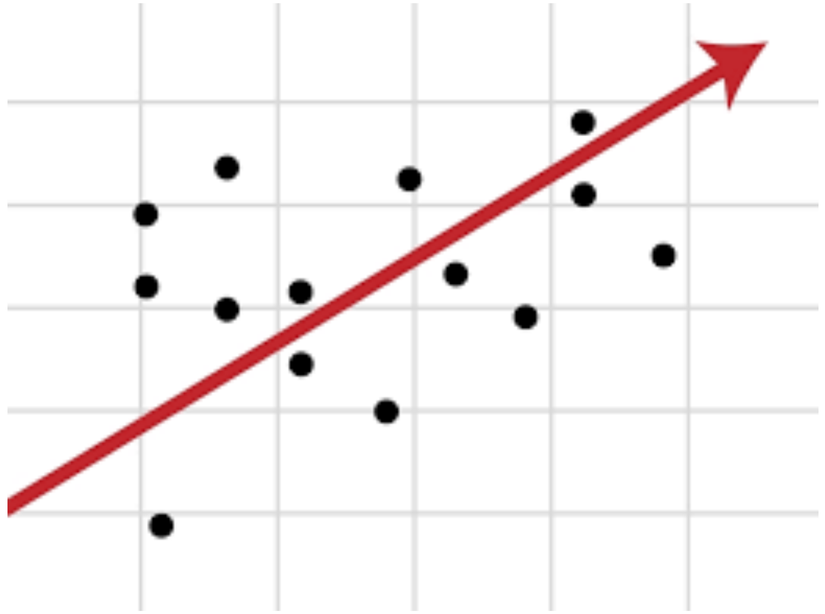
DATA ANALYTICS- UNIT 2

By

DEEPIKA KAMBOJ

Regression modelling

Regression modelling is a statistical method used to predict a continuous outcome variable (also known as the dependent variable) based on one or more predictor variables (also known as independent variables).



Types of Regression Analysis

- Linear regression
- Logistic regression
- Polynomial regression
- Multiple regression
- Non-linear regression
- Regularized regression

