

Week-04-L-03

Select a Statistical Model

Interpretation using Regression Analysis

Prof. J. Ramkumar
Department of ME & Design
Indian Institute of Technology Kanpur



Regression Analysis

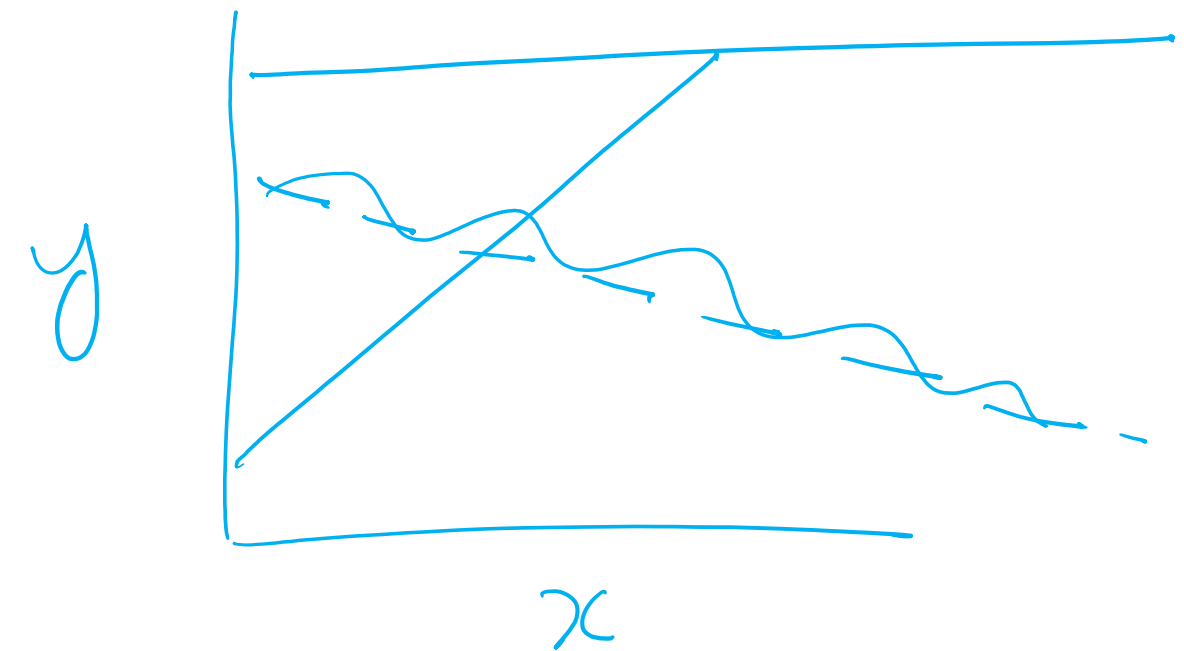


Regression analysis is a reliable method of identifying which variables have impact on a topic of interest.

The process of performing a regression allows you to confidently determine:

- which factors matter most (significant)
- which factors can be ignored (insignificant)
- how these factors influence each other

25°C to 35°C



Important Terms



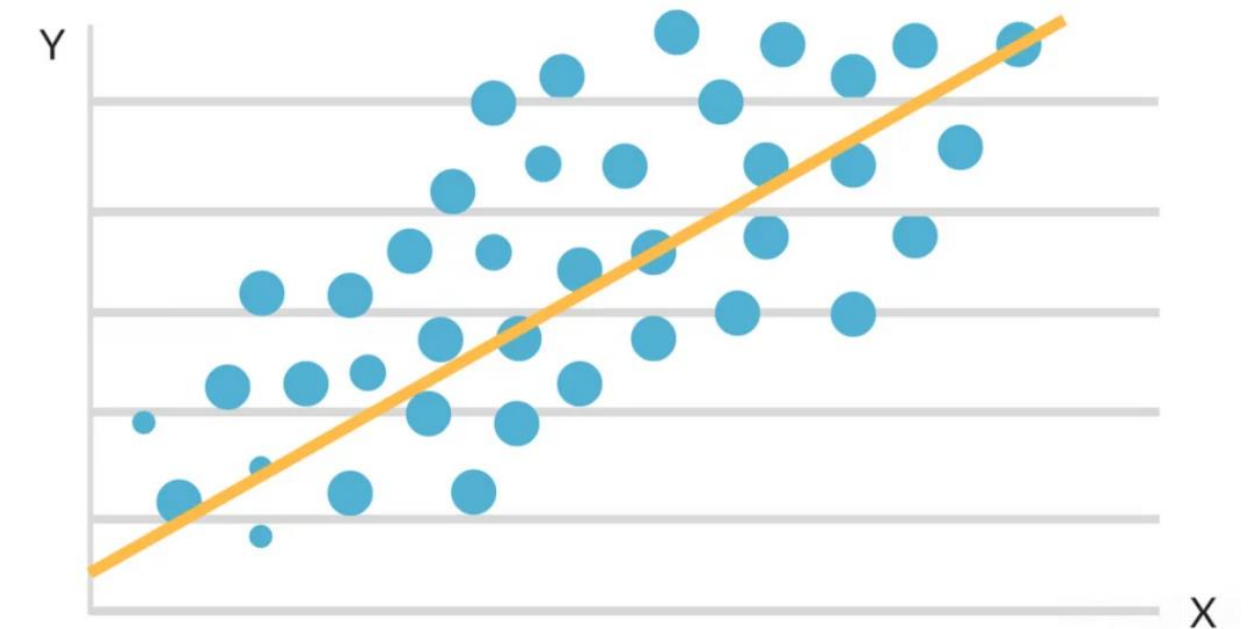
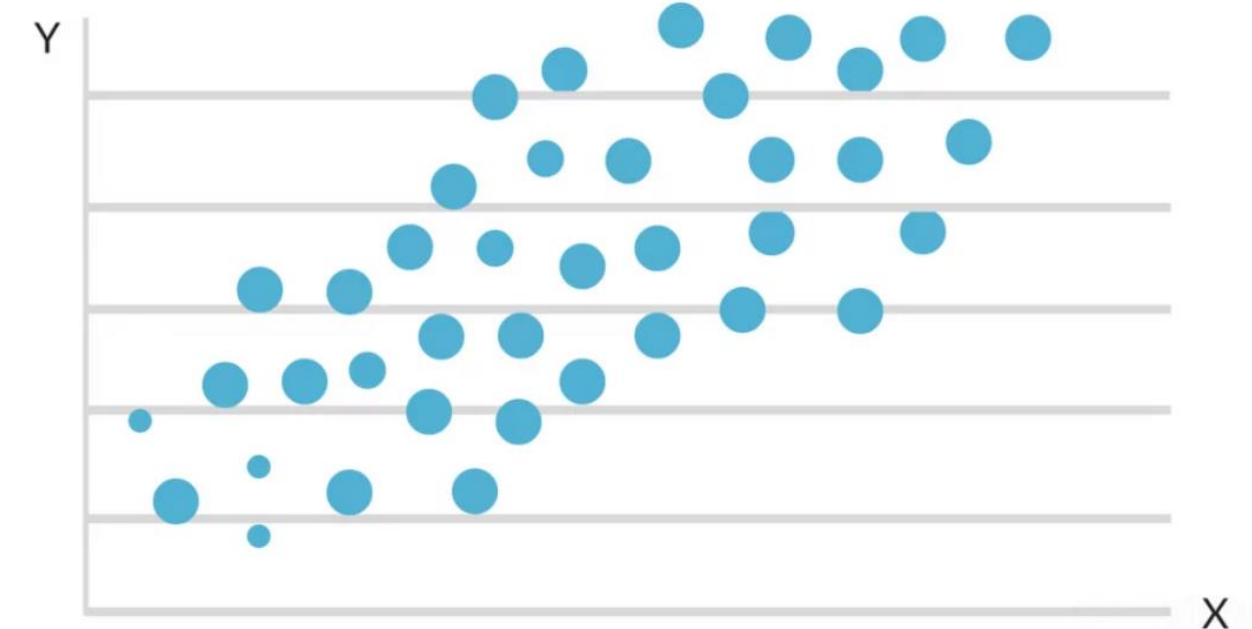
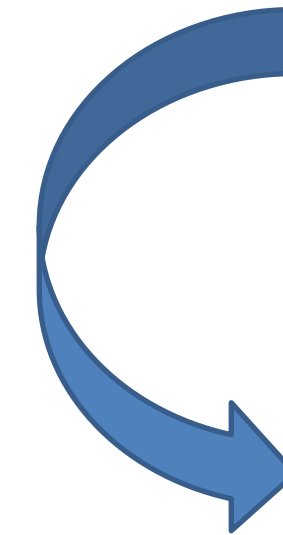
In order to understand regression analysis fully, it's essential to comprehend the following terms:

Dependent Variable: —

- This is the main factor that you're trying to understand or predict.

