Preparation of Calibration Curves

A Guide to Best Practice

September 2003
LGC/VAM/2003/032

Setting standards in analytical science

Preparation of Calibration Curves

A Guide to Best Practice

September 2003

Contact Point:
Liz Prichard
Tel: 020 8943 7553

Prepared by:
Vicki Barwick

Approved by:

_____

Date:

_____

The work described in this report was supported under contract with the Department of Trade and Industry as part of the National Measurement System Valid Analytical Measurement (VAM) Programme

Milestone Reference: KT2/1.3
LGC/VAM/2003/032

Contents

1. Introduction

Instrument calibration is an essential stage in most measurement procedures. It is a set of operations that establish the relationship between the output of the measurement system (*e.g.*, the response of an instrument) and the accepted values of the calibration standards (*e.g.*, the amount of analyte present). A large number of analytical methods require the calibration of an instrument. This typically involves the preparation of a set of standards containing a known amount of the analyte of interest, measuring the instrument response for each standard and establishing the relationship between the instrument response and analyte concentration. This relationship is then used to transform measurements made on test samples into estimates of the amount of analyte present, as shown in Figure 1.

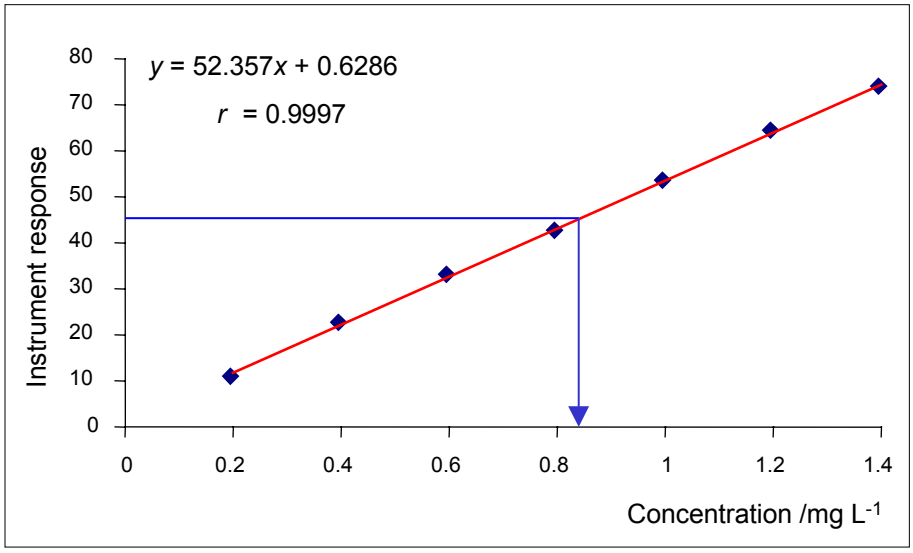


Figure 1: Typical calibration curve

As calibration is such a common and important step in analytical methods, it is essential that analysts have a good understanding of how to set up calibration experiments and how to evaluate the results obtained.

During August 2002 a benchmarking exercise was undertaken, which involved the preparation and analysis of calibration standards and a test sample using UV spectrophotometry. The aim of the exercise was to investigate the uncertainties associated with the construction of a calibration curve, and with using the calibration curve to determine the concentration of an unknown compound in an aqueous solution. In addition, it was hoped to identify any common problems encountered by analysts undertaking calibration experiments.

Members of the Environmental Measurement Training Network (EMTN) and the SOCSA Analytical Network Group (SANG) were invited to participate in the exercise. Five members of EMTN, six members of SANG and three organisations who are members of both EMTN and SANG submitted results. Some participants submitted results from more than one analyst, giving 19 sets of results in total. Full details of the protocol and results sheet circulated to the laboratories can be found in Appendix 1. Appendix 2 contains an ideal set of results from the benchmarking exercise to illustrate how the report should be presented.

The results of the benchmarking exercise were interesting. Although the exercise initially appeared relatively straightforward, a number of mistakes in carrying out the experiments and analysing the data were identified. Since a number of the mistakes occurred in more than one laboratory, it is likely that other laboratories carrying out similar exercises may make the same errors.

The aim of this guide is to highlight good practice in setting up calibration experiments, and to explain how the results should be evaluated. The guide focuses on calibration experiments where the relationship between response and concentration is expected to be linear, although many of the principles of good practice described can be applied to non-linear systems.

With software packages such as Excel, it easy to generate a large number of statistics. The guide also explains the meaning and correct interpretation of some of the statistical terms commonly associated with calibration.

2. The Calibration Process

There are a number of stages in the process of calibrating an analytical instrument. These are summarised below:

- Plan the experiments;
- Make measurements;
- Plot the results;
- Carry out statistical (regression) analysis on the data to obtain the calibration function;
- Evaluate the results of the regression analysis;
- Use the calibration function to estimate values for test samples;
- Estimate the uncertainty associated with the values obtained for test samples.

The guide considers each of these steps in turn.

2.1 Planning the experiments

The issues an analyst needs to consider when planning a calibration study are as follows:

- The number of calibration standards;
- The concentration of each of the calibration standards;
- The number of replicates at each concentration;
- Preparation of the calibration standards;

One of the first questions analysts often ask is, "How many experiments do I need to do?". Due to time and other constrains, this often translates as, "What is the absolute minimum I can do?". When thinking about a calibration experiment, this relates to the number of calibration standards that need to be analysed, and the amount of replication at each calibration level.

For an initial assessment of the calibration function, as part of method validation for example, standards with at least seven different concentrations (including a blank) should be prepared. The standard concentrations should cover, at least, the range of concentrations encountered during the analysis of test samples and be evenly spaced across the range (see Section 2.3.1). Ideally, the calibration range should be established so that the majority of the test sample concentrations fall towards the centre of the range. As discussed in Section 2.7, this is the area of the calibration range where the uncertainty associated with predicted concentrations is at its minimum. It is also useful to make at least duplicate measurements at each concentration level, particularly at the method validation stage, as it allows the precision of the calibration process to be evaluated at each concentration level. The replicates should ideally be independent – making replicate measurements on the same calibration standard gives only partial information about the calibration variability, as it only covers the precision of the instrument used to make the measurements, and does not include the preparation of the standards.

Having decided on the number and concentrations of the calibration standards, the analyst needs to consider how best to prepare them. Firstly, the source of the material used to prepare the standards (*i.e.*, the reference material used) requires careful consideration. The uncertainty associated with the calibration stage of any method will be limited by the uncertainty associated with the values of the standards used to perform the calibration – the uncertainty in a result can never be less than the uncertainty in the standard(s) used. Typically, calibration solutions are prepared from a pure substance with a known purity value or a solution of a substance with a known concentration. The uncertainty associated with the property value (*i.e.*, the purity or the concentration) needs to be considered to ensure that it is fit for purpose.

The matrix used to prepare the standards also requires careful consideration. Is it sufficient to prepare the standards in a pure solvent, or does the matrix need to be closely matched to that of the test samples? This will depend on the nature of the instrument being used to analyse the samples and standards and its sensitivity to components in the sample other than the target analyte. The accuracy of some methods can be improved by adding a suitable internal standard to both calibration standards and test samples and basing the regression on the ratio of the analyte response to that of the internal standard. The use of an internal standard corrects for small variations in the operating conditions.

Ideally the standards should be independent, *i.e.*, they should not be prepared from a common stock solution. Any error in the preparation of the stock solution will propagate through the other standards leading to a bias in the calibration. A procedure sometimes used in the preparation of calibration standards is to prepare the most concentrated standard and then dilute it by, say, 50%, to obtain the next standard. This standard is then diluted by 50% and so on. This procedure is not recommended as, in addition to the lack of independence, the standard concentrations will not be evenly spaced across the concentration range leading to the problem of leverage (see Section 2.3.1).

Construction of a calibration curve using seven calibration standards every time a batch of samples is analysed can be time-consuming and expensive. If it has been established during method validation that the calibration function is linear then it may be possible to use a simplified calibration procedure when the method is used routinely, for example using fewer calibration standards with only a single replicate at each level. A single point calibration is a fast way of checking the calibration of a system when there is no doubt about the linearity of the calibration function and the system is unbiased (*i.e.*, the intercept is not significantly different from zero, see Section 2.5.2). The concentration of the standard should be equal to or greater than the maximum concentration likely to be found in test samples.

If there is no doubt about the linearity of the calibration function, but there is a known bias (*i.e.*, a non-zero intercept), a two point calibration may be used. In this case, two calibration standards are prepared with concentrations that encompass the likely range of concentrations for test samples.

