Investigating Correlation

MATH NSPIRED

Math Objectives

- Students will recognize that the correlation coefficient describes the strength and direction of the linear association between two variables.
- Students will recognize that when two variables are highly linearly correlated, their correlation coefficient will be close to ± 1, and when they have little correlation, the correlation coefficient will be close to 0.
- Students will recognize that two variables with a high correlation coefficient might have a scatter plot that displays a nonlinear pattern.
- Students will recognize that correlation is not affected by the choice of *x* or *y*, that is, by the choice of which variable is explanatory and which is response.

Vocabulary

- correlation coefficient
- explanatory variable
- outlier
- scatter plot
- response variable
- linear

About the Lesson

- This lesson involves investigating the connection between the scatter plot of bivariate data and the numerical value of the correlation coefficient.
- As a result, students will:
  - Consider a scatter plot of points that lie in a straight line and one whose points do not line in a straight line and interpret the correlation coefficient for each plot.
  - Look at pairs of scatter plots to estimate which plot has the higher correlation coefficient.
  - Move points to try to match a given correlation coefficient.
  - Investigate a plot of ordered pairs and a plot of the inverse relation by inspecting the coordinates of the points and, by dragging the points in either plot, observing that the correlation coefficients are the same for both plots.

TI-Nspire™ Navigator™ System

- Send out the *Investigating_Correlation.tns* file.
- Monitor student progress using Class Capture.
- Use Live Presenter to spotlight student answers.
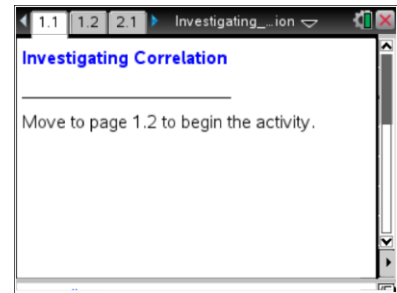
Tech Tips:

- This activity includes class captures taken from the TI-Nspire CX handheld. It is also appropriate for use with the TI-Nspire family of products including TI-Nspire software and TI-Nspire App. Slight variations to these directions may be required if using other technologies besides the handheld.
- Watch for additional Tech Tips throughout the activity for the specific technology you are using.
- Access free tutorials at http://education.ti.com/ calculators/pd/US/Online-Learning/Tutorials

Lesson Files:
Student Activity

- Investigating_Correlation_ Student.pdf
- Investigating_Correlation_ Student.doc

TI-Nspire document

- Investigating_Correlation.tns

Activity Materials

- Compatible TI Technologies: ☐ TI-Nspire™ CX Handhelds, 👆 TI-Nspire™ Apps for iPad®, 💻 TI-Nspire™ Software

Discussion Points and Possible Answers

> 📟 **Tech Tip:** If students experience difficulty grabbing and dragging a point, make sure they have not selected more than one point. Press [esc] to release points. Check to make sure they have moved the cursor until it becomes a hand (✋) getting ready to grab a point. Select [ctrl] [✋] to grab the point and close the hand (✊).

> 👆**Tech Tip:** If students experience difficulty grabbing and dragging a point, have them tap and hold the desired point for a few seconds and then drag the point to the desired location.

> 📟💻**Tech Tip:** Hovering over a point will display the coordinates of that point. To deselect a point, select any white space on the screen. When grabbing and moving points in a data and statistics plot, it is necessary to deselect a point after it has been moved; otherwise, when a new point is selected, both points will move together.

> **Teacher Tip:** This lesson does not assume that students have been introduced to the formula for the correlation coefficient. It could be used as an introduction to the concept of correlation

Move to page 1.2.

1. The scatter plot on the left of this page displays the relationship between the variables ($x2$, $y2$); the scatter plot on the right of this page 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.
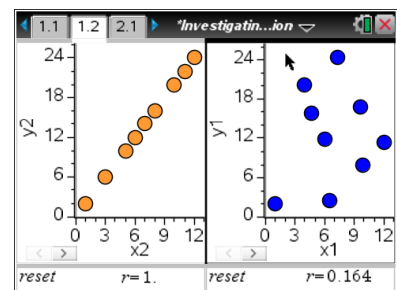
Sample Answer: The points all lie on a line. It would seem reasonable to predict that the ordered pair would be (2, 4) because the pattern suggested by the other points is that twice the first coordinate ($x2$) equals the second coordinate ($y2$).

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.

Sample Answers: It is difficult to predict because there is no clear trend for the relationship between the variables. When the $x1$ coordinate is around 2, the corresponding values for $y1$ go from approximately 2 to 20. At $x1 = 2$ the corresponding $y1$ might even be outside of those values.

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.

Sample Answers: When the correlation coefficient is 1, the points lie on a line, and you can predict what the y-coordinate is likely to be for any given x-value over the given domain. When the correlation coefficient is close to 0, there is no clear trend, and it is difficult to predict the second coordinate for a given first coordinate.

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?

Sample Answers: Students might generate another linear relationship where the correlation is 1 or close to 1. They might also generate a scatter plot where the correlation is close to -1 and note that when the correlation coefficient is -1, you can predict the second coordinate exactly, just as you can when the correlation coefficient is 1.

Teacher Tip: Students might create a nonlinear relationship that has a high correlation. If they do, suggest to the whole class that you will revisit this question later in the activity.

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.

Sample Answers: Correlation is a measure of the strength of a linear relationship. Since the relationship cannot be any stronger than when all of the points are on a straight line and the correlation for collinear data was given as 1, it would seem that the greatest possible correlation value is 1.

Teacher Tip: Note that the square of the correlation coefficient, r^2, is a measure of the percent of variability in the response variable that can be removed by the linear relationship with the explanatory variable. Thus, it does not make sense to have more than 100% of the variability removed. This activity approaches the ±1 boundaries for the correlation coefficient more informally, but you might want to discuss r^2 at this time.

TI-Nspire Navigator Opportunity: *Class Capture.*
See Note 1 at the end of this lesson.

Teacher Tip: Be sure that students investigate correlations that are negative. A common misconception is that perfect correlation is only +1, where as the value of the explanatory variable increases as you move from left to right on a plot so do values of the response variable in some regular linear pattern. If the correlation is –1, the relationship is perfectly linear; however, as values of the explanatory variable increase, values of the response variable decrease.

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.
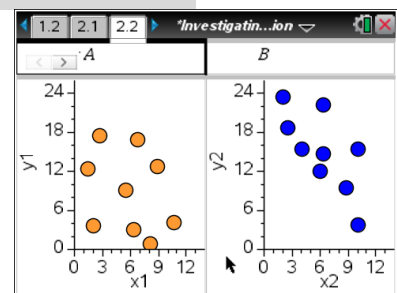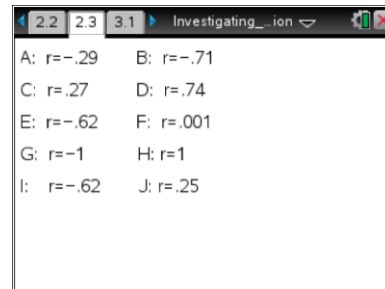
Sample Answers:

	Left Scatter Plot vs. Right Scatter Plot
Pair A, B	**B (actual, *r* = –0.29 vs *r* = –0.71)**
Pair C, D	**D (actual *r* = 0.27 vs *r* = 0.74)**
Pair E, F	**E (actual *r* = –0.62 vs *r* = 0.001)**
Pair G, H	**They are equal in strength; one correlation is –1 and the other +1.**
Pair I, J	**I (actual *r* = –0.62 vs. *r* = 0.25)**

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

Sample Answers: If the points seemed to be in a cloud with no linear pattern, the correlation seems weaker than a scatter plot where the points seemed to have a clear linear trend.

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?


```
2.2  2.3  3.1    Investigating_...ion
A: r=−.29      B: r=−.71
C: r=.27       D: r=.74
E: r=−.62      F: r=.001
G: r=−1        H: r=1
I:  r=−.62     J: r=.25
```

Sample Answers: Students might note they learned: plots with the stronger correlations have a more linear trend; "stronger" can be both positive and negative (–1 is a strong correlation even though it is negative); plots with weak correlations have less of a linear trend; it is hard to tell correlation values from looking at the plot.

