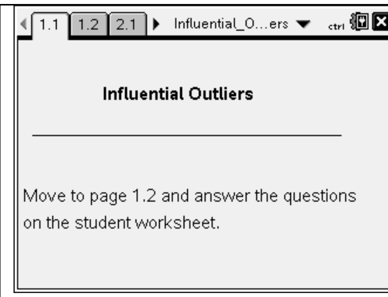
Influential Outliers
Student Activity

Name ________________
Class ________________

Open the TI-Nspire document *Influential_Outliers.tns.*

How can you tell whether an individual outlier is influential with respect to the least-squares regression line? This lesson investigates that question.



Note: Press the Reset up arrow (▲) to return to the original configuration of points. You can also press ctrl Z to return the data to the original state.

.

Move to page 1.2.

Press ctrl ► and ctrl ◄ to navigate through the lesson.

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

   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; identifying a point as an outlier is a subjective decision.

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

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

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

Note: Click ▲ to reset and return to the original configuration of points.

2. Drag the same point to the right so that its *x*-coordinate is around 28. (Be sure the line itself is not highlighted.) Describe 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?

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

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 farther to the right of the main cluster.

4. a. 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.

Prediction:

b. Repeat question 4a using the point (14, 17).

Prediction:

5. Click ▲ 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 cluster of other points, starting from lower right and going around in a counterclockwise direction. Check your prediction by dragging the point.

Prediction:

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

Move to page 2.1.

The ▲ will reset the points to their original configuration. To return the line to its original position, press ctrl Z.

7. The centroid for a set of bivariate data (x, y) is the point whose coordinates $(\bar{x}, \bar{y})$ are the means of each set of data.
   a. Estimate the coordinates of the centroid for the data represented in the plot on page 2.1.

   b. Grab the line near one end and rotate it a bit. Describe any relationship that seems to exist between the centroid and the line as it is rotated.

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?

9. Suppose you were to drag the right-most 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.

   Prediction:

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.

    Prediction:

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

Click the Reset ▲ to restore the points to their original position. Press ctrl Z to restore the line to its original configuration until page 2.1 has all the points directly on the line.

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 to verify your prediction by moving the point.

    Prediction:

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.

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

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