Where there is some doubt about the linearity of the calibration function over the entire range of interest, or the stability of the measurement system over time, the bracketing technique may be useful. A preliminary estimate of the analyte concentration in the test sample is obtained. Two calibration standards are then prepared at levels that bracket the sample concentration as closely as possible. This approach is time consuming but minimises any errors due to non-linearity.

2.2 Making the measurements

It is good practice to analyse the standards in a random order, rather than a set sequence of, for example, the lowest to the highest concentration.

All equipment used in an analytical method, from volumetric glassware to HPLC systems must be fit for their intended purpose. It is good science to be able to demonstrate that instruments are fit for purpose. Equipment qualification (EQ) is a formal process that

provides documented evidence that an instrument is fit for its intended purpose and kept in a state of maintenance and calibration consistent with its use. Ideally the instrument used to make measurements on the standards and samples should have gone through the EQ process.[1, 2, 3]

2.3 Plotting the results

It is always good practice to plot data before carrying out any statistical analysis. In the case of regression this is essential, as some of the statistics generated can be misleading if considered in isolation (see section 2.5).

Any data sets of equal size can be plotted against each other on a diagram to see if a relationship (a correlation) exists between them (Figure 2).

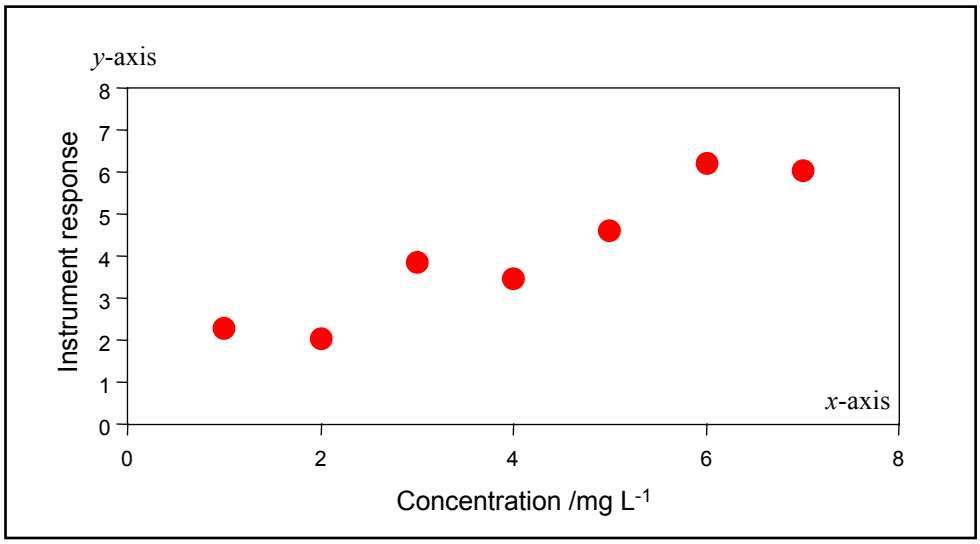


Figure 2: Scatter plot of instrument response data versus concentration

The horizontal axis is defined as the *x*-axis and the vertical axis as the *y*-axis. When plotting data from a calibration experiment, the convention is to plot the instrument response data on the *y*-axis and the values for the standards on the *x*-axis. This is because the statistics used in the regression analysis assume that the errors in the values on the *x*-axis are insignificant compared with those on the *y*-axis. In the case of calibration data, the assumption is that the errors in the instrument response values (due to random variation) are greater than those in the values assigned to the standards. In most cases this is not an unreasonable assumption.

The values plotted on the *y*-axis are sometimes referred to as the dependent variable, because their values depend on the magnitude of the other variable. For example, the instrument response will obviously be dependent on the concentration of the analyte present in the standards. Conversely, the data plotted on the *x*-axis are referred to as the independent variable.

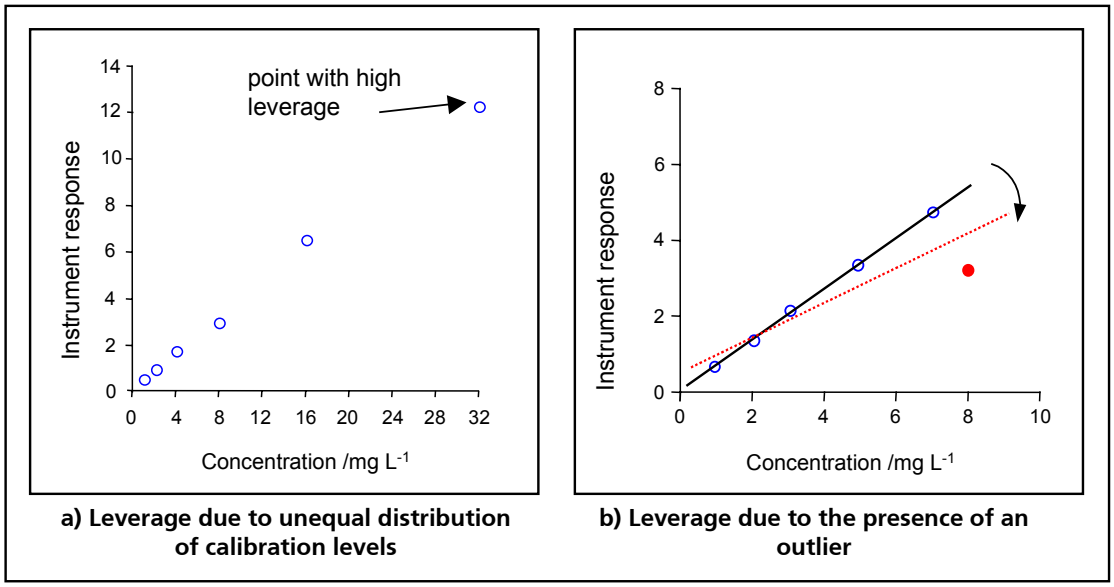
1 P. Bedson and M. Sargent, J. Accred. Qual. Assur., 1996, 1, 265-274.
2 P. Bedson and D. Rudd, J. Accred. Qual. Assur., 1999, 4, 50-62.
3 D. G. Holcombe and M. C. Boardman, J. Accred. Qual. Assur., 2001, 6, 468-478.

2.3.1 Evaluating the scatter plot

The plot of the data should be inspected for possible outliers and points of influence. In general, an outlier is a result which is significantly different from the rest of the data set. In the case of calibration, an outlier would appear as a point which is well removed from the other calibrations points. A point of influence is a calibration point which has a disproportionate effect on the position of the regression line. A point of influence may be an outlier, but may also be caused by poor experimental design (see section 2.1).

Points of influence can have one of two effects on a calibration line – leverage or bias.

Leverage



a) Leverage due to unequal distribution of calibration levels

b) Leverage due to the presence of an outlier

Figure 3: Points of influence – leverage

An outlier at the extremes of the calibration range will change the position of the calibration line by tilting it upwards or downwards (see Figure 3b). The point is said to have a high degree of leverage. Leverage can be a problem if one or two of the calibration points are a long way from the others along the *x*-axis (see Figure 3a). These points will have a high degree of leverage, even if they are not outliers. In other words, a relatively small error in the measured response will have a significant effect on the position of the regression line. This situation arises when calibration standards are prepared by sequential dilution of solutions (*e.g.*, 32 mg L^{-1}, 16 mg L^{-1}, 8 mg L^{-1}, 4 mg L^{-1}, 2 mg L^{-1}, 1 mg L^{-1}, as illustrated in Figure 3a).

Leverage affects both the gradient and intercept of the line with the *y*-axis.

Bias

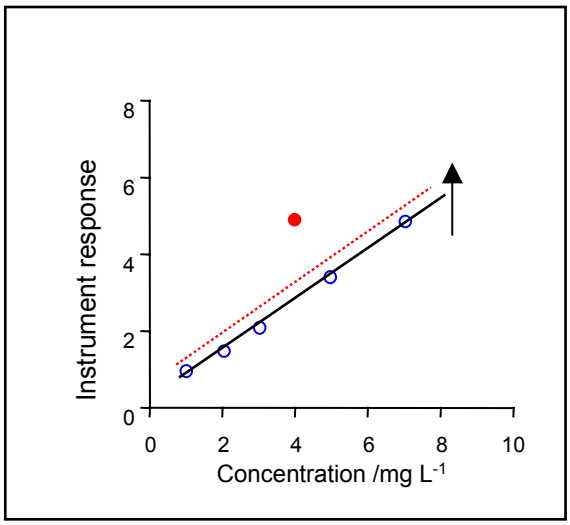


Figure 4: Points of influence – bias

An outlier in the middle of the calibration range (see Figure 4) will shift the regression line up or down. The outlier is a point of influence as it has introduced a bias into the position of the line. The gradient of the line will be approximately correct but the intercept will be wrong.