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

Sample Answers: When the correlation is close to positive 1, the slope of a line describing the relationship is positive, and as the values of the explanatory variable increase so do the values of the response variable. When the correlation is close to negative 1, the slope of a line describing the relationship is negative, and as values of the explanatory variable increase, the values of the response variable decrease.

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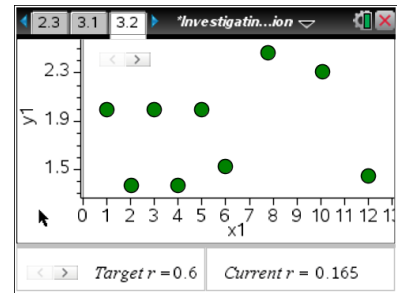
4. Move one or more of the points until you have a correlation coefficient value that approximately matches the target value. Sketch your scatter plot below. Select the arrow at the lower left of the page to produce a new target value.

    Target *r* = 0.6
    Target *r* = 0.24
    Target *r* = –0.7

    Sample Answers: Student sketches will vary.

TI-Nspire Navigator Opportunity: *Class Capture.*
See Note 2 at the end of this lesson.

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.
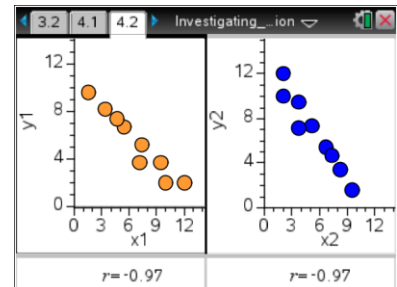
    a. What is the difference between the two plots? Give an example that illustrates your answer.

        Sample Answers: Each point in the plot on the left corresponds to a point in the plot on the right, but the coordinates are exchanged. For example, (7.097, 3.764) from the plot on the left corresponds to (3.764, 7.097) in the plot on the right.

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

        Answer: They are equal.

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?

Sample Answers: No matter what the scatter plots look like, as long as the coordinates of the points in the first plot are just reversed and plotted in the second plot, the correlation coefficient of both plots will be the same.

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?
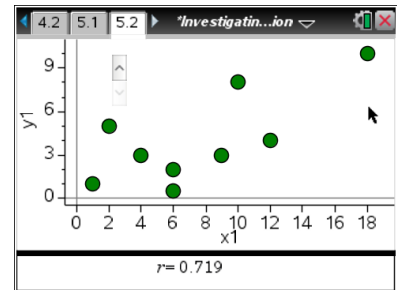
    Sample Answers: The correlation is very close to 1, which implies there is a strong positive linear association between the two variables.

    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.

    Sample Answers: A linear model could be fit to these data and be used to make predictions, but it would not be advised to have much confidence in the prediction. Points in the lower left corner are all jumbled together into a small area, and the one point (18,10) in the upper right makes it seem like there are two points—one in the lower left from the cluster and one in the upper right, which would make a linear relationship (two points make a line). But the line would not be modeling a general trend in data, and, therefore, predictions may not be reliable.

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.

    Sample Answers: The correlation is very close to –1, which indicates a strong negative association—as one variable increases, the other decreases. Though a linear model could be fit to these data, it probably would not make sense to use a line to make predictions since overall, the trend is not linear.

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.

        Sample Answers: Students might say it means a strong positive linear association. Some might indicate they cannot tell much about the association unless they see the actual plot.

    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.

        Sample Answers: Sketches will vary. See example.
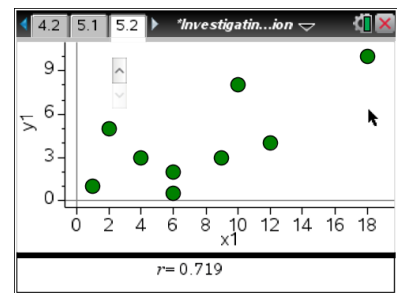
    c. Create a second scatter plot that looks like part of a parabola but has a low correlation coefficient. Make a sketch of your plot, and record the value of the correlation coefficient.

