Following is a simple derivation of the least squares fit.

Suppose the relationship between the two experimental parameters being studied is

$$y = f(x)$$

where x is the independent parameter which is varied, and y is the dependent parameter. If $f(x)$ is a polynomial function, or can be approximated by a polynomial, then the least squares method is a *linear* one, and it will almost always give reliable answers. If $f(x)$ cannot be expressed as a polynomial, but consists of transcendental functions, the least squares method is non-linear, and may or may not work reliably. In some cases, a change of variables may result in a polynomial, as in the exponential example above. A function like

$$y = a + \frac{b}{x} + \frac{c}{x^2}$$

is not a polynomial in x, but it is a polynomial in the variable $z = 1/x$.

Suppose the functional relationship between x and y is a polynomial of degree ℓ:

$$y = a_0 + a_1 x + a_2 x^2 \ldots a_\ell x^\ell \tag{1}$$

or

$$y = \sum_{j=0}^{\ell} a_j x^j \tag{2}$$

and we have a set of N data points x_i, y_i obtained by experiment. The goal is to find the values of the $\ell + 1$ parameters $a_0, a_1 \ldots a_\ell$ which will give the best fit of Equation 1 to our data points. The first piece of information to note is that

$$N \geq \ell + 1 \tag{3}$$

or else we will not be able to make a unique determination. For example, if $\ell = 1$, we need at least two data points to find the equation of the straight line. In order to make any meaningful statistical statements, however, we will need even more than $\ell + 1$ points, as we shall see later. A good rule of thumb: if we wish to fit our data with a polynomial of degree ℓ in a 95% confidence interval, we should choose N such that

$$N - (\ell + 1) \geq 10 \tag{4}$$

The idea behind the linear least squares method is to *minimize* the sum

$$S = \sum_{i=1}^{N} \left(y_i - \sum_{j=0}^{\ell} a_j x_i^j \right)^2 \tag{5}$$

S will be a minimum if

$$\frac{\partial S}{\partial a_k} = 0 \qquad k = 0, 1, 2 \ldots \ell \tag{6}$$

The result will be $\ell + 1$ linear equations in $\ell + 1$ unknowns:

$$\sum_{j=0}^{\ell} a_j \left(\sum_{i=1}^{N} x_i^{j+k} \right) = \sum_{i=1}^{N} x_i^k y_i \qquad k = 0, 1 \ldots \ell \tag{7}$$

which can be solved by standard matrix techniques for the unknown coefficients $a_0, a_1 \ldots a_\ell$. As an example, let us consider the case where $\ell = 1$, or

$$y = mx + b$$

In this case,

$$S = \sum_{i=1}^{N} (y_i - (mx_i + b))^2$$

Expanding Equation 7, we have

$$b(N) + m \left(\sum_{i=1}^{N} x_i \right) = \textstyle\sum_{i=1}^{N} y_i \tag{8}$$

$$b \left(\sum_{i=1}^{N} x_i \right) + m \left(\sum_{i=1}^{N} x_i^2 \right) = \textstyle\sum_{i=1}^{N} x_i y_i \tag{9}$$

Then the intercept b and the slope m can be found from Cramer's rule

$$b = \frac{\left(\sum y_i\right)\left(\sum x_i^2\right) - \left(\sum x_i\right)\left(\sum x_i y_i\right)}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} \tag{10}$$

and

$$m = \frac{N\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} \tag{11}$$