



Lesson 18: Analyzing Residuals

Student Outcomes

- Students use a graphing calculator to construct the residual plot for a given data set.
- Students use a residual plot as an indication of whether the model used to describe the relationship between two numerical variables is an appropriate choice.

Lesson Notes

In this lesson, students continue to construct residual plots. Residual plots are analyzed to determine if the points are scattered at random around the horizontal line (indicating a linear relationship) or if the points have a pattern (indicating a nonlinear relationship).

Classwork

The previous lesson shows that when data is fitted to a line, a scatter plot with a curved pattern produces a residual plot that shows a clear pattern. You also saw that when a line is fit, a scatter plot where the points show a straight-line pattern results in a residual plot where the points are randomly scattered.

Example 1 (10 minutes): The Relevance of the Pattern in the Residual Plot

MP.2

This is another opportunity to reinforce *why* the particular patterns in the scatter plots result in the corresponding patterns in the residual plots. Pose these questions:

- Why does the first scatter plot result in an *arch* shape in the residual plot?
 - *The points in the scatter plot are not in a straight line.*
- Why does the second scatter plot result in a *U* shape in the residual plot?
 - *The points in the scatter plot are not in a straight line.*
- Why does the third scatter plot result in a random scatter of points in the residual plot?
 - *The points seem to be linear.*

Explain how the patterns in the residual plot show us whether a linear model is a good fit for the data.

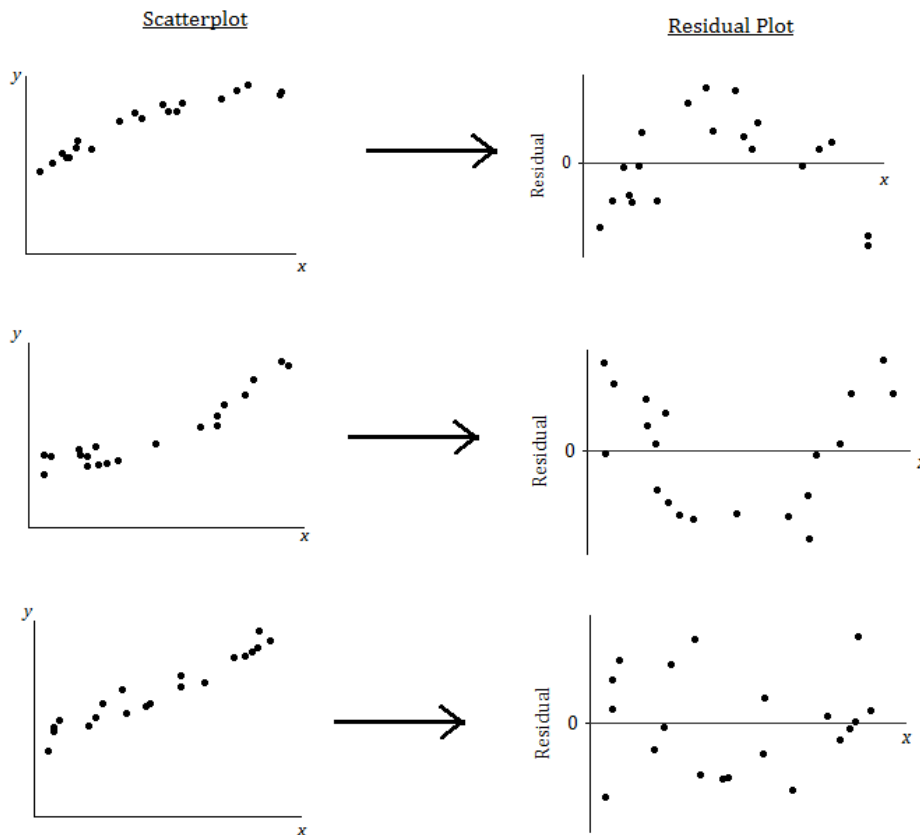
- If the points in a residual plot are random, a linear model is the best fit.
- If the points have a pattern, a line is not the best fit.

Pose the following question to students to connect back to previous lessons:

- If a linear model is not the best fit for the first and second scatter plots, what type of model might be more appropriate?
 - *Exponential or (possibly) quadratic.*

Example 1: The Relevance of the Pattern in the Residual Plot

Our previous findings are summarized in the plots below:



What does it mean when there is a curved pattern in the residual plot?

Allow students to answer this question. Discuss with students that a curved pattern in the residual plot indicates that the relationship would be better described by a nonlinear function. When you see a curved pattern in the residual plot, you should investigate nonlinear models rather than using a line or model to describe the relationship. (See the first two scatter plots and residual plots in the figure above, which show a curved pattern in both the scatter plot and the residual plot.)

What does it mean when the points in the residual plot appear to be scattered at random with no visible pattern?

