STAT 501 – Homework 1 – Fall 2015 – Due Date Aug 30

Instructions: Use Word to type your answers within this document. Then, submit your answers in the appropriate dropbox in ANGEL by the due date. The point distribution is located next to each question.

1. (**4x4 = 16 points**) Data were gathered for y = hours of sleep the previous day and x = hours of studying the previous day for n = 116 college students. The estimated regression equation is found to be $\hat{y}$ = 7.56 – 0.269x.

   (a) What is the estimated slope of the regression line? Write a sentence that interprets the slope in the context of the variables for this problem. That is, explain exactly what the slope indicates about the relationship between hours of sleep and hours of study.
   (b) What is the estimated intercept of the regression line? In the context of these variables, what is the interpretation of the intercept?
   (c) For students who studied 2 hours the previous day, what is the estimated value of average hours of sleep the previous day?
   (d) Suppose that a student studied 2 hours the previous day and slept 6 hours that day. What is the value of the residual for that person? (A residual is defined as the difference between the observed and predicted values of y for an individual.)

2. (**2x4 = 8 points**) Two of the following statements about a population model for a simple regression are correct and two are incorrect. Which two statements are correct? Explain your answer.

   (a) $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
   (b) $Y_i = \beta_0 + \beta_1 X_i$
   (c) $E(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i$
   (d) $E(Y_i) = \beta_0 + \beta_1 X_i$

3. (**4x4 = 16 points**) The fitted line plot below gives results for a straight-line regression between y = left forearm length (cm) and x = left foot length (cm). The data are from n = 55 college students.

Fitted Line Plot
LeftArm = 10.16 + 0.5963 LeftFoot

S	1.65245
R-Sq	45.6%
R-Sq(adj)	44.5%
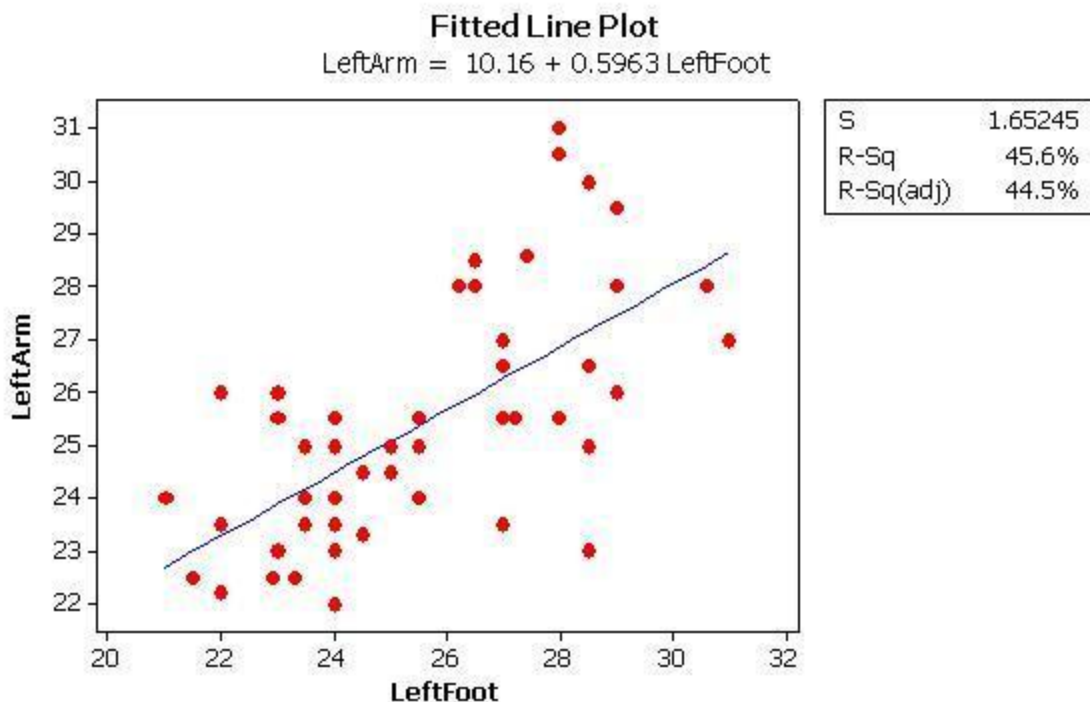
(a) Write a sentence that gives the value of the slope and interprets it in the context of this situation.

(b) Write a sentence that gives the value of R^2 and interprets it in the context of this situation.

(c) Use the equation to estimate the average left forearm length of college students with a left foot length of 25 cm.

(d) The Minitab output includes the information that "**S**=1.65245". Explain what is measured by this statistic.

4. (**10x4 = 40 points**) Infant weights in pounds have an upward linear trend with age in months. Data from a sample of 5 babies in a local community, including one newborn and four others who are 1 month, 2 months, 3 months and 5 months old, were used to obtain an estimated regression equation based on the least squares criterion with a slope of 0.2838 pounds per month. Some of the information is given in the following table.

Age (x_i)	0	1	2	3	5
Weight (y_i)		8.1	8.4	9.3	
Predicted weight ($\hat{y}_i$)				8.8271	
Residual error (e_i)	0.0243				

(a) What is the equation of the population regression line in this setting? *[Hint: There should be no numbers in this equation, just β's.]*

(b) What is the estimated regression equation? *[Hint: There should be numbers in this equation. Use the information in the question and in the table, particularly the 3-month old baby.]*

(c) Based on the estimated regression equation, what is the predicted birth weight of a newborn in this community?

(d) What is the actual birth weight of the newborn in the sample?

(e) Complete the remaining entries in the table above.

(f) Comment on the validity of using the estimated regression equation to predict the weight for a one year old.

(g) Calculate SSE, the sum of residual error squares.

(h) Calculate the sample estimate of the variance, σ^2, for the regression model.

(i) Calculate the value that would be given in Minitab for "**S**=". Write a sentence that interprets this value.

(j) Calculate the value of R^2. To start, you will have to calculate the value of SSTO. Write a sentence that interprets the value of R^2.

5. (**4x5 = 20 points**)
   (a) Briefly describe the four assumptions (or conditions) that underlie the simple linear regression model.
   (b) The scatter plot below shows sample data for y = selling price of a house and x = square foot area of the house.
      (i)   Name one condition that may be satisfied by the selling price vs square foot area data and justify your choice.
      (ii)   Name one condition that may not be satisfied by the data and justify your choice.
      (iii)   Would you expect the magnitude of the sample correlation coefficient to be near 0, closer to +1, or closer to −1? Justify your choice.
      (iv)   Based on the estimated regression equation, $\widehat{Price}$ = 5049 + 64.87 SQFT, what is the y-intercept estimate? Is this value meaningful? Why or why not?

Scatterplot of Selling Price versus Square Feet Area