



Continuing Education for technicians, engineers, and scientists with emphasis on measurement, test, and metrology

Login As: Student Go

[About Us](#) | [Courses](#) | [Programs](#) | [My Courses](#) | [Contact](#) | [Resources](#) | [Articles](#) | [Forums](#) | [Site Map](#) | [Login](#) | [Sign Up!](#)

Quick Links

- Site Help
- The Blog
- The Podcast
- Pricing

Student Registration

Students

- Register
- My Courses
- My Account
- Edit Resume
- Training Record
- Bid on Tasks
- Calculators
- Suggest Course

Employer Registration

Employers

- Search Resumes
- Request Bids
- Assign Courses
- Employee Records
- Reports
- Create Courses
- Suggest Course
- Create Groups

Educator Registration

Educators

- Student Records
- Create Groups
- Assign Courses

Consultants

- Edit Profile
- Bid on Tasks

Least-Squares Line Fits and Associated Uncertainty By David Archer

Introduction

There are several measurement situation where one is trying to determine if there is a linear relationship between a pair of measured values. For instance the relationship between stress and strain, voltage and current, input voltage and output voltage, etc.. In this article the basics of least-squares line fits will be discussed, along with a basic uncertainty analysis.

The General Least-Squares Problem

In the general least-squares problem, one has a set of measured data collected as ordered pairs

$$(x_i, y_i) \text{ where } i = 1..N$$

and one wants to fit this data to the functional form

$$y = f(c_1, \dots, c_M; x)$$

where the c_i are parameters M parameters that define the function.

One can define the residuals of the data set from

$$r_i = y_i - f(c_1, \dots, c_M; x_i)$$

which measures how far each data point deviates from the function along a vertical line. A measure of the goodness of the fit is the root-mean-square (RMS) value of the residuals

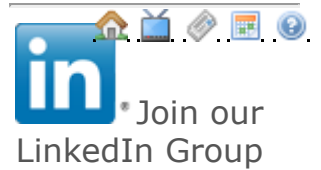
$$r_{RMS} = \frac{1}{N} \sqrt{\sum_{i=1}^N r_i^2}$$

The general least-squares problem is to find the constants c_i that minimizes the RMS value of the residuals. From basic calculus, one knows one has a minimum if

$$\frac{\partial r_{RMS}}{\partial c_i} = 0$$

One generates a series of M such equations for the M unknowns in the functional form. Solving this system of equations results in the least-squares fit for the particular functional form.

The Least-Squares Line Fit Problem



Twitter Updates

follow
LearningMeasure.com on
Twitter



[About Us](#) | [Courses](#) | [Programs](#) | [My Courses](#) | [Contact](#) | [Resources](#) | [Articles](#) | [Forums](#) | [Site Map](#) | [Login](#) | [Sign Up!](#)

$$r_{RMS} = \frac{1}{N} \sqrt{\sum_{i=1}^M (y_i - mx_i - b)^2}$$



Taking the derivative of this expression with respect to b and equating it to zero results in

$$\frac{\partial r_{RMS}}{\partial b} = \frac{1}{N} \frac{\partial}{\partial b} \sqrt{\sum_{i=1}^N (y_i - mx_i - b)^2} = -\frac{1}{N} \frac{\sum_{i=1}^N (y_i - mx_i - b)}{\sqrt{\sum_{i=1}^N (y_i - mx_i - b)^2}} = 0$$

With a little manipulation this reduces to

$$m \left(\frac{1}{N} \sum_{i=1}^N x_i \right) + b = \left(\frac{1}{N} \sum_{i=1}^N y_i \right)$$

Performing the same operation for the parameter m,

$$\frac{\partial r_{RMS}}{\partial m} = \frac{1}{N} \frac{\partial}{\partial m} \sqrt{\sum_{i=1}^N (y_i - mx_i - b)^2} = -\frac{1}{N} \frac{\sum_{i=1}^N (y_i - mx_i - b)x_i}{\sqrt{\sum_{i=1}^N (y_i - mx_i - b)^2}} = 0$$

which can be manipulated into the form

$$m \left(\frac{1}{N} \sum_{i=1}^N x_i^2 \right) + b \left(\frac{1}{N} \sum_{i=1}^N x_i \right) = \frac{1}{N} \left(\sum_{i=1}^N y_i x_i \right)$$

