Residuals Exploration

Amber, Brandon, and Cody have each created a line of best fit for the data below. Now it's time to decide which is the best one. Their teacher has explained the concept of residuals to them, and they want you to test each of their lines to see who constructed the most accurate line of best fit.

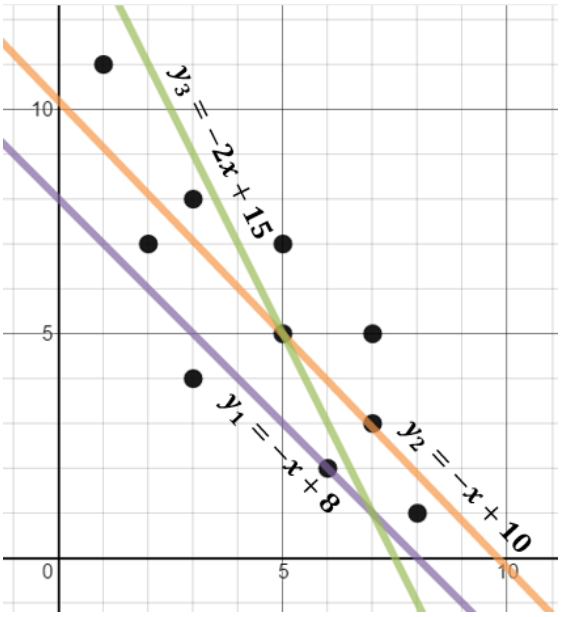
A residual is _____

_____

You can calculate a residual by _____

_____

x	y
1	11
2	7
3	8
3	4
5	7
5	5
6	2
7	3
7	5
8	1



Amber's equation:

$$y_1 = -x + 8$$

Brandon's equation:

$$y_2 = -x + 10$$

Cody's equation:

$$y_3 = -2x + 15$$

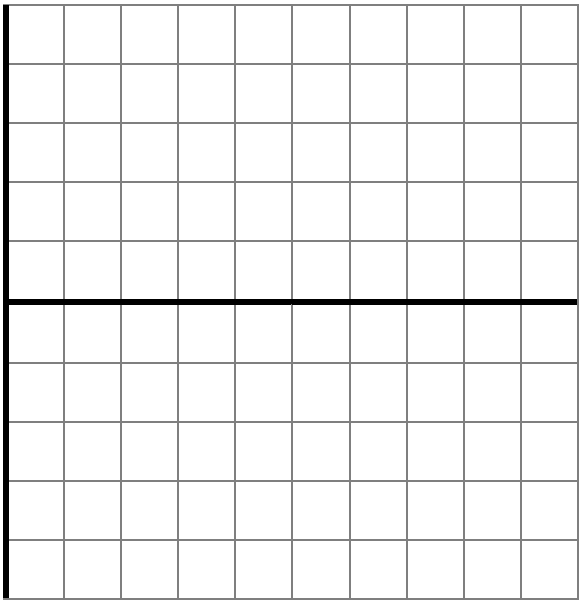
Which line of best fit do you think is the most accurate? Why?

This activity was created by Jessica Thomas (jmthomas891@gmail.com).

First, test Amber's equation by completing the chart below.

x	Observed y	Predicted y_1 $y_1 = -x + 8$	Residual: $y_1 - y$
1	11		
2	7		
3	8		
3	4	$-(3) + 8 = 5$	$5 - 4 = 1$
5	7		
5	5		
6	2		
7	3		
7	5		
8	1		

Then, plot the residual values as points on the graph below. The residual point for the data point (3,4) will be (4,1) because it is the 4th data point and has a residual value of 1.



