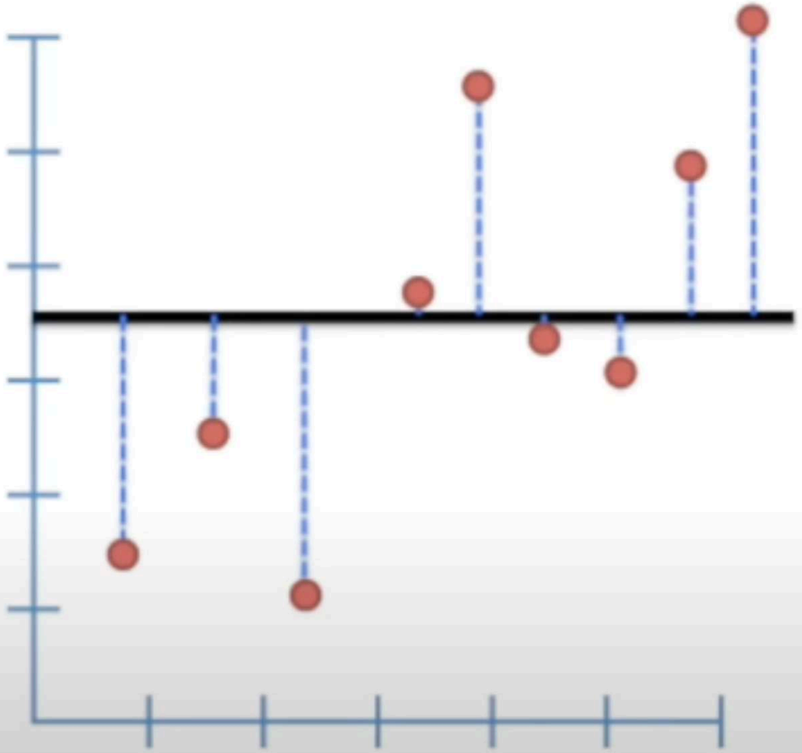
DATA ANALYTICS- UNIT 2

By

DEEPIKA KAMBOJ

Linear Regression

$$(b - y_1)^2 + (b - y_2)^2 + (b - y_3)^2 + (b - y_4)^2 + (b - y_5)^2 + (b - y_6)^2 + (b - y_7)^2 + (b - y_8)^2 + (b - y_9)^2$$



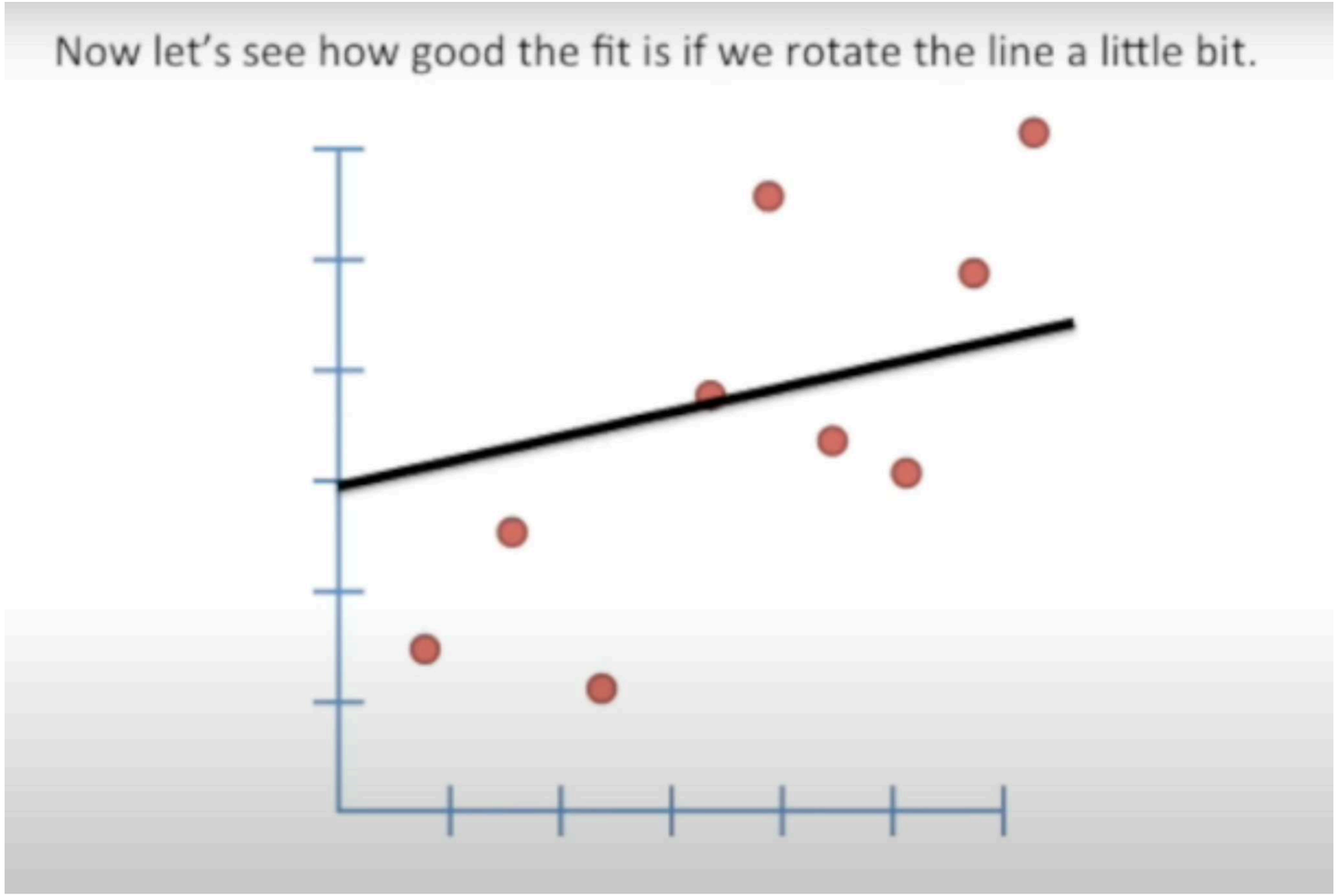
= 24.62

This is our measure of how well this line fits the data.

It's called the "sum of squared residuals, because the residuals are the differences between the real data and the line, and we are summing the square of these values.

