

**Stat 401 A XM Homework 10**  
**On-Campus Due Date:** Wednesday, November 19  
**Off-Campus Due Date:** November, December 3

1. The average and standard deviation of scores on Exam 1 were 85.2 and 7.7, respectively. The average and standard deviation of scores on Exam 2 were 82.9 and 9.8, respectively. This data was computed using the scores of the 38 students who took both Exam 1 and Exam 2. For this problem, imagine that these 38 students are like a random sample of graduate students who would take Stat 401.
  - (a) Name 3 statistical tests that may be appropriate for addressing the question, “Were the scores on Exam 1 significantly higher than the scores on Exam 2?” Explain what additional information you would need in order to conduct these tests.
  - (b) The sample linear correlation between Exam 1 and Exam 2 scores is 0.565. Assuming that the assumptions of least-squares regression are satisfied, compute the equation of the least-squares regression line for predicting Exam 2 performance from Exam 1 performance.
  - (c) Is the slope of the least-squares regression line significantly different from 0? Provide a test statistic, its degrees of freedom, a  $p$ -value, and a conclusion.
  - (d) Estimate the average Exam 2 score for students who score 85 on Exam 1.
  - (e) Give a 95% confidence interval corresponding to your estimate above.
  - (f) Suppose a new student were to take Stat 401 under the same conditions experienced by our class. Give a 95% prediction interval for the student’s score on Exam 2, given that the student scored 85 on Exam 1.
  
2. An experiment was conducted to compare soybean yields of plots treated with one of three herbicides (A, B, or C) or a control (D) . A total of 32 soybean plots were used in the experiment; 8 plots were randomly assigned to each of the four treatments. Statistics summarizing the yields in bushels per acre are provided in the following table.

Treatment	Number of Plots	Mean Yield	Variance of Plot Yields (SD <sup>2</sup> )
A	8	60.3	9.8
B	8	62.1	10.1
C	8	57.6	9.2
D	8	52.0	9.7

Is there any evidence to suggest that all the herbicide means are not identical? Use all of the available data to conduct one  $F$ -test of the null hypothesis  $H_0 : \mu_A = \mu_B = \mu_C$  against the alternative that  $H_0$  is false. Compute the test statistic, state its degrees of freedom, determine an approximate  $p$ -value, and provide an appropriate conclusion.

3. 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 an 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 data can be read using the file *bear.sas* on the course web site. The data set contains 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) Provide an interpretation of the partial regression coefficient associated with the variable *chest*.
- (c) 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.
- (d) Conduct one test of the null hypothesis that says that the partial regression coefficients for both *neck* and *length* are zero. (Hint: You should conduct a test that compares the fit of a reduced model with the fit of a full model. You will need to add a little code to *bear.sas* to get the residual sum of squares for the reduced model.)
- (e) What proportion of variation in bear weights is explained by the multiple regression of *weight* on *chest*, *length*, and *neck*?
- (f) 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?
- (g) 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, an estimate of the partial regression coefficient for each explanatory variable, and an estimate of the standard deviation of log bear weights for any given values of the explanatory variables.

- (h) 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?
- (i) A bear is captured that is 60 inches long, with chest circumference of 35 inches, and neck circumference 24 inches. Use the better of the two multiple regressions considered in this problem to estimate the mean or median weight of all bears that share these characteristics. (*Note that the data for this bear is included as the last observation in the data set. You can get SAS to predict the response and compute confidence intervals and prediction intervals for you by adding “/ p clm cli alpha=0.05” to the end of the model statement in proc reg. See cyclone.sas for example code.*)
- (j) Provide a 95% confidence interval for the mean or median weight estimated in part (i).
- (k) Provide a 95% prediction interval for the weight of the bear described in part (i).