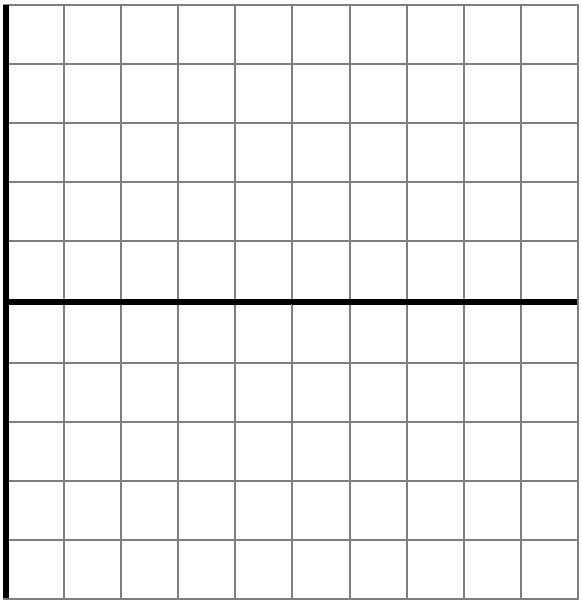
Lastly, calculate the average of the residuals for Amber's line of best fit.

This activity was created by Jessica Thomas (jmthomas891@gmail.com).

Now test Brandon's equation by completing the chart below.

x	Observed y	Predicted y_1 $y_2 = -x + 10$	Residual: $y_2 - y$
1	11		
2	7		
3	8		
3	4	$-(3) + 10 = 7$	$7 - 4 = 3$
5	7		
5	5		
6	2		
7	3		
7	5		
8	1		

Then, plot the residual values as points on the graph below. The residual point for the data point (3,4) will be (4,3) because it is the 4th data point and has a residual value of 3.



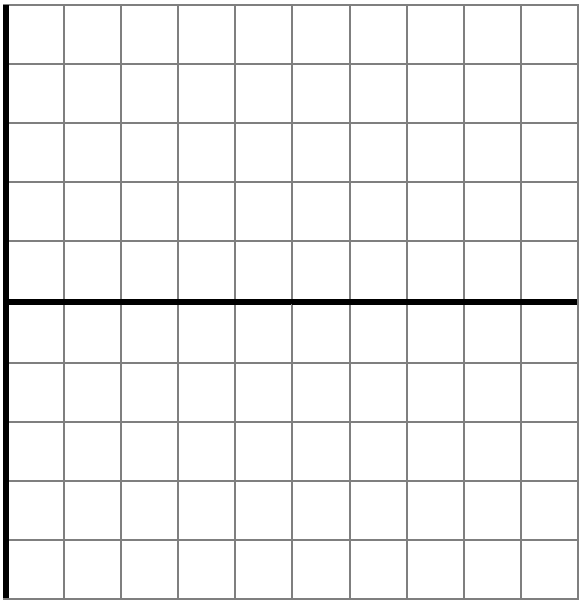
Lastly, calculate the average of the residuals for Brandon's line of best fit.

This activity was created by Jessica Thomas (jmthomas891@gmail.com).

Last, test Cody's equation by completing the chart below.

x	Observed y	Predicted y_1 $y_3 = -2x + 15$	Residual: $y_3 - y$
1	11		
2	7		
3	8		
3	4	$-2(3) + 15 = 9$	$9 - 4 = 5$
5	7		
5	5		
6	2		
7	3		
7	5		
8	1		

Then, plot the residual values as points on the graph below. The residual point for the data point (3,4) will be (4,5) because it is the 4th data point with a residual value of 5.



Lastly, calculate the average of the residuals for Cody's line of best fit.

This activity was created by Jessica Thomas (jmthomas891@gmail.com).

Analysis

Investigate all three residuals plots and residuals averages side-by-side.

It is fairly clear that Amber's line of best fit is a little too low for the data set.

1. What does this "too low" line do to her residual plot compared to the residual plots for the other lines of best fit?

2. What does this "too low" line do to the average of her residuals compared to the other averages?

It's hard to tell if Brandon or Cody's lines of best fit are the most accurate for the data set. However, you know that the line should come as close to possible to all data points.

3. For the line to come as close as possible to all data points, what should the average of the residuals be closest to?

4. Based on your answer for question #3, who has the most accurate line of best fit?

5. Brandon's line of best fit produces a scatterplot with what type of correlation?

6. Cody's line of best fit produces a scatterplot with what type of correlation?

7. What type of correlation should a residual plot have if the line of best fit is as accurate as possible?

This activity was created by Jessica Thomas (jmthomas891@gmail.com).