To simplify these equations, the following quantities are introduced

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i = \text{the mean value of the } x_i$$

$$\langle x^2 \rangle = \frac{1}{N} \sum_{i=1}^N x_i^2 = \text{the mean square value of the } x_i$$

$$\langle y \rangle = \frac{1}{N} \sum_{i=1}^N y_i = \text{the mean value of the } y_i$$

$$\langle xy \rangle = \frac{1}{N} \sum_{i=1}^N x_i y_i = \text{the mean value of } x_i y_i$$

The equations that determine the minimum RMS residual can now be recast in the form

$$m \langle x^2 \rangle + b \langle x \rangle = \langle xy \rangle$$

$$m \langle x \rangle + b = \langle y \rangle$$

or written as a matrix equation



Continuing Education for technicians, engineers,
and scientists with emphasis on measurement,
test, and metrology

Login As:

[About Us](#) | [Courses](#) | [Programs](#) | [My Courses](#) | [Contact](#) | [Resources](#) | [Articles](#) | [Forums](#) | [Site Map](#) | [Login](#) | [Sign Up!](#)

$$m = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2}$$



$$b = \langle y \rangle - m \langle x \rangle$$

which solves the least squares problem.

Uncertainty Analysis

Assuming for the moment that the slope or intercept were the quantities that one was actually trying to measure, one should perform an uncertainty analysis on these parameters by using the "Law of propagation of uncertainty" to determine the uncertainty in the slope and intercept in terms of the uncertainty in the measurement data.

The uncertainty in the slope δm can be written

$$\delta m = \sqrt{\sum_{i=1}^N \left(\frac{\partial m}{\partial x_i} \delta x_i \right)^2 + \sum_{i=1}^N \left(\frac{\partial m}{\partial y_i} \delta y_i \right)^2}$$

where δx_i is the uncertainty associated with x_i , and δy_i is the uncertainty associated with y_i .

It can be shown that the partial derivatives are

$$\frac{\partial m}{\partial x_i} = \frac{1}{N} \frac{\left(\langle x^2 \rangle - \langle x \rangle^2 \right) (y_i - \langle y \rangle) - (\langle xy \rangle - \langle x \rangle \langle y \rangle) (2x_i - 2\langle x \rangle)}{\left(\langle x^2 \rangle - \langle x \rangle^2 \right)^2}$$

and

$$\frac{\partial m}{\partial y_i} = \frac{1}{N} \frac{x_i - \langle x \rangle}{\langle x^2 \rangle - \langle x \rangle^2}$$

Similarly for the uncertainty in the intercept δb is

$$\delta b = \sqrt{\sum_{i=1}^N \left(\frac{\partial b}{\partial x_i} \delta x_i \right)^2 + \sum_{i=1}^N \left(\frac{\partial b}{\partial y_i} \delta y_i \right)^2}$$

and the partial derivatives are

$$\frac{\partial b}{\partial x_i} = -\frac{\partial m}{\partial x_i} \langle x \rangle - \frac{mx_i}{N}$$

and

$$\frac{\partial b}{\partial y_i} = \frac{y_i}{N} - \frac{\partial m}{\partial y_i} \langle x \rangle$$



the uncertainty in the value y is then

$$\delta y = \sqrt{(m\delta x)^2 + (x\delta m)^2 + \delta b^2}$$

Now what has been calculated to this point has been uncertainty associated with the fit itself. It assumes that the underlying phenomenon is linear. This may not be the case. However, one does have a quantity that is a direct measure of the "goodness" of the linearity assumption, namely the residuals.

There will be non-zero residuals for a given fit, even if the underlying phenomenon is perfectly linear, due to the uncertainty in the measurement. That however is predictable. The estimated residuals due to the measurement uncertainty is

$$r_{\text{RMS-predicted}} = \frac{1}{N} \sqrt{\sum_{i=1}^N (m\delta x_i)^2 + \sum_{i=1}^N (x_i\delta m)^2 + N\delta b^2}$$

This can be compared against the actual residuals. If the actual residuals are significantly larger than the predicted residuals, one can say there is some non-linearity in the underlying phenomenon. In such a case one would probably want to combine the RMS residuals with the uncertainty in y for an overall uncertainty statement.

Conclusion

The basics of least-squares line fits was presented, along with a basic uncertainty analysis. Hopefully this article can be useful as a reference if your measurement requires some sort of least-squares line fit. There are many phenomenon, and situations in calibration and measurement, where such a fit is useful.