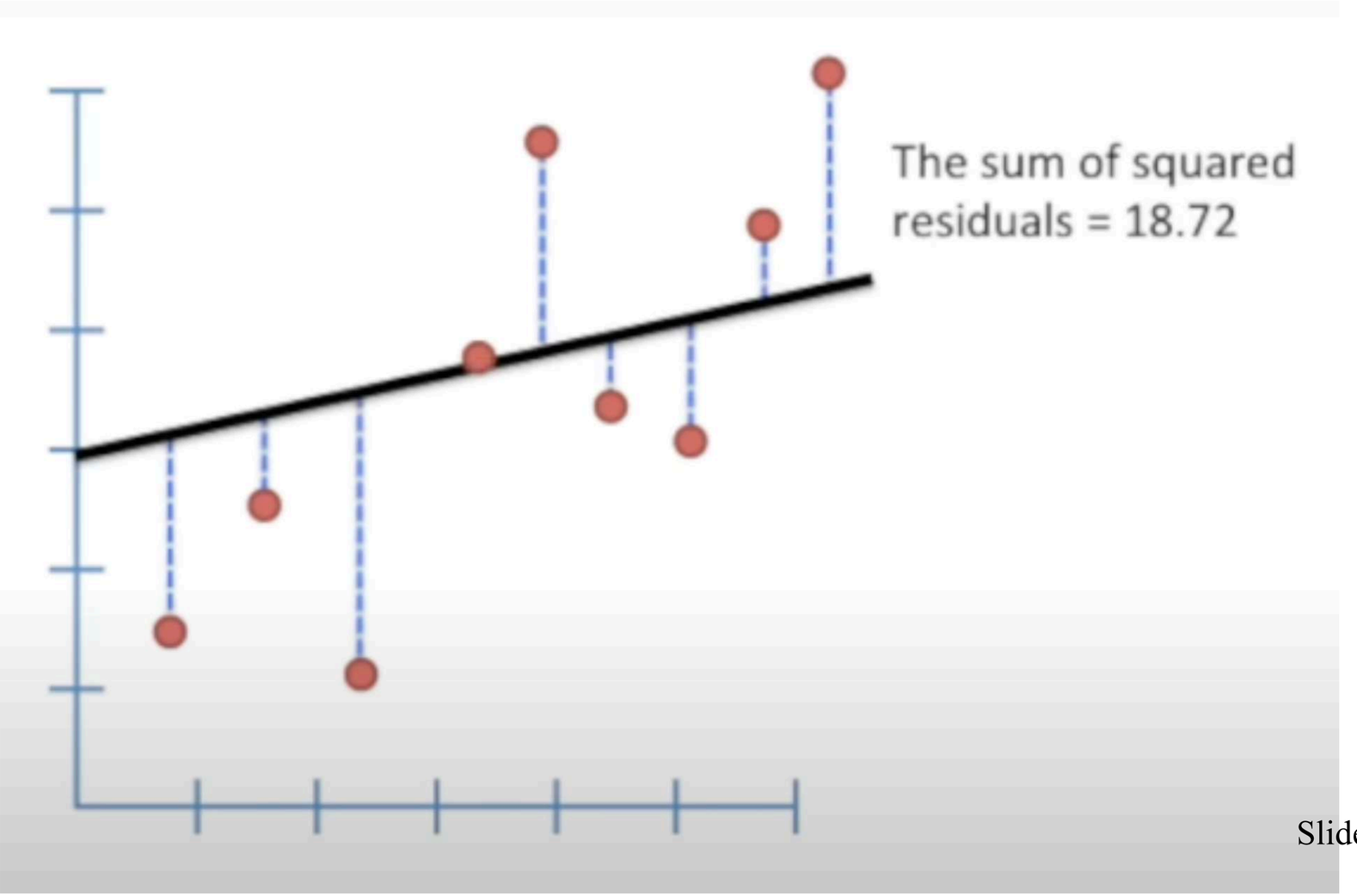
Slide Courtesy: StatQuest

Linear Regression



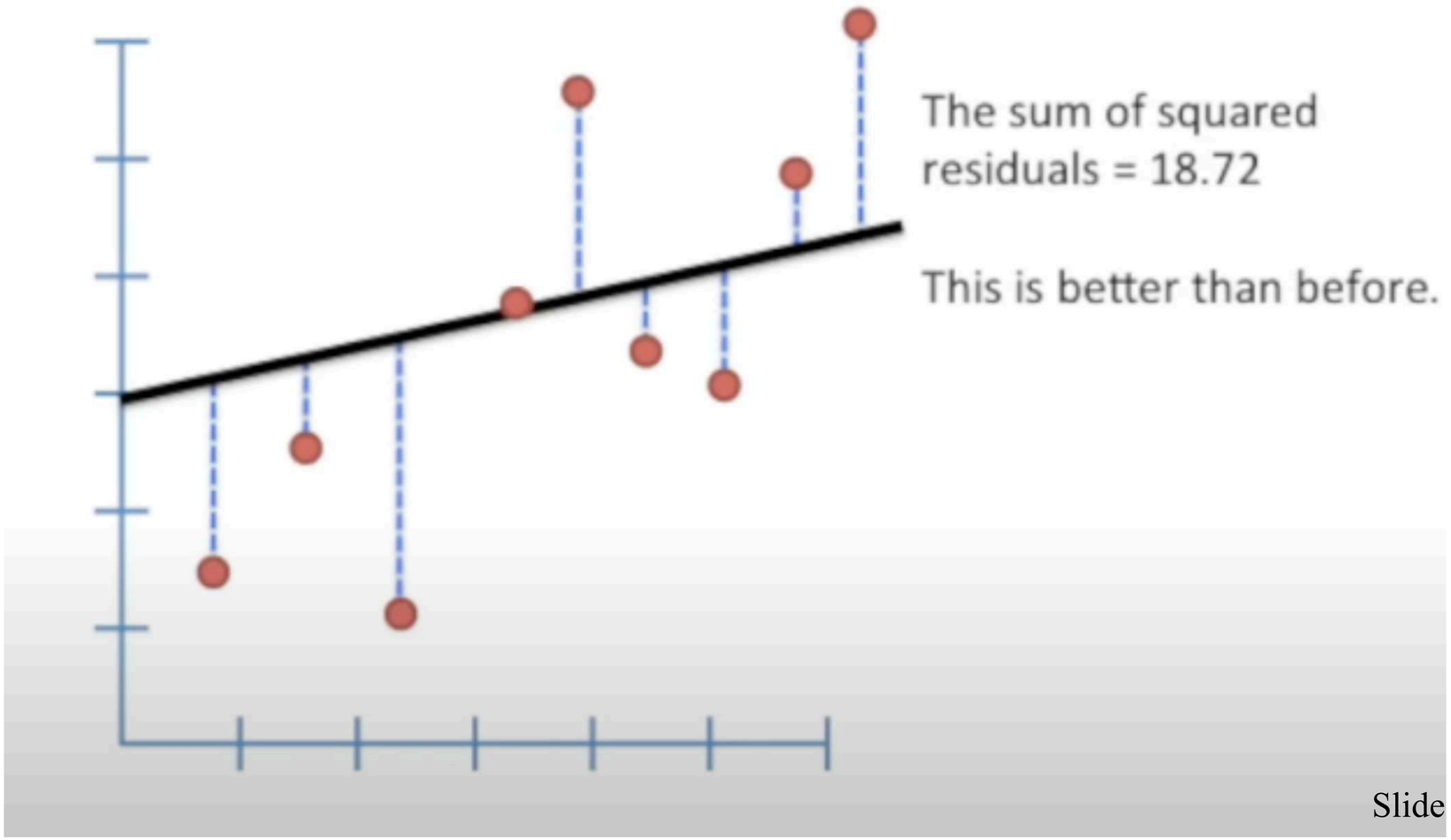
Slide Courtesy: StatQuest

Linear Regression



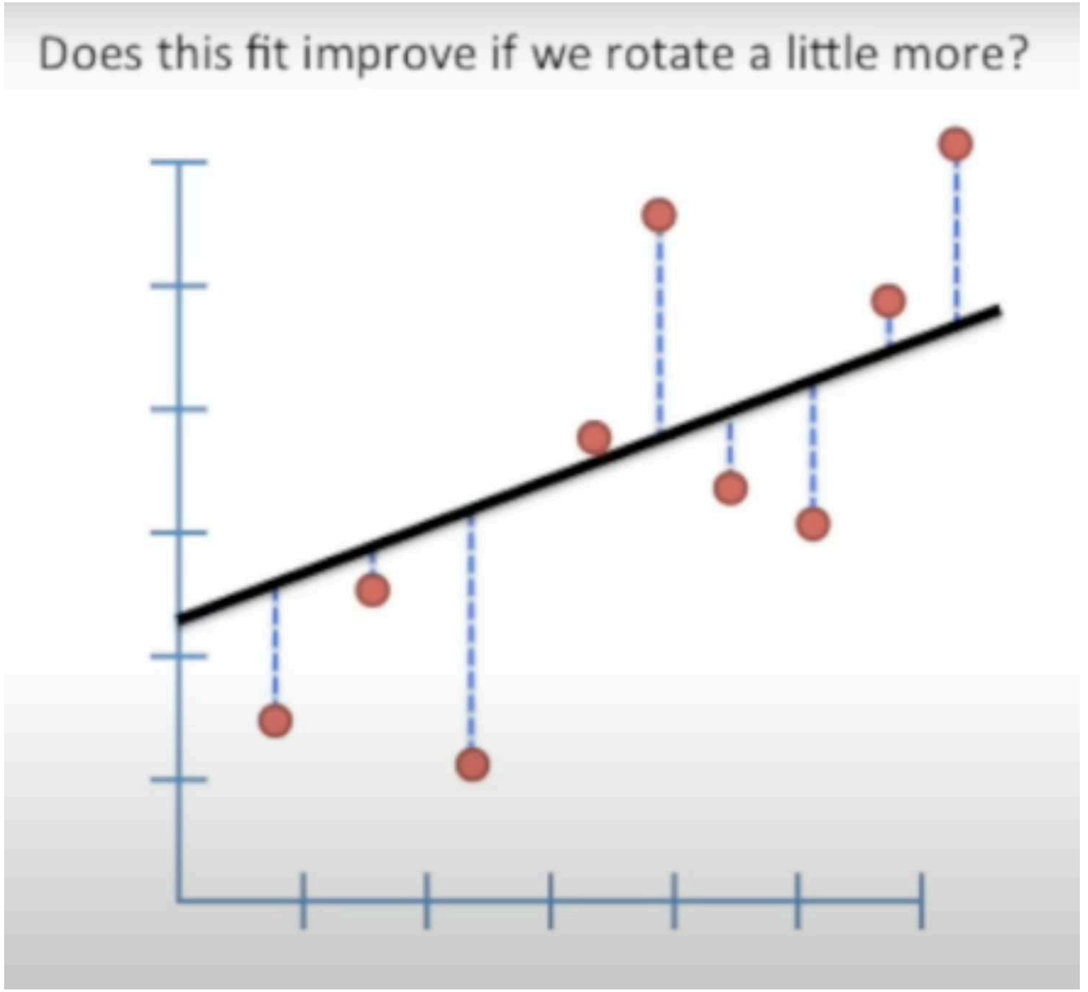
Slide Courtesy: StatQuest

Linear Regression



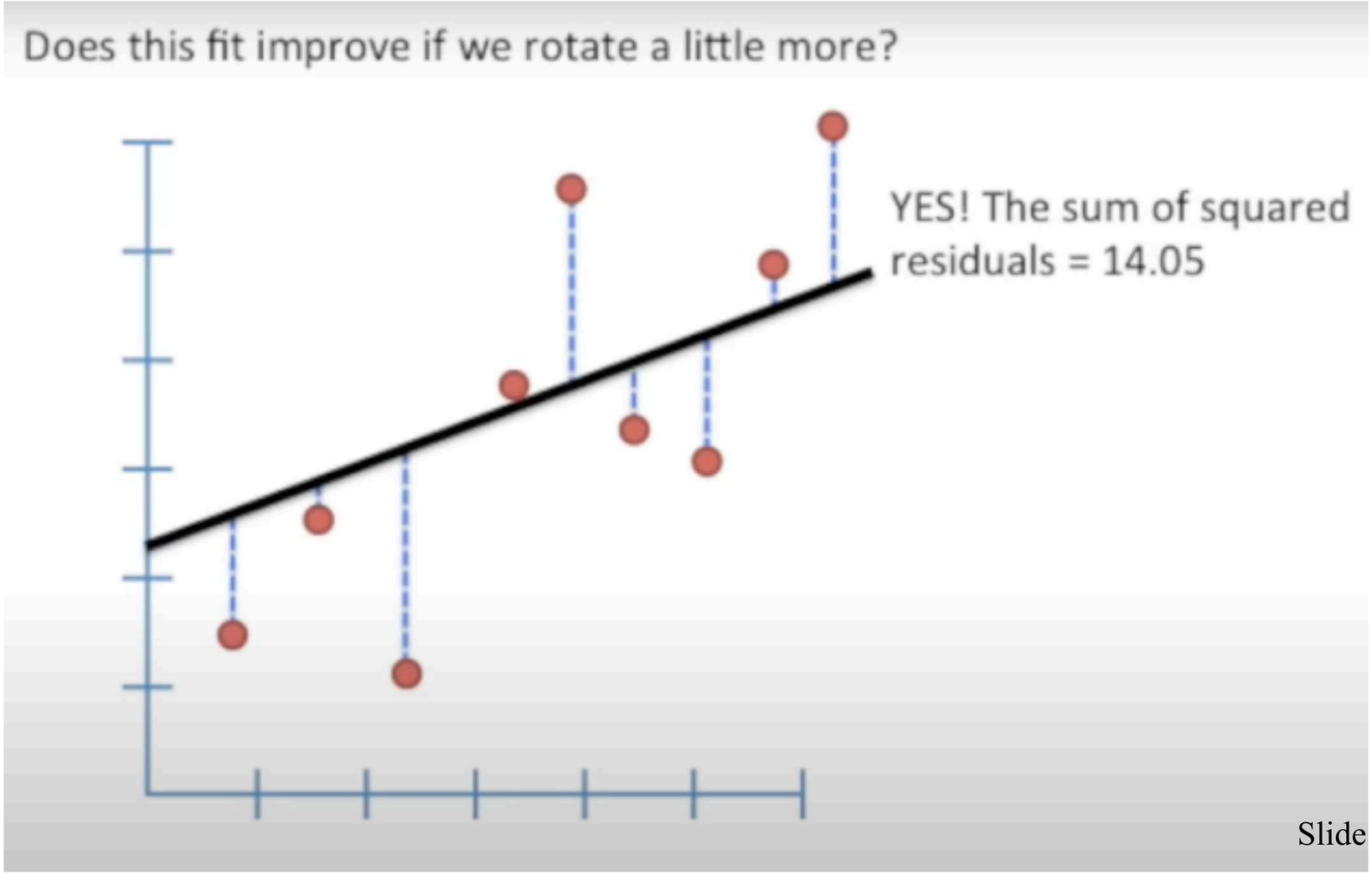
Slide Courtesy: StatQuest

Linear Regression



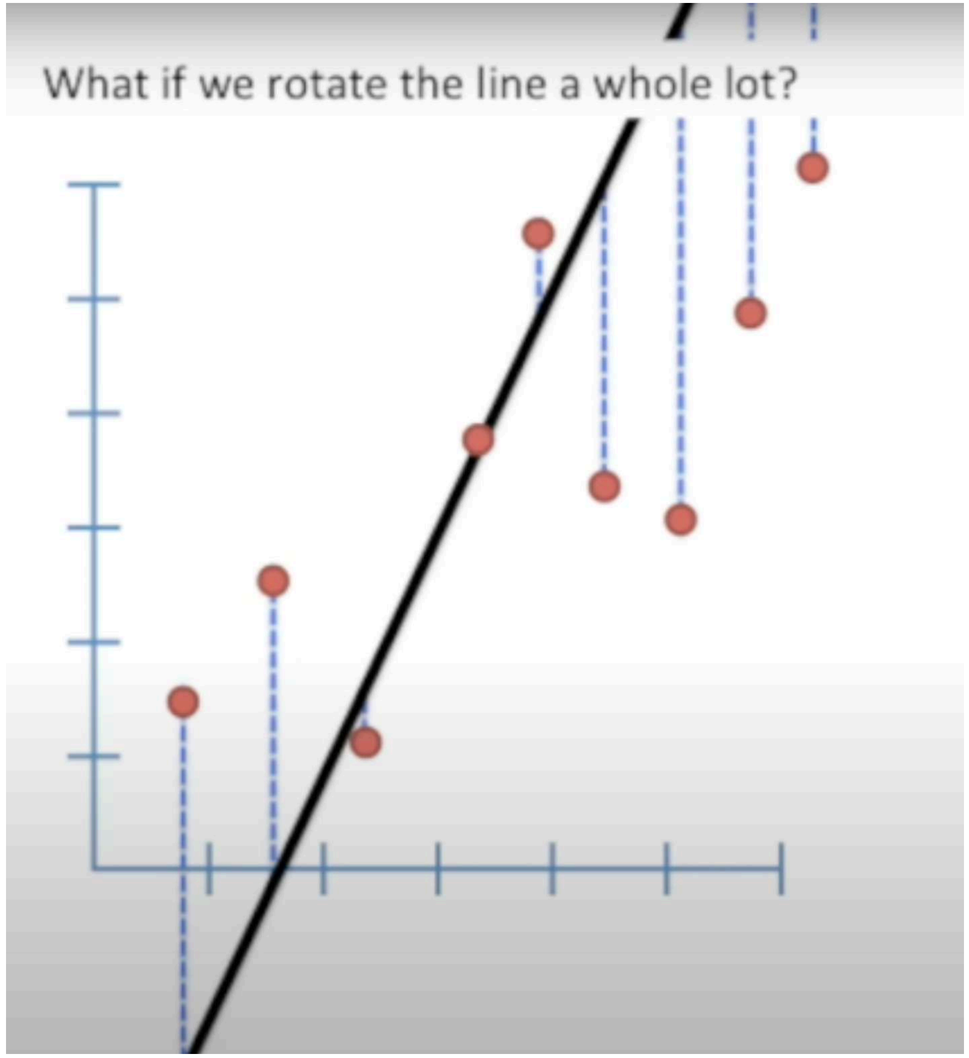