Again, allow students to answer this question. Discuss with students that when the relationship is approximately linear, the points in the residual plot will be scattered at random around the horizontal line at zero. This indicates that a linear model is an appropriate way to describe the relationship. (See the third scatter plot and residual plot in the figure above, which shows a linear pattern in the scatter plot and no pattern in the residual plot.)

Why not just look at the scatter plot of the original data set? Why was the residual plot necessary? The next example answers these questions.

Example 2 (10 minutes): Why Do You Need the Residual Plot?

Ask students the following:

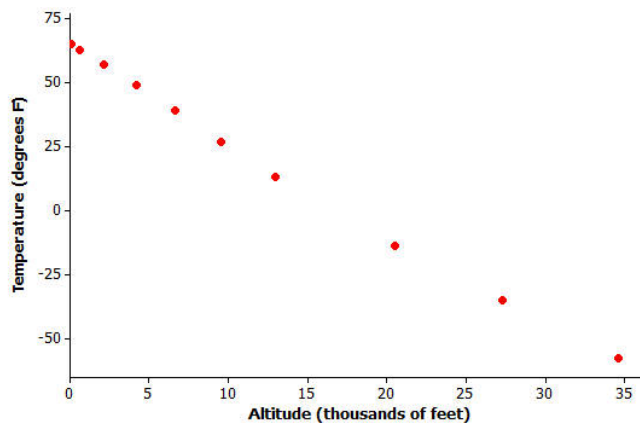
MP.3

- What does the residual plot for Example 2 indicate about using a linear model?
 - *It is not a good fit. There is a nonlinear relationship.*

The point is that because the scale on the vertical axis of the residual plot exaggerates the vertical deviations, the residual plot shows detail of the pattern in the residuals that is not easily seen in the scatter plot.

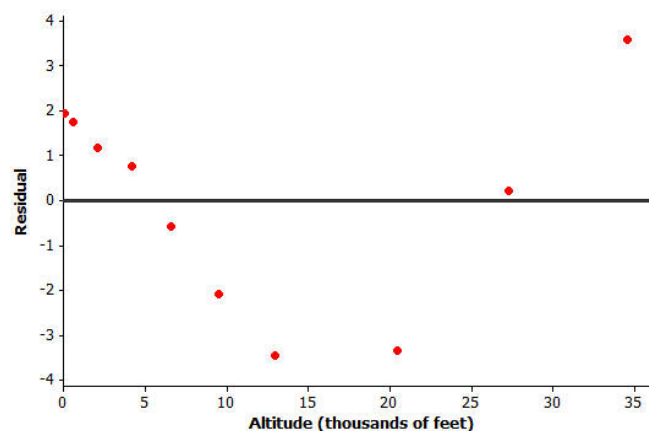
Example 2: Why Do You Need the Residual Plot?

The temperature (in degrees Fahrenheit) was measured at various altitudes (in thousands of feet) above Los Angeles. The scatter plot (below) seems to show a linear (straight line) relationship between these two quantities.



Data source: *Core Math Tools*, www.nctm.org

However, look at the residual plot:



There is a clear curve in the residual plot. So what appeared to be a linear relationship in the original scatter plot was, in fact, a nonlinear relationship.

How did this residual plot result from the original scatter plot?

The residuals are the vertical deviations from the least-squares line in the original scatter plot. Because of the change of scale, tiny vertical deviations can be shown as much larger distances from the zero line in the residual plot. Therefore, a very subtle curvature in the original relationship can be seen more clearly in the residual plot. This is why you draw residual plots.

Exercises 1–3 (15 minutes): Volume and Temperature

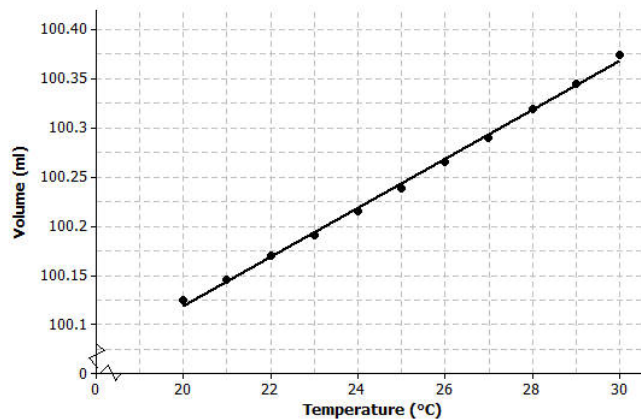
Let students work with a partner on Exercises 1–3. Then discuss and confirm answers as a class.

Exercises 1–3: Volume and Temperature

Water expands as it heats. Researchers measured the volume (in milliliters) of water at various temperatures. The results are shown below.

Temperature (°C)	Volume (ml)
20	100.125
21	100.145
22	100.170
23	100.191
24	100.215
25	100.239
26	100.266
27	100.290
28	100.319
29	100.345
30	100.374

- Using a graphing calculator, construct the scatter plot of this data set. Include the least squares line on your graph. Make a sketch of the scatter plot including the least squares line on the axes below.



