Errors and evaluation of analytical data

Presented by
K.N.S.SWAMI., M.Sc., SET.
Guest Faculty in Chemistry
Department Of Chemistry
P. R. Government College (Autonomous),Kakinada.

CONTENTS:

- ERRORS
- ACCURACY AND PRECISION
- PROPAGATION OF ERRORS
- SIGNIFICANT FIGURES
- COMPUTATION RULES
- CONFIDENCE LIMITS & CONFIDENCE INTERVAL

Errors

The term error has two slightly different meanings.

1) error refers to the difference between a measured value and the "true" or "known" value.
2) error often denotes the estimated uncertainty in a measurement or experiment.

- In quantitative analysis, when numerical data and numerical results are measured with the greatest exactness.
- It has been observed that the results of successive determination differ among themselves to a greater or lesser extent.
- Measurements invariably involve errors and uncertainties.
- it is impossible to perform a chemical analysis that is totally free of errors or uncertainties
- We can only hope to minimize errors and estimate their size with acceptable accuracy
- Errors are caused by faulty calibrations or standardizations or by random variations and uncertainties in results.
- Frequent calibrations, standardizations, and analyses of known samples can sometimes be used to lessen all but the random errors and uncertainties.

" We can only hope to minimize errors and estimate their size with acceptable accuracy"

Classification Of Errors

Generally Chemical analyses are affected by two types of errors:

1. Systematic (or determinate) error, causes the mean of a data set to differ from the accepted value.
2. Random (or indeterminate) error, causes data to be scattered more or less symmetrically around a mean value.

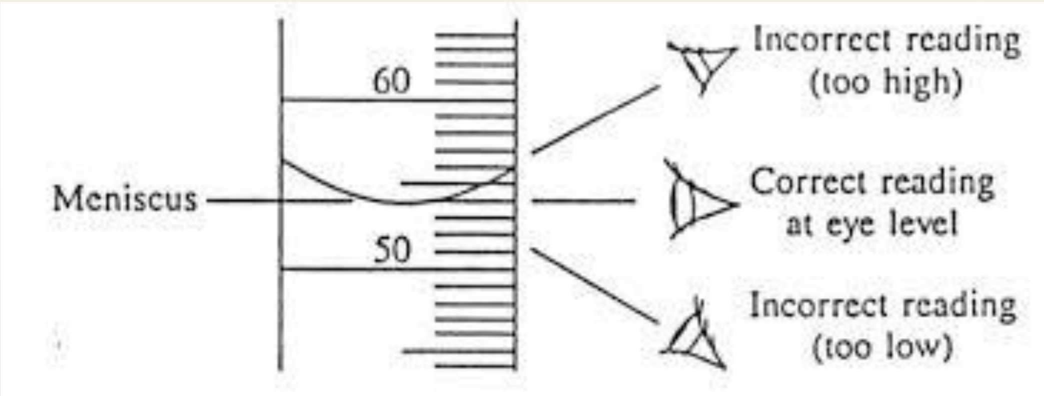
1. Systematic or determinate or constant errors: These errors can be avoided and their magnitude can be determined, thereby correcting the measurements.

or

- have a definite value,
- an assignable cause, and are of the same magnitude for replicate measurements made in the same way.
- They lead to bias in measurement results.

There are Four types of systematic errors:

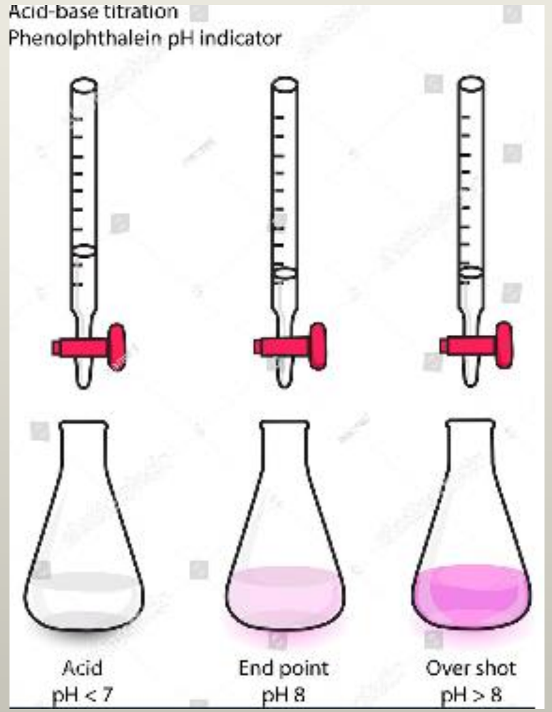
- Personal errors and *Operational errors*
- Instrumental errors *and reagent errors*
- Method errors
- *Additive and proportional errors*



