Calibration and Regression

Partners in Measurement

By Khalid S. Al-Ghamdi

Q How do we use regression analysis for calibration?

A As a statistical term, regression originated with Sir Francis Galton (1822-1911) in his 1866 article, "Regression towards Mediocrity in Hereditary Stature." Originally, Galton used the word reversion instead of regression. Galton studied the relationship between parents' heights and the heights of their offspring. He studied the heights of 928 adults born to 205 sets of parents. He considered the parents' average heights as the predictor variable *x* and their respective offspring's heights as the dependent variable *y*. He found that unusually tall parents, i.e., taller than average (or mediocrity), tended to have children who were shorter than themselves, and parents who were shorter than average tended to have children taller than themselves. So, he formulated his theory "regression toward mediocrity" to represent the tendency in the relationship between heights of parents and children.

There are two types of relationships: functional and statistical. A functional relationship may be expressed as $y = f(x)$, where *x* is the independent variable and *y* is the dependent variable. If the bivariate data points (x,y) are plotted, they fall exactly on the curve described by $y = f(x)$. A statistical relationship may be described as $Y = \mathbf{f}(X) + \varepsilon$, where the symbol ε represents random error, and the capital letters for *X* and *Y* indicate that they are not fixed values but random variables. Depending on the context, *X* may be regarded as either fixed or random. If the bivariate data points generated by observing the (X,Y) process are plotted, they will not lie exactly on the curve described by $y = f(x)$. In this relationship, *X* is called the independent, regressor, or explanatory variable; *Y* is called the dependent variable or the response variable.

Terminology relating to quality and statistics (E456) defines calibration as a "process of establishing a relationship between a measurement device and a known standard value(s)." The International Organization for Standardization (ISO) defines calibration (VIM, JCGM 200:2012) as an "operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication." So, in metrology, measurement device calibration "establishes" or estimates the relationship between known inputs and measured outputs and the random error or uncertainty in the measured outputs. This is accomplished through regression analysis. Whenever there is measurement activity going on, there is calibration behind it, and regression is behind the calibration, except in the case of direct reading instruments.

The simplest form of regression, and the form most often used in calibration, is linear regression. The general regression function $f(X)$ is replaced by a linear function, giving $Y = a + bX + \varepsilon$. The error ε is assumed to have mean 0 and variance σ^2. The parameters *a* and *b* are most often estimated using ordinary least squares (OLS). This is done by finding values of *a* and *b* (denoted as $\hat{a}$ and $\hat{b}$) that minimize the sum of squared residuals,

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{a} - \hat{b}x_i \right)^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2$$

where $\hat{y}_i$ is the estimated average value of the response *y* corresponding to the value of x_i. The residuals are the difference between the observed values of *y* and the expected values from the regression equation. The parameter *b* is the slope of the line, and *a* is the y-intercept.

In classical calibration, the independent variable is the measurand, the property being measured, represented by the set of calibration standard values. The response variable is the reading (indication) from the instrument. The slope estimate $\hat{b}$ is the coefficient of sensitivity, the change in the average response given a unit change in the measurand. The parameter $\hat{a}$ is an estimate of the average reading when the measurand value is zero, such as for a blank sample.

The residuals should be approximately normally distributed with constant spread. Figure 1 shows an example calibration plot. However, Figure 2 reveals the pattern in the plot of residuals versus fitted values and shows increasing spread with increasing level of the standard. It is common for the spread of the residuals to be an increasing function of the measurand (and fitted values). In analytical chemistry, the variance of repeated measurements commonly follows a power law. In these cases, OLS is not appropriate. The

calibration model should be fit using weighted least squares (WLS) with weights inversely proportional to the variance of repeated measurement to assure the greatest possible accuracy in the parameter estimates. The variance of repeated measurement may be estimated using many replicates at each level of the measurand and computing a variance at each level or by estimating a (nonlinear) regression for the variance as a function of the measurand.

Inverse calibration is regression using the instrument reading as the explanatory variable and standard value as the response variable. The measurand value of an unknown test item is then calculated directly from the inverse calibration equation rather than by inverting the ordinary calibration equation. In the case in which the indication from the instrument is multivariate, as in many spectroscopic applications, inverse calibration is the only possible approach. It is in such cases performed using techniques such as multiple linear regression, principal components regression (PCR), or partial least squares (PLS).

A good calibration design is essential for building a good calibration model. The standard levels should cover the anticipated range of measurement. Often, four to seven levels are recommended so that the model fit can be better verified over the working range, so that constancy of variance can be checked and to give better parameter estimates. Sometimes quadratic or nonlinear models are used instead of linear models. Having enough standard levels facilitates selection of the most appropriate model using analysis of variance (ANOVA). There should be enough replicates at each level to estimate the variance of the response.

The frequency of calibration should be no more than required in order to avoid over-adjustment of the measurement system and to reduce the associated costs of calibration. The calibration can be monitored by regularly measuring a chosen calibration standard and control charting the result. A signal on the control chart indicates the need for recalibration. This is condition-based calibration, although a maximum time interval for recalibration should also be used to be compliant with ISO/IEC 17025.

