In these activities you will investigate patterns of association in bivariate data. After completing the activities, discuss and/or present your findings to the rest of the class.

Activity 1 [Page 1.3]

1. a. Refer to the graph you created on page 1.3. Select the point representing a box of cereal, then **Enter**. Explain what the number and vertical segment tell you.

    b. Select the point you think will have the greatest error (difference between the actual and predicted cost). Find the difference and explain what it means.

    c. Fill in the table with the missing values using your line.

Item	Store A Price $	Store B Price $	Store B price predicted by line*($)	Actual price minus predicted price ($)
Whole Grain Cereal (10 oz)				
Raisins				
Peanuts				
Ketchup (24 oz)				

    d. For the items in the table, what is the total of the differences in the actual prices for the items minus the prices predicted by your line? Explain what your answer means in terms of the prices.

2. For an article they were writing, the class wanted to describe how far off the predicted prices were from the actual prices for the items at the two stores.

   a. Marj claimed the sum of the differences in actual price minus predicted price for the items in the table was -0.59, or about $\frac{-60}{4} = -15$ cents per item. Petra disagreed and said if you found the mean the way Marj did, you could have a sum of the differences in actual minus predicted prices to be 0, when there really were a lot of "differences". Whose reasoning makes the most sense and why? Give an example to support your thinking.

   b. Hilary suggested finding the sum of the differences by taking the absolute value of the differences and then adding. What do you think about Hilary's reasoning?

   c. Anita said she would find the mean difference by squaring all the differences and finding the sum of the squared differences. What do you think about Anita's reasoning?

   d. Use one of the methods described in b) and c) to find the sum of the differences in prices. Compare your answer with others. What might explain differences?

3. SAD is the sum of the absolute value of all of the differences. Select All Segments then Show SAD.

   a. How does the value for SAD relate to the scatter plot and line?

b. The predicted cost of all of the items at Store B using the equation B = 1.25 A is $63.38, and the actual cost was $59.58. How does the difference, actual total cost minus predicted total cost compare to the SAD? Explain any differences.

c. Move the line. Describe what happens to the SAD.

d. Find a new line that has a smaller SAD than your original line. Explain what your SAD represents.

4. Identify the following statements as true or false. Be ready to explain your thinking.

a. The line $y = x$ represents a relationship where all of the ordered pairs are of the form (x, x).

b. If the equation for a linear *relationship* is $y = 2x$, the x-values are twice as large as the y-values.

c. Suppose the equation that models a linear relationship is B = A, where A and B are greater than or equal to 0. If the point (A, B) is below the line, then B will be greater than A.

d. If the difference, actual price of an item minus the price predicted by a line, is negative, the point representing that item lies below the line.

 education.ti.com

Activity 2 [Page 2.2]

1. Determine the error in using the line to predict the bite force for the Indian Gharial.

   a. What does the value tell you?

   b. Thom says that the length of the segment from this point to the line is larger than all of the other segments together. How could you decide if he is correct?

   c. Move the line to minimize the total sum of the absolute distances. What is the equation of your line? Compare your line to others in your class. Who has the smallest SAD?

   d. What is the mean absolute deviation of the difference in the actual values for a given weight and the values predicted using your model?

2. Another crocodile, Croc G, with a body mass of 350 pounds had a bite force of 1100 psi.

   a. How do you think the SAD will change if you add Croc G to the plot?

   b. What do you think will happen to the original SAD of 2288.08 psi if a crocodile that weighs 200 pounds and has a bite force of 800 psi is added to the plot?

c. If you add a point representing a crocodile that weighs 200 pounds with a bite force of 1000 psi, do you think the SAD will increase, decrease or remain the same?

3. In 2012, an American alligator named Hercules was weighed and his bite force measured.

a. Use your equation to predict the bite force given that Hercules weighed 665 pounds.

b. Use the MAD you found in 1d) to find an interval that would typically contain the bite force for Hercules.

c. Hercules actually had a bite force of 2125 psi. How well did your model predict his bite force?

d. An adult male lion a weighs about 420 pounds and has a bite force of about 600 psi. Do you think it would be reasonable to use your model to estimate the bite force of other animals? Explain why or why not.

Activity 3 [Page 3.2]

1. Many different lines can be drawn to fit a linear relationship. One method to find a good fit is to keep the sum of all the absolute values of the differences between the predicted and actual values (SAD) as small as possible. Another method is to see if there is any pattern or predictability in the residuals (errors). If you can describe a pattern in the error in predicting the dependent variable, you might be able to find a better model. The "scatter" around a line you are using to model a linear relationship (called a regression line) should appear to be random.

    a. Select Show SAD. What is the SAD for the line on the screen? What does that number represent?

    b. Trey moved the line and the equation for his new line is P = 9.44D+25. Create Trey's line. Did his new line reduce the SAD?

    c. Do the points on the residual plot seem to be randomly scattered? Explain your reasoning.

    d. Move the line until the residuals seem to be random. What is the equation for your line? The SAD?

    e. Compare your line with others. How well does each seem to satisfy the criteria of a small SAD and no patterns in the residuals?

2. Match the sentence starters with an ending that makes a true sentence. Not all parts will have matches.

Sentence Starters

a. If a scatter plot displays a linear pattern between two variables

b. A model is a good fit for the linear pattern in the data

c. A residual is

d. An equation that models the relationship between two variables

e. If a scatter plot shows a clear pattern

Sentence Enders

f. the distance between a point and a line used to model the relationship between the variables.

g. a straight line can be used to model the relationship.

h. if the residual plot is randomly distributed around the horizontal axis.

i. can be used to predict one value given the other.

j. the difference between the actual outcome and the predicted outcome for a given input.

k. if the sum of the absolute value of the residuals is small.