Linear Regression

Fitting a Line to Data

Aka

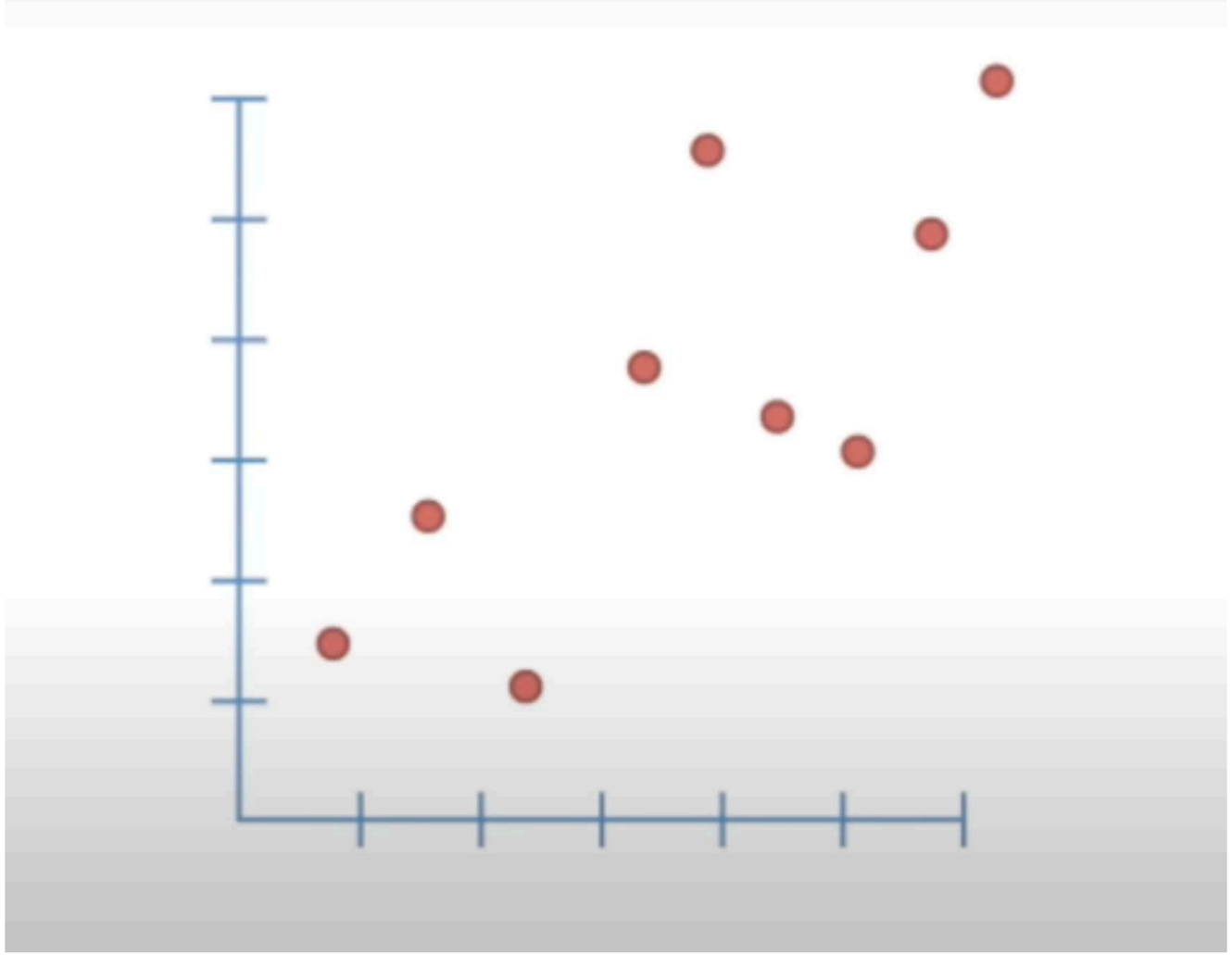
Least Squares

Aka

Linear Regression

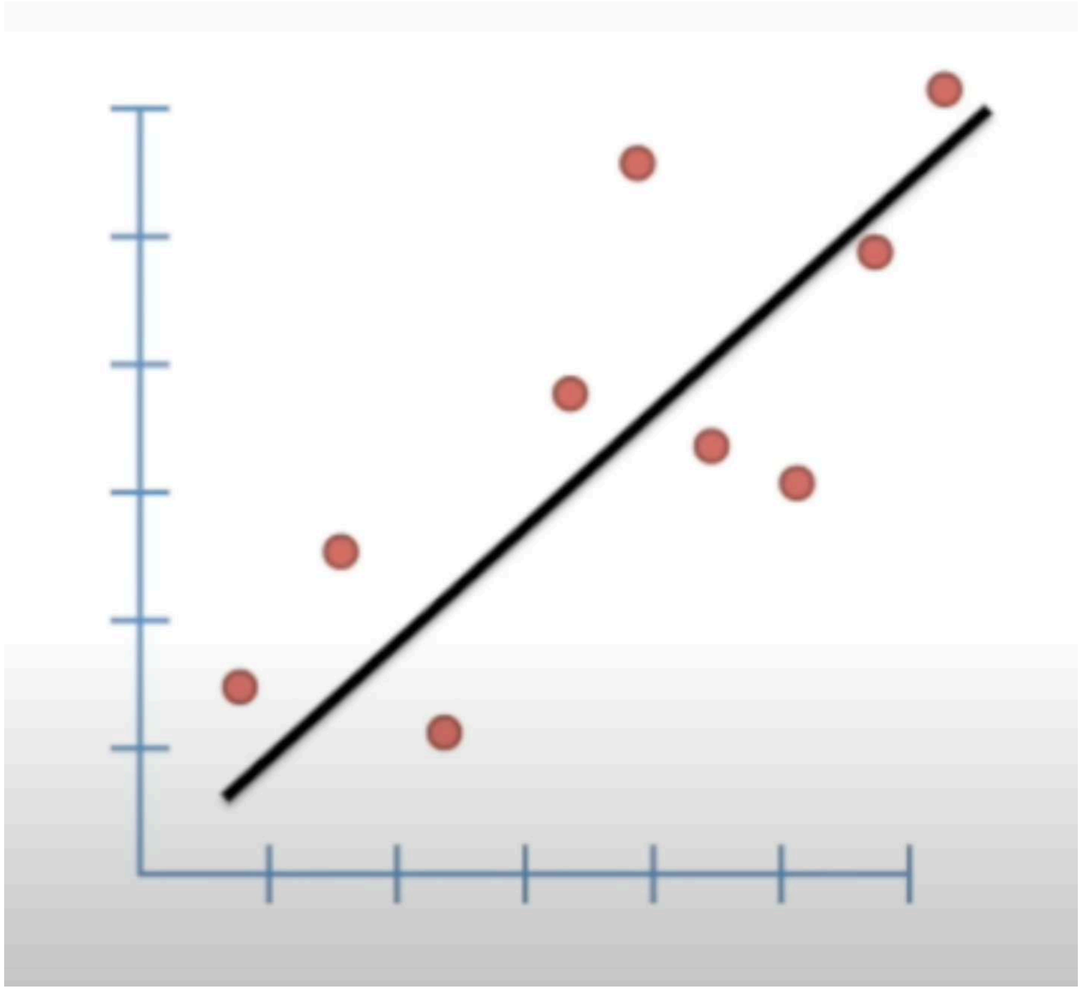
Slide Courtesy: StatQuest

Linear Regression



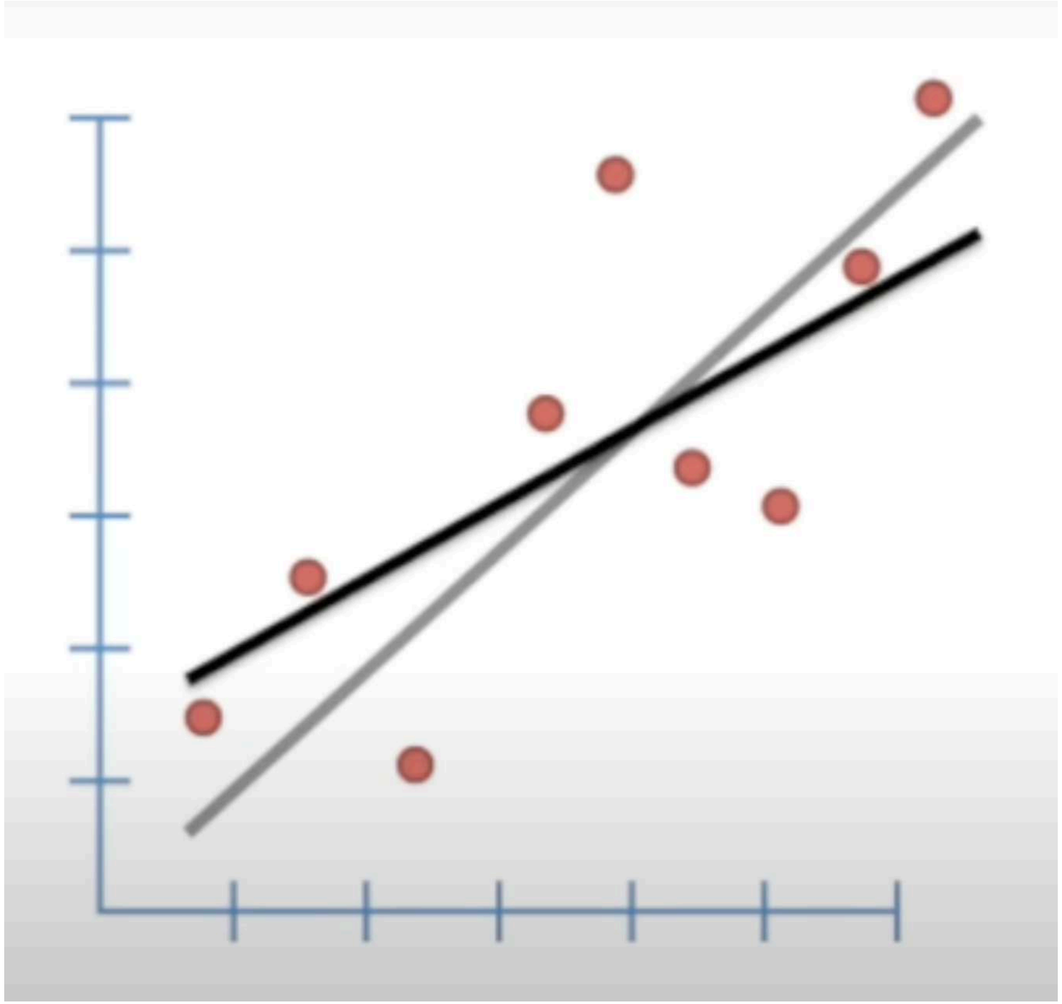
Slide Courtesy: StatQuest

Linear Regression



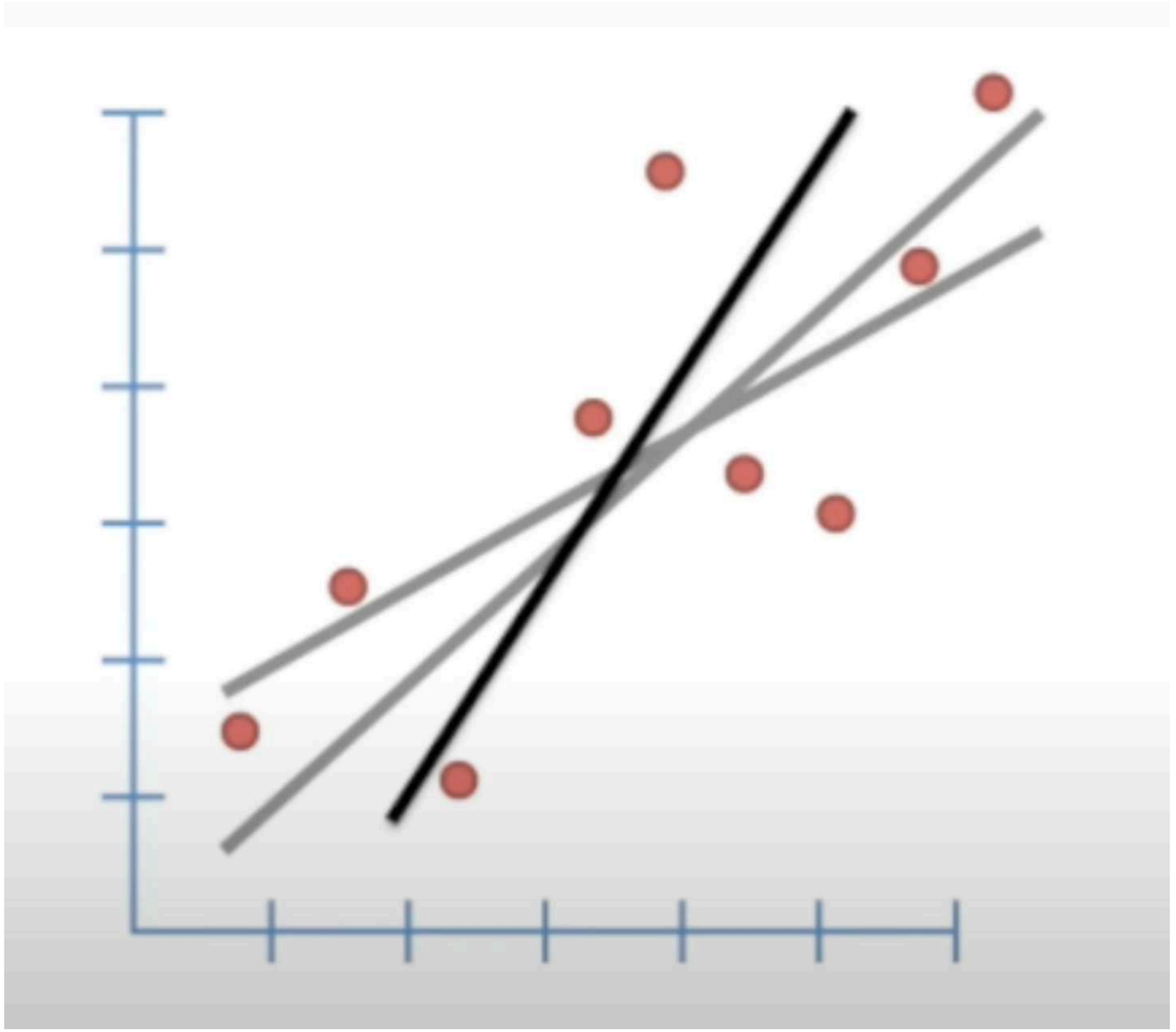
Slide Courtesy: StatQuest

Linear Regression



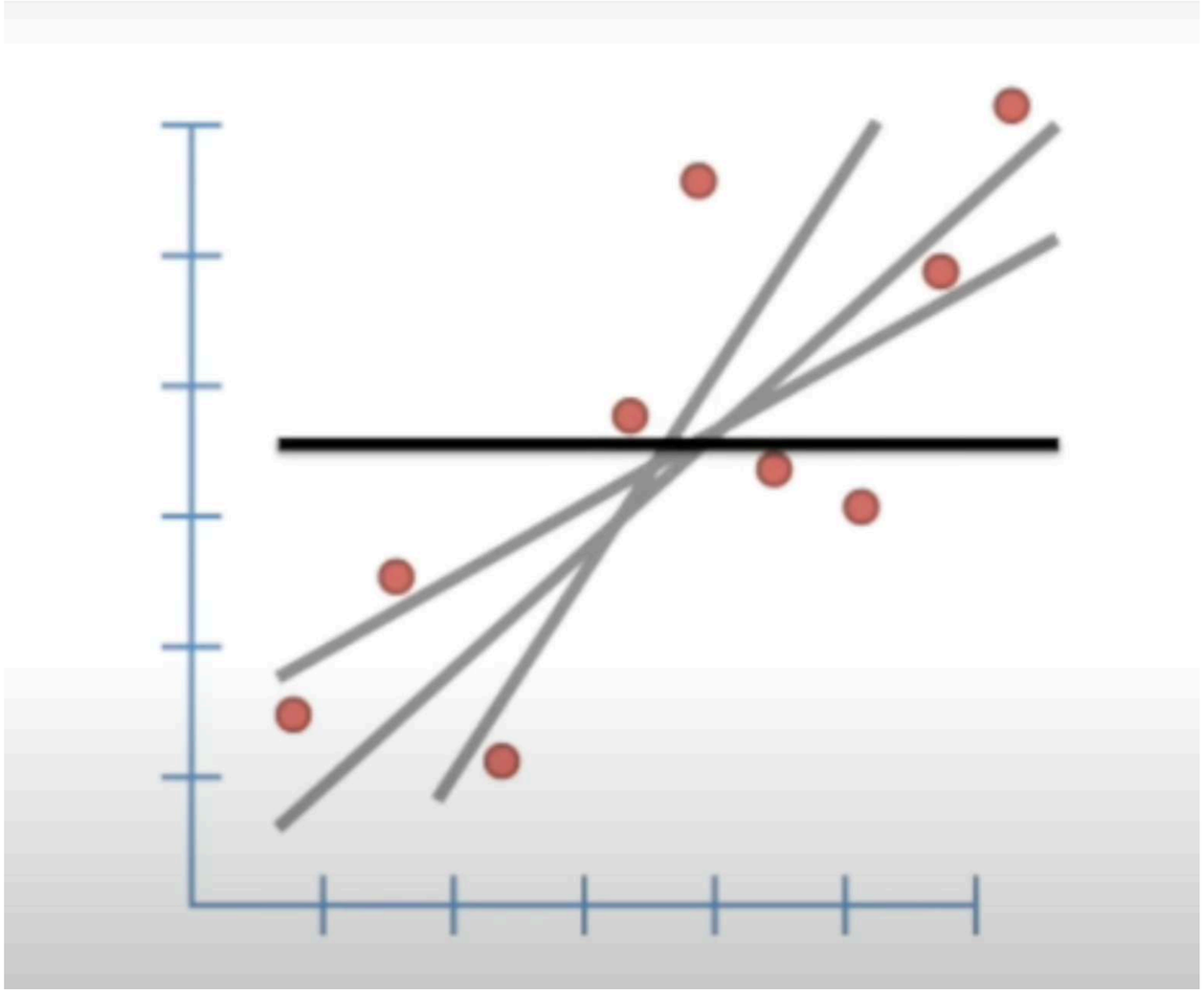
Slide Courtesy: StatQuest

Linear Regression



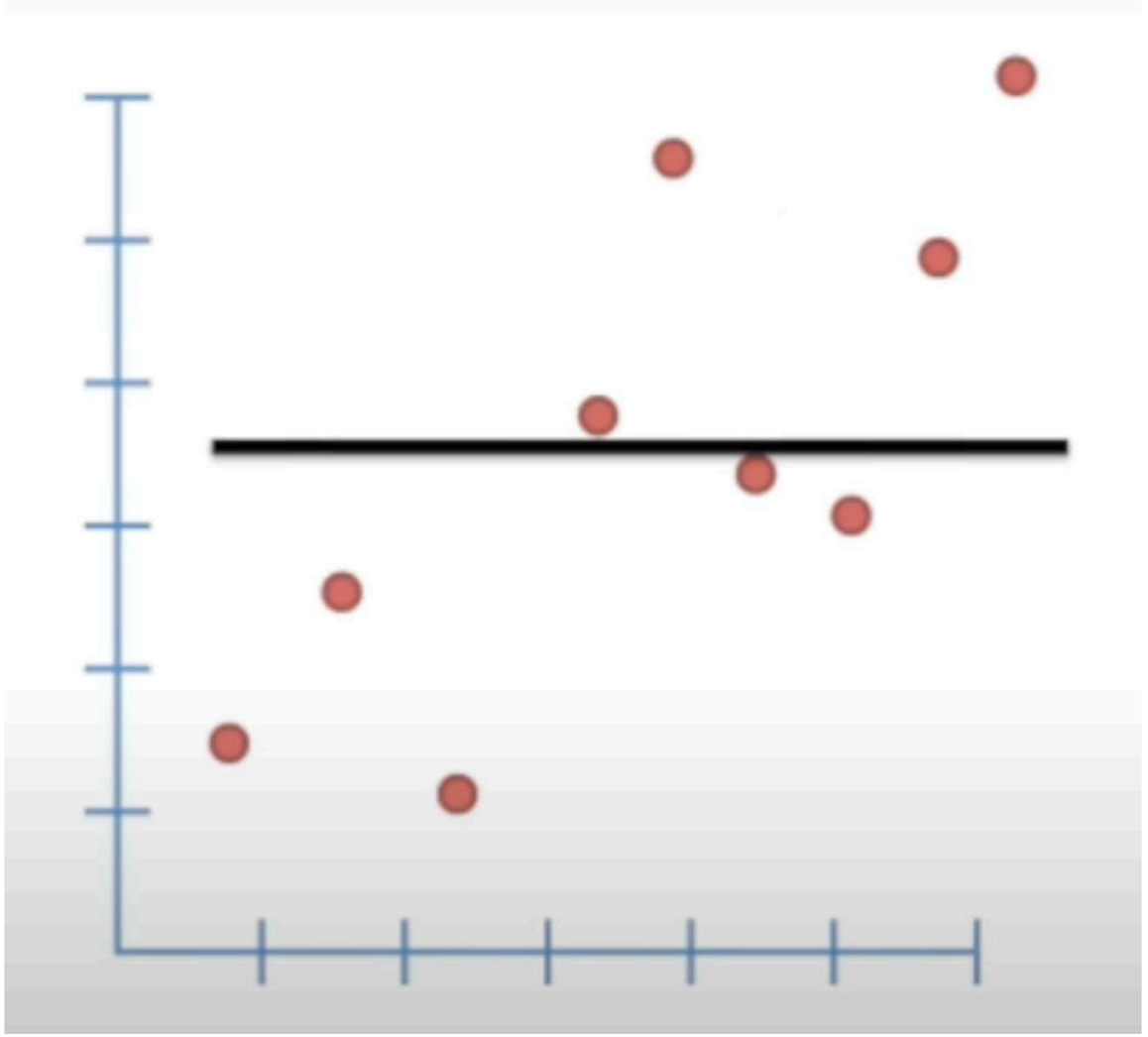
Slide Courtesy: StatQuest

Linear Regression



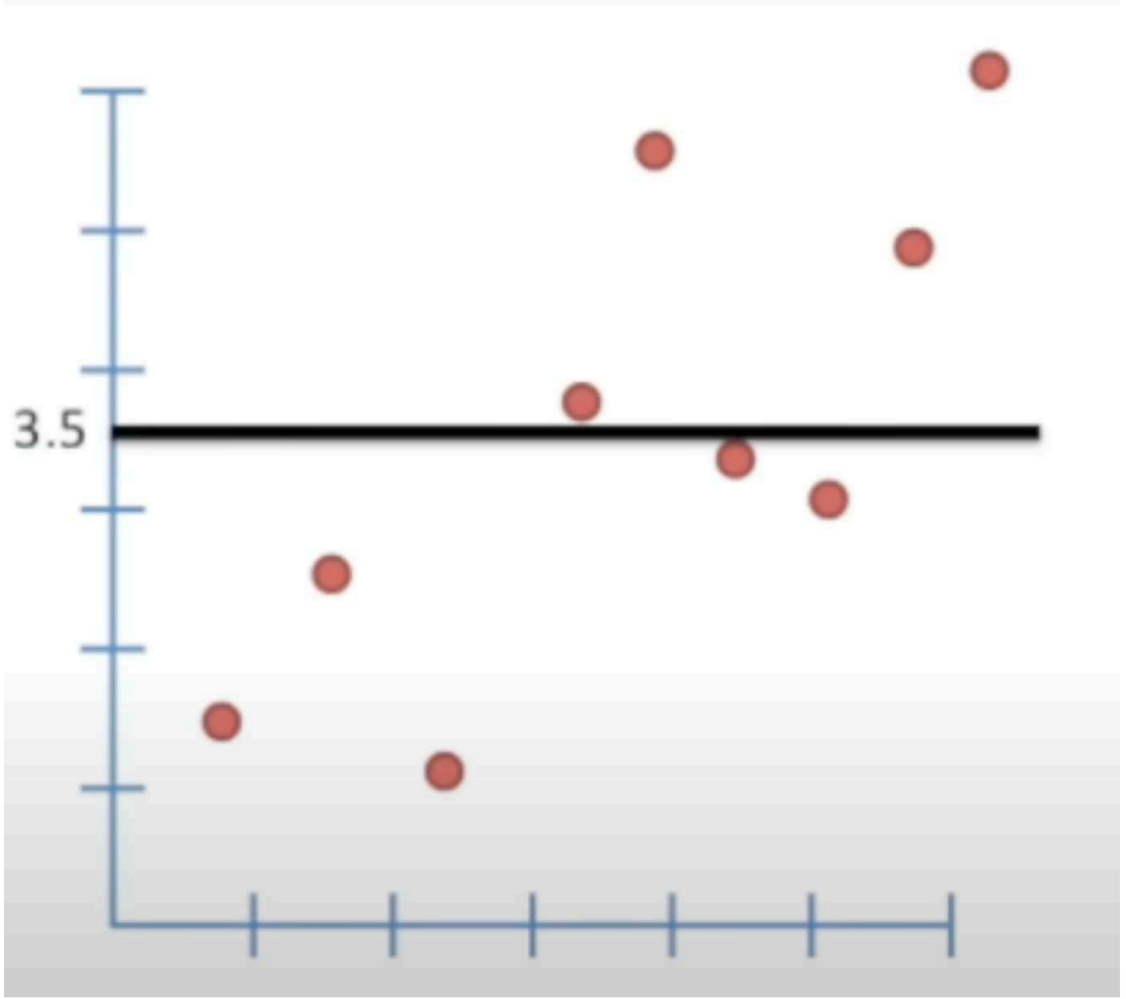
Slide Courtesy: StatQuest

Linear Regression



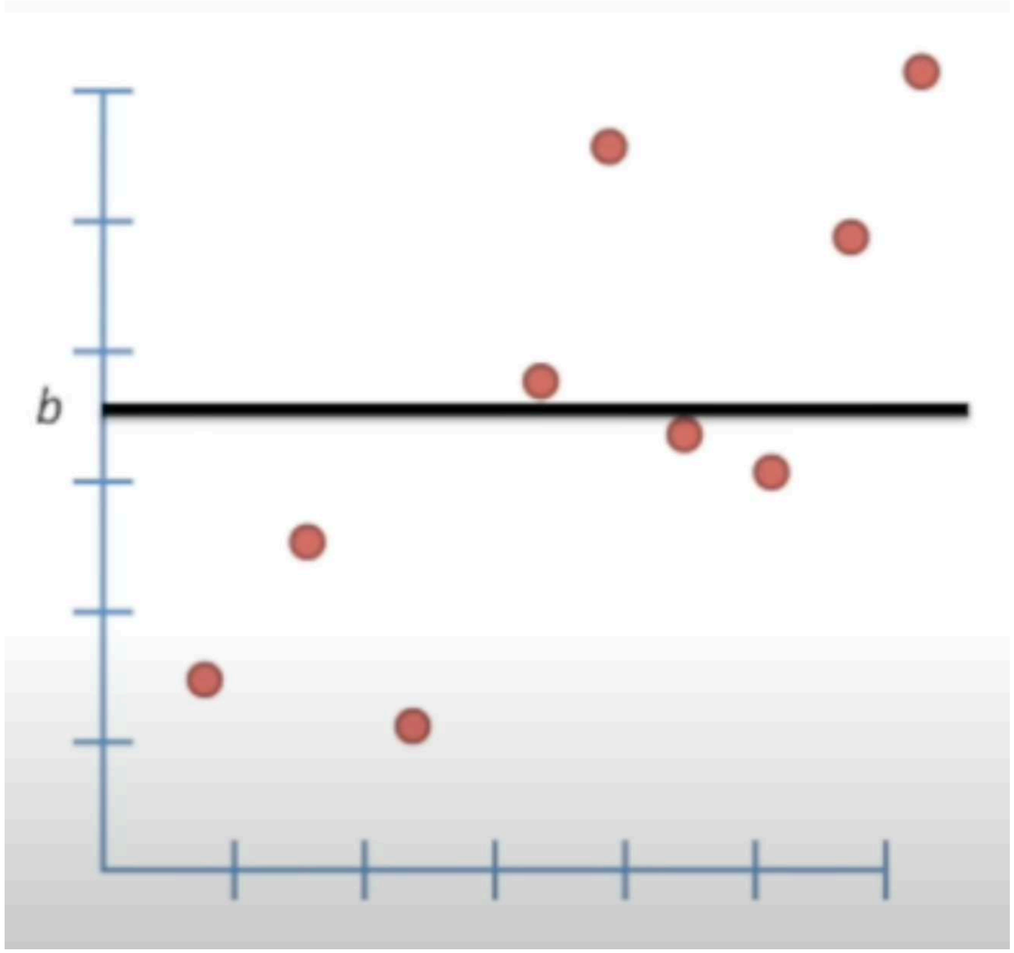
Slide Courtesy: StatQuest

Linear Regression



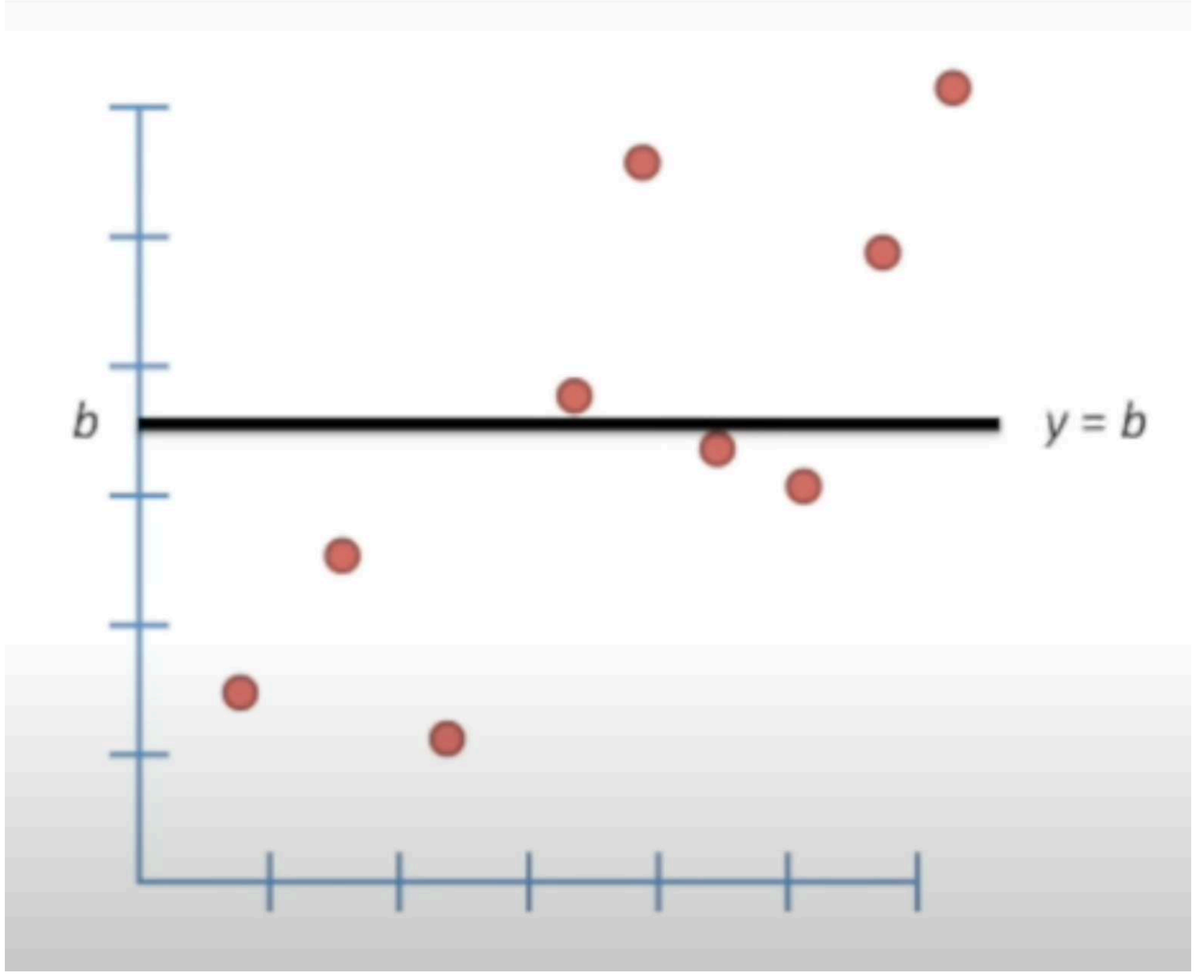
Slide Courtesy: StatQuest

Linear Regression



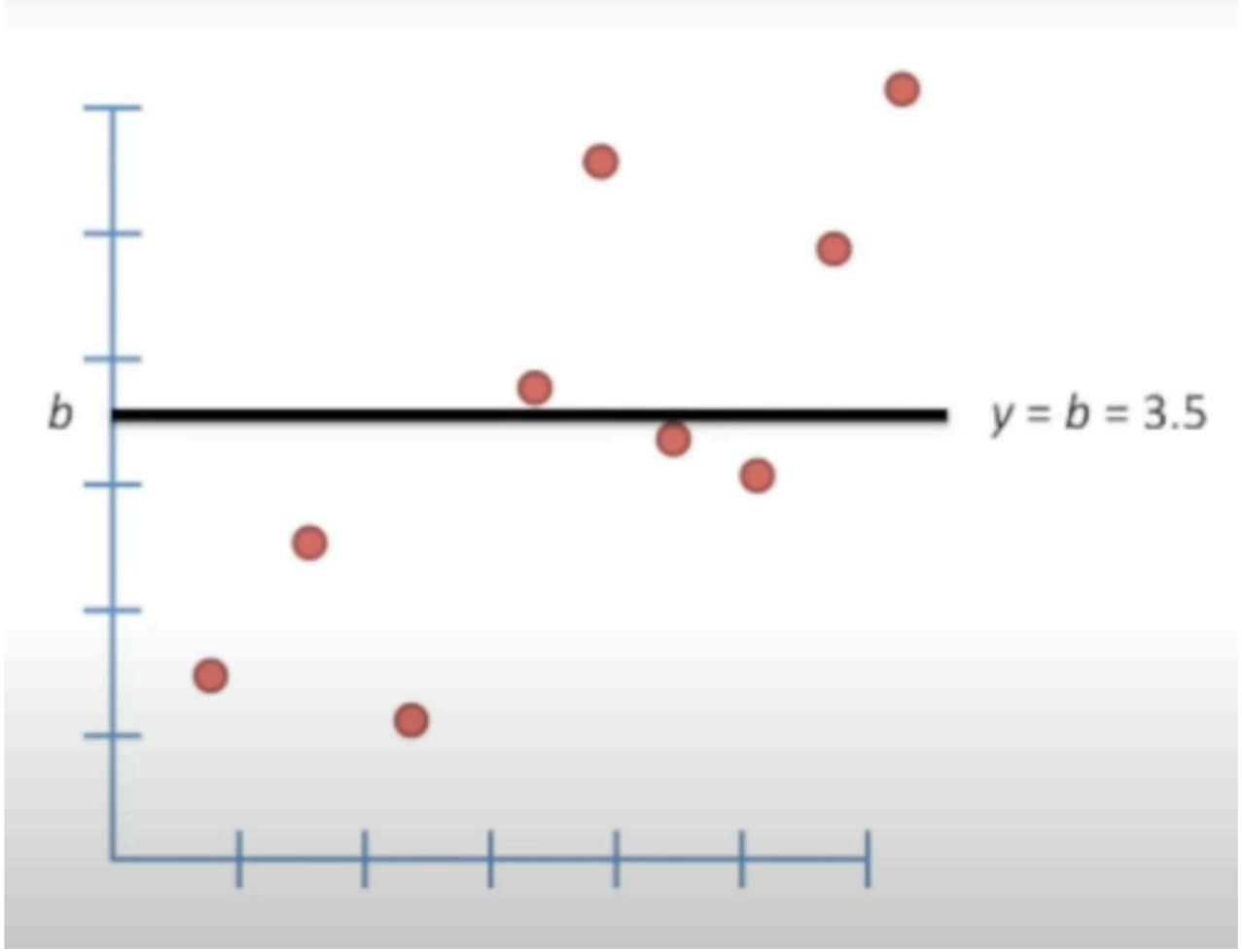
Slide Courtesy: StatQuest

Linear Regression



Slide Courtesy: StatQuest

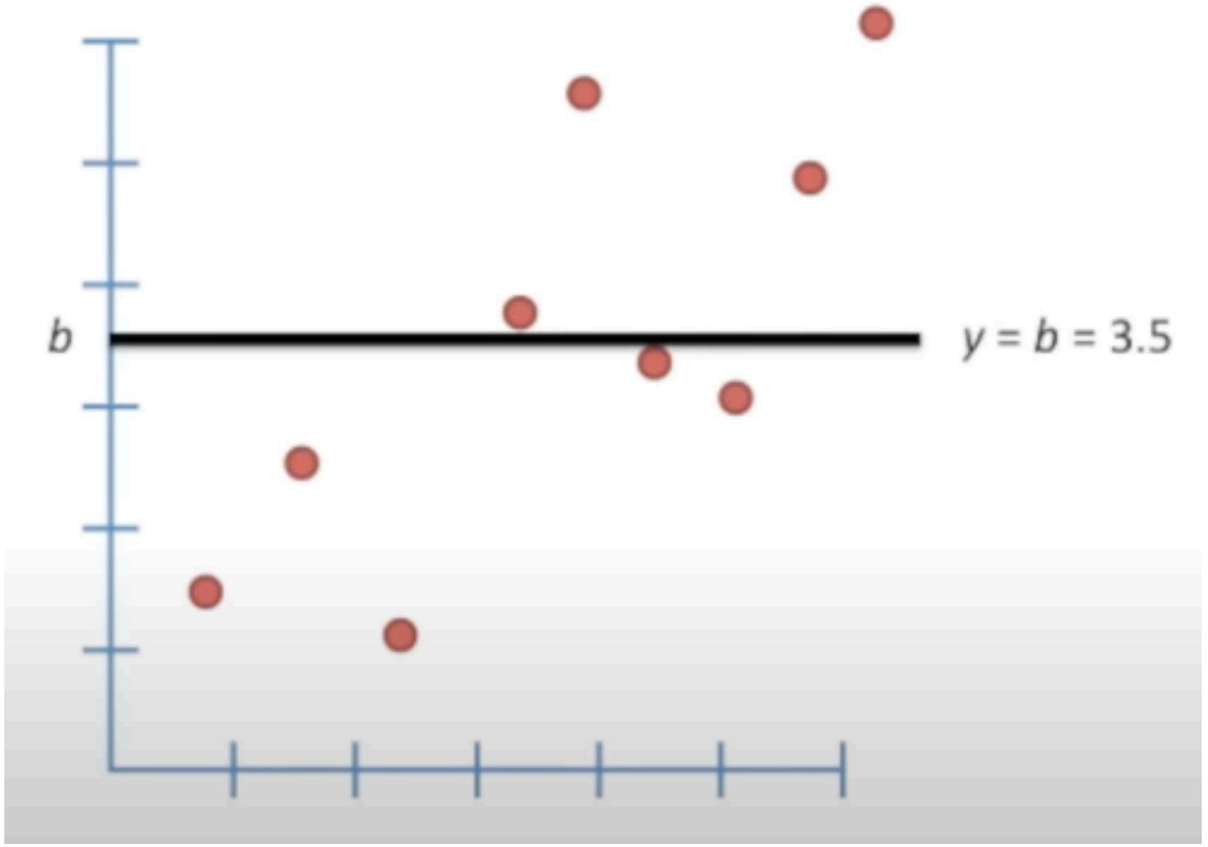
Linear Regression



Slide Courtesy: StatQuest

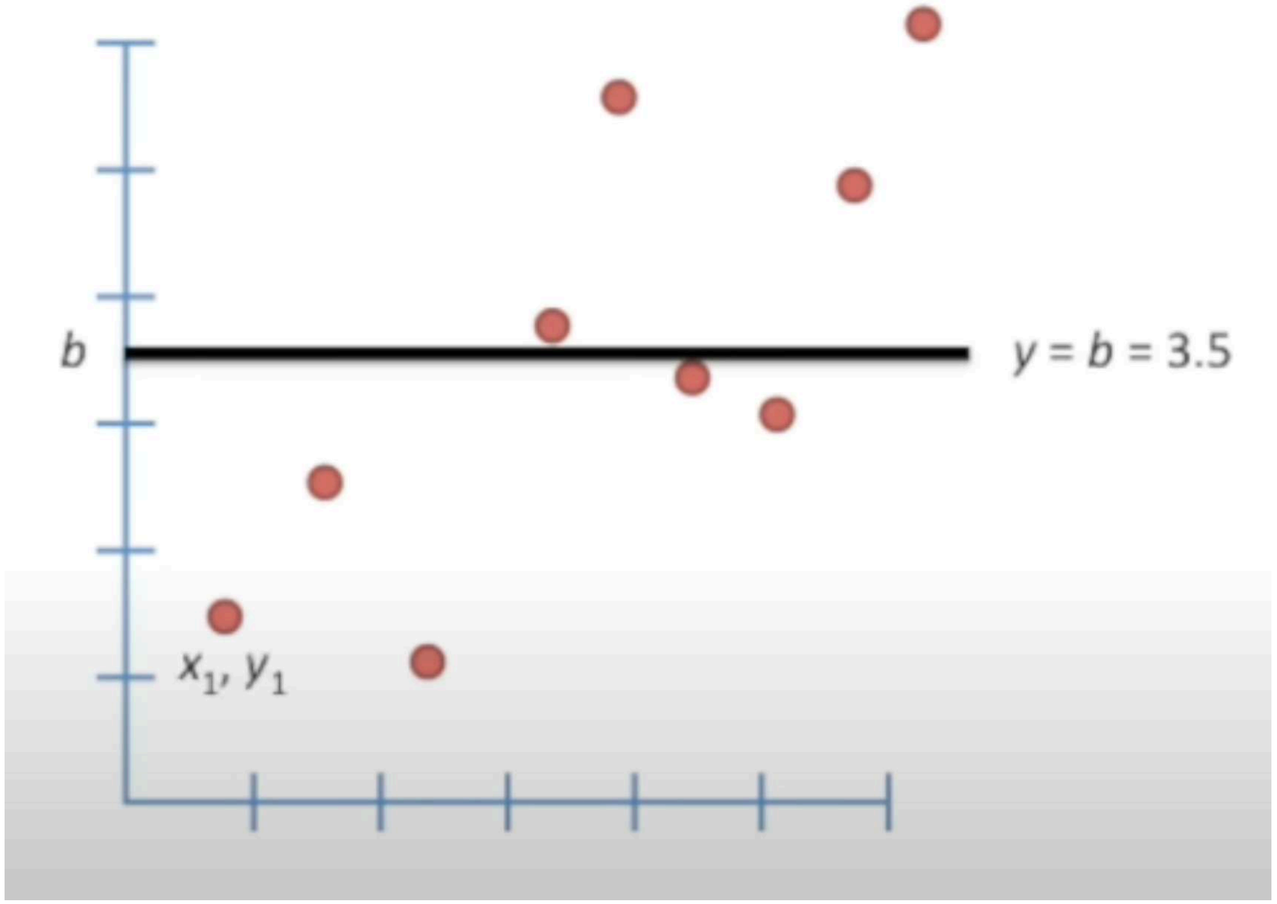
Linear Regression

We can measure how well this line fits the data by seeing how close it is to the data points



