

## Is my calibration linear?

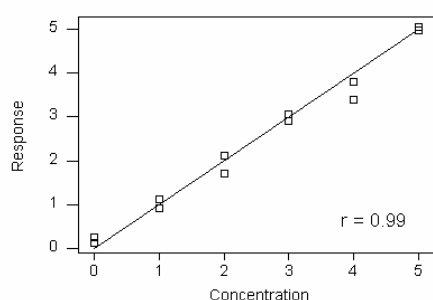
Examining a calibration function for linearity is an everyday task in both validating analytical methods and routine analytical operations. Linearity is an important and desirable feature of an analytical method. For example, if a calibration function is linear, then it is easier to estimate the equation, 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 for the valid use of the method of standard additions. Given the importance of linear calibration, it is strange that most analytical chemists are willing to use the correlation coefficient as an indicator of linearity.

### Properties of the correlation coefficient

The correlation coefficient  $r$ , given by

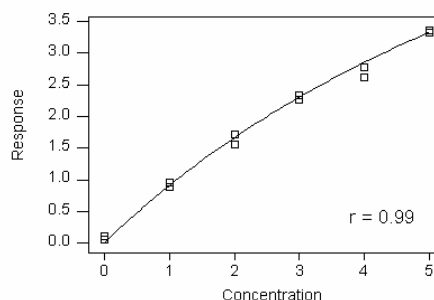
$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

is a measure of relationship between two variables  $x$  and  $y$ . It has several valuable properties. Its use in calibration, however, 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 (Figure 1, Dataset A).



**Figure 1.** Calibration (arbitrary units) with points clustered round a true straight line, with  $r = 0.99$  (Dataset A).

*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 readily visible curve (Figure 2, Dataset B). There is a related problem: 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.



**Figure 2.** Calibration (arbitrary units) with points clustered round a true curve, again with  $r = 0.99$  (Dataset B).

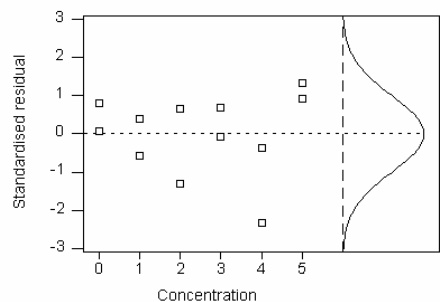
Of course, it's all a matter of degree. A calibration with  $r = 0.9999$  would necessarily be close to a straight line. The problem is that we cannot say how close, or whether it's close enough.

### Testing for lack of fit

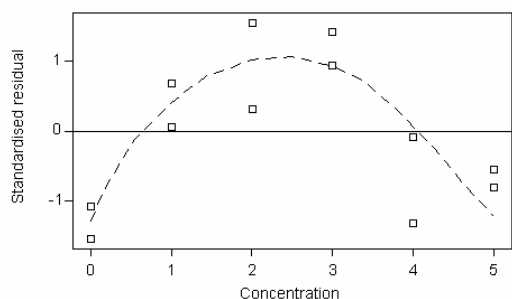
Strictly speaking, we cannot test for linearity as such. The best that we can achieve is to show that a deviation from linearity is too small to detect given our results, that is, it's not statistically significant. One approach along these lines is to examine the residuals from a linear regression. (That 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. As an example, residuals from Dataset A are plotted against concentration in Figure 3 (overleaf).

If there is non-linearity, however, a pattern should be discernible in the residual plot, typically a bow shaped trend in the points (Figure 4, Dataset B). 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 Figure 4, for example, that the systematic deviation of the residuals from zero was reasonably large in relation to the differences between the duplicated measurements, and therefore probably statistically significant.



**Figure 3.** Standardised residuals from linear regression using Dataset A. There is no convincing pattern in the residuals.



**Figure 4.** Standardised residuals from linear regression using Dataset B. A clear trend (dashed line) is apparent.

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 Figure 4, the statistic shows that there is significant lack of fit, with  $p \approx 0.01$ .) 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.

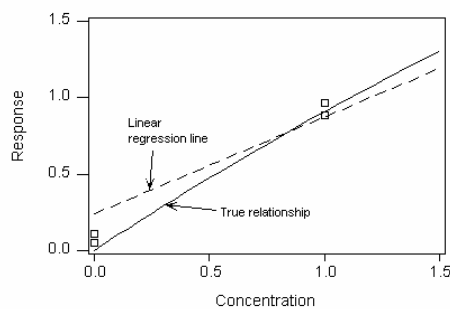
All of these statistical tests (including the production of residual plots) can be easily carried out in standard statistical software packages.

### Good design

Of course, the calibration experiment needs to be well designed. An effective design is to use six or more concentrations of the analyte, equally spaced across the concentration range of interest, and to measure each one twice, but with the measurements executed in a random order. The rationale for these measures is straight-forward: (a) there must be enough calibration points to make a pattern discernible—it's difficult to be exact here but six seems to be a practical minimum; (b) randomising the order of the measurements avoids the problem of confusing non-linearity with temporal effects such as instrumental drift occurring during the calibration.

### Exact linearity?

As a final thought, we can ask whether we really require exact linearity. We could accept a deviation from linearity if, for example, the evaluation uncertainty resulting from the use of a linear calibration function made an insignificant contribution to the overall uncertainty of the measurement result, and that is often the case. However, we have to consider which part of the calibration is relevant to our needs.



**Figure 5.** Part of calibration graph magnified (Figure 2, Dataset B), showing bias at low concentrations between the true relationship (solid line) and the linear regression (dashed line).

Near the bottom of the concentration range, 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. Figure 5 shows the outcome for Dataset B. Low concentrations of analyte could be subject to a relatively large systematic error.

### Recommendation

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.

*This Technical Brief was prepared for the Analytical Methods Committee by the Statistical Subcommittee (Chairman M Thompson)*

AMC Technical Briefs are informal but authoritative bulletins on technical matters of interest to analytical scientists. They are prepared by the Analytical Methods Committee of the Analytical Division of the RSC. More information and AMC products on [www.rsc.org/amc](http://www.rsc.org/amc)