I.Personal errors and Operational Errors:

These errors are not connected with the method or procedure but the individual analyst is responsible for them. This type of errors may arise due to the inability of the individual making observations. Some important personal errors are:

A. Inability in judging color change sharply in visual titrations.
B. Error in reading a burette.
C. Mechanical loss of material in various steps of an analysis.
D. Failure to wash and ignite a precipitate properly.
E. Insufficient cooling of crucible before weighing.
F. Using impure reagents.
G. Ignition of precipitate at incorrect temperatures.
H. Errors in calculations.



II.Instruments and reagent errors:

following factors are responsible for such errors:

A. Balance arms of unequal lengths.

B. Uncalibrated or improperly calibrated weights.

C. Incorrectly graduated burettes.

D. Attack of foreign materials upon glassware.

E. Loss in weight of platinum crucibles on strong heating.

F. Impure reagents.

These errors can be avoided by using calibrated weights, glassware and pure reagents.

III. 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 volatilization 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.

IV. Additive and proportional errors:

- Absolut vale of additive error is independent of the amount of the constituent present in the determination

e.g., loss in weight of a crucible adds error to the weight of precipitate is ignited in it.

- On the other hand, the magnitude of proportional error depends upon the quantity of the constituent.

- e.g., impurity present in a standard substance gives a wrong value for the normality of a standard solution.

Random or Indeterminate Errors:

These errors are accidental and analyst has no control over them.

- These are random in nature and lead to both high and low result with equal probability.
- These cannot be eliminated or corrected and are the ultimate limitation on the measurement.
- These can be treated by statistics repeated measurements of the same variable can have the effect of reducing their importance.

Accuracy

- Accuracy indicates the closeness of the measurement to the true or accepted value and is expressed by the error.
- Accuracy measures agreement between a result and the accepted value.
- Accuracy is often more difficult to determine because the true value is usually unknown. An accepted value must be used instead.
- Systematic errors cause a constant error (either too high or too low) and thus affect the accuracy of a result.
- Accuracy is expressed in terms of either absolute or relative error.

Determination of Accuracy:

Absolute Error

* The absolute error of a measurement is the difference between the measured value and the true value. If the measurement result is low, the sign is negative; if the measurement result is high, the sign is positive.

$$\text{Absolute Error} = \text{Experimental Value} - \text{True Value}$$

Relative Error

- The relative error of a measurement is the absolute error divided by the true value.
- Relative error may be expressed in percent, parts per thousand, or parts per million, depending on the magnitude of the result.

$$E_r = \frac{x_i - x_t}{x_t} \times 100\%$$

$$\% \text{ Error} = \frac{\text{Experimental Value- Theoretical Value}}{\text{Theoretical}} \times 100$$

Precision

* Precision describes the agreement among several results obtained in the same way. Describes the reproducibility of measurements.

*Precision is readily determined by simply repeating the measurement on replicate samples.

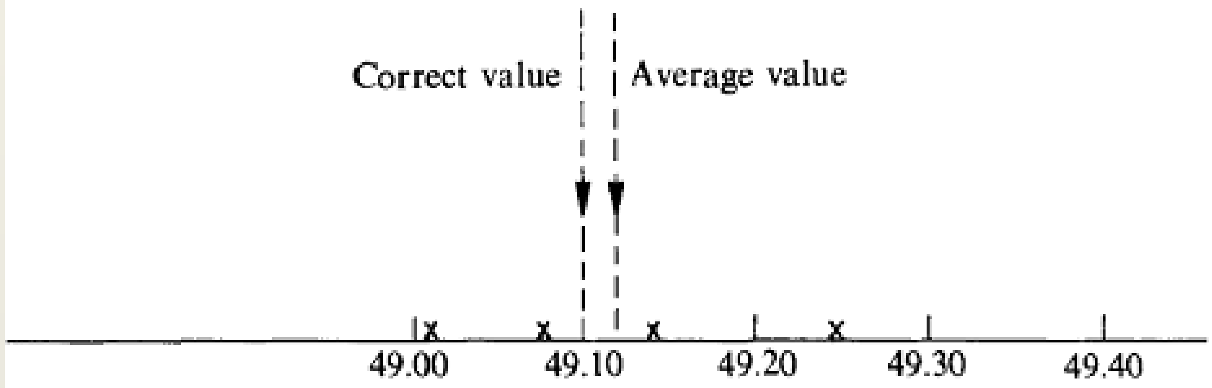
*Precision of a set of replicate data may be expressed as standard deviation, variance, and coefficient of variation.

* deviation from mean is how much x_i the individual result deviates from the mean.

Precision always accompanies accuracy, but a high degree of precision does not imply accuracy

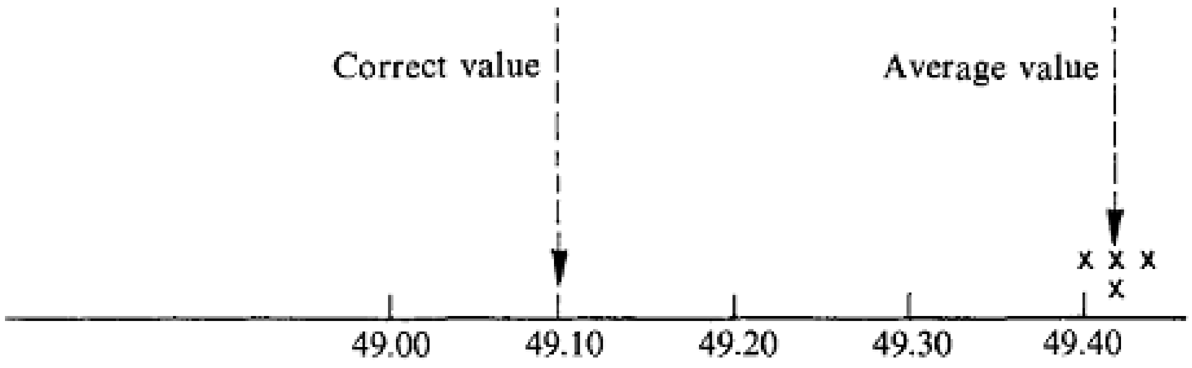
A substance was known to contain 49.10 + or - 0.02 per cent of a constituent A. The results obtained by two analysts using the same substance and the same analytical method were as follows.

