Indira Gandhi National
Open University
School of Sciences

BCHET-141
ANALYTICAL METHODS IN CHEMISTRY

VOL.

1

QUALITATIVE AND QUANTITATIVE ANALYSIS, CHROMATOGRAPHY AND ION EXCHANGE

ignou
THE PEOPLE'S
UNIVERSITY

BCHET-141: ANALYTICAL METHODS IN CHEMISTRY

In our earlier courses especially in lab courses, we mainly emphasised basic analytical methods for the qualitative analysis using titrimetry and qualitative analysis involving identification of inorganic and organic compounds). This course “Analytical Methods in Chemistry” is basically its further extension. In this course, our focus will be on sampling techniques, statistical techniques, separation methods, thermal methods, electroanalytical methods and spectroscopic methods. These methods are commonly used for separation, purification and characterization of organic and inorganic compounds. Analytical methods which you are going to study in this course have wider applications in the field of medicine, industry, environmental quality control, food industry and forensics, archeology and space science.

This course has four blocks. Block 1 has 4 Units. The first two units i,e. Units 1 and 2 deal with sampling, accuracy, precision, types of measurements and evaluation of analytical data. Units 3 to 4 deal with the techniques of solvent extraction and its applications. Block 2 has three Units. In all these three units i.e. Units 5, 6 and 7 there is a detailed description of chromatographic methods of separation.

Block 3 consists of three units. Unit 8 on ‘Thermal Methods of Analysis’ deals with the thermal methods and their applications in characterisation of inorganic and organic compounds. Unit 9 and Unit 10 deal with three important electroanalytical methods i.e. potentiometry, pH metry and conductometry.

The Block 4 titled, ‘spectroscopic methods of analysis’ deals with basic molecular and atomic spectroscopic methods based on the consequences of the interaction of radiation with matter. The block consists of five units. The first of these units deals mainly with the electromagnetic radiation in terms of its nature, characteristics, and properties. The second and third units deal with molecular spectroscopic methods, *viz*., UV-VIS spectrometry and IR spectrometry, respectively in terms of the origin, principles and analytical applications. It also gives an account of the instrumentation and the qualitative and quantitative applications. The fourth and fifth units, on the other hand, deal with the atomic spectroscopic method *viz*., flame atomic absorption and flame atomic emission spectrometry respectively. These cover conceptual background, spectral characteristics, instrumentation and applications of these techniques.

Expected Learning Outcomes

After studying this course, you should be able to:

- describe the sampling of water and soil samples;
- discuss error and type of errors in measurements;
- explain the differentiate between accuracy and precision;
- define some common terms of statistical calculations;
- discuss the principle of liquid-liquid extraction;

- describe various applications of extraction techniques;
- discuss the classification of various chromatographic techniques according to the different criteria;
- state the importance of paper chromatography, column chromatography and ion exchange chromatography;
- explain the principle of paper chromatography, adsorption chromatography and ion exchange chromatography,;
- discuss the applications of paper chromatography, adsorption chromatography and ion exchange chromatography;
- describe basic thermal methods of analysis;
- explain the basic principles of the thermogravimetric analysis;
- illustrate applications of the thermal methods in the characterisation of inorganic and organic compounds;
- classify electroanalytical methods in different groups;
- discuss the basic principles of potentiometry and pH metry;
- illustrate the applications of potentiometric and pH metric methods;
- describe the basic principle of conductometric methods of analysis;
- describe the wave and particle nature of electromagnetic radiation and define the parameters that characterise electromagnetic radiation;
- explain the origin and the characteristics of UV-VIS, Infrared, FAAS and FAES spectra;
- explain the principles of different types of spectrometers used in UV-VIS, Infrared, FAA and FAE spectrometric methods;
- outline the quantitative methodologies for UV-VIS, FAA, and FAE spectrometric methods; and
- enlist the applications of UV-VIS, IR, FAA, and FAE spectrometric methods.

Indira Gandhi National
Open University
School of Sciences

BCHET-141
ANALYTICAL METHODS IN CHEMISTRY

Block

1

QUALITATIVE AND QUANTITATIVE ASPECTS OF ANALYSIS

Course Design Committee

Prof. N.K. Kaushik, (Retd.)
Department of Chemistry,
University of Delhi, New Delhi

Prof. B. S. Saraswat (Retd.)
School of Sciences, IGNOU,
New Delhi

Prof. Nafisur Rehman,
Department of Chemistry,
A.M.U. Aligarh

School of Sciences,
IGNOU, New Delhi 110068

Prof. M.S. Nathawat
Prof. Sunita Malhotra
Prof. B.I. Fozdar
Prof. Javed A. Farooqi
Prof. Sanjiv Kumar
Prof. Lalita S. Kumar
Prof. Kamalika Banerjee

Block Preparation Team

Prof. Sunita Malhotra
School of Sciences, IGNOU
New Delhi

Prof. N.K. Kaushik, (Retd.) (Editor)
Department of Chemistry,
University of Delhi, New Delh

Prof. Javed A. Farooqi
School of Sciences, IGNOU
New Delhi

Prof. Kamalika Banerjee
School of Sciences, IGNOU,
New Delhi

Course Coordinator: Prof. J. A. Farooqi

Print Production

Mr. Rajiv Girdhar
Assistant Registrar (Pub),
MPDD, IGNOU, New Delhi

Mr. Hemant Kumar Parida
Section Officer (Pub),
MPDD, IGNOU, New Delhi

Acknowledgements: Sh. Sarabjeet Singh for CRC preparation

Material partially adapted from the MCH-001 and MCH-002 Course of PGDAC Programme.

March, 2022

ISBN:

Printed and published on behalf of Indira Gandhi National Open University, New Delhi by the Registrar, MPDD, IGNOU.

Printed at:

BLOCK 1: QUALITATIVE AND QUANTITATIVE ASPECTS OF ANALYSIS

This is the first Block of Analytical Methods in Chemistry and addresses to the basics of Sampling techniques and data analysis. The first unit deals with Sampling- the most important initial step crucial for any analytical measurement. In this unit we have also covered concepts like accuracy, precision types of measurements. The second unit is devoted to and evaluation of analytical data. Units 3 and 4 deal with the techniques of 'Solvent Extraction' and cover the general principles, classification of extraction systems, extraction equlibria, factors influencing extraction & metal ion separations as well as choice of organic phase for extractions.

Expected Learning Outcomes

After studying this block, you should be able to:

- explain the importance and significance of small sample that you receive in the laboratory;
- describe the sampling of water and soil samples;
- understand error and type of errors in measurements;
- explain the differentiate between accuracy and precision;
- define some common terms of statistical calculations;
- understand the normal distribution of errors;
- estimate the Statistical test of data namely t-test and F-Test;
- get a guideline concerning whether results should be retained or rejected;
- calculate the confidence limits and confidence interval.
- describe liquid-liquid extraction and its importance;
- discuss the principle of liquid-liquid extraction;
- explain the efficiency of extraction technique;
- differentiate between distribution coefficient and distribution ratio and discuss the mechanism of extraction, *i.e.* extraction by solvation and chelation, and
- describe various applications of extraction techniques.

ignou
THE PEOPLE'S
UNIVERSITY

UNIT 1

SAMPLING AND ERROR IN CHEMICAL ANALYSIS

Structure

1.1 INTRODUCTION

Sampling is not a modern development; it has been practiced from the early age of civilization. The caveman made use of a sampling procedure to select stones from which he made his weapons. Today, a common buyer interested to purchase some food material (say sweets) from a new market place may taste a bit of sweet before buying it. It is sampling in a crude manner, because he was assuming that the remainder of the food material (sweet) would have a

taste similar to that which he had tasted. With the introduction of statistical methods the sampling has been greatly advanced and can now be considered as an accurately controlled process.

The process of sampling consists of drawing a very small portion of the material as a representative of the bulk of the material. It may involve an elaborate array of operations, such as, crushing, grinding, subdivision, etc. The collected sample should be suitable for convenient transport and laboratory testing.

The purpose of this unit is to provide sufficient information to the student to enable him to examine the factors affecting the reliability of results and understand the contributions of errors, their types. In the next section we have discussed common terms use for evaluation of analytical data. Like mean, medium, mode, deviation, average deviation, probable deviation and range. In the last section we will learn some basic concept of analysis of chemical compounds.

Expected Learning Outcomes

After studying this unit and having the experiments performed, you should be able to:

- understand the importance and significance of small sample that you receive in the laboratory,
- get knowledge of sampling process in general for gas, liquid, and solid materials,
- understand the proper transmission and storage of samples,
- understand the sampling of water and soil samples,
- understand error and type of errors in measurements,
- explain the differentiate between accuracy and precision and
- define some common terms of statistical calculations.

1.2 THEORY OF SAMPLING

A sample is a portion of the material taken from the consignment in such a way that it possesses the essential characteristics of the bulk an ideal sample would be identical in all of its intensive properties of the bulk of the material from which it is removed.

Thus, a single drop, removed from a tanker of a completely mixed liquid, may be a sample. In practice a sample is satisfactory if its properties correspond to the characteristic properties of the bulk within limits set by the nature of the test. In a sampling plan some factors are of prime importance to be considered. First of all we should consider nature of the sample, its physical state - gas, liquid or solid. Its toxicity, if any, should also be taken the account for. The cost of the test, the cost of sampling and the value of the product should be compared. It is important to point out that an expensive test should not be run on a large number of samples of material of low value.

The number of defective items will also influence the choice of a sampling plan. For example, selecting portions of material on the basis of particle sizes can introduce serious error. The coarser particles may be different in composition when compared with the finer particles.

The philosophy of sampling has been undergoing a gradual change. The main factor responsible for this change is the use of quality control by many industries. The application of statistical methods is very useful in this regard. It is assumed that the property being investigated has a normal distribution in the population. The use of various statistical formulae can be made for calculating precision etc., but a discussion on statistical calculations is beyond the scope of this chapter. Numerous books on statistics are available that discuss the subject within the limitations of our present day knowledge.

SAQ 1

Name three important factors of prime importance in a sampling plan.

1.3 COLLECTION OF SAMPLES

The overall sampling procedure may be divided into three steps:

(i) collection of gross sample,

(ii) reduction of the gross sample to a suitable size for the laboratory, and

(iii) preparation of the laboratory sample.

The decreasing sample size must be accompanied by a corresponding decrease in particle size by a variety of methods such as: grinding, mixing, pulverizing, etc. In order to prepare a laboratory sample, it requires packaging in a suitable form that ensures protection from changes in composition.

Some of the methods used for collecting samples will now be discussed in general way, for different types of materials depending on the physical state, i.e., for gas' liquid and solid samples.

1.3.1 Gases

Collection of gases varies from simple to very complicated. Collection of pure gases is simple, where as collection of a mixture of gases or of atmospheric gases represents a complicated type of sampling. Thus, the collection of gas samples depends on the type of the gases, environment from which the collection is made and temperature.

However, in many cases, the analysis of gases is frequently carried out coincidentally with the collection.

The gas samples are usually collected in cylindrical containers made of glass and fitted with stopcocks at both ends. At one of the ends some arrangement may be made to evacuate the container or some other device required to filling the container. The stopcocks should be carefully cleaned and lubricated.

The following three methods, in general, are used for collecting gaseous samples.

(i) Flushing

(ii) Displacement with liquid

(iii) Expansion into an evacuated container

When sampling from atmosphere, a single sample would supply little information. Therefore, a large number of samples taken at various locations and at various times will be required in any study of atmospheric pollution. In most cases of atmospheric sampling, large volumes of air are passed through the collecting apparatus. Many solids and liquids may be very finely divided and remain suspended in the air, such solids may be removed by suitable filters and liquids are either adsorbed or reacted with liquids or solids present in the sampling apparatus. In order to know the total volume of air that is represented by the collected sample, a flow meter may be used. Sometimes, a manual pump is used that delivers a definite volume of air through the sample tube with each stroke of the pump.

1.3.2 Liquids

The sampling of homogeneous liquids or pure liquid does not pose any problem. A regular sample can be collected either by pouring the liquid into container or by dipping the container into bulk of the liquid.

However, when the samples are to be collected from the various depths, the equivalent volumes from the top, middle and bottom portions of the well-stirred bulk liquids are collected with the help of a special device, which consists of a bottle with a two-hole stopper. One hole is fitted with a short piece of tubing which is slightly longer than the depth of the bulk liquid. The longer tube is fitted with a stopcock. To collect the sample the stopcock is closed, the bottle is dipped in the liquid and the shorter tube is lowered to the desired depth. The stopcock is then opened. As the air escapes through the long tube (on opening the stopcock) the liquid fills the bottle through the shorter tube. This method can be applied for sampling emulsions and suspensions that are fairly stable.

The accurate sampling of immiscible liquids is very difficult. For sampling immiscible liquids the above method can be applied but each layer is sampled separately. A composite sample is then prepared by mixing the different samples in proportions to the relative volume of different layers in the bulk mixture of liquids.

1.3.3 Solids

It is impossible to give general methods of sampling that are applicable to all solid materials and to all conditions of sampling, since large variations exist within solid materials. Hand sampling is often used to have a gross sample of bulk material such as coal or ores. A laboratory sample is then prepared by reduction of the gross sample to a convenient weight. Hand sampling, though inaccurate, is sometimes the only available means of collecting a sample. It

requires a careful plan for the material and conditions of sampling and also careful supervision of the worker.

Material consisting of discrete lots is sampled by taking a random selection of such lots. The panicle size of the collecting solid material also introduces a variable that does not occur in the sampling of liquids or gases. This factor is of direct importance to the analyst because it may be involved in the laboratory preparation of the sample for analysis. The amount of impurity may vary in different sized particles. The difficulty may be avoided by grinding the large particles to a smaller size and then passing through a sieve, which has a nominal opening of about 0.05 mm.

Sampling of solids depends upon the physical properties and geometry of the material. The procedure of sampling from solids in compact such as slabs, sheets or blocks, can be made on random basis. Sheet of metal, for example, can often be sampled by clamping a number of sheets together the edges flush and milling across the edges to obtain an edge sample. Sometimes, the sampling procedure from a block follows to drill or punch holes at regularly spaced intervals along the diagonal of a block, alternately from one side and then from the other.

1.3.4 Precautions

Sampling requires adequate precautions that can be summarized as follows:

i. Label every bottle containing the sample with appropriate information including the name of collector, date, exact location and any other data relevant to analysis of the sample. Labeling of samples prevent misidentification.

ii. Before filling a liquid sample, rinse sample bottle with the liquid being collected.

iii. While filling the container preferably leave an air space of about 1% of container capacity to allow for thermal expansion.

iv. Take adequate precautions during sampling and sample handling, if there is a chance of sample constituents being toxic, wear gloves, apron or other protective apparel. Always wear eye protection.

v. In a laboratory, make use of fume hood to open the sample container.

vi. Never put food materials near samples.

vii. If inflammable compounds are suspected to be present keep sparks and flames away from samples and sample locations.

viii. Take special safety measures for handling samples with radioactive contaminants.

ix. Operations that change the composition of the sample should be avoided as far as possible.

x. Use plastic container if sample contains fluoride or strong alkalis.

SAQ 2

How the liquid samples are collected from various depths?

1.4 CONTAINERS

The container of a suitable material is of great importance for sampling since the nature of the material that is being sampled and also the tests are to be made on the sample may be affected by the material of the container if it is not inert to the sample. Usually, containers used for sample collection are made of plastic or glasses. One material may be preferred over the other depending on the contamination present in the sample. Avoid glass containers for samples of high pH and containing substances that affect glass. Plastic containers should be used for such samples. However, avoid plastic containers, if samples contain organic compounds that affect plastic.

Metal containers may be used for sampling many materials, but they must also be used with caution. Materials that will corrode the container or that will react with the metal should not be sampled in such containers. For example, for storing acetylene a copper container cannot be used, since it will cause explosion due to the formation of copper acetylide.

Many materials must be protected against light, atmosphere, moisture, temperature etc. Dark bottles are frequently used to protect material from light, use of tight stoppers or sometimes by sealing with paraffin may give a sufficient protection against the atmosphere. In more severe cases it may be necessary to seal in glass in an atmosphere of nitrogen or helium.

SAQ 3

a. Usually, containers used for sample collection are made of --------plastic or --------.glasses

b. --------- containers should be avoided for samples of high pH

c. Dark bottles are frequently used to protect material from ------.

1.5 COLLECTION AND PRESERVATION OF WATER SAMPLES

Collection of water samples and its preservation is very important for the analysis of water, especially for chemical purpose.

1.5.1 Collection of Water Samples

Sampling is necessary to examine waters of a wide range of quality that includes water suitable for domestic use, water suitable for industrial supplies, ground water, surface water from rivers, lakes, rain water, municipal waste water, industrial water, saline water etc.

It is not practical to use the same procedure for sampling all types of water. Collection and preservation of water samples depend on the source, purpose of examination, and the analytical procedure used for determination. Special precautions are necessary for microbiological, organic and trace metal analysis. The constituents present may be totally or partially lost if proper sampling and preservation procedures are not followed.

Collection of water samples may be manual or automatic. In manual sampling no equipment is involved, but this type of method may be time consuming for routine or large scale sampling programs. Whereas in automatic sampling, specifically designed equipment is needed and human errors can be eliminated. This method may provide the means for more frequent sampling.

