CHAPTER 4
ERRORS AND STATISTICS

4.1 LIMITATIONS OF ANALYTICAL METHODS

The function of the analyst is to obtain a result as near to the true value as possible by the correct application of the analytical procedure employed. The level of confidence that the analyst may enjoy in his results will be very small unless he has knowledge of the accuracy and precision of the method used as well as being aware of the sources of error which may be introduced. Quantitative analysis is not simply a case of taking a sample, carrying out a single determination and then claiming that the value obtained is irrefutable. It also requires a sound knowledge of the chemistry involved, of the possibilities of interferences from other ions, elements and compounds as well as of the statistical distribution of values. The purpose of this chapter is to explain some of the terms employed and to outline the statistical procedures which may be applied to the analytical results.

4.2 CLASSIFICATION OF ERRORS

The errors which affect an experimental result may conveniently be divided into 'systematic' and 'random' errors.

Systematic (determinate) errors. These are errors which can be avoided, or whose magnitude can be determined. The most important of them are:

1. *Operational and personal errors.* These are due to factors for which the individual analyst is responsible and are not connected with the method or procedure: they form part of the 'personal equation' of an observer. The errors are mostly physical in nature and occur when sound analytical technique is not followed. Examples are: mechanical loss of materials in various steps of an analysis; underwashing or overwashing of precipitates; ignition of precipitates at incorrect temperatures; insufficient cooling of crucibles before weighing; allowing hygroscopic materials to absorb moisture before or during weighing; and use of reagents containing harmful impurities.

   Personal errors may arise from the constitutional inability of an individual to make certain observations accurately. Thus some persons are unable to judge colour changes sharply in visual titrations, which may result in a slight overstepping of the end point.
2. *Instrumental and reagent errors.* These arise from the faulty construction of balances, the use of uncalibrated or improperly calibrated weights, graduated

glassware, and other instruments; the attack of reagents upon glassware, porcelain, etc., resulting in the introduction of foreign materials; volatilisation of platinum at very high temperatures; and the use of reagents containing impurities.

3. *Errors of method.* These originate from incorrect sampling and from incompleteness of a reaction. In gravimetric analysis errors may arise owing to appreciable solubility of precipitates, co-precipitation, and post-precipitation, decomposition, or volatilisation of weighing forms on ignition, and precipitation of substances other than the intended ones. In titrimetric analysis errors may occur owing to failure of reactions to proceed to completion, occurrence of induced and side reactions, reaction of substances other than the constituent being determined, and a difference between the observed end point and the stoichiometric equivalence point of a reaction.
4. *Additive and proportional errors.* The absolute value of an additive error is independent of the amount of the constituent present in the determination. Examples of additive errors are loss in weight of a crucible in which a precipitate is ignited, and errors in weights. The presence of this error is revealed by taking samples of different weights.

   The absolute value of a proportional error depends upon the amount of the constituent. Thus a proportional error may arise from an impurity in a standard substance, which leads to an incorrect value for the molarity of a standard solution. Other proportional errors may not vary linearly with the amount of the constituent, but will at least exhibit an increase with the amount of constituent present. One example is the ignition of aluminium oxide: at 1200 °C the aluminium oxide is anhydrous and virtually non-hygroscopic; ignition of various weights at an appreciably lower temperature will show a proportional type of error.

Random (indeterminate) errors. These errors manifest themselves by the slight variations that occur in successive measurements made by the same observer with the greatest care under as nearly identical conditions as possible. They are due to causes over which the analyst has no control, and which, in general, are so intangible that they are incapable of analysis. If a *sufficiently large number of observations* is taken it can be shown that these errors lie on a curve of the form shown in Fig. 4.1 (Section 4.9). An inspection of this error curve shows: (*a*) small errors occur more frequently than large ones; and (*b*) positive and negative errors of the same numerical magnitude are equally likely to occur.

4.3 ACCURACY

The accuracy of a determination may be defined as the concordance between it and the true or most probable value. It follows, therefore, that systematic errors cause a constant error (either too high or too low) and thus affect the accuracy of a result. For analytical methods there are two possible ways of determining the accuracy; the so-called absolute method and the comparative method.

Absolute method. A synthetic sample containing known amounts of the constituents in question is used. Known amounts of a constituent can be obtained by weighing out pure elements or compounds of known stoichiometric composition. These substances, primary standards, may be available commercially

or they may be prepared by the analyst and subjected to rigorous purification by recrystallisation, etc. The substances must be of known purity. The test of the accuracy of the method under consideration is carried out by taking varying amounts of the constituent and proceeding according to specified instructions. The amount of the constituent must be varied, because the determinate errors in the procedure may be a function of the amount used. The difference between the mean of an adequate number of results and the amount of the constituent actually present, usually expressed as parts per thousand, is a measure of the accuracy of the method in the absence of foreign substances.

The constituent in question will usually have to be determined in the presence of other substances, and it will therefore be necessary to know the effect of these upon the determination. This will require testing the influence of a large number of elements, each in varying amounts — a major undertaking. The scope of such tests may be limited by considering the determination of the component in a specified range of concentration in a material whose composition is more or less fixed both with respect to the elements which may be present and their relative amounts. It is desirable, however, to study the effect of as many foreign elements as feasible. In practice, it is frequently found that separations will be required before a determination can be made in the presence of varying elements; the accuracy of the method is likely to be largely controlled by the separations involved.

Comparative method. Sometimes, as in the analysis of a mineral, it may be impossible to prepare solid synthetic samples of the desired composition. It is then necessary to resort to standard samples of the material in question (mineral, ore, alloy, etc.) in which the content of the constituent sought has been determined by one or more supposedly 'accurate' methods of analysis. This comparative method, involving secondary standards, is obviously not altogether satisfactory from the theoretical standpoint, but is nevertheless very useful in applied analysis. Standard samples can be obtained from various sources (see Section 4.5).

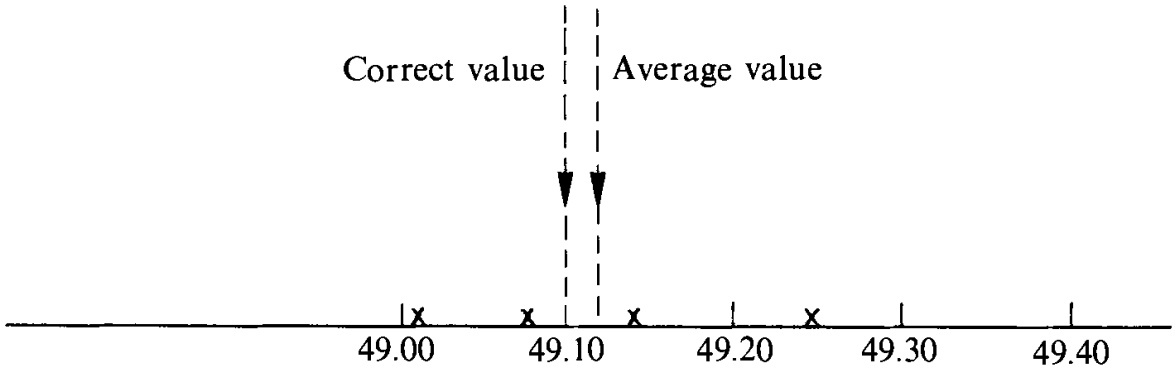
If several fundamentally different methods of analysis for a given constituent are available, e.g. gravimetric, titrimetric, spectrophotometric, or spectrographic, the agreement between at least two methods of essentially different character can usually be accepted as indicating the absence of an appreciable systematic error in either (a systematic error is one which can be evaluated experimentally or theoretically).

4.4 PRECISION

Precision may be defined as the concordance of a series of measurements of the same quantity. Accuracy expresses the correctness of a measurement, and precision the 'reproducibility' of a measurement (the latter definition will be modified later). Precision always accompanies accuracy, but a high degree of precision does not imply accuracy. This may be illustrated by the following example.

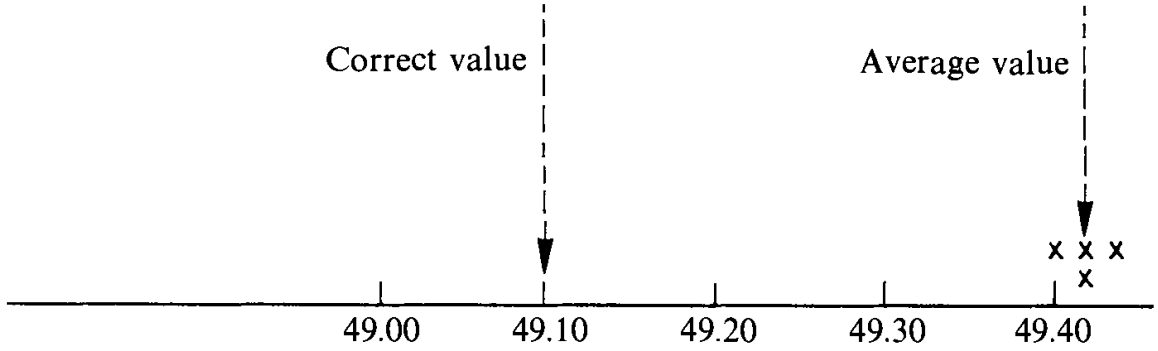
A substance was known to contain 49.10 ± 0.02 per cent of a constituent A. The results obtained by two analysts using the same substance and the same analytical method were as follows.

