Lesson 17A: Analyzing Residuals

Classwork

Example 1: Predicting the Pattern in the Residual Plot

Suppose you are given a scatter plot and least-squares line that looks like this:



The residual plot has an arch shape, like this:



Why is looking at the pattern in the residual plot important?

What does the pattern tell us?

Example 2: 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?

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

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: 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 (curved) relationship.

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

Example 3: The Meaning of Residuals

Suppose that you have a scatter plot and that you have drawn the least-squares line on your plot. Remember that the residual for a point in the scatter plot is the vertical distance of that point from the least-squares line.

In the previous lesson, you looked at a scatter plot showing how fuel efficiency was related to curb weight for five compact cars. The scatter plot and least-squares line are shown below.



Consider the following questions:

* What kind of residual will Point A have?
* What kind of residual will Point B have?
* What kind of residual will Point C have?

You also looked at the residual plot for this data set:

Problem Set 17A

**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 |

1. Construct the scatter plot of this data set. 
2. The equation of the least square line is y = .025x + 99.621, include the least-square line on your graph above.
3. Calculate the residuals. Write all the residuals in the table given below round to the nearest thousandths:

|  |  |  |  |
| --- | --- | --- | --- |
| **Temperature (°C)** | **Volume (ml)** | **Predicted Value** | **Residuals** |
| 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 |  |  |

1. Construct a residual plot for this data set by completing the table below. Make a sketch of the residual plot on the axes given below.



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

Lesson 17B: Analyzing Residuals

Example 1: Using a Graphing Calculator TI-84plus to Construct a Residual Plot

|  |  |
| --- | --- |
| **Shoe Length (*x*)** | **Height (*y*)** |
| inches | inches |
| 8.9 | 61 |
| 9.6 | 61 |
| 9.8 | 66 |
| 10.0 | 64 |
| 10.2 | 64 |
| 10.4 | 65 |
| 10.6 | 65 |
| 10.6 | 67 |
| 10.5 | 66 |
| 10.8 | 67 |
| 11.0 | 67 |
| 11.8 | 70 |

In an earlier lesson you looked at a data set giving the shoe lengths and heights of 12 adult women. This data set is shown in the table below.

Use a calculator to construct the scatter plot (with least-squares line) and the residual plot for this data set.

Calculation of the equation of the least-squares line:

1. From the home screen, go to the statistics editor by pressing STAT, ENTER.
2. Enter the *x*-values into L1 and the $y$-values into L2.
3. Press 2nd, QUIT to return to the home screen.
4. Press STAT, select CALC, select LinReg($a+bx$), and press ENTER.
5. The $y$-intercept, $a$, and the slope, $b$, of the least-squares line are displayed on the screen

Construction of scatter plot:

1. From the home screen press 2nd, STAT PLOT.
2. Select Plot1 and press ENTER.
3. Select “On”, under “Type” choose the first (scatter plot) icon, for Xlist enter L1, for Ylist enter L2, and under “Mark” chose the first (square) symbol.
4. Press 2nd, QUIT to return to the home screen.
5. Set your Window to appropriate x an d y values
6. STAT, CALC, choose 8 enter, scroll to StoreRegEq:, press VARS, Y-VARS, Enter. Enter, Enter, Enter,
7. Press Zoom, select ZoomStat (option 9), ENTER.
8. The scatter plot and the least-squares line are displayed.

Construction of residual plot:

1. From the home screen, press 2nd, STATPLOT.
2. Select Plot2 and press ENTER.
3. Select “On”, under “Type” choose the first (scatter plot) icon, for Xlist enter L1, for Ylist enter RESID, and under “Mark” choose the first (square) symbol. (“RESID” is accessed by pressing 2nd, LIST, selecting NAMES, scrolling down to RESID, and pressing ENTER.)
4. Press 2nd, QUIT to return to the home screen.
5. Press Y=.
6. First, deselect the equation of the least-squares line in Y1 by going to the “=” sign for Y1 and pressing ENTER. Then deselect Plot1 and make sure that Plot2 is selected.
7. Press Zoom, select ZoomStat (option 9), press ENTER.
8. The residual plot is displayed.

Lesson Summary

* After fitting a line, the residual plot can be constructed using a graphing calculator.
* A pattern in the residual plot indicates that the relationship in the original data set is not linear.
* A curve or pattern in the residual plot indicates a curved (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.

1. Consider again a data set giving the shoe lengths and heights of 10 adult men. This data set is shown in the table below. Use your calculator construct the scatter plot of this data set. Include the least-squares line on your graph. To the nearest thousandths give the least squares line equation.

|  |  |
| --- | --- |
| **Shoe Length (*x*)** **inches** | **Height (*y*)** **inches** |
| 12.6 | 74 |
| 11.8 | 65 |
| 12.2 | 71 |
| 11.6 | 67 |
| 12.2 | 69 |
| 11.4 | 68 |
| 12.8 | 70 |
| 12.2 | 69 |
| 12.6 | 72 |
| 11.8 | 71 |

2. Explain what the slope of the least-squares line indicates about shoe length

and height.

3. Use your calculator to construct the residual plot for this data set.

Make a sketch of the residual plot on the axes given below. .



4. Does the scatter of points in the residual plot indicate a linear relationship in the original data set?

Problem Set 17B

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 non-linear relationship. Justify you answer.

* 1.







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:

