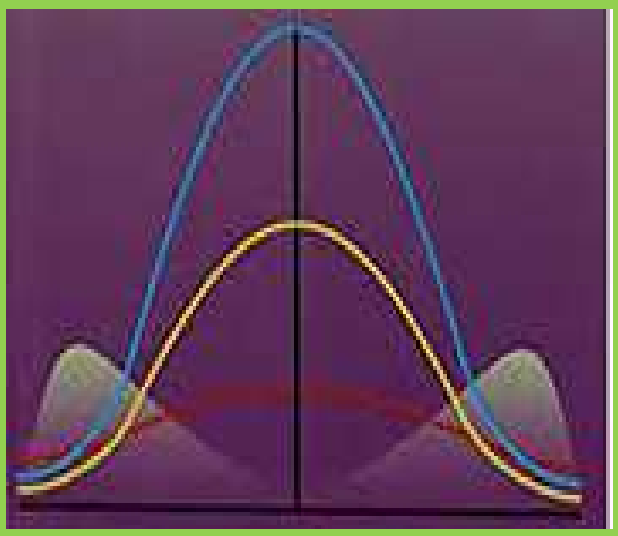
BSCZO- 303

B. Sc. III YEAR

BIOSTATISTICS, INSTRUMENTATION AND TECHNIQUES



DEPARTMENT OF ZOOLOGY
SCHOOL OF SCIENCES
UTTARAKHAND OPEN UNIVERSITY

BSCZO-303

BIOSTATISTICS, INSTRUMENTATION AND TECHNIQUES

DEPARTMENT OF ZOOLOGY
SCHOOL OF SCIENCES
UTTARAKHAND OPEN UNIVERSITY

Phone No. 05946-261122, 261123
Toll free No. 18001804025
Fax No. 05946-264232, E. mail info@uou.ac.in
htpp://uou.ac.in

MEMBER OF THE BOARD OF STUDIES

PROGRAMME COORDINATOR

Dr.Shyam S. Kunjwal
Department of Zoology
School of Sciences
Uttarakhand Open University
Haldwani (Nainital).

Unit Writer — Unit no.

1. Dr.Lakhan Singh — **Unit: 1, 2&3**
Assistant Professor
HAPPRC
HNB Garhwal University,
Uttarakhand

2. Dr.Harish Chandra — **Unit: 4, 5& 6**
Assistant Professor
HAPPRC
HNB Garhwal University,
Uttarakhand

COURSE EDITOR

Dr.Anju Thapliyal
Department of Zoology
BGR Campus Pauri

Dr.Shyam Kunjwal
Department of Zoology
Uttarakhand Open University

Course Title and Code	:	**Biostatistics, Instrumentation and Techniques (BSCZO303)**
ISBN No.	:	**978-93-90845-33-0**
Copyright	:	**UTTARAKHAND OPEN UNIVERSITY**
Edition	:	**2019**
Published by	:	**UTTARAKHAND OPEN UNIVERSITY, HALDWANI, NAINITAL- 263139**
Printed by	:	

CONTENTS

COURSE 1: BIOSTATISTICS, INSTRUMENTATION AND TECHNIQUES

COURSE CODE: BSCZO303 **CREDIT: 3**

UNIT-1 DATA COLLECTION AND PRESENTATION

CONTENTS

1.1 OBJECTIVES

Following are the objectives of this chapter:

1. To know the definitions of Statistics and Biostatistics.
2. To know about some statistical symbols.
3. To know the scope and applications of Biostatistics.
4. To know about data, its collection and collection techniques.
5. To know about organization and representation of data by many graphical techniques such as histogram, pie chart, frequency polygon etc.

1.2 INTRODUCTION

We welcome the reader who wishes to learn biostatistics. In this chapter we introduce you to the subject. First of all we define statistics and biostatistics and then examples are given where bio- statistical techniques are useful. These examples show that biostatistics has an importance in advancing our biological knowledge; biostatistics helps to evaluate many life-and-death issues in medicine.

We advise you to read the examples carefully and then think yourself, "What can be inferred from the information presented?" What would you do with the data after they are collected? How can it be presented and what you can get from it? We want you to realize that biostatistics is a tool that can be used to benefit you and society.

There is no royal road to biostatistics. You need to be involved. You need to work hard. You need to think. If you analyze the actual data, the result will be a powerful tool that has immediate practical uses. Our main purpose is to develop thought patterns in your mind that are useful in evaluating information in all areas of your life.

1.3 DEFINITIONS OF STATISTICS AND BIOSTATISTICS

Much of the joy and pain in life arises in situations that involve considerable uncertainty. Here we are giving two situations which show that the study of statistics and biostatistics is necessary.

1. Parents of a child with a genetic defect consider whether or not they should have another child. They will base their decision on the chance that the next child will have the same defect.

2. To choose the best therapy, a physician must compare the diagnosis or future course, of a patient under several therapies. A therapy may be a success, a failure, or somewhere in between; the evaluation of the chance of each occurrence necessarily enters into the decision.

1.3.1 DEFINITION OF STATISTICS

Statistics is the science which deals with the collection, classifying, presenting, comparing and interpreting numerical data collected to throw light on any sphere of enquiry- Lovitt.

The science of statistics is a most useful servant, but only of great value to those who understand its proper use- W.I.King. Statistics provides tools and techniques for research workers- A.M. Mood. Planning is the order of the day and without statistics planning is inconceivable- L.H.C. Tippet.

Statistics may be defined as a science of numerical information which employs the process of measurement and collection, classification, analysis, decision making and communication of results in a manner understandable and verifiable by other- Cecil H. Meyers

1.3.2 DEFINITION OF BIOSTATISTICS

Biostatistics is the application of statistics methods applied to biological areas. Biological laboratory experiments, medical research (including clinical research), and health services research all use statistical methods. Many other biological disciplines rely on statistical methodology.

There are three reasons for focusing on biostatistics:

1. Some statistical methods are used more deeply in biostatistics than in other fields. For example, a general statistical textbook would not discuss the life-table method of analyzing survival data of importance in many bio-statistical applications. The topics in this book are adapted to the applications in mind.

2. Examples are drawn from the biological, medical, and health care areas; this helps you maintain motivation. It also helps you in understanding how to apply statistical methods.

3. A third reason for a book on biostatistics is to teach the material to the audience of health professionals. In this case, the interaction between students and teacher, but especially among the students themselves, is of great value in learning and applying the subject matter.

1.4 STATISTICAL SYMBOL

1.4.1 STATISTICAL SYMBOL

Some of the statistical symbols which are useful to biostatistics students are:

f: Frequency of the variate

$\bar{x}$: Arithmetic Mean of a given set of values or of a distribution

M_e : Median of a given set of values or of a distribution

M_o : Mode of a given set of values or of a distribution

σ : Standard Deviation of a given set of values or of a distribution

σ^2 : Variance of a given set of values or of a distribution

Σ : Sum of all the values of a given set

Q.D.: Quartile deviation of a given set of values or of a distribution

M.D.: Mean deviation of a given set of values or of a distribution

1.4.2 SCOPE OF BIOSTATISTICS

Biostatistics is the application of statistics in different fields of biology. The science of biostatistics includes the design of biological experiments, especially in medicine, pharmacy, agriculture, forestry, environmental science, fishery etc; the collection, summarization, and analysis of data from those experiments; and execute interpretation and inference from the results. A major branch of this is medical biostatistics, which is exclusively concerned with health and medical sciences.

In current world, the scope of biostatistics is increasing rapidly. If we discuss about biostatistics, we see that almost all educational programmes in biostatistics are at postgraduate level. They are most often found in schools of public health, affiliated with schools of medicine, forestry, or agriculture, or as a focus of application in departments of statistics.

In larger universities where both a statistics and a biostatistics department exist, the degree of integration between the two departments may range from the bare minimum to very close collaboration. In general, the difference between a statistics program and a biostatistics program is twofold: (i) statistics departments will often host theoretical/methodological research which are less common in biostatistics programs and (ii) statistics departments have lines of research that may include biomedical applications but also other areas such as industry (quality control), business and economics and biological areas other than medicine

There is a special need of the subject bio statistics because it related with such areas as medical , pharmacy, forestry, agriculture, etc, which are very necessary for the betterment of society.

1.4.3 APPLICATION OF BIOSTATISTICS

The importance and application of statistics in the field of biology is increasing day by day. Why it is so? The reason is that in biology the interplay of casual and response variables follow the laws that are not in the classic mold of 19^{th} century physical science. In

that century, biologists such as Robert Mayer, Helmholtz, and others in trying to show that biological process were nothing but physicochemical phenomena, helped create the impression that the experimental methods and natural philosophy that had led to such dramatic progress in the physical sciences should be imitated fully in biology.

Many biologists even to this day have retained the tradition of strictly mechanistic and deterministic concepts of thinking, while physicists, as their science became more refined and came to deal with ever more elementary particles, began to resort to statistical approaches. In biology most phenomena are affected by many casual factors, uncontrollable in their variation and often unidentifiable. Statistics is needed to measure such variable phenomena with a predictable error and to ascertain the reality of minute but important differences.

A Biostatistics centre could jointly organize working groups, the seminar series, computing infrastructure and possibly consulting and clinical trials coordinating centre cervices. The main objective of the centre would be to estimate, collaborate on, and circulate results of research in a particular subspecialty in the following reasons:

1. Statistical methods for longitudinal studies;
2. Statistical genetics;
3. Foundations of inference;
4. Bayesian biostatistics
5. Biostatistician practice and education.

The most critical short term problem in the field of biostatistics is the information system. We need to incorporate modern, web-based technologies into the everyday workings of the department of biostatistics. We need reliable and accessible systems that are competitive with those available to departments of statistics and biostatistics. We likely build collaborations with computer science students.

1.5 DATA AND ITS TYPES

1.5.1 DATA

The information collected from census or surveys or from other sources is called raw data. The word data means information. The adjective raw attached to data indicates that the information collected cannot be used directly. It has to be converted into more suitable form before it begins to make sense to be utilized gainfully. Raw data is like raw rice. Raw rice has to be cooked properly and tastefully before it is eaten and digested. Similarly, raw data has to be converted into proper form such as tabulation, frequency distribution form, etc, before any inference is drawn from it.

There are two ways of statistical data;

1. Primary data 2. Secondary data

Primary data: It is the data collected by some person or organization for his own use from any primary source. For example the data of census report collected by the centre government, the data collected by any agency for its own purpose, the gadget of India, etc.

Secondary data: It is the data collected by some other person or organization for their own use but the investigator also gets it for his own use. For example the data collected by any medical agency can be used by some other medical institute students.

In other words, primary data are those data which are collected by you to meet your own specific purpose where as the secondary data are those data which are collected by somebody else. A data can be primary for one purpose and secondary for the other.

1.5.2 TYPES OF DATA

The primary data are of the following types:

1.5.2.1 Nominal Data

In the study of biostatistics, we meet many different types of numerical data. The different types have varying degrees of structure in the relationships among possible values. One of the simplest types of data is **nominal data,** in which the values fall into unordered categories or classes. In a certain study, for instance, males might be assigned the value 1 and females the value 0. Numbers are used mainly for the sake of convenience; numerical values allow us to use computers to perform complex analysis of the data. Nominal data that take on one of two distinct values-such as male and female are said to be **dichotomous or binary,** depending on whether the Greek or the Latin root for two is preferred. However, not all nominal data need be dichotomous. Often there are three or more possible categories into which the observations can fall. For example, persons may be grouped according to their blood type, such that 1 represents type O, 2 is type A, 3 is type B, and 4 is type AB.

1.5.2.2 Ordinal Data

When the order among categories becomes important, the observations are referred to as **ordinal data.** For example, injuries may be classified according to their level of severity, so that 1 represents a fatal injury, 2 is severe, 3 is moderate, and 4 is minor. Here a natural order exists among the groupings; a smaller number represents a more serious injury. A second example of ordinal data is Eastern Cooperative Oncology Group's classification of patient performance status.

Status 0: Patient fully active, able to carry on all predisease performance without restriction.

Status 1: Patient restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature.

Status 2: Patient ambulatory and capable of all self-care but unable to carry out any work activities; up and about more than 50% of waking hours.

Status 3: Patient capable of only limited self-care; confined to bed or chair more than 50% of Waking hours.

Status 4: Patient completely disabled; not capable of any self-care; totally confined to bed.

1.5.2.3 Ranked Data

In some situations, we have a group of observations that are first arranged from highest to lowest according to magnitude and then assigned numbers that correspond to each observation's place in the sequence. This type of data is known as **ranked data.** As an example, consider all possible causes of death in the India. We could make a list of all of those causes, along with the number of lives that each one claimed in. If the causes are ordered from the one that resulted in the greatest number of deaths to the one that caused the smallest and then assigned consecutive integers, the data are said to have been ranked.

1.5.2.4 Discrete Data

For **discrete data,** both ordering and magnitude are important. In this case, the numbers represent actual measurable quantities rather than mere labels. In addition, discrete data are restricted to taking on only specified values-often integers or counts-that differ by fixed amounts; no intermediate values are possible. Examples of discrete data include the number of motor vehicle accidents in Dehradun in a particular month, the number of times a woman has given birth, the number of new cases of tuberculosis reported in the India during a one-year period, and the number of beds available in a particular hospital. Note that for discrete data a natural order exists among the possible values. If we are interested in the number of times a woman has given birth, for instance, a larger number indicates that a woman has had more children. Furthermore, the difference between one and two births is the same as the difference between four and five births. Finally, the number of births is restricted to the nonnegative integers; a woman cannot give birth 3.4 times because it is meaningful to measure the distance between possible data.

1.5.2.5 Continuous Data

Data that represent measurable quantities but are not restricted to taking on certain specified values (such as integers) are known as **continuous data.** In this case, the difference between any two possible data values can be arbitrarily small. Examples of continuous data include time, the serum cholesterol level of a patient, the concentration of a pollutant, and temperature. In all instances, fractional values are possible. Since we are able to measure the distance between two observations in a meaningful way, arithmetic operations can be applied. The only limiting factor for a continuous observation is the degree of accuracy with which it can be measured; consequently, we often see time rounded off to the nearest second and weight to the nearest pound or gram. The more accurate our measuring instruments, however, the greater the amount of detail that can be achieved in our recorded data.

In a study of the effects of maternal smoking on newborns, for example, we might first record the birth weights of a large number of infants and then categorize the infants into three groups: those who weight less than 1500 grams, those who weight between 1500 and 2500 grams, and those who weight more than 2500 grams. Although we have the actual measures of birth weight, we are not concerned with whether a particular child weighs 1560 grams or 1580 grams; we are only interested in the number of infants who fall into each category. From prior experience, we may not expect substantial differences among children within the very low birth weight, low birth weight, and normal birth weight groupings. Furthermore, ordinal data are often easier to handle than continuous data and thus simplify the analysis. There is a consequent loss of detail in the information about the infants, however. In general, the degree of precision required in a given set of data depends on the questions that are being studied.

As we progressed, the nature of the relationship between possible data values became increasingly complex. Distinctions must be made among the various types of data because different techniques are used to analyze them. It does not make sense to speak of an average blood type of 1.8; it does make sense, however, to refer to an average temperature of 4.55°C.

1.6 DATA COLLECTION AND RELATED TERMS

1.6.1 POPULATION, SAMPLE, SAMPLING UNIT AND SAMPLING FRAME

1.6.1.1 Population

Population is a collection of units or objects of which some property is defined for every unit or object. Population may consist of finite or infinite number of units. Population is also called universe by a number of statisticians and scientists.

The inhabitants of a region, number of wheat fields in a state or district, fruit plants in a city, number of students in a institution, insects in a field, persons suffering from any particular disease, workers in a institution, total no of person in city, total households, total no of students in any university, are a few examples of finite populations. All real numbers, all stars in the sky are examples of infinite populations. Generally, the population has a large number of animates and inanimates. Moreover, the units or subjects constituting the population may vary from survey to survey in the same region of activity depending upon the aims and objective of the survey.

In brief, one should very well keep in mind that statistical population is not only the human population which is usually considered in literary sense. It is generally a group or collection of items specified by certain characteristics or defined under certain restrictions.

1.6.1.2 Sample

A sample is the portion of the population that is examined to make inferences about the population or a part or fraction of population, which represent it, is known as sample. Sample

consists of a few items of the population. In principle a sample should be such that it is a true representative of the population.

If poll stars are trying to take the pulse of the nation prior to an election their target population consists of those who will go to the polls and vote, whereas those whose opinions they actually obtain constitute a sample of that population.

1.6.1.3 Sampling Unit

The constituents of a population which are the individuals to be sampled from the population and cannot be further subdivided for the purpose of sampling at a time are called sampling units. For example to know the average income per family, the head of the family is a sampling unit. To know the average marks in a paper of a class a single student is a unit.

1.6.1.4 Sampling Frame

For accepting any sampling procedure it is necessary to have a list or a map identifying each sampling unit by a number. Such a list or map is called sampling frame. For example, a list of students of a class, a list of patients of a particular disease a list of workers in a factory, a list of voters, a list of staff members of a college, a list of villages in a district, etc., are a few examples of sampling frame.

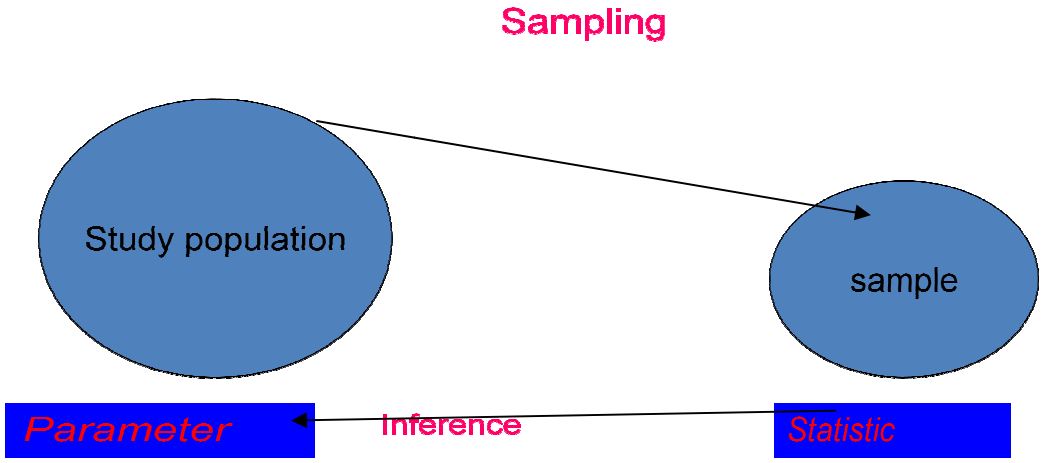


Figure- 1.1 Population v/s Sample

1.6.2 PRINCIPLES STEPS IN A SURVEY

The main steps in the planning and execution of a survey are as follows:

1. Objective of the survey. The first step is to define clearly the objective of the survey. It is generally found that even the sponsoring agency is not quite clear in its thinking as to what it wants and how it is going to use the results. The sponsors of the survey should take care that the objectives should be fulfilled with the available resources presented in the form of time, money and manpower.

2. Defining the population to be sampled. The population from which the sample is to be taken should be defined clearly. For example in sampling of farms clear-cut rules must be framed to define a farm in respect of shape, size, etc., keeping in mind the border- line cases so as to enable the investigating person to decide whether to include or not a particular farm in the population.

3. The frame and sampling unit. The sampling units must cover the entire population and they must be distinct, obvious and non-overlapping in the sense that every element of the population belongs to one and only one sampling unit. For example, in socio-economic survey for selecting people in a town, the sampling unit might be an individual person, a family, a household or a block in a locality.

In order to cover the population decided upon, there should be a list, map or some other acceptable material, called the frame, which serves as a guide to the to the population to be covered. The construction of the frame is one of a main problem since it is the frame which determines the structure of the sample survey. If the frame is not up-to-date, it should be brought up-to-date before using it.

4. Data to be collected. The data should be collected keeping in view the objective of the survey. The tendency should not be to collect too many data some of which are never subsequently examined and analyzed. A practical method is to chalk out an outline of the

tables that the survey should produce. This would help in eliminating the collection of irrelevant information and ensure that no necessary data are omitted.

5. The questionnaire or schedule. Having decided about the type of the data to be collected, the next important step of the sample selection is the construction of the questionnaire (to be filled by the respondent) or schedule of enquiry (to be filled by the interviewer) which requires skill, special technique as well as familiarity with the subject matter. The questions should be clear, brief, non- offending, polite in tone, clear-cut and to the point so that not much scope of guessing is left on the part of the respondent or interviewer. Suitable and detailed instructions for filling up the questionnaire or schedule should also be prepared.

6. Method of collecting information. The two methods commonly used to collect the sample data are:

(i) Interview method. In this method, the investigator goes from house to house and interviews the individuals personally. He asks the questions one by one and fills up the schedule on the basis of the information gained from the individuals.

(ii) Mailed questionnaire method. In this method the questionnaire is mailed to the individuals who are required to fill it up and returns it duly completed.

Whether the data should be collected by interview method or mail questionnaire method or by physical observation has to be decided keeping in view the cost, time, accuracy and money.

7. non – respondents. Due to practical difficulty the data cannot be collected for all the sampled units. For example the selected respondent may not be available at his place when the investigator goes to him or he may refuse to give certain information. This is called non-response problem. Such cases of non-response should be handled with caution in order to draw unbiased and valid conclusions.

8. Selection of proper sampling design. The size of the sample (n), the procedure of selection and the estimation of parameters along with their margins of uncertainty are some of the important problems that should be tackled carefully.

A number of designs for the selection of a sample are available and a good selection will guarantee good and reliable estimates but the relative time and money factors should also be considered for adopting any sampling design.

9. Organization of field work. It is very essential that the investigator should be trained in locating the sample units, recording the measurements, the methods of collection of required data before starting the field work. The success of a survey to a great extent depends upon the reliable field work. It is very necessary to make provisions for adequate supervisory staff for inspection after field work.

10. Pretest. From practical point of view a small pre-test has been found very useful. It always helps to decide upon effecting methods of asking questions and results in the improvement of the questionnaire. In case of large scale surveys it provides the better idea about the cost and time factor.

11. Summary and analysis of the data. The analysis of the data may be classified into the following steps:

(i) Scrutiny and editing of the data. An initial quality check should be done by the supervisory staff when the investigators are in the field. This will help in amending recording errors or in eliminating data that are inconsistent.

(ii) Tabulation of data. Before carrying out the tabulation of the data, we should decide the procedure for tabulation of the data which are incomplete due to non-response. The method of tabulation, hand or machine should be depending upon the size of the data.

(iii) Statistical analysis. A properly scrutinized, edited and tabulated data now prepared for the statistical analysis. There are different methods of estimation; therefore a suitable formula should be used for the estimation of the parameters.

(iv) Information for future surveys. Any completed survey helps in providing a note of caution for designing future surveys. The information in the form of the data means, standard deviation, the nature of the variability and the cost, time, etc., are important which are helpful for future surveys. Any completed sample survey is a lesson for future surveys in recognizing and rectifying the mistakes committed in the post survey.

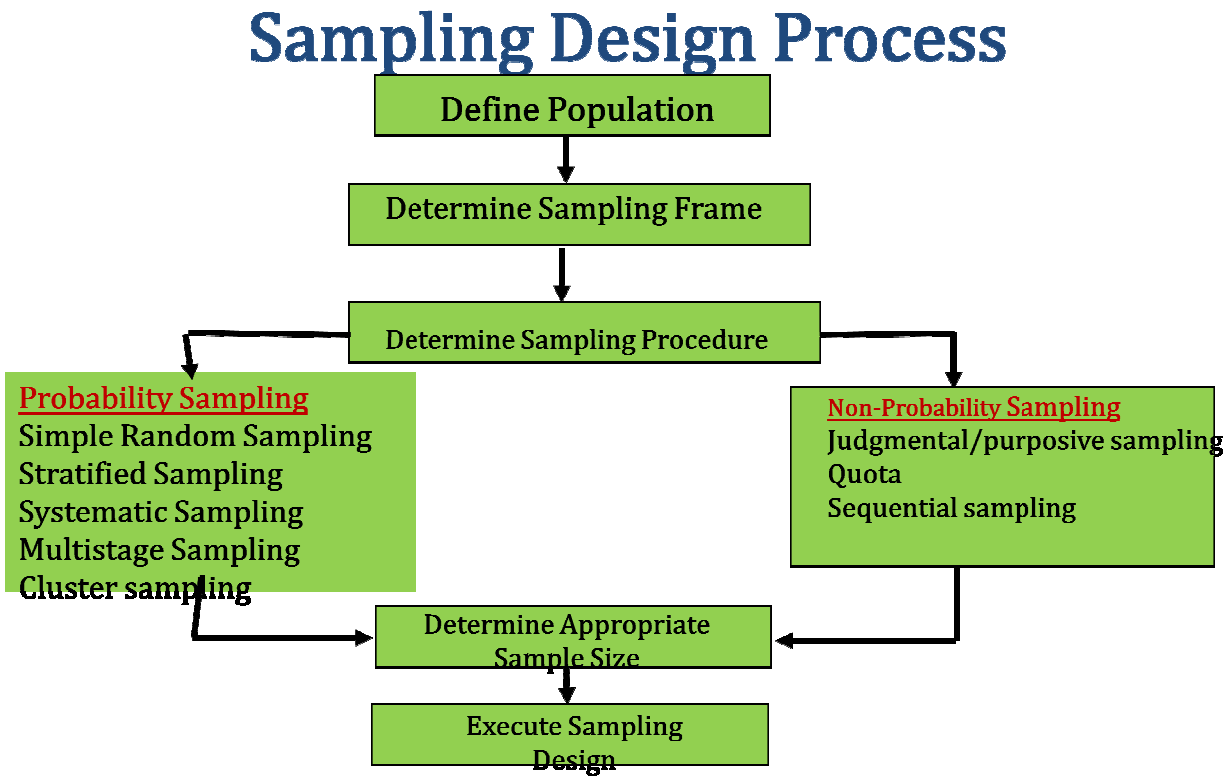


Figure-1.2 Sampling Design Process

1.6.3 SAMPLING AND NON- SAMPLING ERRORS

The errors in the collection, processing and analysis of the data are of two types:

1. Sampling error 2. Non-sampling error

1.6.3.1 Sampling errors

Sampling errors arise in the collection of a sample and the reason is because only a small part of the population is used for getting the population parameter estimates. Therefore these are absent in the complete enumeration. The main reasons of these errors are:

1. Faulty selection of the sample;
2. Substitution of the existing unit;
3. Faulty demarcation of the sampling units;

4. Constant error due to improper choice of the statistics for estimating the population parameters.

1.6.3.2 Non- Sampling errors

The errors due to the inductive process of inferring about the population on the basis of a sample, the non-sampling errors arise at the stages of observation, ascertainment and processing the data and so are present in the complete enumeration and sample survey both. The reasons of these errors are:

1. Faulty planning or definitions
2. Response errors
3. Non-response errors
4. Errors in coverage
5. Compiling errors
6. Publication errors

1.7 TYPES OF SAMPLING SCHEMES

The technique of selecting a sample is of fundamental importance in the theory of sampling and usually depends upon the nature of the data and type of enquiry. The procedure of selecting a sample may be classified into three forms:

1. Subjective or judgment sampling
2. Probability or random sampling
3. Mixed sampling

1.7.1 SUBJECTIVE OR PURPOSIVE OR JUDGMENT SAMPLING

In this scheme of sampling the sample is selected with some definite purpose in mind of the selector and so the selection of the sampling units depends completely on the decision of the selector. This sampling suffers from the drawback of favoritism depending the beliefs and prejudices of the selector and so does not provide a true representative sample of the

population. For example if the selector wants to give the picture that the standard of living has increased in Dehradun city, he may take individuals in the sample from the posh colonies of the city and ignore the colonies where low income group and middle class group live. Another example, suppose a sample of TB patients has to be drawn. Since, it is not possible to ascertain a population of TB patients, the persons turning up to TB sanitorium and having TB are selected in the sample.