Analyst (1) %A 49.01; 49.25; 49.08; 49.14



The arithmetic mean is 49.12% and the results range from 49.01% to 49.25%.

Analyst (2) %A 49.40; 49.44; 49.42; 49.42



The arithmetic mean is 49.42% and the results range from 49.40% to 49.44%.

We can summarise the results of the analyses as follows.

(*a*) The values obtained by Analyst 1 are accurate (very close to the correct result), but the precision is inferior to the results given by Analyst 2. The values obtained by Analyst 2 are very precise but are not accurate.

(*b*) The results of Analyst 1 face on both sides of the mean value and could be attributed to random errors. It is apparent that there is a constant (systematic) error present in the results of Analyst 2.

Precision was previously described as the reproducibility of a measurement. However, the modern analyst makes a distinction between the terms '**reproducible**' and '**repeatable**'. On further consideration of the above example:

(*c*) If Analyst 2 had made the determinations on the same day in rapid succession, then this would be defined as 'repeatable' analysis. However, if the determinations had been made on separate days when laboratory conditions may vary, this set of results would be defined as '**reproducible**'.

Thus, there is a distinction between a within-run precision (repeatability) and a between-run precision (reproducibility).

4.5 MINIMISATION OF ERRORS

Systematic errors can often be materially reduced by one of the following methods.

1. *Calibration of apparatus and application of corrections.* All instruments (weights, flasks, burettes, pipettes, etc.) should be calibrated, and the appropriate corrections applied to the original measurements. In some cases where an error cannot be eliminated, it is possible to apply a correction for the effect that it produces; thus an impurity in a weighed precipitate may be determined and its weight deducted.
2. *Running a blank determination.* This consists in carrying out a separate determination, the sample being omitted, under exactly the same experimental conditions as are employed in the actual analysis of the sample. The object is to find out the effect of the impurities introduced through the reagents and vessels, or to determine the excess of standard solution necessary to establish the end-point under the conditions met with in the titration of the unknown sample. A large blank correction is undesirable, because the exact value then becomes uncertain and the precision of the analysis is reduced.
3. *Running a control determination.* This consists in carrying out a determination under as nearly as possible identical experimental conditions upon a quantity of a standard substance which contains the same weight of the constituent as is contained in the unknown sample. The weight of the constituent in the unknown can then be calculated from the relation:

$$\frac{\text{Result found for standard}}{\text{Result found for unknown}} = \frac{\text{Weight of constituent in standard}}{x}$$

where x is the weight of the constituent in the unknown.

In this connection it must be pointed out that standard samples which have been analysed by a number of skilled analysts are commercially available. These include certain primary standards (sodium oxalate, potassium hydrogenphthalate, arsenic(III) oxide, and benzoic acid) and ores, ceramic materials, irons, steels, steel-making alloys, and non-ferrous alloys.

Many of these are also available as BCS Certified Reference Materials (CRM) supplied by the Bureau of Analysed Samples Ltd, Newham Hall, Middlesborough, UK, who also supply EURONORM Certified Reference Materials (ERCM), the composition of which is specified on the basis of results obtained by a number of laboratories within the EEC. BCS Reference Materials are obtainable from the Community Bureau of Reference, Brussels, Belgium. In the USA similar reference materials are supplied by the National Bureau of Standards.

4. *Use of independent methods of analysis.* In some instances the accuracy of a result may be established by carrying out the analysis in an entirely different manner. Thus iron may first be determined gravimetrically by precipitation as iron(III) hydroxide after removing the interfering elements, followed by ignition of the precipitate to iron(III) oxide. It may then be determined titrimetrically by reduction to the iron(II) state, and titration with a standard solution of an oxidising agent, such as potassium dichromate or cerium(IV) sulphate. Another example that may be mentioned is the determination of the strength of a hydrochloric acid solution both by titration with a standard

solution of a strong base and by precipitation and weighing as silver chloride. If the results obtained by the two radically different methods are concordant, it is highly probable that the values are correct within small limits of error.

5. *Running parallel determinations.* These serve as a check on the result of a single determination and indicate only the precision of the analysis. The values obtained for constituents which are present in not too small an amount should not vary among themselves by more than three parts per thousand. If larger variations are shown, the determinations must be repeated until satisfactory concordance is obtained. Duplicate, and at most triplicate, determinations should suffice. It must be emphasised that good agreement between duplicate and triplicate determinations does not justify the conclusion that the result is correct; a constant error may be present. The agreement merely shows that the accidental errors, or variations of the determinate errors, are the same, or nearly the same, in the parallel determinations.
6. *Standard addition.* A known amount of the constituent being determined is added to the sample, which is then analysed for the total amount of constituent present. The difference between the analytical results for samples with and without the added constituent gives the recovery of the amount of added constituent. If the recovery is satisfactory our confidence in the accuracy of the procedure is enhanced. The method is usually applied to physico-chemical procedures such as polarography and spectrophotometry.
7. *Internal standards.* This procedure is of particular value in spectroscopic and chromatographic determinations. It involves adding a fixed amount of a reference material (the internal standard) to a series of known concentrations of the material to be measured. The ratios of the physical value (absorption or peak size) of the internal standard and the series of known concentrations are plotted against the concentration values. This should give a straight line. Any unknown concentration can then be determined by adding the same quantity of internal standard and finding where the ratio obtained falls on the concentration scale.
8. *Amplification methods.* In determinations in which a very small amount of material is to be measured this may be beyond the limits of the apparatus available. In these circumstances if the small amount of material can be reacted in such a way that every molecule produces two or more molecules of some other measurable material, the resultant amplification may then bring the quantity to be determined within the scope of the apparatus or method available.
9. *Isotopic dilution.* A known amount of the element being determined, containing a radioactive isotope, is mixed with the sample and the element is isolated in a pure form (usually as a compound), which is weighed or otherwise determined. The radioactivity of the isolated material is measured and compared with that of the added element: the weight of the element in the sample can then be calculated.

4.6 SIGNIFICANT FIGURES AND COMPUTATIONS

The term 'digit' denotes any one of the ten numerals, including the zero. A significant figure is a digit which denotes the amount of the quantity in the place in which it stands. The digit zero is a significant figure except when it is the first figure in a number. Thus in the quantities 1.2680 g and 1.0062 g the

zero is significant, but in the quantity 0.0025 kg the zeros are not significant figures; they serve only to locate the decimal point and can be omitted by proper choice of units, i.e. 2.5 g. The first two numbers contain five significant figures, but 0.0025 contains only two significant figures.

Observed quantities should be recorded with one uncertain figure retained. Thus in most analyses weights are determined to the nearest tenth of a milligram, e.g. 2.1546 g. This means that the weight is less than 2.1547 g and more than 2.1545 g. A weight of 2.150 g would signify that it has been determined to the nearest milligram, and that the weight is nearer to 2.150 g than it is to either 2.151 g or 2.149 g. The digits of a number which are needed to express the precision of the measurement from which the number was derived are known as significant figures.

There are a number of rules for computations with which the student should be familiar.

1. Retain as many significant figures in a result or in any data as will give only one uncertain figure. Thus a volume which is known to be between 20.5 mL and 20.7 mL should be written as 20.6 mL, but not as 20.60 mL, since the latter would indicate that the value lies between 20.59 mL and 20.61 mL. Also, if a weight, to the nearest 0.1 mg, is 5.2600 g, it should not be written as 5.260 g or 5.26 g, since in the latter case an accuracy of a centigram is indicated and in the former a milligram.
2. In rounding off quantities to the correct number of significant figures, add one to the last figure retained if the following figure (which has been rejected) is 5 or over. Thus the average of 0.2628, 0.2623, and 0.2626 is 0.2626 (0.2625_7).
3. In addition or subtraction, there should be in each number only as many significant figures as there are in the least accurately known number. Thus the addition

   $168.11 + 7.045 + 0.6832$

   should be written

   $168.11 + 7.05 + 0.68 = 175.84$

   The sum or difference of two or more quantities cannot be more precise than the quantity having the largest uncertainty.
4. In multiplication or division, retain in each factor one more significant figure than is contained in the factor having the largest uncertainty. The percentage precision of a product or quotient cannot be greater than the percentage precision of the least precise factor entering into the calculation. Thus the multiplication

   $1.26 \times 1.236 \times 0.6834 \times 24.8652$

   should be carried out using the values

   $1.26 \times 1.236 \times 0.683 \times 24.87$

   and the result expressed to three significant figures.

4.7 THE USE OF CALCULATORS AND MICROCOMPUTERS

The advent of reasonably priced hand-held calculators has replaced the use of both logarithms and slide-rules for statistical calculations. In addition to the

normal arithmetic functions, a suitable calculator for statistical work should enable the user to evaluate the mean and standard deviation (Section 4.8), linear regression and correlation coefficient (Section 4.16). The results obtained by the use of the calculator must be carefully scrutinised to ascertain the number of significant figures to be retained, and should always be checked against a 'rough' arithmetical calculation to ensure there are no gross computational errors. Microcomputers are used for processing large amounts of data. Although computer programming is outside the scope of this book it should be pointed out that standard programs now exist in BASIC, and other high-level computer languages (see Bibliography, Section 5.7).