Water sampling procedure depends on source also. For example, when a sample is collected from tube well, the collection should be made after the well has been pumped sufficiently so that the sample represents the ground water source. And when a sample is collected from river, the collection should be made at different times from different places and from different depths, since river water has rarely the same composition throughout the year.

1.5.2 Preservation of Water Samples

Analytical determinations may be affected by sample storage before analysis, chemical and biological changes take place continuously after sample collection. Change in pH and in amount of dissolved gases (oxygen, carbon dioxide) is very common. Growth of microorganisms may take place, when the interval between sample collection and analysis is long enough.

To avoid the aforesaid changes, proper preservation of water samples should be made. Nevertheless, complete preservation of water samples is a practical impossibility. At best, preservation techniques can retard the possible chemical and biological changes that occur during the sample storage.

Samples should be kept under conditions that will preserve the sample with as little change as possible. Common preservation techniques used are:

(i) Refrigeration

(ii) Chemical preservation

(iii) pH control

(iv) Keeping in dark

Analyse the samples as quickly as possible on arrival at the laboratory. If immediate analysis is not possible store samples as cool as possible without freezing. Storage at 4°C is recommended for most samples. Cooling minimizes the, potential for volatilization or biodegradation during storage.

Chemical preservatives may be used in certain cases. Before adding such preservatives make sure that they do not interfere with the analysis being made. The choice of the preservative be made taking due regard to determinations. When the preservatives are used, they are added to the

sample bottle initially so that the whole sample is preserved as soon as collected.

Since pH plays an important role in many reactions, therefore, pH control is essential. Sometimes, pH may change significantly in a matter of minutes.

Changes caused by growth of microorganisms are retarded by keeping the sample in the dark and at a low temperature. The use of amber and opaque bottles is quite satisfactory to avoid photochemical reactions required for the growth of microorganisms.

SAQ 4

Name some of the preservation techniques used for water samples.

1.6 SAMPLING OF SOIL

Sampling of soil is by far the most significant part of soil analysis. Sampling of soil depends on various conditions, such as, the purpose for which the sample is required, the nature of the soil, the analytical determinations to be made, the area and the depth from which the samples are drawn etc.

Soil is the material occupying the outermost part of the earth's crust and possessing different mineralogical and chemical properties resulting from its unique position on the surface of the earth and the environmental factors existing at that position. The properties of soil vary with lateral and vertical displacement. The individual layers of soil that exhibit distinct variation with depth are called soil horizons where as the vertical cross-sections as profiles. Soil may be classified on the basis of the size of the particles and, silt and clay. The coarse sand has particles of diameter of 2-0.2 mm, which are taken as fine sand. The silt particles below 0.002 mm diameter size are clay particles.

The soil samples are usually collected in plastic, jute or cloth bags. Calico bags are convenient to collect the soil samples. Such bags are numbered and labeled to give required information, such as, different horizons, location, date, time etc. For most purposes 1-2 kg sample weights are collected.

Samples should be taken from the face of a freshly dug pit sufficiently deep to permit the collections from different horizons. Samples are collected from different areas with the help of a suitable tool by breaking the lumps, if any, into smaller pieces and mixing thoroughly. Pass the collected soil through nylon or stainless steel sieves with holes of 2 mm. Crush the bigger particles again and again till the aggregate particles are fine enough to pass through the sieves. The bigger particles left as stones and gravels are discarded after weighing.

Area for digging should be chosen with the consideration of such factors as, degree of erosion, surface drainage, color, texture, etc. Sampling of soil should be done in broad daylight and the soil should be reasonably dry. Before sampling, all fitter should be removed from the surface of the area chosen for sample collection, but do not remove the organic matter embedded in the

horizon. Collect the samples from different depths as required. The depth of choice depends on the purpose of study. For example, to study the fertility status the samples of upper 6" of 9" depth soil are suitable, whereas for salinity and alkalinity studies samples of low depths are more important than that of surface soil.

To take a composite sample representation of the area collect sample from each important and distinctly different soil stratum at different depths. Thoroughly mix each composite sample and dry in a well ventilated place. Pass the sample through a stainless steel or nylon sieve to collect the particles of diameter less than 2mm.

For sampling soil a suitable tool is essential. An ideal sampling tool should be such that it gives an uncontaminated reproducible sampling unit of approximately uniform cross-section to the desired depth. Augers were often used in past but they are not very satisfactory. Sampling tubes are usually used for sampling surface horizons. Instruments such as towel and knife are frequently used to collect samples from suitable depths after digging a pit.

SAQ 5

What are the factors which affect the soil samples?

1.7 ERRORS AND TYPES OF ERRORS

The data obtained by a physical measurement should always raise the question of errors and their nature. Various types of errors are caused in quantitative chemical analysis. It is, therefore, worthwhile to account for these errors. Now first understand what are error and its types.

1.7.1 Error

The error is an inverse measure of the accuracy of a result. Less the error, more accurate the result is. Error is mathematically defined as the difference between the observed value and the true value:

$$E = O - T$$

where "E" is the error (absolute error),

"O" is the observed value of a measurement, and

"T" is the true value.

It is with regard to sign, and it is reported in the same units as the measurements. Let us consider, for example, for the capacity of a measuring flask whose true value, as given by standard measurements, is 250 cm^3. For a series of 5 measurements done by an analyst the error is represented in Table 1.1.

Table 1.1: Expression of Error in Measurements in Volume of a Flask

Serial Number of Observations	Observed Value	True Value (cm^3)	Error (E) (cm^3)
1	249	250	–1
2	247	250	–3
3	250	250	0
4	248	250	–2
5	251	250	+1
		Total	–5
		Average	–1

The error represented in the above table is the absolute error and the average

$$\text{Error} = \frac{\text{Total error}}{\text{Number of observations}} = \frac{-5\,\mathbf{cm}^3}{5} = -1\,\mathbf{cm}^3$$

However, the absolute error is of little practical significance for a quantitative analysis. It is the relative error, that is, the error relative to the true value (E/T) expressed in suitable units, is of the practical importance as a measure of inaccuracy (or as an inverse measure of accuracy). It is convenient to express relative error in terms of percentage (parts per hundred), or parts per thousand (ppt), preferably.

To understand the importance of relative error let us consider the measurement of the capacities of three standard flasks of 10, 100 and 1000 ml by three analysts A, B and C respectively represented as follows:

Analyst	**A**	**B**	**C**
T =	10 cm^3	100 cm^3	1000 cm^3
O =	11 cm^3	101 cm^3	1001 cm^3
E =	1 cm^3	1 cm^3	1 cm^3
Relative Error = E/T	1/10 = 0.1	1/100 = 0.01	1/1000 = 0.001
%R.E. = ($E \times 100$)/T	10%	1%	0.1%

You see that the absolute error in all the three cases is the same (1 cm^3), but the comparison of the relative errors tells that the error (inaccuracy) in case of the analyst C is the least and hence his result is the most reliable out of all the three.

However, for a finite measurement the true value is, usually, not known and the scatter is measured in terms of *deviation* which is the difference between the observed value and the mean of the given set of data. You will study about the deviation in detail in the sub Sec**. 1.9.6** of this Unit.

1.7.2 Types of Errors

For a practical point of view, it is useful to classify errors into two categories: (i) determinate errors, and (ii) indeterminate errors.

i) Determinate Errors

As the name implies, determinate errors are those whose magnitude can be determined after assigning a definite cause and thereby they can be corrected, for example, weighing of a hygroscopic salt like calcium chloride. Its weight will vary according to water absorbed by it from atmosphere if it were weighed in open. The error caused due to absorption of water by the salt can be corrected for if the salt were weighed after drying and keeping in a desiccators.

The determinate errors may be constant or variable. When the determinate error possessed the same value from one measurement to another under a variety of conditions, is called a *constant error*, for example, error due to uncalibrated weights. On the other hand, in certain cases the determinate errors may vary in magnitude with conditions, for example, the errors caused due to expansion or contraction of volumetric solutions with a change in temperature. These *variable determinate errors* are sometimes called *systematic errors*. Commonly people do not use this designation of systematic error only for variable errors but frequently call both types of determinate errors (constant or variable) as systematic errors and we shall also follow the same nomenclature.

Sources of Determinate Errors

Determinate errors may be caused by numerous sources. It is not needed to enumerate them all but more important ones are given as follows:

a) *Errors due to equipment*: These types of errors arise due to faulty or uncalibrated devices, for example inequality of the balance arms or insufficient accuracy of the balance, uncalibrated measurement vessels, or uncalibrated weights etc.

b) *Errors due to reagents*: The quality of the reagents is very important in the quantitative analysis. Certain reagents may possess impurities that will interfere in a particular quantitative analysis. Errors are also caused by the use of incorrectly standardized solutions for titration.

c) *Personal errors*: These errors are caused due to constitutional inability of an analyst to make certain observations accurately, that is, they are caused due to some natural weakness of the analyst. For example, some persons always detect the end point a little past in titration because their inability to judge colour changes exactly. Personal errors also include the so called *psychological errors*, due to certain bias often met within students, for example, some students often tend to choose in a burette reading the division which is closer to the previous determination or even to those found by his fellow students rather than the actual one. Obviously, this makes the results less accurate.

d) *Operational errors*: The operational errors are associated with the operation of an analysis. These errors are independent of the instrument and the apparatus employed, also these errors are not related to the chemical properties of the system in hand. Their magnitude depends more upon the analyst himself than on any other factor. They are mainly caused by carelessness of the operator in a quantitative work, for example, loss in bumping of uncovered solution while heating, failure to remove precipitate quantitatively from vessels, underwashing or overwashing of precipitate, etc.

e) *Methodic errors*: Sometimes a particular method for the determination of a particular constituent in the given sample may not be accurate because of improper selection of the procedure in the required range and will give the inaccurate result. For example, in the determination of iron (present in traces) in water, the gravimetric method will not give the correct result, and a method suitable for trace contents, say; a spectrophotometric method should be selected. The methodic errors are inherent in the method, and cannot be corrected unless the correct method is applied.

ii) Indeterminate Errors

Indeterminate errors are the errors for those no exact cause can be assigned, hence they cannot be corrected. The sources of these errors may be similar to those for the determinate errors but no definite causes out of these can be assigned for indeterminate errors. Even after applying every correction for the possible determinate errors the replicate observations may vary. Such a variation in observations is due to the indeterminate errors. These errors follow the rules of chance or the laws of probability and are also known as *Random Errors*. These errors, which accompany every determination, are quite irregular and generally small.

Indeterminate errors cannot be prevented or eliminated by corrections. However, they can be considerably reduced by increased care in work, and increase of the number of replicate determinations. Their pattern of occurrence can be analyzed by the techniques of statistics in order to secure a worthwhile insight into their magnitudes, frequencies of occurrence, and effects on the final expression of results. A plot of the normal distribution of occurrence of indeterminate errors can be prepared.

SAQ 6

In an analysis the observed value is 5.24 g compared with the accepted (true) value of 5.28 g. What is the relative error in parts per thousand?

SAQ 7

List the sources of determinate error.

1.8 ACCURACY AND PRECISION

You have experienced that a single measurement cannot be taken as an accurate result. A single result could be in error of one kind or the other. Our confidence in an analytical result is increased by increasing the number of parallel determinations (replicate determinations). When assessing the final results it is necessary to judge (i) their accuracy, and (ii) their precision. It will be worthwhile to understand the meanings of these two terms while evaluating the analytical data.

1.8.1 Accuracy

The term accuracy is expresses as the correctness or nearness of a measurement to its true value (or accepted value). It is expressed in terms of error. Error (defined in section 1.7.1) is an inverse measure of accuracy. Less the error greater is the accuracy. Thus, after knowing the relative error the loss in accuracy can be estimated.

There are various ways and units to express the accuracy of a measurement. The most common being either in terms of percent relative error. Expressed as percent or part per thousand, E/T x 1000.

Example 1.1: A sample was analyzed for desired constituent having 2.62 g as the true value. The results of three measurements were 2.50 g, 2.54 g, and 2.52 g. Find the error of the mean (mean error), the percent relative error and the relative accuracy of the mean of the measurements.

Solution:

Measurement (O)	True Value (T)	Error (O-T)
2.50 g	2.62 g	–0.12 g
2.54 g	2.62 g	–0.08 g
2.52 g	2.62 g	–0.10 g
Total 7.56		Total –0.30 g

Mean 7.56/3 = 2.52 g Mean Error = –0.10g

$$\%\text{Relative Error} = \frac{\text{Mean Error}}{\text{True Value}} \times 100 = \frac{-0.10}{2.62} \times 100 = \text{-3.8\%}$$

$$\text{Relative Accuracy (\%)} = \frac{\text{Mean}}{\text{True Value}} \times 100 = \frac{2.52}{2.62} \times 100 = 96.2\%$$

Also it can be calculated from %R.E. as = 100 – 3.8 = 96.2%

1.8.2 Precision

Precision is defined as the reproducibility of measurements. It tells an agreement between the numerical values of replicate measurements. The

magnitude of random errors determines the precision of the analytical results. It follows that the closer the results of replicate determinations are to each other, the more precise is the analysis considered to be. Precision in a common way is expressed in terms of deviation. Less the deviation more precise the result is. *Deviation or apparent error* is defined as the difference between the measured value and the mean (average) of the series of measurements. The deviation bears a relationship to the mean value of a series similar to that which exists between the absolute error and the true value. Mathematically,

$$d = | X_i - \bar{x} | \quad or \quad d_i = | \bar{X} - \mu |$$

where, d is the deviation, X_i is the observation, and x is the mean of series of measurements.

$$\bar{x} = \frac{x_1 + x_2 + \ldots x_n}{N} = \frac{1}{N} \sum x_i$$

where symbol Σ represents summation (add all). Deviation is, generally, taken without regard to sign.

It is more informative to express the precision in terms of relative deviation which is deviation relative to the mean expressed in suitable units.

Thus, average deviation $(a.d.) = \frac{d_1 + d_2 + d_n}{n} = \frac{\sum (x_1 - \bar{x})}{n}$ and

$$\%a.d. = \frac{a.d.}{\bar{x}} \times 100$$

The most important measures of precision are the standard deviation and the variance. The standard deviation s of a measurement is theoretically given by:

$$s = \sqrt{\frac{d_1 + d_2^2 + \ldots + d_n^2}{n-1}}$$

where, n is the number of observations and $(n - 1)$ is known as the degree of freedom. Variance V is the square of the standard deviation,

$$V = s^2 = \frac{\sum d^2}{n-1}$$

Distinction between Accuracy and Precision

The accuracy should not be confused with the precision. Good agreement in parallel determinations signifies that the determinations have been made under closely similar conditions; it does not guarantee the accuracy of the results. A method may be precise but may not be accurate if a large systematic error is made. On the other hand it is nearly impossible to have accuracy without good precision. The difference of the terms accuracy and precision can be illustrated by considering the shooting of series of bullets on the targets by three riflemen (A, B, C), shown in Fig. 1.1

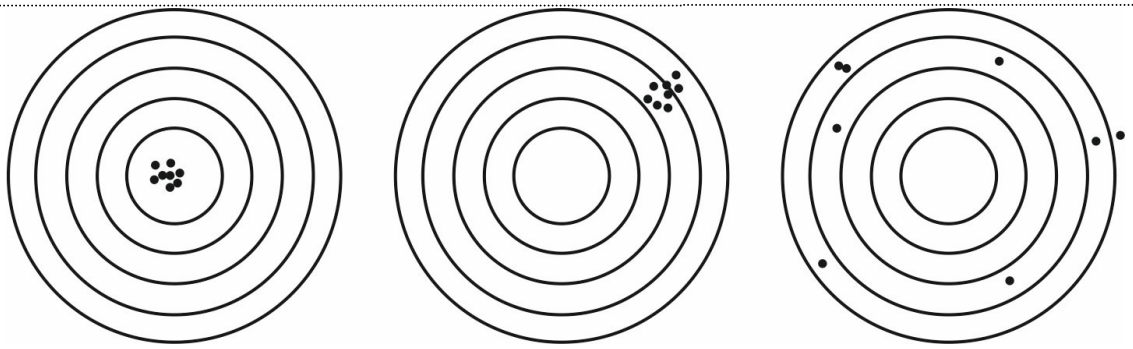


Fig. 1.1: Shooting by three Riflemen.

Rifleman A has the ideal marksmanship. His all hits are in centre. His results are both accurate and precise. The shooting by rifleman B shows a good grouping of hits which indicate that the marksman is undoubtedly consistent, but in this target the grouping is centered at one point away from centre, hence cannot be considered as representing accurate shooting. His results agree well mutually means precise, but the final result (obtained as their mean value) differs somewhat from the actual value, therefore not accurate. This illustrates the effect of a constant error such as poorly adjusted sights and, since precise marksmanship is evidenced, it is reasonable to assume that if the source of error can be located and corrected, accurate shooting should be forthcoming.

The shooting by rifleman C shows that the hits are spotted all over the face of the target in a display of poor reproducibility. It seems that the rifleman has no experience of shooting, his hits result only as an accident. We say that the mean value is of low reliability. Hence, these results are neither precise nor accurate. Also we see that good precision is needed for good accuracy.

