

## 9.3: The Line of Best Fit

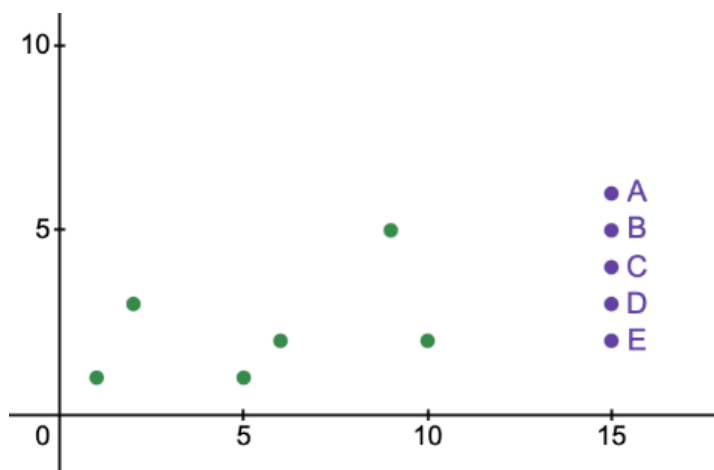
In a previous section, we saw that an  $r$  value close to 1 may indicate a strong linear relationship between an explanatory variable and a response variable. In such situations, we may choose to use an equation of a line to summarize the relationship. We can then use a linear model to make predictions about the values of the response variable.

1. Imagine a line that fits the data best. Which of the purple points, A, B, C, D, or E, does the line intersect? Sketch the line below.



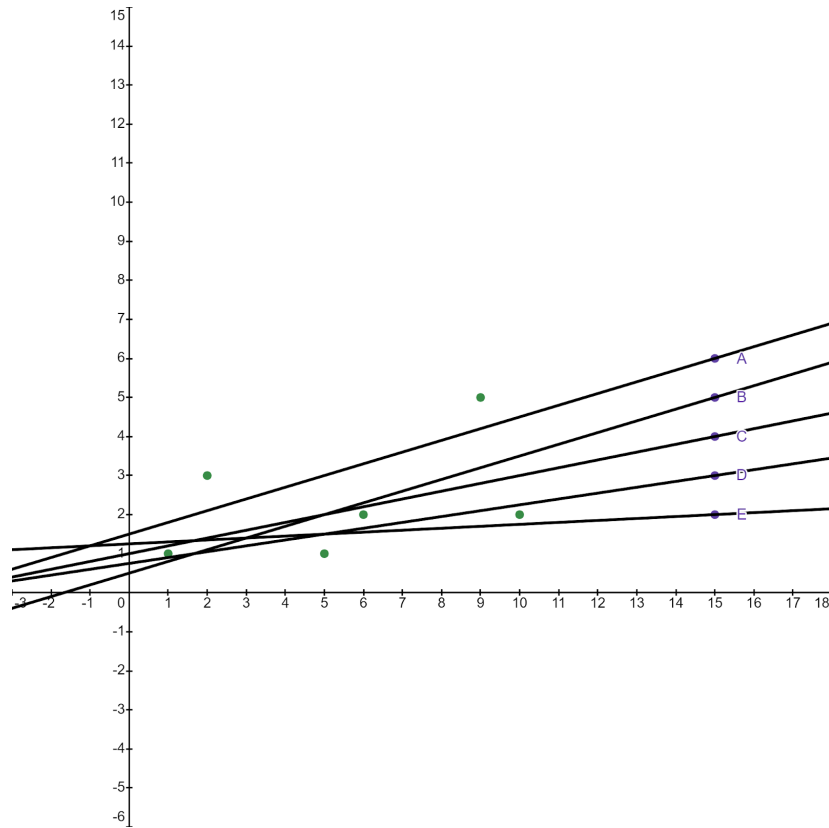
Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

2. Imagine the line in your head continuing out to  $x=100$ . Estimate the  $y$ -coordinate of the point on the line where  $x$  is 100. Explain how you made your guess.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

3. Select the line below that corresponds to the choice you made in question 1. Use it's equation to predict the y-coordinate of the point on the line where x is 100.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

Line A.  $y = 0.3x + 1.5$

Line B.  $y = 0.3x + 0.5$

Line C.  $y = 0.2x + 1$

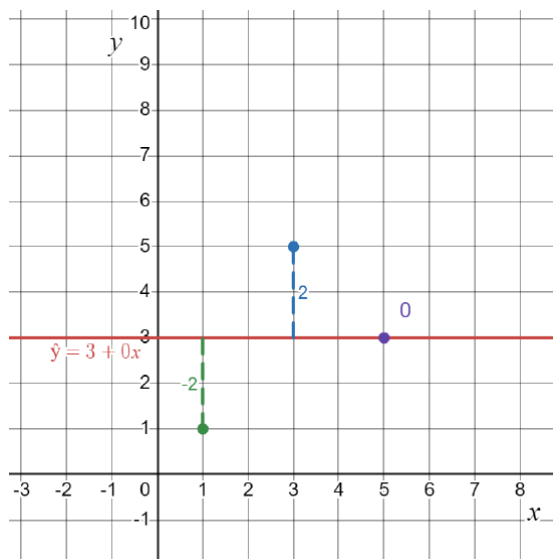
Line D.  $y = 0.15x + 0.75$

Line E.  $y = 0.05x + 1.25$

4. Why do you think this line fits the data best?

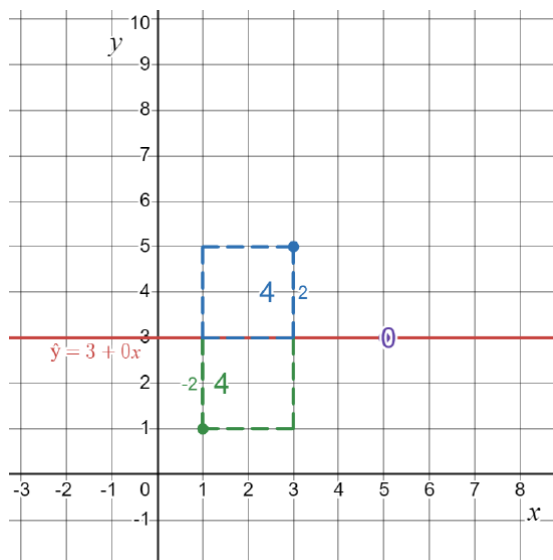
## LSR Line

5. The **least squares regression line (LSR line)** is a linear model that puts roughly half of your scatterplot data above the line and roughly half of your scatterplot data below the line. This line is also called the **line of best fit** as it is the line from which the points in the scatterplot deviate the least from. What do the values -2, 2, and 0 on the graph below represent?



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

6. What do the values 4, 4, and 0 on the graph represent?



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

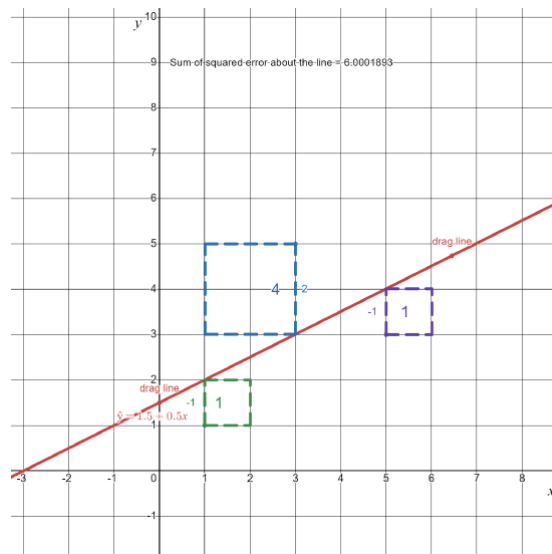
7. Now drag the line by dragging the two red dots at [this desmos graph](#). You can use the QR code below to access the desmos graph.



a. What is the smallest you can make the sum of the squared error about the line?

b. What do you think is true about the line with the smallest sum of squared errors?

For this bivariate set of data, the smallest sum of squared error about the line is 6. The equation of the line of best fit is  $\hat{y} = 0.5x + 1.5$ .



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

## Residuals

The **residual** is the difference (or vertical distance) between the ACTUAL value of a data set ( $y$ ) for an  $x$ -value and the value that your line PREDICTED ( $\hat{y}$ ) for that  $x$ -value.

8. Let's return to the original example. The line of best fit is approximately  $\hat{y} = 0.2x + 1$  (the line that passes through point C). Let's look at how much the data deviates from the linear model. The residual for the point (1,1) has been computed and entered in the table below. Complete the table in the same manner.

x	y	$\hat{y}$	$y - \hat{y}$
1	1	$\hat{y} = 0.2(1) + 1 = 1.2$	$1 - 1.2 = -0.2$
2	3		
5	1		
6	2		
9	5		
10	2		

9. Here's one way to measure how well a line fits a data set:

- Square all the residuals.
- Add up those squares.

The smaller the result, the better the fit. Calculate the square of each residual and enter it in the table. The first squared residual has been computed in the table. Complete the table.

x	Residual: $y - \hat{y}$	Squared Residual
1	-0.2	$(-0.2)^2 = 0.04$
2		
5		
6		
9		
10		

The sum squared residuals is \_\_\_\_\_.

## Calculating and Interpreting Values in the Equation of the LSR Line

10. Use the following steps to calculate the equation of the line of best fit:

- a. Open [this file](#) that contains randomly collected data comparing the number of cricket chirps and the temperature. Copy the data from the file by highlighting the data and clicking copy.

- i. Alternatively, you can access the data set using this QR code:



- b. Open <https://www.desmos.com/calculator> and paste the data into the first line by clicking paste.

- c. In the second line, calculate the linear correlation coefficient,  $r$ , by typing in  $\text{corr}(x_1, y_1)$ .

$r =$  \_\_\_\_\_

- d. Compute the sample means,  $\bar{x} = \text{mean}(x_1)$  and  $\bar{y} = \text{mean}(y_1)$ , and sample standard deviations,  $s_x = \text{stdev}(x_1)$  and  $s_y = \text{stdev}(y_1)$ .

$\bar{x} =$  \_\_\_\_\_

$\bar{y} =$  \_\_\_\_\_

$s_x =$  \_\_\_\_\_

$s_y =$  \_\_\_\_\_

- e. Calculate the slope,  $m = \frac{r \cdot s_y}{s_x}$ , of the line of best fit.

- f. The slope of the line is the predicted change in  $y$  for every unit change in  $x$ . Interpret the slope of the equation of the line of best fit in context (the units of  $x$  are number of chirps, and the units of  $y$  are degrees Fahrenheit).

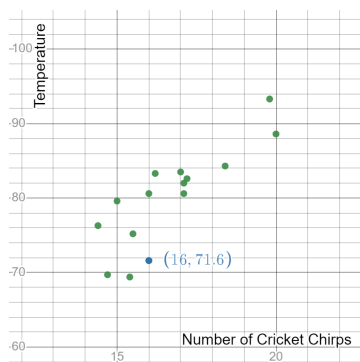
g. Calculate the y-intercept,  $b = \bar{y} - m \cdot \bar{x}$ , for the line of best fit.

h. The y-intercept of the line of best fit is the predicted y-value when the x-value is 0. Interpret the y-intercept in context.

i. Write the equation you found. Use the line of best fit to predict the temperature near a cricket that chirps 24 times.

$$\hat{y} = \underline{\hspace{2cm}}.$$

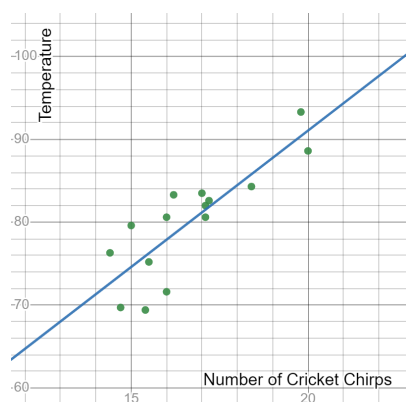
11. One of the crickets chirped 16 times and the temperature near it was 71.6 degrees Fahrenheit. Calculate the residual for this cricket.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

12. In desmos, type in  $y \sim mx + b$  to check your work.

$$\hat{y} = \underline{\hspace{2cm}}$$



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

This page titled [9.3: The Line of Best Fit](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Hannah Seidler-Wright](#).

- [Current page](#) by Hannah Seidler-Wright is licensed [CC BY-NC-SA 4.0](#).
- [1.2: The Statistical Analysis Process](#) by Hannah Seidler-Wright is licensed [CC BY-NC-SA 4.0](#).