The microcomputer may also be interfaced with most types of electronic equipment used in the laboratory. This facilitates the collection and processing of the data, which may be stored on floppy or hard discs for later use.

There is a large amount of commercial software available for performing the statistical calculations described later in this chapter, and for more advanced statistical tests beyond the scope of this text.

4.8 MEAN AND STANDARD DEVIATION

When a quantity is measured with the greatest exactness of which the instrument, method, and observer are capable, it is found that the results of successive determinations differ among themselves to a greater or lesser extent. The average value is accepted as the most probable. This may not always be the true value. In some cases the difference may be small, in others it may be large; the reliability of the result depends upon the magnitude of this difference. It is therefore of interest to enquire briefly into the factors which affect and control the trustworthiness of chemical analysis.

The absolute error of a determination is the difference between the observed or measured value and the true value of the quantity measured. It is a measure of the accuracy of the measurement.

The relative error is the absolute error divided by the true value; it is usually expressed in terms of percentage or in parts per thousand. The true or absolute value of a quantity cannot be established experimentally, so that the observed result must be compared with the most probable value. With pure substances the quantity will ultimately depend upon the relative atomic mass of the constituent elements. Determinations of the relative atomic mass have been made with the utmost care, and the accuracy obtained usually far exceeds that attained in ordinary quantitative analysis; the analyst must accordingly accept their reliability. With natural or industrial products, we must accept provisionally the results obtained by analysts of repute using carefully tested methods. If several analysts determine the same constituent in the same sample by different methods, the most probable value, which is usually the average, can be deduced from their results. In both cases, the establishment of the most probable value involves the application of statistical methods and the concept of precision.

In analytical chemistry one of the most common statistical terms employed is the **standard deviation** of a population of observations. This is also called the root mean square deviation as it is the square root of the mean of the sum of the squares of the differences between the values and the mean of those values (this is expressed mathematically below) and is of particular value in connection with the normal distribution.

If we consider a series of n observations arranged in ascending order of magnitude:

$$x_1, x_2, x_3, \ldots, x_{n-1}, x_n,$$

the arithmetic mean (often called simply the mean) is given by:

$$\bar{x} = \frac{x_1 + x_2 + x_3 \ldots + \ldots + x_{n-1} + x_n}{n}$$

The spread of the values is measured most efficiently by the standard deviations defined by:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots (x_n - \bar{x})^2}{n - 1}}$$

In this equation the denominator is $(n - 1)$ rather than n when the number of values is small.

The equation may also be written as:

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

The square of the standard deviation is called the variance. A further measure of precision, known as the Relative Standard Deviation (R.S.D.), is given by:

$$\text{R.S.D.} = \frac{s}{\bar{x}}$$

This measure is often expressed as a percentage, known as the **coefficient of variation** (C.V.):

$$\text{C.V.} = \frac{s \times 100}{\bar{x}}$$

Example 1. Analyses of a sample of iron ore gave the following percentage values for the iron content: 7.08, 7.21, 7.12, 7.09, 7.16, 7.14, 7.07, 7.14, 7.18, 7.11. Calculate the mean, standard deviation and coefficient of variation for the values.

Results (x)	$x - \bar{x}$	$(x - \bar{x})^2$
7.08	−0.05	0.0025
7.21	0.08	0.0064
7.12	−0.01	0.0001
7.09	−0.04	0.0016
7.16	0.03	0.0009
7.14	0.01	0.0001
7.07	−0.06	0.0036
7.14	0.01	0.0001
7.18	0.05	0.0025
7.11	−0.02	0.0004
$\Sigma x = 71.30$		$\Sigma(x - \bar{x})^2 = 0.0182$
Mean $\bar{x}$ 7.13 per cent		

$$s = \sqrt{\frac{0.0182}{9}}$$

$$= \sqrt{0.0020}$$

$$= \pm 0.045 \text{ per cent}$$

$$\text{C. V.} = \frac{0.045 \times 100}{7.13} = 0.63 \text{ per cent}$$

The mean of several readings ($\bar{x}$) will make a more reliable estimate of the true mean (μ) than is given by one observation. The greater the number of measurements (n), the closer will the sample average approach the true mean. The standard error of the mean s_x is given by:

$$s_x = \frac{s}{\sqrt{n}}$$

In the above example,

$$s_x = \pm \frac{0.045}{\sqrt{10}} = \pm 0.014$$

and if 100 measurements were made,

$$s_x = \pm \frac{0.045}{\sqrt{100}} = \pm 0.0045$$

Hence the *precision* of a measurement may be improved by increasing the number of measurements.

4.9 DISTRIBUTION OF RANDOM ERRORS

In the previous section (4.8) it has been shown that the spread of a series of results obtained from a given set of measurements can be ascertained from the value of the standard deviation. However, this term gives no indication as to the manner in which the results are distributed.

If a large number of replicate readings, at least 50, are taken of a continuous variable, e.g. a titrimetric end-point, the results attained will usually be distributed about the mean in a roughly symmetrical manner. The mathematical model that best satisfies such a distribution of random errors is called the Normal (or Gaussian) distribution. This is a bell-shaped curve that is symmetrical about the mean as shown in Fig. 4.1.

The curve satisfies the equation:

$$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma}}$$

It is important to know that the Greek letters σ and μ refer to the standard deviation and mean respectively of a total population, whilst the Roman letters s and $\bar{x}$ are used for samples of populations, irrespective of the values of the population mean and the population standard deviation.

With this type of distribution about 68 per cent of all values will fall within

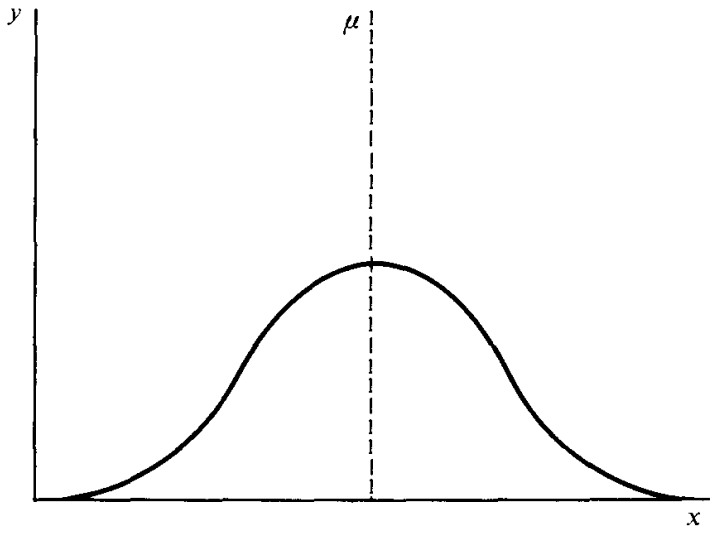


Fig. 4.1

one standard deviation on either side of the mean, 95 per cent will fall within two standard deviations, and 99.7 per cent within three standard deviations.

From the worked example (Example 1 in Section 4.8) for the analysis of an iron ore sample, the standard deviation is found to be ± 0.045 per cent. If the assumption is made that the results are normally distributed, then 68 per cent (approximately seven out of ten results) will be between ± 0.045 per cent and 95 per cent will be between ± 0.090 per cent of the mean value. It follows that there will be a 5 per cent probability (1 in 20 chance) of a result differing from the mean by more than ± 0.090 per cent, and a 1 in 40 chance of the result being 0.090 per cent *higher* than the mean.

4.10 RELIABILITY OF RESULTS

Statistical figures obtained from a set of results are of limited value by themselves. Analysis of the results can be considered in two main categories: (*a*) the reliability of the results; and (*b*) comparison of the results with the true value or with other sets of data (Section 4.12).

A most important consideration is to be able to arrive at a sensible decision as to whether certain results may be rejected. It must be stressed that values should be rejected only when a suitable statistical test has been applied or when there is an obvious chemical or instrumental reason that could justify exclusion of a result. Too frequently, however, there is a strong temptation to remove what may appear to be a 'bad' result without any sound justification. Consider the following example.

Example 2. The following values were obtained for the determination of cadmium in a sample of dust: 4.3, 4.1, 4.0, 3.2 $\mu g\ g^{-1}$. Should the last value, 3.2, be rejected?

The Q test may be applied to solve this problem.

$$Q = \frac{|\text{Questionable value} - \text{Nearest value}|}{\text{Largest value} - \text{Smallest value}}$$

$$Q = \frac{|3.2 - 4.0|}{4.3 - 3.2} = \frac{0.8}{1.1} = 0.727$$

If the calculated value of Q exceeds the critical value given in the Q table (Appendix 14), then the questionable value may be rejected.

In this example Q calculated is 0.727 and Q critical, for a sample size of four, is 0.831. Hence the result 3.2 $\mu g\, g^{-1}$ should be retained. If, however, in the above example, three additional measurements were made, with the results:

4.3, 4.1, 4.0, 3.2, 4.2, 3.9, 4.0 $\mu g\, g^{-1}$

Then

$$Q = \frac{|3.2 - 3.9|}{4.3 - 3.2} = \frac{0.7}{1.1} = 0.636$$

The value of Q critical for a sample size of seven is 0.570, so rejection of the value 3.2 $\mu g\, g^{-1}$ is justified.

