Influential Outliers – r^2

TIMATH.COM: STATISTICS

TEACHER NOTES

Math Objectives

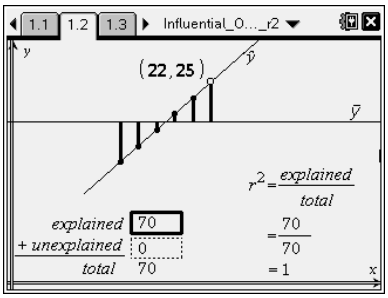
- Be able to identify outliers that are influential with respect to the variation of the data values.
- Describe the role of the location of a point relative to the other data in determining whether that point has influence on the coefficient of determination.

Vocabulary

- scatter plot
- linear regression line
- influence
- deviation
- variation
- sum of squares

About the Lesson

- This lesson is a follow-up activity to the AC lesson *Influential Outliers*.
- Students will manipulate a designated point and observe how its location alters the explained and unexplained deviations of a data value.
- 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 coefficient of determination.



TI-Nspire™ Technology Skills:

- Download TI-Nspire document
- Open a document
- Move between pages
- Grab and drag a point (scatter plot)

Tech Tips:

- *Make sure the **font size** on your TI-Nspire handheld is set to Medium.*
- *The **Class Notes** page at the end of the .tns file should not be deleted.*

Lesson Materials:

Student Activity

Influential_Outliers_r2_Student.pdf

Influential_Outliers_r2_Student.doc

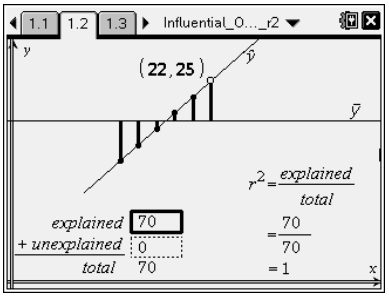
TI-Nspire document

Influential_Outliers_r2.tns

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

Discussion Points and Possible Answers:

TI-Nspire Problem/Page 1.2



Tech Tip: *If students experience difficulty dragging a point, check to make sure that they have moved the cursor (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.*

1. Grab the rightmost point in the plot, (22, 25), and drag it vertically downward. As you move the point, observe the dotted and bold lines. What happens to the explained and unexplained deviations?	**The explained deviation (bold lines) becomes shorter, and the unexplained deviation (dotted lines) appears.** ***Teacher Tip:*** *Students should only be focusing on the lines extending from the data points. The information presented at the bottom of the screen should be ignored until Question 4.*
2. Drag the point in a circle around the other data points. a. When is the deviation for the points mostly explained?	**When the point is near the top right or the bottom left of the data, it is mostly explained (i.e., the length of the bold lines is greater than the length of the dotted lines).**
b. When is the deviation for all of the points mostly unexplained?	**When the point is near the top left or the bottom right of the data, it is mostly unexplained (i.e., the length of the bold lines is less than the length of the dotted lines).**
3. A point is called an **outlier** if it fails to fit the overall pattern of the set to which it belongs. How do your observations in Question 2 correspond to when the point is considered an outlier?	**When the point is an outlier, the deviation is mostly unexplained. When the point is not an outlier, the deviation is mostly explained.**
4. Why do you think 100% of the variation can be explained?	**The regression line looks like it fits the data perfectly.** **The total variation is equal to the explained variation (i.e., only the bold lines are showing).**
5. What would it mean if 0% of the variation could be explained?	**If 0% is explained, then 100% of the total variation is unexplained. This would mean that a linear relationship does not represent the data.**

6. Drag the point in a circle around the other data points.	
a. When is r^2 closest to a value of 1?	**when the point is near the top right or the bottom left of the data (i.e., the length of the bold lines is greater than the length of the dotted lines)**
b. When is r^2 closest to a value of 0?	**when the point is near the top left or the bottom right of the data (i.e., the length of the bold lines is less than the length of the dotted lines)**
7. How do your observations from Question 6 correspond to when the point is considered an outlier?	**When the point is an outlier, r^2 is closer to 0. When the point is not an outlier, r^2 is closer to 1.**
8. Summarize in your own words the effect an outlier has on the coefficient of determination and the percent of total variation that can be explained by the linear relationship.	**As a data point becomes a more significant outlier, it decreases the value of the coefficient of determination and decreases the percentage of the total variation that can be explained.**

Wrap Up:

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

- The difference between deviation and variation.
- Outliers can have a major impact on the explained and unexplained deviation.
- Outliers can have a major impact on the r^2-value and percent of variation that can be explained.