Of course the most favourable methods are those which give precise and at the same time accurate results. In practice results that are precise but subject to small systematic error are often more useful than results with an accurate mean value but low precision, since in first case we can actually find out how much they differ from the true value, while in the second case we know nothing but that the mean value is of low reliability. In the following example you can see the difference of accuracy and precision for the results of burette reading of a titration.

Example 1.2: The burette readings of titrations carried out by three students A, B and C are given below. Compare the accuracy and precision of the three students, if the true reading is 22.22 cm^3.

Student	A	B	C
Burette readings	cm^3	cm^3	cm^3
	22.22	22.28	22.38
	22.24	22.27	22.12
	22.23	22.29	22.32
	22.21	22.28	22.30
	22.20	22.28	22.18
Mean value	22.22	22.28	22.18

You can understand from the observed values and calculated mean values of titrations of three students that the results of student A are reproducible and the mean value resembles the true value. Hence the results of student A are both *precise* and *accurate*. A look of the titration results of student B shows that his results are reproducible but the mean value is slightly on the higher side than the true value. May be he might be taking the end point (colour change) on the higher side. Therefore, the results of student B are precise but not accurate. The readings of students C are spread in a wide range with a poor reproducibility. Hence his results are neither precise nor accurate.

SAQ 8

Define accuracy and precision.

1.9 SOME IMPORTANT TERMS USED FOR EVALUATION OF DATA

You have experienced that a single measurement cannot be taken as an accurate result. Our confidence in an analytical result is increased by increasing the number of parallel determinations, known as *replicate determinations*. Determination of the number of times a measurement should be replicated in order to approach the value of experimental mean around the true mean with a certain degree of probability. That is, the more numerous the number of observations the more their results approach the truth. The replications are useful to us in two ways. First, a reliable central value of the set of analytical data should be evaluated. The mean is the most commonly used measure of the central value and the less common used measures are the median and the mode. Second, an analysis of the variation in results helps us to estimate the uncertainty associated with the central value of the data. You should note that a population is the collection of all measurements (very large number and to infinity to the analyst, while a sample is the subset of these measurements (finite number of measurements) selected from the population and we also call it as finite sample. Statisticians use Greek letters (such as μ, σ) for the population parameters, whereas English letters (such as $\bar{x}$, s) for finite samples known as statistics. The dual set of symbols is a valuable aid in discussing experimental data.

1.9.1 Mean

The mean is **the mathematical average of a set of two or more numbers**. Summing the numbers in a set and dividing by the total number gives you the arithmetic mean. In statistics, **the mean is equal to the total number of observations divided by the number of observations.** The *mean* is the most widely used measure of the central value. For a finite sample (for n < 30) the mean known as *sample mean* is represented by $\bar{x}$ and is the arithmetic average of all the observations in the set of data;

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum x_i}{n}$$

where, x_1, x_2…, x_n are the replicate observations, x_i represents the individual value of x making up the set of n number of observations, the symbol $\sum x_i$ means summation of x individual values from i = 1 to i = n, i.e. to add up all of the individual values of x in the set of replicate analyses.

For the entire population of data or universe of data (the number of observations approaching infinity i.e. $N \rightarrow \infty$) the mean known as *population mean* is represented by μ and is given by

$$\mu = \frac{x_1 + x_2 + x_3 + \ldots x_N}{N} = \frac{\sum x_i}{N}$$

Where N denotes number of observations. We will see later, after studying the probability distribution of data that the population mean μ is the most probable value and is taken to be the *true value* of the measured quantity.

For example, the arithmetic mean of four values: 4, 12, 36 and 44 is:

$$\frac{4+12+36+44}{4} = 24$$

1.9.2 Median

When quick measure of central value is to be decided and when gross errors are suspected the central tendency of a group of results can be expressed in terms of *median* by arranging the observations either in ascending or descending sequence. Median means the middle value. That is, there are equal numbers of results that are larger and smaller than the median. For an odd number of total observations (n is odd), the median is obtained directly as middle value. For a set of even numbered total observations (n is even), the median is the average of the middle pair of observations.

For example:

1, 3, 3, **6**, 7, 8, 9

Median = 6

1, 2, 3, **4**, **5**, 6, 8, 9

Median = (4 + 5) ÷ 2

= 4.5

The median is a less efficient measure of central tendency than is the mean, but in some cases it may give an easy look of central tendency in the sequential arranged data particularly in dealing with small samples. Thus, for small number of observations, the median may be a better estimate of the central value (the true value). Statistically it can be shown that the median of 10 observations is as efficient conveying the information as is the mean from 7 observations.

The median is used advantageously when a set of analytical data contains a probable outlying result, a result that differs significantly from others in the set. An outlying result does not affect the median value since the outlying result lies on the extremes. On the other hand, an outlying result can have a significant effect on the mean of the set since it is included in the calculation of the mean.

1.9.3 Mode

The observation which occurs most frequently (i.e. with maximum frequency) in a series of observations is known as *mode*. It is yet another quick measure of central value if the number of observations is not too small. For example, the mode of the set of data: 12.6, 12.7, 12.9, 12.7, 12.6, 12.8, 13.0, 12.5, 12.6, the value 12.6 is the *mode* since this is occurring with maximum frequency (four times).

1.9.4 Deviation

The error of a measurement cannot be stated if the true value of the quantity is not known. It is meaningful then to take the difference between a particular measured value (observation) and the arithmetical mean of a series of measurements and this difference is called as its *deviation for apparent error*.

A deviation is generally taken without regard to sign. It is defined mathematically as,

$$d = | x - \bar{x} | \ldots \text{ or } D = | x - \mu |$$

where, d is the deviation of the observation x of a finite sample from its mean $\bar{x}$, D is the deviation of an individual measurement from the population mean μ, and | | denotes that the difference is taken as absolute. The reproducibility of measurements is expressed in terms of various types of deviations.

1.9.5 Average Deviation

The average deviation $\bar{d}$ (a.d.) or the mean deviation is the average of individual deviations:

$$\bar{d}\,(\text{a.d.}) = \frac{d_1 + d_2 + d_3 + \ldots + d_n}{n} = \frac{\sum d_i}{n} = \frac{\sum | x_i - \bar{x} |}{n}$$

Where the symbols have their usual meanings.

The ratio of the average deviation to the mean is known as *Relative Average Deviation* (RAD) which can be expressed as *percent average deviation* when multiplied by 100.

$$\text{RAD} = \frac{\text{a.d.}}{\bar{x}}$$

$$\%\ \text{a.d.} = \frac{\text{a.d.}}{\bar{x}} \times 100$$

Historically the average deviation has been widely employed as the estimate of precision. However, it suffers from the disadvantage that the estimate of this statistics depends upon the number of measurements. The larger the number the better will be the estimate.

1.9.6 Probable Average

Deviation

The probable deviation P is an older measure of precision and now only rarely used. It is defined as the deviation having a magnitude such that there are

equal numbers of deviations greater and smaller than itself. In a set of large number of observations it is also known as *probable error*.

1.9.7 Range

The difference between the largest and smallest values in a set of measurements is known as the range. It tells the spread of data. The range is often used, with appropriate factors that depend on the number of measurements, as a quick statistics to a rough estimate of precision.

Example 1.3: The careful determination of iron in an iron ore sample is carried out by gravimetric method by five students. The percentages found are:

67.48, 67.47, 67.37, 67.40 and 67.43

Find the standard deviation, average deviation and probable deviation of a single determination and of the mean.

Solution:

% of iron	$x - \mu$	$(x - \mu)^2$
67.48	0.05	0.0025
67.47	0.04	0.0016
67.37	0.06	0.0036
67.40	0.03	0.0009
67.43	0.00	0.0000
Mean/mean μ = 67.43	Σ0.18	Σ=0.0086

Standard deviation, $\sigma = \sqrt{\frac{(x-\mu)^2}{N-1}} = \frac{\sqrt{0.0086}}{\sqrt{4}}$

$= 0.047$

Average deviation, $S = \frac{\sum(x-\mu)}{N} = \frac{0.18}{5} = 0.036$

Probable deviation, $\rho = \sigma \times 0.67\,(\text{For } 50\%)$

$= 0.047 \times 0.67 = 0.631$

The results are genrally expressed in terms of σ i.e., in the above example 67.43 ± 0.047 percent.

$$\sigma_\mu = \frac{\sigma}{\sqrt{N}} = \frac{0.047}{\sqrt{5}} = 0.021\%$$

$$S_\mu = \frac{S}{\sqrt{N}} = \frac{0.03}{\sqrt{5}} = 0.016\%$$

$$\rho_\mu = \frac{\rho}{\sqrt{N}} = \frac{0.031}{\sqrt{5}} = 0.014\%$$

SAQ 9

Calculate the median for the data: 14.1, 13.8, 14.3, 13.6, 13.4 and 13.5.

1.10 SOME BASIC CONCEPTS OF CHEMICAL ANALYSIS

If you analysing some chemical compound some basic knowledge is necessary to analyse the compounds. In this section we will learn some following concepts of chemical analysis.

1.10.1 Scale of Operation

Analytical procedures can be classified in two manners: one on the size of the sample and the other on the relative amount of the constituent in the sample.

On the basis of the size of the sample taken the procedures of analysis may be classified into four groups: (i) macro (ii) meso or semi-micro (iii) micro, and (iv) ultra-micro or sub-micro.

Macro Procedure: A macro procedure is that procedure in which the sample taken is more than 0.1 g and, generally, not greater than 1.0 g. For example, in the wet scheme used for basic and acid radicals test, the amount of the sample taken lies within this range and thus, this is a macro procedure. These procedures are also known as decigram sample procedures, as the sample mass is a few decigrams, with the lowest limit of one decigram.

Meso Procedure: In a meso (or semi-micro) procedure the amount of the sample taken lies between 0.01 g to 0. 1 g. These procedures are also known as centigram sample procedures, as the sample mass is a few centigrams, with the lowest limit of one centigram.

Micro Procedure: When the amount of the sample taken is less than 0.01g but more than 0.001 g, the procedure is known as the micro procedure. Such a procedure is also termed as a milligram sample procedure, since the sample mass is a few milligrams, with the lowest limit of one milligram.

Ultra-Micro Procedure: The procedure for which the mass of the sample taken is less than 0.001g, the method is known as ultra-micro procedure. It is also called as a sub-micro procedure.

The relative amount of the constituent in the sample is also taken as the basis of methods of determination. These are: major, minor and trace analysis. In major analysis the constituent being determined has a relative amount of 1-100% in the sample. Gravimetry and titrimetry are applied for major constituent determinations. In **minor** analysis the constituent being determined has a relative amount of 0.01-1% and in **trace** less than 0.01%. Physico-chemical methods are applied for last two types of analysis.

A suitable technique of analysis may, thus, be concerned depending on the amount of sample available and the approximate percentage of the constituent (being determined) present in the sample.

1.10.2 The Mole Concept

A student performing a quantitative analysis is usually concerned with the amount of the constituent present in the given sample. The amount may be expressed in various units of mass such as gram, milligram, microgram etc. However, the amount of reacting substances must be expressed in terms of such chemical units, which will be related to the expression of a chemical reaction, that is, the molecules. For example, in the following reaction:

$$NaOH + HCl \longrightarrow NaCl + H_2O \quad \ldots (16.1)$$

1 mole of NaOH + 1 mole of HCl ⟶ 1 mole of NaCl + 1 Mole of water

Supposing 100 molecules of NaOH are to be neutralized by HCI, then 100 molecules of the HCI are required. In other words, I molecule of NaOH will be neutralized by 1 molecule of HCI. In a particular sample of a given amount, the number of molecules can be obtained by the use of Avogadro's number, 6.023×10^{23}. Let us call the amount of 6.023×10^{23} molecules as a mole or mol (Mol in SI Unit).

In general, a mole signifies as a collection of 6.023×10^{23} particles, where by particles we mean molecules, atoms, ions etc. If we consider the amount in grams of 6.023×10^{23} molecules of a particular compound, then this will be equal to its gram molecular weight (gmw). Hence, one mole of a compound signifies 6.023×10^{23} molecules in number on one hand and its gram molecular weight in amount on the other.

The number of moles (n), in a particular amount, m gram of a compound can be easily calculated by the following relationship:

$$n = \frac{m}{\text{gmw}} \quad \ldots (16.2)$$

For example in a 2 g of NaOH, there will be 2/40 = 0.05 moles.

Another way of calculating the number of moles of a substance in solutions if the molarity is known, that is:

Number of moles = volume in $dm^3 \times$ Molarity

Yet, another way is to relate mole to volume of gases:

1 mol = 22.4 litres of a gas at STP

(One mole of all gasses occupies 22.4 dm^3 in volume.)

SAQ 10

What is the number of moles of water in one litre of it at 4°C?

1.10.3 Concentration Units

There are various methods of expressing the concentration of a solution such as: molarity (*M*), normality (*N*), percentage (%), parts per million (ppm), mole fraction etc. Those, which are in common use, will be discussed here.

Normality: This is the most commonly used system of representing the concentration of a solution and is defined as: the number of gew (gram equivalent weight) of that substance present in one dm^3 of solution.

Molarity: It is defined as the number of moles of the solute dissolved in one dm^3 of the solvent.

Molality: It is defined as the number of moles of the solute dissolved in one kilogram of the solvent.

Mole Fraction: It is the ratio of number of moles of that component to the total number of moles of all the components present in that solution.

Percentage Concentration: Percentage (or parts per hundred) is often used to express the approximate concentration. The various ways to express the percentage composition of a solution are:

i) $\text{Volume percent} = \frac{\text{volume of solute}}{\text{volume of solution}} \times 100$

ii) $\text{Weight} - \text{volume percent} = \frac{\text{weight of solute in g}}{\text{volume of solution in ml}} \times 100$

iii) $\text{Weight percent} = \frac{\text{weight of solute}}{\text{weight of solution}} \times 100$

i) Volume percentage is generally used for liquid mixtures to indicate the concentration of a liquid in the solution. For example, 5% alcohol means by volume 5 parts of alcohol in 100 parts of solution.

ii) Weight-volume percentage usually indicates the concentration of dilute aqueous solutions of solid substances.

iii) The percentage of commercial aqueous reagents is given in terms of weight percentage frequently. For example,36% of HCI in water. The weight percent is independent of temperature.

The approximate normal and molar concentration of commercially supplied solution of acids and bases are often calculated by weight percent and specific gravity. The specific gravity of a solution can be used for the weight volume relationship to calculate the weight of one litre of solution. Then by the weight percent concentration the amount of particular substance in one litre of solution can be calculated and thus on dividing by gmw the Molarity of that particular acid or base can be calculated.

Thus, it gives: $M = \frac{\text{Sp.gr.} \times 1000 \times \text{weight\%}}{100 \times \text{gmw}}$

SAQ 11

i) The number iof gram equivalent weight of a substance present in one litre of a solution is known as_____.

ii) The number of moles of solute dissolved in one litre of the solvent is known as _____.

iii) The number of moles of a solute dissolved in one kilogram of the solvent is known as _____.

1.11 SUMMARY

In this unit you learnt that:

- Sampling is to collect a very small portion of material such that it is a representative of the bulk. It should be suitable for convenient transport and laboratory testing.
- In a sampling plan the factors of prime importance are: nature of the sample, its physical state, its toxicity and the cost of the test compared with the value of the product.
- Gas samples are collected by flushing or by displacement with a liquid or expansion into an evacuated container.
- A regular liquid sample can be collected either by pouring the liquid into container or by dipping the container into bulk of the liquid.
- Sampling of solids depends upon the physical properties and geometry of the material.
- Sampling of water depends on source, purpose of examination and the
- Analytical procedure used for determination.
- Sampling of soil depends on the area and depth from which the sample is drawn, the nature of the soil and the purpose for which the sample is required.
- The error is an inverse measure of the accuracy of a result. Less the error, more accurate the result is. Error is mathematically defined as the difference between the observed value and the true value
- There are two types of error (i) determinate errors, and (ii) indeterminate errors.
- Determinate errors are those whose magnitude can be determined after assigning a definite cause and thereby they can be corrected for.
- Indeterminate errors are the errors for those no exact cause can be assigned, hence they cannot be corrected.
- The term accuracy is defined as the nearness of a measurement to its true value (or accepted value). It is expressed in terms of error

- Precision is defined as the reproducibility of measurements. It tells an agreement between the numerical values of replicate measurements.
- Meaning of some important terms used in data analysis and chemical analysis of a compound.

1.12 TERMINAL QUESTIONS

1. What kind of container will be used if a water sample contains fluoride or strong alkalis?
2. For storing acetylene a copper container cannot be used. Why?
3. What is error?
4. What is indeterminate error?
5. What is probable deviation?
6. What is the number of moles of HCl in 250 cm^3 of its 0.1 M solution?

1.13 ANSWERS

Self Assessment Questions

1. (i) Nature of the sample

   (ii) The physical state of the sample

   (iii) The toxicity of the sample

2. The liquid samples are collected from various depths with the help of the special device by which the equivalent volumes from the top, middle and bottom portions of the well stirred bulk liquid are collected.

3. a) plastic or glasses

   b) high pH

   c) light

4. Refrigeration, chemical preservation, pH control and keeping in dark.

5. Sampling of soil depends on various conditions, such as, the purpose for which the sample is required, the nature of the soil, the analytical determinations to be made, the area and the depth from which the samples are drawn etc.

6. E = 5.24 – 5.28 = - 0.04 g

   $$\text{RE (ppt)} = \frac{-0.04}{5.28} \times 1000 \qquad = -7.58 \text{ ppt}$$

7. i) *Errors due to equipment* ii) *Errors due to reagents*
   iii) *Personal errors* iv) *Operational errors* and v) *Methodic errors*

8. The accuracy is nearness to true value and the precision tells the reproducibility of replicate determinations. A result may be precise even without being accurate. On the other hand a result cannot be accurate unless it is precise.

9. Arranging sequentially the observations: 13.4, 13.5, 13, 6, 13.8, 14.1, 14.3 the median will be the mean of 3rd and 4th observation.

Median = (13.6 + 13.8) / 2 = 13.7

10. One liter of water at 4°C weighs = 1000 g.

$$\text{No. of moles} = \frac{\text{wt. in g}}{\text{gmw}}$$

$$= \frac{1000}{18} = 55.6$$

11. i) normality, ii) liter, iii) molality

Terminal Questions

1. Plastic containers should be used for each sample.

2. Since it will cause explosion due to the formation of copper acetylid.

3. The error is an inverse measure of the accuracy of a result. Less the error, more accurate the result is. Error is mathematically defined as the difference between the observed value and the true value:

4. Indeterminate errors are the errors for those no exact cause can be assigned, hence they cannot be corrected.

5. It is defined as the deviation having a magnitude such that there are equal numbers of deviations greater and smaller than itself. In a set of large number of observations it is also known as *probable error*.

6. $$\text{Molarity} = \frac{\text{no. of moles}}{\text{dm}^3}$$

No. of moles = molarity × volume in dm^3

$$= 0.1 \times \frac{250}{1000} = 0.025$$

UNIT 2

TREATMENT OF ANALYTICAL DATA

Structure

2.1 INTRODUCTION

An Analytical Chemist always worries about the quality of analytical results. Whether these results are obtained directly him or from someone else, she (or he) always thinks that the result should be good enough to use as a basis for action and should be of sufficiently accurate as need be. Any laboratory that performs analysis providing a basis for utility must have a solid quality assurance programme. The aspects which can answer the quality assurance are the assessing of analytical data. When assessing certain analytical data we are generally interested most in learning that upto what extent the results are reliable or how far they agree with the actual content of the component analyzed. Later on we wish to know whether statistically the same reliability is achieved in the analyses of different samples.

Statistical calculations are necessary to understand the properties of the data that are collected and, therefore, to set limitations on each step of analysis. The role of statistical calculations is to sharpen the analysts judgment concerning the effects of random errors because their source is not known. We can use statistical methods to evaluate the random errors which follow a normal distribution or Gaussian distribution. Advances in theoretical statistical methods and their application to industrial problems have given many answers in a logical manner. The behaviour of most industrial plants is subject to variations caused due to multiple effects. That is, the individual results are subject to chance variations and in order to draw any worthwhile conclusion it

is necessary to examine a set of data with a proper statistical approach. As more and more results are available the accuracy of estimation improves.

In this unit we shall examine the methods used by analytical chemists scientists in evaluating the significance of analytical data with the knowledge of normal distribution of errors. We shall also be able to make use of the theory of probability to express the reliability of our data. We shall see how the precision of measurements can be determined, how different sets of data may be compared with the help of different tests of significance. Finally we shall learn how to calculate the confidence limits and confidence intervals.

Expected Learning Outcomes

After studying this unit, you should be able to:

- understand the normal distribution of errors;
- estimate the Statistical test of data namely t-test and F-Test;
- get a guideline concerning whether results should be retained or rejected; and
- calculate the confidence limits and confidence interval.

2.2 DISTRIBUTION OF ERRORS

Probability distributions are of fundamental importance for judging the reliability of analytical data. Even after taking all the precautions to avoid systematic errors, the replicate results vary with each other by small to large extent. This variation in results is due to chance or random errors. The word *error* is used somewhat loosely while talking for chance error. Strictly we should speak of *deviation*, since the true value is not known. However, as we shall see below, in general convention the use of term error for chance error should not create any confusion. The chance or random errors cannot be eliminated and are often the major source of uncertainty in a determination. The combined effect of the individual uncertainties, usually, causes replicate measurements to fluctuate randomly around the mean of the set.

What we say is all measurements contain random errors. The elimination of these errors is beyond the power of the observer. However, an analyst can estimate the deviations introduced by random errors. We can apply the laws of probability provided we have sufficient number of observations. These laws of probability or rules of chance in an empirical way are summarized as follows:

i) Small deviations occur much more frequently.
ii) Large deviations occur much less frequently.
iii) Positive and negative deviations are equally probable, so that the arithmetic mean is the most probable value.

The above three rules of chance are applied for a very large number of observations which is known as the entire population or universe of data. You will see below that the rules can be verified graphically when frequency of errors are plotted as a function of their magnitudes. The plot so obtained is known as **NORMAL ERROR CURVE –** and describes the properties of universe of data. As an explanation consider the discussion given below.

If a large number of replicate readings are taken of a continuous variable, e.g. the percentage of iron in an iron ore, the results obtained will usually be distributed around the mean in a roughly symmetrical manner. When frequencies of observations (on y-axis) are plotted against values of observations (along x-axis), we get the distribution as shown in Fig.2.1 where the dots indicate the frequencies of respective values of observations. Such a distribution of *random values* (*also of random errors*) is called the normal frequency distribution. On drawing the boundaries of these frequencies we get a bell shaped curve that is symmetrical around the mean and is known as the normal distribution curve or a Gaussian Distribution Curve (as this follows the normal frequency curve of Gauss). You see that on either side of the curve there is an exponential decrease in frequency as the magnitude of error increases. The inflexion points of the curve are obtained when the result is $\pm 1\sigma$ of the mean μ. With this type of distribution about 68% of all values of the universe of data will fall within $\pm 1\sigma$ of the mean. The value of σ also indicates the precision. By comparing the values of standard deviation s_1, s_2 of two sets of data the precision of two sets of data can also be compared. For example if the standard deviation (s_1) of the first data set is ½ of the value of the standard deviation (s_2) of the second data set as shown in Fig. 2.2a means that the precision of the first data set is twice than that of the second data set.

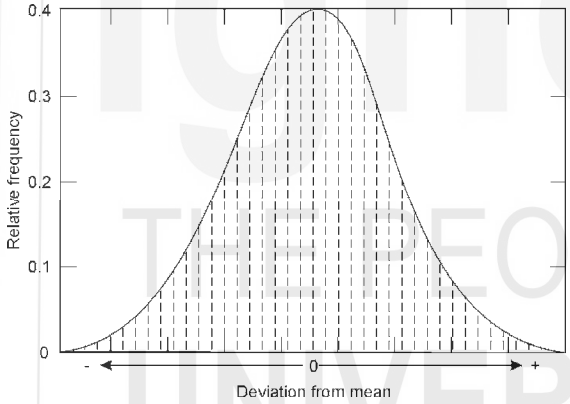


Fig. 2.1: Frequency distribution curve.

The ideal curve is based upon the infinite number of observations. The greater the number of measurements, the more closely the deviations will be represented by the above curve.

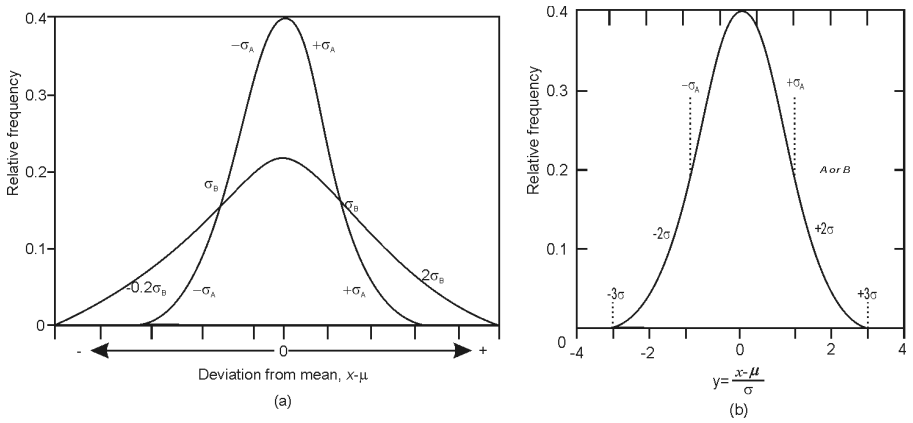


Fig. 2.2: Normal error curves (a) the abscissa is the deviation from the uranium in the units of measurements; (b) The abscissa is the deviation from the mean in units of σ.

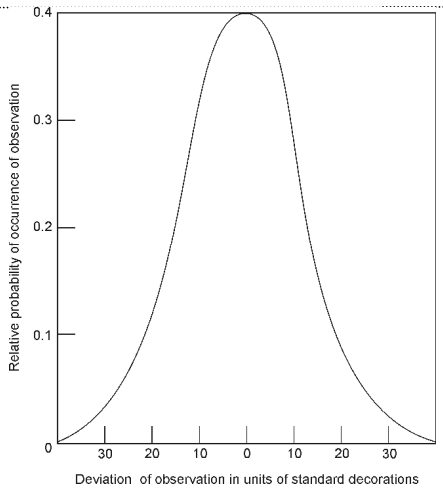


Fig. 2.3: Normal error curve.

2.3 STATISTICAL TESTS OF SIGNIFICANCE

In analytical chemistry we develop new methods of analysis and it is frequently desired to compare the results of a new method with those of an accepted (say from a past experience or a standard, e.g.; from the National Institute of Standards and Technology, NIST) method. The average values obtained from the two methods may show a difference. The question arises whether this difference is due to random fluctuation (chance error) or directional fluctuation (systematic error). The answer is qualified by a degree of certainty involving the method what is known as NULL HYPOTHESIS which considers that there is no significant difference between two sets of data. In null hypothesis procedure a comparison, of statistical parameters (based on mean or standard deviation, or some other property), is made between two sets of replicate measurements obtained by two different methods, one of them being the test method and the other being the usually accepted method. With this comparison the value of the statistical parameter of test of significance is calculated and compared with the value of the parameter given in the statistical tables available. A simple examination of the two values (calculated & tabular) will show how large a difference needs to be in order to be considered for the limit of significant divergence. Thus, if there is not a statistically significant divergence, means the null hypothesis is valid or that there is no source of systematic error and the variation in results follows the law of random errors. And if there is a statistically significant divergence, means the null hypothesis is not valid and a source of systematic error is highly probable.

We shall discuss below two tests of significance: (i) the t-test which is based on comparison of two means terms of frequencies and (ii) the F-test which is based on the comparison of two variances.

2.3.1 The t-Test or Student's t-Test

The t-test is used to test the null hypothesis that the means of two sets of data do not differ significantly. The application of t-test will be considered here only in a simple case.

When Accepted Mean Value is Known: This Eq. 2.1 applicable to the comparison of the finite sample mean *x* and the population mean μ. The quantity t is defined as

$$t = \sqrt{n}(\bar{x} - \mu)/s \quad \ldots (2.1)$$

where $\bar{x}$ = average value for the data set

μ = population mean when the series has been carried to an infinite number of observations, or accepted value given by some national standards

s = standard deviation of the data set, and

n = number of observations in the data set.

Values of *t* with reference to probability levels of 90, 95 and 99 percent are summarized in Table 2.2. In the table ν refers to the degrees of freedom.

In the procedure to test the null hypothesis the quantity t is calculated for the given observations and known as t_{calc}. It is compared with the corresponding value of t (t_{tab}) found in table of t (Table 2.2) at the desired of confidence and corresponding to ν degrees of freedom (ν = n – 1). On comparison,

i) If $t_{calc} < t_{tab}$, the null hypothesis is valid, there is no significant difference between the two means ($\bar{x}$ and μ), the variation in results is just by random errors and no systematic source of error is probable.

ii) If $t_{calc} > t_{tab}$, then the null hypothesis is rejected, a significant difference between the two means is indicated and the difference is due to some of systematic error in the values of finite series.

The above criteria indicate that the smaller the calculated t value, the more confident you are that there is no significant difference between the two means.

Example: Suppose five observations obtained for the determination of atomic mass of cadmium were: 112.25, 112.36, 112.32, 112.21, 112.36. Does the mean of these values differ significantly from the NBS (National Bureau of Standards) accepted value 112.41?

The test is applied as follows after calculating the required quantities

x	$\lvert x - \bar{x} \rvert$	$(x - \bar{x})^2$
112.25	0.05	0.0025
112.36	0.06	0.0036
112.32	0.02	0.0004
112.21	0.09	0.0081
112.36	0.06	0.0036
Σx = 561.5		Sum = 0.0182

$\bar{x}$ = 561.5 / 5

= 112.30

s = $(0.0182)^{1/2}$ = 0.067 and n = 5

$$t = \sqrt{n}(\bar{x} - \mu)/s$$

$$= (112.30 - 112.41)\sqrt{5} / 0.067$$

$$= -0.11 \times 2.236 / 0.067 = -3.67$$

(Or t_{calc} = 3.67 (disregarding the negative sign)

From Table 2.1 at 99% probability level corresponding value of t for 4 degrees of freedom is 4.60. Thus t_{tab} = 4.60.

Comparing two t values we see that

$t_{calc} < t_{tab}$ at 99% confidence level.

It is concluded that the null hypothesis is valid at 99% level of significance and there is no significant difference between the two means. The variation is due to random errors.

2.3.2 F-Test

The F test serves to show whether the precision of two different methods is the same within specified probability limits. It is applied in terms of variance ratio. The F value which is the ratio of two variances in question is determined by the relation

$$F = \frac{s_1^2}{s_2^2} \quad \ldots (2.2)$$

where

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2 \, , \; s_2^2 = \frac{1}{n_2 - 1} \sum_{i=2}^{n_2} (x_i - \bar{x})^2$$

Placing of the larger of the two variances in numerator ($s_1^2 > s_2^2$), so that the value of F is always greater than unity. The value of F determined by the use of Eq. (2.2) for experimental variances s_1^2 and s_2^2 is known as F_{calc}.

Statisticians have compiled tables of F values for various significance levels for various degrees of freedom ($\nu_1 = n_1 - 1$, $\nu_2 = n_2 - 1$), where n_1 is the number of observations for the set of larger variance (i.e. larger standard deviation). Table 2.1 is a brief F table (two sided) for the 95% confidence level. The value of F obtained from such a table is called as F_{tab}.

To test null hypothesis for the two sets of data by F test, the calculated value of F is compared with the corresponding tabular value of F. On comparison

(i) If $F_{calc} < F_{tab}$, that is if the experimental F is smaller than the corresponding tabular value of F, then no statistically significant difference is indicated between s_1 and s_2 (i.e. between two sets of data), and the null hypothesis is valid, and (ii) If $F_{calc} > F_{tab}$, then s_1 is significantly greater than s_2, and the null hypothesis is not valid.

Table 2.1: Values of F at 95% confidence level

υ_2	υ_1					
	2	3	4	5	6	∞
2	19.00	19.16	19.25	19.30	19.33	19.50
3	9.55	9.28	9.12	9.01	8.94	8.53
4	6.94	6.59	6.39	6.26	6.16	5.63
5	5.79	5.41	5.19	5.05	4.95	4.36
6	5.14	4.76	4.53	4.39	4.28	3.67
∞	3.00	2.00	2.37	2.21	2.10	1.00

To illustrate “F” test suppose that two series of observations are made one of 4 observations of standard deviation equal to 0.02 and another of 6 observations of standard deviation equal to 0.04. We have to test whether there is significant difference between the two standard deviations.

For the condition of application of F test we have to consider the greater standard deviation that is, 0.04 as s_1 and the smaller 0.02 as s_2. Hence $\nu_1 = 6 - 1 = 5$, and $\nu_2 = 4 - 1 = 3$

F is calculated as

$$F_{calc} = \frac{s_1^2}{s_2^2} = \frac{(0.04)^2}{(0.02)^2} = 4.0$$

Corresponding tabular value of F for $\nu_1 = 5$ and $\nu_2 = 3$ from Table 3.5 at 95% confidence level is $F_{tab} = 9.01$.

If you find that $F_{calc} < F_{tab}$, then, it is concluded that the null hypothesis is valid and there is no statistically significant difference between the standard deviations of the two sets of data.

SAQ 1

What is the main difference between t-test and F-test?

2.4 CONFIDENCE INTERVAL

From the above discussion you understand that mostly in chemical analysis, the true value of the population mean μ cannot be determined because a very large number of measurements (approaching infinite number) would be required to calculate it. However, with the help of statistics a range can be established surrounding sample mean $\bar{x}$ within which the population mean μ is expected to lie with a certain degree of confidence based on probability

distribution. Thus, "the range (or a numerical interval) around the mean $\bar{x}$ of a set of replicate analytical results within which the population mean μ can be expected to lie, with a certain degree of confidence (i.e., with a certain probability), is known as **Confidence Interval**. "The boundaries of this range are called the **Confidence Limits**. "The likelihood that the true value falls within the range is called the **Probability or Confidence Level** and it is often expressed as a percentage. Consider the example that in a set of iron determination it is 95% probable that the population mean μ lies in the limit ± 0.20% Fe of sample mean $\bar{x}$ = 11.30% Fe. It tells that the confidence interval is 11.10% to 11.50% Fe with 95% probability. The 95% confidence limit for μ = 11.30% ±0.20% Fe, and the confidence level is 95%.