Independent Variables: —

- These are the factors that you hypothesize have an impact on your dependent variable.



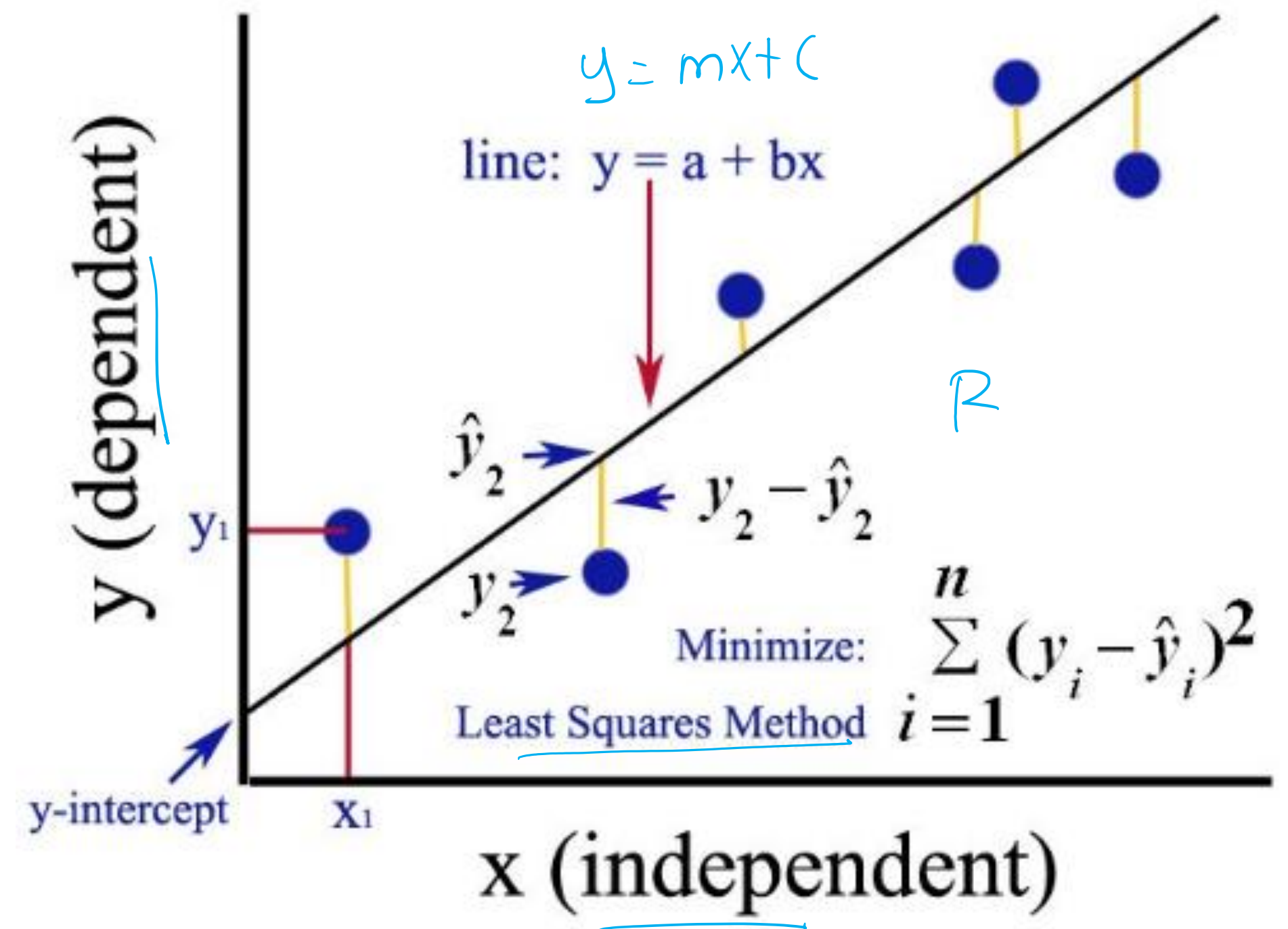
Least squares method

- best fit line

Definition:

The least squares method is a mathematical technique that allows the analyst to determine the best way of fitting a curve on top of a chart of data points.