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



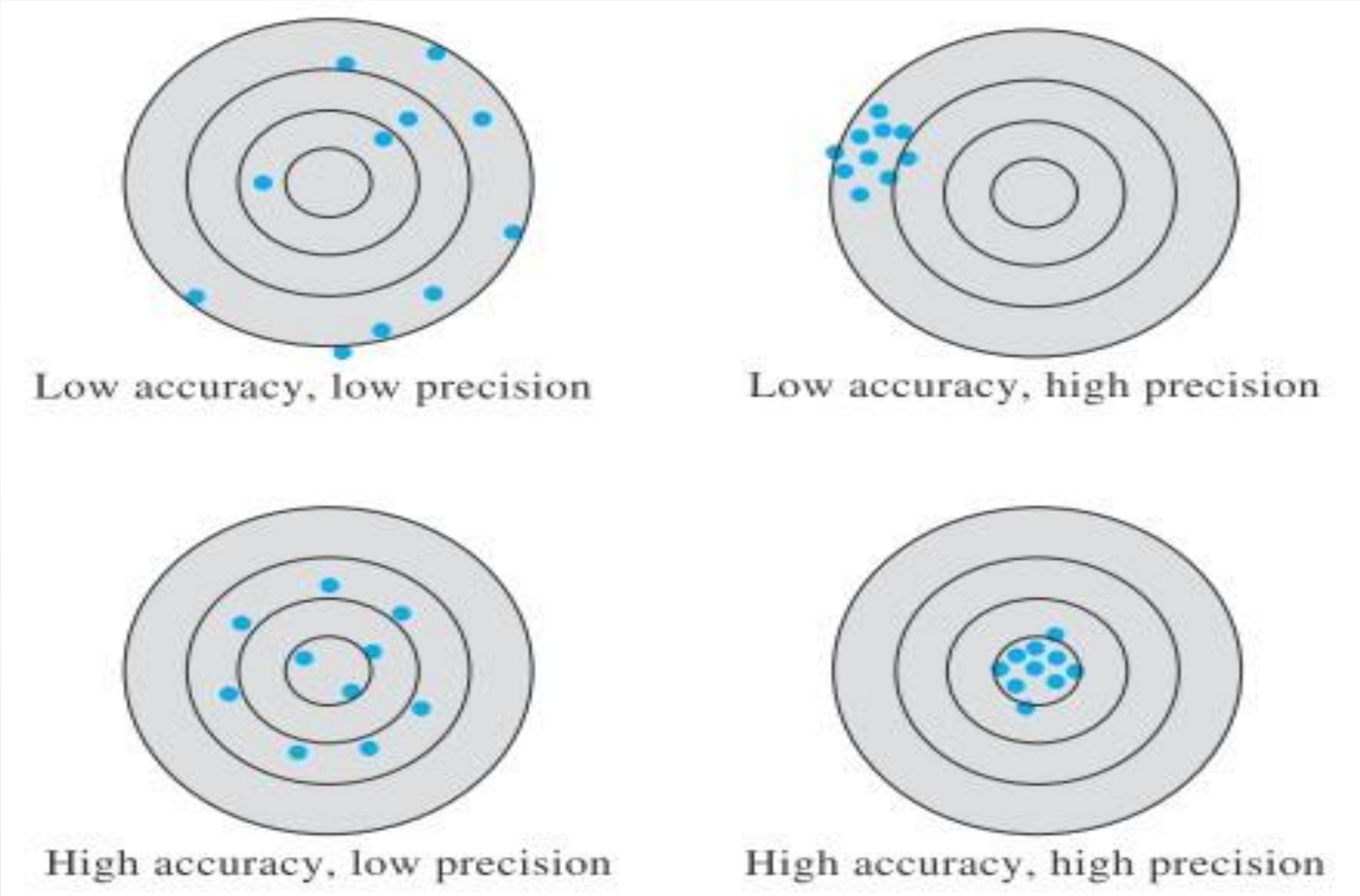
The arithmetic mean is 49.42% and the results range from 49.40% to 49.44%.

We can summarise the results of the analyses as follows.

- 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.
- 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:

- 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).


Low accuracy, low precision
Low accuracy, high precision
High accuracy, low precision
High accuracy, high precision

Methods of expressing Precision:

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.):

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

Propagation Of Errors

The numerical result of a measurement is not of interest in its own right, but rather is used, sometimes in conjunction with several other measurements to calculate the quantity which is actually desired attention is naturally focused upon the precision and accuracy of the final computed quantity, but it is instructive to see how errors in the individual measurements are propagated into this result.

Determinate Errors:

Addition and Subtraction:

Consider a computed result, R based upon the measured quantities A, B and C. Let α, β and γ represents the absolute determinate errors in A, B and C respectively and let **ρ represents** the resulting error in R if the actual measurements are $A+\alpha$, $B+\beta$ and $C+\gamma$ We can see how the errors are transmitted through addition and subtraction .

suppose $R=A+B-C$

$$R+\rho = (A+\alpha)+(B+\beta)-(C+\gamma)$$

$$R+\rho=(A+B-C)+(\alpha+\beta-\gamma)$$

$$R= A+B-C$$

$$\rho = \alpha+\beta-\gamma$$

Multiplication and Division:

Now suppose, on the other hand, that multiplication and division are involved', i.e, let R= AB/C ABC. Again the actual measurements are A+ α,B+β and C+γ. Then

$$R+\rho = \frac{(A+\alpha)\ (B+\beta)}{C+\gamma} = \frac{AB+\alpha B+\beta A+\alpha\beta}{C+\gamma}$$

Let us neglect , α and β, since it may be supposed that the errors are very small compared with the measured values. Then subtracting R=AB/C gives

$$\rho = \frac{AB+\alpha B+\beta A}{C+\gamma} - \frac{AB}{C}$$

Placing the right-hand terms over a common terminator, we get

$$\rho = \frac{\alpha BC+\beta AC-\gamma AB}{C\ (C+\gamma)}$$

It is now convenient to consider the relative error, ρ/R by dividing by $R=AB/C$, which leads, after appropriate cancellation to

$$\frac{\rho}{R} = \frac{\alpha BC+\beta AC-\gamma AB}{AB\,(C+\gamma)}$$

Since γ is very small compared with C, this reduces to

$$\frac{\rho}{R} = \frac{\alpha}{A} + \frac{\beta}{B} - \frac{\gamma}{C}$$

Thus it is found that determinate errors are propagated follow.

