TREATMENT OF ANALYTICAL DATA

Introduction:

Analytical chemistry involve separating, identifying and determining the relative amounts of the components in a sample of matter. Qualitative analysis reveals the chemical identity of the species in the sample. Quantitative analysis establishes the relative amount of one or more of these species or analytes, in numerical terms. Qualitative information is required before a quantitative analysis can be undertaken.

In analytical chemistry much effort is laid in gathering quantitative data. Such quantitative data can be obtained from experimental measurements. These measurements are subjected to error. Hence it is necessary for analytical chemists to study the errors that attend their measurements in order to minimise them. The significance of experimental results can be examined by certain methods. Most of these methods, which are applied on experimental results, are based upon statistical concepts. Being a chemist it is necessary to express the experimental results properly so that maximum information is conveyed to the people.

Accuracy and Precision:

A numerical value obtained in ~~a~~ scientific measurement contains some degree of uncertainty. Though the uncertainty in the measurement is small, it is present in very accurate work. This uncertainty is called the error of the measurement. Two different aspects are to be considered, while discussing about the errors present in a set of analytical data. The first aspect is accuracy and the second aspect is precision.

Accuracy relates to the closeness of approach of a single measurement or of the average of a series of measurements, to the true value. Thus accuracy can be described as the degree of agreement between a measured value and the true value. No measurement of a quantity is completely accurate and the true value is one which is known within certain limits. Hence a true value is one which can be approached but cannot be attained.

The precision of a set of numerical measurements is defined as degree of agreement between it and other values. Repeated measurements of the same quantity will usually not be identical, but will scatter around some common value. Precision describes the reproducibility or scatter of a series of measurements or results, while accuracy is used to express the correctness of a measurement.

Both the terms accuracy and precision have specific meanings, so they should not be used interchangeably and carelessly.

To understand the exact meanings of accuracy and precision, ~~let us~~ consider the two sets of data for the weight of a crucible. The data in set A were ~~be~~ obtained by five different students each weighing the crucible on his own analytical balance using different sets of weights. The second set (set B) were obtained by sixth student, who has weighed the ~~ca~~ crucible five times on his analytical balance. At last, the crucible was weighed by a skilled analyst and he finds a value of 9.2474 g as the average of ten replicate measurements.

	Set A	Set B
Weight of crucible in g	9.2463	9.2483
	9.2480	9.2481
	9.2477	9.2484
	9.2489	9.2480
	9.2455	9.2483
Average weight in g	9.2473	9.2482

The weights obtained in set B are less scattered and much closer to each other than those found in set A. The precision of these data is higher than the precision of the data in set A. The average of the weights obtained from set A is closer to the true value than the average of weights in set B.

These data reveal an interrelation between accuracy and precision. Precise data need not be necessarily accurate. The data in set A, though has accuracy lacks precision. Accuracy without precision is highly unlikely and meaningless and there can be precision without accuracy.

Measure Of Accuracy:

The general expression of accuracy is merely the difference between measured value (or the average value) and the true value. This is known as the absolute error or the mean error. The calculation of absolute error does not provide much information about accuracy. For example, an error of 3.0 mg (low) is committed in the weighing of 30 mg. The error is significant if the weight is 30 mg and is insignificant if the weight is 30 g.

Thus the accuracy of a measurement is usually expressed in relative terms, as a fraction of the true value. Relative accuracy is expressed in terms of parts per thousand (P.P.t). This is obtained by multiplying the fraction by one thousand

$$\therefore \text{Relative accuracy or relative error} = \frac{\text{absolute error}}{\text{true value}} \times 1000$$

The significance of relative error as a measure of accuracy can be understood from the following example.

The relative error of 3.0 mg in 30 mg weight can be given as

$$\text{relative error} = \frac{0.0030}{0.0300} \times 1000 = 100 \text{ ppt}.$$

The relative error of 3.0 mg in 30 g weight is

$$\text{relative error} = \frac{0.0030}{30.0000} \times 1000 = 1 \text{ ppt}.$$

$\therefore$ The error of 3.0 mg in 30 mg is significant whereas it is insignificant in 30 g weight.

Precision Measures:

Precision of a set of data can be expressed in many different ways.

The Range (w)

This is merely the numerical difference between the highest and the lowest values of a set of results. For the data in set A

$$w = 9.2489\,g - 9.2455\,g = 0.0034\,g.$$

The calculation of range as a measure of precision is not much informative. Because Range tells nothing about the distribution of the data. The range for two sets of measurements may be same, but the distribution of measurements in the two sets need not be same. Sometimes the data may be uniformly distributed between the two limits (i.e. maximum and minimum) or sometimes all the data could be clustered around one value, with just a single datum way off. The range is most frequently used for very small sets of data, i.e. two or three items. The range is more informative about the precision of a set of data than the average deviation when there are fewer than eight pieces in the set.

Average deviation :(a or $\bar{d}$)

Average deviation is another measure of precision. The mean deviation or average deviation is the mean of the deviations of all the individual measurements. This can be calculated by determining the arithmetical mean of the results, then calculating the deviation of each individual measurement from the mean and finally dividing the sum of the deviations, regardless of sign, by the number of measurements.

For finite set of measurements average or arithmetic mean m or $\bar{x}$ is $\frac{\Sigma x_i}{n}$. The deviation of each measurement from mean can be given as

$|x_i - m|$

∴ The mean deviation or average deviation is calculated by the formula

$$a \text{ or } \bar{d} = \frac{\sum_i |x_i - m|}{n}$$

Standard deviation (s or σ)

Standard deviation is a statistical term used to express precision. σ would be the parameter used to describe the standard deviation of the members of the set containing infinite measurements. Whereas for finite set of measurements the symbol for standard deviation is S. The calculation of standard deviation is found to be more reliable than mean deviation or relative mean deviation [the relative mean deviation is the mean deviation divided by the mean and is expressed in percentage or parts per thousand].

Standard deviation is defined as the square root of the sum of the squares of the deviation divided by one less than the numb of pieces of data in the set.

$$S = \sqrt{\frac{\sum_i (x_i - m)^2}{(n-1)}}$$

The smaller the value of the standard deviation, the more precise is the data.

Variance (S^2):

Variance is the square of standard deviation. For certain statistical tests it is more convenient to use square of the standard deviation, instead of standard deviation.

$$S^2 = \frac{\sum_i (x_i - m)^2}{(n-1)}$$

Relative standard deviation (~~[illegible]~~)

The relative standard deviation is the standard deviation divided by the mean of ~~[illegible]~~ the set. It expressed as parts per thousand.

$$\text{relative standard deviation} = \frac{S}{m} \times 1000$$

Errors in Quantitative analysis:

When certain numerical data and numerical results are measured in a quantitative analysis, then the successive results may differ among themselves to a greater or lesser extent. The average value of a series of measurements is accepted as the 'most probable value'. The difference between average value an true value may ~~[illegible]~~ be small in some cases and larg in some other cases. Hence, the reliability of the results depend upon the magnitude of the difference between the ~~[illegible]~~ average value and the true valu

The term error refers to the numerical difference between a measured value and true value. The true value of any quantity is really something we never know.

Errors can be divided into the following two categories.

1. Determinate or Systematic errors.
2. Indeterminate or random errors.

Determinate errors:

The errors which occur due to definite cause are termed as determinate or systematic errors. The determinate errors are tangible and magnitude of which can be determined and there by the measurements can be corrected. Determinate errors are generally unidirectional with respect to the true value. The unidirectional nature may lead to both high and low results with equal probability. Hence determinate error is characterised by the fact that it ordinarily affects to the same degree the results of a series of determination. Determinate errors are often reproducible.