It should be noted that the value Q has no regard to algebraic sign.

4.11 CONFIDENCE INTERVAL

When a small number of observations is made, the value of the standard deviation s, does not by itself give a measure of how close the sample mean $\bar{x}$ might be to the true mean. It is, however, possible to calculate a confidence interval to estimate the range within which the true mean may be found. The limits of this confidence interval, known as the confidence limits, are given by the expression:

$$\text{Confidence limits of } \mu, \text{ for } n \text{ replicate measurements, } \mu = \bar{x} \pm \frac{ts}{\sqrt{n}} \qquad (1)$$

where t is a parameter that depends upon the number of degrees of freedom (v) (Section 4.12) and the confidence level required. A table of the values of t at different confidence levels and degrees of freedom (v) is given in Appendix 12.

Example 3. The mean ($\bar{x}$) of four determinations of the copper content of a sample of an alloy was 8.27 per cent with a standard deviation $s = 0.17$ per cent. Calculate the 95% confidence limit for the true value.

From the t-tables, the value of t for the 95 per cent confidence level with $(n - 1)$, i.e. three degrees of freedom, is 3.18.

Hence from equation (1), the 95 per cent confidence level,

$$95\%\text{(C.L.) for } \mu = 8.27 \pm \frac{3.18 \times 0.17}{\sqrt{4}}$$

$$= 8.27 \pm 0.27 \text{ per cent}$$

Thus, there is 95 per cent confidence that the true value of the copper content of the alloy lies in the range 8.00 to 8.54 per cent.

If the number of determinations in the above example had been 12, then the reader may wish to confirm that

$$95\%\text{(C.L.) for } \mu = 8.27 \pm \frac{2.20 \times 0.17}{\sqrt{12}}$$

$$= 8.27 \pm 0.11 \text{ per cent}$$

Hence, on increasing the number of replicate determinations both the values of t and $s/\sqrt{n}$ decrease with the result that the confidence interval is smaller. There is, however, often a limit to the number of replicate analyses that can be sensibly performed. A method for estimating the optimum number of replicate determinations is given in Section 4.15.

4.12 COMPARISON OF RESULTS

The comparison of the values obtained from a set of results with either (*a*) the true value or (*b*) other sets of data makes it possible to determine whether the analytical procedure has been accurate and/or precise, or if it is superior to another method.

There are two common methods for comparing results: (*a*) Student's *t*-test and (*b*) the variance ratio test (*F*-test).

These methods of test require a knowledge of what is known as the number of degrees of freedom. In statistical terms this is the number of independent values necessary to determine the statistical quantity. Thus a sample of n values has n degrees of freedom, whilst the sum $\Sigma(x - \bar{x})^2$ is considered to have $n - 1$ degrees of freedom, as for any defined value of $\bar{x}$ only $n - 1$ values can be freely assigned, the nth being automatically defined from the other values.

(*a*) Student's *t*-test. This is a test[1] used for small samples; its purpose is to compare the mean from a sample with some standard value and to express some level of confidence in the significance of the comparison. It is also used to test the difference between the means of two sets of data $\bar{x}_1$ and $\bar{x}_2$.

The value of t is obtained from the equation:

$$t = \frac{(\bar{x} - \mu)\sqrt{n}}{s} \qquad (2)$$

where μ is the true value.

It is then related to a set of *t*-tables (Appendix 12) in which the probability (P) of the *t*-value falling within certain limits is expressed, either as a percentage or as a function of unity, relative to the number of degrees of freedom.

Example 4. *t*-Test when the true mean is known.

If $\bar{x}$ the mean of the 12 determinations = 8.37, and μ the true value = 7.91, say whether or not this result is significant if the standard deviation is 0.17.

From equation (2)

$$t = \frac{(8.37 - 7.91)\sqrt{12}}{0.17} = 9.4$$

From *t*-tables for *eleven* degrees of freedom (one less than those used in the calculation)

for P =	0.10 (10 per cent)	0.05 (5 per cent)	0.01 (1 per cent)
t =	1.80	2.20	3.11

and as the calculated value for t is 9.4 the result is highly significant. The *t*-table tells us that the probability of obtaining the difference of 0.46 between the experimental and true result is less than 1 in 100. This implies that some particular bias exists in the laboratory procedure.

Had the calculated value for t been less than 1.80 then there would have been no significance in the results and no apparent bias in the laboratory procedure, as the tables would have indicated a probability of greater than 1 in 10 of obtaining that value. It should be pointed out that these values refer to what is known as a double-sided, or two-tailed, distribution because it concerns probabilities of values both less and greater than the mean. In some calculations an analyst may only be interested in one of these two cases, and under these conditions the t-test becomes single-tailed so that the probability from the tables is halved.

***(b)* F-test.** This is used to compare the precisions of two sets of data,[2] for example, the results of two different analytical methods or the results from two different laboratories. It is calculated from the equation:

$$F = \frac{s_A^2}{s_B^2} \tag{3}$$

N.B. The larger value of s is always used as the numerator so that the value of F is always greater than unity. The value obtained for F is then checked for its significance against values in the F-table calculated from an F-distribution (Appendix 13) corresponding to the numbers of degrees of freedom for the two sets of data.

Example 5. F-test comparison of precisions.

The standard deviation from one set of 11 determinations was $s_A = 0.210$, and the standard deviation from another 13 determinations was $s_B = 0.641$. Is there any significant difference between the precision of these two sets of results?

From equation (3)

$$F = \frac{(0.641)^2}{(0.210)^2} = \frac{0.411}{0.044} = 9.4$$

for

$P = 0.10 \quad 0.05 \quad 0.01$

$F = 2.28 \quad 2.91 \quad 4.71$

The first value (2.28) corresponds to 10 per cent probability, the second value (2.91) to 5 per cent probability and the third value (4.71) to 1 per cent probability.

Under these conditions there is less than one chance in 100 that these precisions are similar. To put it another way, the difference between the two sets of data is highly significant.

Had the value of F turned out to be less than 2.28 then it would have been possible to say that there was no significant difference between the precisions, at the 10 per cent level.

4.13 COMPARISON OF THE MEANS OF TWO SAMPLES

When a new analytical method is being developed it is usual practice to compare the values of the mean and precision of the new (test) method with those of an established (reference) procedure.

The value of t when comparing two sample means $\bar{x}_1$ and $\bar{x}_2$ is given by the

expression:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}} \tag{4}$$

where s_p the pooled standard deviation, is calculated from the two sample standard deviations s_1 and s_2, as follows:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{5}$$

It should be stressed that there must not be a significant difference between the precisions of the methods. Hence the F-test (Section 4.12) is applied prior to using the t-test in equation (5).

Example 6. Comparison of two sets of data.

The following results were obtained in a comparison between a new method and standard method for the determination of the percentage nickel in a special steel.

	New method	Standard method
Mean	$\bar{x}_1 = 7.85$ per cent	$\bar{x}_2 = 8.03$ per cent
Standard deviation	$s_1 = \pm 0.130$ per cent	$s_2 = \pm 0.095$ per cent
Number of samples	$n_1 = 5$	$n_2 = 6$

Test at the 5 per cent probability value if the new method mean is significantly different from the standard reference mean.

The F-test must be applied to establish that there is no significant difference between the precisions of the two methods.

$$F = \frac{s_A^2}{s_B^2} = \frac{(0.130)^2}{(0.095)^2} = 1.87$$

The F-value ($P = 5$ per cent) from the tables (Appendix 13) for four and five degrees of freedom respectively for s_A and $s_B = 5.19$.

Thus, the calculated value of F (1.87) is less than the tabulated value; therefore the methods have comparable precisions (standard deviations) and so the t-test can be used with confidence.

From equation (5) the pooled standard deviation s_p is given by:

$$s_p = \sqrt{\frac{(5 - 1) \times 0.0169 + (6 - 1) \times 0.0090}{9}} = \pm 0.112$$

and from equation (4)

$$t = \frac{7.85 - 8.03}{0.112\sqrt{1/5 + 1/6}} = \frac{0.18}{0.112 \times 0.605} = 2.66$$

At the 5 per cent level, the tabulated value of t for $(n_1 + n_2 - 2)$, i.e. nine degrees of freedom, is 2.26.

Since $t_{\text{calculated}} 2.66 > t_{\text{tabulated}} 2.26$, there is a significant difference, at the specified probability, between the mean results of the two methods.

4.14 PAIRED *t*-TEST

Another method of validating a new procedure is to compare the results using samples of varying composition with the values obtained by an accepted method.

The manner of performing this calculation is best illustrated by an example:

Example 7. The *t*-test using samples of differing composition (the paired *t*-test).

Two different methods, A and B, were used for the analysis of five different iron compounds.

Sample	1	2	3	4	5
Method A	17.6	6.8	14.2	20.5	9.7 per cent Fe
Method B	17.9	7.1	13.8	20.3	10.2 per cent Fe

It should be apparent that in this example it would not be correct to attempt the calculation by the method described previously (Section 4.13).

In this case the differences (d) between each pair of results are calculated and $\bar{d}$, the mean of the difference, is obtained. The standard deviation s_d of the differences is then evaluated. The results are tabulated as follows.

Method A	Method B	d	$d-\bar{d}$	$(d-\bar{d})^2$
17.6	17.9	+0.3	0.2	0.04
6.8	7.1	+0.3	0.2	0.04
14.2	13.8	−0.4	0.5	0.25
20.5	20.3	−0.2	−0.3	0.09
9.7	10.2	+0.5	0.4	0.16
		$\sum d = 0.5$		$\sum(d-\bar{d})^2 = 0.58$
		$\therefore \bar{d} = 0.1$		

$$s_d = \sqrt{\frac{0.58}{4}} = \pm 0.38$$

Then t is calculated from the equation

$$t = \frac{\bar{d}\sqrt{n}}{s_d} = \frac{0.10\sqrt{5}}{0.38} = 0.58_9$$

The tabulated value of t is 2.78 ($P = 0.05$) and since the calculated value is less than this, there is no significant difference between the methods.

4.15 THE NUMBER OF REPLICATE DETERMINATIONS

To avoid unnecessary time and expenditure, an analyst needs some guide to the number of repetitive determinations needed to obtain a suitably reliable result from the determinations performed. The larger the number the greater the reliability, but at the same time after a certain number of determinations any improvement in precision and accuracy is very small.

Although rather involved statistical methods exist for establishing the number

of parallel determinations, a reasonably good assessment can be made by establishing the variation of the value for the absolute error Δ obtained for an increasing number of determinations.

$$\Delta = \frac{ts}{n}$$

The value for t is taken from the 95 per cent confidence limit column of the t-tables for $n - 1$ degrees of freedom.

The values for Δ are used to calculate the reliability interval L from the equation:

$$L = \frac{100\Delta}{z} \text{ per cent}$$

where z is the approximate percentage level of the unknown being determined. The number of replicate analyses is assessed from the magnitude of the change in L with the number of determinations.

Example 8. Ascertain the number of replicate analyses desirable (*a*) for the determination of approximately 2 per cent Cl^- in a material if the standard deviation for determinations is 0.051, (*b*) for approximately 20 per cent Cl^- if the standard deviation of determinations is 0.093.

(*a*) For 2 per cent Cl^-:

Number of determinations	$\Delta = \frac{ts}{n}$	$L = \frac{100\Delta}{z}$	Difference (per cent)
2	$12.7 \times 0.051 \times 0.71 = 0.4599$	22.99	
3	$4.3 \times 0.051 \times 0.58 = 0.1272$	6.36	16.63
4	$3.2 \times 0.051 \times 0.50 = 0.0816$	4.08	2.28
5	$2.8 \times 0.051 \times 0.45 = 0.0642$	3.21	0.87
6	$2.6 \times 0.051 \times 0.41 = 0.0544$	2.72	0.49

(*b*) For 20 per cent Cl^-:

Number of determinations	$\Delta = \frac{ts}{n}$	$L = \frac{100\Delta}{z}$	Difference (per cent)
2	$12.7 \times 0.093 \times 0.71 = 0.838$	4.19	
3	$4.3 \times 0.093 \times 0.58 = 0.232$	1.16	3.03
4	$3.2 \times 0.093 \times 0.50 = 0.148$	0.74	0.42
5	$2.8 \times 0.093 \times 0.45 = 0.117$	0.59	0.15
6	$2.6 \times 0.093 \times 0.41 = 0.099$	0.49	0.10

In (*a*) the reliability interval is greatly improved by carrying out a third analysis. This is less the case with (*b*) as the reliability interval is already narrow. In this second case no substantial improvement is gained by carrying out more than two analyses.

This subject is dealt with in more detail by Eckschlager,[4] and Shewell[5] has discussed other factors which influence the value of parallel determinations.

4.16 CORRELATION AND REGRESSION

When using instrumental methods it is often necessary to carry out a calibration procedure by using a series of samples (standards) each having a *known* concentration of the analyte to be determined. A **calibration curve** is constructed by measuring the instrumental signal for each standard and plotting this response against concentration (See Sections 17.14 and 17.21). Provided the same experimental conditions are used for the measurement of the standards and for the test (unknown) sample, the concentration of the latter may be determined from the calibration curve by graphical interpolation.

There are two statistical tests that should be applied to a calibration curve:

(*a*) to ascertain if the graph is linear, or in the form of a curve;
(*b*) to evaluate the best straight line (or curve) throughout the data points.

Correlation coefficient. In order to establish whether there is a linear relationship between two variables x_1 and y_1 the Pearson's correlation coefficient r is used.

$$r = \frac{n\Sigma x_1 y_1 - \Sigma x_1 \Sigma y_1}{\sqrt{[n\Sigma x_1^2 - (\Sigma x_1)^2][n\Sigma y_1^2 - (\Sigma y_1)^2]}} \tag{6}$$

where n is the number of data points.

The value of r must lie between $+1$ and -1: the nearer it is to $+1$, or in the case of negative correlation to -1, then the greater the probability that a definite linear relationship exists between the variables x and y. Values of r that tend towards zero indicate that x and y are not linearly related (they may be related in a non-linear fashion).

Although the correlation coefficient r would easily be calculated with the aid of a modern calculator or computer package, the following example will show how the value of r can be obtained.

Example 9. Quinine may be determined by measuring the fluorescence intensity in $1M$ H_2SO_4 solution (Section 18.4). Standard solutions of quinine gave the following fluorescence values. Calculate the correlation coefficient r.

Concentration of quinine (x_1)	0.00	0.10	0.20	0.30	0.40 μg mL^{-1}
Fluorescence intensity (y_1)	0.00	5.20	9.90	15.30	19.10 arbitrary units

The terms in equation (6) are found from the following tabulated data.

x_1	y_1	x_1^2	y_1^2	x_1y_1
0.00	0.00	0.00	0.00	0.00
0.10	5.20	0.01	27.04	0.52
0.20	9.90	0.04	98.01	1.98
0.30	15.30	0.09	234.09	4.59
0.40	19.10	0.16	364.81	7.64
$\Sigma x_1 = 1.00$	$\Sigma y_1 = 49.5$	$\Sigma x_1^2 = 0.30$	$\Sigma y_1^2 = 723.95$	$\Sigma x_1 y_1 = 14.73$

Therefore

$(\Sigma x_1)^2 = 1.000$; $(\Sigma y_1)^2 = 2450.25$; $n = 5$

Substituting the above values in equation (6), then

$$r = \frac{5 \times 14.73 - 1.00 \times 49.5}{\sqrt{(5 \times 0.30 - 1.000)(5 \times 723.95 - 2450.25)}} = \frac{24.15}{\sqrt{584.75}} = 0.9987$$

Hence, there is a very strong indication that a linear relation exists between fluorescence intensity and concentration (over the given range of concentration).

It must be noted, however, that a value of r close to either $+1$ or -1 does not necessarily confirm that there is a linear relationship between the variables. It is sound practice first to plot the calibration curve on graph paper and ascertain by visual inspection if the data points could be described by a straight line or whether they may fit a smooth curve.

The significance of the value of r is determined from a set of tables (Appendix 15). Consider the following example using five data $(x_1 y_1)$ points: From the table the value of r at 5 per cent significance value is 0.878. If the value of r is greater than 0.878 or less than -0.878 (if there is negative correlation), then the chance that this value could have occurred from random data points is less than 5 per cent. The conclusion can, therefore, be drawn that it is likely that x_1 and y_1 are linearly related. With the value of $r = 0.998_7$ obtained in the example given above there is confirmation of the statement that the linear relation between fluorescence intensity and concentration is highly likely.

4.17 LINEAR REGRESSION

Once a linear relationship has been shown to have a high probability by the value of the correlation coefficient (r), then the best straight line through the data points has to be estimated. This can often be done by visual inspection of the calibration graph but in many cases it is far better practice to evaluate the best straight line by linear regression (the method of least squares).

The equation of a straight line is

$$y = ax + b$$

where y, the **dependent** variable, is plotted as a result of changing x, the **independent** variable. For example, a calibration curve (Section 21.16) in atomic absorption spectroscopy would be obtained from the measured values of absorbance (y-axis) which are determined by using known concentrations of metal standards (x-axis).

To obtain the regression line 'y on x', the slope of the line (a) and the intercept on the y-axis (b) are given by the following equations.

$$a = \frac{n\Sigma x_1 y_1 - \Sigma x_1 \Sigma y_1}{n\Sigma x_1^2 - (\Sigma x_1)^2} \tag{7}$$

and $$b = \bar{y} - a\bar{x} \tag{8}$$

where $\bar{x}$ is the mean of all values of x_1 and $\bar{y}$ is the mean of all values of y_1.

Example 10. Calculate by the least squares method the equation of the best straight line for the calibration curve given in the previous example.

From Example 9 the following values have been determined.

