Math Objectives

- Students will identify outliers that are influential with respect to the least-squares regression line.
- Students will describe the role of the location of a point relative to the other data in determining whether that point has influence on the least-squares regression.
- Student will use appropriate tools strategically (CCSS Mathematical Practices).

Vocabulary

- centroid
- influence
- least-squares regression line
- residual
- scatter plot

About the Lesson

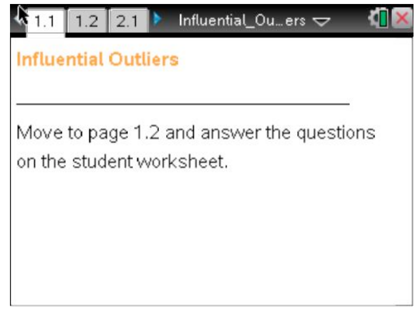
- Students will investigate the relationship between the location of an outlier relative to the main pattern of a scatter plot and the influence that point exerts on the least-squares regression line.
- Students will manipulate a designated point and observe how its location alters an existing least-squares line.
- Students will manipulate a line fixed through the centroid of the scatter plot and observe how various slopes affect the sum of squared residuals.

Vocabulary

- Students should be familiar with outliers and univariate data.

TI-Nspire™ Navigator™ System

- Use Screen Capture to compare student results.
- Send .tns file to students.
- Use Quick Poll to determine student 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.

Lesson Materials:

Student Activity

- Influential_Outliers_Student.pdf
- Influential_Outliers_Student.doc

TI-Nspire document

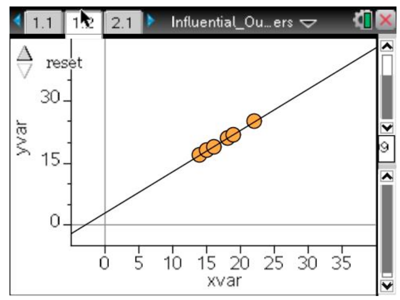
- Influential_Outliers.tns

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

Discussion Points and Possible Answers

Tech Tip: If students experience difficulty dragging the point, check to make sure that they have moved the arrow until it becomes a hand ( ) getting ready to grab the point. Also, be sure that the word *point* appears. Then press ctrl to grab the point and close the hand ( ). When finished moving the point, press esc to release the point.

Move to page 1.2.



1. Page 1.2 displays a scatter plot of six points together with the associated least-squares regression line (LSRL).

   Note: All the points lie directly on the given line.

   A point is called an **outlier** if it fails to fit the overall pattern of the set to which it belongs. With univariate data, the "1.5 IQR rule" can be used to decide whether a point is an outlier or not. There is no such rule for bivariate data. Outliers are in the eyes of the beholder; it is a subjective decision.

   In this activity, you will be intentionally creating an outlier and then experimenting with how it affects a fitted line.

   Grab the rightmost point (22, 25) and drag it vertically up and down. (Be sure the line itself not highlighted. If it is, click in an open space to make the equation and highlight disappear.) Describe how the LSRL changes.

   <u>Answer:</u> The LSRL "follows" the point up and down the page. It almost seems like the line is spinning around a point in the middle of the data. When the point is dragged down, the line rotates clockwise, and when the point is dragged up, the line rotates counterclockwise.

2. Use the up arrow to reset and return to the original configuration of points. Drag the same point to the right so that its *x*-coordinate is around 28. (Be sure the line itself is not highlighted.) Notice how the LSRL changes as the point is dragged vertically up and down.

a. How is the behavior of the LSRL similar to what you observed previously?

Answer: The LSRL still “follows” the point up and down the page by rotating around the center of the plot.

b. How is the behavior of the LSRL different from what you observed previously?

Answer: The LSRL still “follows” the point up and down the page, but more closely now.

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

3 Drag the same point to the right so the *x*-coordinate is around 35. Describe how the LSRL responds to dragging the point vertically up and down. Summarize the effect on the LSRL as the point moves further to the right of the main cluster.

Answer: The LSRL seems always to follow the moving point up and down, but it does so more and more closely as the point moves further to the right.

4. a. Use the arrow to reset the points and return to the original plot. Suppose you were to grab the point (16, 19) and drag it vertically up and down. Before you do so, predict how you think the LSRL will behave. Then drag the point and check your conjecture.

Sample Answer: The line changes but not by very much; it shifts up or down in the direction of the point but generally keeps the same overall trend in the set of data.

b. Repeat part a using the point (14, 17).

Sample Answer: The line follows the point again like it did when I grabbed the point on the right (22, 25).

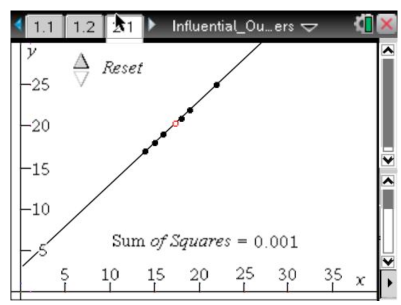
5. Use the up arrow to reset the screen to the original display. Predict what will happen to the LSRL when you drag the point farthest to the right in a circle around the main cluster of other points, starting from lower right and going around in a counterclockwise direction. Check your prediction by dragging the point.

   <u>Sample Answer:</u> The line does not always stay with the point. It rotates to follow the point around as long as the point is farther to the right or to the left than the rest of the data, that is for *x*-values of the coordinate of the point that are much larger or smaller than the *x*-values of the cluster of the other points. When the dragged point's *x*-coordinate is within the domain of the other points, the line actually rotates in the clockwise direction and changes direction but does not follow the point.