The confidence interval is related to the standard deviation of the mean and its size depends on how well the sample standard deviation *s* estimates the population standard deviation σ. When *s* is closer estimate of σ, the confidence interval will be narrower. To give confidence interval at the high confidence levels is one of the best ways of indicating reliability. We shall discuss below two cases for finding the confidence interval: (A) when σ is known or *s* is a good approximation of σ, and (B) when σ is not known.

2.5.1 Confidence Interval of Population Mean μ when σ is Known or *s* is a Good Estimate of σ

The normal error curve can be expressed in a single variable y where *y* is defined as $\pm y = (x - \mu)/\sigma$ given by Eq. (2.3). From the definition of *y* it follows that $(x - \mu) = \pm y\,\sigma$ with the help of normal error curve and the values of *y* with probability given in Table 3.1 we see that the areas represent probabilities for the absolute deviation $|x - \mu|$ to exceed the value of $y\,\sigma$. Since $y\,\sigma$ is the deviation of a single observation *x* from the population mean μ, we can express the probability in terms of *y* as

$$\pm y\,\sigma = x - \mu$$

or $$\mu = x \pm y\sigma \quad \ldots \quad (2.3)$$

Again from Table 2.1, when y = 0.67 there are 50% chances that an observation will lie in the area having a lower deviation than ± .67 σ, & when *y* = 1.00 there is a 68.3% probability that a particular measured value has a deviation ± σ and so on. Thus, based on measuring a single value we can write for population mean μ (from Eq. 2.3) lying within limits.

μ = 0.67 s with 50% confidence (or 50% probability), … (2.3 – i)

μ = x ± 1 σ with 68.3% confidence … (2.3 – ii)

μ = x ± 2 σ with 95.5% confidence … (2.3 – iii)

μ = x ± 3 σ with 99.7% confidence … (2.3 – iv)

Otherwise in more useful way as

μ = x ± 1.96 σ with 95% confidence … (2.3 – v)

$\mu = x \pm 2.58\,\sigma$ with99% confidence … (2.3 – vi)

$\mu = x \pm 3.29\,\sigma$ with 99.9% confidence … (2.3 – vii)

However, it is not advisable to estimate the true mean from a single observation *x*. Instead we generally use the sample mean $\bar{x}$ to take the better estimate of population mean μ. We also know that, if the standard deviation of *x* is σ, then the standard deviation of $\bar{x}$ is $\frac{\sigma}{\sqrt{n}}$. Then in terms of $\bar{x}$, the population mean μ lies within the limits.

$\mu = \bar{x} \pm \frac{y\sigma}{\sqrt{n}}$ with the confidence level corresponding to the value y.

That is, confidence interval for $\mu = \bar{x} \pm \frac{y\sigma}{\sqrt{n}}$ or from $\bar{x} - \frac{y\sigma}{\sqrt{n}}$ to $\bar{x} + \frac{y\sigma}{\sqrt{n}}$, and confidence limits can also be expressed with certain confidence level in the same manner.

In the modern practice it is usual to employ a confidence level of 95% for *y* = 1.96 or 99% for *y* = 2.58. Thus, we can say that the population mean μ lies within the limits.

$\mu = \bar{x} \pm 1.96\,\sigma/\sqrt{n}$ with 95% confidence … (2.4a)

It means that it is 95% probable that the population mean μ lies in the interval $\bar{x} - 1.96\,\sigma/\sqrt{n}$ to $\bar{x} + 1.96\,\sigma/\sqrt{n}$. In a similar way the population mean μ lies within the limits.

$\mu = \bar{x} \pm 2.58\,\sigma/\sqrt{n}$ with 99% confidence … (2.4b)

$\mu = \bar{x} \pm 3.29\,\sigma/\sqrt{n}$ with 99.9% confidence … (2.4c)

Consider the following example,

Example 2.2: Measurements of glucose levels in a patient suffering from diabetes gave the following results: 1.108, 1.100, 1.122, 1.088, 1.115, 1.099 and 1.075 g/L. Calculate the 95% confidence interval when *s* = 0.019 g/L.

Solution: $\bar{x} = \frac{1.108 + 1.100 + 1.122 + 1.088 + 1.115 + 1.099 + 1.075}{7}$

$= 1.101\ \text{g/L}$

For 95% confidence the limits are $\pm \frac{1.96\,\sigma}{\sqrt{7}}$ n

$= \pm \frac{1.96 \times 0.019}{2.646}$

$= \pm 0.01407$

$= \pm 0.014\ g/L$

The population mean μ lies within the limits.

$\mu = (1.101 \pm 0.014)$ g/L with 95% confidence

Thus, it is 95% probable that the population mean lies in the interval from 1.087 to 1.115 g/L.

2.4.2 Confidence Interval of Population Mean μ When σ is Not Known

In the usual practice the population standard deviation is not known and can only be approximated for a finite number of measurements by the sample standard deviations. To overcome this difficulty another method is used, which is based on a statistical factor, "t" (also known as student t), that depends on the number of degrees of freedom and confidence level desired. The quantity t is defined as the difference between the two means divided by its standard deviation:

$$t = \sqrt{n}(\bar{x} - \mu)/s \quad \ldots (2.5)$$

Statisticians have compiled the values of t for the given degrees of freedom ($\upsilon = n - 1$) and for various confidence levels desired. For illustration some of the values of t are listed in Table 2.2.

Table 2.2: Values of *t* for υ degrees of freedom and various confidence levels

υ	Confidence		
	90%	**95%**	**99%**
1	6.314	12.706	63.657
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.500
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
$\propto$	1.645	1.960	2.576

A simpler and approximate procedure, particularly useful for small numbers of observations (n < 10) to express confidence interval, is based on the range R (the difference between the largest and smallest values of observations).

According to this to a certain degree of confidence the population mean μ lies within the limits

$$\mu = \bar{x} + C_n R \quad \ldots (2.6)$$

where C_n is constant depending upon the number of observations and the desired confidence level. A few values of C_n are given in Table 2.3.

Table 2.3: Values of C_n for sample of n observations

N	Confidence level		
	90%	**95%**	**99%**
1	3.196	6.353	31.828
2	0.885	1.304	3.008
3	0.529	0.717	1.316
4	0.388	0.57	0.843
5	0.312	0.399	0.628

Note: All methods based on range should be used with caution, since the outlying results may cause uncertainty.

SAQ 2

A replicate analysis of potassium in blood serum yielded concentration of K^+ inmg/100 mL: 15.30, 15.85, 15.55 and 16.30. Calculate the 90% confidence interval for the set. Assume the value of C_n for 4 observations at 90% level = 0.53.

2.5 REJECTION OF DATA

Sometimes, in a set of analytical data there appears a value which is not fitting in the set as it looks at a wide difference from the rest of the values. Now the question arises how to decide whether to remove a result which appears out of line with others when there are no known reasons to suspect it? The question is not of much importance if the number of replicate observations is large since a single value will have only a small effect upon the mean. But it is, of course, important when the number of replicate measurements is small, since here the divergent observation has a significant effect on the value of mean, while at the same time there are insufficient data to decide the fate of the suspected result. In a small set of data the decision for rejection or retention by the blind application of statistical test is no doubt an arbitrary decision.

The criteria for rejection of an observation are based on the supposition that an outlier is due to some systematic source of error. If it is not, then it falls within the random error and should be retained. So many authors agree that the question of rejecting one divergent value from a small sample cannot satisfactorily be answered. It is unfortunate fact that no universal rule can be invoked to settle the question of retention or rejection. Now let us discuss a test based on the range which is called the “Q” test.

2.4.1 The "Q" Test

The criterion used to check the rejection of suspected result (in a set of 3 – 10 results) is the Q-test which is a simple and widely used statistical test. Q, the rejection quotient, is defined as the ratio of the divergence of the suspected value from its nearest neighbour to the range of the set of measured values. If the value of Q_{calc} is greater than the value of Q given in the table at the desired confidence level for the given number of observations, the suspected value is rejected. The Q test is applied as follows:

i) Arrange the observations either in the increasing or decreasing order. The lowest or the highest or both may be the doubtful values.

ii) Calculate the range, R = highest values – lowest value

iii) Find the difference between the suspected value and its nearest neighbour. Call this difference as Y.

iv) Calculate the rejection quotient, Q = Y/R and call it as Q_{calc}.

v) Consult the table of Q (Table 3.4) for the given number of observations and call it as Q_{tab}.

vi) Criterion: Compare Q_{calc} with Q_{tab}. If $Q_{calc.} > Q_{tab}$, then the suspected value is rejected.

Note: Since both lowest and highest values may be considered to be the suspected values, it is advisable to apply the test for both. Say in above if the lowest value was taken as the suspected value we have to extend the test as below.

vii) If the lowest value is rejected, then the range R for remaining values is calculated and if the lowest value is not rejected, then the same range R is used and apply the above procedure steps (ii) to (vi) for the highest value.

viii) Repeat the process further if necessary. Some Q values are given in Table 2.4.

Table 2.4: Values of Rejection Quotient Q at 90% confidence level

Number of observations (n)	**Q**
3	0.94
4	0.76
5	0.64
6	0.56
7	0.51
8	0.47

9	0.44
10	0.41

The following example will illustrate application of the Q test.

Example 2.5: Apply Q-test to check the rejection of the highest value in the following results:

2.18, 2.19, 2.30, 2.15 and 2.20

R = 2.30 – 2.15 = 0.15

Y = 2.30 – 2.20 = 0.10

Q_{calc} = Y/R = 0.10/0.15 = 0.67

From Table 3.4, for n = 5, $Q_{tab.}$ = 0.64, $Q_{calc} > Q_{tab}$, therefore the value 2.30 is rejected at 90% confidence level.

SAQ 3

In replicate determination of iron the following results of percentage of iron were obtained. Should any of the results be rejected by Q test?
%Fe: 52.40, 52.47, 52.50, 52.51, and 52.46.

2.6 SUMMARY

Statistical analysis is necessary to understand the significance of analytical data. In this unit you have studied the methods used by scientists in evaluating the significance of analytical data with the knowledge of probability distribution of errors.

You have learnt here that the study of probability distributions is of fundamental importance to the use of statistics for judging the reliability of analytical data. We have elaborated cases where all measurement errors follow a probability distribution called normal or Gaussian distribution. It can be expressed by a differential equation. Various confidence limits are used to get the confidence interval. The rejection of an outlying observation can be ascertained by the suitable statistical methods.

In analytical chemistry it is frequently desired to compare the results of two different methods, one of them is the test method and the other is usually an accepted method. This is done by testing the null hypothesis using the tests of significance: the t-test which is based on the comparison of two means, and the F-test which is based on the comparison of two variances

2.7 TERMINAL QUESTIONS

1. What are the usual properties of error curves?

2. Explain the Q test.

3. An analyst got the percent alcohol content in a blood sample: 0.084, 0.089 and0.079. Calculate the 95% confidence limit for the mean assuming t = ± 4.30 for two degrees of freedom and 95% confidence.

2.8 ANSWERS

Self Assessment Questions

1. The t-test is used to test the null hypothesis that the means of two sets of data do not differ significantly. The application of t-test will be considered here only in a simple case.

   The F-test serves to show whether the precision of two different methods is the same within specified probability limits. It is applied in terms of variance ratio. The F value is the ratio of two variances in question.

2. Range R = 16.30 – 15.30 = 1.00

   $\bar{x}$ = (15.30 + 15.85 +15.55 + 16.30)/4 = 15.75

   $\mu = \bar{x} \pm C_n R = 15.75 \pm 0.53 \times 1.00$

   15.75 – 0.53 = 15.22 & 15.75 + 0.53 = 16.28

   Therefore, with 90% confidence the population mean μ lies between 15.22 and 16.28 mg/100 mL.

3. The value 52.40 is not fitting well in the other observation, therefore, this value is the suspected value. Omitting this value, the arithmetic mean of the rest is,

   $$\bar{x} = \frac{52.40 + 52.46 + 52.47 + 52.50 + 52.51}{5} = 52.47$$

   Y = 52.46 – 52.40 = 0.06

   R = 52.51 – 52.40 = 0.11

   Q_{calc} = Y/R = 0.06/0.11 = 0.55

For n = 5, Q = 0.64, so that 52.40 is not the suspected value.

Terminal Questions

1. There are numerous properties of an error curve. If a large number of replicate readings are taken of a continuous variable, e.g. the percentage of iron in an iron ore, the results obtained will usually be distributed around the mean in a roughly symmetrical manner. When frequencies of observations (on y-axis) are plotted against values of observations (along x-axis), we get the distribution (draw Fig.2.1) where the dots indicate the frequencies of respective values of observations. Such a distribution of random values (also of random errors) is called the

normal frequency distribution. On drawing the boundaries of these frequencies we get a bell shaped curve that is symmetrical around the mean and is known as the normal distribution curve or a Gaussian Distribution Curve (as this follows the normal frequency curve of Gauss). On either side of the curve there is an exponential decrease in frequency as the magnitude of error increases.

2. The criterion used to check the rejection of suspected result (in a set of 3 – 10 results) is the Q-test which is a simple and widely used statistical test. Q, the rejection quotient, is defined as the ratio of the divergence of the suspected value from its nearest neighbour to the range of the set of measured values. If the value of Q_{calc} is greater than the value of Q given in the table at the desired confidence level for the given number of observations, the suspected value is rejected. Briefly the steps of Q-test is as follows:

   i) The observations are arranged either in the increasing or decreasing order. The lowest or the highest or both may be the doubtful values.

   ii) The range, R = highest values – lowest value is calculated

   iii) The difference Y between the suspected value and its nearest neighbour is calculated.

   iv) The rejection quotient, Q = Y/R or Q_{calc} is calculated.

   v) Criterion: If $Q_{calc.} > Q_{tab}$, then the suspected value is rejected.

   vii) If the lowest value is rejected, then the range R for remaining values is calculated and if the lowest value is not rejected, then the same range R is used .

3. Use the formula for the confidence interval for a particular set of values as given in section 2.4.

   The mean of the sample:

   $= 0.084+0.089+0.0793=0.084$

   From the t table, the t value for 95 percent confidence level is 1.96.

   Substituting the values for confidence interval we get,

   $= 0.084 \pm 1.96 \times 0.005\sqrt{3} = 0.84 \pm 0.0056\%$

Some Useful Books

1. *Analytical Chemistry* by Cary D. Christian , John Wiley and sons.

2. *Basic concepts of Analytical Chemistry* by S.M. Khopkar, New Age International Publishers.

3. *Vogel's Textbook of Quantitative Chemical Analysis* by J. Menham, R.C. Denney, J.D. Barnes and M.J.K. Thomas, 6th Edn, Low Price Edition, Pearson Education Ltd, New Delhi (2000).

UNIT 3

SOLVENT EXTRACTION TECHNIQUE

Structure

3.1 INTRODUCTION

As you know that in this course, you will be studying a variety of analytical methods. And, the substance to be analysed may be a mixture or an impure substance. Hence, the separation of a mixture into its pure components or the separation of a pure substance from an impure one is also an important part of the process of analysis.

Solvent extraction technique is one such method of separation. And as you progress to know more about it during the study, you will realise that this technique is very convenient, simple and quick to use. In analytical chemistry, this technique can be used for purposes of preparation, purification, enrichment, separation in all scales of working from microanalysis to large scale production.

In this unit, we will begin our discussion with the general principle of this technique and then the classification of different types of solvent extraction techniques. The efficiency of solvent extraction will also be explained. Finally, the mechanism of extraction, *i.e.* extraction by *solvation* and *chelation* will be elaborated.

The applications of solvent extraction will be dealt in the next unit, *i.e.* Unit 4.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ explain the term extraction;
- ❖ list various types of extractions;
- ❖ describe liquid-liquid extraction and its importance;
- ❖ discuss the principle of liquid-liquid extraction;
- ❖ explain the efficiency of extraction technique;
- ❖ differentiate between distribution coefficient and distribution ratio and discuss the mechanism of extraction, *i.e.* extraction by solvation and chelation.

3.2 PRINCIPLE

Any substance whether organic or inorganic, can be separated using the extraction technique because it **distributes** or **partitions** itself between two immiscible phases differently. Thus, when a substance S, present in phase 1, generally aqueous, is brought in equilibrium with phase 2, the organic phase, it distributes itself between the two phases, *i.e.*

$$S_{(\text{phase 1})} \quad S_{(\text{phase 2})} \rightleftharpoons$$

Nernst (1891) defined the partition isotherm, "A solute will distribute itself between two essentially immiscible solvents in such a way that the ratio of the concentration of the solute in the two phases, after equilibration has been achieved, at a particular temperature, is constant, provided the solute has the same molecular weight in each phase".

Such a distribution is governed by the **distribution law,** about which you have already studied in the Unit 3 on **phase equilibrium** in the earlier course, BCHCT-135**.**

You also know that for such equilibrium, *at any given temperature, the ratio of activities or concentrations of solute* S *in the two phases is constant.* This is shown below:

$$\frac{[S]_{\text{phase 2}}}{[S]_{\text{phase 1}}} = \frac{[a_s]_{\text{phase 2}}}{[a_s]_{\text{phase 1}}} = \text{constant} = K_{eq} = K_D \qquad \ldots (3.1)$$

The equilibrium constant is called the **distribution coefficient,** K_D or **partition coefficient.**

It is the distribution coefficient**,** whose value decides the mobility of substance S from phase 1 to phase 2. If the value of K_D is *sufficiently large*, only then such a complete mobility or transfer of substance to phase 2 will take place. However, if this value is small, then the substance will remain in phase 1 only.

