

8.4: Curvilinear, Multivariable, and Multivariate Regression

A straight-line regression model, despite its apparent complexity, is the simplest functional relationship between two variables. What do we do if our calibration curve is curvilinear—that is, if it is a curved-line instead of a straight-line? One approach is to try transforming the data into a straight-line. Logarithms, exponentials, reciprocals, square roots, and trigonometric functions have been used in this way. A plot of $\log(y)$ versus x is a typical example. Such transformations are not without complications, of which the most obvious is that data with a uniform variance in y will not maintain that uniform variance after it is transformed.

Note

It is worth noting here that the term “linear” does not mean a straight-line. A linear function may contain more than one additive term, but each such term has one and only one adjustable multiplicative parameter. The function

$$y = ax + bx^2$$

is an example of a linear function because the terms x and x^2 each include a single multiplicative parameter, a and b , respectively. The function

$$y = x^b$$

is nonlinear because b is not a multiplicative parameter; it is, instead, a power. This is why you can use linear regression to fit a polynomial equation to your data.

Sometimes it is possible to transform a nonlinear function into a linear function. For example, taking the log of both sides of the nonlinear function above gives a linear function.

$$\log(y) = b \log(x)$$

Another approach to developing a linear regression model is to fit a polynomial equation to the data, such as $y = a + bx + cx^2$. You can use linear regression to calculate the parameters a , b , and c , although the equations are different than those for the linear regression of a straight-line. If you cannot fit your data using a single polynomial equation, it may be possible to fit separate polynomial equations to short segments of the calibration curve. The result is a single continuous calibration curve known as a spline function. The use of R for curvilinear regression is included in Chapter 8.5.

Note

For details about curvilinear regression, see (a) Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemometrics*, Wiley-Interscience: New York, 1986; (b) Deming, S. N.; Morgan, S. L. *Experimental Design: A Chemometric Approach*, Elsevier: Amsterdam, 1987.

The regression models in this chapter apply only to functions that contain a single dependent variable and a single independent variable. One example is the simplest form of Beer's law in which the absorbance, A , of a sample at a single wavelength, λ , depends upon the concentration of a single analyte, C_A

$$A_\lambda = \epsilon_{\lambda,A} b C_A$$

where $\epsilon_{\lambda,A}$ is the analyte's molar absorptivity at the selected wavelength and b is the pathlength through the sample. In the presence of an interferent, I , however, the signal may depend on the concentrations of both the analyte and the interferent

$$A_\lambda = \epsilon_{\lambda,A} b C_A + \epsilon_{\lambda,I} b C_I$$

where $\epsilon_{\lambda,I}$ is the interferent's molar absorptivity and C_I is the interferent's concentration. This is an example of multivariable regression, which is covered in more detail in Chapter 9 when we consider the optimization of experiments where there is a single dependent variable and two or more independent variables.

Note

For more details on Beer's law, see Chapter 10 of *Analytical Chemistry 2.1*.

In multivariate regression we have both multiple dependent variables, such as the absorbance of samples at two or more wavelengths, and multiple independent variables, such as the concentrations of two or more analytes in the samples. As discussed in Chapter 0.2, we can represent this using matrix notation

$$\begin{bmatrix} \dots & \dots & \dots \\ \vdots & A & \vdots \\ \dots & \dots & \dots \end{bmatrix}_{r \times c} = \begin{bmatrix} \dots & \dots & \dots \\ \vdots & \epsilon b & \vdots \\ \dots & \dots & \dots \end{bmatrix}_{r \times n} \times \begin{bmatrix} \dots & \dots & \dots \\ \vdots & C & \vdots \\ \dots & \dots & \dots \end{bmatrix}_{n \times c}$$

where there are r wavelengths, c samples, and n analytes. Each column in the ϵb matrix, for example, holds the ϵb value for a different analyte at one of r wavelengths, and each row in the C matrix is the concentration of one of the n analytes in one of the c samples. We will consider this approach in more detail in Chapter 11.

Note

For a nice discussion of the difference between multivariable regression and multivariate regression, see Hidalgo, B.; Goodman, M. "Multivariate or Multivariable Regression," *Am. J. Public Health*, **2013**, *103*, 39-40.

8.4: Curvilinear, Multivariable, and Multivariate Regression is shared under a [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) license and was authored, remixed, and/or curated by LibreTexts.