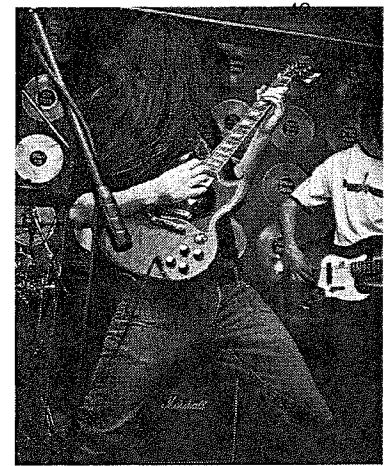


8.8 Rockin' the Residuals

A Solidify Understanding Task

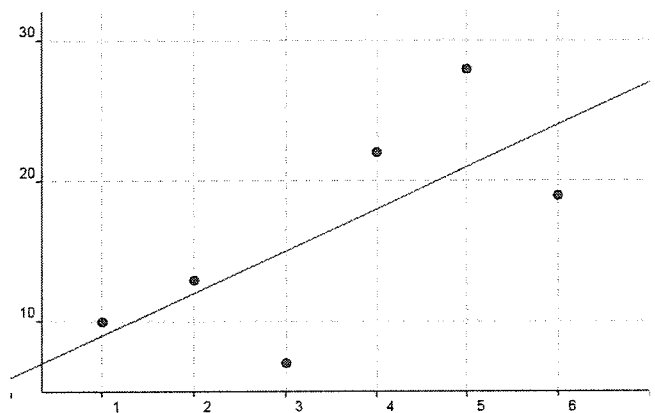


The correlation coefficient is not the only tool that statisticians use to analyze whether or not a line is a good model for the data. They also consider the residuals, which is to look at the difference between the observed value (the data) and the predicted value (the y-value on the regression line). This sounds a little complicated, but it's not really. The residuals are just a way of thinking about how far away the actual data is from the regression line.

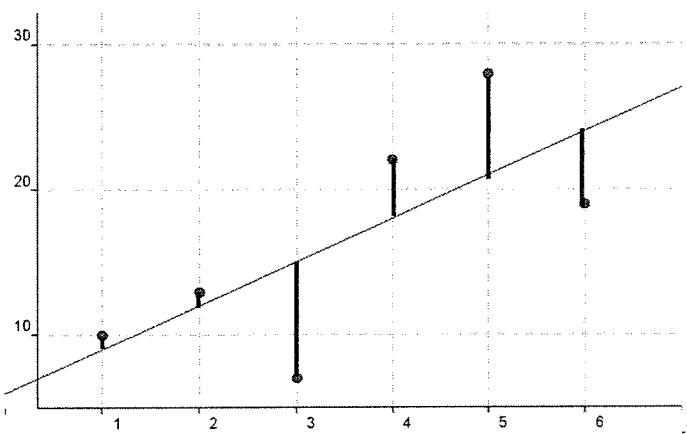
Start with some data:

x	1	2	3	4	5	6
y	10	13	7	22	28	19

Create a scatter plot and graph the regression line. In, this case the line is $y = 3x + 6$.



Draw a line from each data point to the regression line, like the segments drawn from each point below.

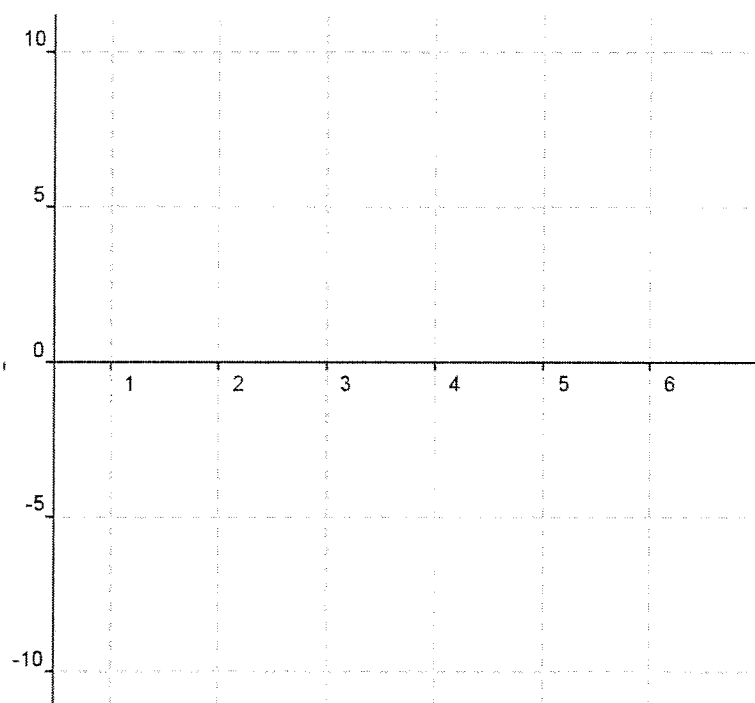


© 2012 Mathematics Vision Project | MVP

In partnership with the Utah State Office of Education

Licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license

1. The residuals are the lengths of the segments. How can you calculate the length of each segment to get the residuals?
2. Generally, if the data point is above the regression line the residual is positive, if the data point is below the line, the residual is negative. Knowing this, use your plan from #1 to create a table of residual values using each data point.
3. Statisticians like to look at graphs of the residuals to judge their regression lines. So, you get your chance to do it. Graph the residuals here.



