2.4 Quantities related to the use of linear calibration functions

When reporting results obtained from linear calibration functions, the following quantities should also be reported: *number of observations*, the *equation representing the functional relation*, the *standard deviation of observations about the line*, and the estimated (fitted) parameters and their standard deviations. The result of chemical analysis should be quoted as the *estimated value of the independent variable* with its *confidence limits*.

Number of Observations

The total number of points measured when obtaining the calibration function. Symbol: m.

Degrees of Freedom

The number of observations minus the number of fitted parameters. Symbol: n. For linear calibration curves $n = m - 2$.

Independent Variable

The quantity (measured or computed), with values chosen arbitrarily or by design when performing the calibration process. Symbol: x (an individual value represented here as x_j). It is supposed that this value carries negligible error.

Dependent Variable

The quantity measured or computed, and plotted as the function of the independent variable. Symbol: y (an individual value represented here as y_j). It is subject to errors and deviations. When performing an analysis, this quantity is measured or computed from the measured signal.

Equation for the Functional (Calibration) Relation

The equation expressing the linear relation between the dependent and independent variables. This takes the form:

$$y = a + bx + e_y \tag{17}$$

where a is the intercept with the y axis and b the slope of the line, both estimated by the method of least squares, and e_y is the error in y.

Comment: When the calibration relation is strictly linear, as shown above, and applies to the entire chemical measurement process, the intercept can be identified directly with the blank *B*, and the slope, with the sensitivity *A*, as done in Chapter 18.

Slope

This is the parameter *b* in the calibration equation. It can be estimated by the formula:

$$\hat{b} = \frac{m \sum x_j y_j - \sum x_j \sum y_j}{m \sum x_j^2 - (\sum x_j)^2} \tag{18}$$

Intercept

This is the parameter *a* in the calibration equation. It can be estimated by the formula:

$$\hat{a} = \frac{\sum y_j \sum x_j^2 - \sum x_j \sum x_j y_j}{m \sum x_j^2 - (\sum x_j)^2} = (\sum y_j - b \sum x_j)/m \tag{19}$$

Correlation coefficient

The terms regression and correlation refer to statistical, as opposed to functional relations among variables. Thus, the statistical parameter estimates for the slope and intercept of the calibration function may be properly described by a correlation coefficient. This is given by:

$$r(a,b) = -\bar{x} / \bar{x}_q \tag{20}$$

Knowledge of this correlation coefficient, and the related confidence ellipse for *mutually* consistent values of ***a*** and ***b***, is necessary in certain circumstances. For example, this confidence ellipse may be used to derive bounds for the entire calibration line, i.e., bounds suitable for use with all future values of *x*. It is important also in two other situations: (a) generating a reduced confidence region for the line, when *a* or *b* is known to be restricted to a limited range, and (b) when confidence intervals for *functions* of the estimated parameters (***a, b***) are desired.

Comment: The rather popular usage of a correlation coefficient as a measure of the co-variation of the dependent variable (*y*) and the independent variable (*x*) is not recommended for calibration curves or other functional relations, because *r* is properly a measure of *statistical* associations.

Standard Deviation of Points About the Fitted Line

An estimate of the precision of the (dependent variable) measurements. Also known as the *Residual Standard Deviation*. Symbol: s or s_y. It can be calculated by the formula:

$$s_y = \sqrt{\frac{\sum y_j^2 - \frac{(\sum y_j)^2}{m} - \frac{\left(\sum x_j y_j - \frac{\sum x_j \sum y_j}{m}\right)^2}{\left(\sum x_j^2 - \frac{(\sum x_j)^2}{m}\right)}}{m - 2}} \tag{21}$$

or, more simply from its definition (following the estimation of a and b),

$$s_y = \sqrt{\sum\left[y - \left(\hat{a} + \hat{b}x\right)\right]^2 /(m - 2)} \tag{22}$$

where m - 2 represents the number of degrees of freedom.

Standard Deviation of the Slope

A quantity related to the precision of the estimated slope of the fitted line. Symbol: s_b. It can be calculated by the formula:

$$s_b = \sqrt{\frac{m s_y^2}{m\sum x_j^2 - \left(\sum x_j\right)^2}} \tag{23}$$

Standard Deviation of the Intercept

A quantity related to the precision of the estimated intercept of the fitted line. Symbol: s_a. It can be calculated by the formula:

$$s_a = s_b \sqrt{\frac{\sum x_j^2}{m}} = s_b \bar{x}_q \tag{24}$$

Confidence Limits About the Slope

Limits ($\pm C_b$) about the estimated value of the slope, corresponding to confidence level $1\text{-}\alpha$. The quantity C_b can be calculated by the formula:

$$C_b = t_{p,n}\, s_b \qquad (25)$$

Confidence Limits About the Intercept

Limits ($\pm C_a$) about the value of intercept corresponding to confidence level $1\text{-}\alpha$. The quantity C_a can be calculated by the formula:

$$C_a = t_{p,n}\, s_a \qquad (26)$$

See comment on *Correlation coefficient* concerning the confidence ellipse.

Estimated Value of the Independent Variable

The value of the independent variable, obtained from a measured or selected value of the dependent variable, y^* through the fitted equation. Symbol: $\hat{x}$. It can be calculated by the formula:

$$\hat{x} = \frac{y^* - \hat{a}}{\hat{b}} \qquad (27)$$

Confidence Limits About the Estimated Value of the Independent Variable

Limits $\pm C_x$ about the estimated value of the independent variable, x, corresponding to confidence level $1\text{-}\ \alpha$. The quantity C_x can be calculated by the formula:

$$C_x \approx t_{p,n}\ \frac{s_y}{b}\sqrt{1 + \frac{1}{m} + \frac{m\left[y^* - \frac{\Sigma y_j}{m}\right]^2}{[m\Sigma x_j^2 - (\Sigma x_j)^2]\, b^2}} \qquad (28)$$

For the meaning of $t_{p,v}$, see *Confidence Limits About the Mean*. The relation is approximate, because random error in *b* introduces some asymmetry (which can be taken into account with a more rigorous expression). Unless the relative standard deviation of *b* is large, the approximation is quite adequate.

Comment: If the value y* is obtained as the arithmetic mean of *n* replicate measurements, the following equation must be used:

$$C_x \approx t_{p,n} \frac{s_y}{b} \sqrt{\frac{1}{n} + \frac{1}{m} + \frac{m\left[\overline{y}* - \sum \frac{y_j}{m}\right]^2}{[m\sum x_j^2 - (\sum x_j)^2]\, b^2}} \qquad (29)$$

Estimated Value of the Dependent Variable

The predicted value of the dependent variable which corresponds to a selected value of the independent variable *x**. Symbol: $\hat{y}$. It can be estimated by the formula:

$$\hat{y} = \hat{a} + \hat{b}x^* \qquad (30)$$

Confidence Limits About the Estimated Value of the Dependent Variable

Limits ($\pm C_y$) about the estimated value of the dependent variable, $\hat{y}$, corresponding to confidence level $1\text{-}\alpha$. The quantity C_y can be calculated by the formula:

$$C_y = t_{p,n}\, s_y \sqrt{\frac{1}{m} + \frac{m\left[x^* - \frac{\sum x_j}{m}\right]^2}{m\sum x_j^2 - (\sum x_j)^2}} \qquad (31)$$

For the meaning of $t_{p,v}$, see *Confidence Limits About the Mean*.

Minimum Significant Signal (Critical Level)

The minimum value of the net signal, $\hat{a}$, that is statistically significant. Symbol: S_C. It can be calculated by the formula:

$$S_C = t_{p,n}\, s_0 \qquad (32)$$

where $t_{p,v}$ is the critical value from the *t*-distribution, and s_0 is the estimated standard deviation of the net signal when $x = 0$:

$$s_o = \left(s_a^2 + s_y^2\right)^{½} \qquad (33)$$

<u>Comment</u>: S_C is employed to make Detection Decisions. If the observed net signal exceeds S_C, it is considered "Detected" at the (***1-p***) significance level since this is a 1-sided test.

Minimum Detectable Concentration or Amount (Detection Limit)

The minimum value of the independent variable that can be confidently detected (probability ***p***), when S_C is employed as the decision threshold. Symbol: x_D. It can be calculated by the formula:

$$x_D \approx \left(2t_{p,n}\, s_0 / b\right)\left(K / I\right) \qquad (34)$$

where:

$$K = 1 + r(a,b)\left(s_a / s_0\right) t_{p,n}\left(s_b / b\right) \qquad (35)$$

$$I = 1 - t_{p,n}^2\left(s_b / b\right)^2 \qquad (36)$$

<u>Comment</u>: x_D, as indicated above, is strictly speaking an *estimate* for the minimum detectable quantity; it is the maximum null-signal upper limit for a *particular realization* of the calibration curve. If s_y were known without error, the relative uncertainty interval of x_D would be no greater than that of the slope. When s_y is used as an estimate of s_y, the uncertainty in x_D is further amplified by the confidence interval for s/s. Note that the ratio K/I » 1 when s_b « $b/t_{p,v}$; when $s_b > b/t_{p,v}$, the uncertainty in the detection limit is unbounded (see Chapter 18, and the comment following the Equation for the Functional (Calibration) Relation).