Types of determinate errors:

Determinate errors can be broadly divided into four types

(a) Instrumental and Reagent errors:
(b) Personal errors.
(c) Methodic errors
(d) Additive and proportional errors:

(a) Instrumental and Reagent errors:

Instrumental errors may be due to the following reasons.

(1) unequal lengths of balance arms.
(2) uncalibrated or improperly calibrated fractional weights.
(3) Incorrectly graduated burettes and pipettes etc.

These errors also arise by the attack of foreign materials upon glass-ware, porcelain etc and by the loss in weight of platinum crucibles when strongly heated. These errors can be corrected by careful calibration.

On the other hand reagent errors arise by utilization of impure reagents. An analysis may be inaccurate, if impure reagents are used. The degree of impurity should be measured inorder to perform a suitable correction in the measurement.

(b) Personal Errors:

These errors arise due to the incapability of the experimenter. For example, if an experimenter is unable to distinguish between shades of colour in a volumetric titration, then he may perform an over titration or undertitration. Some examples of personal errors are

(1) Errors in reading a burette
(2) Mechanical loss of material during transfer
(3) Failure to wash and ignite a precipitate properly.
(4) Insufficient cooling of crucible before weighing
(5) Using reagents containing harmful impurities.
(6) Failure in washing the glass-ware properly before analysis.

The error due to prejudice also come under this category. Sometimes even conscious workers may be influenced unconsciously by prejudice. Such a prejudice of a worker leads to error in determination.

(c) Methodic Errors:

These errors arise due to some faults in the method used for the determination of a constituent. These are serious errors encountered usually in chemical analysis. Some examples of methodic errors are,

(1) Solubility of a precipitate in the solution in which it is precipitated.
(2) Decomposition or volatilization of weighing forms of precipitate on ignition or on heating.
(3) Hygroscopic nature of the weighing form.
(4) Co-precipitation and post-precipitation.
(5) occurence of induced or side reactions.
(6) Difference between the observed end point and the stoichiometric end point of a reaction.

(d) Additive and Proportional Errors:

Loss in weight of a crucible in which a precipitate is ignited and errors in weights ~~are~~ is the well known example for additive error.

The absolute value of an error varies with sample size in such a way that the relative error remains constant. For example, in the iodometric determination of an oxidant like chlorate, another oxidising agent such as bromate would cause high results if its presence was unsuspected and not corrected for. If large samples are employed, the value of absolute error increases but the relative error remains constant.

Indeterminate errors:

Indeterminate errors are the errors which cannot be attributed to any known cause. These errors are accidentally encountered in the process of analysis. These errors are difficult to define and their existence can be identified by the small differences in the successive values of a measured quantity when the measurements are made by the same person.

Indeterminate errors can be conveniently divided into two classes.

(1) Variation with in determinate error:

(2) Erratic errors:

(1) Variation with in determinate error:

A determinate error may be caused, if the knife edge of a balance has become worn. If such a balance is used for weighing then the error varies owing to load and hence the determinate error may become an indeterminate error. On the other hand the variation in temperature and humidity may give rise to determinate error. If they are not controlled the determinate error may transform into an indeterminate one.

(2) Erratic errors:

It is very difficult to identify the presence of such errors by the analyser. Some examples of erratic errors are

(a) Fluctuations in balance room can cause erratic errors in weighing

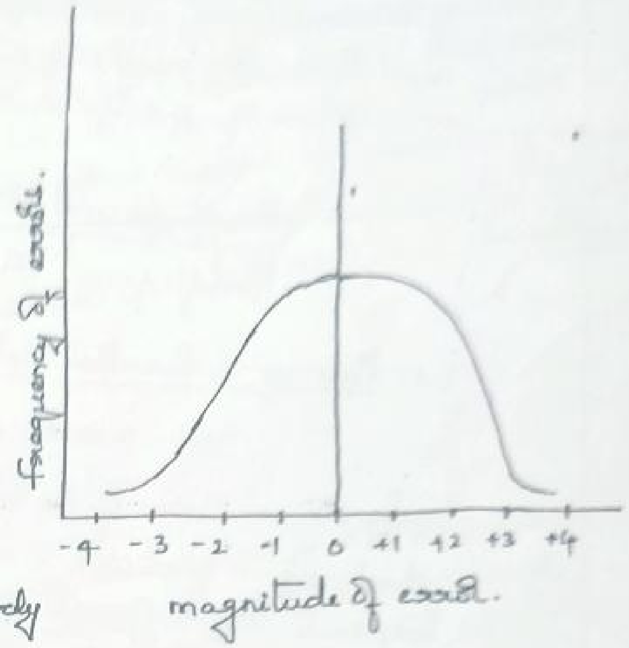
(b) Accidental loss of material during analysis.

The numerical relationship between the magnitude of a random error and the frequency of its occurence can be represented in a curve which is known as normal distribution curve. Few results which can be observed in the normal distribution curve are.

(a) Random errors with small magnitude occur frequently.

(b) Large errors occur at a low frequency rate.

(c) Positive and negative errors are likely to occur.



The shape of the curve reveals the relative precision of the measurements made. A narrow peaked curve with steep slope indicates a relatively high degree of precession, a broad curve indicates a relatively low degree of precision.

Minimisation of Errors:

The determinate errors can be minimised as follows

(a) Running a blank determination

The errors which arise due to utilisation of impure reagents and vessels can be minimised by running a blank determination. A blank is prepared by all the reagents used in the quantitative determination except the unknown component. Thus in making a blank, the sample is omitted. The same process of determination is applied for the blank as used for the sample. The blank measurement is deduced from the sample measurement. Thus the errors due to impure reagents can be minimised.

(b) calibration of apparatus and application of correction:

Proper calibration of all the instruments used in the process of analysis will minimise the errors in the determination. Appropriate correction in the measurements will also minimise errors.

(c) Running a controlled determination:

In this process a standard substance containing the same am of the constituent as in the sample is analysed under the similar experimental conditions. The weight of the constituent X in the unknown can be calculated by making use of the following relationship

$$\frac{\text{Result found for standard}}{\text{Result found for unknown}} = \frac{\text{Wt of constituent in standard}}{X}$$

(d) Running a Parallel determination:

The results obtained in running a parallel determination indica only the precission of the analyser. In this process the analysis is performed in duplicate or triplicate to minimise the errors. If the results of duplicate and triplicate differ from the original by a large extent then extra determinations are carried until concordance is obtained.

(e) standard addition:

This method is applicable for the physical processes used in the quantitative analysis of a chemical constituent. A known amount of the standard constituent which is to be determined is added to the sample and the entire constituent is analysed by the physical process. The difference in the measurement of the ~~sample~~ standard alone and the sample with standard constituent gives the result with minimised error.

(f) Use of independent methods of analysis:

A single constituent can be determined by two or more methods. If the results of the two methods are in good agreement, ~~=~~ it may be said that the values are correct with in small limits of error. For example the strength of hydrochloric acid can be determined volumetrically by a base and gravimetrically by $AgNO_3$ as silver chloride. ~~Th~~ This is a very ~~god~~ good way of verifying the results.

Propagation of Errors:

Sometimes, the numerical results of a measurement is used in conjunction with several other measurements for the quantitative determination of a constituent. The measurement should be precise and the final result should be accurate. If errors are committed in the individual measurements then those errors propogate into the final result. Let us now discuss, how the errors on individual measurements propogate into the final result. An elementary approa of this process was first given by 'Waser' and later a more sophisticated treatme was given by 'Wilson'. But a rigorous treatment useful for analytical chemistry was given by 'Benedetti-pichler'.

Determinate errors:

Addition and subtraction:

Let R be the final result computed basing on the measurements A, B and C. Let α, β and γ be the absolute determinate errors in A, B and C respectively. Let ρ represent the resulting error in R.

