



Math Objectives

- Students will recognize that one point can influence the correlation coefficient and the least-squares regression line.
- Students will differentiate between an outlier and an influential point with respect to the least-squares regression line.
- Students will look for and make use of pattern or structure (CCSS Mathematical Practices).

Vocabulary

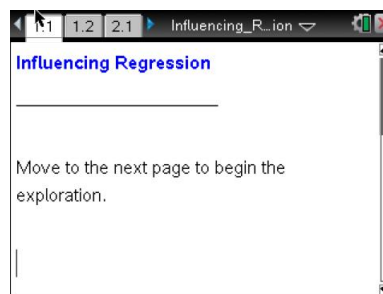
- correlation coefficient
- influential point
- least-squares regression line
- outlier
- scatterplot
- slope
- centroid (optional)

About the Lesson

- This lesson involves a least-squares regression line fit to a set of nine values.
- As a result, students will:
 - Add a value to the data set and observe how the least-squares regression line and the correlation coefficient change.
 - Determine a value that could be added to the data set to make the correlation coefficient close to zero, to make it close to one, or to change the sign of the correlation coefficient from a positive value to a negative value.
 - Determine a value that could be added to the data set without changing the equation of the least-squares regression line or the correlation coefficient.

TI-Nspire™ Navigator™ System

- Transfer a File.
- Use Quick Poll to assess students' understanding.



TI-Nspire™ Technology Skills:

- Download a TI-Nspire document
- Open a document
- Move between pages
- Grab and drag a point

Tech Tips:

- Make sure the font size on your TI-Nspire handhelds is set to Medium.
- You can hide the function entry line by pressing **ctrl** **G**.

Lesson Files:

Student Activity

Influencing_Regression_Student.pdf

Influencing_Regression_Student.doc

TI-Nspire document

Influencing_Regression.tns

Visit www.mathnspired.com for lesson updates and tech tip videos.



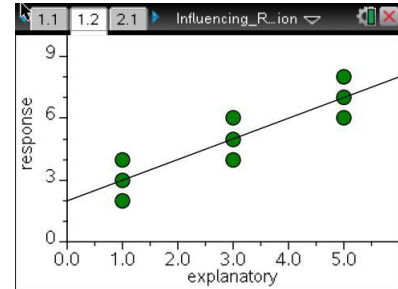
Discussion Points and Possible Answers

Teacher Tip: Students should be familiar with least-squares regression lines and the correlation coefficient before completing this lesson.

Move to page 1.2.

- a. Move the cursor to the scatterplot on Page 1.2, and click on the line. What is the equation of the least-squares regression line?

Answer: $\hat{y} = 2 + 1.0x$



- b. The correlation coefficient for these data is 0.8944. Interpret this correlation coefficient in terms of the graph.

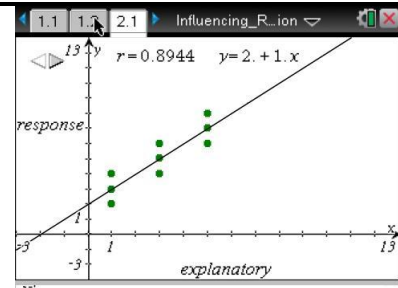
Sample Answer: There is a moderately strong, positive linear association between the explanatory and the response variables.

TI-Nspire Navigator Opportunity: Quick Poll

See Note 1 at the end of this lesson.

Move to page 2.1.

- On this page, click the arrow to add the data point (3,5) from the table to the plot.
 - Fill in the table below with the new correlation coefficient, the equation of the least-squares regression line, the position of the line, and the change from the original plot with respect to the correlation coefficient.
 - Click the arrow to replace the new point with the next ordered pair. Continue until you have completed the table.





Sample Answers:

Point added	Correlation coefficient r	Equation of the Least-Squares Regression Line $\hat{y} = a + bx$ and Position of Line	What changed from original?
none	0.8944	$\hat{y} = 2 + 1.0x$	NA
(3,5)	0.8944	$\hat{y} = 2 + 1.0x$	nothing
(5,7)	0.9063	$\hat{y} = 2 + 1.0x$	r slightly increased
(3,0)	0.6761	$\hat{y} = 1.5 + 1.0x$	r decreased; line pulled down
(4,10)	0.7883	$\hat{y} = 1.95 + 1.14x$	r decreased slightly; line pulled up
(10,0)	-0.1254	$\hat{y} = 4.91 - .11x$	r became negative; slope negative;
(10,5)	0.531	$\hat{y} = 3.70 + .35x$	r much smaller; slope decreased
(10,12)	0.9587	$\hat{y} = 2 + 1.0x$	r increased

b. Did you add any points for which the regression equation did not change? Explain.