Slide Courtesy: StatQuest

Linear Regression



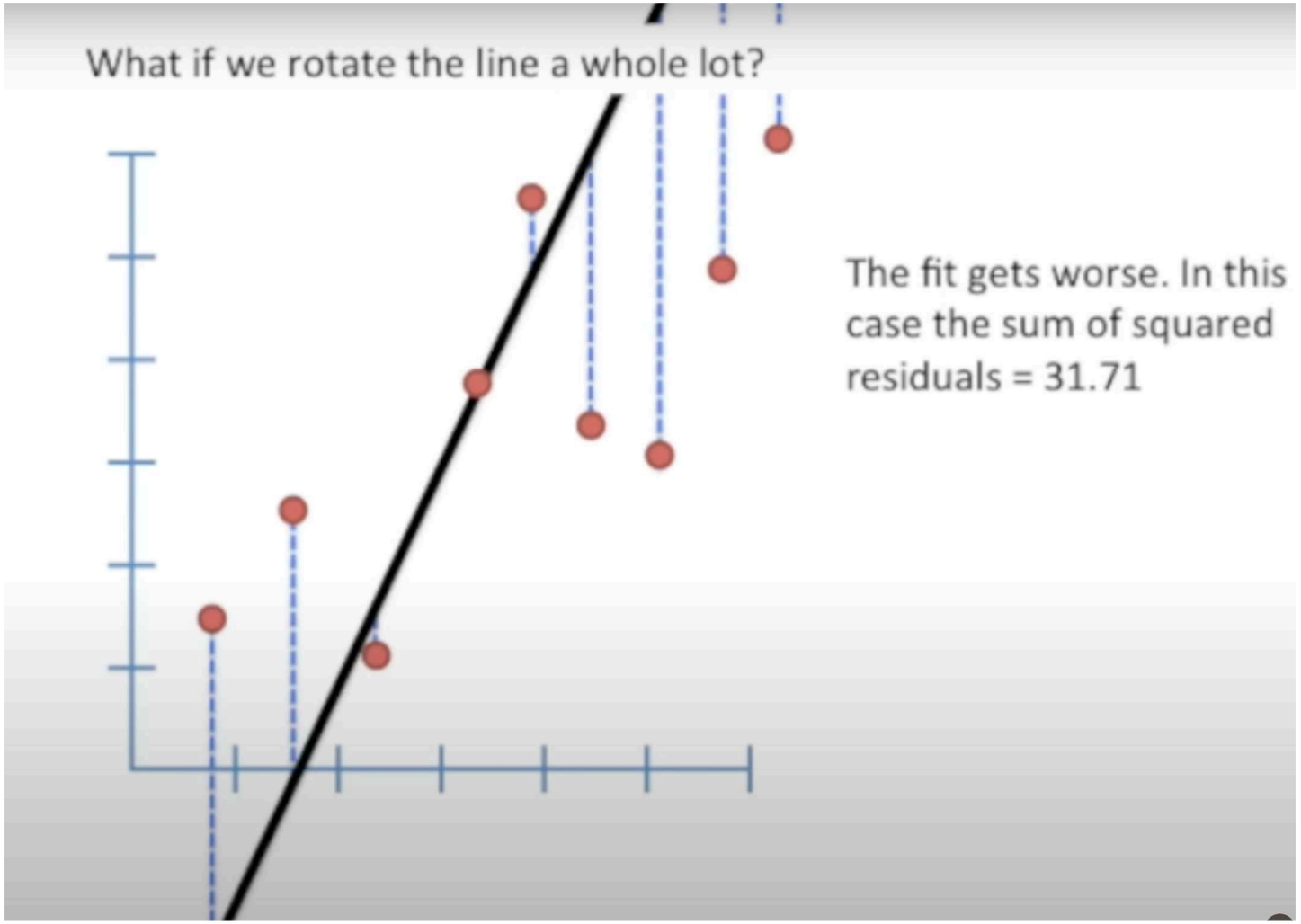
Slide Courtesy: StatQuest

Linear Regression



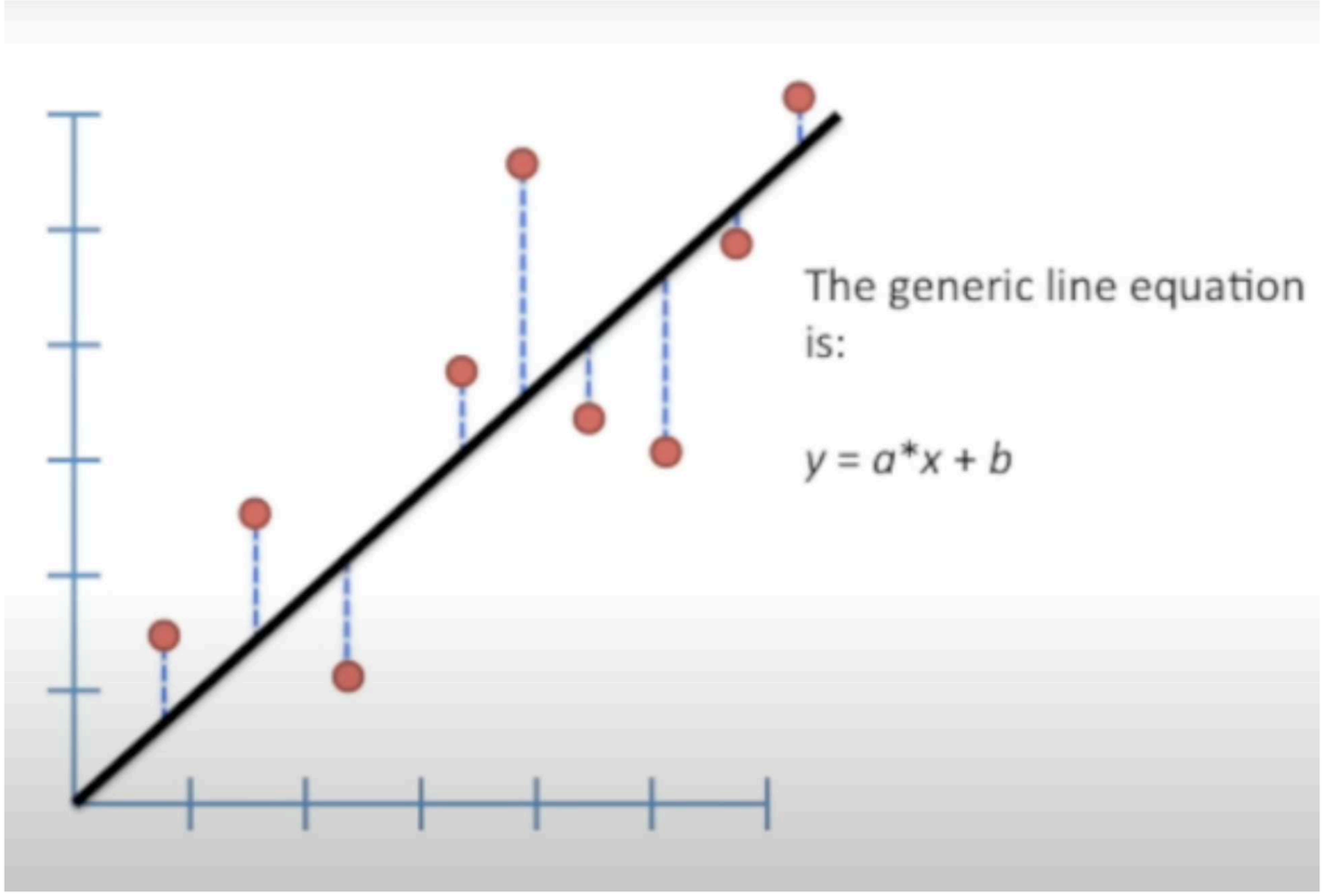
Slide Courtesy: StatQuest

Linear Regression



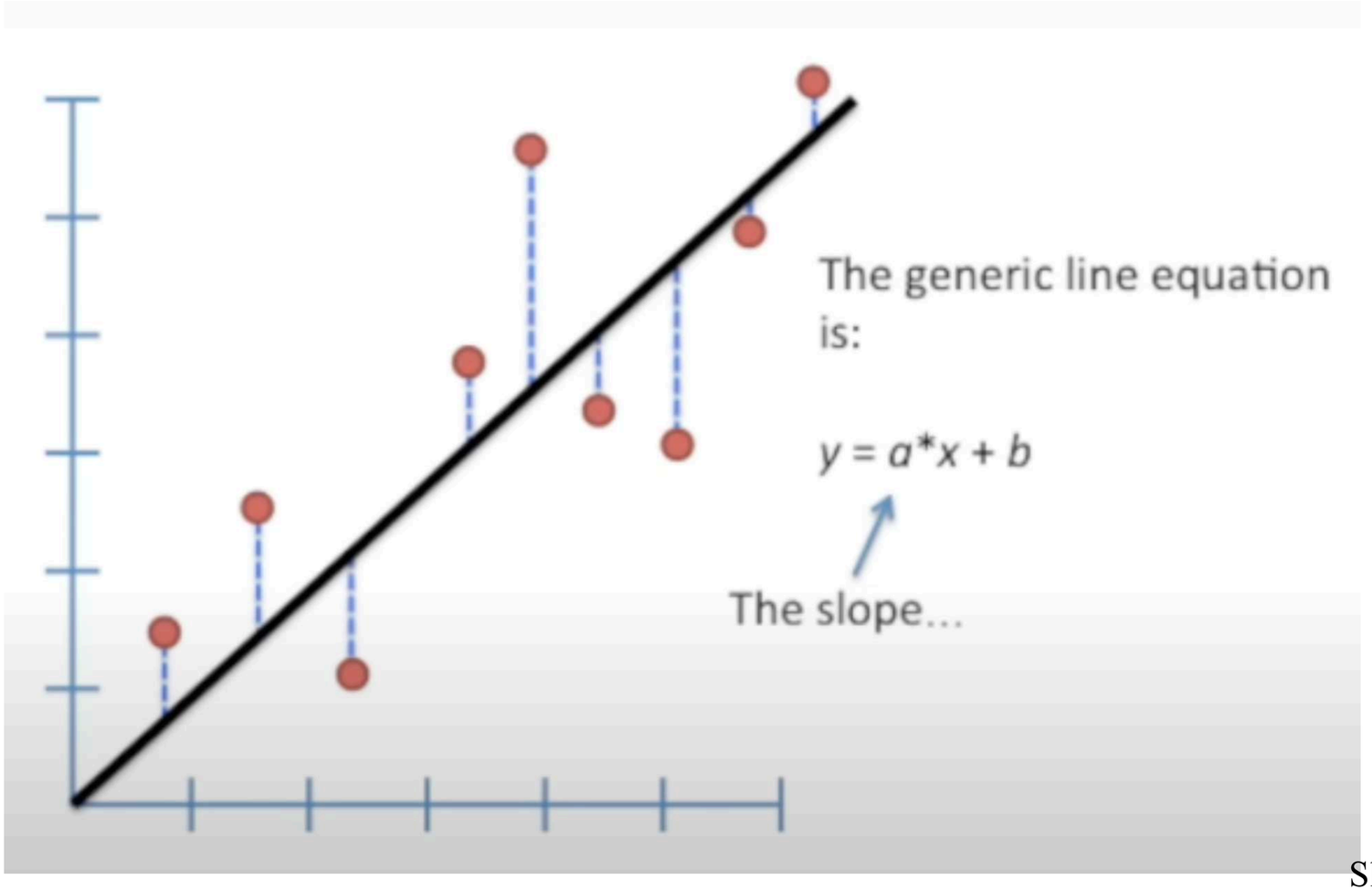
Slide Courtesy: StatQuest

Linear Regression



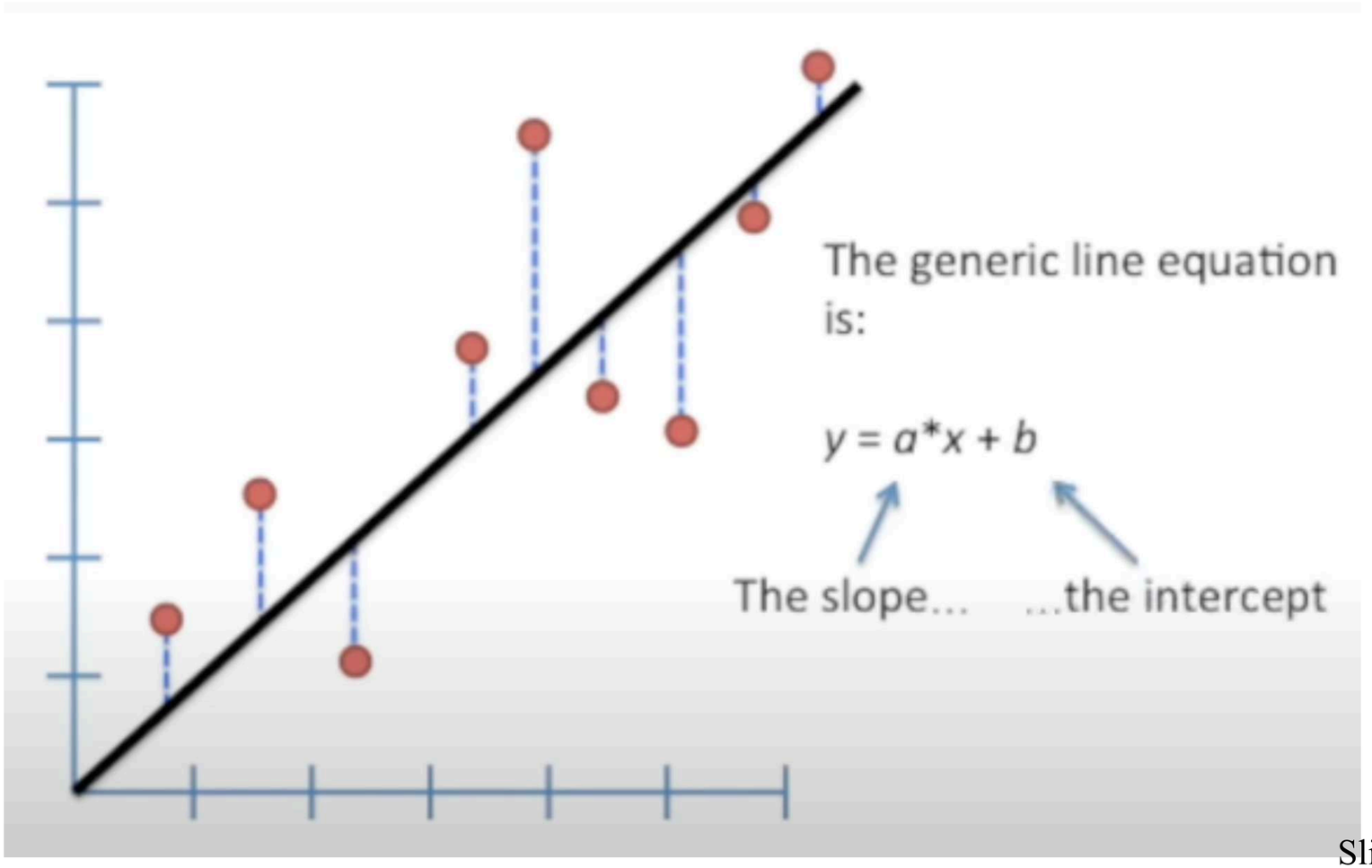