Calibration is used universally in measurement science. Regression is the most powerful and widely used technique in statistics. It is fitting that the two work together so well.

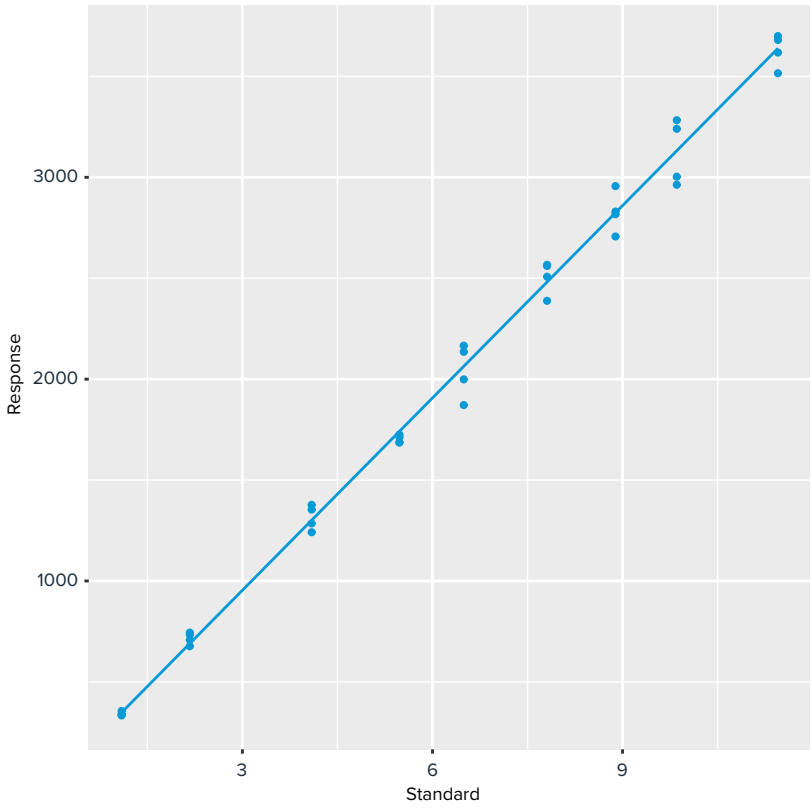

Figure 1 — An Example Calibration Plot

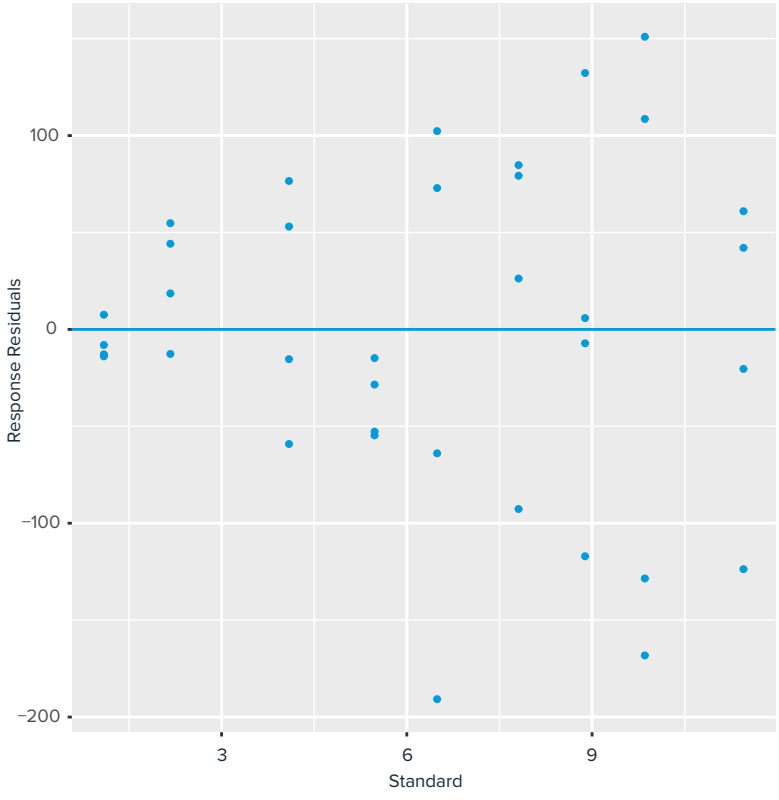

Figure 2 — Residuals versus FItted Values

Khalid S. Al-Ghamdi, unit supervisor at Saudi Aramco, Saudi Arabia, is a member of the committees on quality and statistics (E11) and on water (D19).

John Carson, Ph.D., of P&J Carson Consulting LLC, Findlay, Ohio, is the Data Points column coordinator.He is chairman of the subcommittee on statistical quality control (E11.30), part of the committee on quality and statistics (E11), and a member of the committees on petroleum products, liquid fuels, and lubricants (D02); air quality (D22); and environmental assessment, risk management, and corrective action (E50).