This sampling method is seldom used and cannot be recommended for general use. However, if the selector is experienced and skilled and this sampling is carefully applied, then judgment sampling may provide useful results. This sampling is used in the selection of-national players of a national team and opinion surveys.

1.7.2 PROBABILITY SAMPLING

Probability sampling is the scientific method of selecting samples according to some laws of chance in which each unit of the population has some pre-assigned probability of being selected in the sample. The different types of probability sampling are:

(i) Where each unit has an equal chance of selection.

(ii) Where sampling units have different chances of selection.

(iii) Where chances of selection of unit is proportional to the sample size.

Some techniques which are commonly used in sampling are as follows:

1.7.2.1 Simple random sampling

This is the basic and most commonly used method of sampling. In this method each unit of the population has an equal chance of selection in the sample.

In this method, an equal probability is attached to each unit of the population at the first draw. It also indicates an equal probability of selection for the remaining units at the subsequent draws.

For example, to draw a simple random sample from an outdoor patient register of the department of obstetrics and gynecology, each entry would need to be numbered subsequently. If you want to draw a sample of size 700 out of 3500, a list of 700 random

numbers between 1 and 3500 would need to be prepared using one of the known procedures (described later). The 700 entries made in the register corresponding to 700 random numbers present in the prepared list would make up the required sample.

There are two ways in simple random sampling, if the unit drawn is replaced back before the next unit is drawing, the technique is called simple random sampling with replacement and if the drawing units are not replaced back and the next draws are done without selected units, the technique is called simple random sampling without replacement.

1.7.2.1.1 Selection of a simple random sample

Mainly two approaches are use to draw a simple random sample:

(a) Lottery system method
(b) Mechanical randomization or random numbers method

(a) Lottery system. This is the simplest method of selecting a random sample. The process is given below:

Suppose for a survey we want to select (n) students out of a class of (N) students. We assign the numbers 1 to N; one number to each student and we write these numbers on (N) identical chits which are same in size shape and color. These chits are put in a bag and thoroughly shuffled and then (n) chits are drawn one by one. The (n) students corresponding to the numbers on these chits will make the required sample.

This method is quite independent of the properties of the population. Generally, in place of chits, cards are used. This is one of the most reliable methods of selecting a random sample.

(b) Mechanical randomization or random numbers method. The lottery method is time consuming, if the population is large. In random numbers methods, a randomly generated numbers' table known as random number table is used to draw the required sample. There are many tables of this types prepared by many professors and scientists. These tables are so constructed that each of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 appears same number of times independently of each other. If we want to select a sample from a

population of size N ($\leq$ 99) then the numbers can be combined two by two and we will get pairs from 00 to 99. Similarly if N ($\leq$ 999) or N ($\leq$ 9999), and so on, then we combine the numbers three by three for N ($\leq$ 999) and four by four for N ($\leq$ 9999), and so on. Since each of the digits 0, 1, 2,…..,9 occurs equal number of times independently of each other, so does each of the pairs 00 to 99 or triplets 000 to 999, or quadruplets 0000 to 9999, and so on.

The steps of drawing the random sample are as follows:

(i) Identify the N units in the population with numbers from 1 to N.

(ii) Select at random any page of the random number table and pick up the numbers row wise or column wise or diagonal wise at random.

(iii) The population units corresponding to these selected numbers form the required sample.

Some commonly used random number tables are:

1. Tippet's (1927) random number table. (Tracts for computers No. 15, Cambridge University Press).
2. Fisher and Yates (1938) Tables (in Statistical tables for Biological, Agricultural and Medical Research)
3. Kendal and Babington Smith's (1939) random number tables (Tracts for computers, No. 24, Cambridge, University, Press).
4. Rand Corporation (1955) random number table. (Free Press, Illinois).

Example1. Draw a random sample (without replacement) of size 15 from a population of size 500.

Solution First of all we identify 500 units in the population with numbers from 1 to 500. Then we select at random any page of the random number table discussed above row wise or column wise or diagonally, we select on by one three digited numbers, discarding the numbers over 500, until 15 numbers below 500 are obtained. Since we are using simple random sampling technique, the numbers selected previously will also be discarded. The 15 numbers finally, so selected will constitute the required sample. The following is an extract from the first set of 40 four-digited numbers in Tippet's random number tables:

2952	6641	3992	9792	7669	5911	3170	5624
4167	9524	1545	1396	7203	5356	1300	2693
2370	7483	3408	2762	3563	1089	6913	7691
0560	5246	0112	6107	6008	8126	4233	8776
2754	9143	1405	9025	7002	6111	8816	6446

The table for the selection of 15 units of the sample is as follows:

Table -1

S. No.	Random No.	Unit selected/ not selected
1.	295	Unit selected. Since, $295 \leq 500$
2.	266	Unit selected. Since, $266 \leq 500$
3.	413	Unit selected. Since, $413 \leq 500$
4.	992	Discard random number. Since, $992 \geq 500$
5.	979	Discard random number. Since, $979 \geq 500$
6.	279	Unit selected. Since, $279 \leq 500$
7.	695	Discard random number. Since, $695 \geq 500$
8.	911	Unit not selected. Since, $911 \geq 500$
9.	317	Unit selected. Since, $317 \leq 500$
10.	056	Unit selected. Since, $056 \leq 500$
11.	244	Unit selected. Since,$244 \leq 500$
12.	167	Unit selected. Since, $167 \leq 500$
13.	952	Discard random number. Since, $952 \geq 500$
14.	415	Unit selected. Since, $415 \leq 500$
15.	451	Unit selected. Since, $451 \leq 500$

16.	396	Unit selected. Since, $396 \leq 500$
17.	720	Discard random number. Since, $720 \geq 500$
18.	353	Unit selected. Since, $353 \leq 500$
19.	561	Discard random number. Since, $561 \geq 500$
20.	300	Unit selected. Since, $300 \leq 500$
21.	269	Unit selected. Since, $269 \leq 500$

Starting with first number and moving row-wise, the units in the population with the numbers:

295, 266, 413, 279, 317, 056, 244, 167, 415, 451,

396, 353, 300, 269, will be the 15 units selected in the required sample.

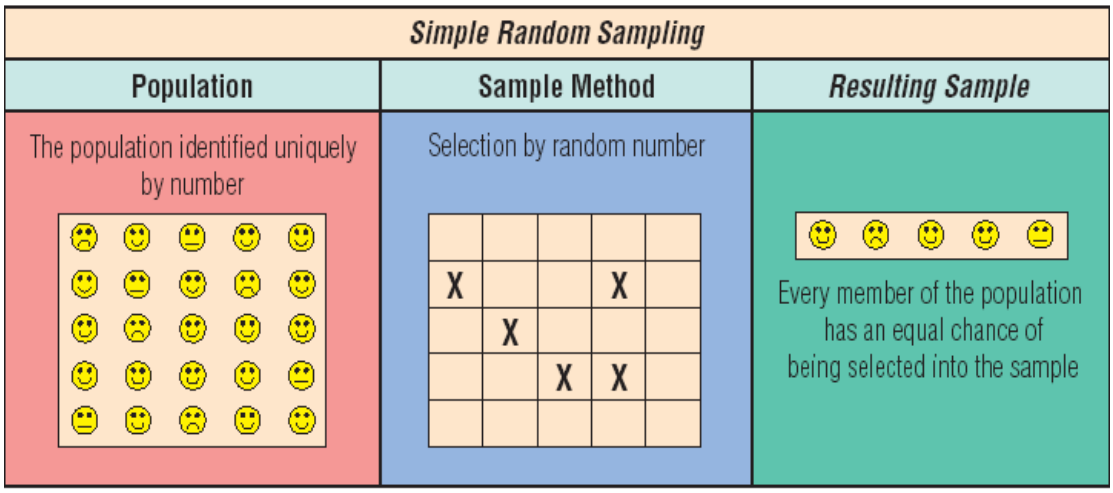


Figure-1.3Simple random sampling

1.7.2.2 Stratified random sampling

Stratified random sampling comes under the category of restricted sampling. When the population is heterogeneous with respect to some major characteristics, applying simple random sampling directly is not suitable. In such a situation, first of all the population is divided into homogeneous groups under certain criteria. These groups are termed as strata or

stratum. From each stratum independent, independent samples are selected using any of the known sampling methods. If the sample selection is carried out using simple random sampling in each stratum, the sample design is called stratified random sampling.

Information about each individual sampling unit is rarely available. Hence, the strata are formed on some broad basis such as localities in a city, districts in a state, etc. If the population is heterogeneous, stratified random sampling is more efficient. This is because a large sample is necessary to get an estimate of a characteristic with the same precision, if we ignore stratification. To be more specific, if every person in the population has the same hemoglobin level, and then a sample of even one individual would be enough to get a precise estimate of the average hemoglobin level. Let us clarify it further: in stratified random sampling, sapling units within each stratum have similar characteristic (e.g., hemoglobin levels) but different from those in other strata (e.g., disease status). In such a case, only a small sample from each stratum may provide a precise estimate of the hemoglobin level for that stratum. The estimates obtained for each stratum may be combined to get a precise estimate of hemoglobin levels for the population. A simple random sampling approach to the entire population without stratification would require comparatively large sample size than the total of stratum-specific samples to obtain an estimate of hemoglobin level with the same level of precision.

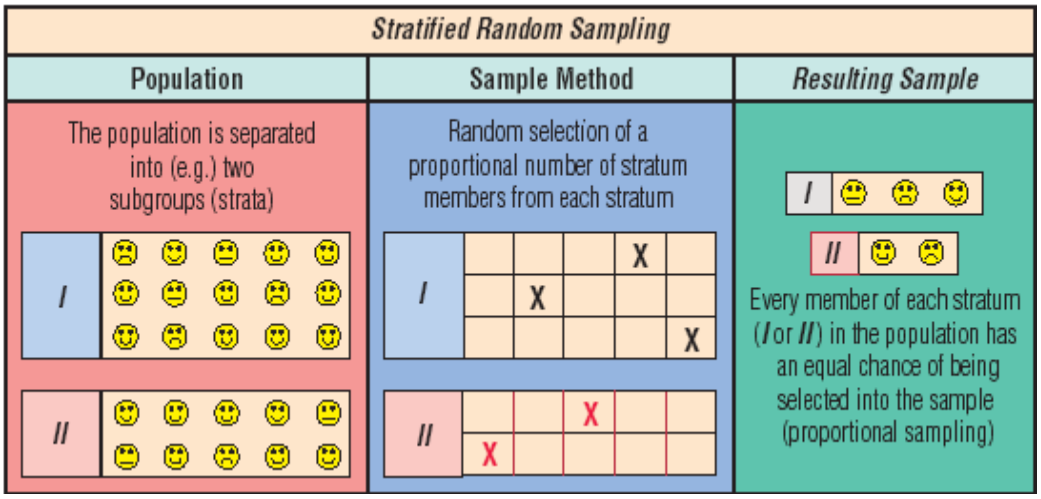


Figure-1.4Stratified random Sampling

1.7.3 MIXED SAMPLING

If the samples are selected partly according to some laws of chance and partly according to a fixed sampling rule, i.e., there is no involvement of probabilities, they are termed as mixed samples and the technique of selecting such samples is known as mixed sampling.

1.8 ORGANIZATION AND REPRESENTATION OF DATA

1.8.1 ORGANIZATION

Data collected as such do not give any meaning. This is divided according to its type explained in 1.5 and then it is consolidated by way of tabulation. Rearrangements and grouping according to requirement and standards are done, thus summarizing tables from data tabulation, which give meaning to the information collected. Data are tabulated by (1)

manual procedure (2) Mechanical procedure (3) Computer feeding. IN preparation of tables following principles are followed:

(i) A rough draft of the table should be prepared first. Before drawing out the final table, rough draft should be examined carefully.

(ii) Headings of the rows and columns should be brief and clear.

(iii) Title, note, row and column are made specific, connoting meaning or expressions.

(iv) Numbers of class intervals are decided as per aims of study which should not be too small or too big.

(v) Symbols used, should be explained.

(vi) Tabulated data should specify the units of their measurements.

(vii) The sources from which data are obtained should be given.

1.8.2 REPRESENTATION OF DATA

Tabulated data will give some information and also allow for further analysis. The columns and rows in a table make eye strain and there are chances of poor visual impression of data presented in a tabular form. Now the well tabulated data can be represented in the form of picture, diagram or figure which will help in good comparison through good visual impression. The representation of quantitative data through charts and diagrams is known as graphical representation of statistical data. A picture is said to be more effective than words for describing a particular thing or phenomenon. Main objective of diagram is to help the eye to grasp series of numbers and to grasp the meaning of series of data and also to assist the intelligence.

There are various types of graphs in the form of charts and diagrams. Some of them are:

1. Bar diagram, 2. Pie chart, 3. Histogram, 4. Frequency polygon and Frequency curve, 5. Pictograms, 6. Line chart, 7. Cumulative frequency curve 8. Scatter diagram

1.8.2.1 Bar diagram

The simplest type of graph that can be used to represent the categorical data is the bar diagram. It is also called a columnar diagram. The bar diagrams are drawn through columns of equal width. In this diagram we show the category of the variable on the X-axis and the frequencies on the Y-axis on a graph paper. A bar of each category is of the variable is drawn and the height of the bar represents the frequency of that category. Since the data is of qualitative nature or quantitative data of discrete type, bars should not be next to each other and there should be an equal gap between two successive bars. Following rules were observed while constructing a bar diagram:

(a) The width of all the bars or columns is similar.

(b) All the bars should are placed on equal intervals/distance.

The following types of bar graphs are possible:

(a) Simple bar graph

(b) Double bar graph

(c) Multiple bar graphs

We will illustrate each of these graphs by the following illustrations:

(a) **Simple Bar Diagram**

A simple bar diagram is constructed for an immediate comparison. It is advisable to arrange the given data set in an ascending or descending order and plot the data variables accordingly. In Base hospital has been found patients in OPD in particular disease as below in year 2012.

Month:	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
Patients:	285	315	250	289	386	410	452	620	421	186	450	500

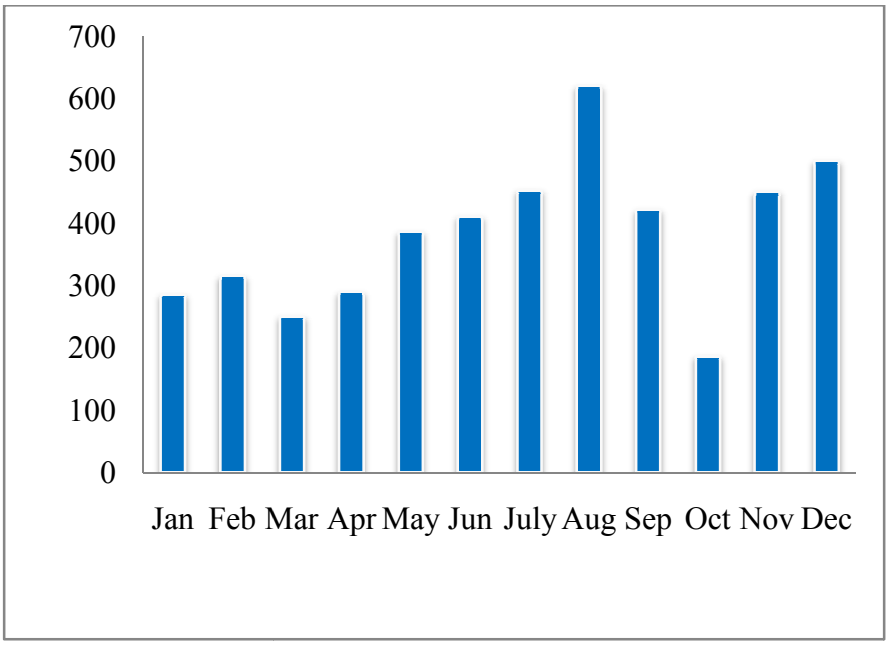


Figure -1.5 simple bar diagram

(b) Double Bar Diagram

When two components are grouped in one set of variable or different variables of one component are put together, their representation is made by a double bar diagram. In this method, different variables are shown in a single bar with different rectangles. From above example, patients were divided in two categories as male and female and the data is given below:

Month:	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Male	100	250	150	189	270	200	350	275	215	86	300	200
Female:	185	115	100	100	116	210	102	345	206	100	150	300

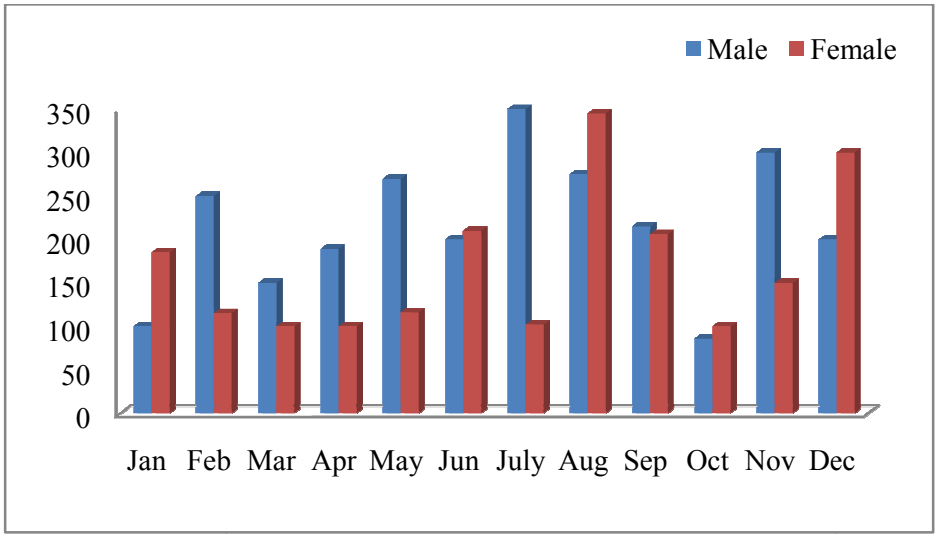


Figure -1. 6 Double Bar Diagram

(c) Multiple Bar Diagram

Multiple bar diagram shows that the proportion of subgroup between two or more categories are represented with a bar giving proportion to each of them within the bar. It is also advisable to make one bar as 100% and each subcategory is given proportion within the graph.

1.8.2.2 Pie Chart

Pie diagram is another graphical method of the representation of categorical data. Pie is a mathematical constant defined as the ratio of the circumference of a circle to the diameter and is equal to 22/7. It is drawn to depict the total value of the given attribute using a circle. In the pie chart, a circle (total 360^{o}) is divided into sectors with areas proportional to the frequencies or the relative frequencies of the categories of a variable. Dividing the circle into corresponding degrees of angle then represent the sub– sets of the data. Hence, it is also called as Divided Circle Diagram.

Example 2. A household with a monthly salary of Rs. 7200 plans his budget for a month as given below:

Item	Food	Rent	Education	Savings	Misc.	Total
Amount (Rs.)	3000	800	1200	1500	700	7200

Make a pie chart for this data.

Solution. First of all we find the angles of each sector as follows:

Total of data corresponds to 360^o. Let x^o = the angle at the centre for item A, then for the data given in above example to draw pie graph, we find the angles of each category.

Calculation of Angles

For Food:

Angle at centre = $\frac{f}{\sum f} \times 360^o = \frac{3000}{7200} \times 360^o = 150^o$. Here f= Frequency of food and $\sum f$ **= Total frequency**

For Rent:

Angle at centre = $\frac{f}{\sum f} \times 360^o = \frac{800}{7200} \times 360^o = 40^o$

Similarly, we can calculate the remaining angles, and the total of angles column should always come to 360^o.

Table-2

Item	Amount (Rs.)	Angle
Food (A)	300	150
Rent (B)	800	40

Education (C)	1200	60
Savings (D)	1500	75
Miscellaneous	700	35
Total	7200	360

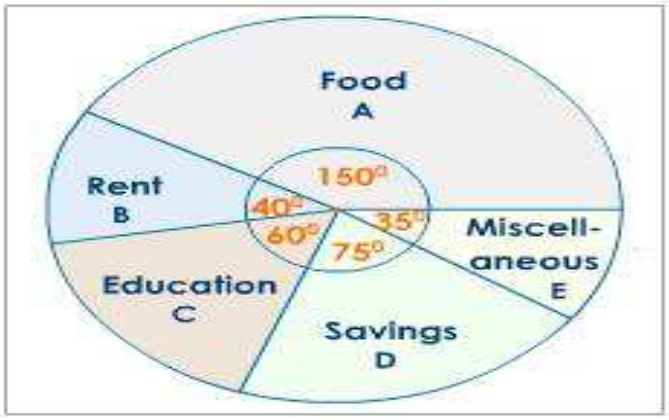


Figure-1.7Pie chart

1.8.2.3 Histogram

A two dimensional frequency density diagram is called a histogram. A histogram is a diagram which represents the class interval and frequency in the form of a rectangle. There will be as many adjoining rectangles as there are class intervals. There are two types of histograms-

(1) Histogram with equal class intervals

(2) Histogram with unequal class intervals

To draw a histogram, you should follow the steps as stated below:

1. Class intervals must be exclusive. If the intervals are in inclusive form, convert them to the exclusive form.

2. Draw rectangles with class intervals as bases and the corresponding frequencies as heights.
3. If the intervals are equal, then the height of each rectangle is proportional to the corresponding class frequency.
4. If the intervals are unequal, then the area of each rectangle is proportional to the corresponding class frequency density.

Example 3. Draw a histogram for the following data showing the class interval and their corresponding frequencies.

Class interval	0-5	5-10	10-15	15-20	20-25
Frequency	4	10	18	8	6

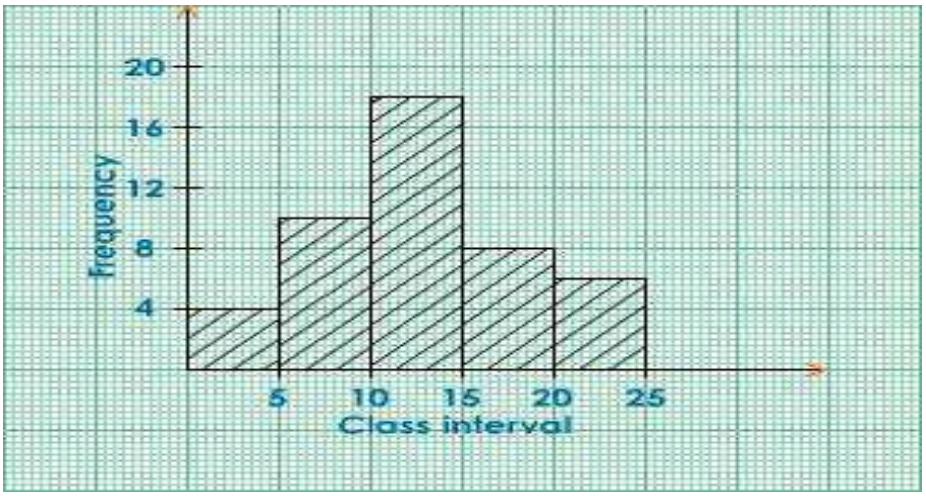


Figure-1.8 Histogram

Example 4. Following is the distribution of shops according to the number of wage - earners employed at a shopping complex.

Table-3 showing the distribution of wage earners

Number of wage earners	No. of shops	Frequency density
Under 5	18	3.6
5 – 10	27	5.4
10 – 20	24	2.4
20 – 30	20	2.0
30 – 50	16	0.8

Illustrate the above table by a histogram, showing clearly how you deal with the unequal class intervals.

Solution. When the class intervals are unequal, we construct each rectangle with the class intervals as base and frequency density as height.