Slide Courtesy: StatQuest

Linear Regression



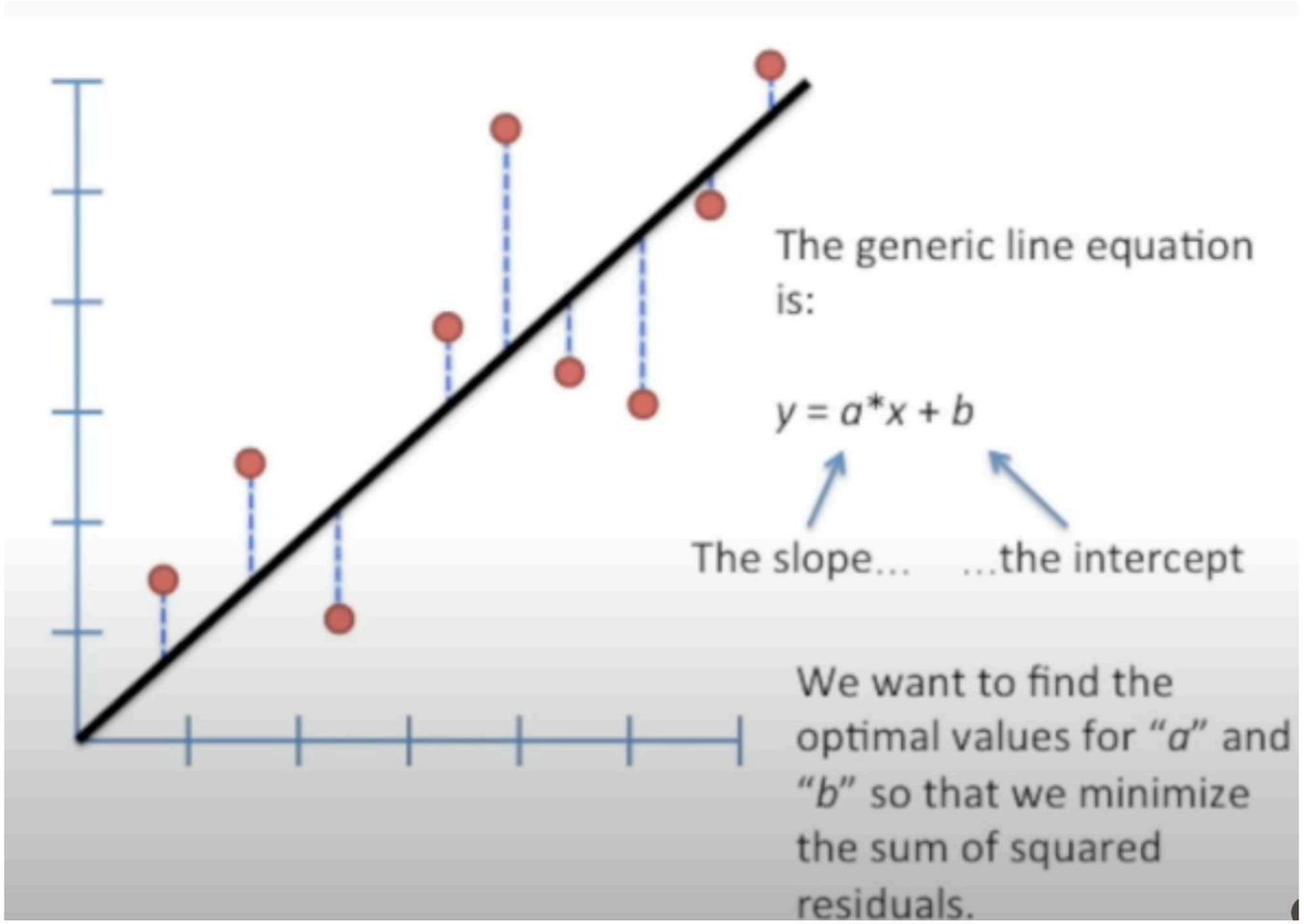
Slide Courtesy: StatQuest

Linear Regression



Slide Courtesy: StatQuest

Linear Regression

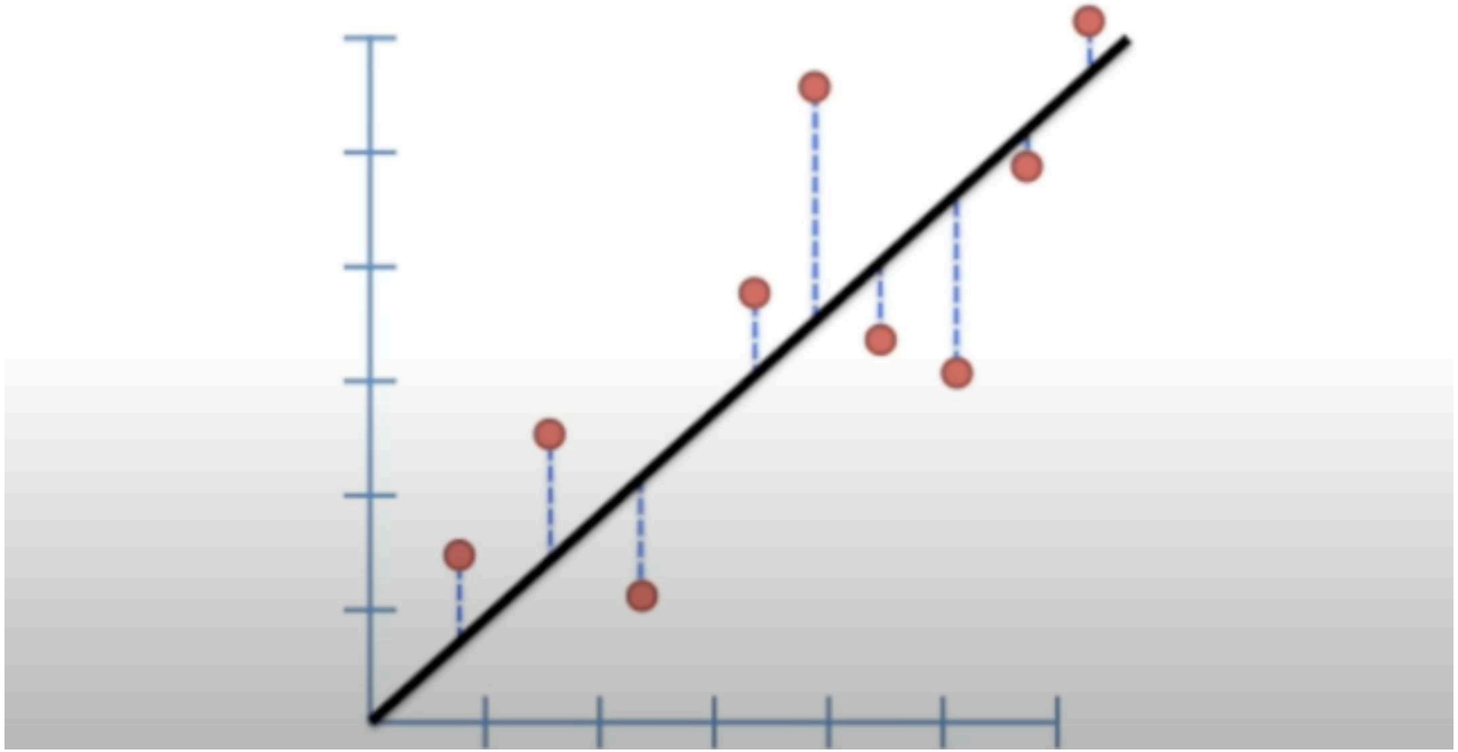


Slide Courtesy: StatQuest

Linear Regression

In more general math terms…

Sum of squared residuals = $((a*x_1 + b) - y_1)^2 + ((a*x_2 + b) - y_2)^2 + \ldots$