Also, when we are considering the **distribution coefficient** here, we are assuming that the substance S is present in both the phases in one form only, *i.e.,* it is neither ionising (nor dissociating) nor associating in either of the phases. We will keep the discussion simple here and take such a situation later at appropriate place. So till then, keep this factor also in your mind.

It is also important to know that the *ratio of distribution* of a substance, S, *in two phases* (*i.e.* K_D) *does not depend on the total quantity of* S.

After understanding about the distribution coefficient, let us now study about the classification of extraction technique.

3.3 CLASSIFICATION

Before taking up the classification of solvent extraction, let us first understand the term *extraction*. When the substance to be purified or isolated is present in *one phase,* and a *second phase* is use to obtain it, then we say that it has been *extracted into the second phase.*

Liquid-liquid extraction is used in many analytical separations and this procedure was used from earliest times in technology for the preparation of essential oils, drugs, dye stuffs and perfumes.

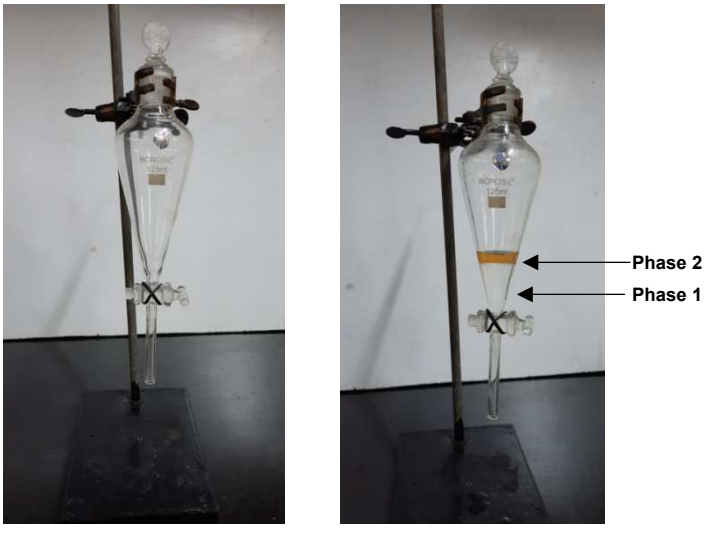
As two phases are involved in the extraction, depending upon their physical states, various types of extractions are possible.

The physical state of the phase in which the sample or substance to be extracted is present is written first and then the physical state of phase used for extraction is written *to denote a particular type of extraction*. A hyphen is used in between the two phases in this representation. For example, if we extract a substance initially present in the **solid phase** by using a **liquid phase**, then we represent it as **solid-liquid extraction**.

The most familiar type of extraction to you may be **liquid-liquid extraction** wherein the substance **to be** separated or extracted is present in the **liquid phase** and it is extracted by using a **second liquid phase.**

The extraction of a substance in a laboratory using a *separatory funnel* is the most familiar and commonly used form of **liquid-liquid extraction.** In this technique, a solute present in phase 1 (say water) is extracted by using the phase 2 (an organic liquid say ether or ethyl acetate).Please note here that the two phases are *immiscible with each other*.

This is shown below in Fig. 3.1. For performing this extraction, the two liquid phases are taken in the separatory funnel and are vigorously shaken. The vigorous shaking increases the surface area between the phases allowing their increased interaction. Then, the contents of the separatory funnel are allowed to rest for sometime so that the two phases get separated.



a) b)

Fig. 3.1: Liquid-liquid extraction using a separatory funnel: a) A separatory funnel and b) Separation of two phases

Can you now guess which phase will be at bottom? Yes, the *denser phase will be at the bottom in the separatory funnel*. Then, by opening the stop cock, the two phases can be separated one by one in different containers.

A solute present in an aqueous solution is extracted in the second phase which is generally organic liquid. Thus, the solute has been extracted from the aqueous phase using a *solvent*. Hence, such an extraction is also called **solvent extraction.**

But has all of the substance has got extracted into the organic phase or phase 2? We will find the answer to this question in the forth coming sections.

Let us now understand the efficiency of the solvent extraction technique.

SAQ 1

In the following types of extractions, identify the phase from which the substance to be extracted is present:

(i) Liquid-Solid (ii) Solid- Liquid (iii) Liquid- Liquid

3.4 EFFICIENCY OF SOLVENT EXTRACTION TECHNIQUE

The distribution of a species in two phases is not governed by a simple equilibrium as discussed in the previous section. In fact, while evaluating the *efficiency* of the solvent extraction, we have to consider the**total concentration** of the substance in each of the phases.

Thus, we need to use another term **distribution ratio, *D.*** The distribution ratio is shown below in Eq. 3.2;

$$D = \frac{[S_{phase\ 2}]_{total}}{[S_{phase\ 1}]_{total}} \quad \ldots (3.2)$$

If phase 2 is an organic solvent and phase 1 is aqueous medium, then for such a system, wecan write the distribution ratio, *D*as shown below:

$$D = \frac{\text{Total conc. of S in organic phase}}{\text{Total conc. of S in aqueous phase}} \quad \ldots (3.3)$$

For simple systems about which we discussed in the previous section, the **distribution ratio** and the **distribution coefficient** are identical.

But for some more complex systems in which the solute or the *substance exists in more than one form* in any one of the phases, the distribution ratio and the distribution coefficient are *not the same***.**

We can understand this with the example of the distribution of a *weak acid* between water and an organic solvent. Weak acids partially ionise in aqueous phase and hence, for a weak acid, HA, we can write the following ionisation in the aqueous medium:

$$HA \rightleftharpoons H^+ + A^-$$

Thus, the concentration of HA in aqueous phase is taken to be the concentration of ***all the species*** of the substance HA, *i.e. the* concentration of *undissociated form* of HA, *i.e.* $[HA]_{aq}$ and *the concentration of conjugate base or anion* of HA, *i.e.*, A^-, represented as $[A^-]_{aq}$.

Then, according to Eq. 3.3, we can write

$$D = \frac{\text{Total conc. of S in organic phase}}{\text{Total conc. of S in aqueous phase}}$$

$$D = \frac{[HA]_{org}}{[HA]_{aq} + [A^-]_{aq}} \quad \ldots (3.4)$$

Looking at Eq. 3.4, you can now appreciate that in such situations, *distribution ratio, D* is ***different*** from the *distribution coefficient*, K_D given by Eq. 3.1.

Remember that the distribution coefficient is equilibrium constant and its value is constant for the distribution of a solute between two phases at a constant temperature. However, the value of distribution ratio depends upon the concentrations of different forms of the species present in the two phases.

The distribution ratio, as explained above is just a ratio of concentrations and *does not depend on the volume ratio of the phases*. However, the *fraction of the substance extracted* or the *extent of extraction* depends on the volume ratios of the two phases used in the extraction.

We will illustrate the efficiency of the extraction by calculating the fraction of the substance transferred from phase 1 to phase 2 is the next unit, *i.e.* Unit 4 in its Sec. 4.3 on Qualitative and Quantitative aspects of Solvent Extraction.

SAQ 2

Which one of the two is larger for a given system involving ionisation of the substance to be extracted in the aqueous phase-distribution constant or distribution ratio?

We will now discuss the mechanism of extraction. But before proceeding to study of mechanism, answer the following SAQ.

3.5 MECHANISM OF EXTRACTION: EXTRACTION BY SOLVATION AND CHELATION

It is important to understand here that the general aspects of solvent extraction discussed so far are equally valid for inorganic and organic compounds. Solvent extraction is a much cleaner way of separating inorganic species as compared to the precipitation methods which involve precipitation followed by filtration and washing of the precipitate.

The use of solvent extraction involves the concept of ‘like dissolves like’. Thus, a substance, whether organic or inorganic, will move to the organic phase when its solubility is higher in the organic phase as compared to the aqueous phase (or medium).

The uncharged organic molecules will tend to dissolve in the organic phase while the anions arising from their ionisation will be charge species and would remain in the polar aqueous phase.

The inorganic species such as metal ions, would like to remain in the aqueous phase as these are charged species. To make them extracted to the organic phase, their charge has to be neutralised so that these behave like the organic species. Two of such approaches are as follows:

(i) Extraction by solvation and

(ii) Extraction by chelation

Let us first understand extraction by solvation and then will discuss extraction by chelation.

(i) Extraction by Solvation

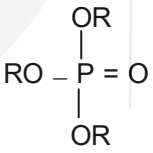
Neutral inorganic molecules or complexes can be solvated using electron donor extractants. The solvation increases their solubility in the organic phase. Such electron donor reagents are of the following two types:

(a) *Reagents containing oxygen linked to carbon, e.g.,* ethers, esters, alcohols and ketones.

(b) *Reagents in which oxygen or sulphur is linked to phosphorous*, *e.g.* alkylphosphates or alkyl thiophosphates.

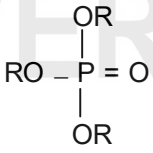
Some examples of such reagents which are extensively used are as follows:

Trialkylphosphate



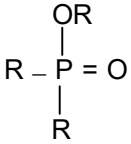
Example: Tri-*n*-butyl phosphate (TBP)

Dialkyl alkylphosphonate



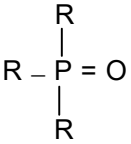
Example: Dibutyl butylphosphonate (TBBP)

Alkyl dialkylphosphinate



Example: Butyl dibutyl phosphinate

Trialkylphosphine oxide



Example: Tri-*n*-octylphosphine oxide (TOPO)

Note that the first three of the above compounds are esters while the last compound is a phosphine derivative. In the last century, another very useful extractant, cyanex 923 was introduced which was a mixture of four trialkylkylphosphine oxides.

Similar to the above reagents containing P=O, reagents with P=S are also possible. These are esters of alkylthiophosphoric, alkylthiophoshonic, alkylthiophosphinic esters and triakylphoshphine sulphides.

You can yourself try writing their structures as given in SAQ below:

Some examples of these reagents are namely cyanamid reagent, triisobutylphoshinesulphide (TIBPS) and cyanex 471 X.

SAQ 3

Write the general structures of following classes of compounds:

(i) Alkylthiophosphoric ester

(ii) Alkylthiophosphonic ester

(iii) Alkylthiophosphinic ester

(iv) Trialkylphosphine sulphide

Since these oxygenated solvents and organophosphorous esters *solvate the proton and the neural inorganic species,* they can be used for the extraction of the acids and salts. Some such compounds along with their extractants are as follows:

- $Ce(NO_3)_4$. $(ether)_2$
- HNO_3. TBP
- VO_2 (NO_3). TBP
- $H^+(TBP)_4.ReO^-_4$

(ii) Extraction of Chelation

Extractions by chelation or using chelating agents are a part of the broad category of *extraction by compound formation*.

Many chelating agents are weak acids which can ionise in water. They react with the metal ions to give the chelates which are uncharged complexes. These are nearly insoluble in water but are highly soluble in organic solvents, *i.e.* ethers, hydrocarbons, ketones, chloroform and carbon tetrachloride.

Many chelating agents are having a limited solubility in water or they may get hydrolysed or may be oxidised by air in their aqueous solutions. Hence, instead of adding them to the aqueous phase, these chelating agents are added to the organic phase. The chelates formed are then extracted to the organic phase.

You can see Fig. 3.1 for these equilibria.

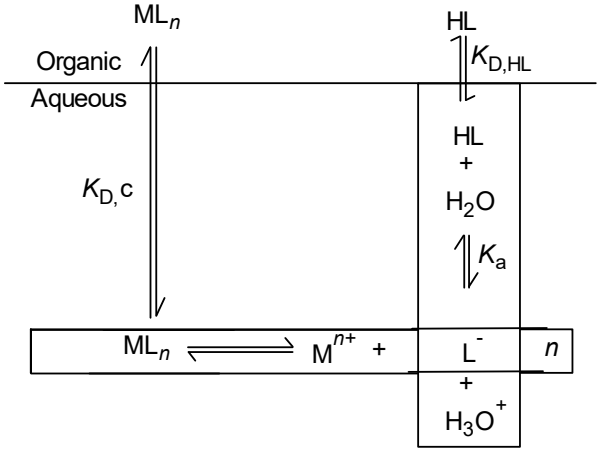


Fig. 3.1: Liquid-liquid extraction of metal ion using a chelating agent

In the aqueous phase, the chelating agent complexes with metal ion to form a stable metal-ligand complex which is then extracted into the organic phase.

The distribution ratio is given as follows:

$$D = \frac{[\text{Conc. of metal chelate in organic phase}]}{[\text{Conc. of metal ion in aqueous phase}} \quad \ldots (3.5)$$

$$= \frac{[ML_n]_{org}}{[M^{n+}]_{aq}} = K \frac{[HL]^n_{org}}{[H^{n+}]^n_{aq}}$$

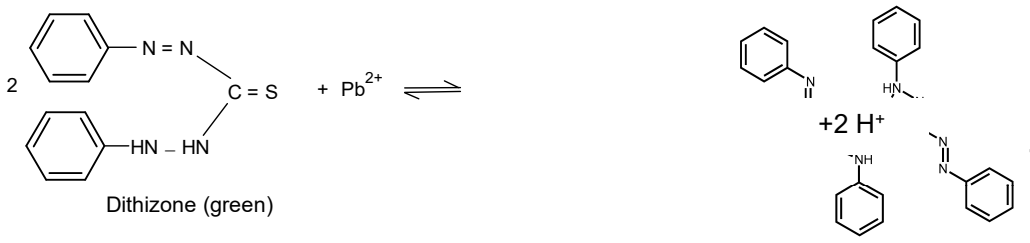
Here, K is a constant containing many terms involved in the equilibria shown in Fig. 3.1.

Normally, the extractions are carried out using a large excess of the chelating agent (HL) so that its concentration remains essentially constant during the extraction.

Some examples of chelating agents are β-diketones, dithizone, 8-hydroxyquinoline,monoximes,dithiocarbamates and hydroxamic acids.

Some other commercial acidic chelating agents are LI X 63,LIX 64 N, LIX65 N, SME 529, P-5000series ,(hydroxyoxime S), Kelex 100 and Hostarex DK16 ,LIX 54 ,X 151 (β-diketones).

Extractions with Diphenylthiocarbazone: Diphenylthiocarbazone, or dithizone, is a useful reagent for separating minute quantities of a dozen or more metal ions. Its reaction with a divalent cation, such as Pb^{2+}, resulting formation of a chelate can be written as shown below:



Dithizone (green)

Both dithizone and its metal chelates are essentially insoluble in water but dissolve readily in solvents such as carbon tetrachloride and chloroform. Solution of the dithizone reagent is deep green, whereas solutions of the metal chelates are intensely red, violet, orange, or yellow and provide a sensitive means for the photometric determination of the separated ions.

The distribution ratio of the various metal dithiionates is described by Eq. 3.5. Thus, for lead,

$$D = \frac{c_{org}}{c_{aq}} = \frac{[PbDz_2]_{org}}{[Pb^{2+}]_{aq}} = \frac{K\, c_{Dz}^2}{[H_3O^+]_{aq}^2} \quad \ldots (3.6)$$

From Eq. 3.6, you can see that the distribution ratio, *D* depends upon the pH of the aqueous phase.

Thus, the pH can be changed or adjusted in a useful way to extract or separate the metal ions if the equilibrium constants of the metal ions under consideration are sufficient different. Thus, the selectivity can be achieved by controlling the pH.

You can see such a selective extraction of metal ions in Fig. 3.2. It shows how the percentages of different metal ions, extracted into a chloroform solution of dithizone, change with the pH of the aqueous medium.

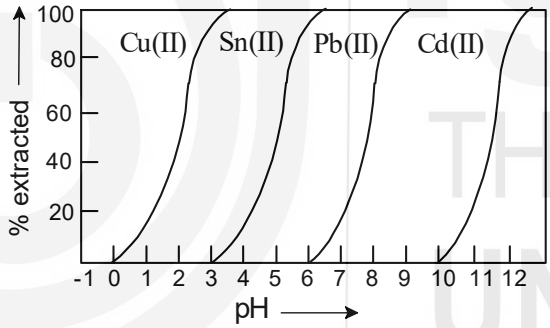


Fig. 3.2: Extraction of different metal ions with change in pH of the aqueous medium.

You will study more about the extraction of metal ions from aqueous solution to the organic phase using different chelating agents in Sec. 4.4 of Unit 4.

If a metal M, present as various species M_1, M_2 M_3 …., is partitioned between an organic and an aqueous phases, then the ***distribution ratio*** or ***extraction coefficient***, *E* can be defined as

$$E = \frac{\text{Total metal conc.in all forms in organic phase}}{\text{Total metal conc.in all forms in aqueous phase}}$$

If polynuclear species are formed, their concentrations are multiplied by appropriate stoichiometric coefficients.