Frequency density = Frequency/ Class width

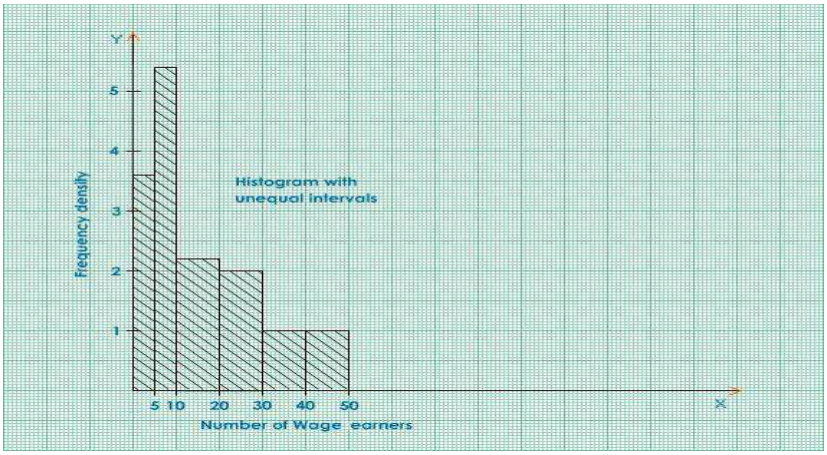


Figure- 1.9

1.8.2.4 Frequency Polygon and Frequency Curve

In a frequency distribution, the mid-value of each class is obtained. Then on the graph paper, the frequency is plotted against the corresponding mid-value. These points are joined by straight lines. These straight lines may be extended in both directions to meet the X - axis to form a polygon. If these points are joined by a free hand smooth curve then it is called Frequency curve.

Example 5. The growth rate of different crops like rice, wheat, birth rates, death rates and life expectancy are given in the following table. Make a frequency polygon from it.

Table-4 Showing class interval and frequency

Class interval	Mid Marks	Frequency
40 – 44	42	3
45 – 49	47	10
50 – 54	52	12
55 – 59	57	15
60 – 64	62	7
65 – 69	67	5

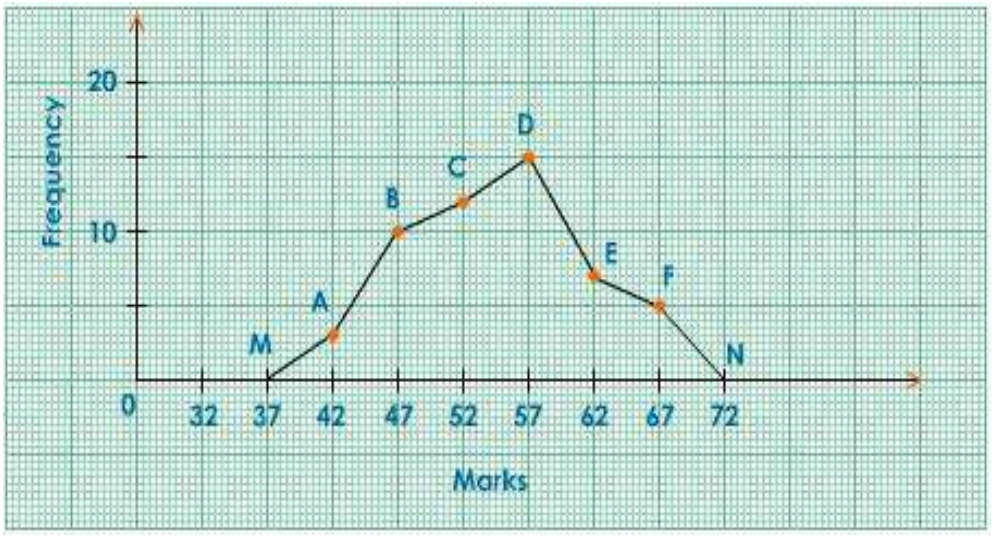


Figure-1.10

1.8.2.5 Pictograms

Pictograph is the use of pictures or images to present data. They will give the quick idea for the frequency of the characteristics and fraction also marks on pictures, e.g., bus for transport, man for cases, cot for hospital beds, etc. It is widely used by government and private organizations. The chief advantage of this method is its attraction.

1.8.2.6 Line chart

It is most widely used in medical science. It shows the trend of times. Data having some order as age –wise incidence of a disease can be represented by a line chart. It is drawn by taking one variable on the horizontal X-axis and the other variable on the vertical Y-axis. This graph shows the effect of one variable on the other variable, e.g., age specific incidence of cancer among males of Delhi.

Cumulative frequency curve

If we plot the less than cumulative frequencies rather than frequencies against the upper limits of the classes, the curve obtained on joining these points by free hand curve is called less than cumulative frequency curve or ogive or less than ogive and If we plot the more than cumulative frequencies rather than frequencies against the lower limits of the classes, the curve obtained on joining these points by free hand curve is called more than cumulative frequency curve. The advantage of this curve is that it enables us to answer the queries related to the frequency distribution of the variable.

1.8.2.8 Scatter diagram

It is the simplest way of the representation of bivariate data. Thus for the bivariate distribution (x, y) ; if the values of the Variable X and Y be plotted as x along X-axis and the y along the Y-axis respectively in the x y plane, the diagram of dots so obtained is called scatter diagram.

1.9 SUMMARY

From the study of this chapter the students came to know the definitions of statistics and biostatistics, the scope and applications of biostatistics. The students studied and learnt about data. What is data? What are the types of data? The classification of different types of data provides knowledge to treat different types of data. We learn from the study of this chapter the different steps necessary for adopting any sampling procedure and the two types of error involved in the collection of sample and complete census. We learn definitions of some terms related to sampling as population, sample, sampling frame and sampling unit. Also we learn many sampling techniques for collecting sample data. In the last of this chapter we study and learn the organization and many representation techniques of data through graphs, pictures and chart.

1.10 GLOSSARY

Data: Collected and recorded information which is either in numerical form otherwise it has no meaning.

Qualitative data: The data collected on some quality characteristic.

Quantitative data: The data collected on quantitative variable.

Variable: Any attribute, any phenomenon or any event that can have different values.

Tabulation: It is the process to [put the data into different groups or classes.

Sampling Frame: It is the list or map of population units on the basis the selection of a sample is carried out.

Srswr: Simple random sampling a technique where each unit of the population has equal chance of selection in the sample and when sampling is done with replacement, i.e., selected unit is replaced before the next unit selection is made.

Srswor: It is also simple random sampling technique but without replacement, i.e., the unit selected is not replaced back.

Stratified random sampling: It is the technique of random sample when the population has heterogeneity by dividing it into homogeneous groups.

Bar diagram: Diagrammatic representation technique of frequency data for nominal classes by bars in which the length of bars is proportional to the class frequencies.

Cumulative frequency curve: The curve in which the cumulative frequencies rather than frequencies are plotted against the class intervals. These are of two types less than and more than types and their intersection represent median.

Histogram: It is a diagrammatic representation of the frequency distribution of a quantitative data with areas of the rectangles proportional to the class frequency.

1.11 SELF ASSESMENT QUESTIONS

Question 1. For the given data draw a bar chart.

Year	1978	1979	1980	1981	1982	1983
Rice (in tons)	4500	5700	6100	6500	4300	7800

Question 2. For the data give below make a pie chart.

Blood group	A	B	AB	O	Total
Frequency	15	25	20	30	90

Question 3. Make a histogram and frequency polygon for the data given below.

Profit per shop	0-100	100-200	200-300	300-400	400-500	500-600
No. of shops	12	18	27	20	17	6

Question 4. What are the different types of data? Explain in brief. Also explain the different sampling schemes of data collection.

Question 5. The data collected directly by the investigator or his/her team is called

(a) Secondary data (b)Primary data

(c) Population data (d) Sample data

Question 6. The data collected from already published material is called

(a) Secondary data (b)Primary data

(c) Population data (d) Sample data

Question 7. In bar diagram the base line is:

(a) Horizontal (b) Vertical

(c) False base line (d) any of the above

Question 8. Less than and more than ogives intersects at;

(a) mean (b) median

(c) mode (d) origin

Question 9. In cases of frequency distribution with classes of unequal width, the heights of bars in a histogram are proportional to:

(a) class frequency (b) class intervals

(c) frequencies in percentage (d) frequency densities

Question 10. Sampling frame is a term used for:

(a) a list of random numbers (b) a list of voters

(c) a list of sampling units of the population (d) none of the above

Question 11. A selection procedure of a sample having no involvement of probability is:

(a) Purposive sampling (b) Judgment sampling

(c) Subjective sampling (d) all the above

Answers: 5- (b), 6- (a), 7- (b), 8- (b), 9- (d), 10- (c), 11- (d).

1.12 REFERENCES

Agarwal, B.L., Programmed Statistics, New Age International (P) Ltd, New Delhi, (2003).

Arora, P.N. and Malhan, P.K., Biostatistics, Himalaya Publishing House (P) Ltd. Mumbai, (1996).

Belle, G., Fisher, L., Heagerty, P.G. and Lumly, T.,Biostatistics- A Methodology for the Health Science, John Willey & Sons, inc. Pub., Hoboken, New Jersey, (2004).

Gupta, S.C. and Kapoor, V.K., Fundamentals of Mathematical Statistics, Sutan Chand & sons, Daryaganj, New Delhi, (1970).

Gupta, S.C. and Kapoor, V.K., Fundamentals of Applied Statistics, Sutan Chand & sons, Daryaganj, New Delhi, (1976).

Pagaro, M. and Gauvreau, K., Principles of Biostatistics, Duxbury Thomson Learning, U.S.A. (2000).

Prabhakara, G.N., Biostatistics, Jaypee Brothers Medical Publishers (P) Ltd, New Delhi, (2006).

Singh, D. and Chaudhary, F.S., Sample Survey Designs, New Age International (P) Ltd, New Delhi, (1986).

Sudaram, K.R., Dwivedi, S.N., Sreenivas, V., Medical Statistics, Wolters Kluwer (P) Ltd, New Delhi. (2010).

1.13 TERMINAL QUESTIONS

1. What is sampling frame?

2. What do you know about histogram?

3. Explain stratified random sampling

4. What is the difference between simple random sampling with and without replacement?

5. What are different types of graphs or charts for data representation? Explain any two of them.

6. Explain the difference between sample and population.

7. What is the need of biostatistics?

Unit-2– MEASURES OF CENTRAL TENDENCY

CONTENT

2.1 OBJECTIVES

From the study of this chapter the students will be able:

1. To know about the measures of central tendency- mean, median and mode.
2. To know the merits and demerits and uses of these measures.
3. To know about different methods of measuring mean, median and mode.
4. To know the situations where which measure is better to use?
5. To know the advantages of short cut methods of computing mean.

2.2 INTRODUCTION

In the previous chapter, we discussed data collection, data organization and data representation techniques. The data representation techniques such as frequency histograms and frequency polygons, introduced the concept of the shape of distributions of data. For example, a frequency polygon illustrated the distribution of body mass index data. We expend chapter 1 on these concepts by defining measures of central tendency.

Measures of central tendency as the name suggests are numerical measurement of the central part of the distribution. Measures of central tendency are also called averages or measures of location because they show the location of the centre of the distribution from which the data were sampled. According to Professor Bowley, averages are, "statistical constants which enable us to comprehend in a single effort the significance of the whole." In other words, these are numbers that tell us where the majority of values in the distribution are located. For example the average marks in a distribution of marks of all the students of a class. The averages which are commonly used in biostatistics are as follows:

1. Mean or arithmetic mean 2. Median 3. Mode.

2.3 MEAN

Mean or arithmetic mean of a series of data is the ratio of the sum of the observations to the number of observations. If $x_1, x_2, ...x_n$ are the observations of a series then their arithmetic mean is given by

$$\bar{x} = \frac{x_1 + x_2 + ...x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (1)$$

And if the corresponding frequencies, $f_1, f_2, ...f_n$ of the variables $x_1, x_2, ...x_n$ are given, then the arithmetic mean is defined as ratio in which the numerator is the sum of products of the variables with their frequencies and denominator is the sum of the frequencies.

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + ...f_nx_n}{\sum f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{N} \qquad (2)$$

where, N = $\sum f_i$ = sum of frequencies.

2.3.1 MEAN OF INDIVIDUAL ITEMS

Mean of individual items is given by the ratio of the sum of items to the number of items as given in formula (1).

Example 1. Find the arithmetic mean of triglycerides present 10 patients in their blood samples in a hospitalas:

25, 30, 21, 55, 47, 10, 15, 17, 45, 35

Solution. Let $\bar{x}$ be the average triglyceride value and since these are individual items, their mean can be computed by formula

$$\bar{x} = \frac{x_1 + x_2 + ...x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$
$$= \frac{25+30+21+55+47+10+15+17+45+35}{10} = \frac{300}{10} = 30$$

2.3.2 MEAN IN DISCRETE FREQUENCY DISTRIBUTION

If $x_1, x_2, ...x_n$ are the observations in a discrete distribution according to some characteristic and $f_1, f_2, ...f_n$ be their corresponding frequencies then the arithmetic mean is given by the formula (2). The computation procedure for mean can be easily understood with the help of the example given below.

Example 2. The distribution of marks of 50 students of B.Sc. class in a botany semester examination is given below. Find the average of marks.

Marks (x)	12	23	25	35	45	15	40
Frequency(f)	3	10	12	10	2	8	5

Solution. Since this is a discrete distribution so the average of marks is given by the formula (2). For the computation of average marks we prepare the following table:

Table1. For calculation of Mean in Discrete Distribution

Marks (x)	Frequency (f)	f x
12	3	36
23	10	230
25	12	300
35	10	350
45	2	90

15 40	8 5	120 200
Total	$\sum f = 50$	$\sum fx = 1326$

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + ...f_nx_n}{\sum f_i} = \frac{\sum_{i=1}^{n} f_ix_i}{N} = \frac{1326}{50} = 26.52$$

This is clear that it is not necessary that average will be a number presenting in the data and also it is not an integer value while the marks in integers.

2.3.3 MEAN IN CONTINUOUS DISTRIBUTION

In case of continuous distribution, there are given class intervals and their corresponding frequencies. First of all we find the mid values of these classes and treat them as the variable values. Now we apply the formula (2) for the calculation of arithmetic mean. The procedure will be clear from the following example.

Example 3. For the data given in the below table on systolic BP of 68 patients, calculate the arithmetic mean.

Table 2.

Systolic BP (mmHg)	Frequency (f)	Systolic BP (mmHg)	Frequency (f)
90-100	3	140-150	11
100-110	5	150-160	9
110-120	7	160-170	6
120-130	10	170-180	2
130-140	15		

Solution. For the calculation of mean we prepare the following table:

Table3. For calculation of Mean in Continuous Distribution

Systolic BP (mmHg)	Frequency (f)	Mid Value (x)	Fx
90-100	3	95	285
100-110	5	105	525
110-120	7	115	805
120-130	10	125	1250
130-140	15	135	2025
140-150	11	145	1595
150-160	9	155	1395
160-170	6	165	990
170-180	2	175	350
Total	$\sum f = 68$		$\sum fx = 9220$

$$\bar{x} == \frac{\sum_{i=1}^{n} f_i x_i}{N} = \frac{9220}{68} = 135.6 mmHg$$

2.3.4 SHORT-CUT METHOD FOR MEAN

For the computation of mean short –cut method is applied when the variable values and their frequencies are large. To make the computations easy we take a middle value in the given values of x as assumed mean and subtract this assumed mean from all the values of x. This assumed mean is also called provisional mean. Then the formula for the calculation of arithmetic mean is given by as follows:

$$\bar{x} = A + \frac{\sum d}{n} \qquad (3)$$

Where A= assumed mean, n= number of observations in the given data

d= x- A= deviation of all the variate values from assumed mean A.

Steps of computation for short- cut method:

Step 1. Take any observation (generally, middle value if we arrange the values in ascending or descending order of magnitude) of the individual series as assumed mean A.

Step 2. Find the deviation of the values of variate x from assumed mean A, i.e., calculate the differences d= x- A

Step 3. Find the sum of d and use above formula (3), we find the value of mean.

If the frequencies corresponding to the variate values are given, then we use the formula for mean as follows:

$$\bar{x} = A + \frac{\sum fd}{N} \qquad (4)$$

Where, N = $\sum f$ = sum of frequencies. Here we find the product of f and d.

If the data is continuous, we find the mid values as x and then d= x- A. Now apply the above formula (4). The procedure will be clear from the examples as give below.

Example 4. The marks of the 7 students of a class in a test are as given below:

12, 15, 22, 25, 35, 40, 45

Find the mean by short-cut method.

Solution. Let us take assumed mean A=25. Now we prepare the table for the computation of mean as given below:

Table-4 Mean for individual data by short cut method

X	d = x- 25
12	-13
15	-10
22	-3
25	0
35	10
40	15
45	20
Total	$\sum d = 19$

Arithmetic mean $\bar{x} = A + \frac{\sum d}{n} = 25 + \frac{19}{7} = 25 + 2.71 = 27.71$

Thus the average of marks of the given 7 students of the class is 27.71

Example 5. Ten patients were examined for uric acid test. The operation was performed 1050 times and the frequencies so obtained for different number of patients (x) are shown in the table given below. Compute the arithmetic mean by short- cut method.

x:	0	1	2	3	4	5	6	7	8	9	10
f:	2	8	43	133	207	260	213	120	54	9	1

Solution. Let 5 be the assumed mean. Now we prepare the table for the calculation of mean.

Table-5 Mean for discrete grouped data by short cut method

X	Frequency (f)	d = x- 5	fd
0	2	-5	-10
1	8	-4	-32
2	43	-3	-129
3	133	-2	-266
4	207	-1	--207
5	260	0	0
6	213	1	213
7	120	2	240
8	54	3	162
9	9	4	36
10	1	5	5
Total	$\sum f = 1050$		$\sum fd = 12$

Arithmetic mean $\bar{x} = A + \frac{\sum fd}{N} = 5 + \frac{12}{1050} = 5 + 0.0114 = 5.0114cm$

Thus the average for uric acid is 5.0114.

2.3.5 STEP DEVIATION METHOD OF MEAN

It can be used in grouped data. When all the classes are of equal width (say h), in continuous data and the values of x are at equal interval in discrete grouped data then the we may simplify the calculations by taking d= (x- A)/ h in short-cut method. Now the formula for the calculation of mean becomes.

$$\bar{x} = A + \frac{\sum fd}{N} \times h$$

Here, the symbols have the same meaning as in short-cut method above and h is the gap between the two values of x or class interval.

Example 6. Find the mean by step deviation method for the data of blood pressure of 68 patients as given in the following table.

BP(mmHg) (x)	90	100	110	120	130	140	150	160	170
Frequency (f)	3	5	7	10	15	11	9	6	2

Solution. We take assumed mean A= 130 and here interval between any two values of x is 10, i.e., h= 10. Now prepare the table for the computation of mean.

Table6. Step Deviation Method of Mean in discrete grouped data

BP (mmHg)	Frequency (f)	$d = \frac{x-130}{10}$	fd
90	3	-4	-12
100	5	-3	-15
110	7	-2	-14
120	10	-1	-10
130	15	0	0
140	11	1	11
150	9	2	18
160	6	3	18
170	2	4	8

Total	$\sum f = 68$		$\sum fd = 4$

Arithmetic mean $\bar{x} = A + \frac{\sum fd}{N} \times h = 130 + \frac{4}{68} \times 10 = 130 + 0.588 = 130.588 mmHg$

Thus the average for BP is 130.588mmHg.

Example7. For example 3, calculate arithmetic mean by step deviation method.

Solution. For the calculation of mean the table is given below:

Table-7. Mean by Step Deviation Method in Continuous Data

Systolic BP (mmHg)	Frequency (f)	Mid Value (x)	$d = \frac{x-135}{10}$	fd
90-100	3	95	-4	-12
100-110	5	105	-3	-15
110-120	7	115	-2	-14
120-130	10	125	-1	-10
130-140	15	135	0	0
140-150	11	145	1	11
150-160	9	155	2	18
160-170	6	165	3	18
170-180	2	175	4	8
Total	$\sum f = 68$			$\sum fd = 4$

Arithmetic mean $\bar{x} = A + \frac{\sum fd}{N} \times h = 135 + \frac{4}{68} \times 10 = 135 + 0.588 = 135.588 mmHg$

Thus the average for uric acid is 135.588mmHg.

Example 8. In a study on patients of typhoid fever the following data are obtained. Find the arithmetic mean.

Age in years	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89
No. of cases	1	0	1	10	17	38	9	3

Solution. This is inclusive type data; first of all we convert it to exclusive type data. The procedure for converting inclusive type data to exclusive type data is as follows:

We see that the upper limit of the first class is 19 and the lower limit of the second class is 20 and their difference is 20-19=1. Now subtract half of the difference, i.e., 0.5 from the upper limit and 0.5 to the lower limit. Also we see that this difference is the same for each of the class. So the new classes are as 9.5-19.5, 19.5-29.5 and so on.

Now for the calculation of mean any method discussed above can be used. Here we apply step deviation method.

Table-8. Mean by Step Deviation Method in Inclusive Data

Age	Frequency (f)	Mid Value (x)	$d = \frac{x - 44.5}{10}$	fd
9.5-19.5	1	14.5	-3	-3
19.5-29.5	0	24.5	-2	0
29.5-39.5	1	34.5	-1	-1
39.5-49.5	10	44.5	0	0
49.5-59.5	17	54.5	1	17
59.5-69.5	38	64.5	2	76

69.5-79.5 79.5-89.5	9 3	74.5 84.5	3 4	27 12
Total	$\sum f = 79$			$\sum fd = 128$

Arithmetic mean $\bar{x} = A + \frac{\sum fd}{N} \times h = 44.5 + \frac{128}{79} \times 10 = 44.5 + 16.2 = 60.7$

2.3.6 WEIGHTED MEAN

In computation of arithmetic mean some items are more important than the others, in such cases the weight age should be given to the items according to their importance. For example if we want to have an idea of the change in cost of living of group of people of a certain locality, then the simple mean of the prices of the commodities consumed by them will not do, since all the commodities are not equally important, e.g., wheat, rice and pulses are more important than cigarettes, tea, confectionery, etc.

If $x_1, x_2, ...x_n$ are the variate values of a distribution and $w_1, w_2, ...w_n$ be their corresponding weights then weighted mean is give by:

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$$

Example 9. The following table gives the platelets count (in lakh/cmm) from the analysis of the blood samples on five different days in a pathology laboratory. Find the average platelets count per patient.

Day	1	2	3	4	5
Platelates count (in lakh/cmm) (w)	0.50	0.75	1.00	1.50	2.00
No. of patients (x)	65	80	95	90	70

Solution. The table for the calculation of weighted mean is given by:

Table 9. Table for Weighted Mean

Platelets count (x)	No. of patients (w)	Wx
0.50	65	32.5
0.75	80	60.0
1.00	95	95.0
1.50	90	135
2.00	70	140
Total	$\sum w = 400$	$\sum wx = 462.5$

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} = \frac{462.5}{400} = 1.156$$

Thus, the average platelets per patient are 1.156 lakh/cmm.

2.3.7 COMBINED MEAN

If $\bar{x}_1, \bar{x}_2, ...\bar{x}_m$ are the means of m series of sizes $n_1, n_2, ...n_m$ respectively, then their combined arithmetic meaning $\bar{x}$ is given by:

$$\bar{x} = \frac{\sum n_i \bar{x}_i}{\sum n_i}; i = 1,2,.....m$$

Example 10. There are 40 male and 10 female employees in a firm. The mean salary of male employees is Rs.520 and that of female employees Rs. 420. Find the combined average salary of all the employees.

Solution. Here, $n_1 = 40, n_2 = 10, \bar{x}_1 = 520, \bar{x}_2 = 420$

$$\text{Combined mean } \bar{x} = \frac{\sum n_i \bar{x}_i}{\sum n_i} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{520 \times 40 + 420 \times 10}{40 + 10} = \frac{25000}{50} = 500$$

Hence, the average salary of all the employees is Rs. 500.

2.3.8 CORRECTED MEAN

Some times there are problems of such type that we used wrong digits while the actual digits were different, then we replace the wrong digits with the correct digits and now we can get the correct mean. The procedure will be clear from the example given below.

Example 11. A student calculates the mean of 20 observations as 25.2. Later on he found that he misread one observation 34 in place of 43, find the correct mean.

Solution. We know the mean of individual series is given by:

$$\bar{x} = \frac{\sum x}{n} or \sum x = n\bar{x} = 20 \times 25.2 = 504$$

But he misread 43 as 34. So the correct total of x= 504-34+43=513.

So correct mean=513/ 20 =25.65

2.3.9 MERITS, DEMERIRS AND USES OF MEAN

Merits:

1. Mean is rigidly defined.
2. It can be calculated easily by a non mathematical person also.
3. It is based upon all the observations.
4. Among all the averages, it is affected least by fluctuations of sampling.
5. It is the best measure to compare two or more series.
6. It is easily understandable.
7. It is the most widely used method of central tendency.

Demerits:

1. It is affected much by extreme values.

2. It cannot be calculated in case of open end classes.

3. It cannot be calculated in case of qualitative data such as intelligence, beauty, etc.

`4. In extremely asymmetrical distribution, mean is not a suitable measure of central tendency.

5. It cannot be calculated if any observation is missing.

6. It may lead to wrong conclusions if the details of the data are not given. For example the marks of two students in three successive tests are respectively 30, 40, 50 and 50, 40, 30. We see that average score of both the students is same, we can say that both students are of same level while first is improving and the second is deteriorating.

Uses of Mean:

1. It is very much used in practical situations.

2. A common man uses it for computing his monthly budget.

3. It is very much used in sampling and inference.

4. A businessman uses it for computing per unit profit, output per person, average expenditure and average profit per week or per month, etc.

2.4 MEDIAN

Median of a distribution is the middle most value of the variable if the values of the variable are arranged in ascending or descending order of their magnitude. The median divides the observations of the variable in such a way that half of the observations of the variable lie above the median and half below this. Median is thus called a positional average because it locates at the middle of the observations. But if the number of observations is even then after arrangement there will be two middle values and the median will be the average of these two middle values.

2.4.1 MEDIAN IN INDIVIDUAL SERIES

Arrange the data observations, (say n) in ascending or descending order of magnitude. Now there can be two cases:

Case 1. If n is odd then middle most, i.e., (n+1/2)th term value is the median.

Case2. If n is even then there are two middle terms (n/2)th and (n+1/2)th, then median is given by:

$$M_e = \frac{\left[\frac{n}{2}th + \left(\frac{n}{2}+1\right)th\right]term}{2}$$

Example 12. The marks of 9 children in a test exam are: 12, 23, 34, 11, 14, 15, 13, 16, 45.

Find the median of the marks.

Solution. Arrange the given observations in ascending order of magnitude, we get

`11, 12, 13, 14, 15, 16, 23, 34, 45

Here the number of observations n =9, i.e., odd.

So the median is the (9+1)/2 th, i.e., 5^{th} term value, i.e., 15.

Example 13. The number of blood LDL (in mg/dl) present the blood samples of 12 patients are: 5, 19, 42, 11, 50, 30, 21, 0, 22, 52, 36, 27

Find the median of the data.

Solution. On arranging the given observations in ascending order of magnitude, we get,

0, 5, 11, 19, 21, 22, 27, 30, 36, 42, 50, 52

Here number of observations = 12, i.e., even. So median is given by

$$M_e = \frac{\left[\frac{n}{2}th + \left(\frac{n}{2}+1\right)th\right]term}{2} = \frac{\left[\frac{12}{2}th + \left(\frac{12}{2}+1\right)th\right]term}{2}$$
$$= \frac{(6th+7th)term}{2} = \frac{22+27}{2} = \frac{49}{2} = 24.5$$

So median is 24.5 mg/dl which does not belong to the data. So in case of even number of observations median is not present in the data observations.

2.4.2 MEDIAN IN DISCRETE FREQUENCY DISTRIBUTION

If $x_1, x_2, ...x_n$ are the observations in a discrete distribution according to some characteristic and $f_1, f_2, ...f_n$ be their corresponding frequencies then for the calculation of median we calculate the cumulative frequencies. The median is calculated with the help of the following steps.

Working steps for median

Step1. Arrange the given values in ascending order of magnitude.

Step2. Find the total of frequencies, called cumulative frequency and denoted by c.f.

Step3. Find $\frac{N}{2}$, where N= $\sum f$.

Step4. Find cumulative frequency just greater than $\frac{N}{2}$. The value of x corresponding to this cumulative frequency is the required median.

Example14. Find the median for the following data.

X	21	15	17	9	5	7	8	10
F	2	5	3	4	5	1	6	12

Solution. For calculating the median we arrange the values of x in ascending order and then prepare the cumulative frequency table as follows:

Table10. Median in Discrete Distribution

x	f	c.f.
5	5	5
7	1	6
8	6	12
9	4	16
10	12	28
15	5	33
17	3	36
21	2	38
	$N = \sum f = 38$	

Here N/2 = 38/2= 19 and cumulative frequency just greater than 19 is 28. The value of x corresponding to cumulative frequency 28 is 10. So the median of the given data is 10.

2.4.3 MEDIAN IN CONTINUOUS FREQUENCY DISTRIBUTION

When the data is in class interval form, the class corresponding to c.f. just greater than N/2 is called the median class and the median is computed by the following formula:

$$M_e = L + \frac{\left(\frac{N}{2} - C\right)}{f} \times h$$

Where L= lower limit of the median class

C= cumulative frequency just before the median class

N= Total of frequency

f = frequency of the median class

h = magnitude of the median class

Example15. The following table gives the distribution of weights of 100 persons. Find the median of this data.

Weight	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90
Frequency	1	3	6	10	15	25	15	10	11	4

Solution. For computing the median we prepare the following table:

Table11. Median in Continuous Distribution

Weight (in kg) (x)	Frequency (f)	Cumulative Frequency (c.f.)
40-45	1	1
45-50	3	4
50-55	6	10
55-60	10	20
60-65	15	35
65-70	25	60
70-75	15	75
75-80	10	85
80-85	11	96

85-90	4	100
Total	$N = \sum f = 100$	

Here N/2 = 10/2= 50 and cumulative frequency just greater than 50 is 60. The class corresponding to cumulative frequency 60 is 65-70. So class 65-70 is the median class. Now median is given by:

$$M_e = L + \frac{\left(\frac{N}{2} - C\right)}{f} \times h$$

here L= lower limit of the median class= 65

C= cumulative frequency just before the median class=35

N= Total of frequency =100

f = frequency of the median class=25

h = magnitude of the median class=5

$$M_e = 65 + \frac{(50-35)}{25} \times 5 = 65 + \frac{15}{25} \times 5 = 65 + 3 = 68$$

So median of the given data is 68kg

2.4.4 MERITS, DEMERIRS AND USES OF MEDIAN

Merits:

1. Median is rigidly defined.
2. It is not affected at all from extreme values.
3. It is easy to understand and to calculate.
4. In case of individual series data it can be located merely by inspection.

5. It can be calculated in case of open end classes.

6. Its graphical representation is also possible.

7. It can be computed even if the classes are of unequal interval.

8. In case of qualitative data, e.g., beauty, honesty, intelligence, etc. it is the best measure of central tendency.

Demerit:

1. It is not amenable to algebraic treatment.

2. It is a positional average and is based only on the middle term. It does not use all the observations of the data.

3. In case of irregular distribution, it is not a good measure.

4. In case of even number of observations it cannot be determined exactly, it can be estimated only by the average of the two middle terms.

5. In comparison to mean it is affected much by fluctuations of sampling.

Uses:

1. It is a good measure if numerical measurements are not possible.

2. In case of qualitative data where the observations cannot be determined quantitatively, it is the only average.

3. It is generally used in studying the average intelligence or average honesty of a group of people.

2.5 MODE

Mode is the most frequent item of the series, i.e., in a given set of observations a item or observation which is repeated maximum number of times an all other observations cluster around this, is called mode. For example, the average height of an Indian male is 5 feet 6

inch; the average size of the shoes of an Indian male is number 7, etc. Mode is also known as norm.

2.5.1 TYPES OF MODE OF A DISTRIBUTION

Unimodal: If the data of a distribution has only one mode then the distribution is called unimodal.

Bimodal: If we find that there are two items in a distribution which have the same number of repetitions, then these two items are the modes and the distribution is called bimodal.

Trimodal: Similarly, in a distribution, if there are three such items that they have the same frequency then these three items are called the modes of the distribution and the distribution is called trimodal.

Ill- defined mode: If there exists more than one mode in a distribution, then mode is called ill-defined.

2.5.2 MODE IN IDDIVIDUAL SERIES

In case individual series mode is the most frequent observation. It is clear from the following example.

Example 16. Find the mode of the series given below:

2, 3, 4, 7, 9, 3, 2, 1, 5, 3, 6, 3, 8, 3

Solution. In the given series the observation 3 is repeated maximum number of times (5) so the mode of the given series is 3.

2.5.3 MODE IN DISCRETE FREQUENCY DISTRIBUTION

In case of discrete frequency distribution, mode is the value of the variable which has the maximum frequency. Consider the following example:

Example 17. Find the mode of the following frequency distribution:

Variable (x)	2	5	7	9	11	25	35	43	52
Frequency (f)	1	3	4	8	25	12	11	10	8

Solution: Here we see that in the given distribution, the variable 1 has the maximum frequency 25. So the mode of this distribution is 11.

2.5.4 GROUPING METHOD OF MODE

When the distribution is irregular, the frequencies are increasing and decreasing in

An irregular pattern or the difference between the maximum frequency and the frequency succeeding or proceeding to it is small and the observations are concentrated on either side, in such a situation mode cannot be determined merely by inspection. In such a case, we apply the grouping method for the computation of mode. The procedure of grouping method will be clear from the following example.

Example 18. Find the mode of the following distribution.

Variable (x)	2	3	4	5	6	7	8	9	10	11	12	13
Frequency (f)	1	3	4	5	7	10	11	10	9	14	7	5

Solution. Here we see that initially the frequencies are increasing from 1 to 11 and then decreasing but the frequency 14 of the variable value 11 is again increasing and then decreasing up to frequency 5. This distribution shows an irregular pattern. So for the calculation of mode we apply the grouping method of mode. For this we prepare a table and the procedure of preparing the table is explained below the table.

Table12. Table for grouping the frequencies

Variable (x)	Frequency(f)					
	Column (i)	Column (ii)	Column (iii)	Column (iv)	Column (v)	Column (vi)
2	1	} 4				
3	3		} 7	} 8		
4	4	} 9			} 12	
5	5		} 12			} 16
6	7	} 17		} 22		
7	10		} **21**		} 28	
8	11	} 21				} **31**
9	10		} 19	} **30**		
10	9	} **23**			} **33**	
11	**14**		} **21**			} 30
12	7	} 12		} 26		
13	5					

Prepare a table from the frequencies of the distribution. In column (i), we have the original frequencies. Mark bold type the maximum frequency in this column. Column (ii) is prepared by adding the frequencies two by two as 1+3 = 4; 4+5 = 9 and so on. Mark bold type the maximum frequency in this column also. Column (iii) is prepared by adding the

frequencies two by two leaving the first frequency. Column (iv) is prepared by adding the frequencies three by three. Column (v) is prepared by adding the frequencies three by three leaving the first frequency and column (VI) is prepared by adding the frequency three by three leaving the first two frequencies. In each column make bold type the maximum frequency. The table is given above:

Now to find the mode we prepare the following analysis table:

Table13. ANALYSIS TABLE

Column number (1)	Maximum frequency (2)	Value(s) of x related to the maximum frequency (3)
i	14	11
ii	23	10, 11
iii	21, 21	7, 8, 11, 12
iv	30	8, 9, 10
v	33	9, 10, 11
vi	31	7, 8, 9

In the analysis table column number (1) shows the columns serially from the above table 12, column number (2) shows the maximum frequency from the same table 12 and column number (3) shows the value of x related to the maximum frequency or the values of x which contributes in the maximum frequency. Finally, in column number (3) of the analysis table we see that the value 11 is repeated maximum number of times. So 11 is the mode of the above distribution.

2.5.5 MODE IN CONTINUOUS FREQUENCY DISTRIBUTION

In case of grouped continuous frequency distribution the maximum frequency shows that the related class is the modal class and for the computation of mode we use the following formula:

$$M_o = L + \frac{(f_1 - f_0)}{(2f_1 - f_0 - f_2)} \times h$$

Where L= lower limit of the modal class

h= magnitude of the modal class

f_1 = frequency of the modal class

f_2 = frequency of the class succeeding the modal class

f_0 = frequency of the class preceding the modal class

For a moderately asymmetrical distribution the mode can be calculated by a formula given by Karl Pearson as follows:

Mode = 3 Median – 2 Mean

Example 19. Following table shows the blood pressure and the frequency related to it. Find the mode of this distribution.

Table 14.

C.I.	Frequency	C.I.	frequency
70-80	2	110-120	32
80-90	4	120-130	28
90-100	14	130-140	12
100-110	35	140-150	5

Solution. From the table it is clear that maximum frequency is 35 and the related class is the 100-110. So 100-110 is the modal class. Now to compute the mode we use the following formula:

$$M_o = L + \frac{(f_1 - f_0)}{(2f_1 - f_0 - f_2)} \times h$$

Here L= lower limit of the modal class= 100

h= magnitude of the modal class= 10

f_1 = frequency of the modal class= 35

f_2 = frequency of the class succeeding the modal class= 32

f_0 = frequency of the class preceding the modal class= 14

$$\therefore \quad M_o = 100 + \frac{(35-14)}{(70-14-32)} \times 10 = 100 + \frac{210}{24} = 100 + 8.75 = 108.75$$

So mode of the given distribution is 108.75.

2.5.6 MERITS, DEMERITS AND USES OF MODE

Merits:

1. Mode is easy to understand and to calculate.

2. It is not affected by extreme values.

3. It can be determined graphically.

4. In some cases it can be located by inspection only.

5. It can be computed for the distributions of unequal class intervals provided the modal class; the class preceding the modal class and succeeding the modal class are of equal width.

6. It represents the most frequent value of the distribution, practically it is very useful.

Demerits:

1. It is not based upon all the observations.

2. It is not subjected to algebraic treatments, i.e., we cannot compute the combined mode if we have the modes of the two series.

3. In some cases mode is ill defined. In some cases it is not possible to find a clear mode. Some series have two modes and some more than two modes.

4. As compared to mean, mode is affected much by fluctuations of sampling; it is an unstable measure of central tendency.

5. If the modal class or the class preceding or succeeding the modal class are of unequal width, it cannot be determined.

6. There are different formulas for the calculation of mode.

Uses:

1. It is used to find the ideal size; it is very useful in business forecasting.

2. It is very useful in ready-made market, e.g., shoes, shirts, jeans etc.

3. It is very useful in commercial management.

2.6 SUMMARY

The study of this chapter provides us the knowledge of central tendency and measures of central tendency. From the study of this chapter we came to know the definitions of the measures of central tendency as mean, median and mode. We studied and learnt different methods of computing mean. We learnt about weighted mean and combined mean. We learnt how we can calculate the mean, median and mode in case of individual series, in case of

discrete distribution and continuous distribution. We studied the grouping method of mode. We studied the merits, demerits and uses of mean, median and mode also. From the study of these methods, merits, demerits and uses we came to know the situations where which method is suitable and also which measure is suitable for the particular situation? Over all we learnt a lot about measures of central tendency.

2.7 GLOSSARY

Measures of central tendency: These are mean, median and mode and they provide us the most important knowledge about the distribution. All these give the study of central part of the distribution.

Inclusive class: A class in which its upper and lower both limits are included is called inclusive class.

Exclusive class: A class in which only one limit (generally lower) is included and other limit (upper) is not included in the class is called exclusive class.

Arithmetic Mean: Arithmetic means or simply mean is ratio of sum of values to the number of values.

Median: On arranging the values according to magnitude, the middle most value of the arranged data is the median.

Mode: The most frequent item in the series is called mode.

Multimodal Distribution: If a distribution has more than one mode, it is called multimodal distribution.

Ill-defined Mode: If in a distribution there exists more than one mode, mode is called ill-defined mode.

Cumulative frequency Table: A table got on adding the successive frequencies is called cumulative frequency table.

2.8 SELF ASSESMENT QUESTIONS

Question 1. Find the arithmetic mean of the following series:

2, 13, 4, 12, 13, 5, 16, 17, 11, 22, 32, 42, 44

Question2. Find the median and mode of the following series:

12, 32, 2, 3, 15, 16, 12, 13, 23, 43, 35, 12

Question 3. Find the mean of the following distribution:

x:	12	22	13	14	15	17	24	25
f:	2	1	3	10	8	5	3	2

Question4. Find the median for the given data.

I.Q.:	92	65	77	72	110	89	112	98
f:	3	5	4	10	6	2	1	5

Question5. Find the mean, median and mode for the data given in the following table:

I.Q.	Frequency	I.Q.	Frequency
90-100	11	130-140	43
100-110	27	140-150	28
110-120	36	150-160	16
120-130	38	160-170	1

Question6. Which of the following is not a measure of central tendency?

a. Mean b. variance c. median d. none of these

Question7. If a constant 10 is added in each observation of a set of data, the effect on mean is:

a. Mean is decreased by 10. b. Mean is increased by 10

c. Mean is multiplied by 10 d. Mean is not affected

Question8. The best measure of central value in case of qualitative data is:

a. Mean b. mode c. Median d. None of these

Question9. The measure which is affected much by extreme observations is:

a. Mean b. mode c. median d. Variance

Question10. If we add 3 in each observation, the effect on median is:

a. Median is decreased by 9 b. Median is increased by 3

c. median is not affected d. Median is multiplied by 3

Question11. The most frequent item in the series is called:

a. median b. mean c. mode d. none of these

Question12. The average of 3 numbers is 2 and the average another 4 numbers is 5. What is the combined average?

a. 7 b. 10 c. 3 d. 3.71

Question13. For a group of 100 students, the mean score in a test was found to be 40. Later on it it was found that a value 45 was misread as 54. The correct mean will be:

a. 40.50 b. 39.85 c. 39.80 d. 39.91

Question14. Arrangement of observations according to magnitude is carried out in:

a. mean b. mode c. median d. none of these

Question15. Grouping method is used for:

a. mean b. mode c. median d. all of these

Question16. Which measure of central tendency is most widely used?

a. mean b. mode c. median d. all of these

Question17. For asymmetrical distribution, the relation between the measures of central values is:

a. mean= mode- 2 median b. mode= 2 median- 3 mean

c. mode= 3 median- 2 mean d. mode = 3 mean- 2 median

Question18. Which measure is affected least by fluctuation of sampling is:

a. mode b. mean c. median d. none of these

Question19. The measure which cannot be calculated with open end classes is:

a. mean b. mode c. median d, all of theses

Answers: 1. 17.92, 2. 14, 12, 3. 16.24, 4. 77, 5. 126.4, 126.84, 122.86, 6. b, 7. b, 8. c, 9. a, 10. b, 11. c, 12. d, 13. d, 14. c, 15. b, 16. a, 17. c, 18. b, 19. a.

2.9 REFERENCES

Agarwal, B.L., Programmed Statistics, New Age International (P) Ltd, New Delhi, (2003).

Arora, P.N. and Malhan, P.K., Biostatistics, Himalaya Publishing House (P) Ltd. Mumbai, (1996).

Belle, G., Fisher, L., Heagerty, P.G. and Lumly, T.,Biostatistics- A Methodology for the Health Science, John Willey & Sons, inc. Pub., Hoboken, New Jersey, (2004).

Gupta, S.C. and Kapoor, V.K., Fundamentals of Mathematical Statistics, Sutan Chand & sons, Daryaganj, New Delhi, (1970).

Pagaro, M. and Gauvreau, K., Principles of Biostatistics, Duxbury Thomson Learning, U.S.A. (2000).

Prabhakara, G.N., Biostatistics, Jaypee Brothers Medical Publishers (P) Ltd, New Delhi, (2006).

Sudaram, K.R., Dwivedi, S.N., Sreenivas, V., Medical Statistics, Wolters Kluwer (P) Ltd, New Delhi. (2010).

2.10 TERMINAL QUESTIONS

Question1. Define arithmetic mean and write down its merits and demerits

Question2. Explain step deviation method of mean by taking an example.

Question3. Which measure is the best in qualitative data? Give its merits and demerits.

Question4. Which measure cannot be calculated with open end classes and why?

Question5. Which measure is best for a industry and why?

Question6. What is the importance of mean?

Question7. Explain grouping method of mode.

Unit-3 MEASURES OF VARIABILITY/ DISPERSION

CONTENTS

3.1 OBJECTIVES

From the study of this chapter the students will be able:

1. To know about the measures of variability such as range, interquartile range, mean deviation and standard deviation.
2. To know the advantages and disadvantages of these measures.
3. To know about the differences of these measurement.
4. To know the situations where which measure is better to use?
5. To know why the coefficient of variation is the best to comparing two or more series.

3.2 INTRODUCTION

In the previous chapter we study various measures of central tendency and the methods of measuring them. The measures of central tendencies give us only an idea of the central part of the distribution. Although these measures provide us important information about the population, i.e., they are necessary but not sufficient for explaining a distribution properly. These measures do not provide the information about the variability of the observations. Variability is an important factor in nature. There is slight variability in heights of a region of persons but there is more variation from the heights of the persons of other region. To understand a distribution more clearly we should study about variability.

Let us consider the example of distribution of marks 5 students of three sections of a class. The marks of section A are 5, 5, 5, 5, 5; the marks of section B students 3, 4, 5, 6, 7 ; and the marks of section C students are 0, 1, 5, 9, 10. We see that the total score of all the three section students is 25 and the average score of all the three section students is also same 5. But we see that there is no variability in section A marks; it is a constant series. There is less variability in marks of section B; but there is more variability in section C students marks. It is clear from the example that the distribution of marks of the three sections is quite different while the average is same.

The entire concept of statistics is based on variation. If there is no variation and all individuals are alike, there is no need for collecting any sample data because one individual can provide all the information of the complete data which we want to study on the basis of the complete data set. Bur variation in any phenomenon is inherent and statistical techniques help us to explain this variation. Thus we have to know dispersion. Dispersion means scatteredness and this tells us about the homogeneity and heterogeneity of the distribution. To know a distribution more clearly we study the variability or scatteredness.

3.3MEASURES OF VARIABILITY

Commonly used measures for variability are as follows:

1. Range
2. The interquartile range
3. Mean deviation
4. Standard deviation

3.3.1 RANGE

The simplest measurement of variability is range. It is defined as the difference between the extreme observations of the distribution, i.e., the difference between the largest and the smallest observation of the distribution. This uses the two extreme observations and does not provide the information on the middle values of the distribution. It is simple measurement of variability but its results are misleading if the two extreme observations are unusual.

There are four situations of occurrence in range value.

i. There may be no variability at all (see table A Below).
ii. There may be less variation between the extreme observations (see table B below) .
iii. There may be very much variation between the extreme observations (see table C below).
iv. None of the observation is representative of the mean (see table D below).

Table1. Situations in range value

A	B	C	D
7 7 7 7 7 7 7	6 6 8 7 8 6 8	2 5 6 7 8 9 12	49 0 0 0 0 0 0
Mean= 7	Mean= 7	Mean= 7	Mean= 7
Range= 5-5	Range= 6-8	Range= 2-12	Range= 0-49
Variation= 0	Variation= 2	Variation= 10	Variation= 49

3.3.1. Merits and Demerits of Range

Merits:

1. It is simple understand and to calculate
2. Unit of measurement is same as the unit of variable under study.
3. It does not require mathematical calculation.

Demerits:

1. It uses only two extreme observations and does not use all other information so there is no importance for collection of all other data observations. A good

measure should use all the data observations for providing better estimate of variability.

2. It is not an ideal measure.
3. It cannot be calculated if the distribution contains open-end classes.
4. It is not amenable to further mathematical treatment.

3.3.2 INTERQUARTILE RANGE

It provides more knowledge about the distribution as it includes middle 50% observations of the distribution. It is defined as the difference between the first and third quartile of the distribution. The third or upper quartile (Q_3) divides the distribution in such a way that 75% observations lie below it and 25% observations of the distribution above it. It's just opposite first quartile or lower quartile (Q_1) divides the distribution in such a way that 25% observations of the distribution lie below it and 75% observations above it. Thus, interquartile range gives us information about 50% central observations of the distribution. If interquartile range is high, it means that the middle 50% observations are a long distance and if it is low, it means that middle 50% observations are closely near to each other. Quartiles divide a series into four equal parts.

There are different methods for calculating interquartile range, depending upon the type of distribution. For calculating interquartile range we have to calculate the two quartiles.

3.3.2.1 Interquartile Range for Individual Series

Let the variable under study X takes the values $x_1, x_2, ...x_n$, then for the calculation of interquartile range we apply the following steps;

Step1. Arrange the given observations in ascending or descending order of their magnitude.

Step2. Now calculate first quartile (Q_1) and third quartile (Q_3) by the method:

Q_1 = Value of $\left(\frac{n+1}{4}\right)$th term in the arranged series.

Q_3 = Value of $3\left(\frac{n+1}{4}\right)$ th term in the arranged series.

Step 3. The value of $Q_3 - Q_1$ is the interquartile range.

Example1. Find the interquartile range for the following series:

2, 12, 23, 4, 34, 25, 21, 17, 14

Solution. Arrange the given observations in ascending order of magnitude, we get,

2, 4, 12, 14, 17, 21, 23, 25, 34

Here n= 9

$\therefore$ Q_1 = Value of $\left(\frac{n+1}{4}\right)$ th term in the arranged series.

= Value of $\left(\frac{9+1}{4}\right)th$ = 2.5 th term in the arranged series.

= Value of 2^{nd} term + ½ (3^{rd} term - 2^{nd} term)

= 4+ ½ (12 – 4) = 4+ 8/2 = 4+4 = 8

Q_3 = Value of $3\left(\frac{n+1}{4}\right)$ th term in the arranged series.

= Value of 3(2.5)th term = value of 7.5th term

= Value of 7^{th} term + value of ½ (8^{th} term – 7^{th} term)

= 23+ ½ (25 – 23) = 23+ 1 = 24

Interquartile range = $Q_3 - Q_1$ = 24 – 8 =16

This shows that 50% observations of the given series lie between 8 and 24.

3.3.2.2 Interquartile Range for Discrete Distribution

If $x_1, x_2, ...x_n$ are the observations in a discrete distribution according to some characteristic and $f_1, f_2, ...f_n$ be their corresponding frequencies then for the calculation of quartiles, we calculate the cumulative frequencies. The interquartile range is calculated with the help of the following steps.

Working steps for interquartile range

Step1. Arrange the given values in ascending order of magnitude.

Step2. Find the total of frequencies, called cumulative frequency (denoted by c.f)

.Step3. Find $\frac{N}{4}$ for first quartile, where N=$\sum f$.

Step4. Find cumulative frequency just greater than $\frac{N}{4}$. The value of x, corresponding this cumulative frequency is the value of first quartile.

Step5. Find $3\left(\frac{N}{4}\right)$ for third quartile.

Step6. Find cumulative frequency just greater than $3\frac{N}{4}$. The value of x, corresponding to this cumulative frequency is the value of third quartile.

Step7. For interquartile range, find $Q_3 - Q_1$.

Example2. Find the interquartile range for the following data.

X	21	15	17	9	5	7	8	10
F	2	5	3	4	5	1	6	12

Solution. For calculating the interquartile range, first of all we calculate first and third quartile and for this, we arrange the values of x in ascending order and then prepare the cumulative frequency table as follows:

Table2. Quartiles in Discrete Distribution

x	F	c.f.
5	5	5
7	1	6
8	6	12
9	4	16
10	12	28
15	5	33
17	3	36
21	2	38
	$N = \sum f = 38$	

Here $\frac{N}{4} = \frac{38}{4} = 9.5$. The cumulative frequency just greater than 9.5 is 12 and the value of x, corresponding to cumulative frequency 12 is 8. So the first quartile is 8.

Here $3\frac{N}{4} = 3 \times \frac{38}{4} = 28.5$. The cumulative frequency just greater than 28.5 is 33 and the value of x, corresponding to cumulative frequency 33 is 15. So the third quartile is 15.

Interquartile range = $Q_3 - Q_1 = 15 - 8 = 7$.

This shows that 50% observations of the given series lie between 8 and 15.

3.3.2.3 Interquartile Range for Continuous Distribution

When the data is in class interval form, the class corresponding to c.f. just greater than $\frac{N}{4}$ is the first quartile class and the value of first quartile is given by:

$$Q_1 = L + \frac{\left(\frac{N}{4} - C\right)}{f} \times h$$

Where L= lower limit of the first quartile class

C= cumulative frequency just before the first quartile class

N= Total of frequency

f = frequency of the first quartile class

h = magnitude of the first quartile class

When the data is in class interval form, the class corresponding to c.f. just greater than $3\frac{N}{4}$ is third quartile class and the value of third quartile is given by:

$$Q_3 = L + \frac{\left(3\frac{N}{4} - C\right)}{f} \times h$$

Where L= lower limit of the third quartile class

C= cumulative frequency just before the third quartile class

N= Total of frequency

f = frequency of the third quartile class

h = magnitude of the third quartile class

Example3. The following table gives the distribution of weights of 100 persons. Find the interquartile range for this data.

Weight	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90
Frequency	1	3	6	10	15	25	15	10	11	4

Solution. For calculating the interquartile range, first of all we calculate first and third quartile and for this we prepare the cumulative frequency table as follows:

Table3. Quartiles in Continuous Distribution

Weight (in kg) (x)	Frequency (f)	Cumulative Frequency (c.f.)
40-45	1	1
45-50	3	4
50-55	6	10
55-60	10	20
60-65	15	35
65-70	25	60
70-75	15	75
75-80	10	85
80-85	11	96
85-90	4	100
Total	$N = \sum f = 100$	

Here $\frac{N}{4}=\frac{100}{4}=25$. The cumulative frequency just greater than 25 is 35 and the corresponding class is 60 – 65. So 60 – 65 class is the first quartile class and the first quartile is given by:

$$Q_1 = L + \frac{\left(\frac{N}{4} - C\right)}{f} \times h$$

Where L= lower limit of the first quartile class = 60

C= cumulative frequency just before the first quartile class = 20

N= Total of frequency = 100

f = frequency of the first quartile class = 15

h = magnitude of the first quartile class = 5

$$\therefore \quad Q_1 = L + \frac{\left(\frac{N}{4} - C\right)}{f} \times h = 60 + \frac{(25-20)}{15} \times 5 = 60 + 1.66 = 61.66$$

Here $3\frac{N}{4}=3\frac{100}{4}=75$. The cumulative frequency just greater than 75 is 85 and the corresponding class is 75 – 80. So 75 – 80 class is the third quartile class and the third quartile is given by:

$$Q_3 = L + \frac{\left(3\frac{N}{4} - C\right)}{f} \times h$$

Where L= lower limit of the third quartile class = 75

C= cumulative frequency just before the third quartile class = 75

N= Total of frequency =100

f = frequency of the third quartile class = 10

h = magnitude of the third quartile class = 5

$$\therefore \quad Q_3 = L + \frac{\left(3\frac{N}{4} - C\right)}{f} \times h = 75 + \frac{(75-75)}{10} \times 5 = 75$$

So the third quartile is 75.

Interquartile range = $Q_3 - Q_1 = 75 - 61.66 = 13.34$.

This shows that 50% observations of the given series lie between 61.33 and 75.

3.3.2.4 Merits and Demerits of Interquartile Range

Merits:

1. It is better measure than range.
2. It provides information on 50% observations of the distribution.
3. It is not affected from extreme observations
4. It can be calculated in case of open- end classes.
5. Unit of measurement is same as the unit of variable under study.

Demerits:

1. It provides information only on middle 50% observations and does not give information on rest of the 50% observations. A good measure should use all the data observations for providing better estimate of variability.
2. It is not an ideal measure.
3. Its value is quite variable from sample to sample.
4. It is not amenable to further mathematical treatment.

3.3.3 MEAN DEVIATION

Arithmetic mean, the measure of central tendency plays an important role in the measure of variability also. A measure of variability in the data can be obtained as the average deviation of observations from the arithmetic mean. This is called mean deviation. A measure of variability can be obtained by taking the deviations of the observations from their mean, ignoring the signs. This deviation can be taken from any mean but usually arithmetic mean, median and mode is used for this. Mean deviation is least when deviation is taken from median.

Let us consider the marks of 5 students of a class. The marks are 6, 7, 9, 10, and 18. The arithmetic mean of the marks is 10. The deviations of marks from their arithmetic mean are – 4, – 3, –1, 0, 8. We see that the sum of deviations is zero. This is the property of arithmetic mean, i.e., the sum of deviations taken from mean is zero. So in calculating mean deviation come across two problems:

1. Which average should be taken for deviation?
   The answer to this problem is that any of the averages – mean, median or mode can be used for deviation. But generally the arithmetic mean is used.
2. What should be the sign of the deviations?
   The answer to this problem is that we should use the positive sign for the deviations, i.e., the absolute deviations should be taken. This is done because the sum of deviations taken from mean is zero. This is clear from the example of marks of 5 students taken in the above paragraph.

3.3.3.1 Mean Deviation for Individual Series

For an individual series if $x_1, x_2, ...x_n$ are the values of the variable x, then the mean deviation is given by the following formula:

$$\text{Mean Deviation (M.D.)} = \frac{1}{n}\sum |x_i - A|$$

When A is taken as arithmetic mean, it is called mean deviation about arithmetic mean. When A is taken as median, it is called mean deviation about median and when A is taken as mode, it is called mean deviation about mode.

Example4. Find the mean deviation about mean for the following series:

2, 4, 5, 7, 10, 13, 9, 15, 25

Solution. For finding the mean deviation about mean first of all we find the arithmetic mean of the given series.

$$\bar{x} = \frac{\sum x}{n} = \frac{90}{9} = 10$$

Now we form the following table:

Table4. Mean deviation in Discrete Series

x	$\lvert x - \bar{x} \rvert$
2	8
4	6
5	5
7	3
10	0
13	3
9	1
15	5
25	15
$\sum x = 90$	$\sum \lvert x - \bar{x} \rvert$ **=46**

Mean Deviation (M.D.) = $\frac{1}{n}\sum|x-\bar{x}| = \frac{1}{9}\times 46 = 5.11$

Hence the mean deviation of the given series is 5.11

3.3.3.2 Mean Deviation for Discrete Distribution

If $x_1, x_2, ...x_n$ are the observations in a discrete distribution according to some characteristic and $f_1, f_2, ...f_n$ be their corresponding frequencies then the mean deviation about mean is given by the following formula:

Mean Deviation (M.D.) = $\frac{1}{N}\sum f|x-\bar{x}|$, where N = $\sum f$ = Total frequency

Example5. Ten patients were examined for uric acid test. The operation was performed 1050 times and the frequencies so obtained for different number of patients (x) are shown in the table given below.

x:	0	1	2	3	4	5	6	7	8	9	10
f:	2	8	43	133	207	260	213	120	54	9	1

Compute the Mean deviation about arithmetic mean.

Solution. Let 5 be the assumed mean. Now we prepare the following table for the calculation of mean deviation about mean.

Table5. Mean Deviation for Discrete Distribution

X	Frequency (f)	d = x- 5	fd	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
0	2	-5	-10	5.01	10.02
1	8	-4	-32	4.01	32.08
2	43	-3	-129	3.01	129.43
3	133	-2	-266	2.01	267.33

4	207	-1	--207	1.01	209.07
5	260	0	0	0.01	2.6
6	213	1	213	0.99	210.87
7	120	2	240	1.99	238.8
8	54	3	162	2.99	161.46
9	9	4	36	3.99	35.91
10	1	5	5	4.99	4.99
	$\sum f = 1050$		$\sum fd = 12$		$\sum f\|x - \bar{x}\| = 1302.56$

Arithmetic mean $\bar{x} = A + \frac{\sum fd}{N} = 5 + \frac{12}{1050} = 5 + 0.0114 = 5.01cm$

Now, Mean deviation (M.D.) = $\frac{1}{N}\sum f|x - \bar{x}| = \frac{1}{1050} \times 1302.56 = 1.24cm$

3.3.3.3 Mean Deviation for Continuous Distribution

In case of continuous distribution, there are given class intervals and their corresponding frequencies. First of all we find the mid values of these classes and treat them as the variable values. Then we calculate mean and then mean deviation about mean. The procedure will be clear from the following example.

Example 6. For the data given in the below table on systolic BP of 68 patients, calculate the mean deviation about arithmetic mean.

Systolic BP (mmHg)	Frequency (f)	Systolic BP (mmHg)	Frequency (f)
90-100	3	140-150	11
100-110	5	150-160	9
110-120	7	160-170	6
120-130	10	170-180	2
130-140	15		

Solution. For the calculation of mean deviation we prepare the following table:

Table6. For calculation of Mean in Continuous Distribution

Systolic BP (mmHg)	Frequency (f)	Mid Value (x)	fx	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
90-100	3	95	285	40.6	121.8
100-110	5	105	525	30.6	153
110-120	7	115	805	20.6	144.2
120-130	10	125	1250	10.6	106
130-140	15	135	2025	0.6	9.0
140-150	11	145	1595	9.4	103.4
150-160	9	155	1395	19.4	174.6
160-170	6	165	990	29.4	176.4
170-180	2	175	350	39.4	78.8
Total	$\sum f = 68$		$\sum fx = 9220$		$\sum f\lvert x-\bar{x}\rvert = 866.6$

$$\bar{x} == \frac{\sum_{i=1}^{n} f_i x_i}{N} = \frac{9220}{68} = 135.6 mmHg$$

Mean deviation about mean (M.D.) is given by

$$\frac{1}{N} \sum f|x - \bar{x}| = \frac{1}{68} \times 866.6 = 12.74 mmHg$$

Hence the mean deviation about mean of the systolic BP is 12.74mmHg.

3.3.3.4 Merits, Demerits and uses of Mean Deviation

Merits:

1. Mean deviation is easy to understand and to calculate.

2. Mean deviation is least when taken from median.

3. Mean deviation is less affected from extreme values as compared to range and standard deviation.

4. Mean deviation is better measure of variability than range and interquartile range as it uses all the observations of the data.

Demerits:

1. It is rarely used in social sciences.

2. Its results are not accurate because it is least when taken from median but median itself is not a good measure of central value when the variations in the series are large.

3. It is not suitable to further algebraic treatment as the negative signs of the deviations are also taken as positive.

Uses:

It is used in studying the economic problems.

3.3.4 STANDARD DEVIATION

For describing the scatteredness of the data values the best measure of variability is the standard deviation. It is denoted by σ. If standard deviation in a data is small, it means there is high degree of homogeneity in the data values and vice versa if the value of standard deviation is large, it means there is a large heterogeneity in the data values.

It is defined as the positive square root of the arithmetic mean of the deviations of values when the deviations are taken from their arithmetic mean.

3.3.4.1 Standard Deviation for Individual Series:

Let the variable under study X takes the n values $x_1, x_2, ...x_n$, their standard deviation is given by the following formula:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \qquad \text{or} \qquad \sigma = \sqrt{\frac{\sum d^2}{n}} \qquad \text{where } d = x_i - \bar{x}$$

The steps of the procedure are as follows:

Step1. Compute the arithmetic mean of the given series.

Step2. Compute the deviations of the series values from the mean, i.e., compute $d = x_i - \bar{x}$

Step3. Compute the square of the values got in step 2, i.e., compute $d^2 = (x_i - \bar{x})^2$.

Step4. Find the sum of values got in step 3 and divide it by the number of values, i.e., compute $\frac{\sum d^2}{n} = \frac{\sum (x_i - \bar{x})^2}{n}$.

Step5. Take the square of the value got in step 4. This is the required value of the standard deviation.

The procedure will be clear from the example given below:

Example7. Compute the standard deviation of the following series:

12, 15, 17, 21, 28, 27,

Solution. For the computation of standard deviation we prepare the following table:

Table7. S.D. for Individual Series

x	$d = (x_i - \bar{x})$	$d^2 = (x_i - \bar{x})^2$
12	-8	64
15	-5	25
17	-3	9
21	1	1
28	8	64
27	7	49
$\sum x = 120$		$\sum d^2 = \sum (x_i - \bar{x})^2 = 212$

Arithmetic mean $\bar{x} = \frac{\sum x}{n} = \frac{120}{6} = 20$

Standard deviation $\sigma = \sqrt{\frac{\sum d^2}{n}} = \sqrt{\frac{212}{6}} = \sqrt{35.33} = 5.94$

Hence the standard deviation of the given series is 5.94

3.3.4.2 Short-Cut Method of Standard Deviation

This method is applied when mean is in fractional form because in that case the deviations and their squares make the calculations difficult. So in this case we take the deviations of the values from an assumed mean.

Let d = x – A, here A is the assumed mean, then in this case the formula for standard deviation is given by as :

$\sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$ Where n is the number of observations.

We follow the following steps for the commutation of S.D. in this case:

Step1. Take any value of the series as assumed mean A.

Step2. Compute the deviations of the series values from the assumed mean, i.e.,

Compute $d = x_i - A$

Step3. Find the total of step 2 values, i.e., find total of d, i.e., Σ d.

Step4. Divide the value of step 3 by number of values 'n' and find its square, i.e,

$\left(\frac{\sum d}{n}\right)^2$.

Step5. Compute the square of the values got in step 2, i.e., compute $d^2 = (x_i - A)^2$.

Step5. Find the sum of values got in step 5 and divide it by the number of values, i.e.,

Compute $\frac{\sum d^2}{n} = \frac{\sum (x_i - \bar{x})^2}{n}$.

Step6. Subtract the value of step 4 from value of step 5 and then take its square root.

This is the required value of the standard deviation.

The procedure will be clear from the example given below:

Example8. Find the standard deviation in the above example 7 by short- cot method.

Solution. Let us take 21 as assumed mean A. Now we prepare the following table for the computation of standard deviation.

Table8. For Short –Cut Method of S.D.

x	$d=(x-21)$	d^2
12	-9	81
15	-6	36
17	-4	16
21	0	0
28	7	49
27	6	36
Total	$\sum d=-6$	$\sum d^2=218$

$$\therefore \qquad \sigma=\sqrt{\frac{\sum d^2}{n}-\left(\frac{\sum d}{n}\right)^2}=\sqrt{\frac{218}{6}-1}=\sqrt{35.33}=5.94$$

3.3.4.3 Standard Deviation in Discrete Frequency Distribution

If $x_1, x_2, ...x_n$ are the observations in a discrete distribution according to some characteristic and $f_1, f_2, ...f_n$ be their corresponding frequencies then standard deviation can be calculate with the help of these methods:

1. Actual mean method
2. Assumed mean method
3. Step deviation method

The procedures of the three methods will be clear with the help of the examples.

1. **Actual mean method:**

   For this we use the following formula:

$$\sigma = \sqrt{\frac{\sum f(x-\bar{x})^2}{N}}$$

Where $N = \sum f$ =Total frequency

Example9. Calculate the standard deviation of the distribution of marks of the B.Sc. botany class students. The data is given below:

x: 12, 18, 17, 15, 20, 25, 32, 42

f: 2, 1, 3, 5, 1 2 10 1

Solution. For the calculation of standard deviation we prepare the following table:

Table9. S.D. for Discrete Frequency Distribution by Actual Mean

X	f	f x	$d = (x-\bar{x})$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
12	2	24	-12	144	288
18	1	18	-6	36	36
17	3	51	-7	49	147
15	5	75	-9	81	405
20	1	20	-4	16	16
25	2	50	1	1	2
32	10	320	8	64	640
42	1	42	18	324	324

	$N = 25$	$\sum fx = 600$			$\sum f(x-\bar{x})^2 = 1858$

Arithmetic mean $\bar{x} = \frac{\sum fx}{\sum f} = \frac{600}{25} = 24$

Standard deviation $\sigma = \sqrt{\frac{\sum f(x-\bar{x})^2}{N}} = \sqrt{\frac{1858}{25}} == \sqrt{74.32} = 8.62$

Hence the standard deviation of the given distribution is 8.62

2. Assumed Mean Method:

In this method we take a middle vale of x as the assumed mean A and the apply the following formula:

$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$ **;** where symbols are in their usual meaning.

For the procedure of this method we take a example below:

Example10. For the above example 9 apply assumed mean method for computing the standard deviation.

Solution. Let us assume A= 20. Now for the calculation of standard deviation we prepare the following table:

Table10. S.D. for Discrete Frequency Distribution by Assumed Mean

X	f	$d = (x - A)$	fd	fd^2
12	2	-8	-16	128
18	1	-2	-2	4

17	3	-3	-9	27
15	5	-5	-25	125
20	1	0	0	0
25	2	5	10	50
32	10	12	120	1440
42	1	22	22	484
	$N = 25$		$\sum fd = 100$	$\sum fd^2 = 2258$

Standard Deviation $\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\frac{2258}{25} - \left(\frac{100}{25}\right)^2}$

$$= \sqrt{90.32 - 16} = \sqrt{73.68} = 8.62$$

Hence the standard deviation of the given distribution is 8.62

3. Step Deviation Method:

This method is applied when the values have some common interval (say h), we divide the deviations by this common interval and apply the following formula:

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times h$$

The procedure of the method will be clear from the example given below:

Example11. Daily high blood pressure of a patient on 100 days is given below:

Bp (mmHg):	102	106	110	114	118	122	126
No. of days:	3	9	25	35	17	10	1

Calculate the standard deviation of the above data.

Solution. Let us take the assumed mean A= 114. Here common interval h= 4. Now we prepare the following table for the calculation of standard deviation.

Table11. S.D. for Discrete Frequency Distribution by Step Deviation

BP (mmHg)	f	$d=\frac{(x-114)}{4}$	fd	fd^2
102	3	-3	-9	27
106	9	-2	-18	36
110	25	-1	-25	25
114	35	0	0	0
118	17	1	17	17
122	10	2	20	40
126	1	3	3	9
Total	$N=100$		$\sum fd=100$	$\sum fd^2=2258$

$$\text{S.D.}=\sigma=\sqrt{\frac{\sum fd^2}{N}-\left(\frac{\sum fd}{N}\right)^2}\times h=\sqrt{\frac{154}{100}-\left(\frac{-12}{100}\right)^2}\times 4$$

$$=\sqrt{1.54-.0144}\times 4=1.235\times 4=4.94\,\text{MmHg}$$

3.3.4.4 Standard Deviation in Continuous Frequency Distribution

In case of continuous distribution we find the mid values of classes and treated them as the variable values x. In this case we can apply all the three methods discussed in previous

section. But generally step deviation is applied. The formula is the same as in case of discrete distribution discussed. The procedure is described in the example given below.

Example12. Calculate the standard deviation for the following table giving the age distribution of 542 persons of a city.

Age in years:	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
No. of members:	3	61	132	153	140	51	2

Solution. For the calculation of standard deviation, let us take $d = (x - 55)/10$. Here we let assumed mean A = 55 and common interval (h) = 10. Now we prepare the following table:

Table12. S.D. in Continuous Distribution by Step Deviation Method

Age group	Mid-value (x)	Frequency (f)	$d = \frac{(x-55)}{10}$	fd	fd^2
20-30	25	3	-3	-9	27
30-40	35	61	-2	-122	244
40-50	45	132	-1	-132	132
50-60	55	153	0	0	0
60-70	65	140	1	140	140
70-80	75	51	2	102	204
80-90	85	2	3	6	18
Total		N=542		$\sum fd = -15$	$\sum fd^2 = 765$

$$\text{S.D.} = \sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times h$$

$$= \sqrt{\frac{765}{542} - \left(\frac{-15}{542}\right)^2} \times 10 = \sqrt{1.334} \times 10 = 11.55$$

Hence the standard deviation of age of the given distribution is 11.55 years.

3.3.4. Merits, Demerits and Uses of Standard Deviation

Merits:

1. It is rigidly defined.
2. It uses all the observations of the data in calculation.
3. It is used in correlation.
4. It is affected least by fluctuation of sampling.
5. It is suitable for further mathematical treatments.
6. It is the best measure of variability.

Demerits:

1. Its calculation is difficult in comparison to other measures of dispersion.
2. It is sensitive to extreme values.
3. It is not easily understandable for a common person.

Uses:

1. It is best measure of comparison of variability.
2. It is used in partitioning between groups and within groups in analysis of variance and design of variance.
3. It is used with mean in normal distribution for finding the areas.
4. It shows best dispersion of values from the mean.
5. It is very much used in medical field.

3.4 VARIANCE AND COEFFICIENT OF VARIATION

3.4.1 VARIANCE

It is just the square of the standard deviation. It is denoted by σ^2. In other words variance is the arithmetic mean of the squares of the deviations, when deviations are taken from their arithmetic mean.

3.4.2 COEFFICIENT OF VARIATION

It is the best measure of the comparison of variability of the two series or populations. The units of measurement of the two populations may be different. This comparison is possible because it is a unit free measure. It is presented in percentage and
is expressed as:

Coefficient of variation (C.V.) = $\frac{\sigma}{\bar{x}} \times 100$; where notations have their usual meaning.

A series having lesser c.v. is called more consistent or more homogeneous, i.e., the values of the series are closer to the mean of the series and if the c.v. of a series is larger, it is called more variable or in other words more heterogeneous series, i.e., the values of the series far apart from the mean of the series.

Example13. Calculate the coefficient of variation of the distribution of marks of the B.Sc. botany class students. Given the following information:

Average marks Standard deviation of marks

$\bar{x} = 24$ $\sigma = 6$

Solution. Coefficient of variation (C.V.) = $\frac{\sigma}{\bar{x}} \times 100 = \frac{6}{24} \times 100 = 25\%$

Hence the C.V. of the marks is 25%.

Example14. The following data shows the mean and standard deviation on systolic BP and weight of 10 persons as:

BP		Weight	
Mean	S.D.	Mean	S.D.
120	15	60	4.5

Compare the two characteristics.

Solution. For comparison of the two characteristics we find the C.V. of these characteristics.

$$\text{C.V for BP} = \frac{\sigma}{\bar{x}} \times 100 = \frac{15}{120} \times 100 = 12.5\%$$

$$\text{C.V for Weight} = \frac{\sigma}{\bar{x}} \times 100 = \frac{4.5}{60} \times 100 = 7.5\%$$

We see that the coefficient of variation of BP is more than the coefficient of variation of weight so BP is more variable than the weight of the given persons.

3.5 SUMMARY

The study of this chapter provides us the knowledge of dispersion or variability and measures of variability. From the study of this chapter we came to know the definitions of the measures of variability - range, interquartile range, mean deviation and standard deviation. We learnt the different method for calculating interquartile range, mean deviation and standard deviation for individual series, in case of discrete distribution and continuous distribution. We studied the merits, demerits and uses of these measures. From the study of these methods, merits, demerits and uses we came to know the situations where which method is suitable and also which measure is suitable for the particular situation? Coefficient of variation is the best measure for comparison on the basis of variability as it uses all the values and is a unit free measure.

3.6 GLOSSARY

Range: It is the difference between the largest and the smallest observation of the data.

Interquartile Range: It is based on the first and third quartile of the distribution and it express the limit for the middle 50% observations of the data.

Dispersion: It shows the spread of the data values around the central value of the distribution.

Mean Deviation: It is the average of the absolute deviations of the data values. Generally deviations are taken from mean.

Variance: It is the arithmetic mean of squares of the deviations of the values when deviations are taken from their mean.

Standard Deviation: It is the positive square root of the variance.

Coefficient of Variation: It is the best measure of comparison on the basis of variability of two or more series. It can also be used if the units of measurements are different because it is a unit free measure.

3.7 SELF ASSESMENT QUESTIONS

Question1. Find the range of the following series:

1, 23, 32, 24, 45, 42, 35, 37

Question2. Find interquartile range for the given series:

2, 4, 3, 5, 7, 10, 12, 13, 15, 14, 9, 1, 8, 6, 11, 16

Question3. Find the standard deviation of erythrocyte sedimentation rate (ESR) of the data 3, 4, 5, 4, 2, 4, 5, 3 found in 8 normal persons.

Question4. Calculate the mean and standard deviation of the following data on the length of fishes (in cm) .

Length:	10	20	30	40	50	60	70
No. of Fishes:	4	6	10	15	20	15	10

Question5. A firm has selected a random sample of 100 from its production line and has obtained the data shown in the table below:

Class interval	Frequency	Class interval	Frequency
130- 134	3	150-154	19
135- 139	12	155-159	12
140- 144	21	160-164	5
145-149	28		

Question6. Compute the standard deviation for data of hypo B.P. (in mmHg).

B.P. :	50	60	70	80	90	100	110	120
No. of Patients;	14	40	54	46	26	12	6	2

Question7. The C.V. of two series of observations are 20% and 25% with S.D. 5 and 8 respectively. What are their arithmetic means?

Question8. Which of the following is not a measure of dispersion?

a. Range b. Mean deviation c. median d. standard deviation

/**Question9.** The easiest measure of dispersion is:

a. Mean deviation b. S.D. c. interquartile range d. range

Question10. The range of the series 2, 3, 7, 9, 12, 17, 21 is:

a. 2 b. 21 c. 19 d. None of these

Question11. Which measure of dispersion is the best?

a. Range b. mean deviation c. C.V. d. S.D.

Question12. For comparison of two different series, the best measure of dispersion is:

a. Range b. S.D. c. mean deviation d. C.V.

Question13. Mean deviation is least when taken about:

a. Median b. mean c. zero d. mode

Question14. The mean of a series is 10 and C.V. 40%. Its S.D is:

a. 4 b. 16 c. 40 none of these

Question15. If all the values of a series are 4, the S.D. of this series is:

a. Zero b. 16 c. 4 d. 2

Answers: 1. 44, 2. 8, 3. 0.97, 4. 14.79, 5. 7.21, 6. 15mmHg, 7. 25, 32, 8. c, 9. d. 10. c, 11. d, 12. d, 13. a, 14. a, 15. a.