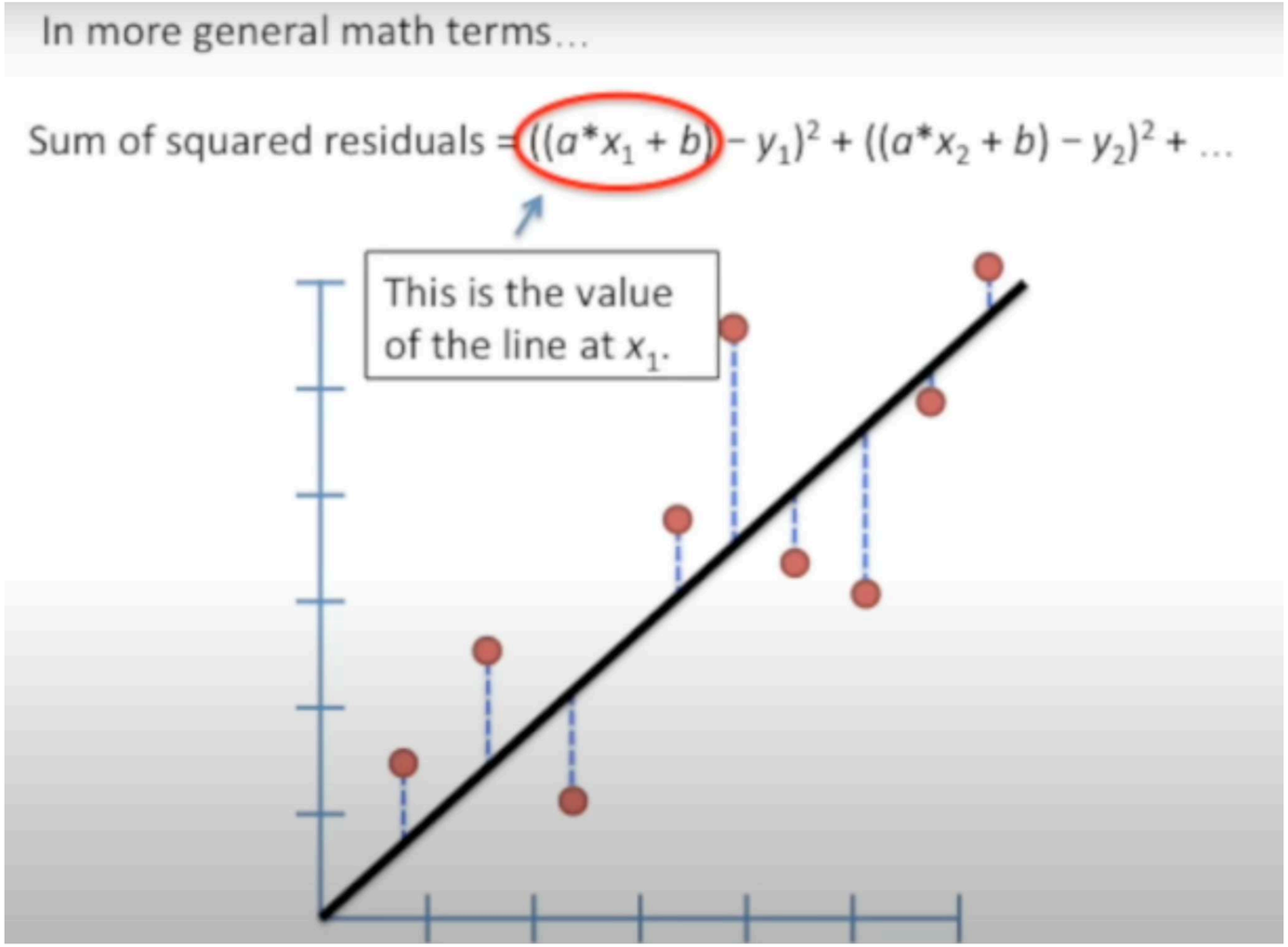


Slide Courtesy: StatQuest

Linear Regression

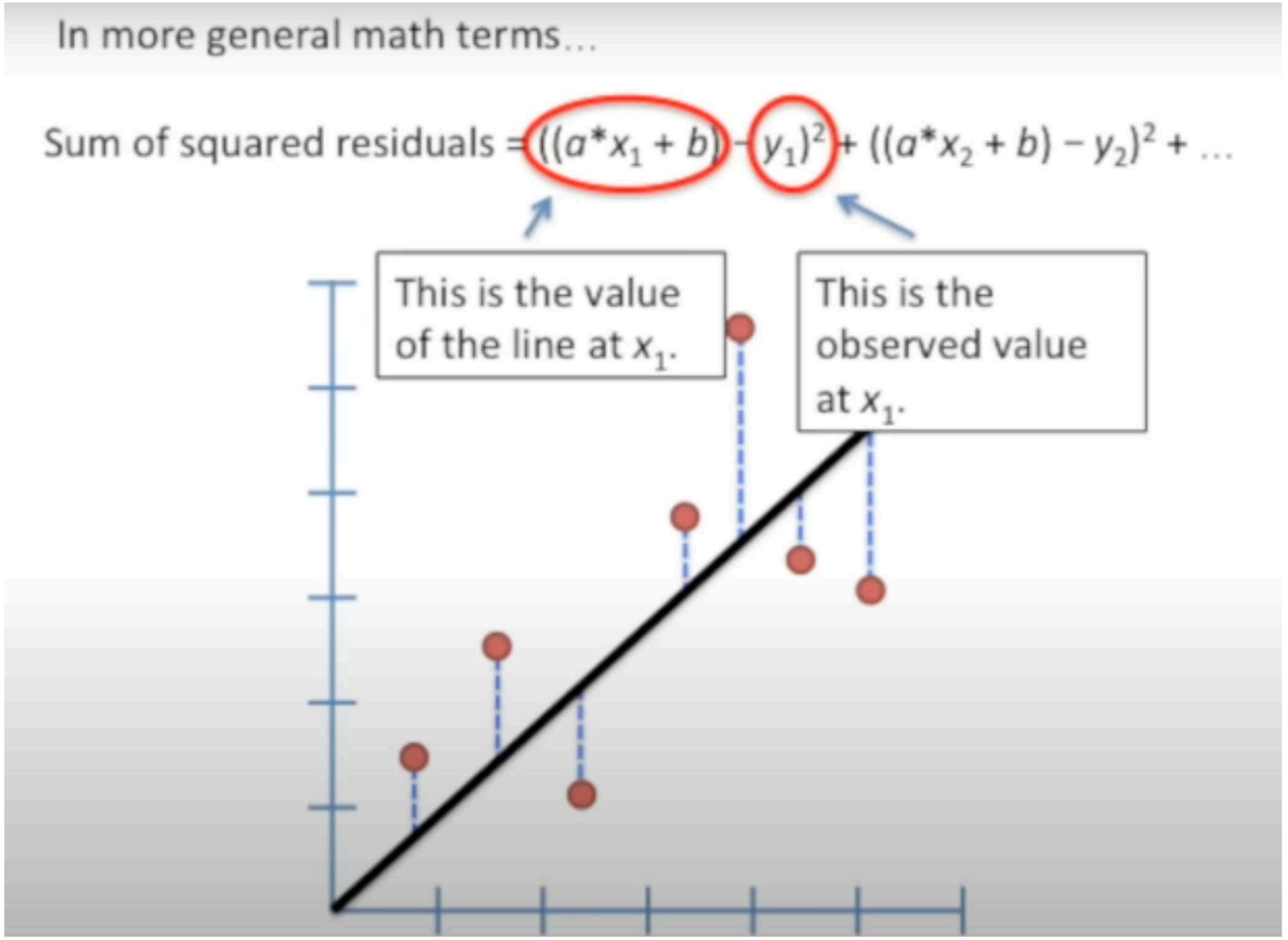
In more general math terms…

Sum of squared residuals = $((a*x_1 + b) - y_1)^2 + ((a*x_2 + b) - y_2)^2 + \ldots$



Slide Courtesy: StatQuest

Linear Regression



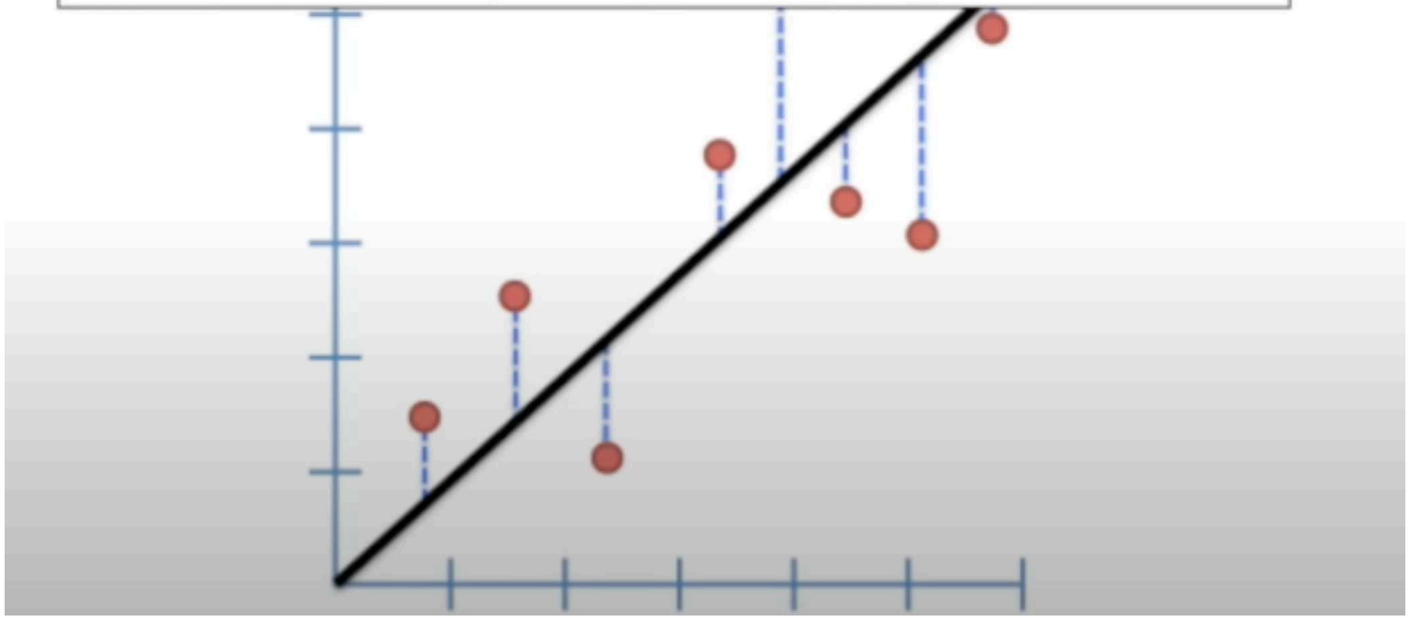
Slide Courtesy: StatQuest

Linear Regression

In more general math terms…

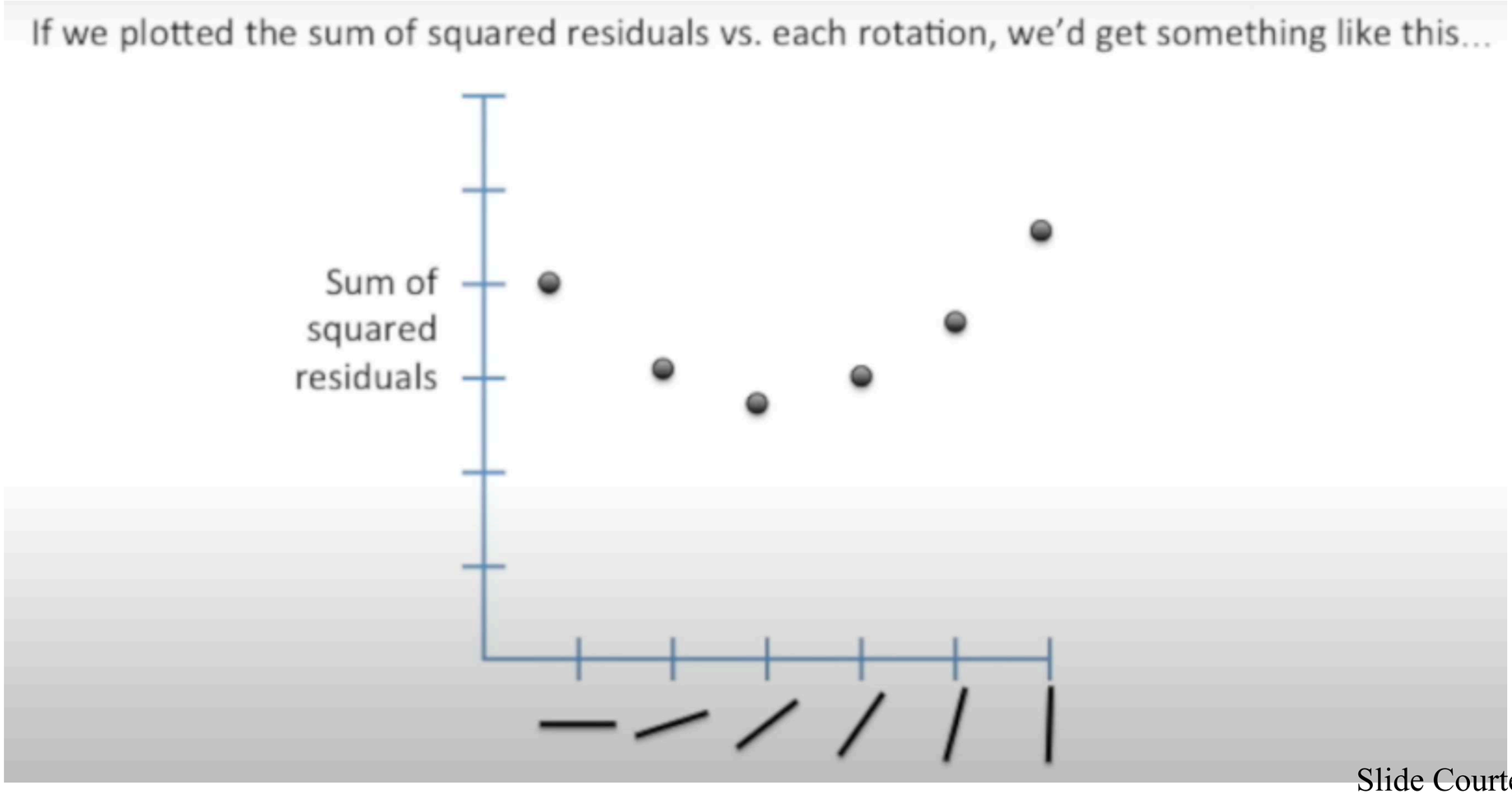
Sum of squared residuals = $((a*x_1 + b) - y_1)^2 + ((a*x_2 + b) - y_2)^2 + \ldots$

Since we want the line that will give us the smallest sum of squares, this method for finding the best values for "*a*" and "*b*" is called "Least Squares".