If the actual measurements are $A+\alpha$, $B+\beta$ and $C+\gamma$ then the error in individual measurements propogate as follows.

Suppose $R = A+B-C$, changing each quantity by the amount of its error we can write.

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

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

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

Multiplication and Divison:

If multiplication and divison are involved in the computation of the result from the measurements A, B and C then the propogation of error into the result can be given as follows

$$\text{Let } R = \frac{AB}{C}$$

If $A+\alpha$, $B+\beta$ and $C+\gamma$ are the actual measurements then

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

As α and β are small, the product $\alpha\beta$ is very small and hence it can be neglected.

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

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

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

If we consider the relative error, $\frac{\rho}{R}$ obtained by dividing ρ with $R = \frac{AB}{C}$

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

As γ is very small compared with C, the ratio reduces to

$$\frac{\rho}{R} = \frac{\alpha BC+A\beta C-AB\gamma}{ABC}$$

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

Thus it can be concluded that the determinate errors are propogated as follows

(a) When addition & subtraction is involved, the absolute determinate errors are transmitted directly into the result.

(b) When multiplication & division is involved, the relative determinate errors are transmitted directly into the result.

Indeterminate Errors:

In case of determinate errors, it is possible to calculate the magnitude & error committed in each individual measurement, but in case of indeterminate errors it is not possible to calculate the magnitude of error. The uncertainty in each result can be expressed as the standard deviation.

Addition and Substraction:

If addition and subtraction are involved in the calculation of the final result from the individual measurement then the error propogates into the final result as follows

Suppose $R = A + B - C$.

$$\therefore S_R^2 = S_A^2 + S_B^2 + S_C^2$$

Where S^2 is variance and S is standard deviation.

Multiplication and Divison:

In case of multiplication and divison, the error propogates as follows.

$$\left(\frac{S_R}{R}\right)^2 = \left(\frac{S_A}{A}\right)^2 + \left(\frac{S_B}{B}\right)^2 + \left(\frac{S_C}{C}\right)^2$$

Confidence Intervals - Tests of Significance:

In quantitative analysis the numerical results & numerical data obtained in successive measurements may differ among themselves. The difference may be great sometimes even when an exact instrument, method and a capable observer is involved in the process of analysis. The average value of these observations is considered as the most probable value. The most probable value so obtained may not coincide with the true value. The deviation may be due to inaccurate method & apparatus & due to inefficiency of the observer. The difference between most probable value and true value, called the absolute error, may sometimes be with in permissible limits. If the absolute error is beyond permissible limits, such most propable value may not be a representative of the true value.

The significance of absolute error or precision can be tested by certain statistical procedures called as "tests of significance". Student's -t Test can be used to define the confidence intervals of true value. Some important test of significance are.

(i) t - test

(ii) chi square (χ^2) test

(iii) Comparison of two means

(iv) F - test.

(i) t - test or Student's - t : This test was developed by W.S. Gosset writing under the pen name Student.

Let $x_1, x_2, x_3 \ldots x_n$ be a set of n successive observations in the chemical analysis. The significance of the difference between sample mean and assigned true value (μ) can be given as follows

$$\pm t = \frac{(\bar{x}-\mu)}{s}\sqrt{n} \quad \text{———} ①$$

$$\text{where } \bar{x} \text{ or } m = \frac{\sum x_i}{n} \text{ and } s^2 = \left[\frac{1}{(n-1)} \sum (x_i - m)^2\right]$$

After calculating the absolute value of t i.e. $|t|$ in the above equation, it is compared with a table value, say $t_{(n-1)}\alpha$ obtained for $(n-1)$ degrees of freedo and specified level of significance (α)

If $|t| \le t_{n-1}(\alpha)$, then it can be concluded that the difference between sample mean and true value is insignificant and the given average value is within permissible limits.

If $|t| > t_{n-1}(\alpha)$, then the difference between sample mean and true value is significant and the given observations do not represent the true value.

It is possible to define the confidence limits by rearranging equation ①

$$\mu = \bar{x} \pm \frac{ts}{\sqrt{n}}$$

This relation can be used to estimate the population mean (μ) and the probability with which it lies with in certain region centered at $\bar{x}$.

(ii) Chi square (χ^2) test:

If an experiment is performed under identical conditions, then the resulting observations cluster around the mean value. If it is performed under non-identical conditions the results deviate greatly from the mean value. Hence a test can be used to specify the variability in the observations to some assigned value, say σ^2

Let $x_1, x_2, x_3 \ldots\ldots, x_n$ be the n observations. To test the variability (σ^2) in these observations the following test is performed.

$$\chi^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sigma^2}$$

If χ^2 value less than & equal to $\chi^2_{(n-1)}(\alpha)$ [from tables] then the variability in the measurements may be σ^2. If $\chi^2 > \chi^2_{(n-1)}(\alpha)$ then the variability in the measurements may not be σ^2

(iii) Comparison of two means:

In a quantitative chemical analysis, a sample can be analysed by two different methods. Each method is repeated several times and the corresponding mean values of the measurements obtained for two methods are calculated. The significance of the difference between two mean values can be explained by a statistical approach called 'null hypothesis'

As the sample is analysed by two methods, let $\bar{x}_1$ and $\bar{x}_2$ be the mean the measurements obtained for two different methods. Let S_1 and S_2 and n_1 & n_2 be the standard deviations and number of measurements obtained for the two metho

The t value for the null hypothesis can be calculated by using the formula

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{S}\sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

Treating S_1 and S_2 as same.

This calculated t value is compared with table 't' values i.e $t_{(n_1+n_2-2)}^{(\alpha)}$ obtained for (n_1+n_2-2) degrees of freedom and specified level of significance.

If the calculated t values is less than or equal to table 't' value then the difference between the two means is insignificant and the two means are same wit certain probability.

If the calculated 't' value is greater than table 't' value then the difference between two means is significant and null hypothesis is incorrect.

(iv) F - test:

The analysis of a quantity may be carried out on two different apparat or in two different laboratories or different environmental conditions. In such a situation for testing the equality of variability in observations recorded on two different occasions we make use of F-test or variance ratio test.

Let $x_1, x_2, x_3, \dots x_{n_1}$ be n_1 observations recorded one one occasion and let $y_1, y_2, y_3 \dots y_{n_2}$ be n_2 observations recorded on other occasion. For testing the equality of variability, calculate

$$F = \frac{S_1^2}{S_2^2} \quad \left[S_1^2 > S_2^2 \right]$$

Where $S_1^2 = \left[\frac{1}{(n_1-1)}\Sigma(x_i-\bar{x})^2\right]$ and $S_2^2 = \left[\frac{1}{(n_2-1)}\Sigma(y_i-\bar{y})^2\right]$

If the calculated value of F is less than the table value of F for (n_1-1) and (n_2-1) degrees of freedom at α level of significance then the observations recorded on the two occasions are equally variable. If $F > F_{(n_1-1)(n_2-1)}(\alpha)$ then we say that the observations have variability.

Rejection of data:

It has been observed that in a series of measurements, one of the numerical value may deviate from the other values to a greater extent. In such a situation, it is rejected, Inorder to get a mean value nearer to true value.