3.8 REFERENCES

Agarwal, B.L., Programmed Statistics, New Age International (P) Ltd, New Delhi, (2003).

Arora, P.N. and Malhan, P.K., Biostatistics, Himalaya Publishing House (P) Ltd. Mumbai, (1996).

Belle, G., Fisher, L., Heagerty, P.G. and Lumly, T.,Biostatistics- A Methodology for the Health Science, John Willey & Sons, inc. Pub., Hoboken, New Jersey, (2004).

Gupta, S.C. and Kapoor, V.K., Fundamentals of Mathematical Statistics, Sutan Chand & sons, Daryaganj, New Delhi, (1970).

Pagaro, M. and Gauvreau, K., Principles of Biostatistics, Duxbury Thomson Learning, U.S.A. (2000).

Prabhakara, G.N., Biostatistics, Jaypee Brothers Medical Publishers (P) Ltd, New Delhi, (2006).

Sudaram, K.R., Dwivedi, S.N., Sreenivas, V., Medical Statistics, Wolters Kluwer (P) Ltd, New Delhi. (2010).

3.9 TERMINAL QUESTIONS

Question 1. Define mean deviation and write down its merits and demerits

Question 2. Explain standard deviation and write down its merits and demerits.

Question 3. Which measure of dispersion is the best and why?

Question 4. Which measure is used if measurement units are different?

UNIT4: PRINCIPLES AND USES OF ANALYTICAL INSTRUMENTS

CONTENT

4.1- OBJECTIVES

In this chapter you will learn about the basic or common instrument used in the Laboratory such as pH meter, UV-visible spectrophotometer, Centrifuges (clinical, High-speed and ultra- centrifuge), Geiger Muller and scintillation counters

4.2. INTRODUCTION

In biological science there is basic requirement of having knowledge of instrument. Instrument play important role in laboratories such as soil analysis lab, diagnostic lab, molecular biology lab, microbiology lab, chemistry lab, physics lab and diagnostic lab. We cannot imagine the laboratories without basic instrument such as pH meter, centrifuge and spectrophotometer. These instruments are very necessary for analysis. Instrument not only speedup the analysis but also have accuracy, specificity and sensitivity. pH meter is one of the important instruments used in basic laboratory to measure pH of water, soil and food etc. Centrifuge is used to separate the mixture of two liquid depending on the density of liquid. UV spectrophotometer is generally used in analytical chemistry for the quantitative determination of different analyses, such as transition metal ions, highly conjugated organic compounds, and biological macromolecules. G-M counter and Scintillation counter is used for the detection of radioactivity.

4.3. PRINCIPLES AND USES OF ANALYTICAL INSTRUMENTS

4.3.1 pH METER

pH meter is generally used to determine the pH of soil, water and culture medium used for the cultivation of fungi and bacteria. There is presence of electrode which very sensitive to detect the change in H ion concentration. The electric circuit measure the electromotive force developed across the electrode pair. pH is abbreviation of "Pondus Hydrogenii"and it was proposed by Sorenson in 1909 in order to express the small concentration of hydrogen ions. pH is a unit of measure that describe acidity and alkalinity of solution. It is measured on a scale 0 to 14 and defined as the negative logarithm of hydrogen ion activity.

$pH= -\log [H^+]$

The pH value of a substance is directly related to the ratio of the hydrogen ion and hydroxyl ion concentrations. If the H+ concentration is higher than OH- the material is acidic. If the OH^- concentration is higher than H^+ the material is basic.

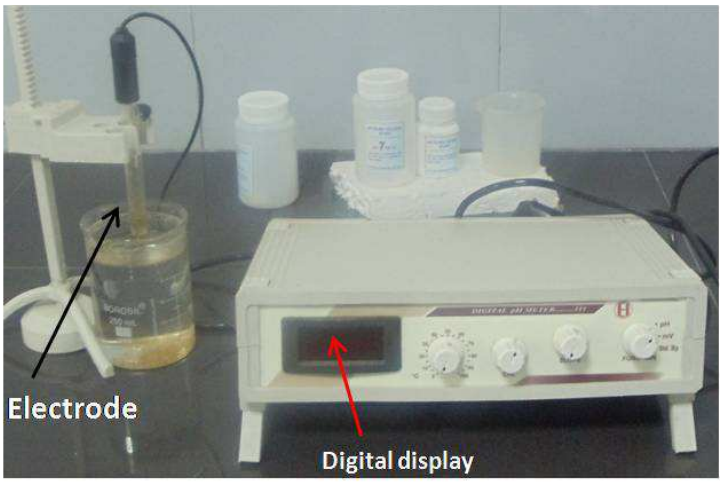


Figure 4.1. Digital pH meter

Principle of pH meter

The pH electrode consist of the pH-sensitive electrode which is a thin glass membrane whose outside surface contacts the solution to be tested. The inside surface of the glass membrane is exposed to a constant concentration of hydrogen ions (0.1 M HCl) (Fig.8).

Inside the glass electrode assembly, a silver wire, coated with silver chloride and immersed in the HCl solution known as Ag/AgCl electrode. This electrode carries current through the half-cell reaction. The potential between the electrode and the solution depends on the chloride ion concentration, but, since this is constant (0.1 M), the electrode potential is also constant. To complete electrical circuit reference electrode is needed. So, Ag/AgCl electrode is immersed in an 0.1 M KCl solution which makes contact with the sample through a porous fiber which allows a small flow of ions back and forth to conduct the current. The potential created at this junction between the KCl solution and the test solution is nearly zero and nearly unaffected by anything in the solution, including hydrogen ions. A voltmeter in the probe measures the difference between the voltages of the two electrodes. The meter then translates the voltage difference into pH and displays it on the screen .

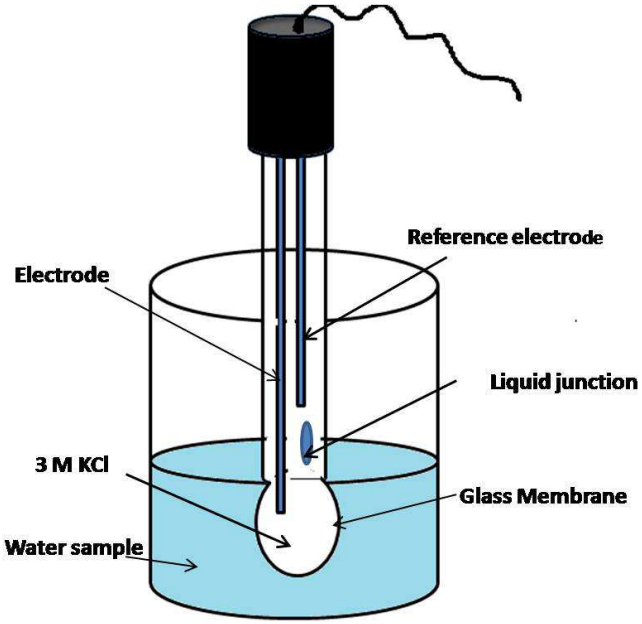


Figure 4.2: Working of pH meter

Calibration of pH meter

1. Switch on the pH meter and move the knob from stand by to pH.
2. Rinse the electrode with double distilled water and wipe out the tip gently with tissue paper,
3. Place the electrode into the solution of pH 7.0 and adjust the pH by adjustment given in pH meter.
4. After adjusting again wash the tip with distilled water and wipe out with tissue paper.
5. Dip the electrode in to solution of pH 4.0 (if your solution is acidic) and pH 9.0 (when your solution is basic)
6. Turn on the knob to slope position till it shows 4.0 or 9.0.
7. After reading the pH of sample. Again wash the electrode with distilled water.
8. Immerse the electrode in to the distilled till used for the next time.

4.3.2- UV-visible spectrophotometer

Spectroscopy is the measurement of (spectrum of light) electromagnetic radiation, absorbed, scattered, or emitted by atoms, molecules, or other chemical species. Each chemical species has unique energy states; spectroscopy can be used to identify the interacting species separately. Moreover; interaction of light with different compounds offers several possibilities of qualitative and quantitative measurements. Electrons in an atom move around the nucleus in orbitals and possess characteristic energy also known as *ground state*. Energy transfer to these electrons energizes them substantially to let them leave their orbital to jump over the next level of energy orbital or level; this is called as *excited state*. Generally, on ceasing the energy supply, electrons emit the absorbed energy in the form of radiation giving rise to the *atomic spectrum* or simply *"line spectrum"*, which is represented as graph of the amount of energy absorbed or emitted by a system against wavelength or similar electromagnetic parameters. Visible light forms a part of the electromagnetic spectrum with γ rays at one end having wavelength of the order of 10^{-14} m, and radio waves at the other end having wavelength 3×10^{3} m or greater. As atoms, each of the electrons in a molecule usually occupies the available lowest energy level (ground state). Electrons in a molecule change their energy level only after the absorption or emission of the distinct quanta (particular energy radiation) of radiation by the molecule. Depending upon the absorption and emission of energy quanta, there occur two viz., absorption and emission spectra respectively. For molecules, each ground and excited state is subdivided into a number of vibrational and rotational energy sublevels, molecular spectra is therefore seen as "*band spectra"*.

Types of spectra

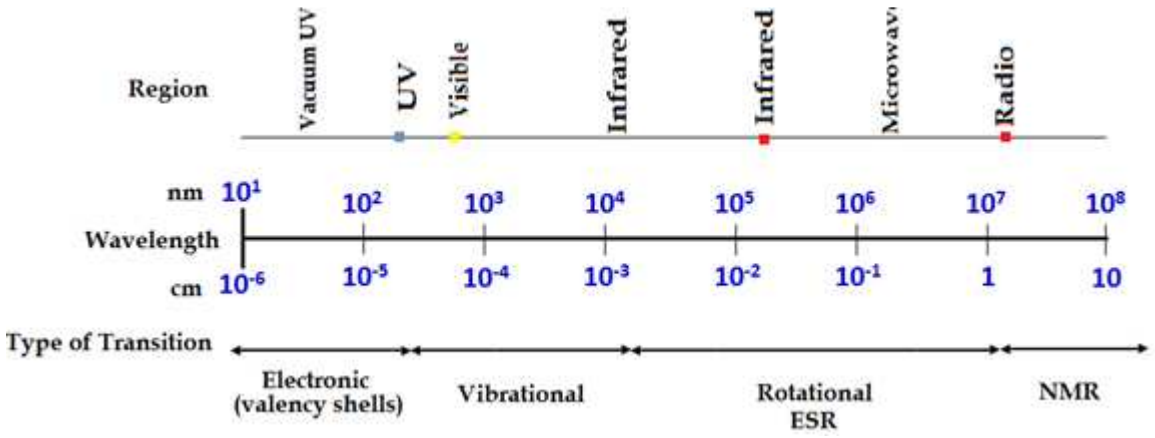


Figure 4.3. Types of spectra and their wavelengths

Electronic Spectra

Electronic spectra arise due to the outer electrons of atoms changing between major electronic energy levels. Such spectra occur in the visible and ultraviolet regions and are usually accompanied by changes in the rotational and vibrational energy levels. These spectra are used routinely in biochemistry. Fluorescence spectra may also arise owing to these transitions.

Vibration – Rotation Spectra

Vibration spectra caused by changes in the vibration energy levels. They occur in the near infra-red region and may be accompanied by changes in the rotational energy levels. Such spectra are sometimes used in studies of the detailed structure of biological macromolecules in non-aqueous environments.

Electron Spin Resonance (ESR) Spectra and Nuclear Magnetic Spectra

These spectra arise due to changes in the directions of the spins of electrons and nuclei respectively in a magnetic field. These two types of spectra are valuable for studying the structure of biological macromolecules.

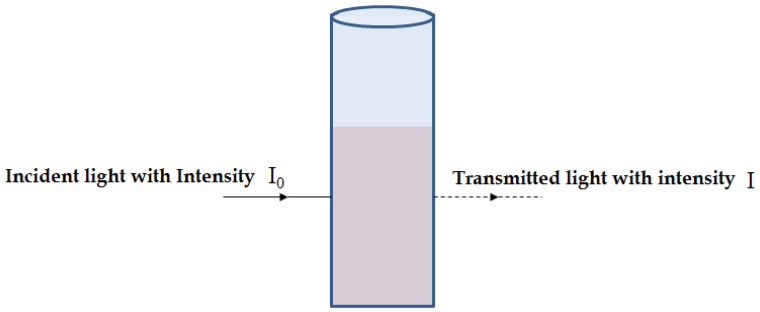
Molecular Band Spectra

Molecular band spectra may be resolved into a number of very close line spectra, corresponding to the vibrational and rotational energies of the electrons only at extremely high resolution.

3.0.1 Absorption

Photons of UV and Visible light may sometimes impart their energy to materials by interaction with individual atoms or molecules. Energy is imparted to the atoms or molecules, causing the excitation of valence electrons. Molecules with excited electronic states represent an unstable state will relax by allowing their electrons to fall to the ground state as soon as possible (10^{-16} sec.).

How much radiation is absorbed?



The absorbance (A) is defined as the $\log_{10}$ of the ratio of the incident to transmitted light intensities:

$A = log_{10} (\frac{I_0}{I})$ - The intensity is measured in the unit of power; therefore the absorbance is a unit less quantity. The intensity of light can be detected by photo-multiplier tubes. The collision of a photon of appropriate energy with the suitable molecule results in absorption of light.

Beer-Lambert Law or Law of absorption

The Beer-Lambert law relates the absorption of most molecular species to the concentration (C_n), the path length (l) and the molar absorptivity (ε).

$A = \varepsilon C_n l$

The wavelength of maximum absorption is known as λ max and is usually used as the wavelength for molar absorptivity (ε). ε is also sometimes known as extinction coefficient and is

measured in $dm^3mol^{-1}cm^{-1}$. The Beer-Lambert law is sometimes expressed in terms of transmittance T as:

T = 1/A

Pathlength of 1 cm is normally chosen to simplify the calculation of the absorbance or molar absorptivity. All modern spectrophotometers are designed to comply the Beer-Lambert law. A plot of absorbance versus wavelength is known as UV-Visible Spectrum and is measured by a UV-Visible Spectrophotometer.

3.0.2. Spectrometers /Spectrophotometer

A spectrometer or spectrophotometer is a device which measures the absorbance by molecules in a given sample. Spectrometers are a monochromator equipped with a photo-transducer. Single and double beam spectrophotometers are readily available in the market whereas double beam spectrophotometers are commonly used in scientific laboratories because of the ease of function and precision in readings of absorbance.

A

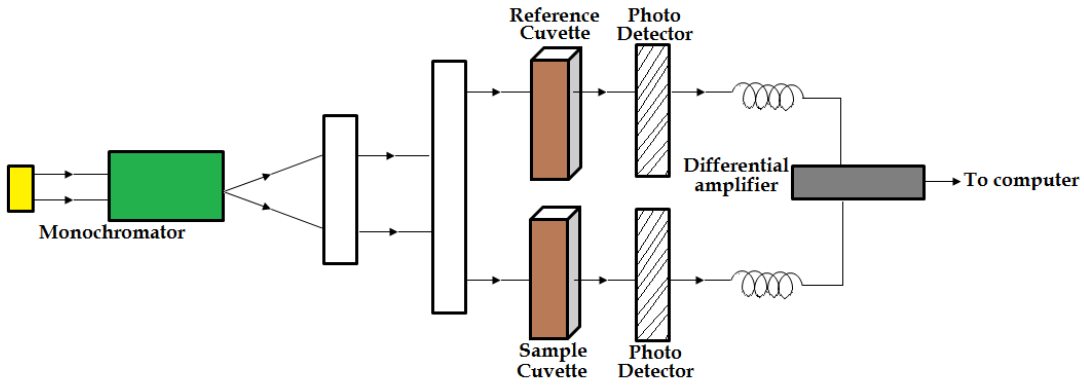


Figure 4.4. Schematic of Double Beam Spectrophotometer

spectrophotometer comprised of several devices (figure 3.1) such as monochromatic for monochromatic light of particular wavelength, grits, photo-detectors, and a processor that send signals to the interfaced computer. To understand and to define a spectrophotometer, it is recommended to study each and every component in details.

Light Source

The irradiation of a sample for UV and Visible spectroscopy requires a light source with constant output intensity. Tungsten (W) filament lamp are used for UV and visible light (320 – 2500) nm while hydrogen and deuterium lamps UV radiation. The light source should also be sufficiently intense so as to allow sufficient transmitted radiation to be detected when the absorption falls within a range of 0 -2. In more advanced spectrophotometers, tungsten-halogen lamps are frequently used. Such lamps contain small quantity of iodine and the lamp is enclosed in *quartz* housing. The iodine (a halogen) raises the temperature of lamp to 3500K, which permits the intensity of the output radiation down to ~ 190 nm (UV range). Quartz is permeable for UV radiation whereas glass blocks the UV radiation.

Hydrogen and deuterium lamps produce high intensity UV radiation. Electrical excitation of hydrogen atoms at low pressure may produce two hydrogen of either low energy with the high energy photon or higher energy with low energy photons. As a result of this unequal uniform output the deuterium or hydrogen lamps give rise to the outputs of radiation over a wavelength range of 160 – 375 nm.

Monochromator

A monochromator is the most commonly used device to select a wavelength of light for the irradiation of a sample. Monochromators use a series of lenses, mirrors, slit and windows together with either prisms and /or diffraction gratings to isolate a narrow band of wavelengths. It is impossible to select one wavelength range. The wavelength of the radiation source will tend to follow a *Gaussian* distribution around a mean value of wavelength known as the nominal wavelength. The effective *bandwidth* is defined as the wavelength range which corresponds to the half peak height width of

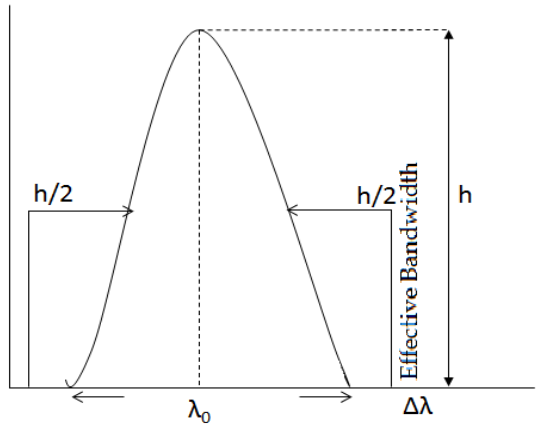


Figure 4.5 Effective Bandwidth

The wavelength distribution profile. There are two common types of monochromators viz., Monochromators based on refracting prisms and Monochromators based on diffraction grating

are used in the modern age spectrophotometers. A monochromator with a bandwidth of less than ~ 0.5 nm is perfectly acceptable.

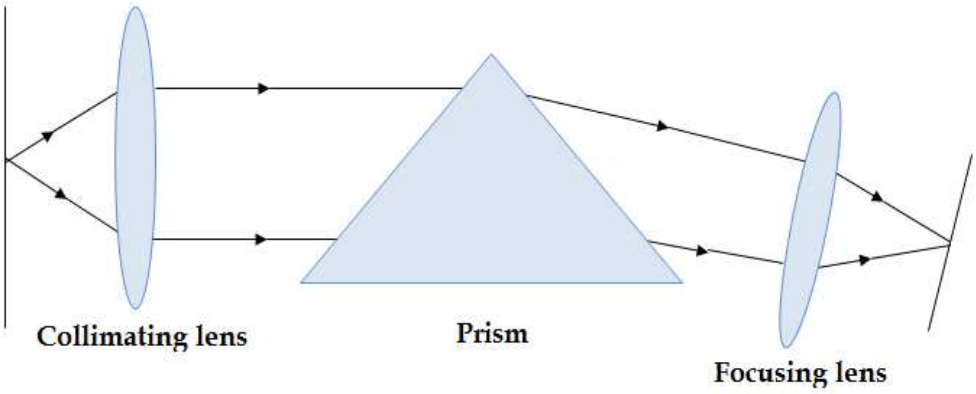


Figure 4.6 Simplified Schematic of a Prism Monochromator

In refracting prism monochromators, the white light enters via a slit passing through a collimator into the monochromator. A collimator is equipped with collimating lens, produce a parallel beam of radiation as shown in Figure 6. The light then passes through a refraction prism that disperses the light into its component wavelengths. The light is focused by another lens on the focal point where another slit is placed. By rotating prisms, radiation of different frequency is selected.

Diffraction grating based monochromator

White light passes through an entrance slit and is focused on a diffraction grating through a concave mirror. The diffraction grating disperses the light into its component wavelengths and reflects the light onto a second concave mirror. The grating can be rotated by a stepper motor. Light is then focused on the exit slit by this second concave mirror. By rotating the grit, light of different wavelengths can be selected.

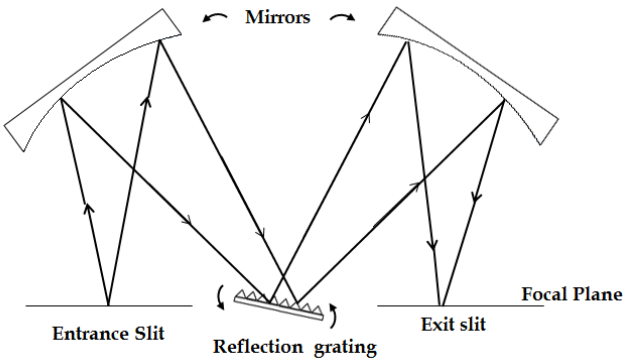


Figure4. 7. Optics of Diffraction Based

Light Detector

The absorbance of an analyte is performed by a photon tube or light detector which measures the intensity of transmitted light. A photo-multiplier tube or photon tube works on the principle of photo-emission or electronic transition involving the counting of photons enabling the monitoring of the intensity of light. Figure 3.5 show a photon tube where by the photons enter into the tube to emit electrons from the large cathode that are captured by a wire anode placed in front of the cathode. The entire circuit is encased in quartz housing under high vacuum. The circuit gets completed and the current generated is directly

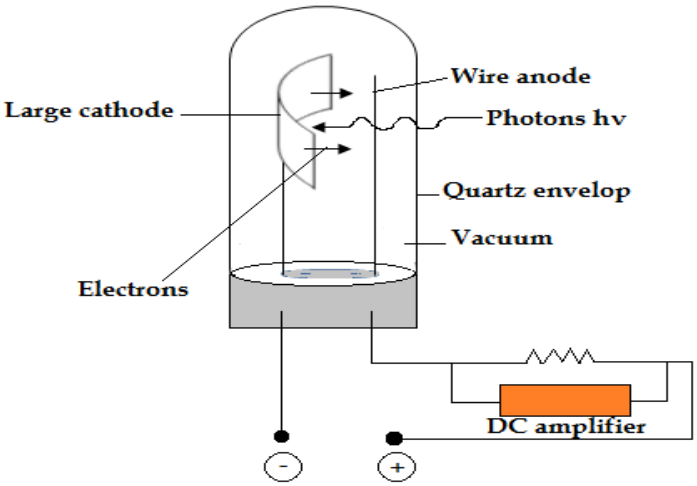


Figure4. 8. A photo-multiplier tube

Proportional to the intensity of light entering into the photon tube. Photon tube operates as photo-multiplier tube whereby a cascade effect is produced, which enables the collection of $10^6 – 10^7$ electrons per photon entering into the tube.

4.4 CENTRIFUGE

Centrifugation is based on the principle of centrifugal force in which liquid are subjected to high speed to separate solid from liquid or liquid from liquid depending on the density. In centrifugation heavy particle settled down and light particle will rises to the top. The substance which is settled down is called “Pellet” and the remaining fluid or overlying fluid is called “Supernatant”. Centrifugation process separates two substance of different density.

Basic principle of centrifugation

When the tube is filled with fluid and allowed to spin, as the rotor spin the apparent centrifugal force act on the sample of fluid and the analyte inside the fluid is pushed both radially outwards to the side of centrifuge tube.

Relative centrifugal force $F = m \omega^2 r$

Where,

F= Relative centrifugal force

m= Mass of particle

r= radius of the circular motion of the centrifuge (unit meter).

ω= angular velocity of the centrifuge (unit radian per sec)

The rate of sedimentation depends upon the applied centrifugal field (G), which depend upon the radical distance of the particle from the rotation axis and square of the angular velocity.

$G= \omega^2 r$

Where angular velocity (ω) = $\frac{2\pi}{60}$ revolution/min (one revolution is equal to 2 radian)

Putting the value in G , therefore

$$G = \frac{2\pi (\text{revolution/min})^2 \times r}{(60)^2}$$

$$\mathbf{G} = \frac{4\pi r}{3600}$$

The relative centrifugal force (RCF) in g units (g=9.8065 M/sec^2 or 980 cm /sec ^{2})

$$RCF = \frac{4\pi r}{3600 \, x \, 980} = 1.11 \times 10^{-5} (\text{revolution min}^{-1})^2 \, r$$

The rate of settling of particle depends upon (a) the relative centrifugal force (b) the density of the specimen, (c) the density of the fluid, (d) frictional forces, and (e) the size and shape of the specimen

Consider if the suspending particle is spherical, then the rate of settling of spherical particle can be expressed by following equation.

$$R = \frac{2r^2\, g\, (dp - df)}{9n}$$

Where,

R= Rate of settling

r= Radius of the particle

g= acceleration due to gravity

dp= Density of the particle

df=Density of the fluid

n=viscosity of the fluid

The rate of sedimentation depend on the applied centrifugal field i.e. RCF in g units (980 cm/sec^2).

Types of rotor

The centrifuge is equipped with two types of rotor

1. Fixed angle rotor: In this type of rotor, the centrifuge tube containing samples are placed in the shield in the rotor at fixed pre-set angle. The rotor holds the centrifuge at fixed inclination i.e. 35 degree to the vertical. The solute during centrifugation forced against the side of the tube resulting in faster separation of solute from the suspension. The disadvantage of this rotor is that there are chances of abrasion due to striking the particle to the wall of centrifuge tube. Another disadvantage of this rotor is there is formation of

smear like sedimentation than clear pellet formation. This type of rotor has short run time (Fig 9).