3.6 SUMMARY

In this unit, you have learnt the following main points:

- Various types of extractions can be classified as solid-liquid, liquid-solid, liquid-liquid and gas-liquid extractions.
- Solvent extraction, *i.e.* liquid-liquid extraction is very commonly used in the laboratories for purification of organic compounds. It is also an important technique in industries.
- The liquid-liquid extractions involve partitioning of a substance in two phases which are immiscible with each other.

- For the mobility of any substance from the aqueous phase to the organic phase, its distribution coefficient should be sufficiently large in favour of the organic phase.
- The efficiency of the extraction also takes into account the distribution of the substance.
- Extraction by solvation and by chelation are two important mechanisms of extraction.
- Extraction by solvation can be performed using ethers, esters, alcohols and ketones which are oxygen containing electron donor extractors. The other reagents used contain oxygen and sulphur linked to phosphorous and these are compounds of the type alkylphosphateesters or alkylthiophosphate esters.
- Extraction of metal ions using chelating agents such as *β*-diketones, dithizone and 8-hydroxyquinoline is useful in metal ion separations and it is it is dependent on the pH of the aqueous phase.

3.7 TERMINAL QUESTIONS

1. What is the principle of liquid-liquid extraction?
2. Explain the process of solvent extraction used in laboratories to purify organic compounds.
3. Differentiate between the terms-distribution coefficient and distribution ratio.
4. For which type of systems, the distribution coefficient and distribution ratio are same?
5. Write the structures of the following:

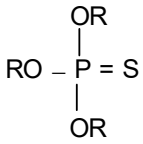
   (i) TBP

   (ii) TBBP

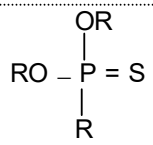
   (iii) TIBPS

3.8 ANSWERS

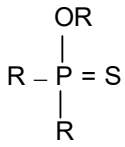
Self Assessment Questions

1. (i) Liquid (ii) Solid (iii) Liquid
2. Distribution coefficient, Refer to Eq. 3.4.
3. (i): Alkylthiophosphoric ester – 

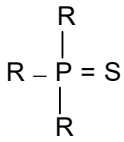
(ii): Alkylthiophosphonic ester –



(iii): Alkylthiophoshinic ester –



(iv): Phosphine sulphide –



Terminal Questions

1. If any substance (solute) is soluble in two immiscible solvents, then it will distribute or partition itself between the two phases (solvents) in a definite proportion. In certain cases, a solute dissolved in one phase i.e. aqueous phase is more or less completely transferred to the second phase which is often an organic liquid such as chloroform, carbon tetrachloride or benzene etc.

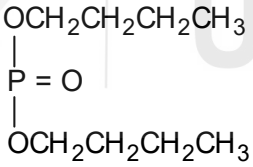
2. Ref. to Fig. 3.1 and the explanation given in Sec. 3.2.

3. *Distribution coefficient* is equilibrium constant and it does not depend upon the quantity of the substance to be extracted.

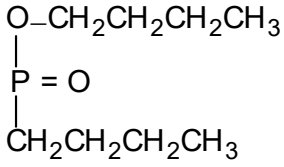
   Distribution ratio is ratio of total concentrations of the substance in the two phases at a particular temperature.

4. For systems which are simple, i.e. in which the substance to be separated exists in one form only and it does not associate or dissociate in any of the phases.

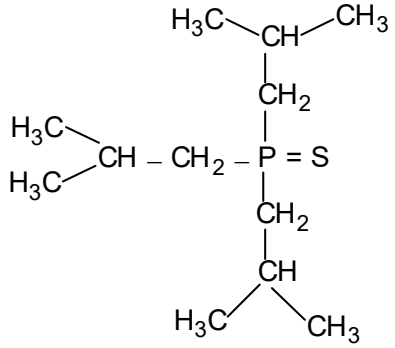
5. (i) TBP –



(ii) TBBP –



(iii) TIBPS –



UNIT 4

EXTRACTION TECHNIQUES: APPLICATIONS

Structure

4.1 INTRODUCTION

In the previous unit, you have studied about the basic aspects of extraction techniques and there we mainly discuss liquid-liquid extraction, *i.e*. solvent extraction. In continuation to that, we will describe the applications of solvent extraction in this unit.

We will begin this unit, by discussing about different ways of carrying out such type of extractions. After that we will deliberate on some qualitative and quantitative aspects of solvent extraction. You will also know how we can efficiently use the available amount of solvent to effectively extract the solute present in one of the liquid phases.

Then, we will discuss the extraction of metal ions from aqueous solution in detail picking up the thread from what are discussed in Unit 3.

Finally, we will describe the extraction of organic species from aqueous and non-aqueous media.

Expected Learning Outcomes

After studying this unit, you should be able to:

- describe the batch, continuous and counter current extraction techniques;
- discuss the efficient use of solvent by carrying out extractors using smaller portion of solvent in more number of lots;
- explain the extraction metal ions using solvent extraction, and
- describe how to extract organic species from aqueous and non-aqueous media.

4.2 TECHNIQUES OF EXTRACTION

You have studied about the term *distribution ratio* in Unit 3. You also know that the liquid-liquid extractions depend upon the *distribution equilibrium between two immiscible phases.*

When a large distribution ratio is readily obtainable for a solute, *i.e.* it is *very high* for the solute (or compound) to be separated from a mixture or from the impurities, then simple extraction called **batch extraction** can be used. Such an extraction is very commonly used in the chemistry laboratories and involves the use of a separatory funnel as shown in Fig. 4.1 below:

		
a) separatory funnel	b) compound in the aqueous phase	c) organic phase added
		
d) shaking of the two phases	e) separation of two phases	f)removal of aqueous phase

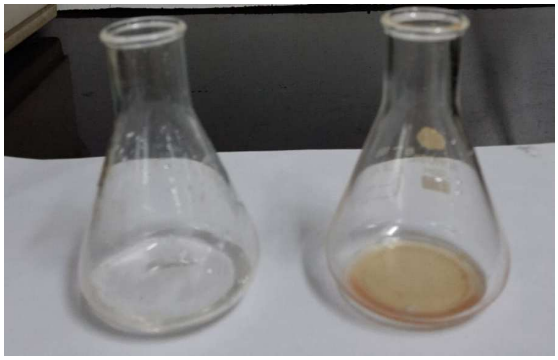
	
g) transfer of organic phase	h) aqueous phase and separated compound in organic phase

Fig. 4.1: Batch Extractions: Steps a) to h) show the extractions process of a compound from aqueous phase using the organic phase.

The length of time required to attain equilibrium must be established by preliminary experiment.

A small number of equilibration steps are sufficient to remove the desired component completely.

The solute which is present in a mixture of compounds (or the impure solute), is taken in one of the phases (say) aqueous phase in the separatory funnel and another phase, *i.e.* organic phase is added in which large distribution ratio of the desired compound (solute) is readily obtainable, see Fig. 4.1 a) to c).

The two phases in the separatory funnel are vigorously shaken so that the solute distributes in the two phases, see Fig. 4.1 d). The separatory funnel is then allowed to rest for some time so that the two phases clearly separate, see Fig. 4.1 e).

The two phases are then taken out from the separatory funnel one by one as shown in Figs. 4.1 f) and g). Thus, the two phases have been separtely obtained and the compound has been separated in the organic layer as shown in Fig. 4.1 h).

If the distribution ratio is greater than 4, then these extractions are adequate for removal of 99% of the substance (solutes) from the aqueous phase.

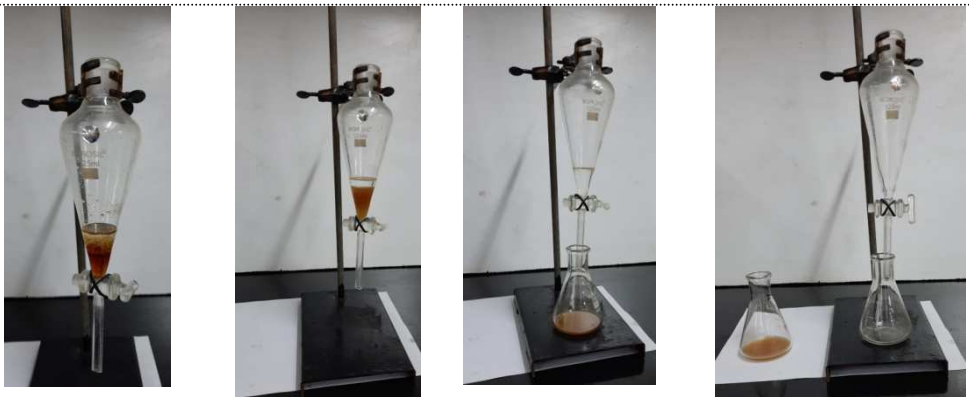
The solute extracted by using such a *single step is called a batch of extraction*.

Still, if some portion of the compound to be extracted is left in the aqueous phase, then the same steps are repeated by using the second lot of the solvent and a second batch of the extracted solute is obtained.

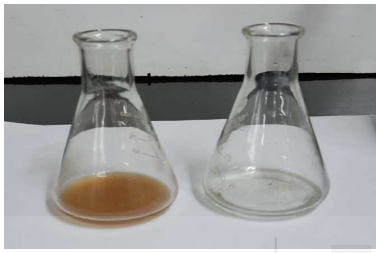
The two batches are then mixed and after the removal of the solvent, the desired solute is obtained as a pure compound.

Here, normally the heavier organic phase, say $CHCl_3$, or CCl_4 is preferred so that when the compound (solute) gets distributed in it favorably; then, it can be taken out from the separatory funnel easily.

However, if the compound (solute) is more soluble in solvents such as ether which is lighter than water, then the solute gets extracted in the upper layer in the separatory funnel. The two layers are then separated and the pure compound is then obtained from the organic layer by removal of the solvent, see Fig. 4.2.



a) b) c) d)



e)

Fig. 4.2: Extraction of a compound from aqueous phase to an organic phase (solvent) which is less dense than water).

In both the situations, the aqueous layer contains the remaining components of the mixture or the impurities which are water soluble.

The second type of extraction called **continuous extraction** is used when the distribution ratio of the desired component is not very favorable, i.e. *low* but the separation factor is high. In such a situation even the usage of large amounts of solvent (second phase) for extraction will not yield appreciable amount of the solute. So a simple extraction is not viable in such situations. A better way is to use *continuous extraction* which is normally carried out using a Soxhlet apparatus as shown in Fig. 4.3.

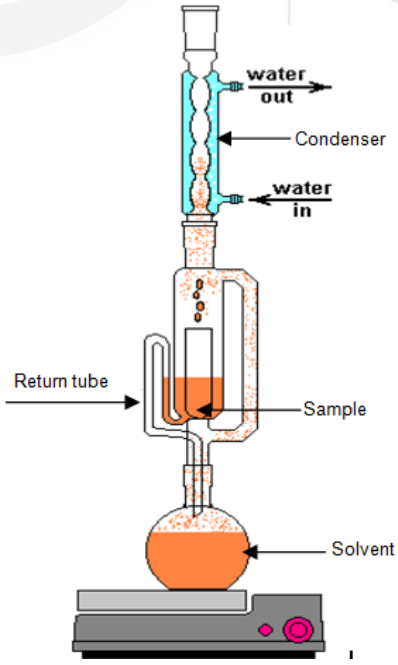


Fig. 4.3: Soxhlet apparatus.

The extracting solvent in this apparatus is taken in a round bottom flask which is heated using a heating mantle. The vapours of the solvent rise and are passed to another tube where these condense. The condensate then continuously passes through the aqueous layer which contains the solute to be extracted.

Thus, continuous passage of condensate through the aqueous layer keeps on *repeating* and the *fresh lot of solvent* enters every time in the form of condensed vapours and extracts the desired solute. Thus, several hundred such extractions occur in a very small time using a very less amount of the solvent.

Baffles are flow-directing or obstructing vanes or panels used to direct a flow of liquid or gas.

The continuous extraction makes use of continuous flow of immiscible solvent phase through the solution to be extracted. If the solvent in volatile, it can be stripped and recycled by distillation. Here, the solute is being removed continuously with the extracting phase In this case, partition equilibrium may not be achieved during the limited time of contact of the two phases. .

The value of partition ratio, the viscosity of the phases, the relative phase volume, and the area of contact between the phases will determine the efficiency of this procedure. As the extractant passes through the solution, baffles and stirrers may be used to bring the two phases into closer, more effective and prolonged contact.

Many continuous extraction devices operate on this general principle. These consist of a flask that serves as a boiler in which the extracting solvent evaporates, condenses and the condensate passes through the solvent to be extracted. (e.g. as in Soxhlet apparatus).

Liquid-liquid partition methods can be extended to the separation of solutes possessing only small differences in their partition coefficients by a method called ***counter current extraction*** or ***distribution***.

The *counter current extraction* uses automated devices. In this methods, *both the phases with fresh portions interact* and the solute distributes between them. This is in contrast to the continuous extraction in which only *the fresh portion of one of the phases* interacts with the solute present in the another phase.

As the name of the *counter current extraction* indicates the two phases move in the *opposite directions in this technique*.

Counter current distribution is a multiple partition process with a large number of stages entirely discontinuous and stepwise in nature. It requires equipment with 100, 200, 300 or more interlocking separating chambers in which equilibrations and phase transfers can be carried out automatically. The mode of separation is an automatic mixing of two phases and transfer of one phase which is moving phase to the subsequent separation chamber.

Compounds with higher K_d value will have the higher migration speeds followed by the compounds with smaller K_d values. All material put into the distribution machine can be recovered.

This method is generally applicable when column or gas chromatography fails. It is not possible to separate compounds of limited solubility, with salts or other strongly polar compounds.

Also, the advantage of this technique is that almost identical solutes having similar distribution ratios can be separated by using this method. The components of the complex organic mixtures, e.g. the naturally occurring substances, are separated by this method.

SAQ 1

What is advantage of using continuous extraction?

4.3 QUALITATIVE AND QUANTITATIVE ASPECTS OF SOLVENT EXTRACTION

The distribution coefficient (distribution ratios) provides us meaningful ways to extract a solute from one phase to another phase.

You know that

$$D = \frac{C_{org}}{C_{aq}}$$ From Unit 3, Eq. (3.3)

If V_{aq} millilitres of water having $a_0$millimoles of a solute HA are extracted with V_{org} milliliters volume of organic phase; then, at equilibrium, if $a_1$millimoles of HA remain in aqueous phase and the number of millimoles transferred to organic phase will be a_0-a_1.

Then, $c_{aq} = \frac{a_1}{V_{aq}}$ and $c_{org} = \frac{a_0 - a_1}{V_{org}}$

We can substitute these concentration expressions in the above equation and rearrange the terms to get the following:

$$D = \frac{a_0 - a_1}{V_{org}} / \frac{a_1}{V_{aq}} \quad \text{… (4.1)}$$

$$= \frac{(a_0 - a_1)V_{aq}}{V_{org} \times a_1}$$

$$a_1 = \frac{(a_0 - a_1)V_{aq}}{V_{org} \times D} = \frac{a_0 V_{aq} - a_1 V_{aq}}{V_{org} \times D}$$

$$a_1 V_{org} D = a_0 V_{aq} - a_1 V_{aq}$$

$$a_1(V_{org} D + V_{aq}) = a_0 . V_{aq}$$

$$a_1 = \frac{V_{aq}}{(V_{org} D + V_{aq})} \times a_0 \quad \text{… (4.2)}$$

If a second extraction is performed using the same volume of the extracting solvent, *i.e*. V_{org}, then the millimoles of a_2 remaining in the aqueous layer can be obtained similar to Eq. 4.2 above as follows:

$$a_2 = \left[\frac{V_{aq}}{(V_{org}D + V_{aq})}\right] = \left[\frac{V_{aq}}{(V_{org}D + V_{aq})}\right]^2 \times a_0 \quad \ldots (4.3)$$

Similarly, after n extractions, we can write the number of millimoles of a, *i.e.*a_n present in the aqueous layer as given in the following expression:

$$a_n = \left[\frac{V_{aq}}{(V_{org}D + V_{aq})}\right]^n \times a_0 \quad \ldots (4.4)$$

If in Eq. 4.4, a_n and a_0 are expressed in terms of concentrations of HA in aqueous layer after n extractions and first extraction, respectively as $a_n = (c_{aq})_n V_{aq}$ and $a_0 = (c_{aq})_0 = V_{aq}$; then Eq. 4.4 becomes

$$(c_{aq})_n = \left[\frac{V_{aq}}{(V_{org}D + V_{aq})}\right]^n \times (c_{aq})_0 \quad \ldots (4.5)$$

Let us perform a separation exercise by using Eq. 4.5.

Suppose, the distribution coefficient for a solute between two solvents-one organic and another water is 85. If 50 mL of organic solvent is available for extraction of I.00 x 10^{-2}M solute from 50 mL of its aqueous solution, then we have the following three options to use the available 50 mL of organic solvent:

(i) All 50 mL of the organic solvent can be used in single step.

(ii) The 50 mL of organic solvent can be divided into two portions of 25 mL each and the extraction is done twice-each time using 25 mL of the solvent.

(iii) The50 mL of organic solvent can be divided into still smaller portions, say, 5 portions of 10 mL each and then, five extractions are done using 10 mL of solvent each time.

Using Eq. 4.5, we can determine the amount of solute X remaining in aqueous layer and from that we can calculate the amount of solute X which has been extracted in the organic solvent.

Let us will work out these values for all the three options listed above, one by one.

(i) $$(c_{aq})_1 = \left[\frac{(V_{aq})}{(V_{org}D + V_{aq})}\right]^n \times (c_{aq})_0$$

$$= \frac{50}{50D + 50} \times 1.00 \times 10^{-2}\text{M}$$

$$= \frac{50}{50 \times 85 + 50} \times 1.00 \times 10^{-2}\text{M}$$

$$= 1.16 \times 10^{-4}\text{ M}$$

So, the concentration of X in organic layer will be = 1.00 x 10^{-2} M^{-1}.16 x 10^{-4}M

= 9.8837 x 10^{-3}M

(ii) $$(c_{aq})_2 = \left[\frac{(V_{aq})}{(V_{org}D + V_{aq})}\right]^2 \times (c_{aq})_0$$

$$= \left[\frac{50.0}{25 \times 85 + 50}\right]^2 \times 1.00 \times 10^{-2}\text{M} = 5.28 \times 10^{-6}\text{ M}$$

Thus, the concentration of X In organic layer will be

$$= 1.00 \times 10^{-2} - 5.28 \times 10^{-6}\text{M}$$

= 9.9947 x 10^{-3} M

(iii) $$(c_{aq})_5 = \left[\frac{(V_{aq})}{(V_{org}D + V_{aq})}\right]^5 \times (c_{aq})_0$$

$$= \left[\frac{50.0}{10.0 \times 85 + 50.0}\right]^5 \times 1.00 \times 10^{-2}$$

$$= 5.29 \times 10^{-9}\text{ M}$$

Thus, the concentration of X In organic layer will be

=1.00 x 10^{-2} M - 5.29 x 10^{-9} M

= 9.9999 x 10^{-3} M

From the above calculations, we can see that *the more number of extractions using the smaller lots of the solvent, the more will be the amount of the solute extracted in the organic layer.*

Thus, by using the same amount of solvent (e.g.50 mL in the above case), we can plan more efficient extraction using smaller lots of solvent and carrying out extractions in multiple steps.

Separation factor, β is $= \frac{D_A}{D_B}$, the ratio of distribution of solutes A and B phases.

If D_A = 10 and D_B = 0.1, a single extraction will remove 90.91% of A and 9.09% B. The second extraction of the same phase will extract the total amount of A up to 99.18% but increase the of B to 18.0%. More complete extractions of A, thus, involves an increased contamination by B.

Deduction of Degree of Extraction, R

The *total weight of metal extracted into the organic phase* is called degree of extraction, *R*.

$$R = \frac{w_{org}}{w_{org} + w_{aq}}$$

This can be written in the following form:

$$R = \frac{EV_{org}}{EV_{org} + V_{aq}} = \frac{E}{E + V_{aq}/V_{org}}$$

The values of *R* between 90 and 100% denote complete extraction and corresponding values of *E* range from 99 to infinity (∞) (for equal phase volumes). So, *R* is more useful than *E* for practical purposes.

% extraction (*E*) is related to ***distribution ratio, D*** by the expression given below:

$$D = \frac{(V_{aq}/V_{org})E}{100 - E} \text{ where } \begin{matrix} V_{aq} - \text{Volume aqueous phase} \\ V_{org} - \text{Volume organic phase} \end{matrix}$$

Further, the extraction is considered to be quantitative when E = 100.

$$D = \frac{100}{100 - 10} = \frac{100}{0} = \infty \text{ (infinity)}$$

If D = 10, then $10 = \frac{E}{100 - E}$ *or* $E = 90.9\ \%$

$$D = 0.1,\ 0.1 = \frac{E}{100 - E},\ \textit{or}\ E = 9.09\ \%$$

SAQ 2

50 mL of an aqueous solution of substance Y are extracted with 40.0 mL of an organic solvent. The distribution coefficient for this system is 9.6. What will be quantity of the substance Y remaining in the aqueous phase if the extraction is carried out by using the solvent in the following portions:

(i) 20.0 mL portion of solvent two times, and

(ii) 5.00 mL portion of solvent eight times ?

SAQ 3

In the extraction of cerium (IV) with 2-thenonyl trifluoroacetone in benzene, the volumes of aqueous and organic phases were 25 mL when the % extraction was 99.8%. Calculate the distribution ratio.

4.4 EXTRACTION OF METAL IONS FROM AQUEOUS SOLUTION

You have already studied about the extraction of metal ions using solvation and chelation in Sec. 3.5 of Unit 3. In continuation to that we will explore more about the extraction of metals in this section.

We will first discuss the *two cases* of *extraction of metal chlorides and nitrates using solvation* and then, the *extraction of metal ions by chelation* will be elaborated.

(a) Extraction of metal chlorides and nitrates using solvation

(i) Extraction of Metal Chlorides

A large number of metal chlorides can be extracted from 6 M hydrochloric acid solution into diethyl ether. These include iron (III), titanium (III), gold (III), tin (IV), antimony (V) and molybdenum (VI) which are extracted to the extent of more than 50%.The iron (III) is extracted to the extent of 99% whereas gallium(III) is 97% extracted.

The species extracted in case of iron (III) is the ion-pair H_3O^+ $FeCl^-_4$. The percentage of iron extracted is also dependent on the *molarity* of HCl solution as little iron is removed when solutions below 3 M HCl or above 9 M HCl are used.

Instead of ether, methyl isobutyl ketone can also be used for the extraction of iron from HCl solutions. Note that methyl isobutyl ketone is *less flammable* than ether.

It may also be important to know that many other ions such as Al (III) and cobalt, lead, manganese and nickel in the form of their divalent cations are *not* extracted by ether.

(ii) Extraction of Nitrates

Diethyl ether and some other solvents selectively extract some nitrate salts. For example, Uranium (VI) can be separated from lead and thorium by ether extraction of an aqueous solution saturated with ammonium nitrate and having a nitric acid conc. of 1.5. M. under these conditions, bismuth and iron (III) are also extracted to some extent.

(b) Extraction of metal ions by Chelation

As you have studied in Unit 3, a large number and variety of chelating agents is available. Some such examples being oxime and dithizone.

Dithizone and its metal ion complexes are water insoluble but are soluble in chloroform and carbon tetrachloride. The metals Fe, Co, Ni, Cu, Zn, Ag, Cd, Pb and Sn form metal chelates and can then be extracted.

4.5 EXTRACTION OF ORGANIC SUBSTANCES

In the last section, you have studied about the extraction of metal ions from aqueous solution. The extraction of organic compounds using solvent extraction is very important technique used in the chemistry laboratories. Here we will deal with extraction of organic compounds from aqueous and non-aqueous media. We will keep the examples simple to show the utility of this technique in the separation of organic compounds.

Let us first take the typical example of extraction of an organic compound from aqueous medium.

4.5.1 Extraction from Aqueous Medium

Suppose we have carried out a synthesis of an organic compound and it is present in the aqueous phase along with the impurities.

For such a compound which is soluble in organic solvent and on the other hand, *the impurities* being *not soluble* in the organic solvent, we can obtain the pure compound by extracting it into an organic solvent. The contents of the reaction mixture after the reaction*, i.e*. the product obtained and impurities are shaken with the organic solvent in the separatory funnel. The desired product will get transferred to the organic solvent while the impurities will remain in the

aqueous layer. The two layers can then be separated and pure product is obtained after the removal of the organic solvent, see Figs 4.1 and 4.2.

Let us study another situation when more than one compound is present in the non-aqueous medium and we want to get each of these compounds in the pure form.

4.5.2 Extraction from Non-aqueous Medium

Such a separation of organic compounds involves acid-base chemistry also in addition to the use of the solvent extraction.

Let us take the simplest situation of an organic phase containing two compounds-one being *acidic* and the other being *neutral*, say benzoic acid and hexane. Here, we can take advantage of the acidic nature of benzoic acid. We can take the mixture of the two compounds in the separatory funnel and add dilute NaOH solution to it. Benzoic acid will form sodium benzoate and it will pass to the aqueous layer. The two layers can then the separated.

Hexane will be present in the separate organic layer. From the aqueous phase, sodium benzoate can then be converted to benzoic acid by treatment with mild acid. Benzoic acid from the aqueous layer can then be extracted again using an organic solvent to obtain pure benzoic acid.

Similarly, *basic organic compounds* such as amines, can be separated from the neutral ones when both of them are present in a mixture. In such a situation, the organic solvent is added to such a mixture. The mixture is then shaken with dilute HCl and the aqueous layer containing salt of the amine can be separated using liquid-liquid extraction.

The neutral compound will remain in the organic layer. Again, the original amine can be regenerated from its salt by treating the aqueous layer obtained above with mild alkali. The amine can then be extracted from the aqueous layer by using the organic solvent. The amine is then obtained in the pure form after removal of the organic solvent.

Similarly, we can extend this situation further when in an organic solvent *phenol, a carboxylic acid* and *a neutral compound* are present as a mixture. Let us see how we can proceed for the separation of the different components of such a mixture.

Here, we can first take this mixture in the separating funnel and shake it with sodium bicarbonate solution. The benzoic acid will be transferred to the aqueous layer as its sodium salt. The other two compounds will remain in the organic layer. The benzoic acid can be recovered from its sodium salt after acidification with the mild acid.

The organic layer can then be treated with dilute NaOH and this time, phenol will pass on to the aqueous layer as its sodium salt. The phenol can then be obtained from its sodium salt by acidification of the sodium salt. The phenol so obtained in the aqueous layer can then be extracted with an organic solvent. The solvent can then be removed to yield the phenol.

The third compound, the neutral component in this mixture has remained in the original organic layer and can be obtained from there.

Thus, we can separate different organic compounds from a mixture depending upon the solubility of these compounds and using acidic/basic behaviour.

SAQ 4

How will you separate salicylic acid from 1,4-dibromobenzene?

4.6 SUMMARY

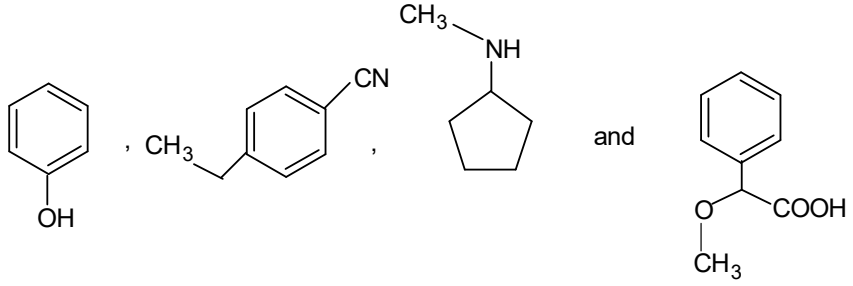
In this unit, we have learnt that

- liquid-liquid extraction can be done in the following three ways:
  - Simple batch extraction
  - Continuous extraction
  - Counter-current extraction
- Extractions can be carried out more efficiently by using small portions of the solvent and carrying out multiple extractions rather than using whole of the solvent in a single extraction. It is true when the distribution ratio is not very favourable.
- Metal ions can be extracted using solvation mechanism and chelate formation. Many reagents are available for this purpose and different reagents are used for different metal ion extractions.
- Organic substances can be extracted from aqueous medium. The mixture of organic substances can be successfully separated into pure components by using their acidic or basic behaviour and solubility characteristics using the technique of solvent extraction.
- Solvent extraction is a simple technique having a wide range of applications in different areas.

4.7 TERMINAL QUESTIONS

1. Differentiate between continuous and counter-current extractions.
2. 2.50 mL of water containing 0.1 g of I_2 is extracted with 25 mL of carbon tetrachloride. The distribution ratio, *D* of I_2 between water and CCl_4 is 1/85. Calculate the amount of I_2 remaining in the aqueous solution after extraction with 25 mL of CCl_4.

   Also calculate the amount of I_2 remaining in the aqueous solution after three extractions using 8.33 mL of CCl_4each time.
3. In solvent extraction of uranium with 8-hydroxyquinoline in chloroform, the distribution ratio was 999.0.The volume of aqueous and organic phases used are 25 mL and 10 mL, respectively. Calculate the % extraction.
4. You are given a mixture of the following four compounds present in dichloromethane.



How will you separate them using solvent extraction? Draw a flow chart for the same.

4.8 ANSWERS

Self Assessment Questions

1. It involves small amount of solvent and takes very less time to execute several hundred extractions.

2. (i) 6.40×10^{-3} M and (ii) 6.89×10^{-4} M

3. $V_{aq} = V_{org} = 25\,mL, \quad E = 99.8\%$

$$D = \frac{(V_{aq}/V_{org})E}{100 - E} = \frac{25/25 \times 99.8}{100 - 99.8} = \frac{99.8}{0.2} = 499$$

4. We can do this by taking this mixture in a separatory funnel and then adding an organic solvent to it. Then, the contents can be treated with dil. NaOH (say 5%) solution when sodium salt of salicylic acid will pass into the aqueous phase leaving 1, 4-dibromobenzene in the organic layer.

   The aqueous layer can then be separated from the organic layer. The acidification of the aqueous layer will yield salicylic acid which can then be extracted using an organic solvent. The removal of solvent will then yield pure salicylic acid.

   Similarly, removal of organic solvent from the organic layer containing 1,4-dibromobenzene will yield 1,4-dibromobenzene.

Terminal Questions

1. The continuous extraction involves interaction of fresh portion of one of the phases with the other phase while in counter-current extraction fresh portions of each of the phases interact and both the phases interact with each other in the *counter-current way*.

2. Suppose x_1 g of I_2 remains in 50 mL of water, its concentration is

   $x_1/50$ g per mL.

   The concentration of I_2 is CCl_4 layer will be $= (0.1 - x_1)/25$ g per mL

$$D = \frac{x_1/50}{(0.1 - x_1)/25} = \frac{1}{85}$$

   or $x_1 = 0.0023$ g

In the second case, the concentration of I_2 in aqueous solution after three extraction taking 8.33 mL each time, is given by

$$x_3 = 0.1\left(\frac{1/85 \times 50}{(50/85 + 8.33}\right)^3 \text{ or } \qquad x_3 = 0.0000145\ \text{g}$$

This extraction is regarded as virtually complete in this case.

3. $D = 999$ and $V_{aq} = 10$ mL

$V_{org} = 25$ mL

Let $E = x$ (% of extraction)

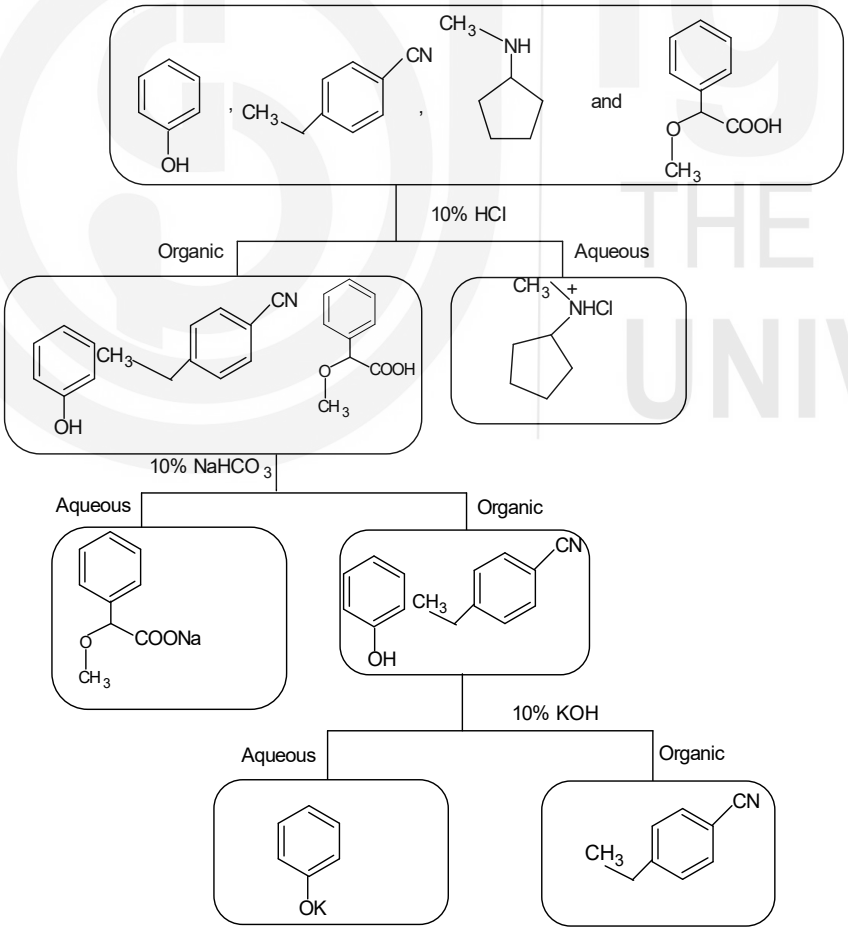
$$D = \frac{(V_{aq}/V_{org})E}{100 - E} = 999$$

$$999 = \frac{(25/10)\,x}{100 - x} \quad \text{or } 2.5\,x = 999 \times 100 - 999\ x$$

or $1001.5\,x = 999 \times 100$

$$x = \frac{999 \times 100}{1001.5} = 99.75\%$$

4.



ignou
THE PEOPLE'S
UNIVERSITY