$\Sigma x_1 = 1.00$; $\Sigma y_1 = 49.5$; $\Sigma x_1^2 = 0.30$; $\Sigma x_1 y_1 = 14.73$; $(\Sigma x_1)^2 = 1.000$; the number of points $(n) = 5$

and the values

$$\bar{x} = \frac{\Sigma x_1}{n} = \frac{1.00}{5} = 0.2$$

and

$$\bar{y} = \frac{\Sigma y_1}{n} = \frac{49.5}{5} = 9.9$$

By substituting the values in equations (7) and (8), then

$$a = \frac{5 \times 14.73 - 1.00 \times 49.5}{(5 \times 0.30) - (1.00)^2} = \frac{24.15}{0.5} = 48.3$$

and

$$b = 9.9 - (48.3 \times 0.2) = 0.24$$

So the equation of the straight line is

$$y = 48.3x + 0.24$$

If the fluorescence intensity of the test solution containing quinine was found to be 16.1, then an estimate of the concentration of quinine (x μg mL^{-1}) in this unknown could be

$$16.10 = 48.3x + 0.24$$

$$x = \frac{15.86}{48.30} = 0.32_8 \ \mu\text{g mL}^{-1}$$

The determination of errors in the slope a and the intercept b of the regression line together with multiple and curvilinear regression is beyond the scope of this book but references may be found in the Bibliography, page 156.

4.18 COMPARISON OF MORE THAN TWO MEANS (ANALYSIS OF VARIANCE)

The comparison of more than two means is a situation that often arises in analytical chemistry. It may be useful, for example, to compare (*a*) the mean results obtained from different spectrophotometers all using the same analytical sample; (*b*) the performance of a number of analysts using the same titration method. In the latter example assume that three analysts, using the same solutions, each perform four replicate titrations. In this case there are two possible sources of error: (*a*) the random error associated with replicate measurements; and (*b*) the variation that may arise between the individual analysts. These variations may be calculated and their effects estimated by a statistical method known as the Analysis of Variance (ANOVA), where the **square of the standard deviation**, s^2, is termed the **variance**, V. Thus $F = \frac{s_1^2}{s_2^2}$ where $s_1^2 > s_2^2$, and may be written as $F = \frac{V_1}{V_2}$ where $V_1 > V_2$.

An Analysis of Variance calculation is best illustrated by using specific values in situation (*b*) just referred to.

Example 11. Three analysts were each asked to perform four replicate titrations using the same solutions. The results are given below.

Titre (mL)		
Analyst A	**Analyst B**	**Analyst C**
22.53	22.48	22.57
22.60	22.40	22.62
22.54	22.48	22.61
22.62	22.43	22.65

To simplify the calculation it is sound practice to subtract a common number, e.g. 22.50, from each value. The sum of each column is then determined. *Note*: this will have no effect on the final values.

	Analyst A	**Analyst B**	**Analyst C**
	0.03	−0.02	0.07
	0.10	−0.10	0.12
	0.04	−0.02	0.11
	0.12	−0.07	0.15
Sum =	0.29	−0.21	0.45

The following steps have to be made in the calculation:

(*a*) *The grand total*

$$T = 0.29 - 0.21 + 0.45$$
$$= 0.53$$

(*b*) *The correction factor* (*C.F.*)

$$\text{C.F.} = \frac{T^2}{N} = \frac{(0.53)^2}{12} = 0.0234$$

where N is the total number of results.

(*c*) *The total sum of squares.* This is obtained by squaring each result, summing the totals of each column and then subtracting the correction factor (C.F.).

	Analyst A	**Analyst B**	**Analyst C**
	0.0009	0.0004	0.0049
	0.0100	0.0100	0.0144
	0.0016	0.0004	0.0121
	0.0144	0.0049	0.0225
Sum =	0.0269	0.0157	0.0539

Total sum of squares $= (0.0269 + 0.0157 + 0.0539) - \text{C.F.}$

$= 0.0965 - 0.0234 = 0.0731$

(*d*) *The between-treatment (analyst) sum of squares.* The sum of the squares of each individual column is divided by the number of results in each column, and then the correction factor is subtracted.

Between sum of squares $= \frac{1}{4}(0.29^2 + 0.21^2 + 0.45^2) - 0.0234$

$= 0.0593$

(*e*) *The within-sample sum of squares.* The between sum of squares is subtracted from the total sum of squares.

$0.0731 - 0.0593 = 0.0138$

(*f*) *The degrees of freedom* (v). These are obtained as follows:

The total number of degrees of freedom $v = N - 1 = 11$
The between-treatment degrees of freedom $v = C - 1 = 2$
The within-sample degrees of freedom $v = (N - 1) - (C - 1) = 9$

where C is the number of columns (in this example, the number of analysts).

(*g*) A table of Analysis of Variance (ANOVA table) may now be set up.

Source of variation	Sum of squares	d.f.	Mean square
'Between analysts'	0.0593	2	$0.0593/2 = 0.0297$
'Within titrations'	0.0138	9	$0.0138/9 = 0.00153$
Total	0.0731	11	

(*h*) The F-test is used to compare the two mean squares:

$$F_{2,9} = \frac{0.0297}{0.00153} = 19.41$$

From the F-tables (Appendix 13), the value of F at the 1 per cent level for the given degrees of freedom is 8.02. The calculated result (19.41) is higher than 8.02; hence there is a significant difference in the results obtained by the three analysts. Having ascertained in this example there is a significant difference between the three analysts, the next stage would be to determine whether the mean result is different from the others, or whether all the means are significantly different from each other.

The procedure adopted to answer these questions for the example given above is as follows:

(*a*) Calculate the titration means for each analyst. The mean titration values are $\bar{x}(\text{A}) = 22.57$ mL; $\bar{x}(\text{B}) = 22.45$ mL; and $\bar{x}(\text{C}) = 22.61$ mL.

(*b*) Calculate the quantity defined as the 'least significant difference', which is given by $s\sqrt{2/n}\, t_{0.05}$ where s is the square root of the Residual Mean Square, i.e. the 'within-titration' Mean Square. Hence $s = \sqrt{0.00153}$; n is the number of results in each column (in this example, 4); t is the 5 per cent value from the t-tables (Appendix 12), with the same number of degrees of

freedom as that for the Residual term, i.e. the 'within-titration' value. In this example the number of degrees of freedom is 9, so the least significant difference is given by

$$\sqrt{0.00153} \times \sqrt{2/4} \times 2.26 = 0.06\,\text{mL}$$

If the titration means are arranged in increasing order, then $\bar{x}(B) < \bar{x}(A) < \bar{x}(C)$, and $\bar{x}(C) - \bar{x}(B)$ and $\bar{x}(A) - \bar{x}(B)$ are both greater than 0.06, whereas $\bar{x}(C) - \bar{x}(A)$ is less than 0.06. Hence there is no significant difference between analysts A and C, but the results of analyst B are significantly different from those of both A and C.

It should be noted that in this example the performance of only *one* variable, the three analysts, is investigated and thus this technique is called a one-way ANOVA. If *two* variables, e.g. the three analysts with four *different* titration methods, were to be studied, this would require the use of a two-way ANOVA. Details of suitable texts that provide a solution for this type of problem and methods for multivariate analysis are to be found in the Bibliography, page 156.

4.19 THE VALUE OF STATISTICS

Correctly used, statistics is an essential tool for the analyst. The use of statistical methods can prevent hasty judgements being made on the basis of limited information. It has only been possible in this chapter to give a brief resumé of some statistical techniques that may be applied to analytical problems. The approach, therefore, has been to use specific examples which illustrate the scope of the subject as applied to the treatment of analytical data. There is a danger that this approach may overlook some basic concepts of the subject and the reader is strongly advised to become more fully conversant with these statistical methods by obtaining a selection of the excellent texts now available.

In addition there is the rapidly developing subject of Chemometrics, which may be broadly defined as the application of mathematical and statistical methods to design and/or to optimise measurement procedures, and to provide chemical information by analysing relevant data. Space does not permit an inclusion in this book of such topics as experimental design and instrumental optimisation techniques or more sophisticated subjects as pattern recognition. There is no doubt however, that a knowledge of the scope of Chemometrics will be increasingly important for any competent analytical chemist. Details of some useful texts, both introductory and more advanced, are given in the Bibliography (Section 5.8). The reader should be aware, however, that some signal-processing techniques are included in this book, e.g. information will be found on derivative spectroscopy (Section 17.12) and Fourier transform methods (Section 19.2).

For References and Bibliography see Sections 5.7 and 5.8.

CHAPTER 5
SAMPLING

5.1 THE BASIS OF SAMPLING

The purpose of analysis is to determine the quality or composition of a material; and for the analytical results obtained to have any validity or meaning it is essential that adequate sampling procedures be adopted. Sampling is the process of extracting from a large quantity of material a small portion which is truly representative of the composition of the whole material.

Sampling methods fall into three main groups:

1. those in which all the material is examined;
2. casual sampling on an *ad hoc* basis;
3. methods in which portions of the material are selected based upon statistical probabilities.

Procedure (1) is normally impracticable, as the majority of methods employed are destructive, and in any case the amount of material to be examined is frequently excessive. Even for a sample of manageable size the analysis would be very time-consuming, it would require large quantities of reagents, and would monopolise instruments for long periods.

Sampling according to (2) is totally unscientific and can lead to decisions being taken on inadequate information. In this case, as the taking of samples is entirely casual, any true form of analytical control or supervision is impossible.

For these reasons the only reliable basis for sampling must be a mathematical one using statistical probabilities. This means that although not every item or every part of the sample is analysed, the limitations of the selection are carefully calculated and known in advance. Having calculated the degree of acceptable risk or margin of variation, the sampling plan is then chosen that will give the maximum information and control that is compatible with a rapid turnover of samples. For this reason, in the case of sampling from batches the selection of individual samples is carried out according to special random tables[6] which ensure that personal factors do not influence the choice.

