

Linear Regression

1. Use the following dataset to create a linear regression model, scatter the dataset before.

X values represents the revenue of a company in millions of Euros, and Y represents the percentage of investment by this company in Research and Development. We are interested here to find if there is a correlation between these two variables and predict the percentage of investment in R&D knowing the revenue of these companies.

X	Y
60	3.1
61	3.6
62	3.8
63	4
65	4.1

To simplify, you can use $y=ax+b$

$$a = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2}$$

$$b = \frac{\sum Y - b \sum X}{N}$$

Compute a and b (learning process).

2. Compute the MSE error (Mean Square Error)
3. Predict the value of Y then X = 64 (millions of euros)