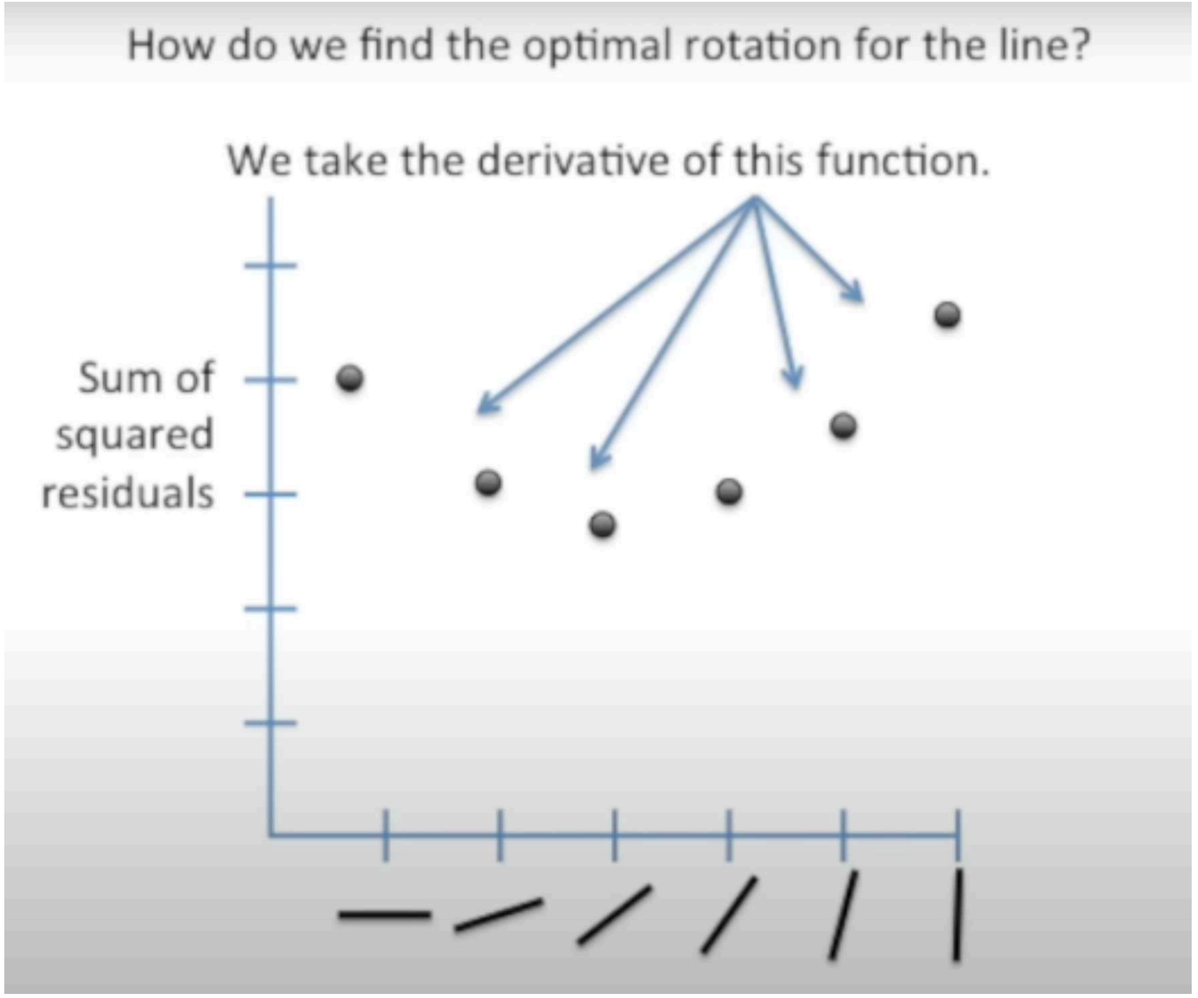
Slide Courtesy: StatQuest

Linear Regression



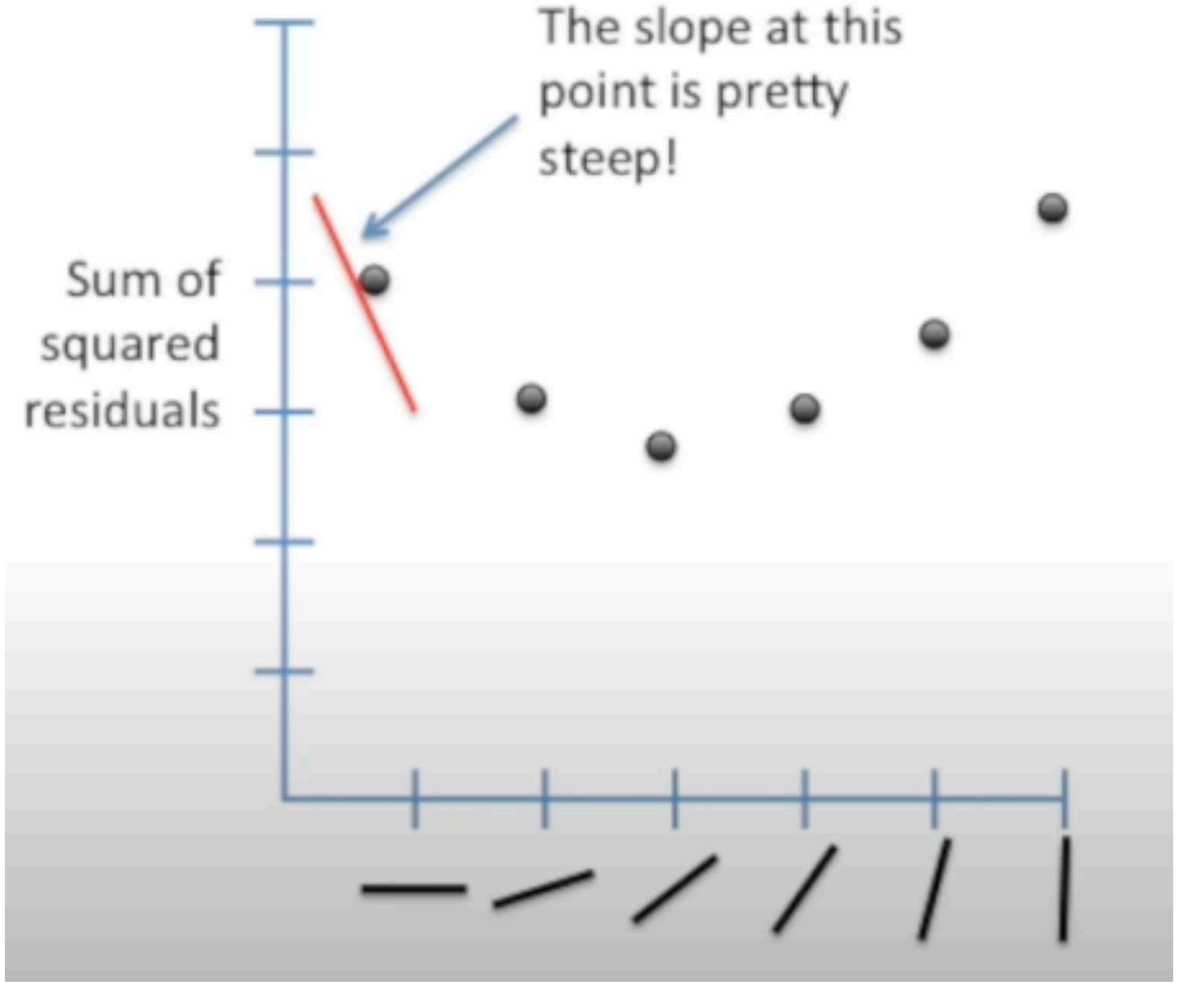
Slide Courtesy: StatQuest

Linear Regression



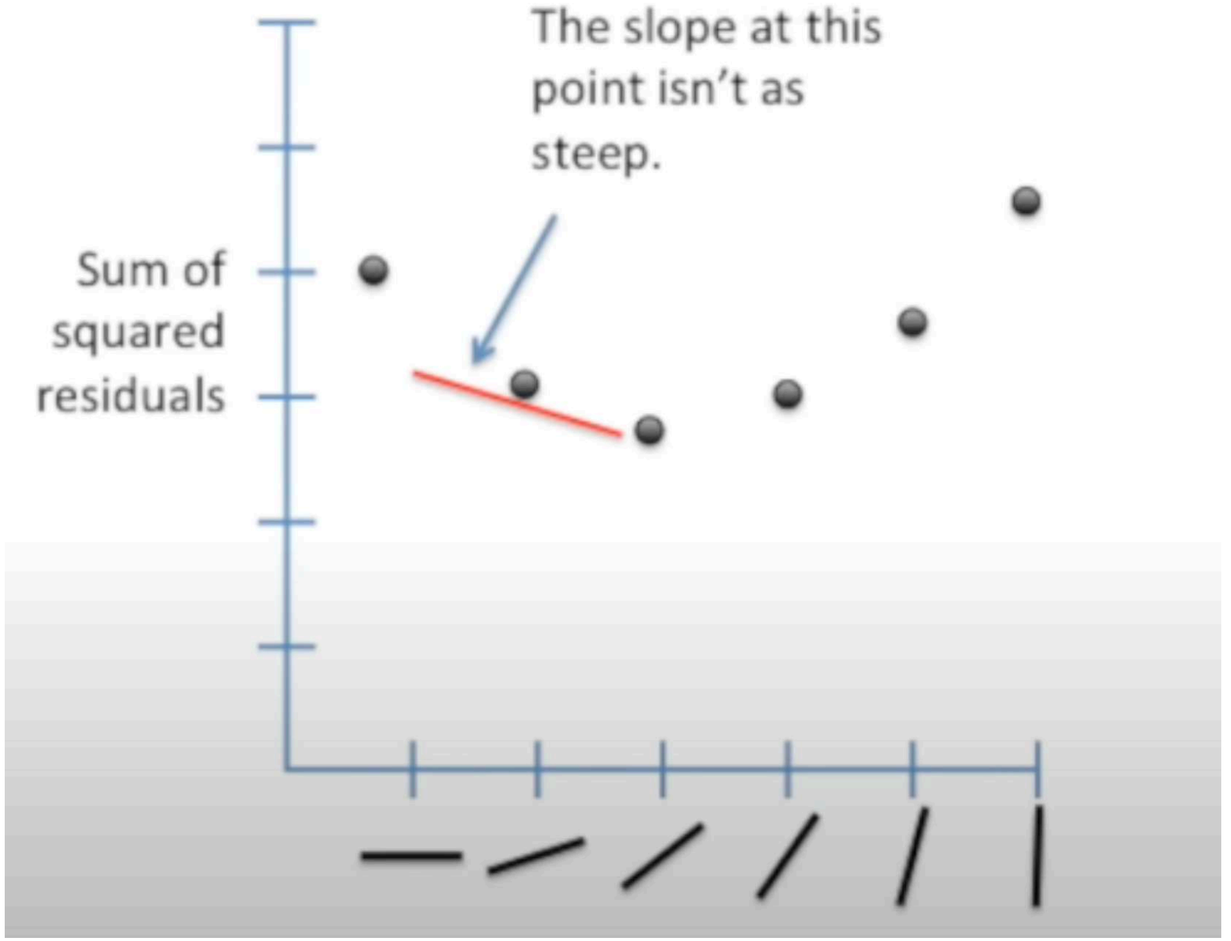
Slide Courtesy: StatQuest

Linear Regression



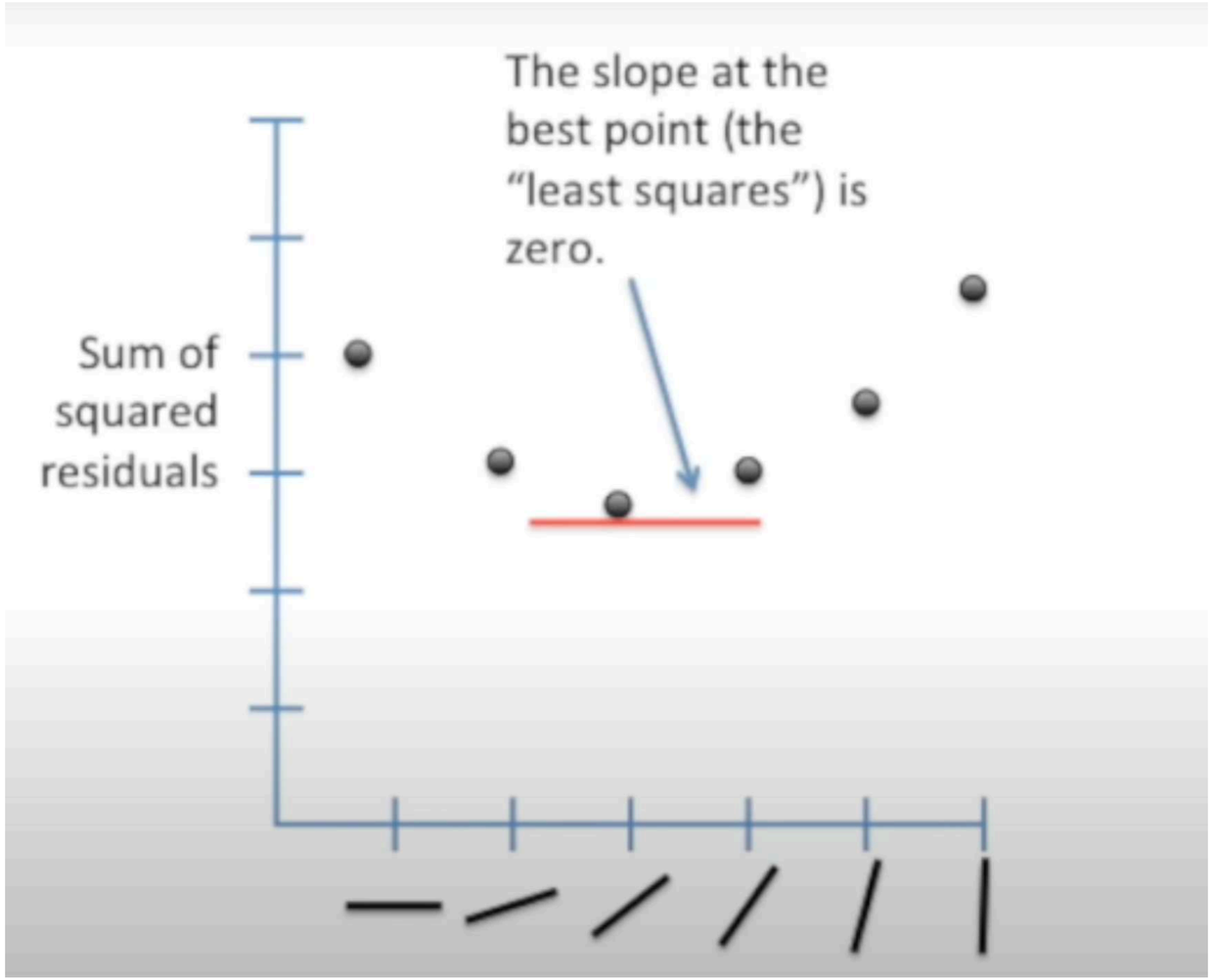
Slide Courtesy: StatQuest

Linear Regression



