Electronic Supplementary Material (ESI) for Journal of Analytical Atomic Spectrometry.

Supplementary Material

Fundamentals and new approaches to calibration in atomic spectrometry

George L. Donati and Renata S. Amais

Estimation of linear function coefficients by least-squares regression

Eqn (S1) represents the linear relationship between the instrument response for an individual measurement, y_i, and its respective analyte concentration, x_i.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \text{(S1)}$$

where β_0 and β_1, are the y-intercept and slope of the calibration curve, and ε_i represents the difference (also known as residual) between an individual experimental value, y_i, and the expected instrument response, $\hat{y}$, according to the linear regression equation.

In least-squares regression, one seeks to minimize the difference between each experimental point and its corresponding expected value. The residual can be positive or negative, *i.e.* the experimental value can be larger or smaller than the expected one. Therefore, it is more convenient to minimize the square of the residual, ε_i^2. Hence, the "best" regression line will be the one for which the sum of the squares of the residuals, S, is closest to zero (*i.e.* $S = \sum \varepsilon_i^2 \rightarrow 0$).[S1]

The values for β_0 and β_1 are unknown. Even more difficult to determine is ε_i, as it varies with each measurement.[S1,S2] However, estimates of slope (m) and intercept (b) can be obtained following the deduction presented below.

$$\varepsilon_i = y_i - \beta_0 - \beta_1 x_i \qquad \text{(S2)}$$

$$\sum_{i}^{n} \varepsilon_i = \sum_{i}^{n} (y_i - \beta_0 - \beta_1 x_i) \tag{S3}$$

$$S = \sum_{i}^{n} \varepsilon_i^2 = \sum_{i}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{S4}$$

To minimize the differences between each (y_i, $\hat{y}$) pair, we need to minimize the sum of the squares of the residuals (S). If we take the partial derivative of eqn (S4) in function of β_0 and β_1, and make it equal to zero, we can estimate the values of b (estimate of β_0) and m (estimate of β_1) that will lead to the lowest S. For estimating b:

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i}^{n} (y_i - n\beta_0 - \beta_1 x_i) = 0 \tag{S5}$$

We can then replace β_0 and β_1 with b and m, and divide both sides of eqn (S5) by the total number of measurements, n.

$$\frac{\sum_{i}^{n} y_i}{n} - \frac{b}{n} - \frac{m \sum_{i}^{n} x_i}{n} = \frac{0}{n} \tag{S6}$$

Because $\frac{\sum_{i}^{n} y_i}{n} = \bar{y}$ and $\frac{\sum_{i}^{n} x_i}{n} = \bar{x}$, eqn (S6) becomes:

$$\bar{y} - b - m\bar{x} = 0$$

$$b = \bar{y} - m\bar{x} \tag{S7}$$

where $\bar{x}$ and $\bar{y}$ are the average values for analyte concentration and instrument response.

The same way, to estimate m:

$$\frac{\partial S}{\partial \beta_1} = -2\sum_{i}^{n} x_i\left(y_i - \beta_0 - \beta_1 x_i\right) = 0 \tag{S8}$$

$$\sum_{i}^{n} x_i y_i - b\sum_{i}^{n} x_i - m\sum_{i}^{n} x_i^2 = 0 \tag{S9}$$

Substituting eqn (S7) in eqn (S9):

$$\sum_{i}^{n} x_i y_i - (\bar{y} - m\bar{x})\sum_{i}^{n} x_i - m\sum_{i}^{n} x_i^2 = 0$$

$$\sum_{i}^{n} x_i y_i - \bar{y}\sum_{i}^{n} x_i + m\bar{x}\sum_{i}^{n} x_i - m\sum_{i}^{n} x_i^2 = 0$$

$$\sum_{i}^{n} x_i y_i - \bar{y}\sum_{i}^{n} x_i - m\left(-\bar{x}\sum_{i}^{n} x_i + \sum_{i}^{n} x_i^2\right) = 0$$

$$\sum_{i}^{n} x_i y_i - \bar{y}\sum_{i}^{n} x_i = m\left(-\bar{x}\sum_{i}^{n} x_i + \sum_{i}^{n} x_i^2\right) \tag{S10}$$

Because $\sum_{i}^{n} x_i = n\bar{x}$, eqn (S10) becomes:

$$\sum_{i}^{n} x_i y_i - n\bar{x}\bar{y} = m\left(\sum_{i}^{n} x_i^2 - n\bar{x}\bar{x}\right)$$

$$\sum_{i}^{n} x_i y_i - n\bar{x}\bar{y} = m\left(\sum_{i}^{n} x_i^2 - n\bar{x}^2\right) \tag{S11}$$

Now, we just need to prove eqns (S12) and (S13), and use them to simplify eqn (S11).

$$\sum_{i}^{n} \left(x_i - \bar{x}\right)^2 = \sum_{i}^{n} x_i^2 - n\bar{x}^2 \tag{S12}$$

$$\sum_{i}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i}^{n} x_i y_i - n\bar{x}\bar{y} \tag{S13}$$

Proof of eqn (S12)

$$\sum_{i}^{n}(x_i - \bar{x})^2 = \sum_{i}^{n} x_i^2 - 2\bar{x}\sum_{i}^{n} x_i + n\bar{x}^2$$

Remember that $\frac{\sum_{i}^{n} x_i}{n} = \bar{x}$, therefore:

$$\sum_{i}^{n}(x_i - \bar{x})^2 = \sum_{i}^{n} x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2 = \sum_{i}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

$$\sum_{i}^{n}(x_i - \bar{x})^2 = \sum_{i}^{n} x_i^2 - n\bar{x}^2$$

Proof of eqn (S13)

$$\sum_{i}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i}^{n} x_i y_i - \bar{y}\sum_{i}^{n} x_i - \bar{x}\sum_{i}^{n} y_i + n\bar{x}\bar{y}$$

Again, $\frac{\sum_{i}^{n} x_i}{n} = \bar{x}$ and $\frac{\sum_{i}^{n} y_i}{n} = \bar{y}$, therefore:

$$\sum_{i}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i}^{n} x_i y_i - \bar{y}n\bar{x} - n\bar{x}\bar{y} + n\bar{x}\bar{y}$$

$$\sum_{i}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i}^{n} x_i y_i - n\bar{x}\,\bar{y}$$

Finally, from eqns (S11), (S12) and (S13):

$$m = \frac{\sum_{i}^{n} x_i y_i - n\bar{x}\bar{y}}{\left(\sum_{i}^{n} x_i^2 - n\bar{x}^2\right)}$$

$$m = \frac{\sum_{i}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i}^{n} (x_i - \bar{x})^2} \quad \text{(S14)}$$

References

S1. D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A*, Ch. 8, Straight Line Regression and Calibration, Elsevier, Amsterdam, 1997, pp. 171-230.

S2. N. R. Draper and H. Smith, *Applied Regression Analysis*, 2nd ed., John Wiley & Sons, New York, 1981, 709p.