CHAPTER 24: INFERENCE IN REGRESSION

Chapters 4 and 5:   Relationships between two quantitative variables.

 Be able to

- Make a graph (scatterplot)

- Summarize

    o   the **form** (linear or curved or other)

    o   If it is linear, the **direction** (positive or negative)

    o   the **strength** (how close the points are to the line or curve that summarizes the pattern)

- If it is at least somewhat linear, compute and interpret a numerical summary of the strength (correlation coefficient)

- If the relationship is at least somewhat linear, find the equation of the best-fitting line (the regression line)

- Use that regression line to make predictions of y when we are given the value for x.

Chapter 24: Make inferences about the population from which the sample data came.

1. Find and interpret a confidence interval for the slope coefficient.

2. Test the hypothesis of no linear relationship between the variables (test the slope and / or the correlation coefficient)

3. Find and interpret interval estimators for the predicted value of y for a given value of x. (Two different estimators)

4. Discuss the conditions needed to do regression inference.   Discuss whether these conditions are met. Do that by using (1) information in the description, (2) a residual plot of the residuals versus the explanatory variable and (3) a histogram of the residuals.

5. Note: Each of the processes in this chapter has an SE associated with it.   We will find ALL of these we need by reading them from the output. Despite the fact that our text has formulas for these, we will not compute them from the formulas.
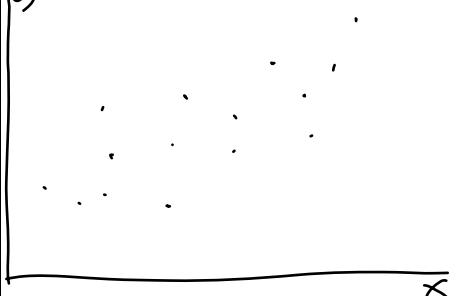
Review of Chapters 4 and 5, mainly using software.

Activity 1:   Look at Example 24.1 in Moore's text on Crying and IQ.

1. Decide which variable is the explanatory variable and which is the response variable.

2. Use Minitab to
    a.   Make a scatterplot.
    b.   Find the correlation coefficient by using Stat > Basic Statistics > Correlation
    c.   Find the equation of the regression line.

3. Interpret the slope in context.

4. Interpret the intercept in context.

5. Interpret the r^2 value in context.

6. Use the equation of the regression line to predict the IQ for a child whose CryCount was 10.

7. Use the "Options" in the Minitab   Stat > Regression > Regression command to predict the IQ for a child whose CryCount was 10.   Notice that it gives you some intervals.   We will learn to interpret these.   Copy the prediction and the intervals (with the labels CI and PI on the intervals) below.

Conditions: Discuss the population model and the linear regression equation in the sample data.

	Population	Sample
Data (Repeated responses of y are independent of each other.)		
line	$\mu_{\hat{y}} = \alpha + \beta x$	$\hat{y} = a + bx$
variability (standard deviation)	σ = Variation of y values for a given x-value. **It is the same for all x-values.** (Fig. 24.2 on p. 590 in "Estimating the parameters.")	s = sample standard deviation. (Estimate of σ using residuals.)
shape	Population: y-values for a given x are normally distributed, with center on the line. 	
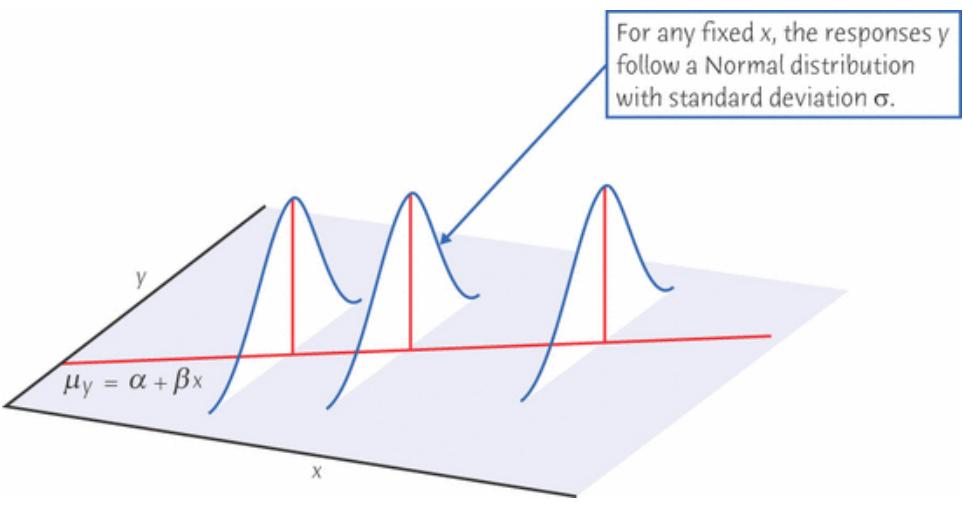
Activity 2:   In the output from Stat > Regression > Regression, find the value of s which is the estimator of σ and the SE for the slope coefficient. (Example in text: Exercise 24.2   See Minitab output. S = 7.90308, SE = 33.02)

Now, for the CryCount and IQ data, estimate σ and find the SE_b of the slope coefficient. (Answers: s =17.50 and SE_b = 0.4870.)

Testing hypothesis of no linear relationship between the variables.

Text: See Ch. 24 Section "Testing the hypothesis of no linear relationship"

Steps:

1. Understand why the null hypothesis of beta = 0 means "no linear relationship."

2. Write hypotheses and choose your significance level.

3. Use Stat > Regression > Regression to find the equation of the regression line and the table of values including the slope coefficient, the SE next to the slope coefficient, the t-score, and the P-value (two-sided p-value).

4. Find the t-score and P-value: (Practice both ways)

    o   Use Minitab to see these.

    o   Use the sample slope coefficient and the SE_b from the output. Also use df = n – 2 and the t-table to find the P-value. When the df you need is not in the table, you may use the closest one.   Or you may use the next smaller one, as the textbook says.

5. Write your conclusion in context.

Activity 3:   Do Exercise 24.5.

Text: See Ch. 24 Section "Testing lack of correlation."

This tests exactly the same question as the test of the slope coefficient. But it uses the correlation coefficient for the test statistic instead.   When you use Minitab to compute the correlation coefficient, it also gives you a P-value, which is for these hypotheses:

Ho: population correlation coefficient is 0
Ha: population correlation coefficient is not equal to 0.

Activity 4: Do Exercise 24.7b using Minitab instead of using Table E. (You will not be expected to use Table E at all in this class.) Use the four-step process.

Finding and interpreting a confidence interval for the slope coefficient.

Text: Confidence intervals for the regression slope.

See the formula and Example 24.6.   Notice that the SE_b needed here is the SAME ONE we used in the hypothesis test of the slope coefficient.
 We use Table C to find t-scores, not software.   When the df you need is not in the table, you may use the closest one.   Or you may use the next smaller one, as the textbook says.
Pay careful attention to the interpretation.

Activity 5: Do both exercises 24.9 and 24.11 in this section. Be sure to write an interpretation of your interval before checking your answer.

Find and interpret interval estimators for the predicted value of y for a given value of x.

Moore text: Chapter 23, section "Inference about prediction"

1. Read example 23.7 about Beer and blood alcohol.

2. Look at Minitab output to find the conclusion of this problem.

3. Learn how to distinguish between the two intervals that are given.

4. Use Minitab to perform the regression and compute the fitted value you need (You tell Minitab the confidence level you want.)   That will always provide two intervals. (Crunch-It claims to do this, but the answers don't seem to always be accurate.)

5. Read the question VERY CAREFULLY to see which of these is requested:

   a. the prediction for ONE y value when x is that given value (wide interval, PI)

   b. the prediction for the MEAN y value for all points where x is equal to that given value (narrow interval, CI)

Activity 6: Do the following problems.

24.13 and then 24.12a.   Notice that these DO NOT ask for the same interval. Look at the language very carefully to see how they differ.

Checking the conditions

Text: See Ch. 24 section "Checking the conditions for inference"

There are four things to check, listed on the first page of this section.   The third (independent observations) is perhaps the trickiest one.   I will not expect you to do that on a test.   The others can all be checked using various graphs.   Read those. Then review Example 24.9 (Climate change and fish) in this section.

Activity 7:   Exercise 24.15.   For part a, do those two parts for Exercise 24.2 about this same data.   For part b, do not do a stemplot, but instead, choose the "Graphs" button in the Minitab Regression Regression and choose to do a histogram of the residuals.   For part d, do a residual plot of the residuals versus the explanatory variable.   In order to become skilled in interpreting all of this, you would need to work with more data sets than we have time to do at this point in the course.   You are expected to know what graphs are needed to check each condition, to produce the appropriate graphs, and to interpret them if the interpretation is pretty clear.   If it is "borderline" I will not expect you to necessarily interpret it in the way I would.

 Quiz 14: Due Wed. May 1 **at the beginning of class.**   Submit a word-processing document in this quiz assignment in Blackboard with your solutions for this quiz, as if it were a test question.   Include your answers in the main part and your computer output in the Appendix.   In the Appendix, identify the output for each of the individual problems separately.   You have ONE opportunity to submit this, just as you will for the test.

24.44   (40 points) In the four-step process, list two questions in the "state" part.   As you work the problem, make clear which parts are relevant to answering which question.

The following parts of problems are intended to make sure you understand the various types of intervals we learned in this chapter. These are NOT four-step problems. In the solution part of your document, It is only necessary that you give the particular answers requested.

24.32c, 24.34, 24.38c, 24.40c