Quantitative Chemical Analysis

Seventh Edition

Daniel C. Harris

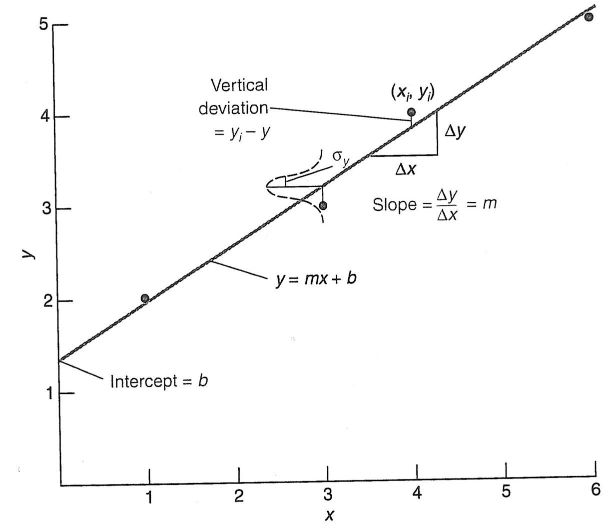


Figure 4-9 Least-squares curve fitting. The points (1,2) and (6,5) do not fall exactly on the solid line, but they are too close to the line to show their deviations. The Gaussian curve drawn over the point (3,3) is a schematic indication of the fact that each value of y_i is normally distributed about the straight line. That is, the most probable value of y will fall on the line, but there is a finite probability of measuring y some distance from the line.

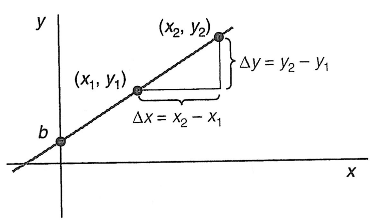
estimate the uncertainty in a chemical analysis from the uncertainties in the calibration curve and in the measured response to replicate samples of unknown.

Finding the Equation of the Line

The procedure we use assumes that the errors in the y values are substantially greater than the errors in the x values.[7] This condition is usually true in a calibration curve in which the experimental response (y values) is less certain than the quantity of analyte (x values). A second assumption is that uncertainties (standard deviations) in all the y values are similar.

Suppose we seek to draw the best straight line through the points in Figure 4-9 by minimizing the vertical deviations between the points and the line. We minimize only the vertical deviations because we assume that uncertainties in y values are much greater than uncertainties in x values.

Let the equation of the line be



Equation for a straight line: $y = mx + b$

Slope $(m) = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$

y-Intercept (b) = crossing point on y-axis

Equation of straight line: $$y = mx + b \qquad (4\text{-}14)$$

in which m is the **slope** and b is the **y-intercept.** The vertical deviation for the point (x_i, y_i) in Figure 4-9 is $y_i - y$, where y is the ordinate of the straight line when $x = x_i$.

$$\text{Vertical deviation} = d_i = y_i - y = y_i - (mx_i + b) \qquad (4\text{-}15)$$

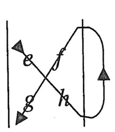
Some of the deviations are positive and some are negative. Because we wish to minimize the magnitude of the deviations irrespective of their signs, we square all the deviations so that we are dealing only with positive numbers:

$$d_i^2 = (y_i - y)^2 = (y_i - mx_i - b)^2$$

Because we minimize the squares of the deviations, this is called the *method of least squares.* It can be shown that minimizing the squares of the deviations (rather than simply their magnitudes) corresponds to assuming that the set of y values is the most probable set.

Finding values of m and b that minimize the sum of the squares of the vertical deviations involves some calculus, which we omit. We will express the final solution for slope and intercept in terms of *determinants,* which summarize certain arithmetic operations. The **determinant** $\begin{vmatrix} e & f \\ g & h \end{vmatrix}$ represents the value $eh - fg$. So, for example,

To evaluate the determinant, multiply the diagonal elements $e \times h$ and then subtract the product of the other diagonal elements $f \times g$.



$$\begin{vmatrix} 6 & 5 \\ 4 & 3 \end{vmatrix} = (6 \times 3) - (5 \times 4) = -2$$

Table 4-6 Calculations for least-squares analysis

x_i	y_i	x_iy_i	x_i^2	$d_i(= y_i - mx_i - b)$	d_i^2
1	2	2	1	0.038 46	0.001 479 3
3	3	9	9	−0.192 31	0.036 982
4	4	16	16	0.192 31	0.036 982
6	5	30	36	−0.038 46	0.001 479 3
$\Sigma x_i = 14$	$\Sigma y_i = 14$	$\Sigma(x_iy_i) = 57$	$\Sigma(x_i^2) = 62$		$\Sigma(d_i^2) = 0.076\ 923$

The slope and the intercept of the "best" straight line are found to be

Least-squares "best" line

$$\text{Slope:} \quad m = \begin{vmatrix} \Sigma(x_iy_i) & \Sigma x_i \\ \Sigma y_i & n \end{vmatrix} \div D \qquad (4\text{-}16)$$

$$\text{Intercept:} \quad b = \begin{vmatrix} \Sigma(x_i^2) & \Sigma(x_iy_i) \\ \Sigma x_i & \Sigma y_i \end{vmatrix} \div D \qquad (4\text{-}17)$$

where D is

$$D = \begin{vmatrix} \Sigma(x_i^2) & \Sigma x_i \\ \Sigma x_i & n \end{vmatrix} \qquad (4\text{-}18)$$

and n is the number of points.

Translation of least-squares equations:

$$m = \frac{n\Sigma(x_iy_i) - \Sigma x_i\Sigma y_i}{n\Sigma(x_i^2) - (\Sigma x_i)^2}$$

$$b = \frac{\Sigma(x_i^2)\Sigma y_i - (\Sigma x_iy_i)\Sigma x_i}{n\Sigma(x_i^2) - (\Sigma x_i)^2}$$

Let's use these equations to find the slope and intercept of the best straight line through the four points in Figure 4-9. The work is set out in Table 4-6. Noting that $n = 4$ and putting the various sums into the determinants in Equations 4-16, 4-17, and 4-18 gives

$$m = \begin{vmatrix} 57 & 14 \\ 14 & 4 \end{vmatrix} \div \begin{vmatrix} 62 & 14 \\ 14 & 4 \end{vmatrix} = \frac{(57 \times 4) - (14 \times 14)}{(62 \times 4) - (14 \times 14)} = \frac{32}{52} = 0.615\ 38$$

$$b = \begin{vmatrix} 62 & 57 \\ 14 & 14 \end{vmatrix} \div \begin{vmatrix} 62 & 14 \\ 14 & 4 \end{vmatrix} = \frac{(62 \times 14) - (57 \times 14)}{(62 \times 4) - (14 \times 14)} = \frac{70}{52} = 1.346\ 15$$

The equation of the best straight line through the points in Figure 4-9 is therefore

$$y = 0.615\ 38x + 1.346\ 15$$

We tackle the question of significant figures for m and b in the next section.

Example Finding Slope and Intercept with a Spreadsheet

Your scientific calculator has a procedure for computing the slope and intercept of a set of (x,y) data, and you should learn how to use that procedure. Alternatively, Excel has functions called SLOPE and INTERCEPT whose use is illustrated here:

	A	B	C	D	E	F
1	x	y			Formulas:	
2	1	2		slope =		
3	3	3		0.61538	D3 = SLOPE(B2:B5,A2:A5)	
4	4	4		intercept =		
5	6	5		1.34615	D5 = INTERCEPT(B2:B5,A2:A5)	

The slope in cell D3 is computed with the formula "= SLOPE(B2:B5,A2:A5)", where B2:B5 is the range containing the y values and A2:A5 is the range containing x values.

How Reliable Are Least-Squares Parameters?

To estimate the uncertainties (expressed as standard deviations) in the slope and intercept, an uncertainty analysis must be performed on Equations 4-16 and 4-17. Because the uncertainties in m and b are related to the uncertainty in measuring each value of y, we first estimate the standard deviation that describes the population of y values. This standard deviation, σ_y, characterizes the little Gaussian curve inscribed in Figure 4-9.

We estimate σ_y, the population standard deviation of all y values, by calculating s_y, the standard deviation, for the four measured values of y. The deviation of each value of y_i from