2.4 Carrying out regression analysis

In relation to instrument calibration, the aim of linear regression is to establish the equation that best describes the linear relationship between instrument response (y) and analyte level (x). The relationship is described by the equation of the line, *i.e.*, $y = mx + c$, where m is the gradient of the line and c is its intercept with the y-axis. Linear regression establishes the values of m and c which best describe the relationship between the data sets. The equations for calculating m and c are given in Appendix 3, but their values are most readily obtained using a suitable software package. Note that regression of y on x (as is usually done in a calibration study) is not the same as the regression of x on y. This is because the procedures used in linear regression assume that all the errors are in the y values and that the errors in the x values are insignificant. This is a reasonable assumption for many analytical methods as it is possible to prepare standards where the uncertainty in the concentration is insignificant compared with the random variability of the analytical instrument. It is therefore essential to ensure that the instrument response data and the standard concentrations are correctly assigned.

Understanding the principles of linear regression requires an understanding of residuals. A residual is the difference between an observed y value, and the y value calculated using the equation of the fitted line (see Figure 5). The residuals give an indication of how well the line fits the data. In Figure 5, the sum of the squared residuals for the poorly fitting dashed line will be much larger than for the solid best-fit line. It can be shown that the line that gives the smallest sum of the squared residuals best represents the linear relationship between the x and y variables. Software for linear regression simply calculates the values for m and c that minimise the sum of the squared residuals. For this reason, this type of regression is often referred to as, “least squares regression”.

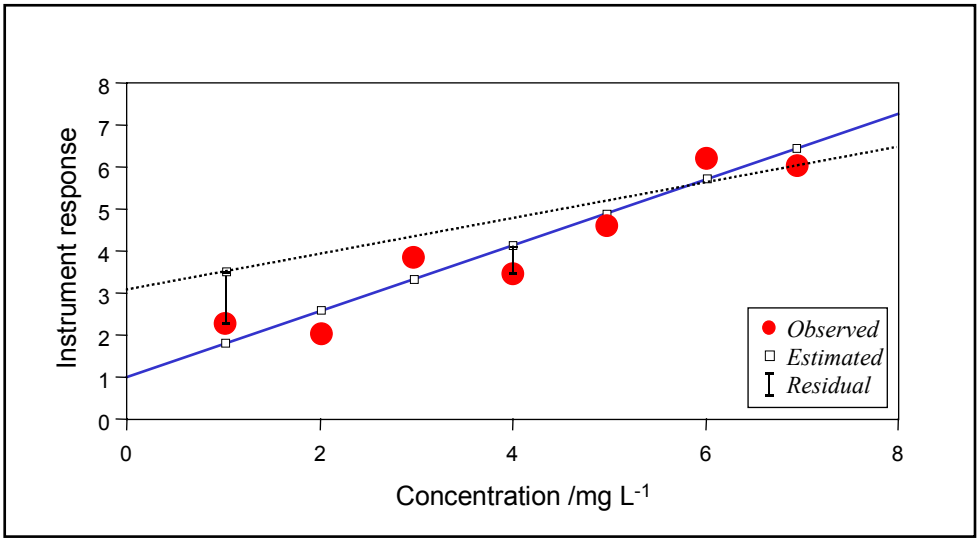


Figure 5: Least squares linear regression – calculating the best straight line

2.4.1 Assumptions

Basic least squares linear regression relies on a number of assumptions. A 'best-fit' line will only be obtained when the assumptions are met. Significant violation of any of the assumptions will usually require special treatment which is outside the scope of this guide.

The first assumption was mentioned in Section 2.4, that is that the error in the x values should be insignificant compared with that of the y values. In addition, the error associated with the y values must be normally distributed. Normality is hard to test for statistically with only small data sets. If there is doubt about the normality it may be sufficient to replace single y values with averages of three or more for each value of x, as mean values tend to be normally distributed even where individual results are not. The magnitude of the error in the y values should also be constant across the range of interest, *i.e.* the standard deviation should be constant. Simple least squares regression gives equal weight to all points – this will not be appropriate if some points are much less precise than others. In many chemical measurement systems the standard deviation increases with concentration, *i.e.*, it is the *relative* standard deviation that remains approximately constant rather than the standard deviation. The general solution to this problem is to use *weighted* regression, which takes account of the variability in the y values.[4]

Both the x and y data must be continuous valued, *i.e.*, not restricted to integers, significantly truncated or categorised (*e.g.*, sample numbers, days of the week, *etc*.). This assumption should be met in the case of instrument calibration as both the instrument response and the concentrations of the standards can, in theory, take any value on a continuous scale.

2.4.2 Carrying out regression analysis using software

The equations for carrying out a linear regression are given in Appendix 3, but regression is usually carried out using software supplied with the instrument or packages such as Excel. Many software packages allow a regression analysis to be carried out without first plotting the data, however it is good practice to produce a plot before carrying out the statistical analysis (see Sections 2.3 and 2.5.2). If the option is available, it is also useful to obtain a plot of the residuals (see Section 2.5.1).

4 Statistics and chemometrics for analytical chemistry, J. N. Miller and J. C. Miller, 4th Edition, Prentice Hall, 2000, ISBN 0-130-22888-5.

Most software also allows the intercept with the *y*-axis to be set to zero when carrying out the regression. This option should not be selected unless it has been proved that the intercept is consistently not significantly different from zero (see Section 2.5.2). Finally, ensure that the *x* and *y* data have been correctly assigned (see Section 2.4).

2.5 Evaluating the results of the regression analysis

Using software to carry out the linear regression will result in a number of different statistical parameters and possibly (depending on the software used) a table and/or plot of the residuals. The meaning and interpretation of each of the common outputs from a regression analysis is discussed below.

2.5.1 Plot of the residuals

Obtaining a plot of the residuals is strongly recommended as it can highlight problems with the calibration data that may not be immediately obvious from a simple scatter plot of the data. The construction of a residual plot is illustrated in Figure 6.

