

Problem.

Find the linear regression equation for the data given below.

x	y
1	3
2	5
5	5
7	8
10	9
11	12

Solution.

Let's solve this problem in few steps.

The number of points equals to $N = 6$.

Let's fill such a table:

x	y	xy	x^2
1	3	3	1
2	5	10	4
5	5	25	25
7	8	56	49
10	9	90	100
11	12	132	121

Sum of each column:

$$\sum_{i=1}^N x_i = 36$$

$$\sum_{i=1}^N y_i = 42$$

$$\sum_{i=1}^N x_i y_i = 316$$

$$\sum_{i=1}^N x_i^2 = 300$$

Slope of the line is calculated using the formula:

$$Slope = \frac{N \sum_{i=1}^N x_i y_i - (\sum_{i=1}^N x_i)(\sum_{i=1}^N y_i)}{N \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2} = \frac{6 \cdot 316 - 36 \cdot 42}{6 \cdot 300 - 36^2} = \frac{16}{21} \approx 0.762$$

Interception term is determined using such expression:

$$Intercept = \frac{\sum_{i=1}^N y_i - Slope \cdot (\sum_{i=1}^N x_i)}{N} = \frac{42 - \frac{16}{21} \cdot 36}{6} = \frac{17}{7} \approx 2.429$$

So the regression line equation is

$$y = \frac{16}{21}x + \frac{17}{7} \approx 0.762x + 2.429$$

Scatter plot with the regression line:

