amc technical brief

Analytical Methods Committee No. 3. Dec 2000

Is my calibration linear?

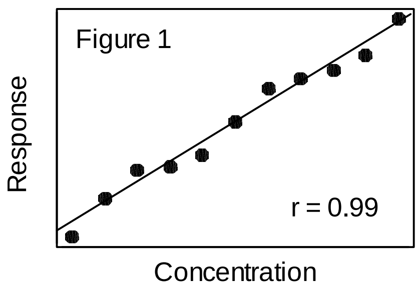
Examining a calibration function for linearity is an everyday task in routine analytical operations and in validating an analytical method. Linearity is an important and desirable feature of an analytical method. For example, if a calibration function is linear, then it is easier to estimate the equation, and evaluation errors (errors in estimating unknown concentrations from the calibration function) are likely to be smaller. Moreover, the assumption of calibration linearity is implicit in the valid use of the method of standard additions. Given the importance of linear calibration, it is strange that most analytical chemists are prepared to use flawed statistical methodology in linearity tests.

Problems with the correlation coefficient

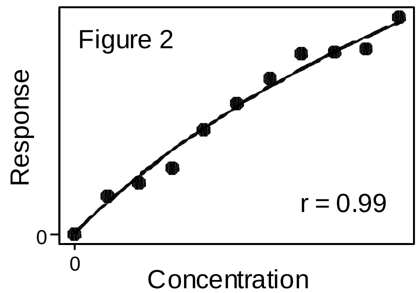
The flawed methodology, which is written into innumerable standard operating procedures, is the use of the correlation coefficient r, given by

$$r = \sum_i (x_i - \bar{x})(y_i - \bar{y}) \Big/ \sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}$$

Its use is based on a widespread misunderstanding. Certainly it is true that, if the calibration points are tightly clustered around a straight line, the experimental value of r will be close to unity (Fig.1).

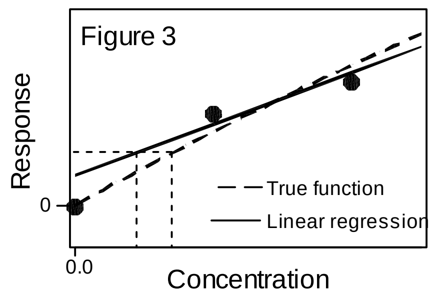


But the converse is not true. A value of r close to unity is not necessarily the outcome of a linear relationship but could, for example, result from points clustered around a slight curve (Fig. 2).



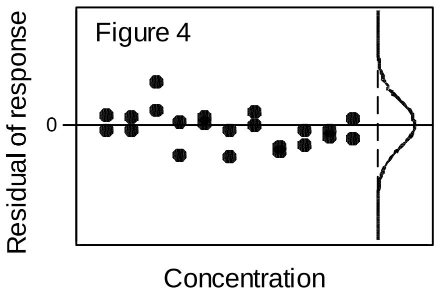
There is a further problem with the correlation coefficient. Values of r cannot properly be compared: we cannot say validly that a data set with r = 0.99 is 'more linear' than one with r = 0.95. The same problems affect the use of the R^2 statistic produced by regression software.

The use of a linear relationship to represent a truly curved calibration function might result in lack of fit that may produce seriously misleading results, especially near the bottom of the concentration range. Fig. 3 shows the outcome for the same data as in Fig. 2. Low concentrations of analyte could be subject to a relatively large systematic error.

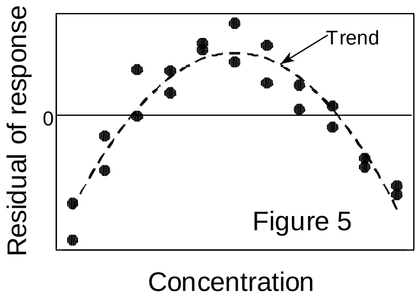


Testing for lack of fit

Strictly speaking, we cannot test for linearity as such. The best that we can achieve is to show that a deviation from linearity is too small to measure, that is, it's not statistically significant. An approach to linearity that is statistically sound is to examine the residuals from a linear regression. (This is a good thing to do anyway.) The residuals are the distances of the experimental points from the fitted regression line, measured in a direction parallel to the response axis. If there is no lack of fit, (that is, the calibration is inherently linear) the residuals plotted against concentration will look like a random sample from a normal distribution with zero mean. An example is shown in Fig. 4.



If there is lack of fit, a pattern might be discernible in the residual plot. But how can we tell if an apparent pattern is significant? That can be achieved by replicating the measurements at each calibration point. This gives us information about the inherent variability of the response measurements (called the 'pure error'). So we could see in Fig. 5, for example, that the systematic deviation of the residuals from zero was large in relation to the differences between the duplicated measurements, and therefore probably statistically significant.



In cases of doubt, a statistical test (analysis of the variance of the residuals into lack of fit and pure error) can be applied to the data as part of the regression. If there is significant lack of fit and the pattern of the residuals supports such an interpretation, we have demonstrated significant non-linearity in an unequivocal manner. (Given the data in Fig 5, the statistic shows that there is highly significant lack of fit.) If there is evidence of uneven variance in response, which is common in long range calibration, weighted regression should be used for the best results.

All of these statistical tests (including the production of residual plots) can be easily carried out in statistical software packages.

Of course, the experiment needs to be well designed. An effective design is to use eight or more concentrations of the calibrators, equally spaced across the concentration range of interest, and to measure each one twice, but in a random order. (There must be enough calibration points to make a pattern discernible - eight seems to be a practical minimum. Randomness avoids the problem of confusing non-linearity with temporal effects such as calibration drift.)

As a final thought, we can ask whether we really require exact linearity. In many instances we could accept a deviation from linearity. That would be appropriate, for example, if the evaluation errors resulting from the use of a linear calibration function made an insignificant contribution to the overall uncertainty of the measurement result, and this is often the case.

Recommendations

The correlation coefficient in the context of linearity testing is potentially misleading, and should be avoided. Testing for lack of fit by examining the residuals after linear regression is statistically sound and easily executed.

Further reading

AMC, Uses (Proper and Improper) of Correlation Coefficients, *Analyst*, 1988, **113**, 1469.

AMC, Is My Calibration Linear?, *Analyst*, 1994, **119**, 2363.