1. Where addition or subtraction is involved, the absolute determinants errors are transmitted directly into the result.

2. Where multiplication or division is involved, the relative determinate errors are transmitted directly into the result.

Significant Figures

- A figure of digit denotes any one of the ten numerals (0,1,2,3,4,5,6,7,8,9). A digit alone or in combination serves to express a number.
- A significant figure is a digit having some practical meaning, *i.e.* it is a digit, which denotes the amount of the quantity in the place in which it stands.
- For example in 0.456, 4.56 and 456 there are three significant figures in each number.
- Zero may or may not be a significant figure. A zero is a significant figure except when it serves to locate the decimal point, while it is a significant figure when it indicates that the quantity in place in which *i.e.* in 1.3680 and 1.0082, zero is significant but in 0.0035, zeros are not the significant figures as they serve only to locate the decimal point. Thus, first two numbers contain five but the third one contains two significant figures.

Computation Rules

- Rule 1 → In expressing an experimental measurement, never retain more than one doubtful digit. Eliminate all the digits that are not significant.
- Rule 2 → Retain as many significant figures in a result or in any data as will give only one uncertain figure. *e.g. a* volume between 30.5 ml and 30.7 ml should be written as 30.6 ml. and not as 30.60 as it would be between 30.59 and 30.61.
- Rule 3 → Two rules are given for rejecting superfluous (Unnecessary)digits.

1. When the last digit dropped is greater than 5, the last digit retained is increased by one. e.g. in rejecting the last digit in 8.947, the new value will be 8.95 as 7 is greater than 5. But when 4.863 is rounded up to two digits, it gives 4.9 as the first digit discarded is 6 which is greater than 5. This is known as rounding up.

2. If the first digit discarded is less than 5, leave the last digit unchanged. It is known as rounding down. e.g. when the number 5.64987 is rounded to two digits, we get 5.6 as the first digit, discarded is 4, which is less than 5. Rounding never changes the power of 10. Thus, it is better to express numbers in exponential notation before rounding. e.g. in rounding 57832 to four figures, result 5.783 X 10^{4}.

- Rule 4. In addition or subtraction, there should be in each number only as many significant figures as there are in the least accurately known number.

  e.g. sum of three values 35.6, 0.162 and 71.41 should be reported only to the

  first decimal place as the value 35.6 is known only to the first decimal place. Thus, the answer 107.172 is rounded to 107.2

- Rule 5. 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.
  e.g. the product of the three figures 0.0121, 25.64 and 1.05782 is 0.0121 X 25.6 X 1.06 = 0.328

- When a calculation involves both addition or subtraction and multiplication or division, addition is done first so as to determine the number of significant figures in the answer.

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 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

Confidence limits of μ , for n replicate measurements $\mu = \bar{x} \pm \frac{ts}{\sqrt{n}}$

,
where t is a parameter that depends upon the number of degrees of freedom(ν) and the confidence level required. A table of the values of t at different confidence levels and degrees of freedom (ν)

Example :

The mean (2) of four determinations of the copper content of a sample of an alloy was 8.27 percent with a standard deviation s = 0.17 percent. 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 - l), i.e. three degrees of freedom, is 3.18.

Hence from equation ,the 95 percent confidence level,

$$95\%(\text{C.L.}) \text{ for } \mu = 8.27 \pm \frac{3.18 \times 0.17}{\sqrt{4}}$$
$$= 8.27 \pm 0.27 \text{ per cent}$$

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

$$95\%(\text{C.L.}) \text{ for } \mu = 8.27 \pm \frac{2.20 \times 0.17}{\sqrt{12}}$$
$$= 8.27 \pm 0.11 \text{ per cent}$$

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

Thank You