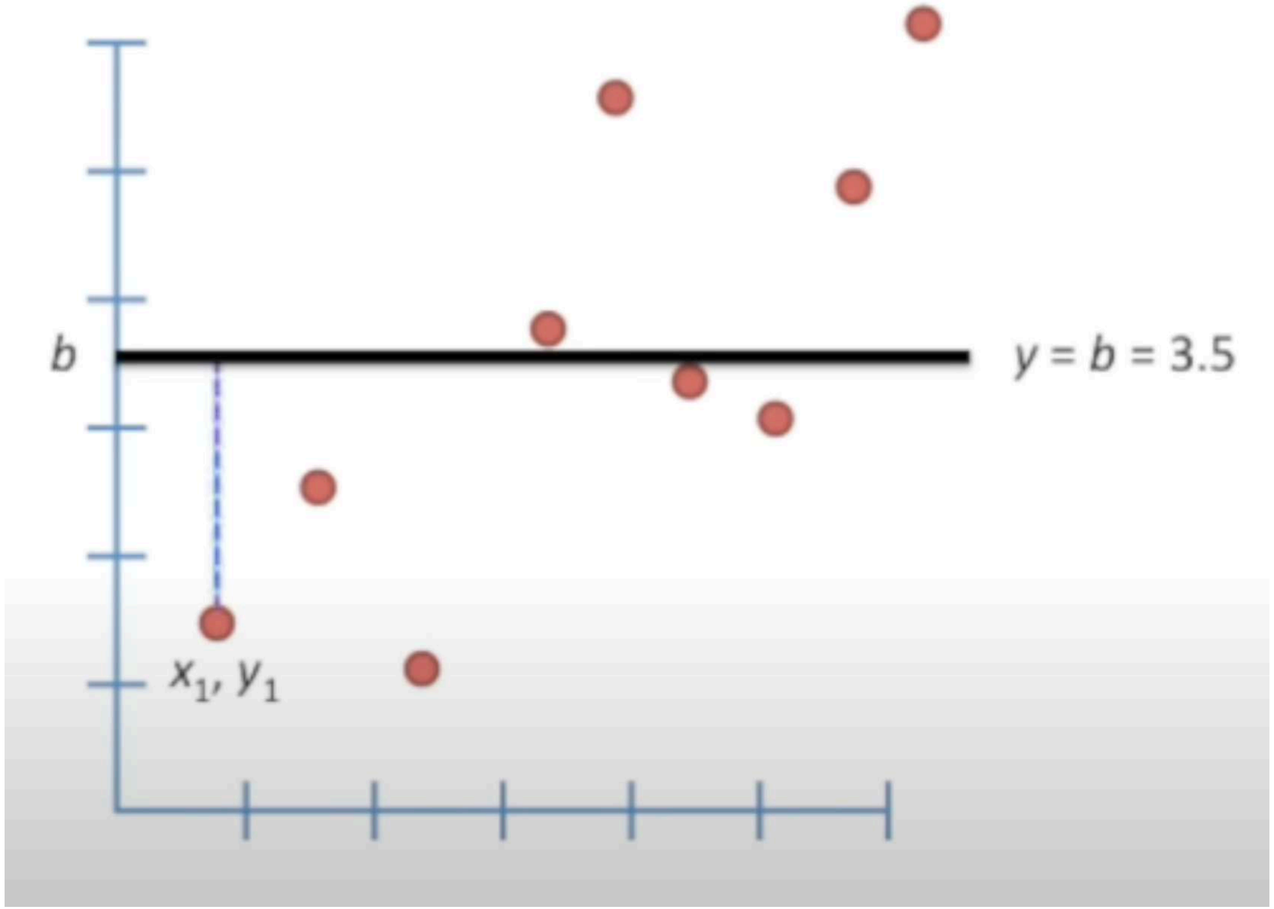
Slide Courtesy: StatQuest

Linear Regression



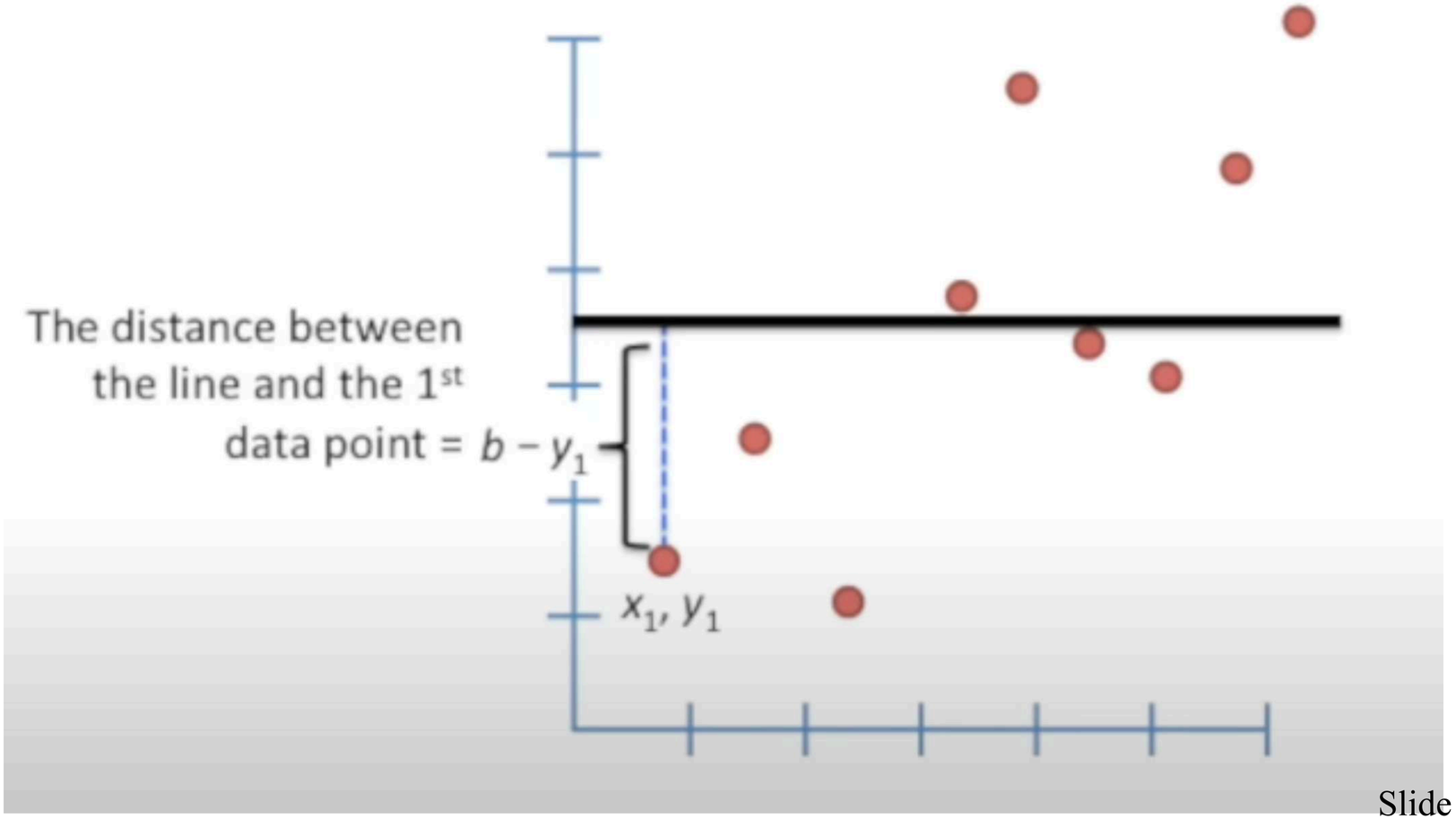
Slide Courtesy: StatQuest

Linear Regression



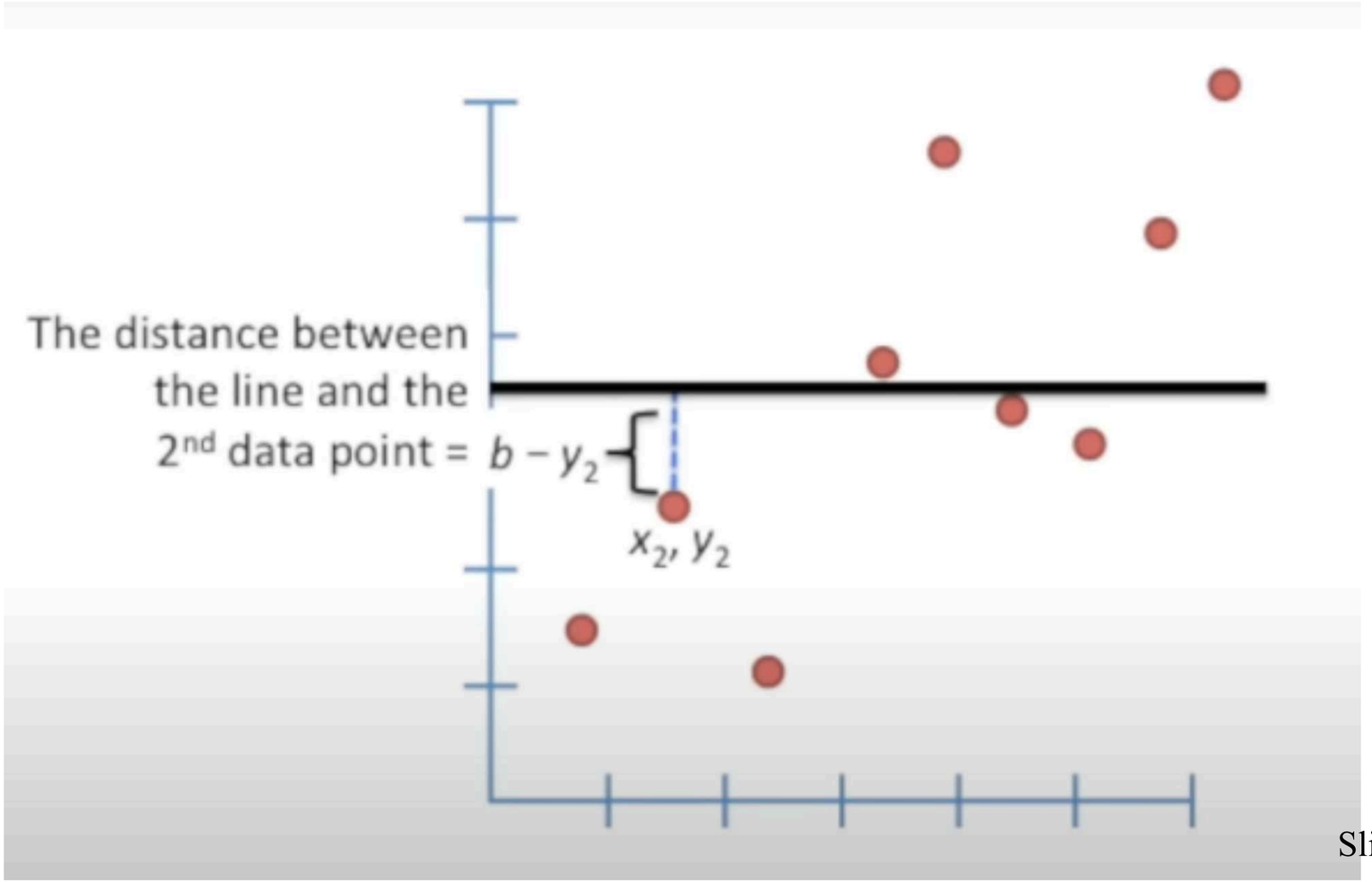
Slide Courtesy: StatQuest

Linear Regression



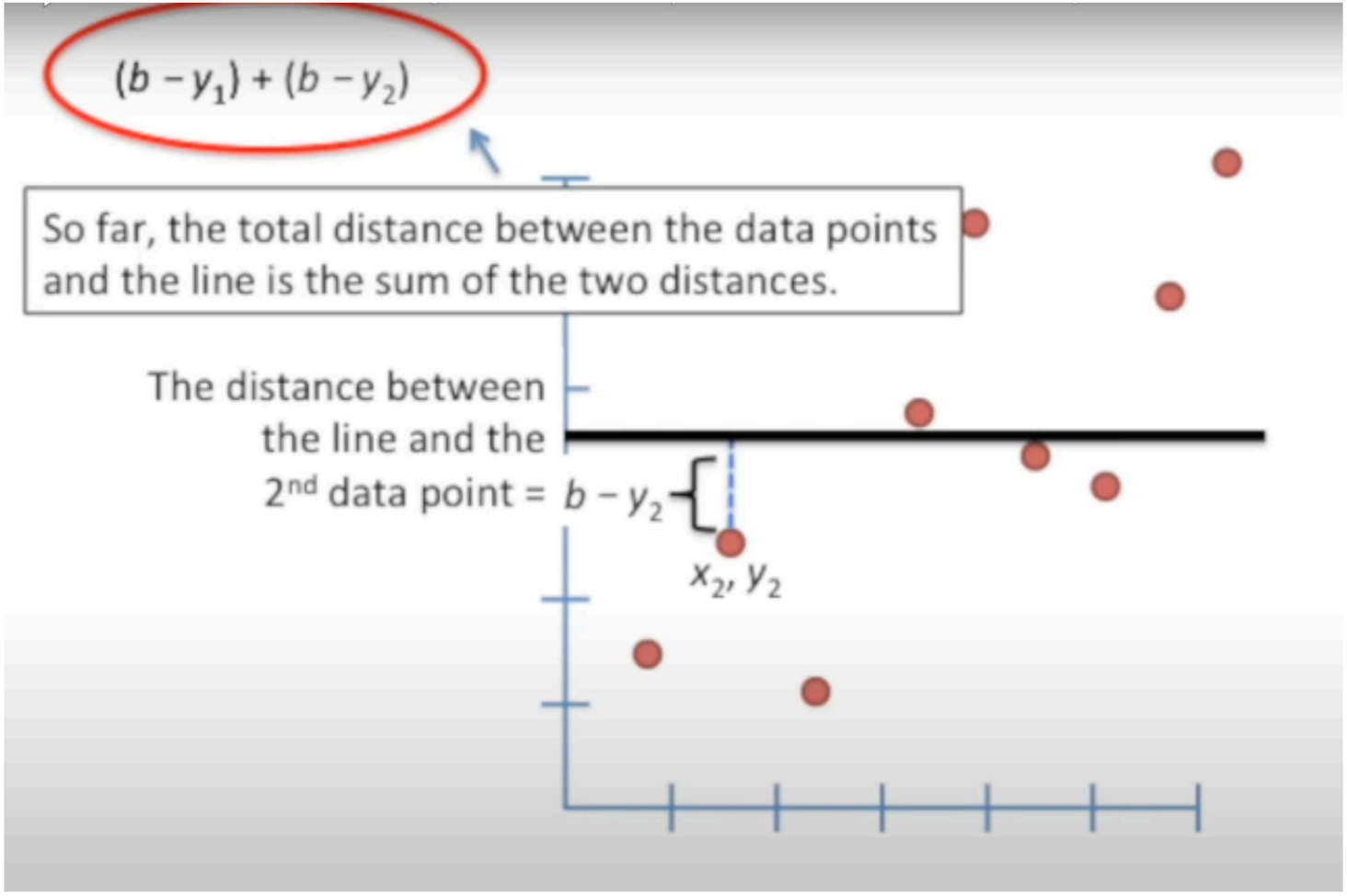
Slide Courtesy: StatQuest

Linear Regression