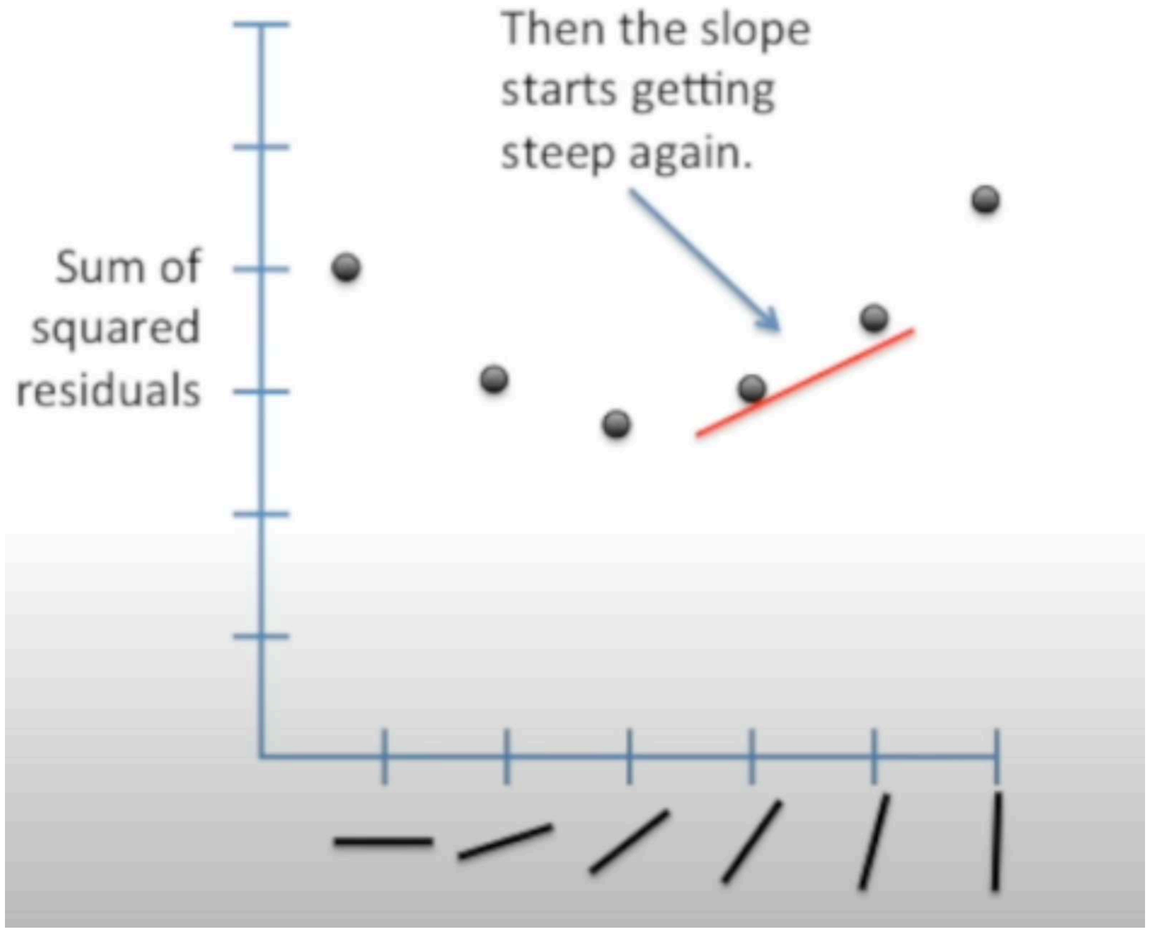
Slide Courtesy: StatQuest

Linear Regression



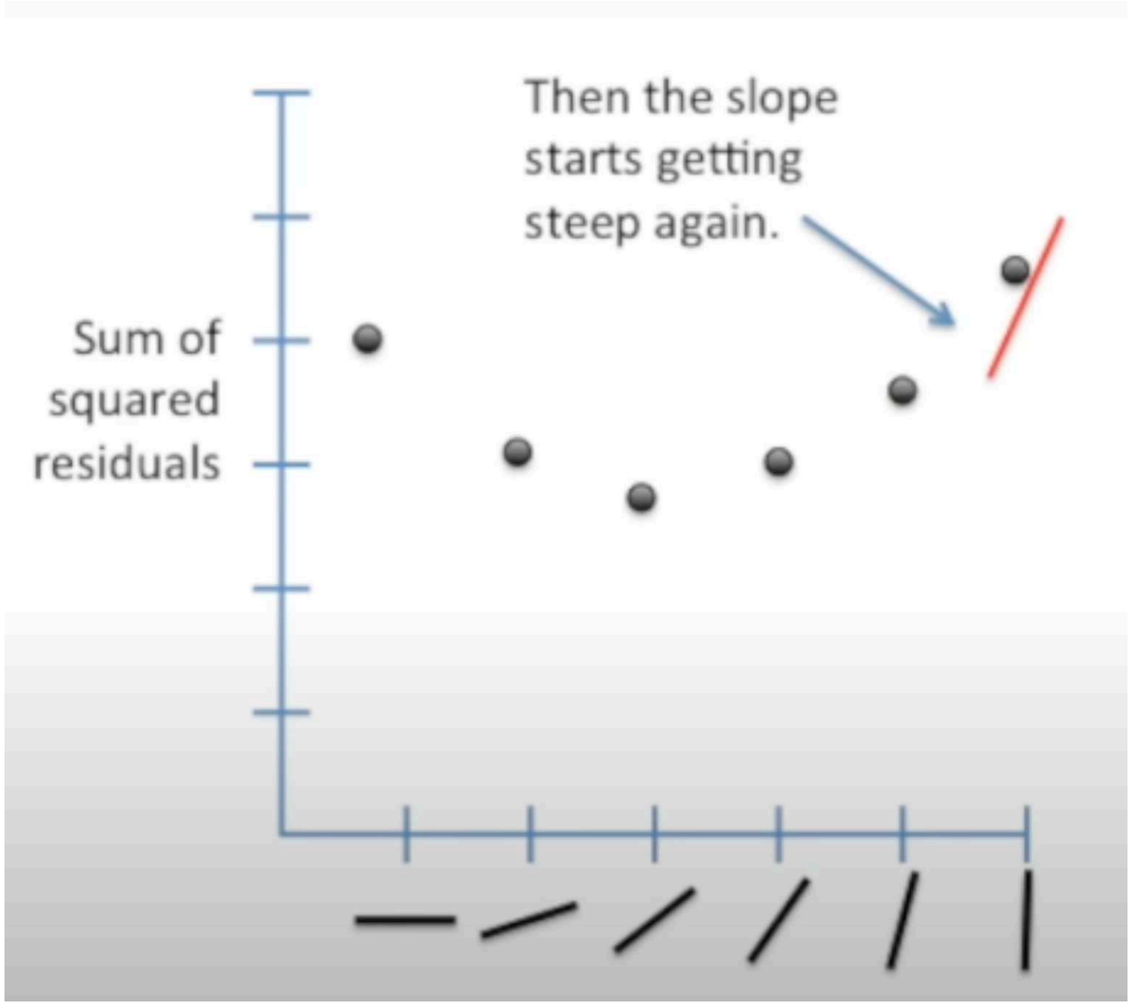
Slide Courtesy: StatQuest

Linear Regression



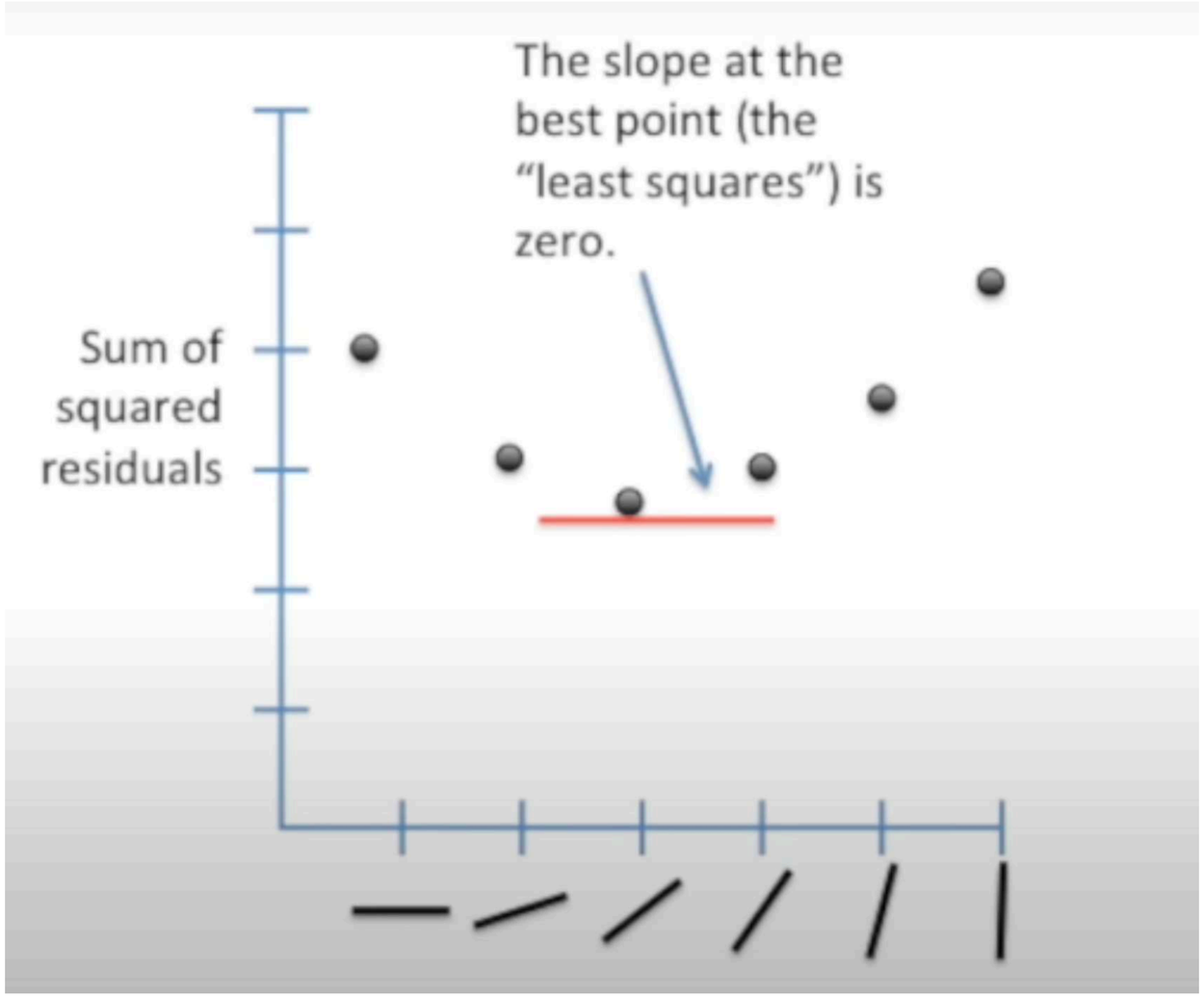
Slide Courtesy: StatQuest

Linear Regression



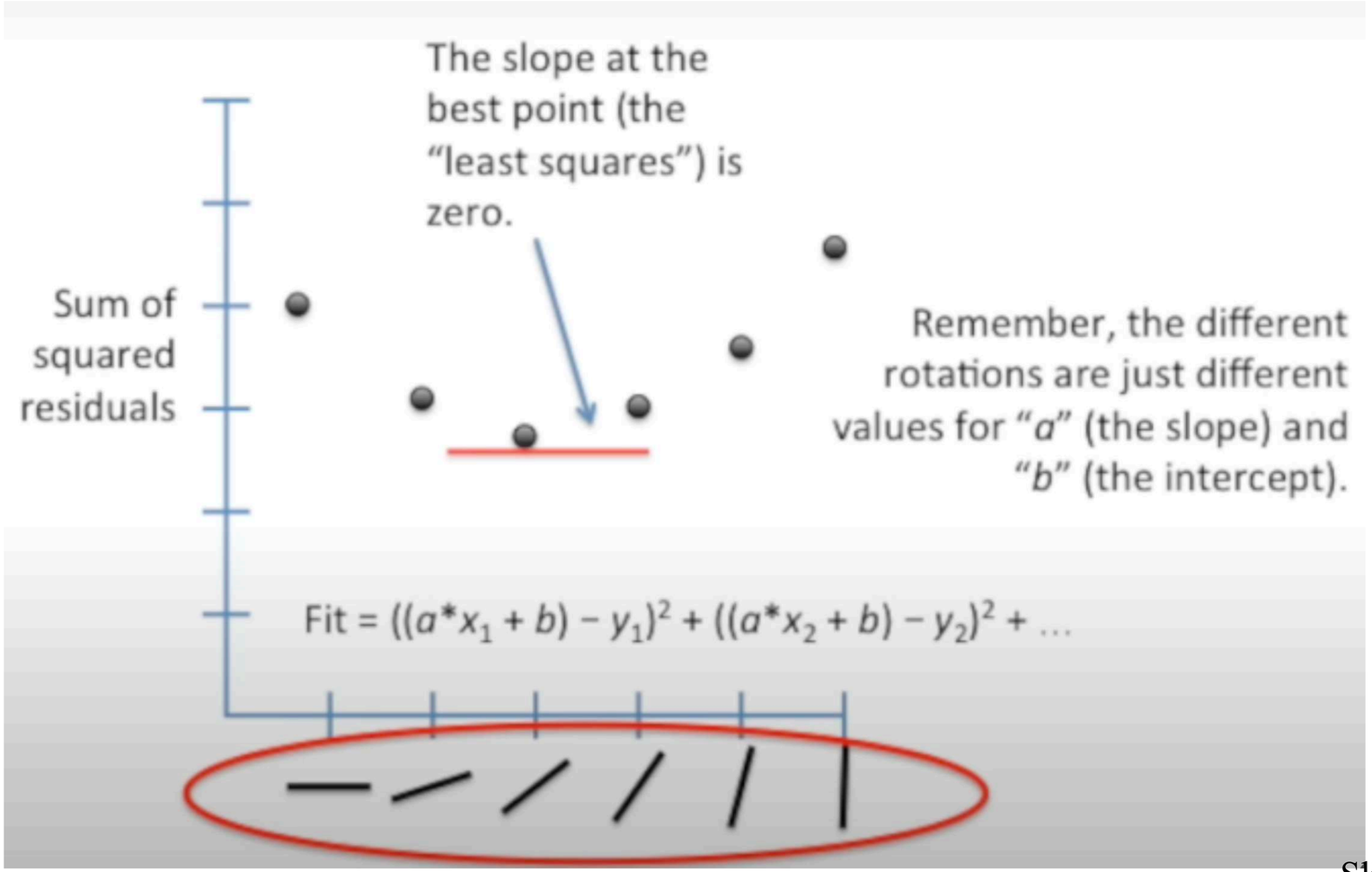
Slide Courtesy: StatQuest

Linear Regression



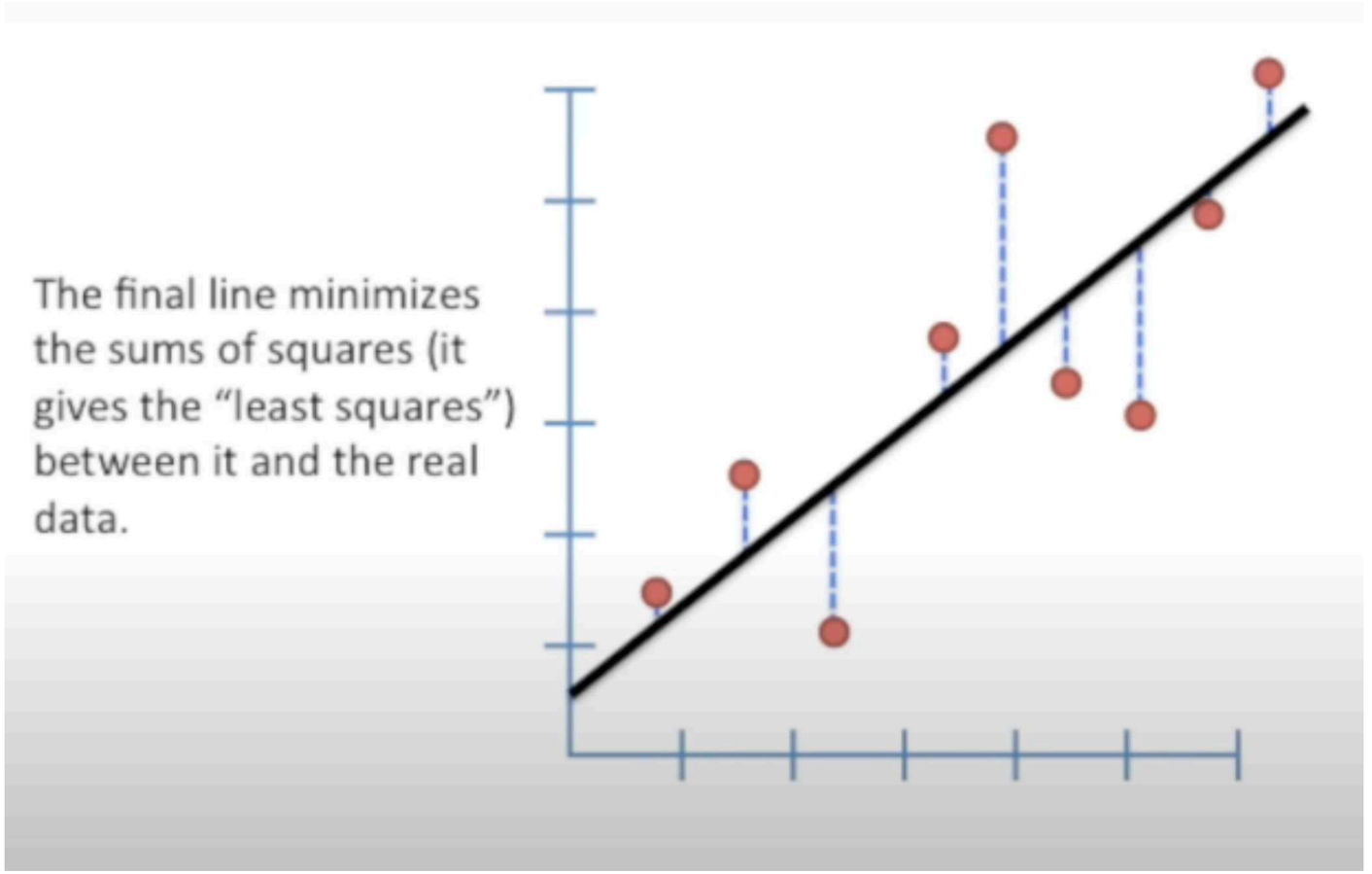