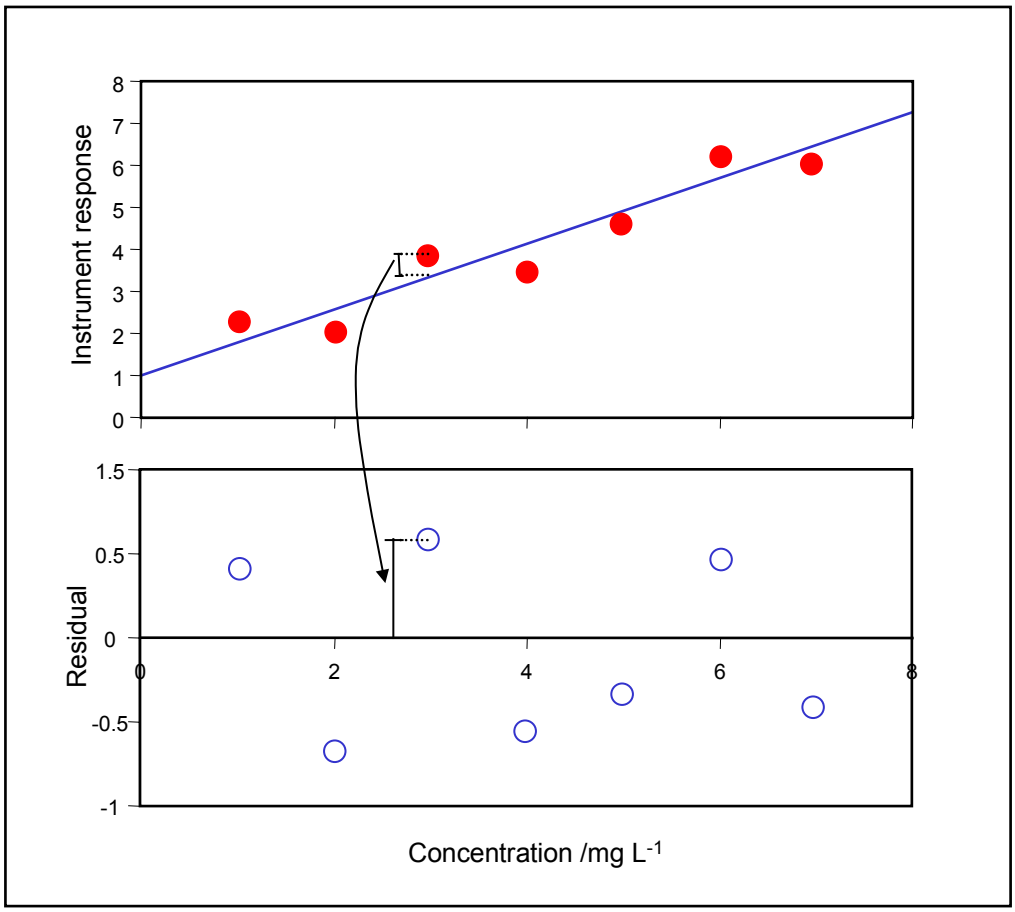


Figure 6: The residuals plot

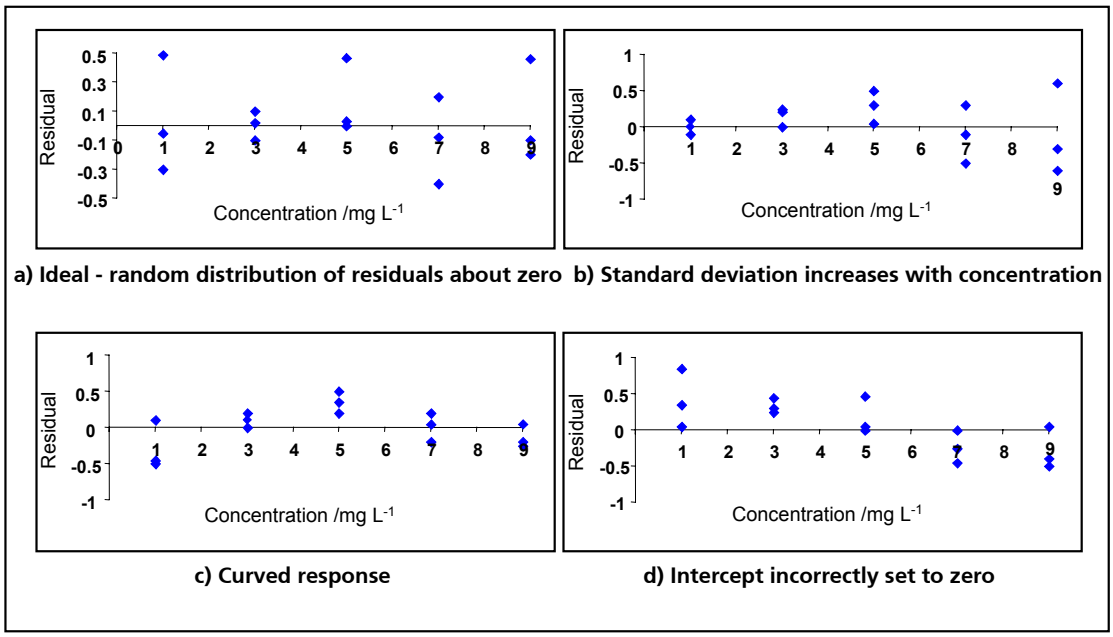


Figure 7: Examples of residuals plots

Figure 7a shows an ideal residual plot. The residuals are scattered approximately randomly around zero, and there is no trend in the spread of residuals with concentration. Figure 7b shows the pattern of residuals that is obtained if the standard deviation of the instrument response increases with analyte concentration. Figure 7c illustrates a typical residual plot that is obtained when a straight line is fitted through data that are non-linear. Finally, Figure 7d shows a possible pattern of residuals when the regression line has been incorrectly fitted through zero (see Sections 2.4.2 and 2.5.2). Figures b) to d) should all cause concern as the pattern of the residuals is clearly not random.

2.5.2 Regression statistics

A typical output from a regression analysis is shown in Figure 8. The output shown is from Excel, but similar information is obtained from other software. The different parts of the output are described in more detail in the following sections.

Regression Statistics	
Multiple R	0.999955883
R Square	0.999911768
Adjusted R Square	0.999889709
Standard Error	0.005164622
Observations	6

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1.2091	1.2091	45330.79	2.93×10^{-9}
Residual	4	0.00010669	2.67×10^{-5}		
Total	5	1.2092			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.0021129	0.0037548	0.56270	0.60368	-0.008312	0.012538
X Variable 1	0.10441	0.00049038	212.91	2.92×10^{-9}	0.10304	0.10577

Figure 8: Typical output from a regression analysis using Excel

The correlation coefficient, *r*

The correlation coefficient, r (and the related parameters r^2 and adjusted r^2) is a measure of the strength of the degree of correlation between the y and x values. In Excel output it is described as 'Multiple R'. r can take any value between +1 and –1; the closer it is to 1, the stronger the correlation. The correlation coefficient is one of the statistics commonly used in analytical measurement. Unfortunately, it is easily (and frequently) misinterpreted. The r value is easily misinterpreted because:

- correlation and linearity are only loosely related. The coefficient r is a measure of correlation *not* a measure of linearity;
- it is relatively easy to generate data with apparently good correlation. However, a plot of the data may well reveal that the data would be unsatisfactory for the purposes of calibration (see Figure 9);
- for predictions made from the calibration curve to have small uncertainties, r needs to be very close to 1 (see Section 2.7);
- A low r value does not necessarily mean that there is no correlation. There could be a relationship between the y and x values, but not a linear one (see Figure 9).

For these reasons, it is essential to plot calibration data, and not just rely on the statistics, when assessing the fitness-for-purpose of a calibration curve. Figure 9 shows some examples of how the correlation coefficient can be misleading.

