Name: _____ Group Members: _____

Exploration 8-2a: Sums of Squares of Residuals

Date: _____

Objective: Find the sum of the squares of the residuals for a function found by linear regression.

Suppose that these data have been measured for the related variables x and y.

x	y
3	41
5	37
7	29
9	28
11	23
13	26
15	19
17	9
19	8
21	4

1. Show by linear regression that the best-fitting linear function is $\hat{y} = -2x + 46.4$. How do you interpret the fact that the correlation coefficient is negative?

2. Plot the given data on a scatter plot. On the same screen, plot the regression line. How well does the linear function fit the data?

3. On your grapher, make three lists, one showing $\hat{y}$ for each point, one showing the residual $y - \hat{y}$, and a third showing the squares of the residuals, $(y - \hat{y})^2$. Copy the results into the table.

4. Calculate SS_{res} from the squares of the residuals. See if you can discover a time-efficient way to find the sum using built-in features on your grapher.

5. This table has the same data as before. Find SS_{res} again using $y_2 = -2x + 46$. How do you interpret the fact that the answer is *greater* than SS_{res} using the regression equation $\hat{y} = -2x + 46.4$?

x	y
3	41
5	37
7	29
9	28
11	23
13	26
15	19
17	9
19	8
21	4

6. Calculate $\overline{x}$ and $\overline{y}$, the averages of x and y. Show algebraically that the point $(\overline{x}, \overline{y})$ is on the regression line $\hat{y} = -2x + 46.4$.

7. The line $y_3 = -2.1x + 47.6$ also contains the "average-average" point $(\overline{x}, \overline{y})$, but it has a slope of -2.1 instead of -2. Plot the line on the same screen as in Problem 2. Can you tell from the graphs which line fits the data better?

8. Find SS_{res} using the equation in Problem 7. Based on your answer, how can you tell that this line does *not* fit the data as well as does the regression line?

9. What did you learn as a result of doing this Exploration that you did not know before?