THE METHOD OF LEAST SQUARES **Chemistry BC3252y**

According to the principle of least squares, the best curve of a given form through a set of **n** experimental points, assuming a normal or Gaussian distribution of random (indeterminate) errors, is the curve for which the sum of the squares of the residuals is a minimum.

Suppose we have a set of n data points (x_i, y_i) where i = l, …n. Assume that the values of x are either exact or have negligible error. In an *unweighted* least squares, the values of y have the same estimated standard deviation, with actual (unknown) errors distributed randomly. In a *weighted* least squares, we have information (w_i) about the relative statistical weight that should be assigned to each point in the fitting. For example, if each point has standard deviation σ_i, then the assigned weight should be $w_i = 1/\sigma_i^2$. More certain points thus have a greater weight in the fitting. An unweighted least squares is mathematically equivalent to a weighted least squares with all points having equal weight.

We confine our attention to the simple case in which the relation between the values of x and the values of y is linear (polynomial of order one):

$$\mathbf{y = mx + b} \qquad (1)$$

The residuals Δ_i are given by

$$\Delta_i = (y_i)_{experimental} - (y_i)_{calculated} = y_i - (mx_i + b) \qquad (2)$$

For a set of n values of (x_i, y_i, w_i) we seek the linear parameters slope, **m,** and y-intercept, **b,** that minimize the sum S defined as follows:

$$S = \sum_{i=1}^{n} w_i \Delta_i^2$$

All summations above and hereafter are understood to run from i = l to i = n. After substitution and expansion, this becomes

$$S = m^2 \sum w_i x_i^2 + 2mb \sum w_i x_i - 2m \sum w_i x_i y_i + b^2 \sum w_i - 2b \sum w_i y_i + \sum w_i y_i^2 \qquad (3)$$

The conditions for a minimum in **S** with respect to variations in **m** and **b** are

$$(\partial S/\partial m)_b = 0 = 2m \sum w_i x_i^2 + 2b \sum w_i x_i - 2 \sum w_i x_i y_i \qquad (4)$$

and $$(\partial S/\partial b)_m = 0 = 2m \sum w_i x_i + 2b \sum w_i - 2 \sum w_i y_i \qquad (5)$$

Solving these two simultaneous equations gives the expressions for the best values of m and b, in the least squares sense:

slope: $$m = \frac{\sum w_i \sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i}{\sum w_i \sum w_i x_i^2 - \left(\sum w_i x_i\right)^2} \qquad (6)$$

y-intercept: $$b = \frac{\sum w_i x_i^2 \sum w_i y_i - \sum w_i x_i \sum w_i x_i y_i}{\sum w_i \sum w_i x_i^2 - \left(\sum w_i x_i\right)^2} \qquad (7)$$

For an unweighted least squares (all weights are 1), the equations are:

$$m = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2} \quad \text{and} \quad b = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2}$$

Automatic Least Squares Fitting using Excel

TRENDLINE

Starting with an Excel graph, you can make at least-squares fit to a line or to variety of other functions using the trendline function..

a) Click on your graph. Click on the data so it is highlighted.
b) Under the **Chart** heading, select **Add trendline**. Under **Trend/Regression type → linear**.
c) Under **Chart**, select **Display equation on chart** and **Display R-squared value on chart**. Press ok.
d) Excel makes a very dark heavy trendline: alter it so it is finer.

LINEST( )

Linest is an Excel function which gives the values of all the parameters in a multilinear regression, their standard deviations, the correlation coefficient, and various other things. It is fully explained in Excel help. It had two major advantages over a related Excel function, regression: (1) it is dynamic: when you change the variables the results change automatically, and (2) you can pre-select just the parameters you want to see.

To do a linear fit (y = mx + b), with the results nicely organized and labeled:

1) Create the following 4×3 block in an empty portion of your spreadsheet
(it prepares a labeled area for the results; the last row is optional):

LINEST()	m	b
value		
SD		
%RSD		

2) Perform the linest() function:

a) Highlight the 2×2 array of empty cells in the second and third rows and second and third columns of this block.
b) Click on the f_x button on the standard toolbar
c) Select the function LINEST(X)
d) Enter the locations of the cells for the y and x data
e) Set the value of Stats to 1
f) *Do NOT press ok*, instead press Crtl+Shift+Enter (all three at once!)

3) (Optional) Add the equations to calculate % RSD (relative standard deviation) values.