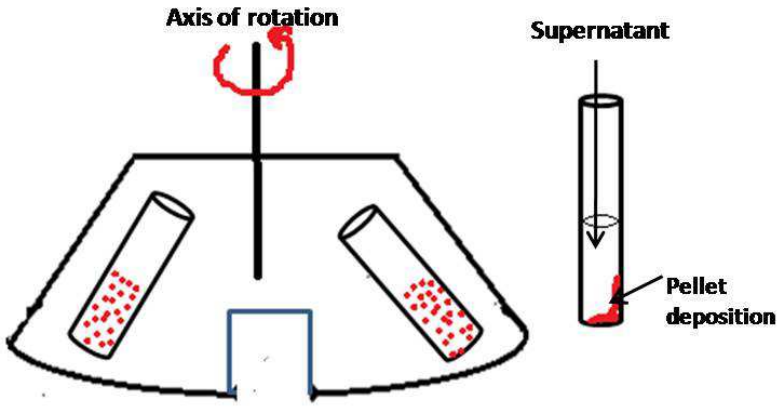


Figure 4.9: Fixed angle rotor and formation of pellet on side wall

2. Swinging bucket rotor: In this type of rotor the sample centrifuge tube are placed vertically and when the machine is started the bottom of the tube swing outward (horizontally) as the shaft rotates. This rotor has advantage than fixed rotor having clear pellet at the bottom of the tube (Fig.10).

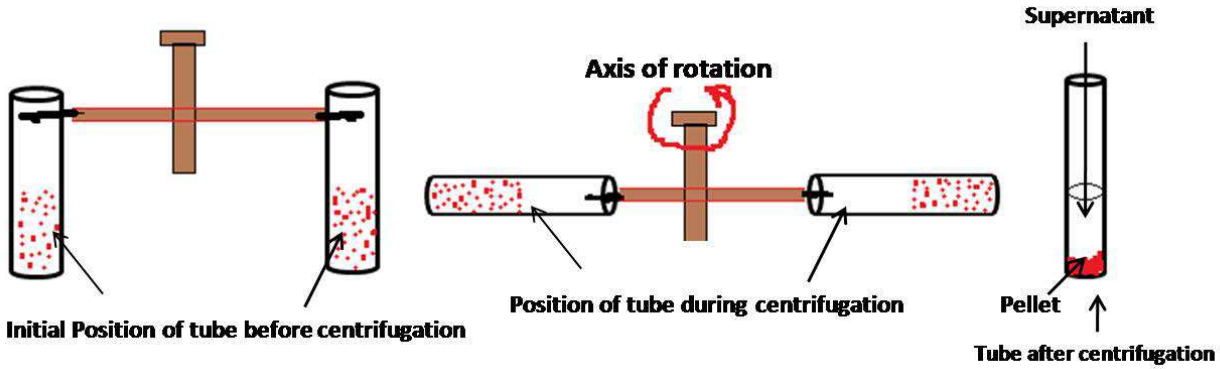


Figure 4.10: Working of swing bucket rotor before and after centrifugation

3. Vertical Rotor: In this type of rotor angle of placement of rotor is fixed but not at the slanting position it is vertical i.e. perpendicular to centrifugal field and the rotation axis as shown in Fig. 4.11. Band separate across the diameter of the tube rather than down the length of tube.

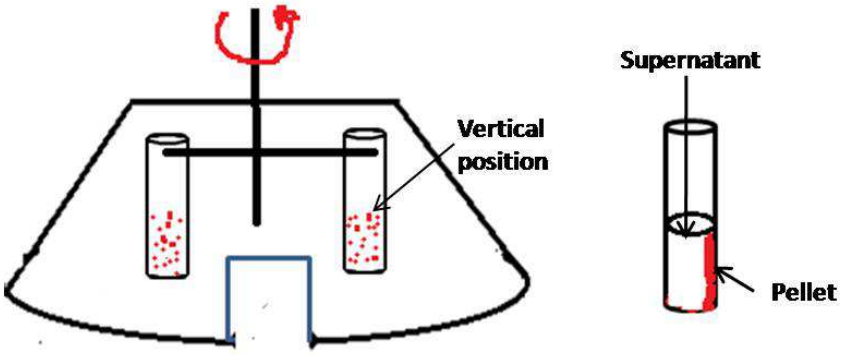


Figure 4.11: Working of vertical rotor before and after centrifugation

Types of centrifuge

Bench centrifuge: The most common type of centrifuge which generally used for common purposes such as separation of serum, plasma from blood sample required for serological reaction. It is not used for the fine substance such as organelle etc. The maximum speed of this type of centrifuge is 4000-5000 rpm and they operate is at ambient temperature. Now a day's small microfuge are available which can easily put into refrigerator to keep temperature cool for sedimentation to prevent denaturation heat sensitive substance such as Protein. .

Refrigerated centrifuge (Large capacity):

Refrigerated centrifuge have inbuilt cooling system for rotor such as compressor. It has speed of 6000 rpm with relative centrifugal field 6500g. Refrigerated centrifuge is used for the heat sensitive substance. This type of centrifuge can be run with fixed type rotor or swinging bucket type rotor. The balancing and placing of sample is very important in centrifuge. Always keep balance by placing centrifuge tube opposite to each other so that the load is distributed equally around the axis of rotor as shown in Fig. The sample tubes or centrifuge tube of 10, 50 and 100 ml can be used for centrifugation. The machine comes with changeable rotor according to the requirement of researcher (Fig. 4.12).

High Speed refrigerated Centrifuge: As name indicates this type of centrifuge work on high speed i.e. 25000 rpm generating RCF of about 60000 g. It is also having flexibility of using both types of rotor i.e. fixed rotor or swinging bucket type rotor. It is used for separating bacteria, cell organelles and precipitated proteins. It is not used for the sedimentation of viruses, and small organelles such as ribosome, for this purpose ultracentrifuge is required having maximum speed.

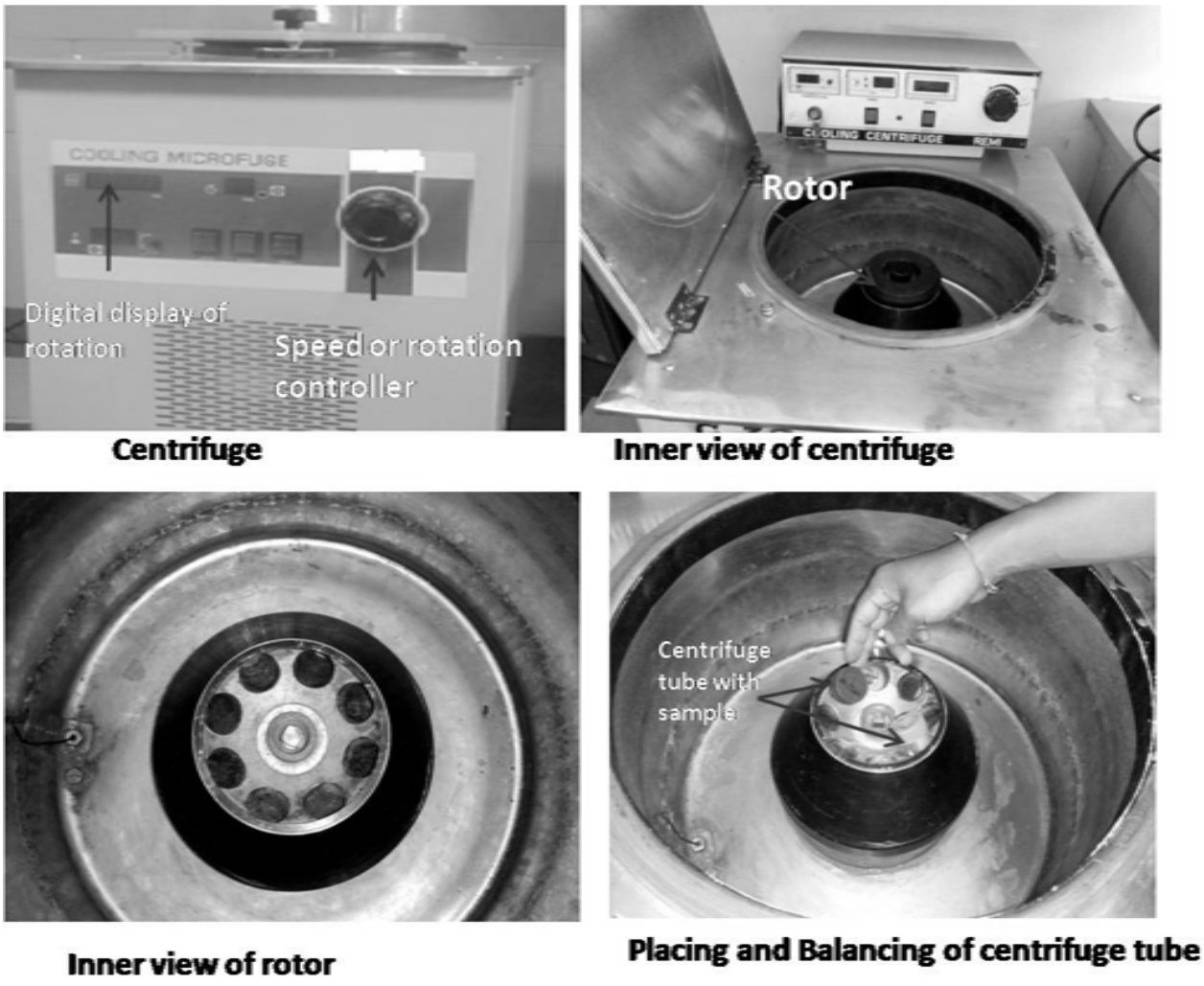


Figure 4. 12: Working and parts of Refrigerated centrifuge

Ultracentrifuge: Centrifuge which is used to sediment viruses and small organelles like ribosome have speed of 80000 rpm with a relative centrifugal field of upto 600000 g. it was developed by Svedberg in 1929 and demonstrated the subunit of proteins. Ultracentrifuges are of two types:

a. Preparatory ultracentrifuge:. Preparatory centrifuge is generally used for the centrifugation or separation of cellular organelles such as mitochondria, ribosome, microsome and viruses. It is also used for the gradient separation of solution containing increasing concentration of dense substance. For e.g. sucrose solution is used for the separation of cell organelles. Caesium salt is used for the separation of nucleic acid. In this centrifugation the sample is run at high speed and then rotor is allowed to come to smooth stop and gradient is taken out to isolate the separated component.
b. Analytical centrifuge (AUC): A typical analytical centrifuge generate centrifugal field of 2500000 g. Analytical centrifuge is used for the determination of the purity of macromolecule, determination of relative molecular mass of solute in their native state, for the examination changes in molecular mass of super molecular complex. It consists of optical detecting system to monitor the material during sedimentation for concentration distribution in the sample at any time during centrifugation using ultra violet light absorption and optical refractive index sensitive system. Measurement of sample concentration at wavelengths from 200 to 800 nm detection of macromolecules containing strong chromophores. It is used to analyse protein, polysaccharide, nucleic acid, drug, ligands, gases, organelles and viruses.

   When the machine switched on or rotor moves the image of analyte (for e.g cells or protein) are projected by an optical system on to film or computer. The concentration of the solution at various points in the cell is determined by absorption of light of the appropriate wavelength and them measuring the degree of blackening of photographic film. This will facilitate to observe the separation of sample concentration versus the axis of rotation due to applied centrifugal force. AUC are generally used for the two types of experiment

a. Sedimentation equilibrium experiment has aim to determine the total time course of sedimentation and report on molar mass and size distribution of dissolved macromolecule
b. Sedimentation velocity experiment is dealing with the final steady state of the experiment where sedimentation is balanced by diffusion opposing the concentration gradient

1.4.1- GEIGER MULLER COUNTERS (GM COUNTER)

The GM counter was named after its inventor Hans Geiger and Walther Muller. This instrument was first developed by Geiger in 1908 with Ernest Rutherford and name Geiger counter. At that time this device has limited used i.e. it can detect only alpha particle. Later on the research scholar of Geiger, Walther Muller improved this device for detecting more types of ionization radiation. Now a day this instrument is known as Geiger- Muller Counter which is used as particle detector that measure ionization radiation i.e. beta particle and alpha particles except gamma particle because it does not ionize the gas.

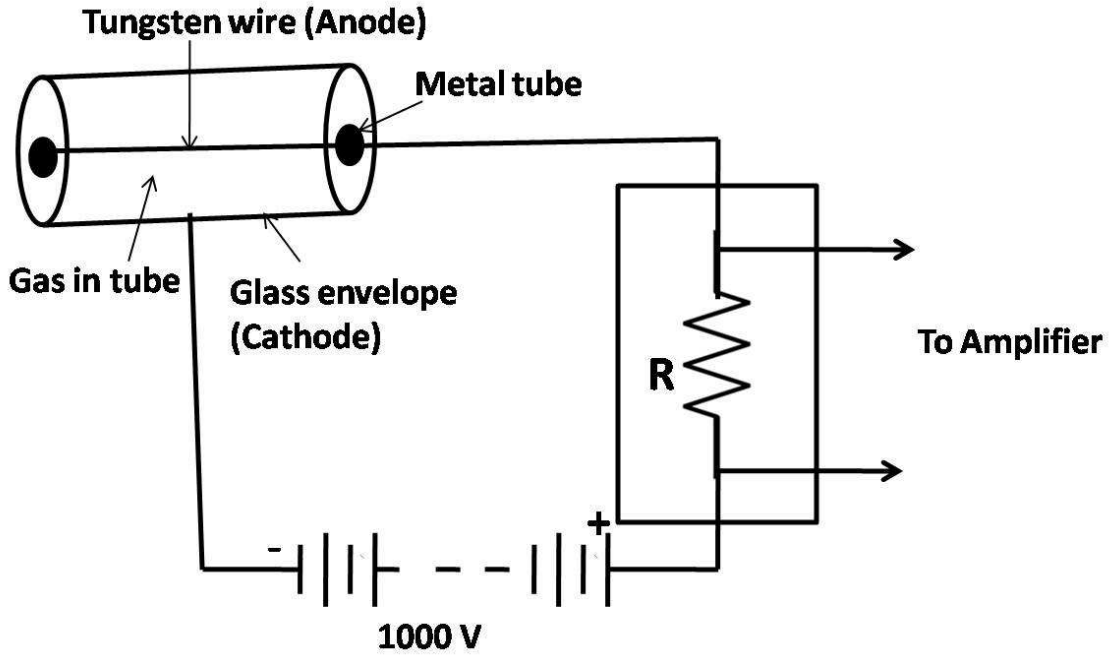


Figure 4.13. Working principle of G-M counter

GM counter consist of a metal tube covered with glass which acts as cathode and along the axis there is thin wire made up of tungsten which acts as anode. The tube is filled with inert gas such as Argon, Helium, Neon with halogen added at low pressure. There is window at one end where mica sheet is fitted, from this window radiation enter into the tube. A high potential difference of about 1000 V is applied between the electrodes through a high resistance R of about 100 mega ohm. (Fig.4.13)

Working of G-M counter

It is based on the principle of ionization as the charged particle passing through the gas medium present in the tube ionizes the atoms of gas by energy transfer.

High potential difference i.e. 1000 V is applied across the anode and cathode of the tube so that a high radial electric field nears the central wire is obtained (Fig.13). Due to the formation of high electrical field electrons generated by ionizing collisions between a high-speed particles entering the tube and the inert gas atoms are accelerated towards the anode wire by the strong electric field and acquire within a very short distance a high speed of their own. Because of this speed, they too can ionize other atoms and free more electrons. The object of the counter is to produce a single pulse for each particle entering the tube. This can only be achieved if spurious pulses due to secondary electrons released from the cathode surface by the bombardment of ions are completely suppressed so that the tube can recover as quickly as possible to be in a state when it is able to record the next entering particle. A quenching gas (halogen or organic vapours) introduced into the tube is to serve this purpose. The idea is to allow the inert gas ions on their way to the cathode to collide with the heavy molecules thereby transfer

Their charges to the molecules and become neutralized - a process known as quenching. The molecular ions thus produced move slowly to the cathode and on reaching there, capture electrons from the cathode surface to become neutral molecules. The multiplication of charges repeats itself in rapid succession producing within a very short interval of time an avalanche of electrons. The electron avalanche is concentrated near

the central wire while the positive ions, being much heavier, drift slowly toward the cathode. The number of electrons striking the wire can be measured to detect the presence of radioactive emission.

1.4.2- SCINTILLATION COUNTERS

Scintillation counter is important instrument generally used for the detection and measurement of radiation. The scintillation counter in which radioactive materials exposed to atoms within the detector that temporarily absorb the radiated energy. These excited atoms return to their unexcited state and emit photons that are detected by the scintillation counter.

Working of Scintillation Counter

In scintillation counter there is lining of phosphor in one end of the photomultiplier tube. Its inner surface is coated with photoemission called photocathode. It acts as negative terminal. There is presence of several numbers of electrodes called dynodes which are arranged in the tube at increasing positive potential. When charge particles reach and strike the photocathode in photomultiplier tube releasing an electron. These electrons accelerate toward the first dynode and strike it. More numbers of secondary electron are emitted which accelerates towards the second dynode so on. Finally the chain continues multiplying the affect of the first charged dynode. Due to this process there is release of a voltage pulse across the external resistance. This voltage pulse is amplified and electronic counter (Fig.4.14).

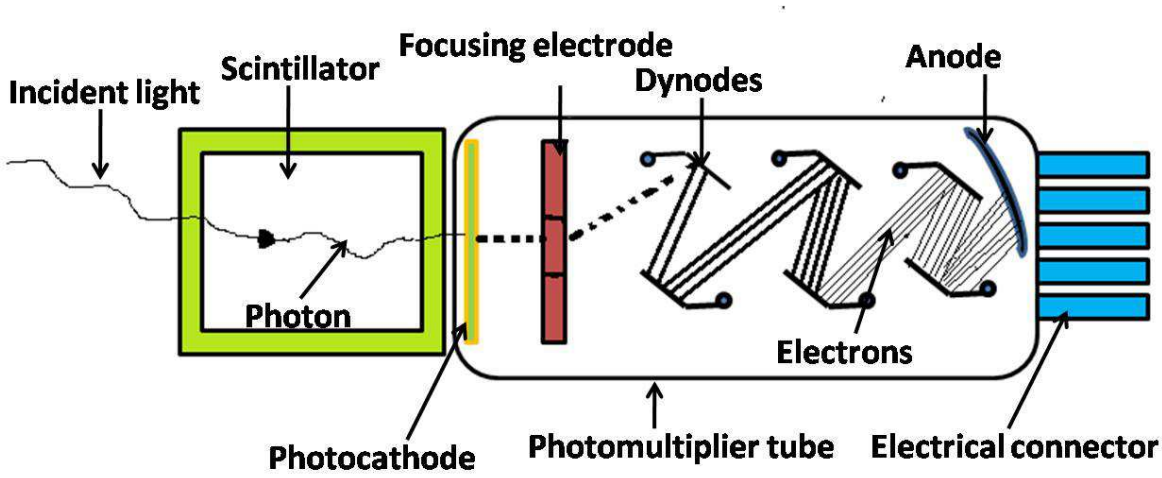


Fig.4.14 Scintillation Counter

Scintillation counter consist of two parts

1. Scintillator: Scintillation is a material may be solid, liquid (organic or inorganic) or gaseous which give luminescence or exhibit scintillation when struck by ionization radiation. The NaI crystal is the most widely used scintillation material which under UV light glows blue.
2. Photo multilplier Tube: PMT is generally used to creates strong electrical output from a weak signal developed by striking of photon on photocathode. This electron accelerates toward the first dynode and strikes it. More numbers of secondary electron are emitted which accelerates towards the second dynode so on. Lastly electron hits the anode and creates current which flow to ground through a resistor where it creates a voltage drop that can be counted.

4.5. SUMMARY

Modern laboratories of biological sciences should equipped with basic instrument like pH meter, UV-Vis spectrophotometer, centrifuge etc. pH meter is generally used to determine the pH of soil, water and culture medium used for the cultivation of fungi and bacteria. There is presence of electrode which very sensitive to detect the change in H ion concentration. The electric circuit measure the electromotive force developed across the electrode pair. Spectroscopy is the measurement of (spectrum of light) electromagnetic radiation, absorbed, scattered, or emitted by atoms, molecules, or other chemical species. Each chemical species has unique energy states; spectroscopy can be used to identify the interacting species separately. Centrifugation is based on the principle of centrifugal force in which liquid are subjected to high speed to separate solid from liquid or liquid from liquid depending on the density. In centrifugation heavy particle settled down and light particle will rises to the top. Geiger- Muller Counter which is used as particle detector that measure ionization radiation i.e. beta particle and alpha particles except gamma particle because it does not ionize the gas. Scintillation counter is important instrument generally used for the detection and measurement of radiation. The scintillation counter in which radioactive materials exposed to atoms within the detector that temporarily absorb the radiated energy.

1.6. GLOSSARY

Bandwidth: The frequency span where constant amplitude input will produce a meter reading within a specified limit (usually 3db). In controllers, the region around the set point where control occurs.

Controller: A device capable of receiving a signal from a process and regulating an input to that process in order to maintain a selected operating condition

Electromotive force (emf): An electrical potential difference which produces or tends to produce an electric current.

Double Beam: In a double beam spectrophotometer the beam from the light source is split in two. One beam illuminates the reference cell holder and the other illuminates the sample.

4.7. SELF ASSESSMENT QUESTIONS AND POSSIBLE ANSWERS

Multiple Choice Questions:

1. Which filament lamp is used for UV and visible light:

(a) Tungten (b) Argon

(c) Platinum (d) Lithium

2. G-M Counter is used to measure the

(a) RBC (b) WBC

(c) Intensity of the radioactive radiation (d) intensity of visible light

3. Centrifugation is based on the principle of

(a). Gravitational force (b) Centrifugal force

(c) Frictional force (d) Vander wall force

4 pH scales ranges from

(a) 0 to 6 (b) 0 to 7

(c) 0 to 10 (d) 0 to 14

4.8. REFERENCES

---

1. Bair, E.J. (1962). Introduction to chemical Instrumentation, McGraw hill, New York

2. R.C.Dubey and D.K. maheshwari.(2012). Text book of Microbiology. S. Chand & company.

4.9 TERMINAL QUESTIONS

Q1. Write principle, procedure and working of pH meter.

Q2. Draw labeled diagram of GM counter.

Q3. Write different types of centrifuge used in laboratory

Q4. Write short note on

a. Bench centrifuge

b. Cooling centrifuge

c. Ultra centrifuge

Q5. Describe the working and principle of Scintillation counter?

Q6. Explain Beer-Lambert law?

Q7. Write principle and working of UV-Vis spectrophotometer

Unit 5: MICROTOMY AND MICROSCOPY

CONTENT

5.1OBJECTIVES

In this chapter, students will able to learn how tissue or cells are fixed and stain to study the histopathology. The chapter also deals with the different types of microscope used for the study of the microscopic world.

5.2 INTRODUCTION

At the beginning of science, there was no sophisticated and advanced instrument available to explore the world of an invisible entity. It was made possible by the development of a microscope in which the microscopic cell, bacteria, fungi, and other microscopic things was seen and detailed structure studied. As the science progress, there was the development of Microbiology has been progressed and now we studied the ultramicroscopic microorganism such as viruses and bacteriophages. Due to the development of the microscope it is possible that we can differentiate between different types of cells. The importance of the microscope in the histopathological investigation cannot be ignored. The abnormal changes in the cell and tissue of human being can be easily seen with the help of the microscope. There are different types of microscope has been developed these days according to the need of researchers and scientists i.e. Compound microscope, Stereo microscope, phase contrast microscope, Bright field microscope, Dark field microscope, Transmission electron microscope, Scanning electron microscope.

The morphological changes in the cell or tissues can be visualized through microscope when it is correctly processed and stained. This chapter focuses on different types of microscope used in the laboratories for various investigations**.**

5.3MICROTOMY

The trimmed block of paraffin is then fixed on a block holder. This block holder is then secured to the microtome and oriented appropriately with respect to the knife. With each revolution of the microtome handle, the specimen moves through the blade and a section of desired thickness is produced. Each successive section adheres to the preceding are forming a continuous ribbon. The ribbon is cut into small piece and transferred onto clean albumenized slides. The ribbon is flattened by putting a few drops of water. These floating sections are stretched using a hot plate or warming table.

Staining

Next, the paraffin is removed with xylene or another appropriate solvent and the specimen is rehydrated. It is then stained, dehydrated, cleared (made transparent) with xylene, covered with a suitable mounting medium, and topped with a cover-slip. Various stains are available to the histologist. Hematoxylin and eosin (H&E) is a frequently used combination of stains. Haematoxylin imparts a purple colour to substances, but must be linked to a metallic salt called a mordant before it can function effectively. This combination, called a lake, carries positive charges and behaves as a basic (cationic) stain. The lake combines electro statically with negatively charged radicals such as phosphate groups of nucleoproteins. Substances that become colored by a basic stain are said to be basophilic. Methylene blue, toluidine blue, and basic fuchsine are basic stains. Unlike haematoxylin, these stains have molecules that carry a positive charge of their own and do not require a mordant. Acidic (anionic) stains carry a negative charge and colour cell or tissue components that bear positive charges. Eosin is an acid stain. It imparts orange or red colorto acidophilic substances. Other commonly used acid stains are orange G, phyloxine, and aniline blue.

In addition to the widely used H & E staining procedure, numerous other stain combinations and techniques are available. Some are especially useful for identifying certain tissue

elements. For example, trichrome procedures such as Mallory's and Masson's specifically stain collagenous fibers within connective tissue. Orcein and Weigert'sresorcinfuchsin are stains used to color elastic fibers, providing a means of distinguishing them from other fibrous elements. Reticular fibers and nervous tissue components such as neurons, myelin, and cells of the neuroglia can be stained by procedures employing the use of silver. There are also special histochemical and immune histochemical procedures that make possible the localization of various carbohydrates, lipids, and proteins found in tissue. Stains such as Wright's and-Giemsa's (Romanovsky stains) are available for differentiating the various cells found in blood and bone marrow.