Sample Answer: Adding (3,5), (5,7) and (10,12) did not change the regression equation. These points are on the least-squares regression line, so they satisfy the equation.

Teacher Tip: Adding a point **concurrent** with a point on the line will not change the regression equation.

c. Did you add any points for which the correlation coefficient did not change? Explain.

Sample Answer: Adding a point at (3,5) doesn't change the correlation coefficient or the equation. The point is on the regression line and seems to be in the "center" of the data.

Teacher Tip: The point (3,5) is the **centroid** of the data set: (\bar{x}, \bar{y}) . Students should know that the least-squares regression line passes through (\bar{x}, \bar{y}) and might suspect that (3,5) is that point. You might want to discuss why that point will not affect the correlation coefficient, reminding students of how the correlation coefficient is defined.

d. Did you add any points for which the slope did not change?

Sample Answer: In addition to the points mentioned in part b, adding a point at (3,0) changed the y-intercept of the equation but not the slope. Adding a point at (4,10) changed the slope by a small amount.

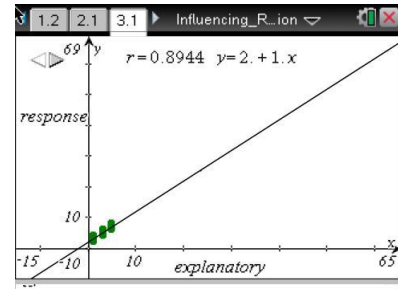


Teacher Tip: To help students understand why these statements are true, you might use the Statistics Nspired activity **Influential Outliers**.

Move to page 3.1.

The question below will repeat the process from question 2 with more extreme data values.

3. In the table below, **predict** how this new data point will change the slope of the original least-squares regression line, and its effect on the correlation coefficient (e.g., change sign, decrease a little, decrease a lot, no change, increase a little, increase a lot).
 - Check your prediction by using the arrow to add the point in your TI-Nspire document.
 - Record the new equation in the table below.
 - Repeat the process with the remaining points.



Sample Answers:

Point added	Change in slope	Change in correlation coefficient	Actual correlation coefficient	Equation of least-squares regression line
none	NA	NA	0.8944	$\hat{y} = 2 + 1.0x$
(50,0)	decreased a lot	Become negative and decreased some	-0.5769	$\hat{y} = 5.22 + -0.09319x$
(50,5)	decreased and close to 0	close to 0	0.0977	$\hat{y} = 4.91 + 0.01193x$
(50,53)	become steeper and close to 1	Close to 1.0	0.9986	$\hat{y} = 1.94 + 1.02102x$

Use the information in the tables from Questions 2 and 3 to answer the questions below.

4. Points that lie within the x-range of the data and change the y-intercept but do not change the slope of the least-squares regression line are called *outliers in regression*. Which points did you add that would be considered outliers?

Sample Answer: (3,0) and (4,10) both changed the y-intercept but had little or no effect on the slope.



Teacher Tip: Outliers tend to have large residuals—they do not pull the line up or down to include them and are consequently farther from the line.

5. Points that do not lie within the x -range of the data and change the slope of the regression line are called *influential points*. Which points did you add that would be considered influential?

Sample Answer: The points (10,0), (10,5), (50,0) and (50,5) all pulled the regression lines down, changing the slope of the regression lines and would be considered influential.

Teacher Tip: Influential points tend to have small residuals because the line changes to include them.

Wrap Up

Upon completion of the lesson, the teacher should ensure that students are able to understand:

- A single point can influence the correlation coefficient and the regression line.
- How to distinguish between an outlier and an influential point with respect to the least-squares regression line.

Assessment

1. Where would you place a point to make the correlation coefficient as close to 0 as possible?

Sample answer: To make the correlation coefficient close to 0: place a point far away from the other data values in the x -direction with a y -value = \hat{y} or within the domain of the x -values but with an extreme y -value

2. Where would you place a point to make the correlation coefficient as close to 1 as possible?

Sample answer: To make the correlation coefficient close to 1: place a point far away from the other data values in the x -direction with a y -value on the given line.

3. Where would you place a point to not change the correlation coefficient or the equation of the least-squares regression line?

Sample answer: Place the point at the center of the data values—the centroid, (\bar{x}, \bar{y}) .



TI-Nspire Navigator

Note 1

Name of Feature: Quick Poll

Use Quick Poll to determine how well students understood the lesson. You might ask them to send you points in response to each of the questions in the Assessment above.

Adapted from Matching Descriptions to Scatter Plots, (1996). *Activity Based Statistics*, by Richard Scheaffer, Mrudulla Gnanadesikan, Ann Watkins, & Jeffrey Witmer. Springer.