Slide Courtesy: StatQuest

Linear Regression



Slide Courtesy: StatQuest

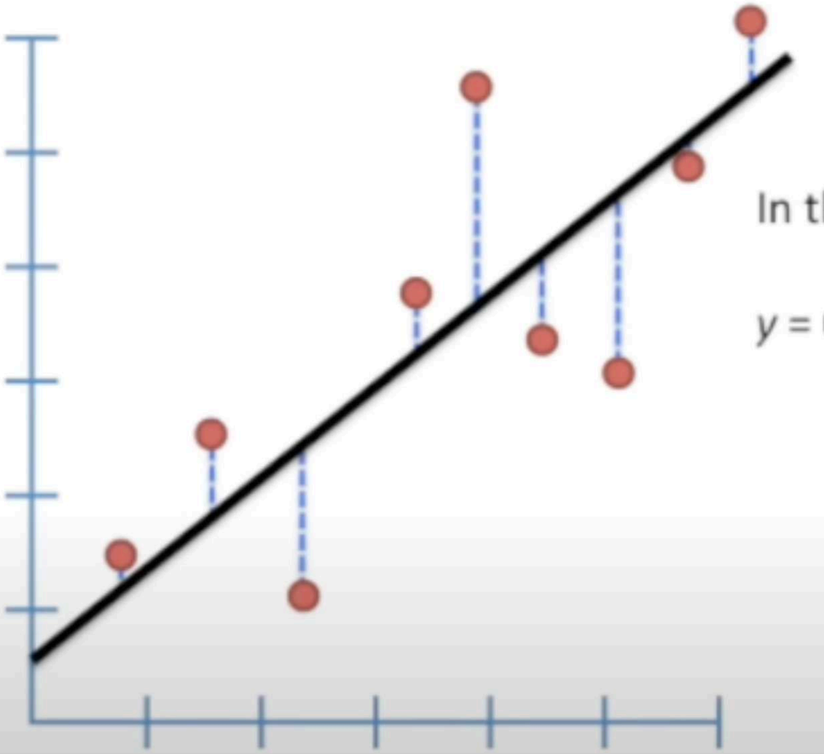
Linear Regression



Slide Courtesy: StatQuest

Linear Regression

The final line minimizes the sums of squares (it gives the "least squares") between it and the real data.



In this case, it is

$y = 0.77*x + 0.66$

Slide Courtesy: StatQuest