6. Summarize your findings about the location of an outlying point relative to the main cluster of points and the amount and kind of influence it exerts on the LSRL.

   <u>Answer:</u> A point exerts the most influence when it is at the extreme left or right end and is "out of line" with the other points of the plot.

Move to page 2.1.



The centroid for a set of bivariate data (*x*, *y*) is the point whose coordinates are the means of each set of data (x,y).

7. a. Estimate the coordinates of the centroid for the data represented in the plot on 2.1.

   <u>Sample Answers:</u> Students' estimates will vary, but will probably be around (17.5, 20.5) if they see the open circle or, more generally, somewhere around (17, 20)..

> **Teacher Tip:** The centroid is the open circle on page 2.1 in the .tns file.
>
> *Note that in a real context when a point is moved, the centroid will also change. This does not happen in this activity to keep the focus on the intended message about influential points.*

b. Grab the line near one end and rotate it a bit. Does it appear that it is rotating around the centroid?

Answer: Yes. The point of rotation seems pretty much in the middle of the points in the plot.

Tech Tip: To restore the line to the way it originally looked, press ctrl Z several times until it has been restored to the original position.

8. As you changed the slope, you should have seen something new in the plot. Carefully describe what you saw. What information does this new display give you?

   Answer: As the line is rotated, some squares showed up. Each square has one side that goes from the line to one point in the scatter plot. There is also something called "sum of squares" that changes as the slope changes. The bigger the squares look, the bigger this number, so maybe it's the sum of the areas.

9. Suppose you were to drag the rightmost point of the scatter plot down to somewhere near (22, 18). Predict what will happen to the sum-of-squares measure of closeness if the line does not move at all. Give a specific numerical prediction if possible. Then check your answer by actually dragging the point.

   Sample Answer: The sum will be pretty big. The residuals for the first five points won't change, since the line does not change. But the point I drag will be off the line by about 7 units, so its square should be about 49 units in area. Since that's the only point off the line, the sum should be around 49. When I moved the point, it worked like I thought it should. You cannot see the whole square.

Teacher Tip: Students should not try to get the point exactly at (22, 18) but close enough to see that the sum of squares is about 49.

10. Based solely on your current plot, discuss how you would have to change the slope of your line to make it become the LSRL for these data. Explain your thinking carefully.

    Answer: I think the slope needs to be smaller. When the point was dragged down, it made a very large residual square. If I rotate the line a little clockwise, its area will decrease a lot and the areas of the other squares will all get larger, but not by much since they are near the centroid.

11. Adjust the movable line so that it has as small a sum of squares as possible. Comment on your prediction in Question 10.

    Sample Answer: I was able to reduce the sum of squares from 49 to about 22. I was right about the direction of rotation, but it actually rotated further than I thought it might.

    > **Teacher Tip:** Before moving on to Question 12, students need to restore page 2.1 to the way it originally looked.

12. Predict what will happen with the residual squares if you were to drag the rightmost point of the scatter plot to somewhere around (17, 25) using the original line. Return to the original configuration and verify your prediction by moving the point.

    Answer: Again, since only one point changes, the sum of squares will reflect only that one point. That point will be about 5 points above the line, so the sum should be around 25.

13. a. Discuss how you would have to change the slope of your line to make it become the LSRL for these data. Then rotate the line to try to minimize the sum of the squared residuals. Explain what you found.

    Sample Answer: I thought I would have to make the slope greater to move the line more toward the new point. But when I tried this, the sum of the squares increased, first by a little and then by a lot.

    b. Make a conjecture about why the centroid might have some relationship to your answer in part a.

    Sample Answer: The new point has an *x*-coordinate nearly the same as the *x*-coordinate of the centroid. Rotations around the centroid will have very little effect on the residual of the "bad" point; the square stays pretty much the same. All you can do is try to decrease some of the other squares, but they were pretty small to start. A point that is an "outlier" farther to the right (or left) of the main cluster of points will have a large squared residual, but when the line is moved, that residual can be made to decrease or increase by quite a lot..

14. From all your previous work, explain where outliers seem to have the most influence on the slope of a least-squares regression line.

    Answer: Points that seem to have the most influence on the slope of the LSRL are those whose *x*-coordinates are "outside" the domain of most of the points in the cluster.

Teacher Tip: Class discussion might raise the following points. The centroid tends to be in or near the main cluster of data. Generally, if a point is outside the domain and does not follow the pattern of the remaining data, then it will influence the slope of the LSRL because the line is pulled towards that point. If a point is within the domain but does not follow the pattern, then the LSRL will be shifted vertically towards that point with little change in the slope.

TI-Nspire Navigator Opportunity: ***Quick Poll***
See Note 2 below.

Wrap Up:

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

- The difference between an outlier and influential point.
- Outliers and influential points can have a major impact on the model equation.
- Influential points have a larger effect on the regression equation and a smaller effect on the *r*-value.
- Outliers have a larger effect on the *r*-value and a smaller effect on the regression coefficient.

TI-Nspire Navigator

Note 1

Questions 1, 2, *Screen Capture*

You might want to use Screen Capture to verify students are moving the point correctly, and to inform the discussion of the answers to the questions 1 and 2.

All Questions, *Quick Poll*

Use Quick Poll to check student responses to the questions presented, or additional related questions.