5.2 SAMPLING PROCEDURE

The sampling procedure may involve a number of stages prior to the analysis of the material. The sampling stages are outlined in Fig. 5.1.

For the most part, bulk materials are non-homogeneous, e.g. minerals, sediments, and foodstuffs. They may contain particles of different composition which are not uniformly distributed within the material. In this case, a number

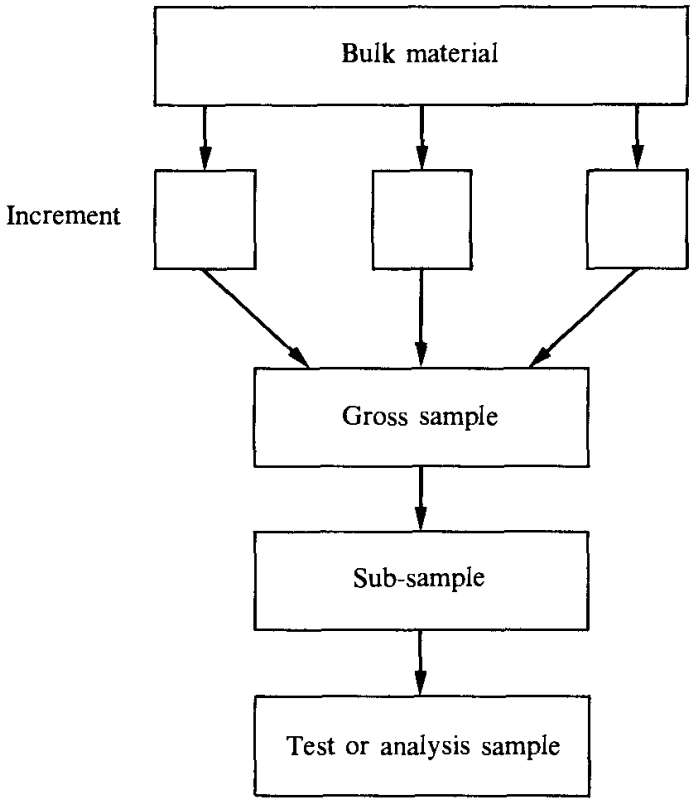


Fig. 5.1

of increments is taken in a random manner from points in the bulk material, so that each part has an equal chance of being selected. The combination of these increments then forms the gross sample. The gross sample is often too large for direct analysis and must be divided further to produce a sub-sample. The sub-sample may require treatment, for example reduction in particle size or thorough mixing, before the analytical sample can be obtained. The analytical sample should retain the same composition of the gross sample.

It must be stressed, however, that the whole object may be the analytical sample, e.g. a specimen of moon-rock. Ideally this sample would be analysed by non-destructive methods. Occasionally the bulk material may be homogeneous (some water samples) and then only one increment may be needed to determine the properties of the bulk. This increment should be of suitable size to provide samples for replicate analyses.

5.3 SAMPLING STATISTICS

The errors arising in sampling, particularly in the case of heterogeneous solids, may be the most important source of uncertainty in the subsequent analysis of the material. If we represent the standard deviation of the sampling operation (the sampling error) by s_S and the standard deviation of the analytical procedures (the analytical error) by s_A, then the overall standard deviation s_T (the total error) is given by

$$s_T = \sqrt{s_S^2 + s_A^2} \tag{1}$$

or

$$s_T = \sqrt{V_S + V_A}$$

where V represents the appropriate variance. The separate evaluation of both V_S (the sampling variance) and V_A (the analytical variance) may be achieved by using the analysis of variance procedure (See Section 4.18). A comparison can be made of the between-sample variance — an estimate of the sampling error — and the within-sample variance — an estimate of the analytical error.

Example 1. If the sampling error is ± 3 per cent and the analytical error is ± 1 per cent, from equation (1) we can see that the total error s_T is given by

$$s_T = \sqrt{3^2 + 1^2} = \pm 3.16 \text{ per cent}$$

If, in the above example, the analytical error was ± 0.2 per cent then the total error s_T would be equal to ± 3.006 per cent. Hence the contribution of the analytical error to the total error is virtually insignificant. Youden[7] has stated that once the analytical uncertainty is reduced to one-third of the sampling uncertainty, further reduction of the former is not necessary. It is most important to realise that if the sampling error is large, then a rapid analytical method with relatively low precision may suffice.

In designing a sampling plan the following points should be considered:[8]

(*a*) the number of samples to be taken;
(*b*) the size of the sample;
(*c*) should individual samples be analysed or should a sample composed of two or more increments (composite) be prepared.

If the composition of the bulk material to be sampled is unknown, it is sensible practice to perform a preliminary investigation by collecting a number of samples and determining the analyte of interest.

The confidence limits (see Section 4.11) are given by the relationship

$$\mu = \bar{x} \pm t s_s \Big/ \sqrt{n} \qquad (2)$$

where s_s is the standard deviation of individual samples, $\bar{x}$ is the mean of the analytical results and serves as an estimate of the true mean μ, and n is the number of samples taken.

Example 2. An estimate of the variability of nickel in a consignment of an ore, based on 16 determinations, was found to be $\pm 1.5\%$. How many samples should be taken to give (at the 95 per cent confidence level) a sampling error of less than 0.5 per cent nickel?

The value 0.5 per cent is in fact the difference between the sample mean $\bar{x}$ and the actual value μ. If this value is represented by E, then equation (2) may be written as

$$E = t\, s_s/\sqrt{n}$$

and, therefore,

$$n = \left(\frac{t\, s_s}{E}\right)^2$$

From the tables (Appendix 12) the value of t for $(n-1)$, 15 degrees of freedom

at the 95 per cent confidence level is 2.13.

$$\therefore n = \left(\frac{2.13 \times 1.5}{0.5}\right)^2 \approx 41$$

Hence, from this test it has been shown that at least 41 samples are required if the specifications given in the above example are to be satisfied.

The other major problem concerned with sampling is that of the sample size. The size of the sample taken from a heterogeneous material is determined by the variation in particle size, and the precision needed in the results of the analysis.

The major source of error in sampling can arise from the taking of increments from the bulk material. It can be shown from random sampling theory that the accuracy of the sample is determined by its total size. Hence, the sampling variance, V, is inversely proportional to the mass of the sample. However, this statement is not true if the bulk material consists of varying particle sizes; then the number of increments taken will influence the sampling accuracy. The sampling variance, V, is inversely proportional to the number of sampling increments (n):

$$V = \frac{k}{n} \tag{3}$$

where k is a constant dependent on the size of the increment and variation within the bulk material.

5.4 SAMPLING AND PHYSICAL STATE

Many of the problems occurring during sampling arise from the physical nature of the materials to be studied.[9] Although gases and liquids can, and do, present difficulties, the greatest problems of adequate sampling undoubtedly arise with solids.

Gases. Few problems arise over homogeneity of gas mixtures where the storage vessel is not subjected to temperature or pressure variations. Difficulties may arise if precautions are not taken to clear valves, taps and connecting lines of any other gas prior to passage of the sample. Similarly care must be taken that no gaseous components will react with the sampling and analytical devices.

Liquids. In most cases general stirring or mixing is sufficient to ensure homogeneity prior to sampling. Where separate phases exist it is necessary to determine the relative volumes of each phase in order to compare correctly the composition of one phase with the other. The phases should in any case be individually sampled as it is not possible to obtain a representative sample of the combined materials even after vigorously shaking the separate phases together.

Solids. It is with solids that real difficulties over homogeneity arise. Even materials that superficially have every appearance of being homogeneous in fact may have localised concentrations of impurities and vary in composition. The procedure adopted to obtain as representative a sample as possible will depend greatly upon the type of solid. This process is of great importance since, if it is not satisfactorily done, the labour and time spent in making a careful analysis

of the sample may be completely wasted. If the material is more or less homogeneous, sampling is comparatively simple. If, however, the material is bulky and heterogeneous, sampling must be carried out with great care, and the method will vary somewhat with the nature of the bulk solid.

The underlying principle of the sampling of material in bulk, say of a truckload of coal or iron ore, is to select a large number of portions in a systematic manner from different parts of the bulk and then to combine them. This large sample of the total weight is crushed mechanically, if necessary, and then shovelled into a conical pile. Every shovelful must fall upon the apex of the cone and the operator must walk around the cone as he shovels; this ensures a comparatively even distribution. The top of the cone is then flattened out and divided into quarters. Opposite quarters of the pile are then removed, mixed to form a smaller conical pile, and again quartered. This process is repeated, further crushing being carried out if necessary, until a sample of suitable weight (say, 200–300 g) is obtained.

If the quantity of material is of the order of 2–3 kg or less, intermixing may be accomplished by the method known as 'tabling'. The finely divided material is spread on the centre of a large sheet of oilcloth or similar material. Each corner is pulled in succession over its diagonal partner, the lifting being reduced to a minimum; the particles are thus caused to roll over and over on themselves, and the lower portions are constantly brought to the top of the mass and thorough intermixing ensues. The sample may then be rolled to the centre of the cloth, spread out, and quartered as before. The process is repeated until a sufficiently small sample is obtained. The final sample for the laboratory, which is usually between 25 and 200 g in weight, is placed in an air-tight bottle. This method produces what is known as the 'average sample' and any analysis on it should always be compared with those of a second sample of the same material obtained by the identical routine.