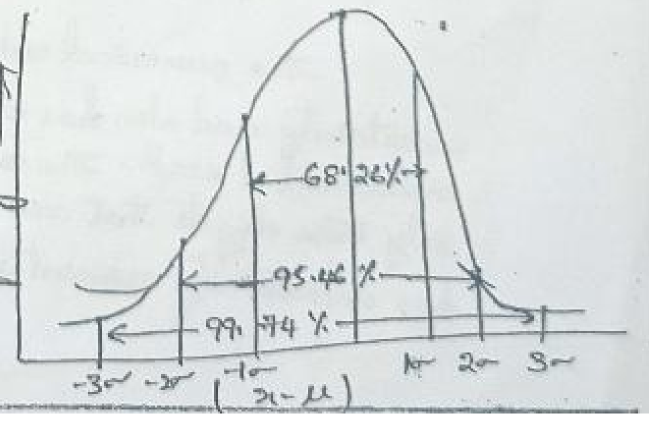
If the number of measurements are more, then there is a greater probability for the average of measurements to be nearer to the true value. It is not so easy to reject a measurement from a series. For rejecting a measurement, the following rules should be followed,

(a) Average deviation:

In a series of measurements, omit the doubtful value and determine the mean ~~and~~ of average deviation of the retained values. If the deviation of the suspected value from the mean is atleast 2.5 times the average deviation i.e. $x \geq 2.5\bar{d}$, then the rejection can be regarded as mathematically justified.

(b) Standard deviation:

If 100 measurements are made using a sample then the normal distribution curve is prepared as shown in the figure. The curve is plotted taking frequency of occurrence of a measurement along y-axis and the magnitude $(x-\mu)$ in terms of σ along x-axis. From the curve it is clear that most of the values (68.26%) fall with in the range of $\pm 1\sigma$ and 95.46% of measurements fall with in the range of $\pm 2\sigma$



and 99.74% of measurements fall in the range of ±3σ. Thus the value of can be used as criterion for rejecting a measurement.

(c) Q-test: This test is used to ~~of~~ justify whether a data can be rejected or not. The Q-test is carried for a population of 3 to 10 measurements and is applied at 90% confidence level.

The measurements are ~~fast~~ first arranged in ascending order and then the range of the measurements [Mn-M1] is calculated. Find the difference betw the suspected result and its nearest neighbour. Divide this difference by the range.

$$\text{i.e. } Q = \frac{\text{Difference of the suspected result from its nearest neighbour}}{\text{Range.}}$$

After calculating Q, it is compared with table value. If computed value is greater than table value, then the measurement can be discarded with 90% confidence.

Significant figures and Computation rules:

The measure of deviation is very significant when it is used to express the degree of precision. It has less significance when only two or three values have been obtained in the process of analysis. In such a situation the precision of a numerical value is best indicated by the number of significant figures used in expressing the value. This is only a rough idea of precision of the results.

The numerical value of an experimental result is associated with some uncertainty and also has a range of reliability that what we call the experimental error. The ~~ratio~~ reliability of a number is indicated by including only those digits that are known with certainty, plus one additional digit that has uncertainty associated with it.

For example when the mass of an object can be measured nearest to 0.1g using a balance then there is an amount of uncertainity in measuring 0.1g. If the weight of an object measured from such balance is 27.3g then 2 and 7 are known exactly but the measurement associated with 3 has some uncertainity. If the same object is weighed on an analytical balance then the uncertainity is associated with the fourth digit after decimal

A number is an expression of a quantity. A figure or digit denotes any one of the ten numerals. A digit alone or in combination, serves to express a number. A significant figure is a digit which has some practical meaning i.e. it is a digit which expresses the amount of the quantity in the place. Hence, significant figures in a number are all the digits whose values are known with reasonable certainity plus one additional digit whose value is uncertain.

A zero is not significant when it serves to locate the decimal point. For example, in the number 0.05800, the zeroes to the left of the digit 5 are not significant, since they are used only to locate the decimal point. But the zeros beyond the digit 5 are significant.

Examples:

(a) 2600 $= 2.6 \times 10^{3}$ (two)

(b) 80400 $= 8.04 \times 10^{4}$ [Three]

(c) 0.004 $= 4 \times 10^{-3}$ [one]

(d) 0.1046 $= 1.046 \times 10^{-1}$ [Four]

(e) 7050 $= 7.05 \times 10^{3}$ [Three]

(f) 0.0002008 $= 2.008 \times 10^{-4}$ [Four]

(g) 10772000 $= 1.0772 \times 10^{7}$ [Five]

(h) 0.010250 $= 1.0250 \times 10^{-2}$ [Five]

(i) 108642000 $= 1.08642 \times 10^{8}$ [Six]

(j) 0.002340 $= 2.340 \times 10^{-3}$ [Four]

(k) 56730 $= 5.673 \times 10^{4}$ [Four]

Computational Rules:

While expressing an experimental result only significant figures should be used. In expressing the result by means of significant figures certain rules are followed, which are known as Computational rules.

Rule 1: Never retain more than one doubtful digit, in expressing an experimental result.

Rule 2: Retain as many significant figures in a result which will contain only one uncertain figure.

For example, a volume between 30.5 ml and 30.7 ml should be expressed as 30.6 ml, but not as 30.60 ml. Since the latter would indicate that the value lies between 30.59 and 30.61 ml.

Rule 3: The rules for rejecting superflous [unnecessary] digits are as follows

(a) If the last digit dropped is greater than 5, then increase the last retained digit by one.

For example in 4.863, if 6 is to be discarded, the retained digit 8 is increased by one i.e. 4.9. This is known as rounding up

(b) If the last discarded digit is less than 5, then leave the last retained digit unchanged. This is known as rounding down.

(c) If the discarded digit is 5 followed by zeros only or by no other digit at all then round up if the last digit retained is odd and round down if the last digit retained is even.

For example, the number 6.7450 is rounded to 6.74
" " 6.8550 " " to 6.86.

Rule 4: In addition or subtraction, the final result should posses as many digits after decimal as there in one of the numbers with fewest decimals.

For example; 0.01 + 55.64 + 1.0572 = 56.71.

Rule 5: In multiplication & division the percentage precision of ~~experimental~~ the product & quotient cannot be greater than the percentage precision of the least precise factor entering into the computation.

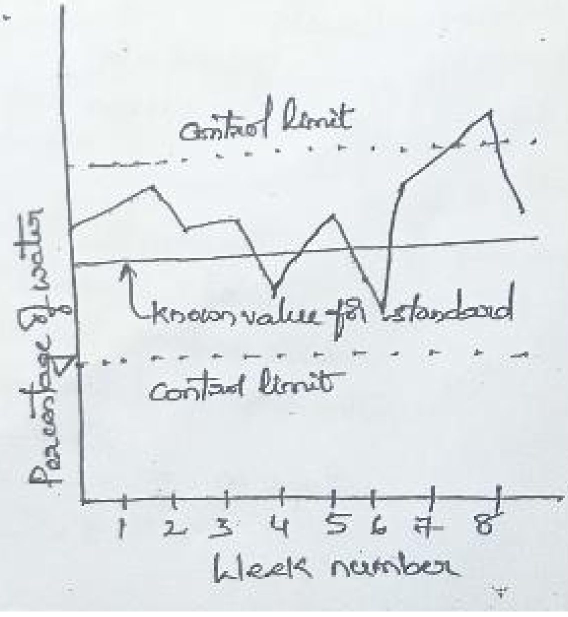
For example the product of three terms 0.0121, 25.64 and 1.05782 can be written as $0.0121 \times 25.6 \times 1.06 = 0.328$

i.e. In multiplication & division of experimental numbers, the final result has only as many significant figures as the factor with least number of significant figures.

Control charts:

The method of control chart was developed for assuming the quality during large-scale manufacturing operations. In large scale productions, it is not possible to inspect each item (say, razor blades & ball bearings) and in some cases the quality test is destructive and hence cannot be applied to each specimen. In such a situation the quality of the products can be assessed by making use of control charts. The control chart method can be briefly discussed as follows.

Suppose a company manufactures some chemical material which consists of some amount of water. The quality of the chemical material depends upon the percentage of water in the product. The quality of the product should be analysed every day and the report is represented in the form of a chart called control chart. The control chart is drawn taking percentage of water along y-axis and no of weeks along x-axis. The horizontal line in the chart represents the percentage of water in the standard sample. The dotted lines represent the control limits. Analytical results falling outside these limits are considered to be resulting from certain disorders in the manufacturing unit. When results fall with in the limits, the fluctuations are only random and indeterminate. When results fall outside the control limit then the process is shut down and the trouble is rectified.



Method of least Squares

Certain analytical procedures involve the usage of physical parame for the quantitative estimation. The measure of physical parameter is directly proportional to the concentration of sample under analysis. One example of such physical method involves the measure of absorption of a solution using spectropho meter.

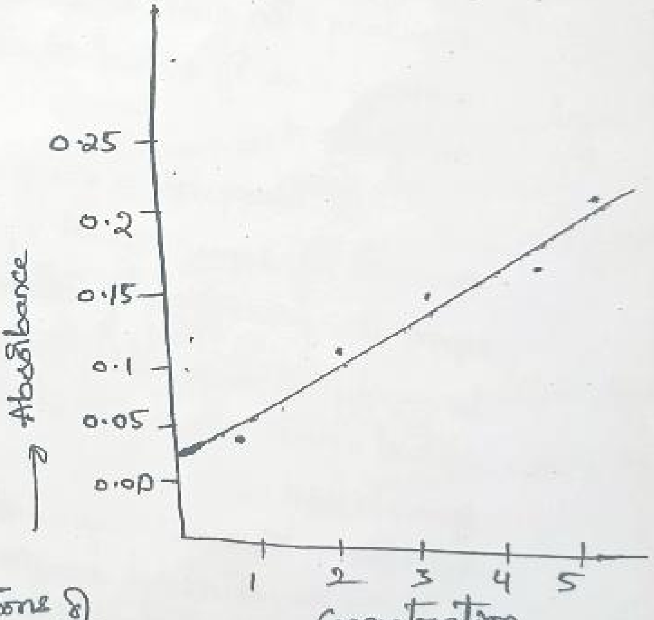
A series of solutions of known concentration are prepared and the corresponding absorption of each solution is measured using spectrophotometer. The absorption of each solution is plotted against the corresponding concentration t give a standard curve or calibration graph. In many cases there is a linear relationship between concentration and instrument response. The points in the curv may be sometimes zig-zag and not fall exactly on a straight line due to some indeterminate errors. Then it becomes difficult to draw a straight line throug these points. Inorder to minimise the error in determining the concentration of an unknown using the calibration graph a best straight line should be drawn through the points of standard solutions.

Alternately a statistical method can be used to calculate the slope and intercept of the best straight line from the data obtained. This statistical method is called 'method of least squares'.

The equation of a straight line is

$$y = mx + b$$

Where m is slope
b is intercept



The sum of the squares of the deviations of the actual instrument readings from the correct values are minimized by adjusting the values of the slope 'm' and the intercept 'b'. The slope and intercept are adjusted using the formulas

$$\text{Slope } m = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} \qquad \text{Intercept } b = \frac{\Sigma y - m \Sigma x}{n}$$

Regression analysis:

Regression ~~at~~ analysis is a branch of statistical theory that is widely used in almost all the scientific disciplines. The dictionary meaning of the term 'regression' is the act of returning or going back. The method of regression analysis reveals average relationship between two or more variables in terms of the original units of the data. This method is useful in estimating the values of dependent variable from the values of independent variable.

Regression line:

If a case of two variables x and y is considered, then ~~two~~ two regression lines are possible. The two regressions are, regression of x on y and the regression of y on x. The regression of y ~~and~~ on x gives the most probable values of y for given values of x.

It is important to note that the regression lines are drawn on least squares assumption method, which stimulates that the sum of the squares of the deviations of the observed y values from the fitted line shall be minimum.

Consider the regression equation of y on x

i.e. ~~[illegible]~~ $y = b + mx$

where b and m are constants. These constants are called the parameters of the line. If the values of either or both of them is changed, another line is determined. The parameter 'b' determines the level of the fitted line (i.e. the distance of the line directly above or below the origin). The parameter 'm' determines the slope of the line. If the values of b and m are obtained, the line is completely determined. The values of b and m can be obtained by the method of least squares. The method least squares states that the line should be drawn through the plotted points in such a manner that the sum of the squares of the deviations of the actual y values from the calculated y values is the least.

i.e. $\Sigma(y-y_c)^2$ should be minimum

where y_c is computed value from the relationship for a given x.

$\therefore \Sigma(y-y_c)^2$ should be minimum

or $\Sigma(y-b-mx)^2$ should be minimum

Let $S = \Sigma(y-b-mx)^2$

Differentiating partially w.r.to b.

$$\frac{\partial S}{\partial b} = 2\Sigma(y-b-mx)(-1)=0$$

$$\Rightarrow \Sigma(y-b-mx)=0.$$

$$\Rightarrow \Sigma y = \Sigma b + m\Sigma x \Rightarrow \Sigma y = Nb + m\Sigma x \quad \text{———} \ ②$$

Differentiating partially S w.r.to m

$$\frac{\partial S}{\partial m} = \Sigma(y-b-mx)(x)=0.$$

$$\rightarrow \Sigma(y-b-mx)(x)=0$$

$$\Rightarrow \Sigma xy = b\Sigma x + m\Sigma x^2 \quad \text{———} \ ③$$

Equations ② and ③ are known as normal equations.

$$\Sigma y = Nb + m\Sigma x \Rightarrow \Sigma y - m\Sigma x = Nb$$

$$\Rightarrow b = \frac{\Sigma y - m\Sigma x}{N}$$

Substituting the value of b in $\Sigma xy = b\Sigma x + m\Sigma x^2$

$$\Sigma xy = \left[\frac{\Sigma y - m\Sigma x}{N}\right]\Sigma x + m\Sigma x^2$$

$$\Rightarrow \Sigma xy = \frac{\Sigma x\Sigma y - m(\Sigma x)^2}{N} + m\Sigma x^2$$

$$\Rightarrow \Sigma xy = \frac{\Sigma x\Sigma y}{N} - m\frac{(\Sigma x)^2}{N} + m\Sigma x^2$$

$$\Rightarrow \Sigma xy - \frac{\Sigma x\Sigma y}{N} = m\left[\Sigma x^2 - \frac{(\Sigma x)^2}{N}\right]$$

$$\Rightarrow m = \frac{\Sigma xy - \frac{\Sigma x\Sigma y}{N}}{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}$$

By knowning the values b and m, it is possible to draw a best line from the obtained data.

Distribution of random Errors:

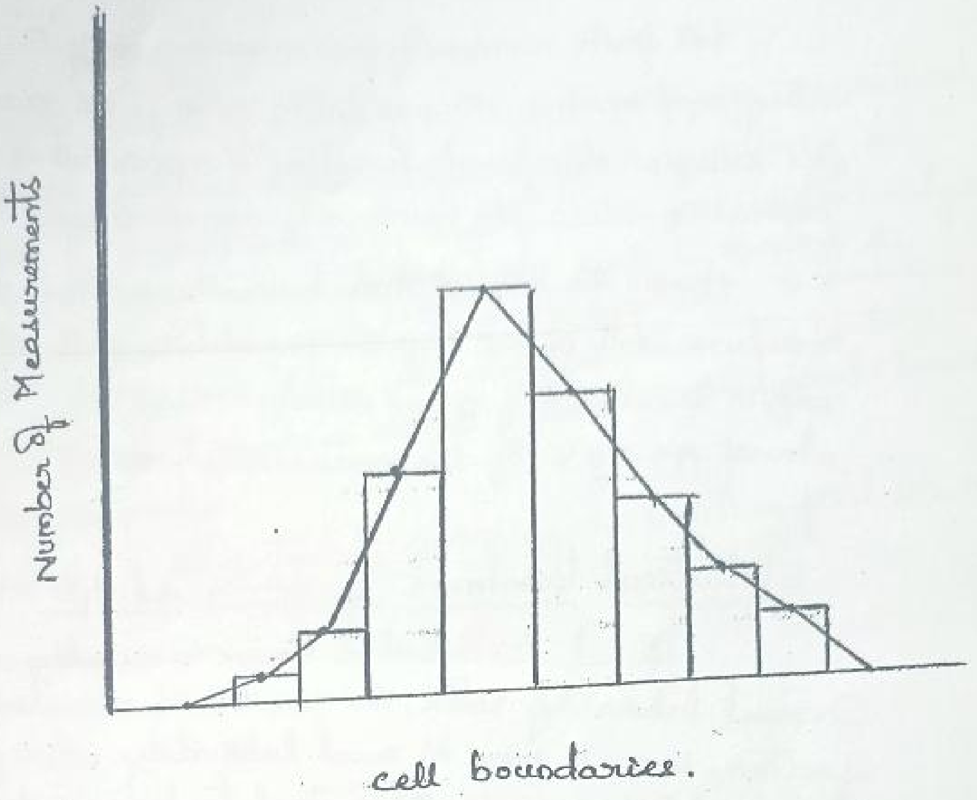
The determinate errors which occur in the measurement due to definite cause can be corrected. Even after correcting the error some deviations were observed in the measurements. These deviations in the measurements are due to random errors. Results that are scattered from a minimum to maximum are best treated by certain statistical techniques.

Frequency Measurement:

(1) Suppose if a data consists of 60 measurements, then it is not easy to remember all the measurements as they are arranged randomly. In such situation, the measurements are arranged in an order from lowest to highest. Even, when they are arranged in an order it is not possible to remember all the measuremen Hence, some more compactness is required in the data.

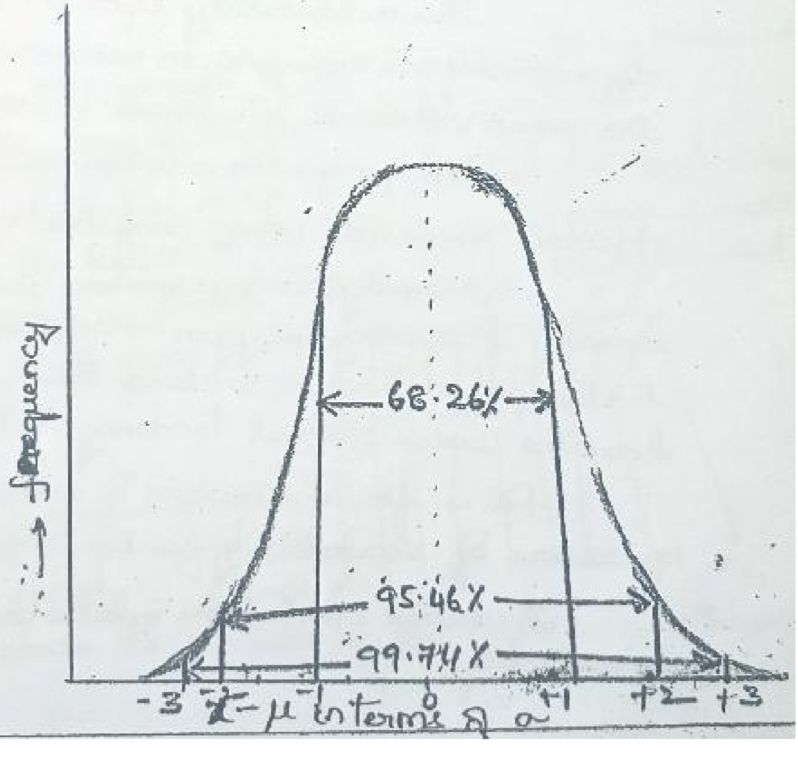
(2) The data can be condensed by grouping the measurements into cells. In this process the range from the lowest to highest value is divided into a convenient number of intervals or cells. After ~~diving~~ dividing them into cells, number of values falling in each cell are counted. A cell is usually represented by boundaries and a confusion may arise if the measured values ~~at~~ are directly taken as boundaries for the cells. Such confusion is avoided by choosing cell boundaries ~~halfy~~ halfway between possible observed values. Though there is some loss of information in this process, it is being compensated by the increased efficiency with which one can understand the significance of the data.

(3) The number of measurements falling ~~in~~ in each cell can be diagramaticall represented ~~as~~ a histogram. A histogram is drawn taking number of measurements along y-axis and cell boundaries along x-axis. A frequency polygon is constructed by plotting frequencies at cell midpoints and connecting the points with straight lines.



Normal Error Curve.

The limiting case approached by the frequency polygon as more and more replicate measurements are performed is the normal or Gaussian distribution curve. Normal error curve usually pertains to infinite population. A finite number of measurements are considered to be sample drawn from large hypothetical population.



The equation for the normal curve can be written as follows

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

Where y represents the relative frequency with which the value 'x' is found in the data. μ and σ are population parameters. μ is mean of infinite population, σ is standard deviation of infinite population.

For finite number of measurements, say 'n', the sample mean is $\bar{x}$. The $\bar{x}$ value approaches the population mean μ as n approaches to infinity. Similarly, the sample standard deviation 's' approaches the population standard deviation when the number of measurements become infinite.

From the normal distribution curve, it is clear that nearly 68.26% of the measurements of an infinite population fall with in the limits $\pm 1\sigma$, whereas with in the limits of $\pm 2\sigma$ nearly 95.46% of the measurements exist and almost 99.74% of the measurements are within the limits $\pm 3\sigma$.

Statistical treatment of finite sample:

Normal distribution curve is usually obtained for infinite population. In real laboratory work, the process of obtaining infinite measurements is only fiction. Hence in case of real laboratory work the following statistical methods are used to measure the central tendency and variability.

Measure of central tendency and variability:

In case of finite number of measurements, the mean value is the best measure of central tendency. Because around this mean value all the measured values will cluster around.

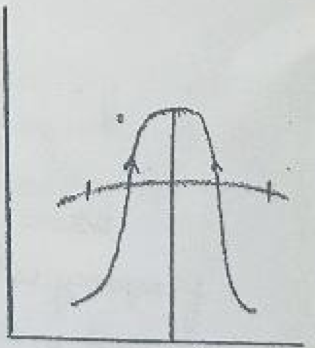
The reliability upon the mean as a measure of central tendency decreases with increase in number of measurements. Nearly for n measurements the mean value is $\sqrt{n}$ times reliable as a measure of central tendency.

Median is also another measure of central tendency. But it is a less efficient measure when compared with mean.

Generally two parameters μ and σ are required to specify frequency polygon. From the frequency distribution of the two populations, it is clear that, both the populations may have the same central tendency, but differs in variability.

For a finite number of measurements, the simplest measure of variability is the range (R).

Average deviation is also a measure of variability.



Two populations wi same central tendenc m, but different variabilities.

Sampling:

Sampling is the process of extracting a small portion of the material from a large quantity, which is a true representative of the composition of the whole material. The process of sampling is very significant in quantitative analysis. If it is not carried out properly, the time and ~~labo~~ labour spent in making the analysis of the sample may be completely wasted. It is also important to remember that the reputation of the analytical ~~chemists~~ chemist ultimately rests on his ability to produce correct results. Inorder to produce correct results the analytical chemist must carry out sampling properly.

Basing on the nature of the material, sampling can be classified into the following types.

(1) Sampling of solids.
(2) Sampling of liquids
(3) Sampling of gases.

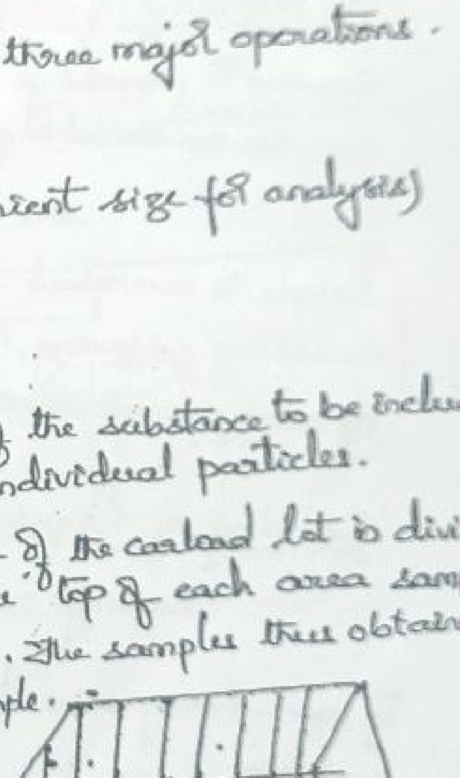
The sampling procedure mainly includes three major operations.

(a) Collection of gross sample
(b) Reduction of gross sample (i.e. to the convenient size for analysis)
(c) Preparation of the sample for analysis.

(1) Sampling of solids:

In case of solid substances, the amount of the substance to be included in the gross sample depends upon the size of the individual particles.

(i) From car load lot: The upper surface area of the carload lot is divided into 8 equal areas. Beginning from the bottom to the top of each area sample is ~~colted~~ collected at ten positions using a shovel. The samples thus obtained from the eight areas are combined to get gross sample.



(ii) From Bags, ~~cart~~ trays etc:

If the material is uniformly composed then the gross sample may be obtained by taking ten equal specified increments uniformly distributed along the diagonals across the surface of the container. otherwise, the entire content of every n^{th} container is taken and combined to get the gross sample.

(iii) Powdered samples:-

The powdered materials are usually sampled using a specially designed ~~the~~ thief or by means of a shovel or dipper etc.

(iv) From wet materials:

If a solid material is uniformly suspended in a liquid then the substance can be sampled easily. If the material is non-uniformly suspended in a liquid then the sample is collected from several vertical sections by means of a specially designed thief and combined to get the gross sample. The collected portions must be kept in a moisture tight container.

(v) From hygroscopic or Deliquescent materials:- The sample is collected by a specially designed thief or a small scoop etc from an air tight container.

(vi) Sampling of non ferrous metals and alloys: The saw dust obtained by sawing or punchings or drillings obtained during the process of cutting the metal can be combined to get gross sample.

The next operation after sampling is the reduction of gross sample to convenient size. The reduction of sample to convenient size includes crushing, sieving, coning, quartering and long pile & alternate shovel method

Sampling of liquids:

Sampling incase of homogeneous liquids is easy and is difficult in case of non-homogeneous or heterogeneous liquids. Sampling of liquids is divided into -3- classes.

(i) Sampling of liquids in quiescent state
(ii) Sampling of liquids in motion.
(iii) Sampling of special liquids.

(i) Sampling of liquids in quiscent state:

Liquids are usually contained in various types of containers and samples can be obtained from different potions by means of a suitable sampli apparatus.

(a) from storage tanks, cans etc:-

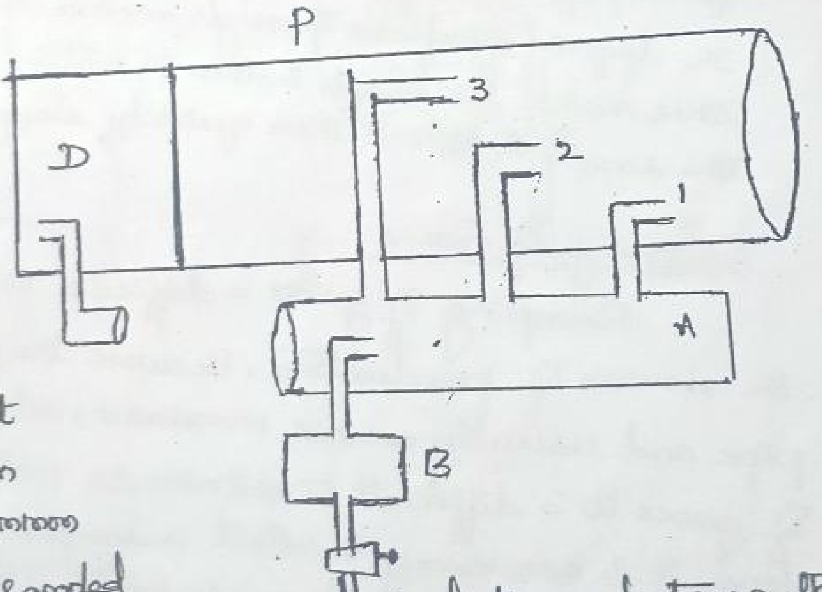
1. If the liquid is thoroughly agitated & is reasonably uniform then samples can be collected from specified potions and can be transferred to a container.

2. If the liquid is heterogeneous i.e. if it permitts the formation of zones of different compositions then the samples are ~~co~~ collected from all the zones (top to the bottom) by means of a sampling bottle having a lid.

(b) From drums, carboys, barrels etc: In case of drums, carboys, barrels etc samples can be obtained by means of a thief & a scoop from various specified potions.

(ii) Sampling of liquids in motion: Sampling of liquids in motion can be carried out by a special apparatus.

Using this apparatus, a definite fraction of varying flow can be collected over a required period of time.



The apparatus consists of 3 pipes 1, 2, 3 bent at 90° angle with opening, facing the flow ~~[illegible]~~ at different levels ~~[illegible]~~ inserted in pipe 'p' and are connected to a common pipe A. The flow through A is sampled by B the opening of which faces the flow at a point halfway between the centres and bottom of the pipe A. Pipe B is fitted with a stop cock C where the amount of the sample flow through B may be regulated.

(iii) Sampling of Special liquids:

(a) Liquids containing insoluble matter in suspension:— usually such liquids are of two types, one containing uniform suspension during the period for of sampling and second consists of sediments settling rapidly, it impossible of uniform distribution.

In the first type, the liquid is thoroughly mixed and the sample can be obtained by a scoop or thief. In the second type, the liquid is thoroughly mixed and complete column of the liquid is rapidly taken by a special thief By repeating this operation several times a fairly representative sample is obtained.

(b) Liquids Composed of immiscible layers:— The layers of immiscible liquids are well separated and portions of each layer are combined in proper ratios of the material in a container of uniform cross-section. Then proper proportions are taken from each layer after measuring the depth of each layer.

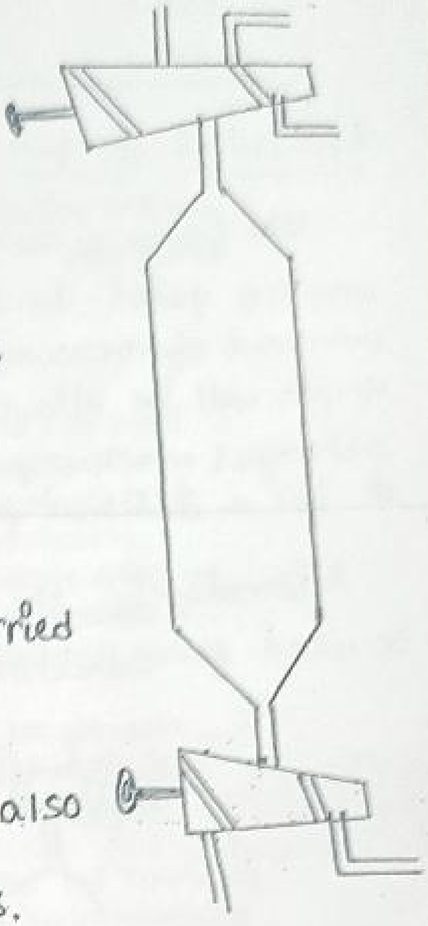
(c) Volatile liquids: volatile liquids are sampled by continuous flow method. In this the liquid is discharged by means of a 'siphon' from container. A portion of the liquid may be seen to the bottom of the sampling bottle until the liquid overflows through another tube from the top into another bottle. This makes the sample bottle to contain maximum amount of the volatile liqu
The sample bottle is then quickly stoppered and sealed.