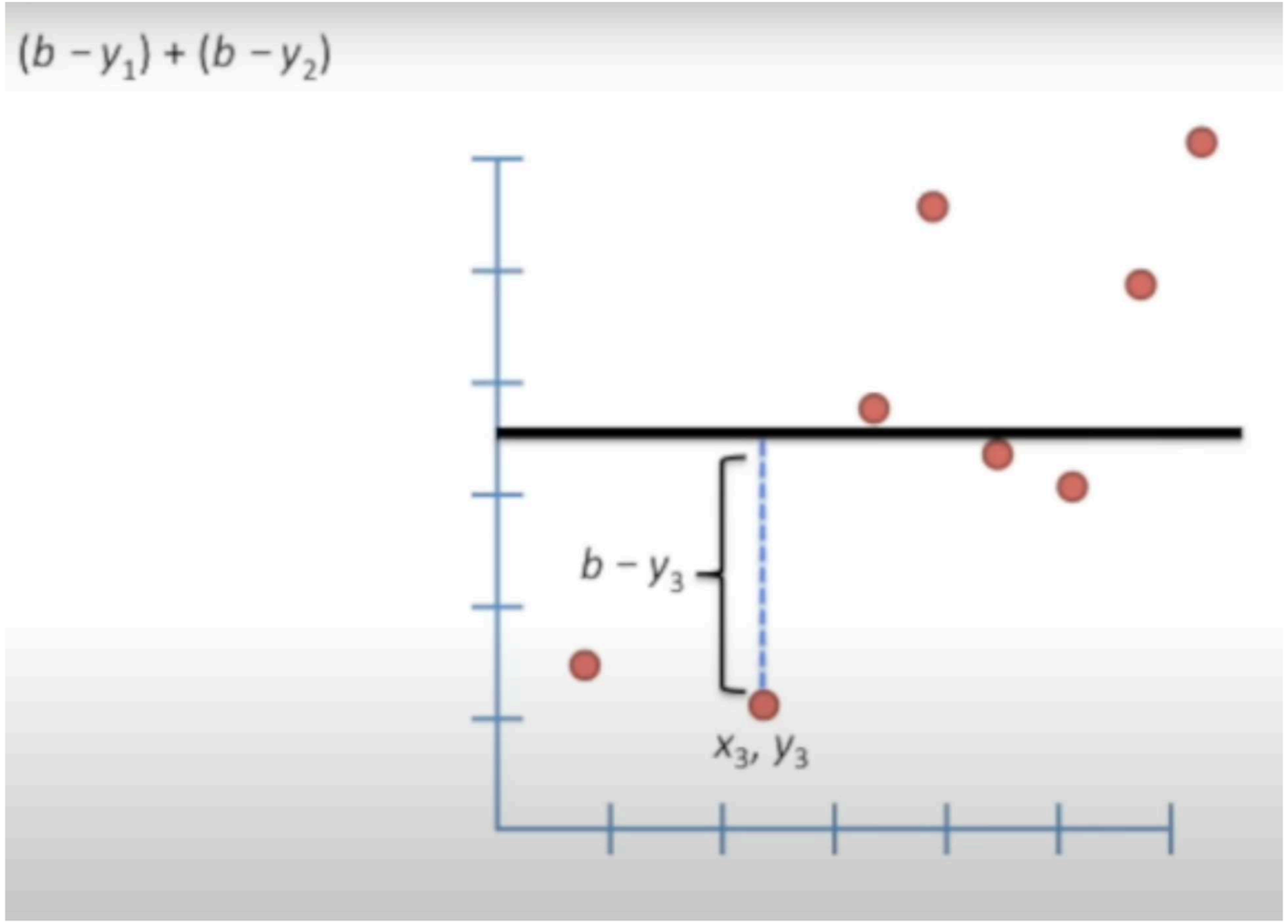
Slide Courtesy: StatQuest

Linear Regression



Slide Courtesy: StatQuest

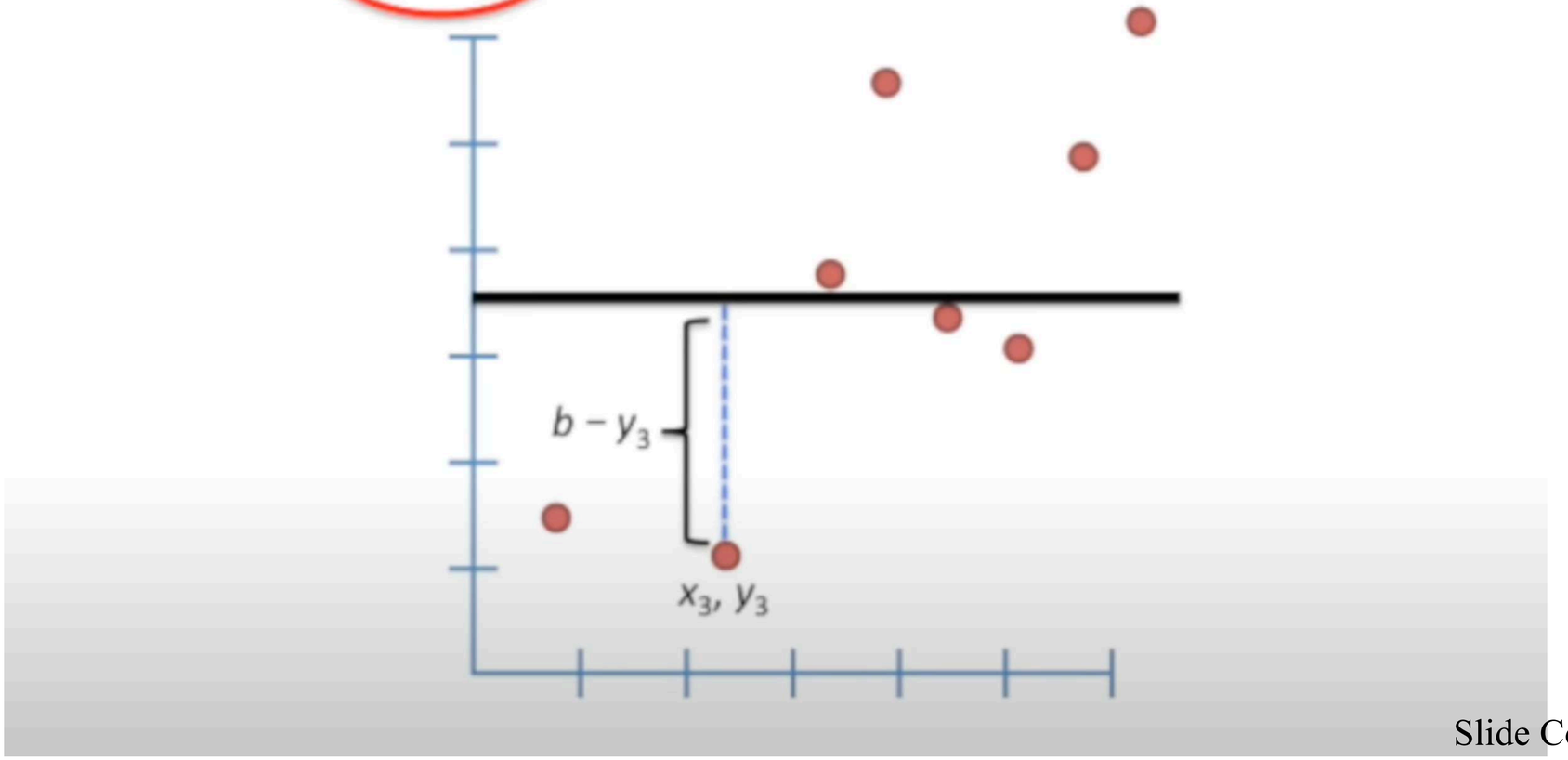
Linear Regression



Slide Courtesy: StatQuest

Linear Regression

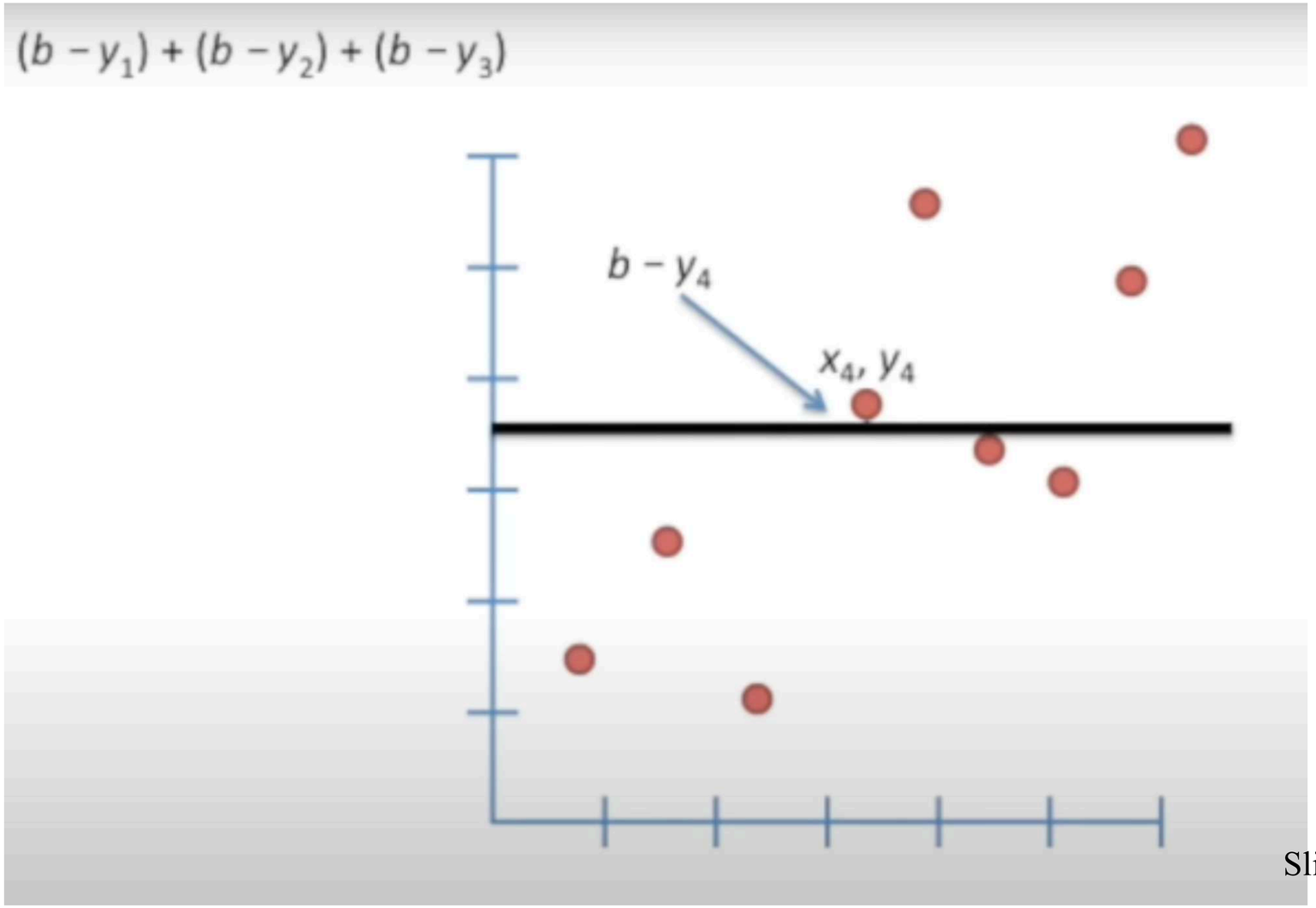
$(b - y_1) + (b - y_2) + (b - y_3)$ Now we've add the 3rd distance to our total sum.