© 2012 Mathematics Vision Project | MVP

In partnership with the Utah State Office of Education

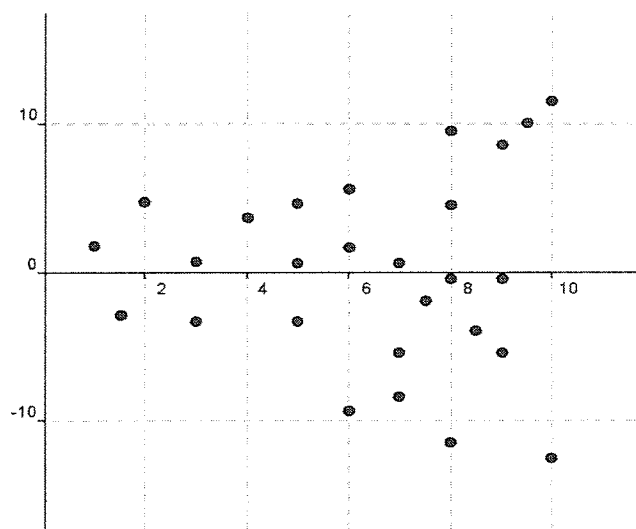
Licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license

Now, that you have constructed a residual plot, think about what the residuals mean and answer the following questions.

4. If a residual is large and negative, what does it mean?
5. What does it mean if a residual is equal to 0?
6. If someone told you that they estimated a line of best fit for a set of data points and all of the residuals were positive, what would you say?
7. If the correlation coefficient for a data set is equal to 1, what will the residual plot look like?

Statisticians use residual plots to see if there are patterns in the data that are not predicted by their model. What patterns can you identify in the following residual plots that might indicate that the regression line is not a good model for the data? Based on the residual plot are there any points that may be considered outliers?

8.

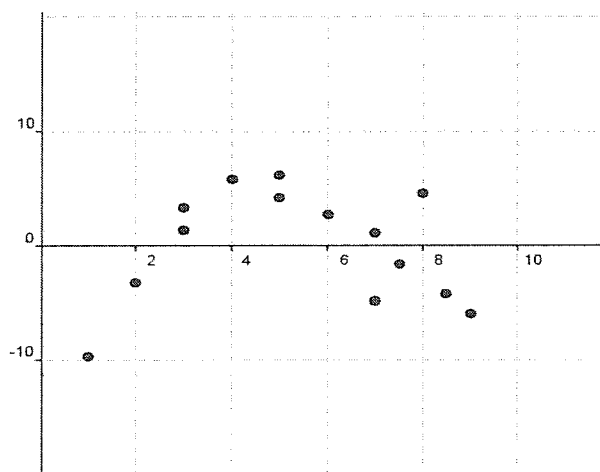


© 2012 Mathematics Vision Project | MVP

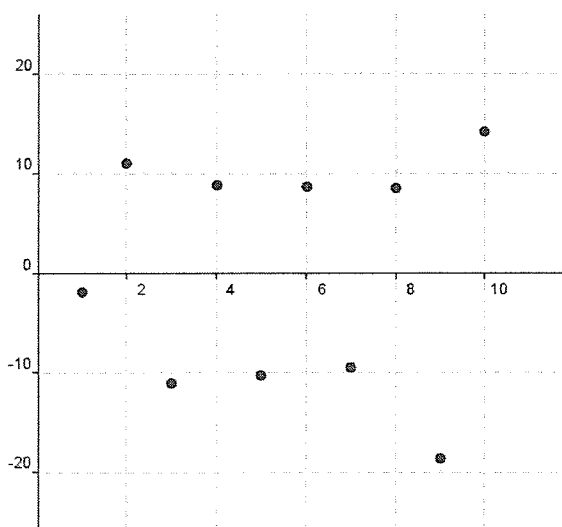
In partnership with the Utah State Office of Education

Licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license

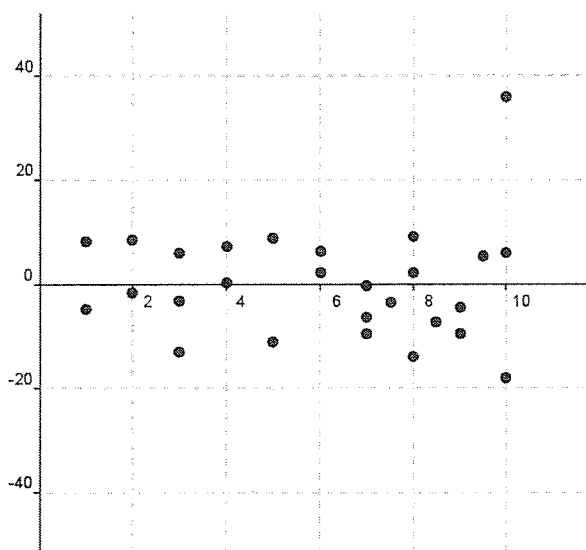
9.



10.



11.



© 2012 Mathematics Vision Project | MVP

In partnership with the Utah State Office of Education

Licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license