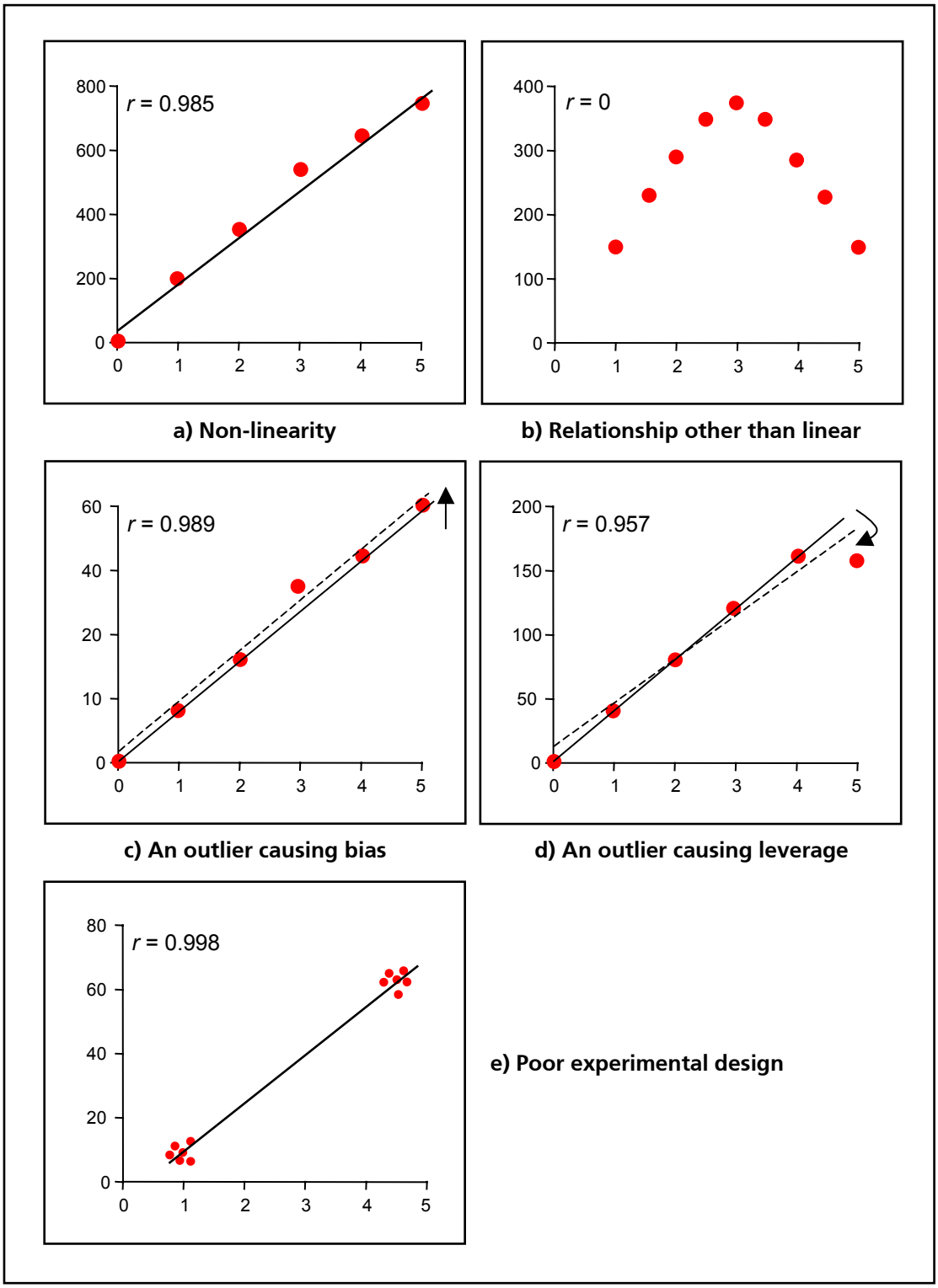


Figure 9: Interpreting the correlation coefficient

Figure 9a shows a case where the relationship is clearly not linear across the entire range of x values. In Figure 9b, the r value of zero indicates that there is no linear correlation. However, there is clearly a significant non-linear relationship. Figures 9c and 9d show the effect of individual outliers. In c) the intercept of the line fitted with the outlier present (broken line) will be incorrect compared to the line fitted with the outlier removed (solid line). In d) both the gradient and the intercept will be incorrect. In Figure 9e, the r value of 0.998 indicates a strong correlation. However, the two groups of points are distinct and neither shows any

significant correlation. Although there are 12 data points, the distribution of the points indicates that this is in effect a two-point calibration (which will always give $r = 1$).

The question which analysts often ask is, "How close to 1 does the correlation coefficient have to be for a 'good' calibration curve?". What the analyst is really after is a calibration line that will result in a satisfactory level of uncertainty in the values predicted from the fitted line. The particular value of r that shows a statistically significant correlation between y and x depends on the number of data points used to calculate it. Figure 10 shows the value of r that would indicate statistically significant correlation for different numbers of data points. Absolute values of r within the shaded area indicate a statistically significant correlation, at the 95% confidence level. Remember that statistical significance only indicates some evidence for correlation. It does not necessarily mean that the data would be appropriate for calibration. For example, with 10 data points a value of $r = 0.6$ would be statistically significant. However, it is highly unlikely that a calibration curve with a correlation coefficient of 0.6 would be of any use, as the uncertainties associated with predicted values obtained from such a line would be prohibitively large (see section 2.7).

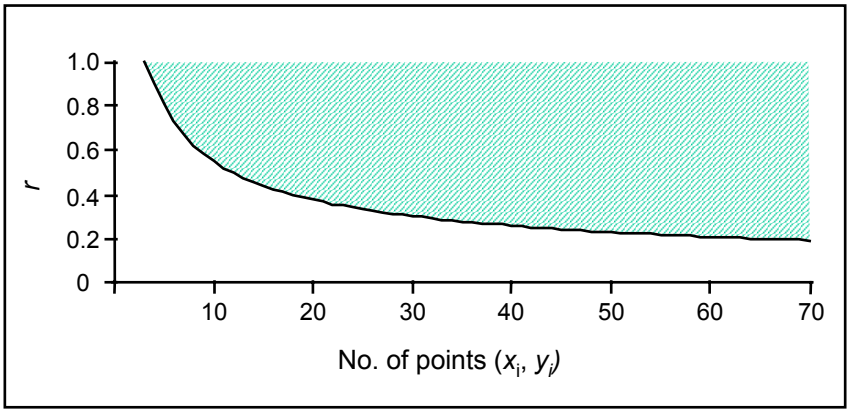


Figure 10: Statistically significant values of *r* (shaded area) at the 95% confidence level

The parameters related to r are r^2 and adjusted r^2. r^2 is often used to describe the fraction of the total variance in the data which is contributed by the line that has been fitted. Ideally, if there is a good linear relation, the majority of variability can be accounted for by the fitted line. r^2 should therefore be close to 1.

The adjusted r^2 value is interpreted in the same way as r^2 but is always lower. It is useful for assessing the effect of adding additional terms to the equation of the fitted line (*e.g.*, if a quadratic fit is used instead of a linear fit). The r^2 value always increases on the addition of an extra term to the equation, but this does not mean that the extended equation is necessarily a better fit of the data. The adjusted r^2 value is more useful in such cases as it takes account of the reduction in the degrees of freedom which occurs each time an additional term is added to the equation of the line (see section below on the residual standard deviation), and therefore does not automatically increase on addition of extra terms. This guards against 'overfitting', which occurs when the equation fitted has more terms than can be supported by the amount of data available (*i.e.*, there are insufficient degrees of freedom).

Residual standard deviation (or standard error)

The residual standard deviation (also known as the residual standard error) is a statistical measure of the deviation of the data from the fitted regression line. It is calculated using Eq. 1:

$$s(r) = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}} \qquad \text{Eq. 1}$$

where y_i is the observed value of y for a given value of x_i, $\hat{y}_i$ is the value of y predicted by the equation of the calibration line for a given value of x_i, and n is the number of calibration points.

ANOVA table

Software such as Excel produces an analysis of variance (ANOVA) table for the regression. The sum of squares terms (SS) represent different sources of variability in the calibration data. Figure 11 illustrates the origin of these terms. The regression term represents the variability in the data that can be accounted for by the fitted regression line. Ideally this should be large; if there is a good linear relationship, the fitted line will describe the majority of the variability in response with concentration. The residual term is the sum of the squared residuals (see section 2.4). This value should be small compared to the regression sum of squares terms because if the regression line fits the data well, the residuals will be small.

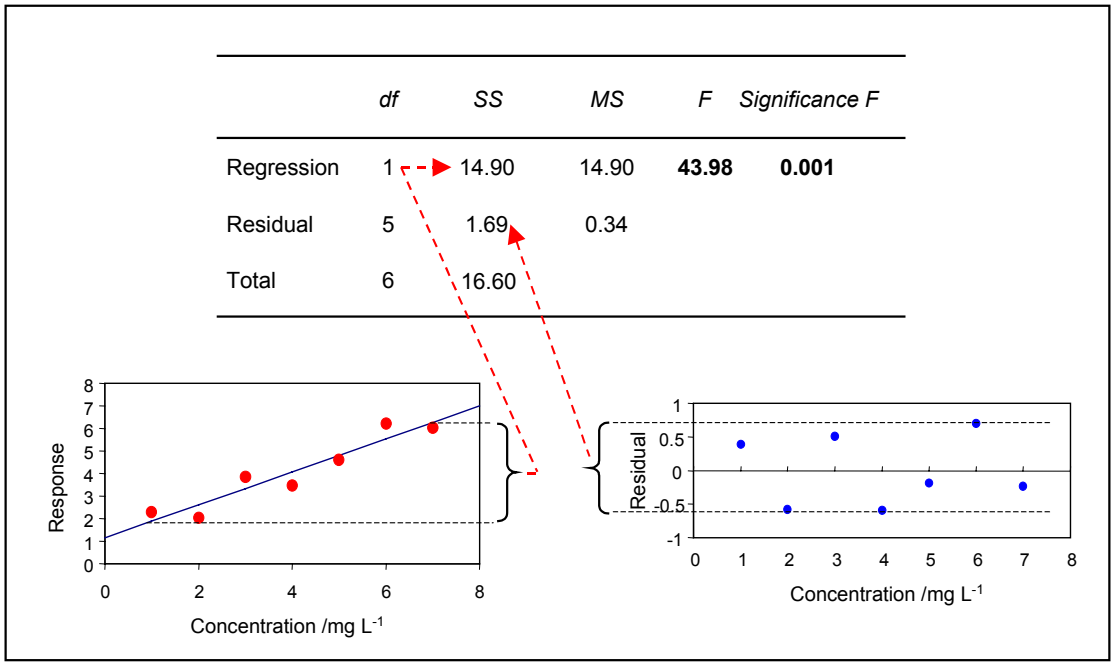


Figure 11: Origin of sum of squares terms in regression analysis

Each mean square (MS) term is simply the sum of squares term divided by its degrees of freedom. The F value is the ratio of the regression MS term to the residual MS term. Ideally this ratio should be very large; if there is a good linear relationship the regression MS term will be much greater than the residual MS term (see Figure 8). The significance F value represents the probability of obtaining the results in the ANOVA table if there is no correlation between y and x values, *i.e.*, obtaining the results by chance. A small value indicates that the results were unlikely to have happened by chance, indicating that it is highly

likely that there is a strong relationship between the y and x values. For a calibration curve to be of any use the significance F value should be extremely small (see Figure 8). This value is also known as the p-value.

Regression coefficients

The final section of the regression output shown in Figure 8 gives information about the regression coefficients m (the gradient of the line) and c (the intercept of the line with the y-axis). The first column of numbers gives the values of the coefficients. In Excel, the gradient is described as "X Variable 1". The next column gives the standard errors (also know as the standard deviations) for each coefficient. These values give an indication of the ranges within which the values for the gradient and intercept could lie. Related to these values are the lower and upper confidence limits for the gradient and intercept (final two columns of the table). These represent the extremes of the values that the gradient and intercept could take, at the chosen level of confidence (usually 95%). The equations for calculating these values are given in Appendix 3.

The t-stat and p-value relate to the significance of the coefficients, *i.e.* whether or not they are statistically significantly different from zero. In a calibration experiment we would expect the gradient of the line to be very significantly different from zero. The t-value should therefore be a large number (for a calibration with 7 data points the t-value should be much greater than 2.6, the 2-tailed Student t value for 5 degrees of freedom at the 95% confidence level) and the p-value should be small (much less than 0.05 if the regression analysis has been carried out at the 95% confidence level). Typical values are shown in Figure 8.

Ideally, we would like the calibration line to pass through the origin. If this is the case then the intercept should not be significantly different from zero. In the regression output we would expect to see a small value for t (less than 2.6 for a calibration with 7 data points) and a p-value greater than 0.05 (for regression at the 95% confidence level). Whether the calibration line can reasonably be assumed to pass through zero can also be judged by inspecting the confidence interval for the intercept. If this spans zero, then the intercept is not statistically different from zero, as in the example shown in Figure 8.

2.6 Using the calibration function to estimate values for test samples

If, after plotting the data and examining the regression statistics, the calibration data are judged to be satisfactory the calibration equation (*i.e.*, the gradient and the intercept) can be used to estimate the concentration of the analyte in test samples. This requires each sample to be analysed one or more times, under the same conditions that the calibration standards were measured. It is also useful to obtain an estimate of the uncertainty associated with the predicted concentration values for test samples. This is described in Section 2.7.

2.7 Estimating the uncertainty in predicted concentrations

Figure 12 illustrates the confidence interval for the regression line. The interval is represented by the curved lines on either side of the regression line and gives an indication of the range within which the 'true' line might lie. Note that the confidence interval is narrowest near the centre (the point $\bar{x}, \bar{y}$) and less certain near the extremes.

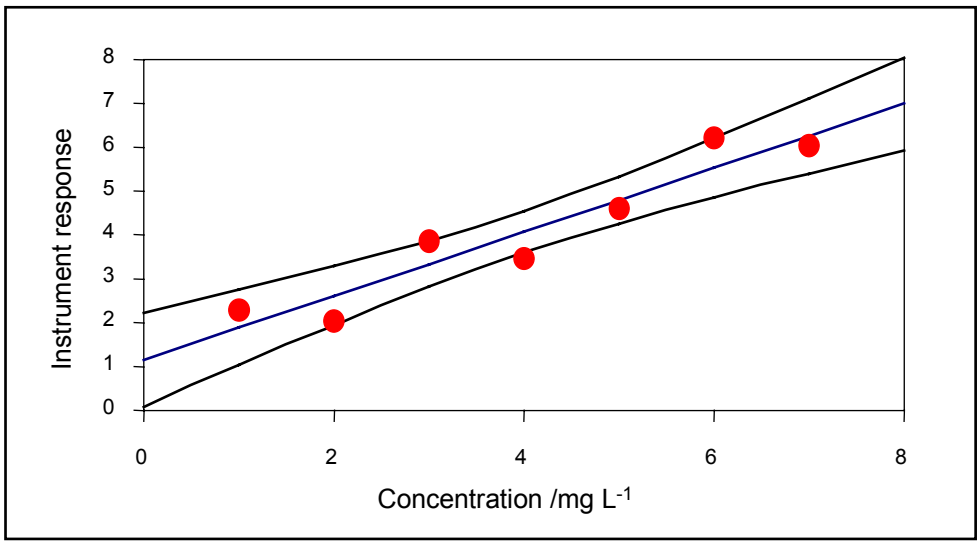


Figure 12: 95% confidence interval for the line

In addition, it is possible to calculate a confidence interval for values predicted using the calibration function. This is sometimes referred to as the 'standard error of prediction' and is illustrated in Figure 13. The prediction interval gives an estimate of the uncertainty associated with predicted values of x.

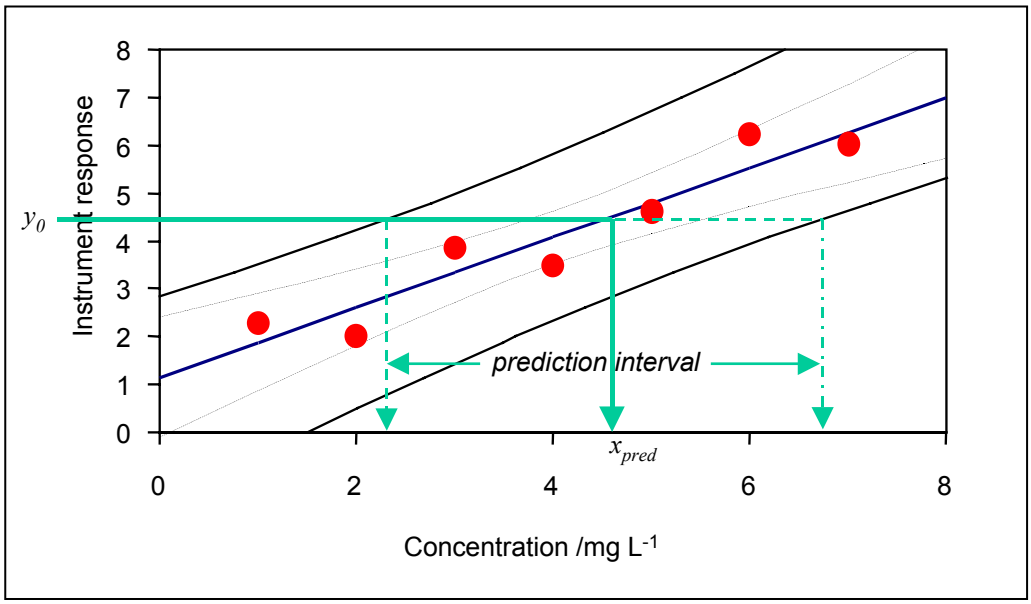


Figure 13: Prediction interval

The prediction interval s_{x_0} is calculated using Eq. 2:

$$s_{x_0} = \frac{s(r)}{m}\sqrt{\frac{1}{N}+\frac{1}{n}+\frac{(\bar{y}_o-\bar{y})^2}{m^2\sum_{i=1}^{n}(x_i-\bar{x})^2}} \qquad \text{Eq. 2}$$

Where:

$s(r)$ is the residual standard deviation (see Eq. 1)

n is the number of paired calibration points (x_i,y_i)

m is the calculated best-fit gradient of the calibration curve

N is the number of repeat measurements made on the sample (this can vary from sample to sample and can equal 1)

$\bar{y}_o$ is the mean of N repeat measurements of y for the sample

$\bar{y}$ is the mean of the y values for the calibration standards

x_i is a value on the x-axis

$\bar{x}$ is the mean of the x_i values

A confidence interval is obtained by multiplying s_{x_0} by the 2-tailed Student t value for the appropriate level of confidence and n-2 degrees of freedom.

2.8 Standard error of prediction worked example

Table 1 and Figure 14 show a set of calibration data which will be used to illustrate the calculation of a prediction interval.

Concentration /mg L^{-1}	Absorbance		
2.56	0.320		
5.12	0.591		
8.192	0.920	0.918	0.920
10.24	1.135		
12.80	1.396		

Table 1: Calibration data for standard error of prediction example

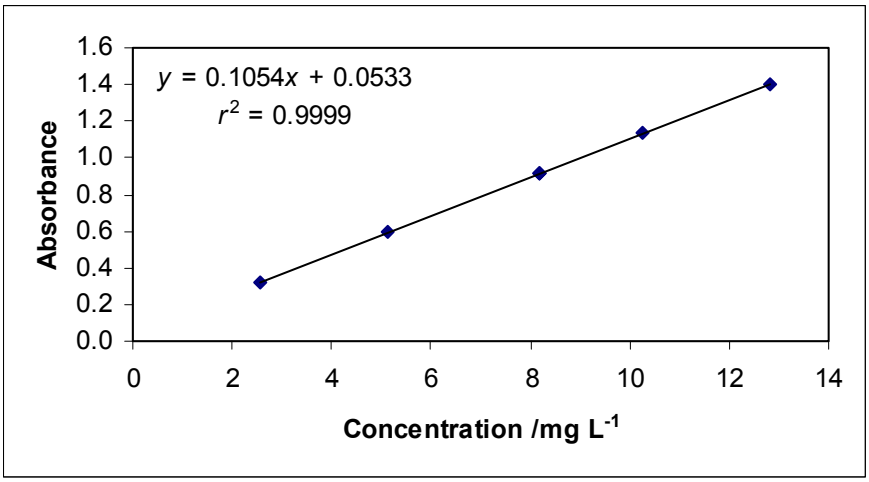


Figure 14: Plot of data for standard error of prediction example

The data required to calculate a prediction interval are shown in Table 2.

Conc x_i	**Absorbance** y_i	$(x_i - \bar{x})^2$	**Predicted** $\hat{y}_i = \mathrm{m}x_i + \mathrm{c}$	**Residuals** $y_i - \hat{y}_i$	**Residuals²** $(y_i - \hat{y}_i)^2$
2.56	0.320	28.509	0.323	-0.00306	$9.344\text{x}10^{-6}$
5.12	0.591	7.725	0.593	-0.00182	$3.327\text{x}10^{-6}$
8.192	0.920	0.0856	0.917	0.00346	$1.194\text{x}10^{-5}$
8.192	0.918	0.0856	0.917	0.00146	$2.118\text{x}10^{-6}$
8.192	0.920	0.0856	0.917	0.00346	$1.194\text{x}10^{-5}$
10.24	1.135	5.478	1.132	0.00264	$6.977\text{x}10^{-6}$
12.80	1.396	24.016	1.402	-0.00613	$3.753\text{x}10^{-5}$
$\bar{x}$	$\bar{y}$	$\sum_{i=1}^{n}(x_i - \bar{x})$			$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
7.899	0.886	65.985			$8.317\text{x}10^{-5}$

Table 2: Data required to calculate a prediction interval

Using Eq. 1 the residual standard deviation is calculated as:

$$s(r) = \sqrt{\frac{8.317\text{x}10^{-5}}{7-2}} = 0.00408$$

Applying Eq. 2, the prediction interval for a sample which gives an instrument response of 0.871, is:

$$s_{x_0} = \frac{0.00408}{0.1054}\sqrt{\frac{1}{1}+\frac{1}{7}+\frac{(0.871-0.886)^2}{0.1054^2 \times 65.985}} = 0.0414 \text{ mg L}^{-1}$$

Note that a single measurement is made on the sample so N = 1.

$$x_{\text{pred}} = \frac{0.871 - 0.0533}{0.1054} = 7.76 \text{ mg L}^{-1}$$

Expressed as a % of x_{pred}, s_{x_0} = 0.53%.

At the 95% confidence level, the 2-tailed Student *t* value for 5 degrees of freedom is 2.571. The 95% confidence interval for x_{pred} is 0.0414 x 2.571 = 0.106 mg L^{-1} (1.4%).

The uncertainty in predicted values can be reduced by increasing the number of replicate measurements (N) made on the test sample. Table 3 shows how s_{x_0} changes as N is increased.

N	s_{x_0} /mg L^{-1}	Uncertainty /%relative
1	0.041	0.53
2	0.031	0.40
3	0.027	0.35
4	0.024	0.31
5	0.023	0.30

Table 3: Standard error of prediction for different values of N

3. Conclusions

The benchmarking exercise highlighted a number of problems associated with carrying out instrument calibration. Some common pitfalls encountered in calibration studies include:

- the concentration range covered by the calibration standards does not adequately cover the range of concentrations encountered for test samples;
- the concentrations of the calibration standards are not evenly spaced across the calibration range;
- the uncertainty associated with the concentrations of the calibration standards is too large (e.g., inappropriate glassware is used to prepare the standards, the material used to prepare the standards is not of an appropriate purity);
- the wrong regression is carried out (*i.e.*, regression of x on y rather than y on x);
- the calibration line is fitted through zero when the intercept is, in fact, significantly different from zero;
- instrument software is used to carry out the regression and automatically calculate the concentration of test samples but a plot of the calibration data is not obtained;
- the residual standard deviation is used as an estimate of the uncertainty in predicted concentration values, rather than carrying out the full standard error of prediction calculation;
- the performance of the instrument used to make the measurements is not within specification.

Following the steps listed below should avoid these problems:

- plan the calibration study so that the concentration range of interest is covered and the concentrations of the calibration standards are evenly distributed across the range;
- include a standard with zero analyte concentration (*i.e.*, a blank);
- ensure that appropriate materials and apparatus are used to prepare the calibration standards;
- ensure that the instrument used to make the measurements is fit for purpose (*i.e.*, carry out equipment qualification);
- plot and examine the results;
- use validated software to perform the linear regression;

- do not set the intercept to zero unless there is evidence that the intercept is not statistically different from zero;
- plot and examine the residuals;
- calculate the uncertainty (prediction interval) for test sample concentrations predicted using the calibration line.

Appendix 1: Protocol and results sheet

Background

Participating organisations were supplied with a solid sample of a photographic chemical, Z, and a waste water sample, S1, containing Z at an unknown concentration. The participants were required to use compound Z to prepare a set of calibration solutions, construct a calibration curve and then use the curve to predict the concentration of Z in solution S1. Each analyst taking part was asked to repeat the exercise three times. Participants were given the option of calculating the standard error of prediction for one of their estimates of the concentration of S1. An Excel spreadsheet for entering the results of the study was supplied to each participant.

LABORATORY PROTOCOL

Construction of a Calibration Curve and the Determination of the Concentration of a Substance in Water by UV Analysis

The aim of this exercise is to investigate the uncertainties associated with the construction of a calibration curve, and with the determination of the concentration of an unknown solution using the calibration curve. In addition, by requesting participants to repeat the exercise in triplicate, it will be possible to evaluate the effect of any inhomogeneity in the material used to prepare the calibration standards.

Participants will be supplied with a solid sample of a photographic chemical, Z, and a waste water sample (S1) containing Z at an unknown concentration. Participants are required to use the solid sample to prepare calibration solutions, construct a calibration curve and then use the curve to determine the concentration of Z in solution S1.

Procedure

Test 1

Prepare a stock solution (solution A) with a concentration of 100 mg L^{-1} by weighing out 25 mg of the solid sample, transferring to a 250 mL volumetric flask and diluting to the mark with de-ionised water.

Through appropriate dilutions of solution A, prepare 7 calibration solutions as follows:

Concentration /mg L^{-1}	Dilution of solution A required
12.5	25 ml diluted to 200 ml
10	10 ml diluted to 100 ml
8 (a)	20 ml diluted to 250 ml
8 (b)	20 ml diluted to 250 ml
8 (c)	20 ml diluted to 250 ml
5	5 ml diluted to 100 ml
2.5	5 ml diluted to 200 ml

Calculate the actual concentrations of the calibration solutions, taking into account the amount of material used (W1). Measure the absorbance of each calibration solution at 306 nm and produce a plot of absorbance *vs*. actual concentration. Carry out a linear regression analysis to determine the equation of the relationship between absorbance and concentration (*i.e.*, $y = mx + c$).

Measure the absorbance of solution S1 at 306 nm. Use the calibration function to calculate the concentration of Z in S1 in mg L^{-1}.

Test 2

Repeat Test 1, except for the replication of the preparation of the 8 mg L^{-1} standard. Therefore, only 5 calibration standards are prepared as shown in the table:

Concentration /mg L^{-1}	Dilution of solution A required
12.5	25 ml diluted to 200 ml
10	10 ml diluted to 100 ml
8	20 ml diluted to 250 ml
5	5 ml diluted to 100 ml
2.5	5 ml diluted to 200 ml

Test 3

As Test 2.