        Sample Answers: Sketches will vary. See example.
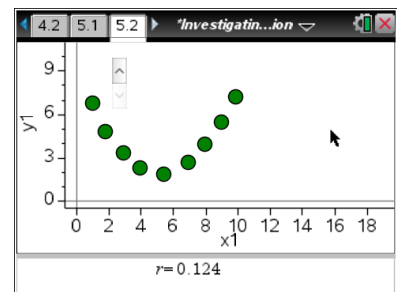
    d. Do the plots you made for parts b and c contradict your answer in part a? Why or why not?

        Sample Answers: Some students might be surprised to find that it is possible to have a high correlation for a nonlinear trend.

> **TI-Nspire Navigator Opportunity:** *Class Capture.*
> **See Note 3 at the end of this lesson.**

---

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

Answer: A graph of the relationship between two variables to verify that the relationship is linear should always precede any use of the correlation coefficient to describe the strength of a possible linear relationship.

Wrap Up

At the end of the activity, students should be able to understand:

*   The correlation coefficient measures the strength and direction of the linear association between two variables, but it is important to see the scatter plot in order to understand whether the relationship should be modeled by a linear function.
*   When two variables are highly linearly correlated, their correlation will be close to ±1; and when they have little linear correlation, the correlation coefficient will be close to 0.
*   Two variables with a high correlation coefficient might have a scatter plot that displays a nonlinear pattern.
*   Correlation is not affected by the choice of *x* or *y*, that is, by the choice of which variable is explanatory and which is response.

Assessment

The following questions can be used as an assessment either from the paper worksheet or if you have the TI-Nspire Navigator System, using Quick Poll. You might ask students to draw a sketch or two to support their answers.

Which of the following statements are always true, which are sometimes true, and which are never true. Use the answers you found in your work for questions 1 to 8 to help justify your reasoning.

a. Two variables can have a strong correlation that is negative.

Answer: Always true. A strong correlation is one that is close to either 1 or –1. See possible answers to question 2b or 3.

b. Correlation measures both the strength and direction of a linear relationship.

Answer: Always true. See the plots in questions 3 and 4 to support the claim.

c. Correlation measures both the strength and direction of a linear relationship.

Answer: Always true. See the plots in questions 3 and 4 to support the claim.

d. If the value of the correlation coefficient is near +1 or –1, you could use a linear model to predict with confidence unknown outcomes for a given value of the explanatory variable.

Answer: Sometimes true. Use plots created in question 8 where the correlations are high but the plots represent non-linear relationships and in question 1 where several of the plots are linear to support answer.

> **Teacher Tip**: Students should recognize that extrapolating from any model can be very misleading even if the relationship over the given domain seems clear.

e. The value of the correlation coefficient for the relationship between the variables (x, y) will be different than the correlation for the relationship between the variables in reverse order, (y, x).

Answer: Never true. See work on Question 5 to support answer.

TI-Nspire Navigator Opportunity: *Quick Poll (Open Response)*
See Note 4 at the end of this lesson.

TI-Nspire Navigator

Note 1

Question 2c, *Class Capture*

Display the different plots students have that have a correlation close to 1 or –-1. Discuss how these are alike or different. (Most should have a strong linear trend.)

Note 2

Question 4, *Class Capture*

Display the different plots students have that meet the target correlation coefficients. Discuss how these are alike or different. Have students experiment to see if there are any critical points that "sway" the correlation by their location.

Note 3

Question 8b, c, *Class Capture*

Display the different plots students create for questions 8b and 8c. Have students discuss how the plots support their answer to part a and what the value of the correlation coefficient without the graph indicates about the strength of a linear relationship.

Note 4

Assessment *Quick Poll (Open Response)*

Assess student understanding by using various parts of question 8 as Quick Polls. In a class discussion of the answers, have students use the tns file to create scatter plots that support their reasoning, and then have different students be Live Presenters to share their thinking with the class.