100%
x y z



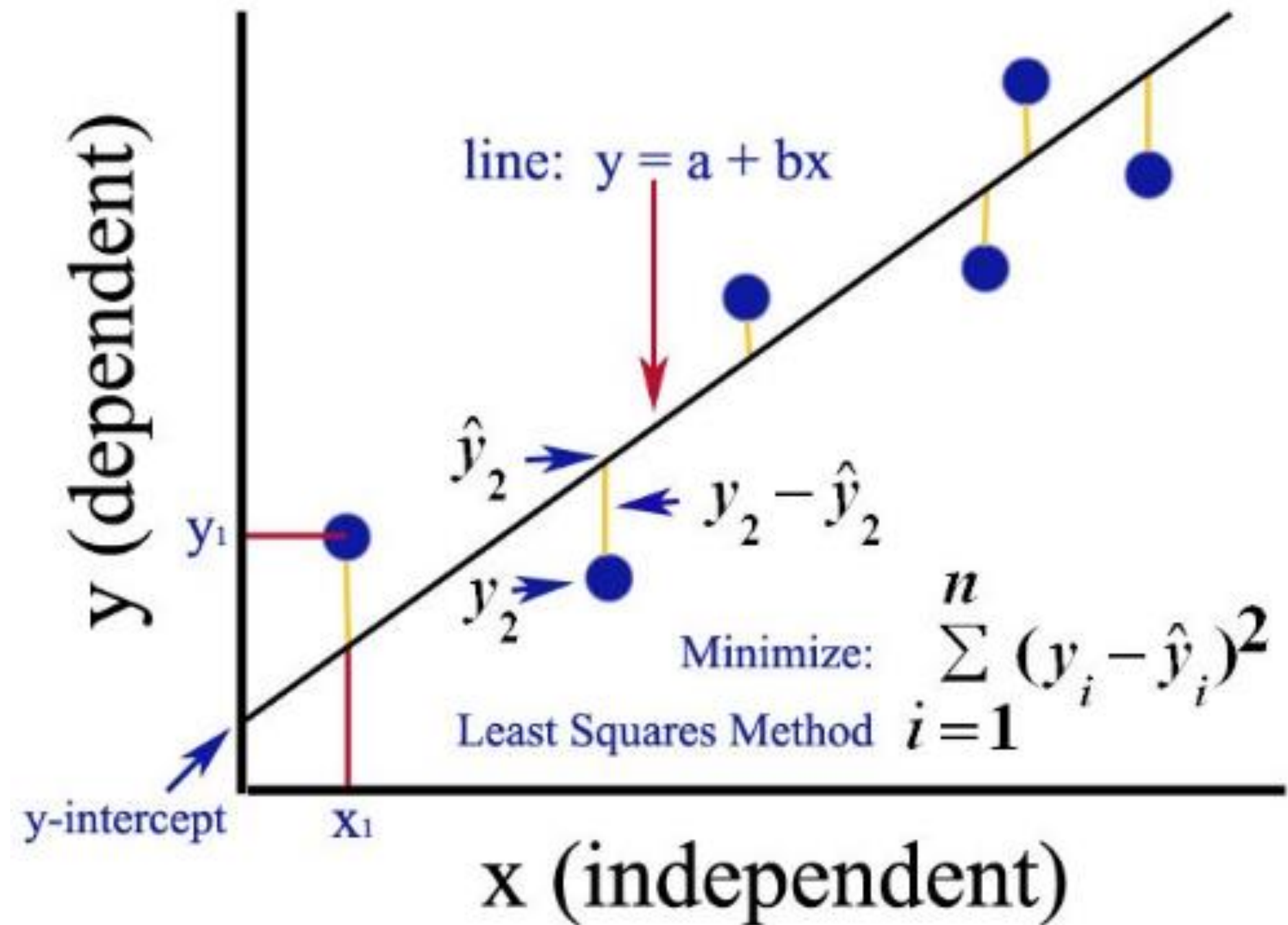
Least squares method

- best fit line

Definition:

Given any collection of pairs of numbers (except when all the x-values are the same) and the corresponding scatter diagram, there always exists exactly one straight line that fits the data better than any other, in the sense of minimizing the sum of the squared errors. |

It is called the least squares regression line. Moreover there are formulas for its slope and y-intercept.



