bear.txt* contains the data, which has 54 rows and 9 columns. There is one row for each bear that was anesthetized and measured carefully using a tape measure and scale. There is one column for each variable in the data set. The variables (from left to right) are

   age: age in months

   length: length of body in inches

   sex: 1=male 2=female

   weight: weight in pounds

   chest: chest circumference in inches

   headlth: length of head in inches

   headwth: width of head in inches

   month: month of capture (1=January, 2=February, etc.)

   neck: neck circumference in inches

   (a) Estimate the parameters in the multiple regression model with *weight* as the response variable and *chest*, *length*, and *neck* as explanatory variables. Specifically, provide an estimate of the intercept, an estimate of the partial regression coefficient for each explanatory variable, and an estimate of the standard deviation of bear weights for any given values of the explanatory variables.

   (b) Conduct one test of the null hypothesis that says that the partial regression coefficients for chest, neck and length are all zero against the alternative that at least one coefficient is not. State the hypotheses, the test statistic, the degrees of freedom, p-value and conclusion.

   (c) Provide an interpretation of the partial regression coefficient associated with the variable *chest*.

   (d) Compute a 95% confidence interval for the partial regression coefficient associated with the variable *length*. Is the partial regression coefficient associated with the variable length significantly different from zero? Explain how your confidence interval can be used to answer this question.
   Hint: Think confidence interval is: "$estimate$" $\pm\ t_{df} \times SE($"$estimate$"$)$

   (e) What proportion of variation in bear weights is explained by the multiple regression of *weight* on *chest*, *length*, and *neck*?

   (f) Provide an estimate of the mean weight of a captured bear that is 60 inches long, with chest circumference of 35 inches, and neck circumference 24 inches.

   (g) Provide a 95% confidence interval for the mean weight estimated in part (f).

   (h) Provide a 95% prediction interval for the weight of the bear described in part (f).

(i) Examine a residual plot and a normal probability plot, boxplot, and histogram of the residuals from the fit of the multiple regression model with *weight* as the response variable and *chest*, *length*, and *neck* as explanatory variables. Which of the four assumptions of multiple linear regression should be questioned from what is depicted in these plots?

(j) Estimate the parameters in the multiple regression model with *logweight* as the response variable and *logchest*, *loglength*, and *logneck* as explanatory variables. These variables are simply the natural logs of the variables considered previously.) Specifically, provide an estimate of the intercept and an estimate of the partial regression coefficient for each explanatory variable.

(k) Examine a residual plot and a normal probability plot, boxplot, and histogram of the residuals from the fit of the multiple regression model with *logweight* as the response variable and *logchest*, *loglength*, and *logneck* as explanatory variables. Do these plots indicate any inconsistencies between the data and the model assumptions?

3. An exercise physiologist wanted to compare the body-fat-reducing effects of three exercise programs (A, B, and C) and varying levels of a dietary supplement (0, 1, 1.5, and 2 g/day). A total of 24 males between 28 and 35 years of age were used for the study. The males were assigned in a completely randomized manner to the 12 combinations of exercise program and dietary supplement level. After completing 12 months of training according to the assigned exercise program and consuming the dietary supplement at the assigned rate, the body fat of each male was measured. The data set `bfchange.txt` posted on the web site consists of three variables: the exercise program, the level of the dietary supplement and the body fat before treatment minus body fat after treatment for each subject. For full credit, you will need to write out the general form of the model, as well as the form of the model for each of the exercise types.

   Assume that the following multiple regression model fits the data well and that all the assumptions of multiple regression are satisfied:

$$\mu(bfchange|X_1, X_2, X_3, X_4, X_5) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

   where X_1 is the supplement amount for each subject is a value, X_2 and X_3 are indicator variables so that $X_2 = 1$ if the subject used exercise program-A ($X_2 = 0$ otherwise) and $X_3 = 1$ the subject used exercise program-B ($X_3 = 0$ otherwise. Finally, $X_4 = X_1 X_2$ and $X_5 = X_1 X_3$.

   (a) Estimate the mean change in body fat for each exercise program when no dietary supplement is consumed. Also provide a 95% confidence interval for the exercise-program-C/no-supplement mean in addition to the estimated mean.

   (b) Is there a statistically significant difference between the estimated means for exercise programs A and C computed in part (a)? Provide an appropriate test statistic, p-value, and conclusion.

   (c) By how much does the mean change in body fat for exercise-program-C subjects taking 2 g/day of the supplement differ from the mean change in body fat for excercise-program-C subjects taking 1 g/day of the supplement? Provide a 95% confidence interval for the difference in means.