

Linear Regression Tutorial

1 Introduction

Often in the physical sciences one is asked to fit the best line to a set of experimental data points. This is called linear regression. Most plotting programs include this linear regression capability, but often it is more beneficial to plot the experimental data on graph paper for the determination of unknown quantities. For that reason, it is useful to have the equations available that lets one mathematically fit the best line to a set of experimental observations.

2 Background

The equations for linear regression can be derived from the Calculus of minimization. We will skip this derivation and proceed straight to the equations. Remember that the equation for a line is $y = mx + b$. The coefficients m and b can be solved for explicitly by solving a set of simultaneous equations.

$$\begin{aligned}\sum y &= bN + m \sum x \\ \sum xy &= b \sum x + m \sum x^2\end{aligned}\tag{1}$$

In these equations the summations would be constructed from experimental data. The value of N is the number of (x, y) data pairs. The only things left after the summations are completed are m and b . These can be computed by algebraically solving the simultaneous equations to obtain explicit values for the slope (m) and y-intercept (b).

$$m = \frac{N \sum xy - \sum x \sum y}{N \sum x^2 - (\sum x)^2}\tag{2}$$

$$b = \frac{\sum x^2 \sum y - \sum x \sum xy}{N \sum x^2 - (\sum x)^2}\tag{3}$$

3 Example

Assume one has the following data from the Ni²⁺ spectrophotometry experiment.

Solution #	Concentration	% T	Absorbance
1	0.1526	75.0	0.1250
2	0.0610	83.0	0.0809
3	0.0305	89.5	0.0482
4	0.0153	95.3	0.0209
5	0.0061	97.7	0.0101

On the graph absorbance (Y-Axis) is plotted vs. concentration (X-axis). For that reason, the absorbance values will be associated with the y values in the summation equations and the concentrations will be associated with the x values in the summations. The table of summations should look like:

Symmmation	Value
N	5
$\sum x$	0.2655
$\sum y$	0.2851
$\sum x^2$	0.02821
$\sum xy$	0.02586

Plugging these values into the simultaneous equations we get...

$$(0.2851) = b(5) + m(0.2655) \quad (4)$$

$$(0.02586) = b(0.2655) + m(0.02821) \quad (5)$$

Using the exact expressions given in equations 2 and 3, the values of m and b can be found.

$$m = 0.7599 \quad (6)$$

$$b = 0.01667 \quad (7)$$

Or, in the equation of a line,

$$y = 0.7599x + 0.01167 \quad (8)$$