Slide Courtesy: StatQuest

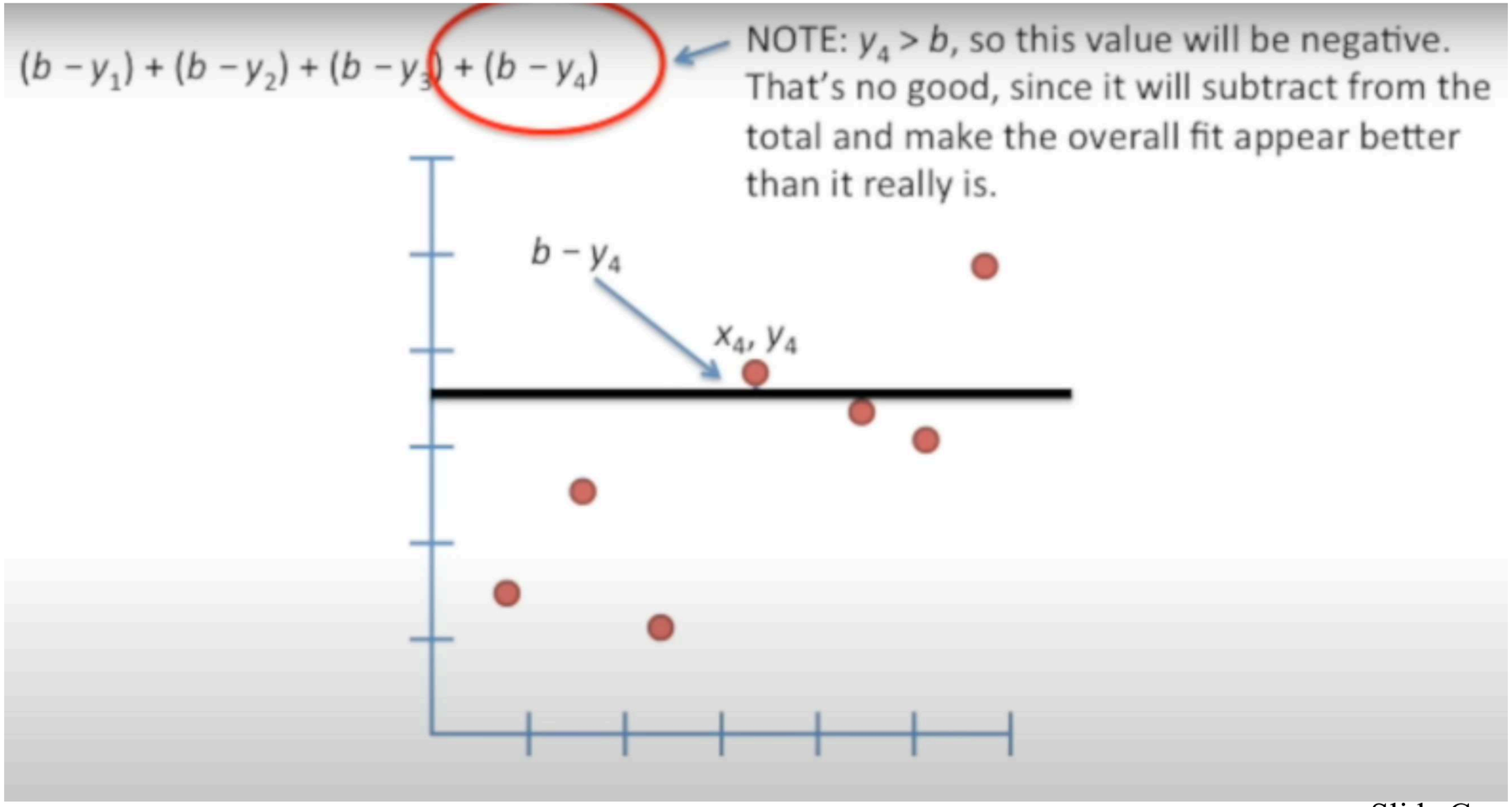
Linear Regression

$(b - y_1) + (b - y_2) + (b - y_3)$



Slide Courtesy: StatQuest

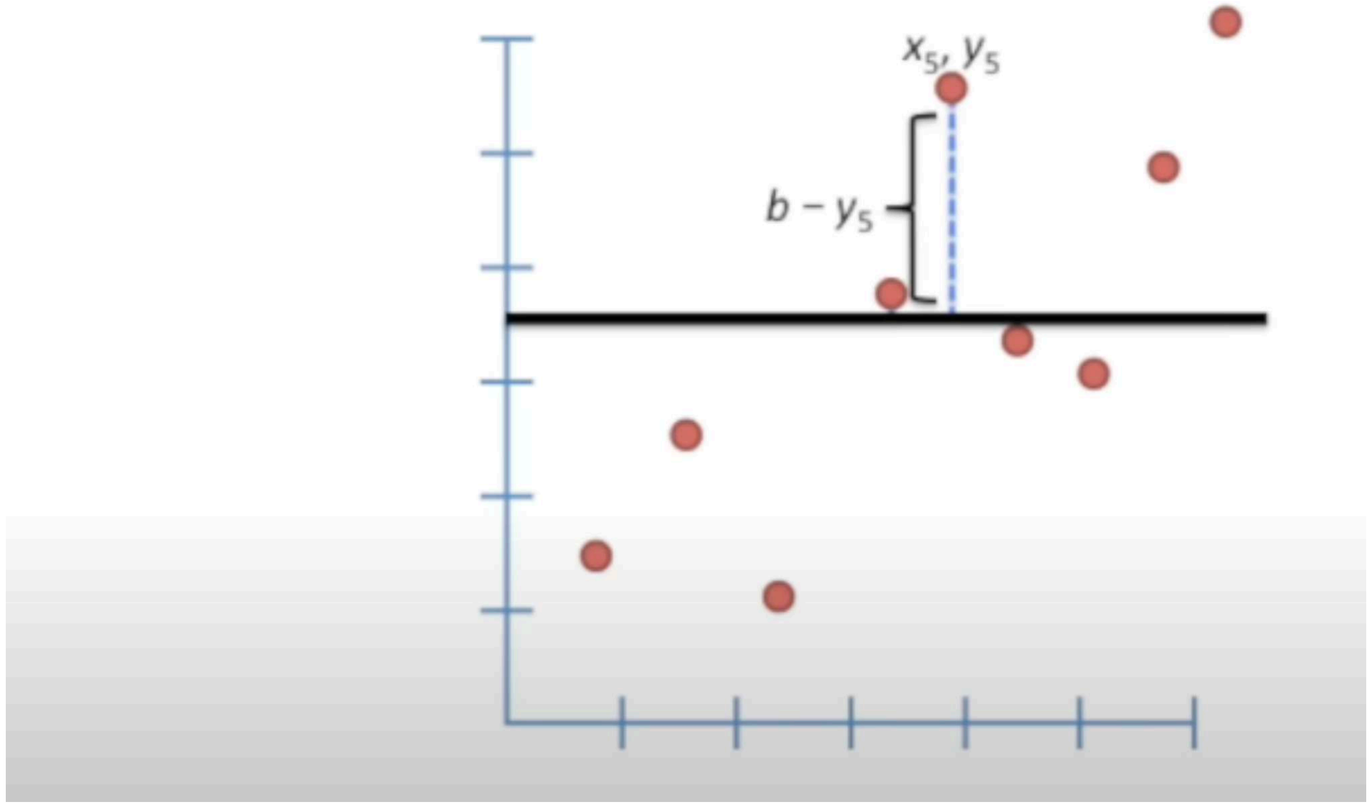
Linear Regression



Slide Courtesy: StatQuest

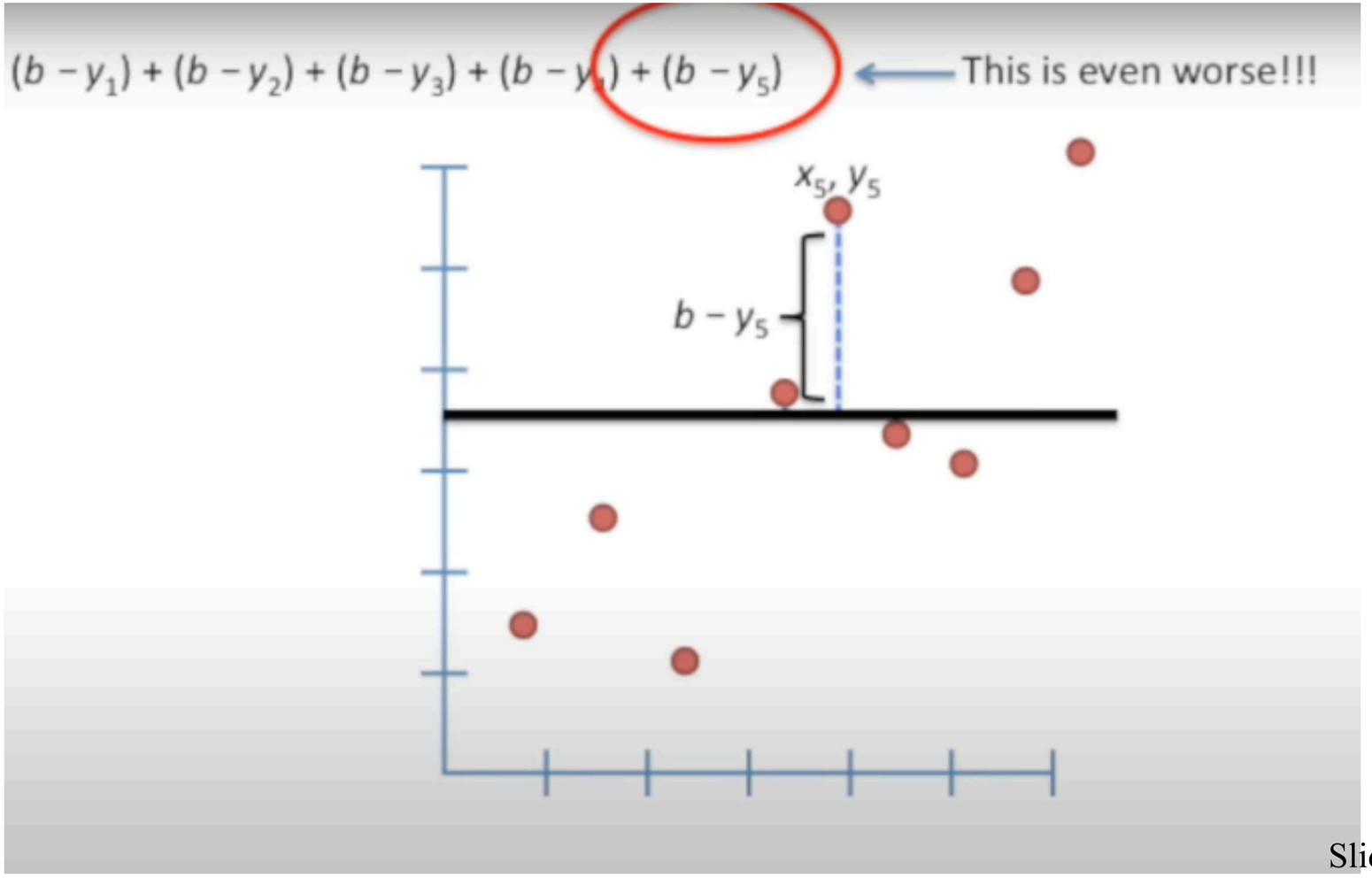
Linear Regression

$$(b - y_1) + (b - y_2) + (b - y_3) + (b - y_4)$$



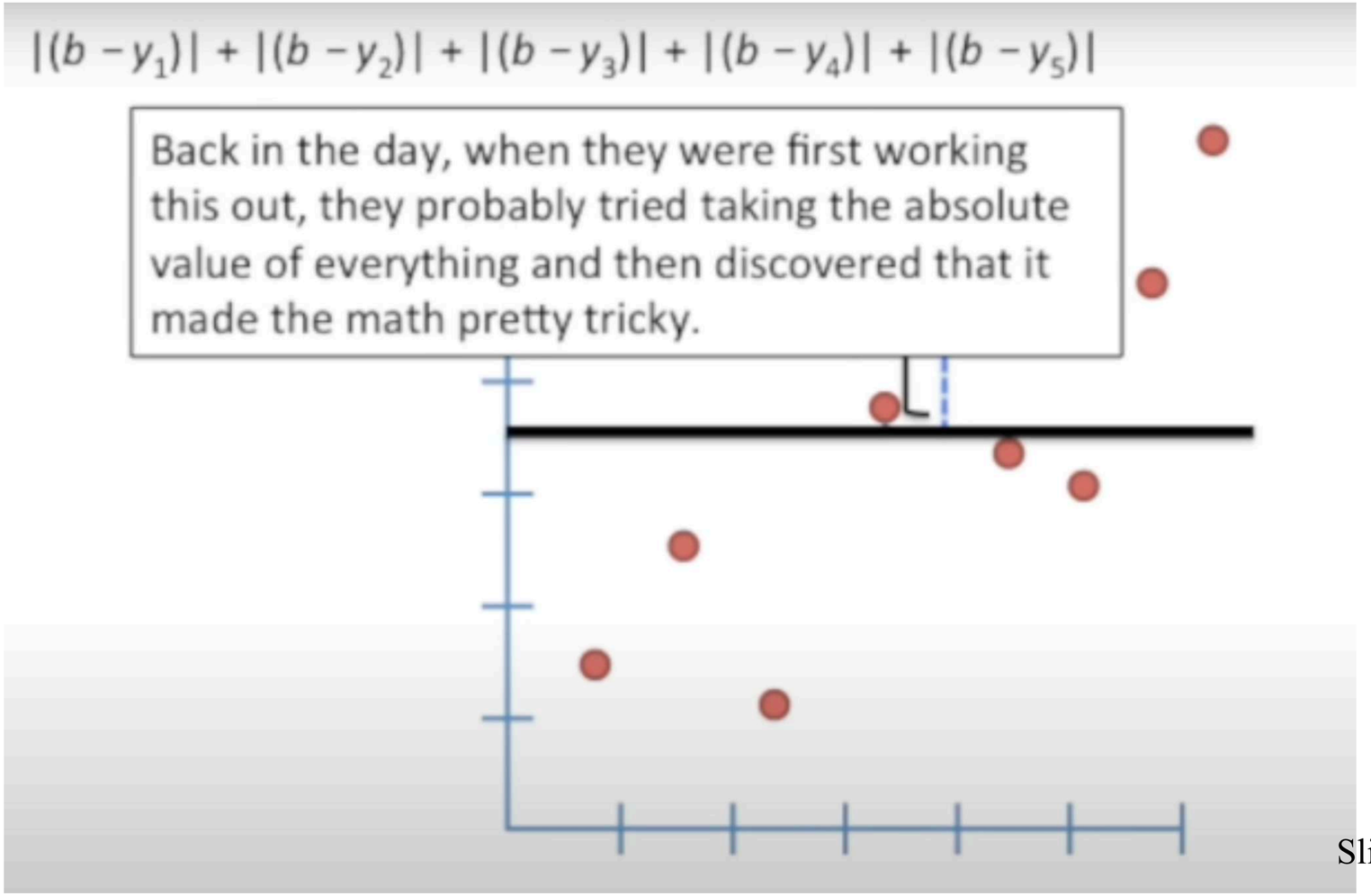
Slide Courtesy: StatQuest

Linear Regression



Slide Courtesy: StatQuest

Linear Regression

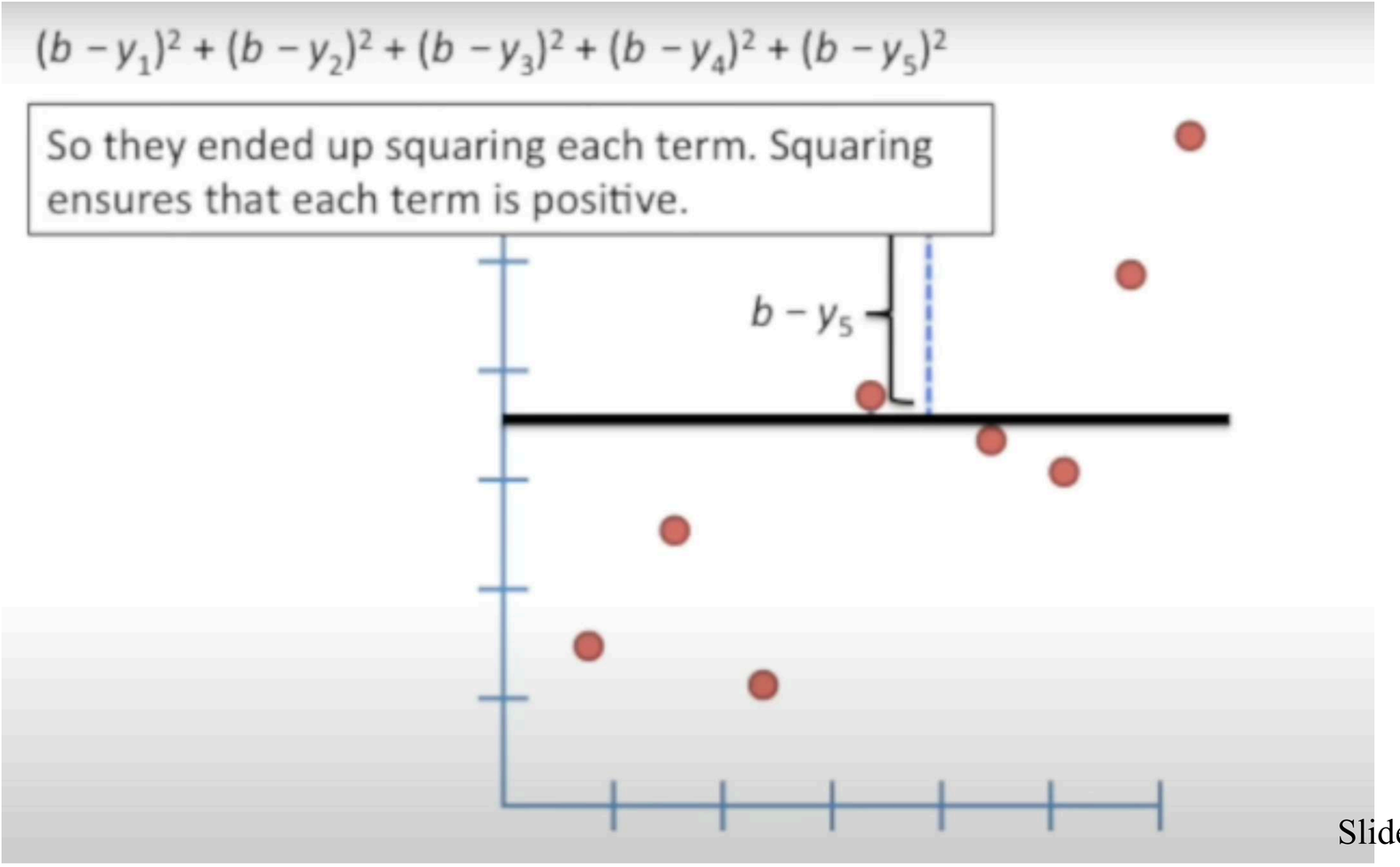


Slide Courtesy: StatQuest