Reporting results

For each test, please report the following information in the spreadsheet supplied:

Laboratory name
Analyst name
Date of analysis
Wavelength used
Cell path length
Amount of solid sample used to prepare the stock solution
Concentrations of the calibration solutions
Absorbance reading for each calibration solution
Correlation coefficient for the calibration curve
Equation of the calibration curve (*i.e.*, gradient and intercept of the line)
Absorbance reading for solution S1
Concentration of Z in solution S1
Further comments – please include a brief description of the equipment used, the temperature at which the measurements were made, the order in which the solutions were analysed and any deviations from the protocol.

In addition, please include a copy of the calibration graph from each test, and any additional statistical information relating to the determination of the equation of the calibration line (*e.g.*, output from software used to carry out the analysis, residual plots).

Optional

For at least one of the tests, use the standard error of prediction equation to calculate the standard error associated with the concentration determined for Z in solution S1.

Evaluation of Results

The results supplied by each analyst were evaluated against the following criteria:

- Carrying out the three tests as specified by the protocol, in particular:
  - preparation of three solutions with a concentration of 8 mg L^{-1} in test 1, not making repeat measurements on the same solution;
  - preparing independent sets of calibration solutions for each test;
  - measuring the absorbance of the solution S1 for each test.
- Correct calculation of the concentrations of the calibration standards;
- Correct regression analysis and correct reporting of the correlation coefficient, *r*, the gradient and intercept of the fitted line;
- Linearity of the calibration curve;
- Correct calculation of the concentration of S1;
- Correct calculation of the standard error of prediction if reported.

For each set of results, the mean and standard deviation of three estimates of the concentration of S1 was calculated, along with the standard deviation of the three absorbances reported for the 8 mg L^{-1} standards in test 1.

Excel Results Sheet (Test 1)

Lab name	
Analyst name	
Date of analysis	
Wavelength used	
Path length of cell	

Calibration Results

Amount of standard used W1 (mg)	

Target concentration (mg L^{-1})	Actual concentration (mg L^{-1})	Absorbance
2.5		
5		
8 (a)		
8 (b)		
8 (c)		
10		
12.5		

Correlation coefficient (*r*) for calibration curve	

	Gradient (m)	Intercept (c)
Equation of calibration curve ($y = mx + c$)		

Concentration of Z in solution S1

Absorbance	Concentration (mg L^{-1})	Standard error (mg L^{-1}) (optional)

Further comments (e.g., instrument used, type of cell, temperature, order of analysis, dilution scheme if different from protocol)

Appendix 2: Example set of results

Test 1

Lab name	
Analyst name	
Date of analysis	
Wavelength used	306 nm
Path length of cell	1.0 cm
Calibration Results	
Amount of standard used W1 (mg)	25.6 mg

Target concentration (mg L^{-1})	Actual concentration (mg L^{-1})	Absorbance
2.5	2.56	0.320
5	5.12	0.591
8 (a)	8.192	0.920
8 (b)	8.192	0.918
8 (c)	8.192	0.920
10	10.24	1.135
12.5	12.80	1.396

Correlation coefficient (*r*) for calibration curve	0.9999

	Gradient (m)	Intercept (c)
Equation of calibration curve (*y* = *mx* + *c*)	0.1054	0.0533

Concentration of Z in solution S1

Absorbance	Concentration (mg L^{-1})	Standard error (mg L^{-1}) (optional)
0.871	7.76	0.041

Further comments (e.g., instrument used, type of cell, temperature, order of analysis, dilution scheme if different from protocol)

Philips 8800 UV/VIS instrument. Absorbance recorded at 25°C

All volumetric solutions prepared at 20°C. Grade A volumetric glassware used.

Standard solutions analysed in random order.

Sample solution analysed last.

The same 1 cm glass cuvette used for all measurements and always placed in the cell holder the same way round.

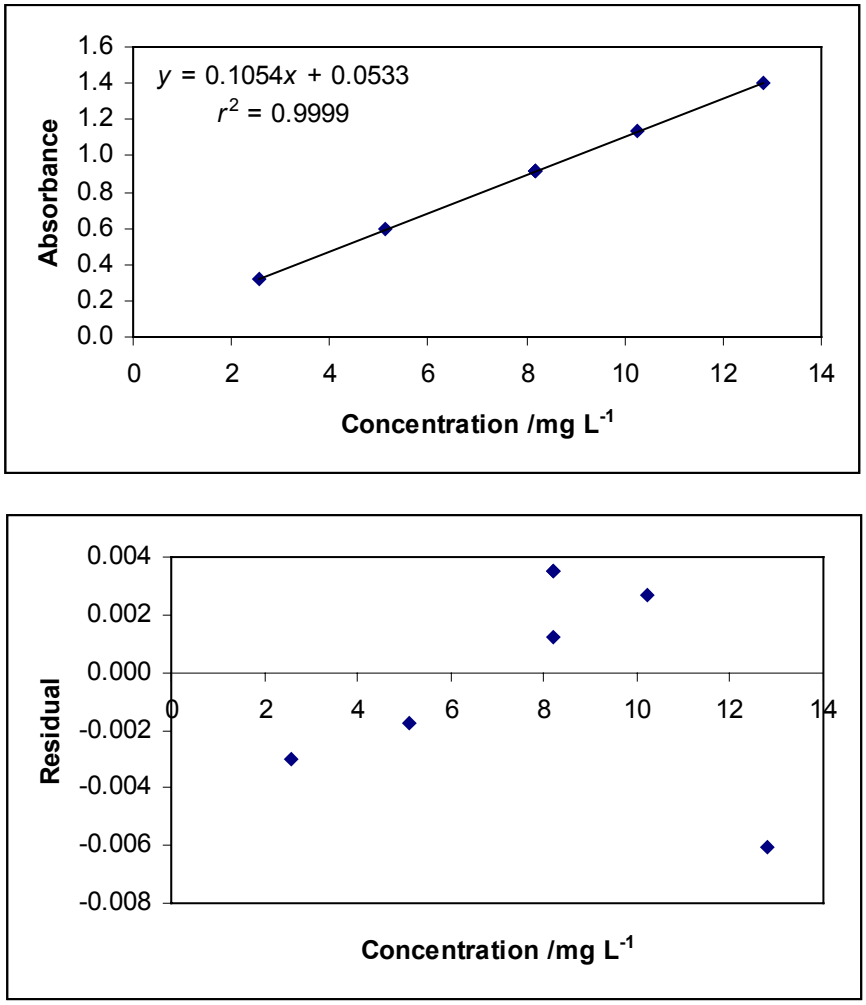


Figure 15: Scatter plot and residual plot for typical benchmarking data set (test 1)

Appendix 3: Linear regression equations

Parameter	Equation
Gradient of the least squares line, m	$m = \dfrac{\sum_{i=1}^{n}\{(x_i - \bar{x})(y_i - \bar{y})\}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$
Intercept of the least squares line, c	$c = \bar{y} - m\bar{x}$
Correlation coefficient, r	$r = \dfrac{\sum_{i=1}^{n}\{(x_i - \bar{x})(y_i - \bar{y})\}}{\left\{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]\right\}^{\frac{1}{2}}}$
Residual standard deviation, $s(r)$	$s(r) = \sqrt{\dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}}$
Standard deviation (error) of the gradient, s_m	$s_m = \dfrac{s(r)}{\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^{\frac{1}{2}}}$
Standard deviation (error) of the intercept, s_c	$s_c = s(r)\left\{\dfrac{\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}\right\}^{\frac{1}{2}}$
Confidence interval of the gradient, c_m	$c_m = ts_m$
Confidence interval of the intercept, c_c	$c_c = ts_c$
Standard deviation of the regression line, s_L	$s_L = s(r)\sqrt{\dfrac{1}{n} + \dfrac{(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$
Confidence interval for the regression line, c_L	$c_L = ts_L$
Prediction interval for predicted values of x, s_{x_0}	$s_{x_0} = \dfrac{s(r)}{m}\sqrt{\dfrac{1}{N} + \dfrac{1}{n} + \dfrac{(\bar{y}_o - \bar{y})^2}{m^2\sum_{i=1}^{n}(x_i - \bar{x})^2}}$
Confidence interval for predicted values of x, c_{x_0}	$c_{x_0} = ts_{x_0}$

x_i value on the x-axis

y_i observed value on the y-axis

$\bar{x}$ mean of x_i values

$\bar{y}$ mean of y_i values

n number of calibration points

N number of repeated measurements made on the test solution

$\bar{y}_0$ mean of N repeat measurements of y for the test solution

$\hat{y}_i$ predicted value of y for a given value x_i

t 2-tailed Student's t value for n-2 degrees of freedom