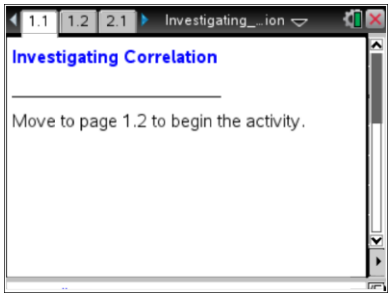
Investigating Correlation

Student Activity

Name _____

Class _____

Open the TI-Nspire™ document *Investigating_Correlation.tns.*



The correlation coefficient is a measure of the strength and direction of the linear association between two variables. In this activity, you will investigate the connection between the scatter plot of bivariate data and the numerical value of the correlation coefficient.

Tech Tip: Hovering over a point will display the coordinates of that point. To de-select a point, click in any white space on the screen. When grabbing and moving points in a Data and Statistics plot, it is necessary to de-select a point after it has been moved; otherwise, when a new point is selected, both points will move together.

Move to page 1.2.

1. The scatter plot on the left of page 1.2 displays the relationship between the variables ($x2$, $y2$); the scatter plot on the right of page 1.2 displays the relationship between the variables ($x1$, $y1$).
   a. Look at the scatter plot in the left screen, and predict a value of $y2$ you think will correspond to an $x2$ value of 2. Explain your reasoning.

   b. Look at the scatter plot in the right screen, and predict a value of $y1$ you think will correspond to an $x1$ value of 2. Explain your reasoning.

The correlation coefficient is a measure of the strength of the linear association between two variables. When the linear correlation is strong, knowing the value of one variable allows you to use a linear model to predict the value of the other variable with more confidence than when the correlation is weak. The linear correlation coefficient is usually represented by the symbol r.

2. The correlation coefficient for both plots is given below the scatter plots.
   a. Explain how you think the value of the correlation coefficient relates to each scatter plot.

b. Grab and drag the points in the right plot until you get a pattern different from the one in the left plot, but one that will allow you to be fairly confident in making a prediction for the second coordinate of a point when the value of the first coordinate is given. What is the correlation coefficient for your new plot?

c. Grab and drag the points in the left plot, and notice what happens to the value of the correlation coefficient. Do you think it is possible to have a correlation coefficient greater than 1? Explain why or why not.

Move to page 2.2.

3. The arrow in the upper left corner will generate pairs of scatter plots. In each case, observe the two screens, and decide which scatter plot appears to show the stronger correlation between the explanatory and response variables.
   a. Record your answers for each pair in the table below.

	Scatter Plot Left vs. Scatter Plot Right
Pair A, B	
Pair C, D	
Pair E, F	
Pair G, H	
Pair I, J	

   b. Explain the strategies you used to determine your answers to question 3a.

Move to page 2.3.

   c. Values of the correlation coefficients are provided for each plot you examined above. Use the values to check your answers to part a. What did you learn from checking your answers?

d. Describe the difference in a scatter plot with a correlation coefficient close to positive 1 and one with correlation coefficient close to negative 1.

Move to page 3.2.

The bottom of the screen on the left displays a target value for a correlation coefficient. On the right, the current value of the correlation coefficient for the scatter plot on the screen is displayed.

4. Move one or more of the points until you have a correlation coefficient value that approximately matches the target value. Sketch your scatter plot below. Select the arrow at the lower left of the page to produce a new target value.

   Target $r = 0.6$

   Target $r = 0.24$

   Target $r = -0.7$

Move to page 4.2.

5. Select the point farthest to the left in the left plot and the point farthest to the right in the right plot, and observe the coordinates. Continue to investigate each of the two plots by selecting other points in each plot and observing the coordinates.
   a. What is the difference between the two plots? Give an example that illustrates your answer.

b. What do you notice about the correlation coefficients for the two plots?

c. Grab and drag one or more points in either plot, and notice that the corresponding point in the other plot also moves. Observe the effect on the two correlation coefficients. What conjecture can you make?

Move to page 5.2.

6. Leave the point (18, 10) unchanged. Grab and drag the other points so that both coordinates for each point are less than four and the scatter plot has no pattern.
   a. What is the value of the correlation coefficient? What information does it provide?

   b. Would you be able to use a linear model to predict with confidence a response value for a given value of the explanatory variable? Explain why or why not.

7. Use the reset arrow to return to the original scatter plot. Grab and drag the data points into two clusters: one at the upper left of the screen, and one at the lower right of the screen. Answer questions 6a and b again for the new scatter plot.

8. Think over the work you have done in questions 1 to 7, and then answer the following questions.
   a. A scatter plot had a correlation coefficient of $r = 0.9$. Describe what this means about the relationship between the variables.

b. Create a scatter plot that looks like part of a parabola with a high correlation coefficient. Make a sketch of your plot, and record the value of the correlation coefficient.

c. Create a second scatter plot that looks like part of a parabola but has a low correlation coefficient. Make a sketch of your plot, and record the value of the correlation coefficient.

d. Do the plots you made for parts b and c contradict your answer in part a? Why or why not?

e. When is it appropriate to use the correlation coefficient to describe the strength and direction of a relationship?