Linear Regression

$$(b - y_1)^2 + (b - y_2)^2 + (b - y_3)^2 + (b - y_4)^2 + (b - y_5)^2$$

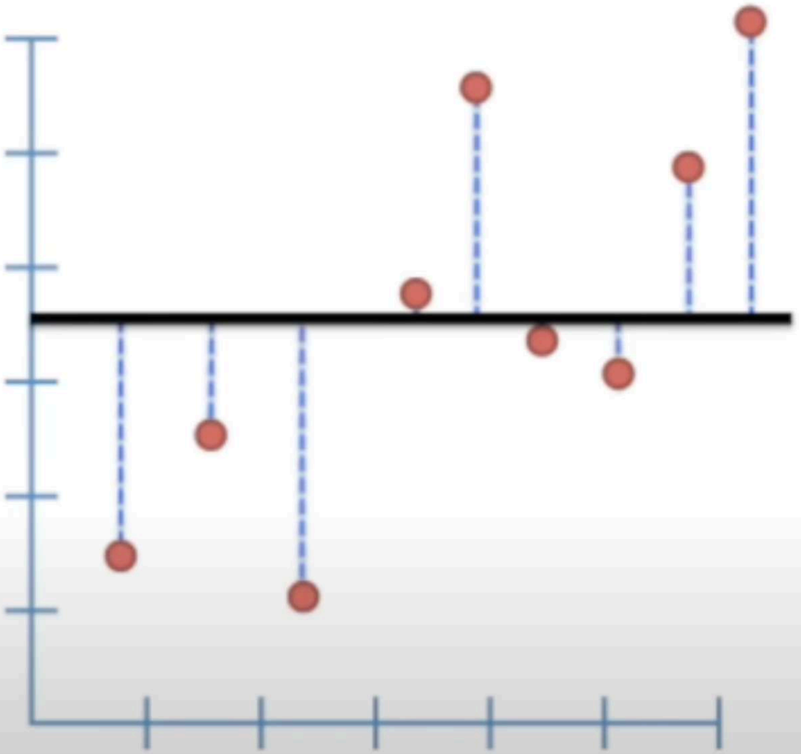
So they ended up squaring each term. Squaring ensures that each term is positive.



Slide Courtesy: StatQuest

Linear Regression

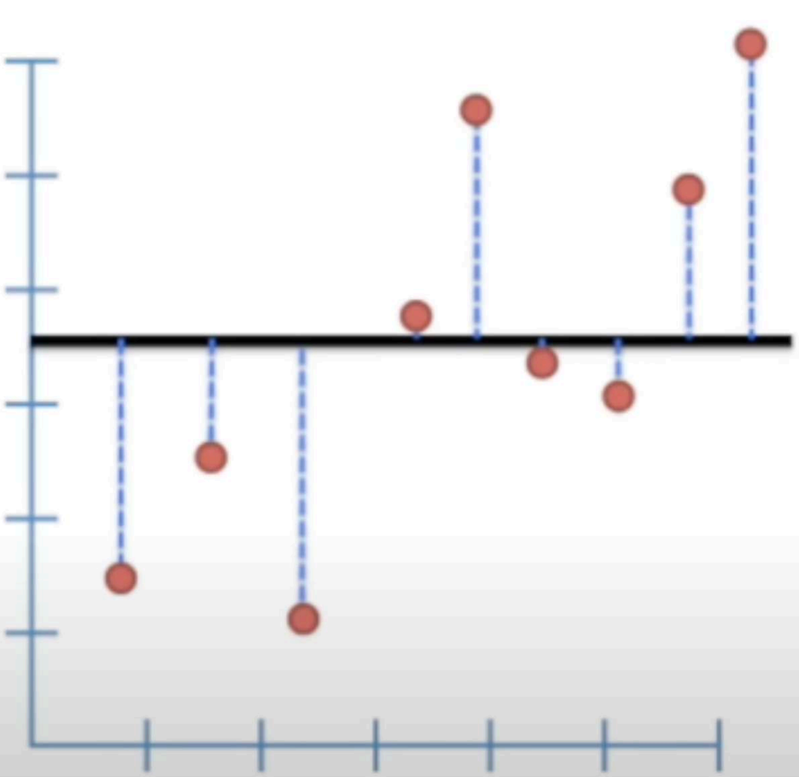
$$(b - y_1)^2 + (b - y_2)^2 + (b - y_3)^2 + (b - y_4)^2 + (b - y_5)^2 + (b - y_6)^2 + (b - y_7)^2 + (b - y_8)^2 + (b - y_9)^2$$



Slide Courtesy: StatQuest

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.

Slide Courtesy: StatQuest