Mechanical methods also exist for dividing up particulate material into suitably sized samples. Samples obtained by these means are usually representative of the bulk material within limits of less than ± 1 per cent, and are based upon the requirements established by the British Standards Institution. Sample dividers exist with capacities of up to 10 L and operate either by means of a series of rapidly rotating sample jars under the outlet of a loading funnel, or by a rotary cascade from which the samples are fed into a series of separate compartments. Sample dividers can lead to a great deal of time-saving in laboratories dealing with bulk quantities of powders or minerals.

The sampling of metals and alloys may be effected by drilling holes through a representative ingot at selected points; *all* the material from the holes is collected, mixed, and a sample of suitable size used for analysis. Turnings or scrapings from the outside are not suitable as these frequently possess superficial impurities from the castings or moulds.

In some instances in which grinding presents problems it is possible to obtain a suitable homogeneous sample by dissolving a portion of the material in an appropriate solvent.

Before analysis the representative solid sample is usually dried at 105–110 °C, or at some higher specified temperature if necessary, to constant weight. The results of the analysis are then reported on the 'dry' basis, viz. on a material dried at a specified temperature. The loss in weight on drying may be determined, and the results may be reported, if desired on the original 'moist' basis; these

figures will only possess real significance if the material is not appreciably hygroscopic and no chemical changes, other than the loss of water, take place on drying.

In a course of systematic quantitative analysis, such as that with which we are chiefly concerned in the present book, the unknowns supplied for analysis are usually portions of carefully analysed samples which have been finely ground until uniform.

It should be borne in mind that although it is possible to generalise on sampling procedures, all industries have their own established methods for obtaining a record of the quantity and/or quality of their products. The sampling procedures for tobacco leaves will obviously differ from those used for bales of cotton or for coal. But although the types of samples differ considerably the actual analytical methods used later are of general application.

5.5 CRUSHING AND GRINDING

If the material is hard (e.g. a sample of rock), it is first broken into small pieces on a hard steel plate with a hardened hammer. The loss of fragments is prevented by covering the plate with a steel ring, or in some other manner. The small lumps may be broken in a 'percussion' mortar (also known as a 'diamond' mortar) (Fig. 5.2). The mortar and pestle are constructed entirely of hard tool steel. One or two small pieces are placed in the mortar, and the pestle inserted into position; the latter is struck lightly with a hammer until the pieces have been reduced to a coarse powder. The whole of the hard substance may be treated in this manner. The coarse powder is then *ground* in an agate mortar in small quantities at a time. A mortar of mullite is claimed to be superior to one of agate: mullite is a homogeneous ceramic material that is harder, more resistant to abrasion, and less porous than agate. A synthetic sapphire mortar and pestle (composed essentially of a specially prepared form of pure aluminium oxide) is marketed; it is extremely hard (comparable with tungsten carbide) and will grind materials not readily reduced in ceramic or metal mortars. Mechanical (motor-driven) mortars are available commercially.

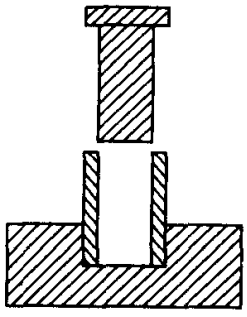


Fig. 5.2

5.6 HAZARDS IN SAMPLING

The handling of many materials is fraught with hazards[10] and this is no less so when sampling materials in preparation for chemical analysis. The sampler must always wear adequate protective clothing and if possible have detailed prior knowledge of the material being sampled. When dangers from toxicity

exist the necessary antidotes and treatment procedures should be available and established before sampling commences.[11] In no instances should naked flames be allowed anywhere near the sampling area.

Apart from the toxic nature of many gases, the additional hazards are those of excessive release of gas due to pressure changes, spontaneous ignition of flammable gases and sudden vaporisation of liquefied gases.

With liquids, dangers frequently arise from easily volatilised and readily flammable liquids. In all cases precautions should be greater than under normal circumstances due to the unpredictable nature and conditions of taking samples. The sampler must always be prepared for the unexpected, as can arise, for example, if a container has built up excess pressure, or if the wrong liquid has been packed. Toxic and unknown liquids should never be sucked along tubes or into pipettes by mouth.

Even the sampling of solids must not be casually undertaken, and the operator should always use a face mask as a protection until it is established that the powdered material is not hazardous.

It should be borne in mind that sampling of radioactive substances is a specialist operation at all times and should be carried out only under strictly controlled conditions within restricted areas. In almost all instances the operator must be protected against the radioactive emanations from the substance he is sampling.

Correct sampling of materials is therefore of importance in two main respects; firstly to obtain a representative portion of the material for analysis, and secondly to prevent the occurrence of accidents when sampling hazardous materials.

5.7 REFERENCES FOR PART B

1. W G Gosset 'Student', *Biometrika*, 1908, **6**, 1
2. J Mandel *The Statistical Analysis of Experimental Data*, Interscience, New York, 1964
3. C J Brookes, I G Betteley and S M Loxston *Mathematics and Statistics for Chemists*, John Wiley, New York, 1966, p. 304
4. K Eckschlager *Errors, Measurements and Results in Chemical Analysis*, Van Nostrand Reinhold, London; 1969
5. C T Shewell *Anal. Chem.*, 1959, **31**(5), 21A
6. J Murdoch and J A Barnes (1970). *Statistical Tables for Science, Engineering and Management*, 2nd edn, Macmillan, London, pp. 30–33
7. W J Youden *J. Assoc. Off. Anal. Chem.*, 1967, **50**, 1007
8. B Kratochvil and J. K. Taylor. *Anal. Chem.*, 1981, **53**, 925A
9. C R N Strouts, J H Gilfillan and H N Wilson *Sampling in Analytical Chemistry: The Working Tools*, Vol I, Oxford University Press, London, 1955, Chapter 3
10. N Irving Sax (Ed) *Dangerous Properties of Industrial Materials*, 3rd edn, Reinhold, New York, 1968
11. G D Muir (Ed) *Hazards in the Chemical Laboratory*, Royal Institute of Chemistry, London, 1971

5.8 SELECTED BIBLIOGRAPHY FOR PART B

1. J C Miller and J N Miller *Statistics for Analytical Chemistry*, 2nd edn, Wiley, Chichester, 1988
2. C Chatfield *Statistics for Technology*, 2nd edn, Chapman and Hall, London, 1984

3. O L Davies and P L Goldsmith *Statistical Methods in Research and Production*, Longman, London, 1982
4. M J Moroney *Facts from Figures*, 3rd edn, Penguin Books, Harmondsworth, 1965
5. G E P Box, W G Hunter and J S Hunter *Statistics for Experimentalists*, Wiley, New York, 1978
6. V Barnett and T Lewis *Outliers in Statistical Data*, Wiley, New York, 1978
7. R Caulcutt and R Boddy *Statistics for Analytical Chemists*, Chapman and Hall, London, 1983
8. P Sprent *Quick Statistics*, Penguin Books, Harmondsworth, 1981
9. B E Cooper *Statistics for Experimentalists*, Pergamon Press, Oxford, 1975
10. B R Kowalski (Ed) *Chemometrics: Theory and Application*, American Chemical Society, Washington, DC, 1977
11. D L Massart, A Dijkstra and L Kaufman *Evaluation and Optimisation of Laboratory Methods and Analytical Procedures*, Elsevier, Amsterdam, 1978
12. C R Hicks *Fundamental Concepts in the Design of Experiments*, 2nd edn, Holt, Rinehart and Winston, New York, 1973
13. K Varmuza *Pattern Recognition in Chemistry*, Springer-Verlag, Berlin, 1980
14. D L Massart, B G M Vandeginste, S N Deming, Y Michotte and L Kaufman *Chemometrics: A Textbook*, Elsevier, Amsterdam, 1987
15. S N Deming and S L Morgan *Experimental Design: A Chemometric Approach*, Elsevier, Amsterdam, 1987
16. J D Lee and T D Lee *Statistics and Computer Methods in BASIC*, Van Nostrand, London, 1983
17. D McCormick and A Roach *Measurements, Statistics and Computation*, ACOL–Wiley, Chichester, 1987
18. R G Brereton Chemometrics in analytical chemistry, a review, *Analyst*, 1987, **112**, 1635
19. C L Wilson and D W Wilson (Eds) Sampling, In *Comprehensive Analytical Chemistry*, Vol 1A, *Classical Analysis*, Elsevier, Amsterdam, 1959, Chapter II.3
20. H A Laitinen Sampling, In *Chemical Analysis: An Advanced Text and Reference*, McGraw-Hill, New York, 1960, Chapter 27
21. N V Steere (Ed) *Handbook of Laboratory Safety*, Chemical Rubber Co., Cleveland, Ohio, 1967
22. G Kateman and F W Pijpers *Quality Control in Analytical Chemistry*, Wiley, Chichester, 1981
23. R Smith and G V James *The Sampling of Bulk Materials*, Royal Society of Chemistry, London, 1981
24. B W Woodget and D Cooper *Samples and Standards*, ACOL–Wiley, Chichester, 1987
25. R Anderson *Sample Pretreatment and Separation*, ACOL–Wiley, Chichester, 1987