Least squares method

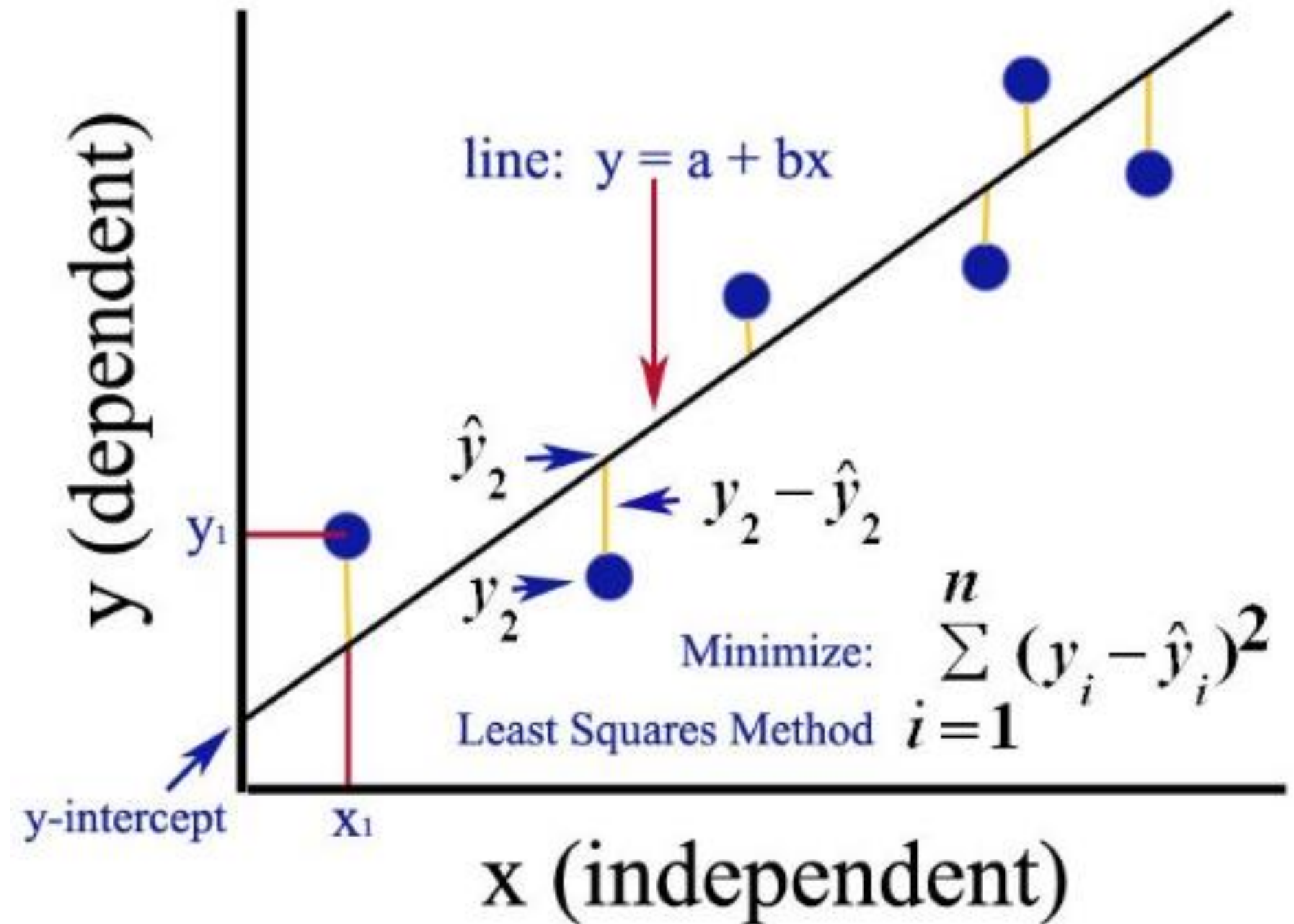
- best fit line

Regression model generated while finding the 'best fit line'.

This just means that we're using the smallest sum of squared errors.

The error is the difference between the predicted y value subtracted from the actual y value.

The difference is squared so there is an absolute difference, and summed.



Least squares method

- best fit equation



The line of best fit determined from the least squares method has an equation that tells the story of the relationship between the data points.