Although it might be clear from the scatter plot given here that there is a curved relationship between the two variables, this detail cannot however be seen in the low-resolution graph produced by a calculator.

2. Using the calculator, construct a residual plot for this data set. Make a sketch of the residual plot on the axes given below.



3. Do you see a clear curve in the residual plot? What does this say about the original data set?

Yes, there is a clear curve in the residual plot. This indicates a nonlinear (curved) relationship between volume and temperature.

Closing (5 minutes)

Discuss the Lesson Summary with students.

Lesson Summary

- After fitting a line, the residual plot can be constructed using a graphing calculator.
- A curve or pattern in the residual plot indicates a nonlinear relationship in the original data set.
- A random scatter of points in the residual plot indicates a linear relationship in the original data set.

Exit Ticket (5 minutes)

Name _____

Date _____

Lesson 18: Analyzing Residuals

Exit Ticket

1. If you see a clear curve in the residual plot, what does this say about the original data set?
2. If you see a random scatter of points in the residual plot, what does this say about the original data set?

Exit Ticket Sample Solutions

1. If you see a clear curve in the residual plot, what does this say about the original data set?

A clear curve in the residual plot shows that the variables in the original data set have a nonlinear relationship.

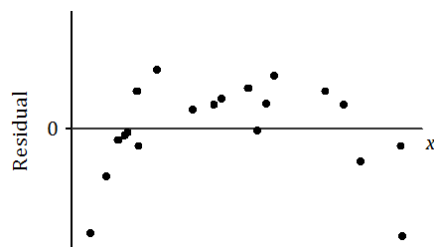
2. If you see a random scatter of points in the residual plot, what does this say about the original data set?

A random scatter of points in the residual plot shows that a straight line is an appropriate model for the relationship between the two variables in the original data set.

Problem Set Sample Solutions

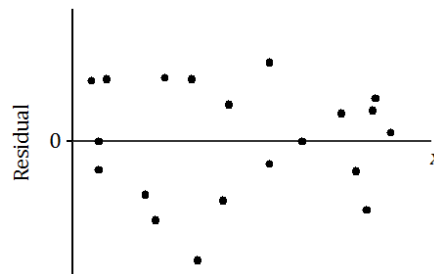
1. For each of the following residual plots, what conclusion would you reach about the relationship between the variables in the original data set? Indicate whether the values would be better represented by a linear or a nonlinear relationship.

a.



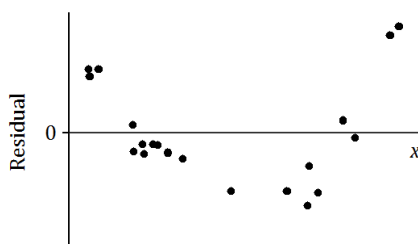
There is a nonlinear relationship between the variables in the original data set.

b.



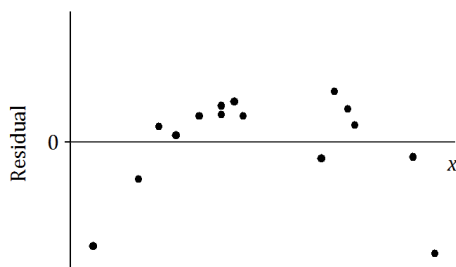
There is a linear relationship between the variables in the original data set.

c.

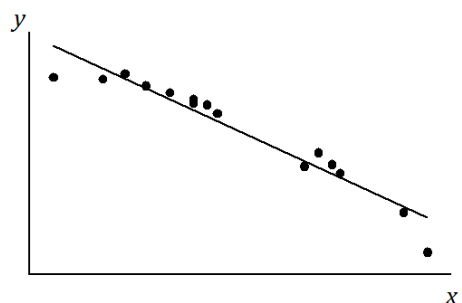
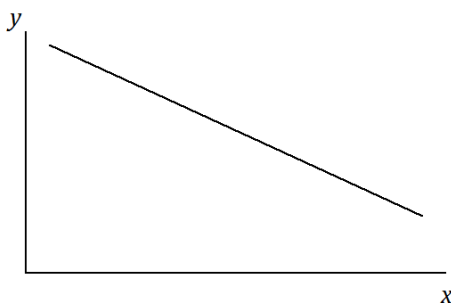


There is a nonlinear relationship between the variables in the original data set.

2. Suppose that after fitting a line, a data set produces the residual plot shown below.



An incomplete scatter plot of the original data set is shown below. The least squares line is shown, but the points in the scatter plot have been erased. Estimate the locations of the original points and create an approximation of the scatter plot below.



This problem should challenge students. There are many correct answers. The student should show points in the scatter plot whose vertical deviations from the given least squares line are in proportion to the values of the residuals in the residual plot. A student's answer is only acceptable if it shows a general pattern like the one shown above. Confirm that each student's points could result in (approximately) the given least squares line.