Sampling of Gases:

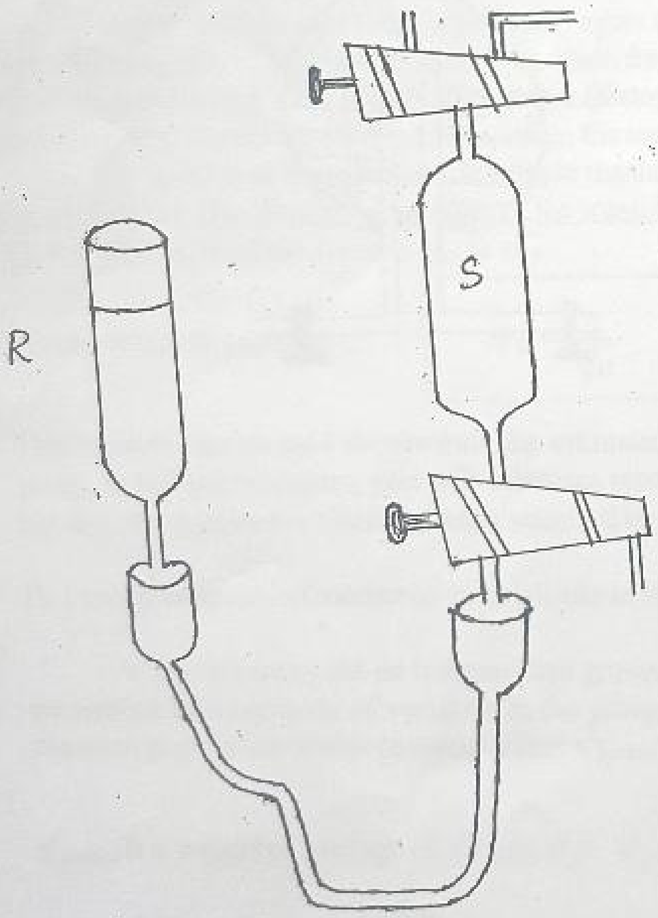
Sampling of gases is difficult when it passes through a pipe from the source of production. Because the gases flow with varying velocities in. pipe and usually possess maximum velocity at the centre of the pipe. In case of gases it is difficult to obtain an instantaneous representative sample. Hence it is necessary to collect a sample over an extended period of time, such that there is high probability for a true sample.

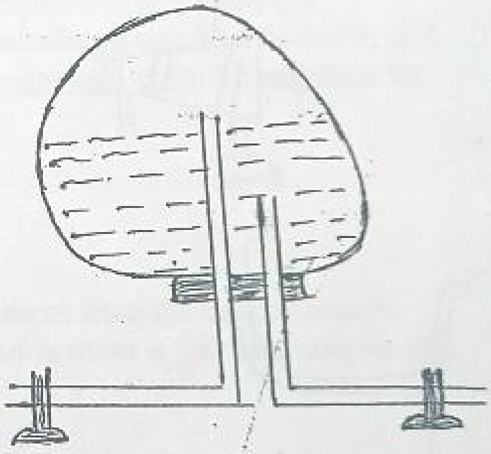
Appratus for gas Sampling :-

The figure represents a sample pipette for sampling gases. In the first step the pipette is evacuated by means of mercury. In the next step the gaseous sample will be allowed into the pipette and which discharges mercury. The sampling can be carried out for a desired period of time.



Alternatively the following appratus can also be used for collecting gaseous samples.



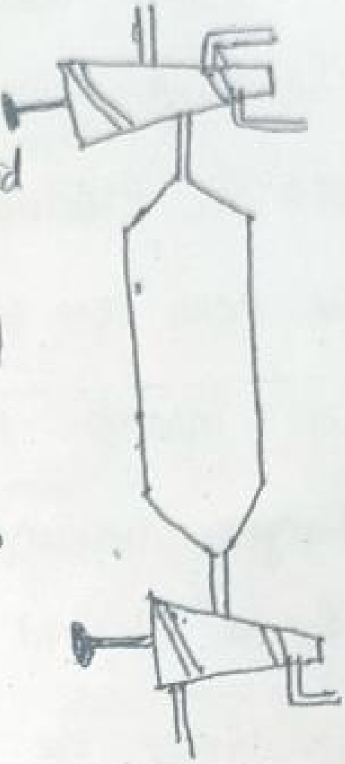


R - Reserviour of mercury ;

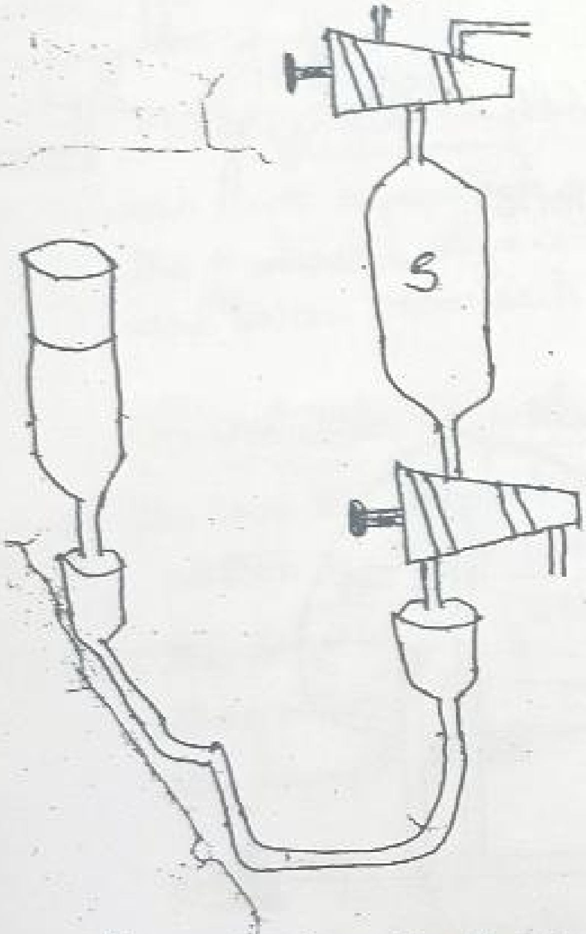
S - Sampling tube

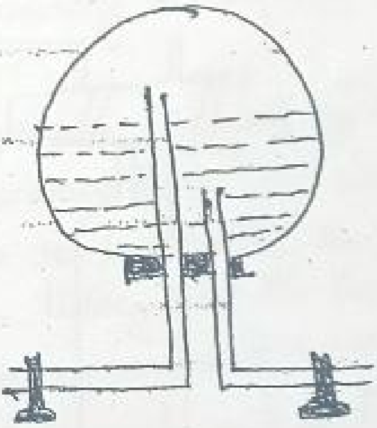
Apparatus for gas Sampling:-

The figure represents a sample pipette for mpling gases. In the first step the pipette is evacuated means of mercury. In the next step the gaseous mple will be allowed into the pipette and which scharges mercury. The sampling can be carried t for a desired period of time.



Alternatively the following apparatus can also be used for collecting gaseous samples.





R — Reservoir of Mercury

S — Sampling tube.