Line of best fit equations may be determined by computer software models:

- which include a summary of outputs for analysis.
- where the coefficients and summary outputs explain the dependence of the variables being tested.

equation of a line

$$y = mx + b$$

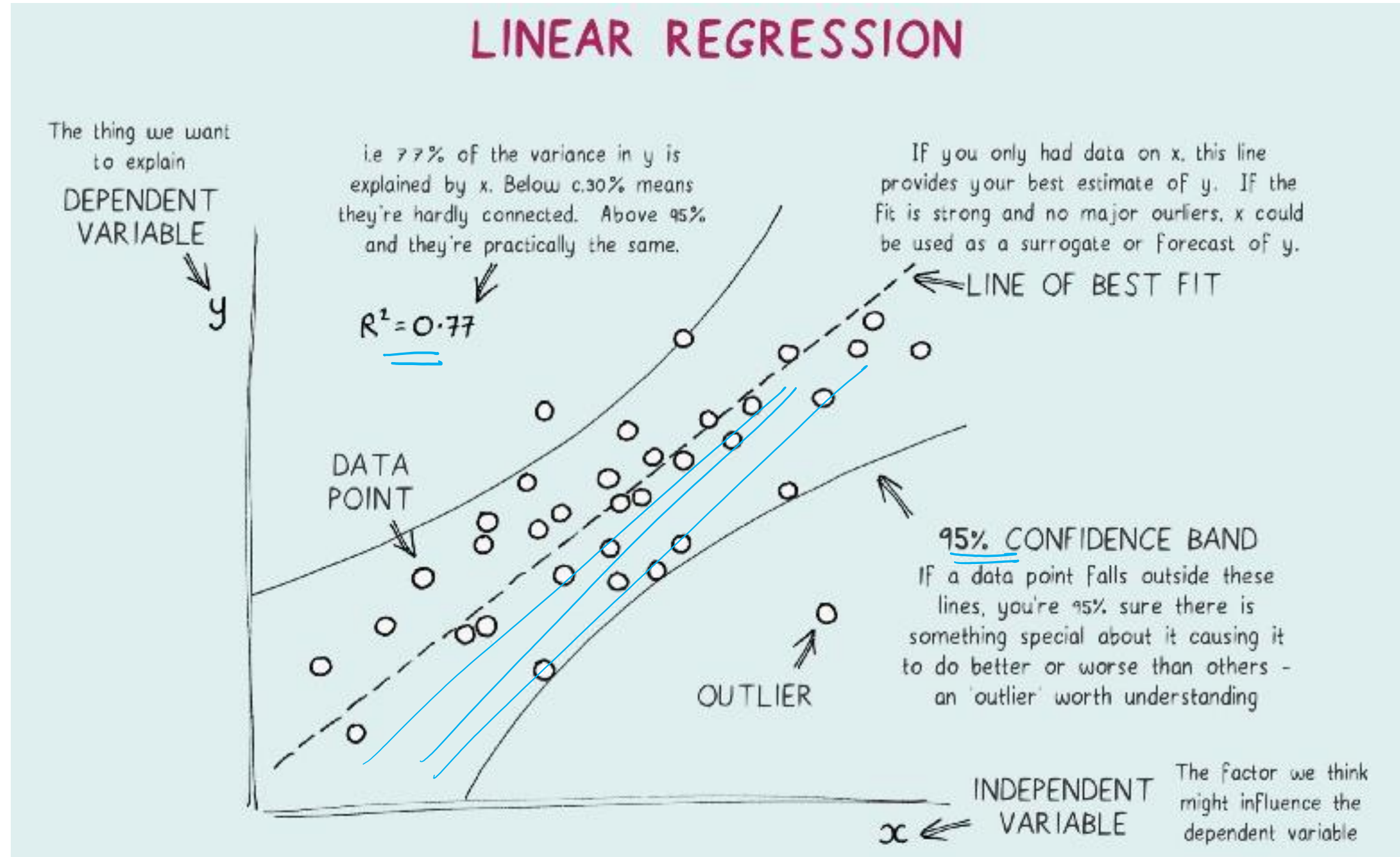
(m = slope b = y intercept)

$$\text{slope} = \frac{\text{change in } y}{\text{change in } x}$$

Source:

www.dept.harpercollege.edu

Regression Analysis



$R^2 = 1$
 $R^2 = 0.95$
 $R^2 = 0.5$

Thank you