Fig 5.1Microtomy

Fixation

As discussed in the preceding chapter, fixation of a sample is the first and most important step in microscopical examination. After being removed from an animal/plant, a tissue or organ is cut into pieces. These are placed in a fixative for suitable time at room temperature or at $4^{o}C$ in a refrigerator. Commonly used fixatives are 4-10% formalin, aqueous or alcoholic Bouin, acetic alcohol formalin, Comoy's fluid. The purpose of fixation is to preserve normal morphology of the tissue and prepare it for further processing.

Dehydration

After fixation, the next step is to remove the fixative. The aqueous fixative may be removed by washing the fixed tissue in running water. Alcoholic fixative is removed by washing the tissue in 70% alcohol several times.

Removal of fixative is followed by dehydration. Dehydration is performed by transferring the sample through a series of alcohols of increasing concentrations upto 100% alcohol.

Clearing

Dehydration is followed by clearing. Several clearing agents such as xylene, toluene and benzene can be used. These agents are miscible with alcohol as well as paraffin wax. This intermediate step is essential before infiltrating the dehydrated tissue with paraffin because alcohol and paraffin do not mix.

Embedding

Cleared tissue is embedded in a suitable embedding medium. Paraffin wax is the most preferred medium. Several plastics are also available nowadays and they offer better sections. Paraffin embedding is carried out in an oven at a temperature just above the melting point of the paraffin. When infiltration is complete, the specimen is transferred to an embedding mold of fresh paraffin which is allowed to harden. Then the mold is removed and excess paraffin is trimmed away.

5.4. TYPES OF MICROSCOPES

The type of microscope was designed as per the requirement and nature of the specimen. So, there exist huge variations in microscopes required as per the desired magnification and visualization. However; some most common types of microscopes include compound (having a combination of lenses) light microscopes, bright and dark field, phase contrast, fluorescence, and

electron microscopes. The modern optical science of microscopy is based on the designs of German physicist Ernst Abbé. The applicability of various microscopes depends upon the most important feature, the resolution power.

RESOLUTION OF MICROSCOPES

To ascertain the distinguished minimum distance between two closely located objects with the help of a lens is called resolution which can be measured using the Abbé equation:

$$d = \frac{0.5\,\lambda}{nSin\theta}$$

Where, λ is the wavelength of the light used to illuminate the sample and nSinθ is the numerical aperture (NA). Angular aperture (θ) as in figure 1 is ½ the angle of the cone of light (illuminating a specimen) entering the lens of an objective piece of the microscope.

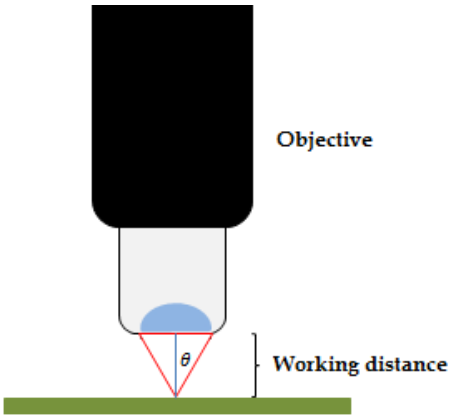


Figure 5.2Angular aperture θ of the cone formed on lens by specimen

The numerical aperture is nSinθ. NA depends upon the working distance which influences the angular aperture. For higher values of NA, working distance should be less for achieving higher angle of cone resulting in the higher resolution. It should be remembered that angle of the cone depends on the refractive index (n) of the medium and the optical material of lens. The refractive index of air is 1.0 and the maximum value of angle is 90^0. Sin90 is 1 therefore; lenses working in air cannot have a numerical aperture greater than 1.0. To achieve higher resolution greater angle of cone is required for which *immersion oil* (colorless) having a refractive index

Equal to that of medium can be used. It also helps to collect the otherwise dispersing light (from the edges of specimen) on the objective lens.

LIGHT MICROSCOPE

In the modern age of advanced scientific endeavor, sophisticated microbial examinations are being deliberated. All the modern age microscopes are compound and light microscope is the simplest compound microscope, which is very well understood by students of science. Several variants of light microscopes viz.Bright field, dark field, phase contrast, fluorescence and many others are available for microscopic examination. A simple compound light microscope has two sets of lenses i.e., an objective which form magnified image, which is further enlarged by another lens. All the light microscopes have a source of white light.

BRIGHT FIELD MICROSCOPES

This is an ordinary microscope, commonly used by students of preliminary science. The bright field microscope forms a dark image against a bright field. Two types of lens sets viz., objective (commonly 4 in number of different magnifying power) and eyepiece (ocular) are used (figure 2). The total magnification power of microscope using particular objective is calculated by multiplying the power of objective to the power of eyepiece for e.g., 45X objective and 10X ocular yield 450X magnification.

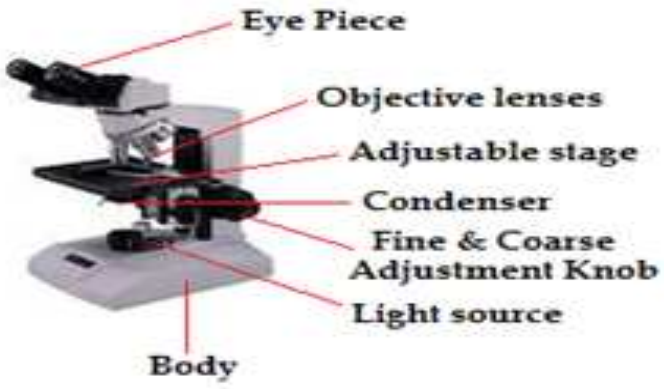


Figure 5.3. Bright Field

The objective lens forms an enlarged real image which is further magnified by eyepiece and the image is virtual formed 25 cm beyond the stage. Sub-stage condenser is also used some times for the purpose to increase the resolution. The specimen can be brought in focus of objective lens by fine and coarse adjustment knob (figure 2). It should be remembered that a microscopes are par focal i.e., even if the objective is changed the specimen remain in focus. To visualize by bright field microscopes, staining of the specimen is often required and it is very difficult to observe living cells through this microscope.

DARK FIELD MICROSCOPE

Dark field microscopes are used to magnify the images of living and unstained living microorganisms. In fact this is also technically and structurally similar as bright field microscopy, only difference is the addition of a dark field stopper between light source and condenser (figure 3). This light stopper (figure 4) let the light to pass from the open space and eventually condensers form a hollow cone of light and focused on the specimen to illuminate. The light that is reflected and refracted by the specimen forms an image, rest of the part around the Specimen appears black. Since the background of the specimen image is dark, the microscopy is called dark field.

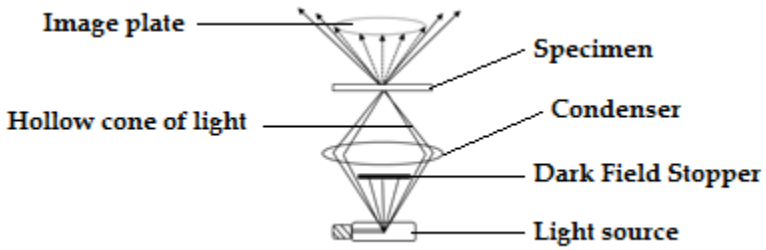


Figure 5.3. Optics of a Dark Field microscope

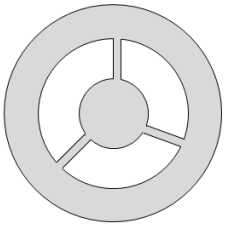


Figure 5.4. Dark Field Stopper

PHASE-CONTRAST MICROSCOPE

Phase-contrast microscopes produce images as a result of slight differences in refractive index and cellular density. The images so formed let the observer to have idea of the thickness of the organism under study. Like dark field, phase-contrast microscopes also have an annular stop disk with a slight difference of thin transparent ring which produces a hollow cone of light. As the cone passes through a living organism (bacterial cell or any other living cell) some light rays are

bent because of the density variations and refractive index. The light wavelength gets retarded by ¼. This deviation focuses to form an image of specimen. The un-deviated light strike a phase

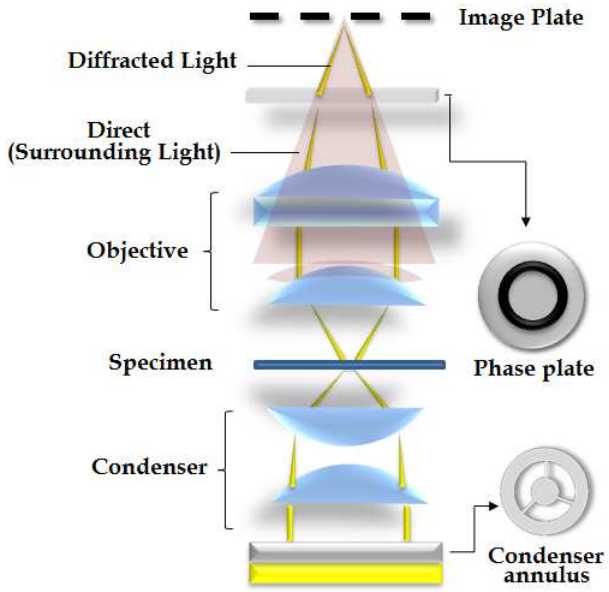


Figure 5. 5 Optics of a phase-contrast microscope

ring in a phase plate (figure 5), a disk placed in the objective. The deviated light (from the specimen) miss the phase plate and pass through the rest of the plate. The deviated and un-deviated lights have ½ and ¼ wavelengths of the light from the source. The un-deviated passing through the plate cancels each other during the formation of image.The un-deviated light render a bright background and deviated light form dark image giving a sharp contrast resulting in the peculiar image of specimen. This is dark-phase-contrast microscope.

FLUORESCENCE MICROSCOPE

Usually, light source is required to illuminate the specimen, which is the basis of image formation by microscopes. However; some objects emit light giving a base for fluorescence microscopy. In fact, some molecules reach to an excited state by absorbing radiant energy and re-emit the same energy in the form of light of lower wavelength than what was absorbed earlier, this is called fluorescence.

In fluorescence microscope, usually specimens are stained with certain dye molecules called *fluorochromes*such as acridine orange, Lucifer yellow, ethidium bromide, TOTO – 1(vital stain), DAPI (diamidino – 2 – phenylindole a DNA specific stain),which are excited by intense light produced by *mercuryvaporarc* lamp.

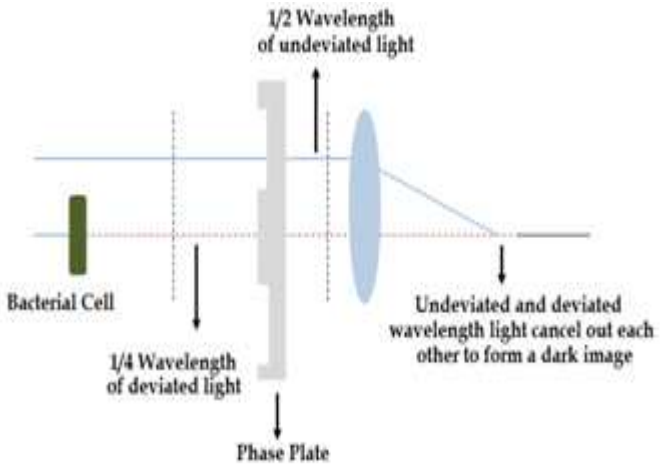


Figure 5.6. Wavelength deviations through phase plate in a phase contrast microscope

Since the light intensity is very high hence an infrared filter is usually used to avoid likely damages to the microscope itself. Like dark field microscopes, a dark field condenser renders a dark background to the image so that fluorescence coming out from the specimen could be seen. Light of specific wavelength excite the stained specimen, the image is formed from the light emitted. A barrier filter is also placed to remove the ultraviolet light that could otherwise damage the viewer's eye.

ELECTRON MICROSCOPY

In practical naked eyes can resolve only up to 10^{-3} m (thickness of hairs) while light microscopes can resolve up to 10^{-7}m, although the objects having lesser size exist and cannot be visualized by light microscopy. In the year 1931 German physicist Ernst Ruska and electrical engineering Max Knoll constructed the first prototype of electron microscope. Modern age electron microscopes have magnification power 10,000,000X and resolution up to 50 Pico-meters because electron

beam having wavelength about 100,000 times shorter than visible light is used to illuminate the specimen. Two basic type viz., transmission and scanning electron microscopes are commonly used these days.

TRANSMISSION ELECTRON MICROSCOPE (TEM)

In the transmission electron microscopes, an electron beam is generated by a heated tungsten filament positioned in the electron gun. This electron beam is focused on the specimen, which scatters the electron as per the density. Since the electrons cannot pass through the glass lens therefore magnetic lenses are used to focus. The entire set of magnetic lenses and specimen are worked under high vacuum as the air inside the column may get charged and moisture along with other molecules may scatter the electron beam. The electrons passing through the specimen are scattered, the dense part of the specimen scatter more as compared to the thin region. Thus the image formed by the scattered electrons will be dark indicating the density and the brighter region indicate the electron transparent regions (thin regions).

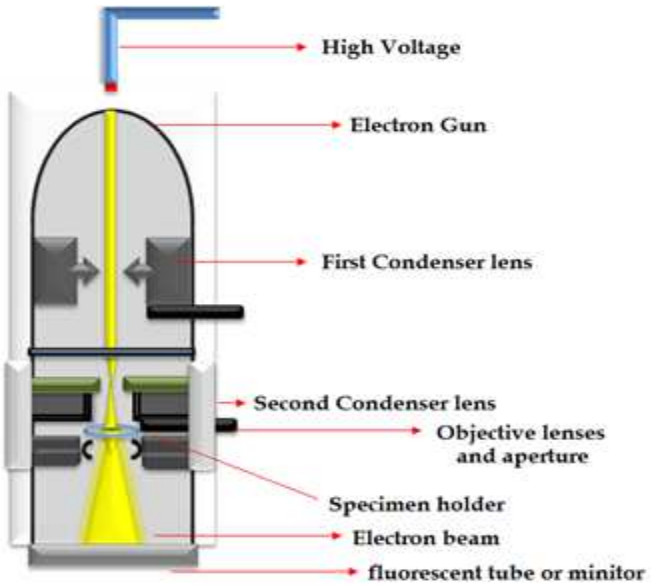


Fig 5.7 Transmission electron microscopes

The scattered electrons form enlarged image on the fluorescent screen and on the photographic plates, which can be used for reporting purposes.

Scanning Electron Microscope (SEM)

Scanning electron microscopes are the recent advancement in electron microscopy facilitating ease in the microbiological and other life science related examinations. Principally same as TEM except having advancements such as specimen preparation is easy, more well defined images, 3D direct viewing and photography and ease of working. Images in SEM are formed by the secondary electron produced as a result of shower of primary condensed electron beam under high vacuum. Figure 7 present the optics of a SEM. An Electron beam is produced by an electron gun so positioned that primary electrons are focused on the shielded specimen (metal coated). As soon as electron beam strikes the surface secondary electrons are produced (more from elevated and fewer from depressed region of the specimen surface), which are detected by a detector as lighter (indicating elevation because of more secondary electron) and darker (indicating depression because of few secondary electron production) regions. The electronic signals are sent to the photomultiplier which is further amplified. A cathode ray tube receives signals from the photo-multiplier and forms a picture. The image can be seen just as a computer monitor or television and can be photographed.

SPECIMEN PREPARATION FOR ELECTRON MICROSCOPY

TEM

Since electrons are rapidly absorbed and scattered by solid matter, therefore, extremely finely cut slices of 20 to 100 nm irrespective of material can only be observed by TEM. The slicing of the object intended to be observed through TEM need technical practice whereby specimen are fixed with the fixer chemical for e.g., osmium tetra oxide followed by rigorous dehydration with organic solvents such as ethanol, acetone etc. Usually, paraffin or any other liquid epoxy plastic are used to soak the specimen and then solidified to form a block. The thin slices are usually cut using ultra microtome's. It should be remembered that prior to use electron microscopes, complete dehydration of the specimen is a must.

SEM

The specimen preparation for SEM is easy however; it involves some of the basic requirement as that of TEM for e.g., dehydration and preservation. In scanning electron microscopes the entire procedure occurs at high vacuum condition, this may also damage the specimen and since secondary electrons are required for image formation, specimen are shielded with a thin layer of metal which prevent the damage and also helps emission of secondary electron from the specimen surface.

5.5 SUMMARY

The secret of the microscopic world cannot be deciphered until the microscope has not been discovered. The microscopy is a very useful instrument for the study of materials and can be used to gain valuable information about a large variety of specimens such as bacteria, fungi, protozoa etc. Some knowledge of the material and the information that is required is essential to determine the best techniques to employ when preparing and examining specimens. In Microscopy sample preparation is the most important of microscopy, as this determines the quality of the images produced. Disadvantage of the optical microscope is its resolution. This limitation is overcome by a scanning electron microscope (SEM).In addition, for 'transparent' specimens; in particular those required for motility, polarized light microscopy is best method for this type sample.

5.6. GLOSSARY

Analyzer: A analyzer is used with a polarizer to provide polarizing light.

Achromatic Lens: A lens that helps to correct the misalignment of light that occurs when it is refracted through a prism or lens.

Abbe Condenser: A lens that is specially designed to mount under the stage and which typically moves in a vertical direction.

Eyepiece: The eyepiece is the lens nearest to your eye.

Compound Microscope:Originally used to describe a microscope with more than one objective lens, a compound microscope is now generally understood to be a high power microscope with multiple, selectable objective lens of varied magnifications.

Condenser:A lens that concentrates the light on a specimen and increases the resolution. Found in or below the stage on compound microscopes.

Cover Slip:A thin, square piece of glass or plastic placed over the specimen on a microscope slide. It flattens out liquid samples and helps single plane focusing.

Electron Microscope:A type of microscope that uses electrons rather than light to create an image of the target.

Darkfield Microscopy: A technique used to enhance the contrast in unstained specimens. It works on the principle of illuminating the sample with light that will not be collected by the objective lens, so not form part of the image.

Monocular Head: A *microscope* head having a single eyepiece lens.

Nosepiece: The part of the *microscope* that holds the objective lenses also called a revolving nosepiece or turret.

Numerical Aperture (N.A.): This is a number that expresses the ability of a lens to resolve fine detail in an object being observed.

5.7. SELF ASSESSMENT QUESTIONS AND POSSIBLE ANSWERS

Multiple Choice Questions:

1. Numerical aperature is:

(a) nSin$\boldsymbol{\theta}$ (b) nTan$\boldsymbol{\theta}$

(c) nSec$\boldsymbol{\theta}$ (d) nCos$\boldsymbol{\theta}$

2. In fluorescence microscope, usually specimens are stained with certain dye molecules

(a) Simple stain (b) Crystal voilet

(c) Alberts Stain (d) Acridine orange

3. Transmission Electron Microscope magnify upto

(a). 100X (b) 1000X

(c) 400,000X (d) 400X

4 3D image of specimen was obtained with

(a) Compound Microscope (b) SEM

(c) TEM (d) Phase contrast Microscope

5. Image can be seen in SEM

(a) Fluorescent Microscope (b) Anode

(c) As Graph (d) Phosphorescent screen

5.8. REFERENCES

1.. R.C.Dubey and D.K. maheshwari.(2012). Text book of Microbiology. S. Chand & company.

2. P.K.Bajpai. (2006). Biological instrumentation and Methodology. S.Chand and Company.

3. Bruce Alberts. (2014). Molecular Biology of the Cells.W. W. Norton & Company; Sixth edition

5.9 TERMINAL QUESTIONS

Q1. Write principle, procedure, and working of Light Microscope.

Q2. Draw labeled diagram of a compound microscope.

Q3. Write principle and working of Fluorescent Microscope

Q4. Write short note on

d. Resolution

e. Numerical apertures

f. Magnification

Q5. Describe the working and principle of the Transmission Electron Microscope?

Q6. Write principle and working of Scanning Electron Microscope?

Unit 6: Separation techniques and cryopreservation

CONTENT

6.1. OBJECTIVE

In this chapter you will learn different techniques which are related to purification and separation of desired molecules and develop a basic understanding of Electrophoresis theory and also become familiar with the procedure involved in agrose gel electrophoresis to visualize DNA.

This chapter also deals with the preservation of cells, organelle, tissue, organs and microbes to very low temperature i.e. cryopreservation.

6.2. INTRODUCTION

As the biological sciences progress there is need of some sophisticated, accurate and sensitive method used for the purification and separation of desired component from the complex or mixture of two or more chemicals. The chromatography techniques have two parts i.e. mobile phase and stationary phase. In a chromatography sample mixture is placed in liquid or gas called mobile phase and the mobile phase carries the sample through a solid support called the stationary phase. In case of Thin Layer Chromatography (TLC) the stationary phase is silica coated on glass plate. These techniques have disadvantage such as it does not have long stationary phase so the length of separation is limited and another problem with this method it operates in open system so the chromatogram may influenced by temperature and humidity. Now days, there is some modern methods which is based on the principle of TLC such as HPLC, GC, HPTLC and Electrophoresis. HPLC is one of the important instruments used in the pharmaceutical industry for the analysis of drug, plant metabolite, and toxins and in medical science it is also use for the detection of hazardous chemicals in blood. Electrophoresis is most commonly used technique in molecular biology for the analysis of nucleic acid and protein. It is based on method of separating molecules on

the basis of charge and molecular weight when applied electrical field. On the basis of molecular weight of nucleic acid we can predict the nucleic acid of unknown microorganism.

6.3CHROMATOGRAPHY

6.3.1CHROMATOGRAPHY

Chromatography is another chemical procedure to find out and to separate the contents of a mixture of two or more chemicals. The literal meaning of chromatography is the study of colors first developed and described by Russian scientist Mikhail Tswett in 1900 which was further developed by Martin and Synge in early 1950s. In fact the chromatography is a separation technique based on the "partition or distribution coefficient or more appropriately partition constant or partition ratio (K_d) that describe the way in which a compound distribute itself between two immiscible phases".

K_d = concentration of compound in phase A / concentration of compound in phase B

The term effective distribution coefficient is defined as the total amount, as distinct from the concentration, of substance present in one phase divided by the total amount present in the other phase. It is in fact the distribution coefficient multiplied by the ratio of the volumes of the two phases. If the K_d of a compound between two phases e.g., A & B; is 1 and if the compound distribution is of 10 cm^3 and 1 cm^3 respectively, the concentration is the two phases will be same however; the total amount of the compound in phase A will be 10 times the amount in phase B.

There are two basic and essential phases' viz., stationary and mobile in chromatography. Stationary phase remain stable and do not move at all and provide base matrix for eg., solid, gel, liquid solid / liquid mixtures etc. on which compounds are separated. Mobile phase are either liquid or gaseous which flows over or through the stationary phase. If base matrix support is polar (e.g., paper, silica etc.) it is forward phase or if the matrix is non-polar (C-18) it is reverse

phase chromatography. The choice of stationary and mobile phase is made so as to have different distribution coefficient for the mixture sample. This is also achieved by setting up following:

1) Adsorption equilibrium between stationary and mobile phase
2) Partition equilibrium
3) Ion exchange equilibrium
4) An equilibrium between a liquid phases trapped inside the pores of a stationary porous structure and a mobile liquid phase as in permeation or molecular exclusion chromatography.
5) Equilibrium between a stationary immobilized ligand and a mobile liquid phase.

Chromatography may be preparative (separating the components of mixture or in advanced terms "purification") or analytical (measuring the relative proportion of analyte in a mixture).Furthermore, there exist two basic types of chromatographic viz., column chromatography and planar chromatographic techniques.

6.3.2. COLUMN CHROMATOGRAPHY

Column chromatography is a separation technique in which the stationary bed is within a tube (glass, plastic or metal). The stationary phase comprised of various polar or non-polar porous materials. The filled column is packed by a liquid mobile phase carrying the mixture to be separated through the column. There are two types of column used: 1) packed column which is characterized by the complete filling of the volume of column (figure 1), 2) open tubular column characterized as stationary phase concentrated along the column wall giving an unrestricted path to the mobile phase.

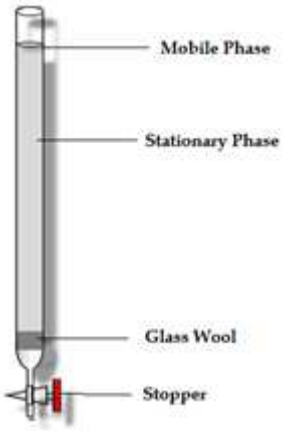
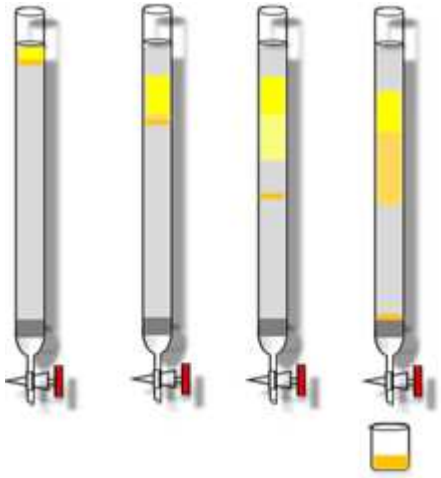


Fig: Packet Column

The separation is done in the column shown in the figure normally to separate the components of the mixture as for e.g., in the plant extract and not to quantify them. The packed column is gently handled following by the repeated washing with the solvent used as mobile phase. Sample is poured in to the column carefully and packed with the liquid mobile phase as in figure 2. The components of sample mixtures immediately get separated depending upon their partition coefficient (stationary to mobile) in the column forming band of different color having negligible differences therefore extra care is needed to separate each component. Separated components are called as *eluent*. The washing of column with the solvent to remove the components is called *elution*.



Separation of component from mixture sample through column chromatography

GAS CHROMATOGRAPHY

Gas chromatography is a column chromatography where the mobile phase is a carrier gas, usually inert gasses such as helium and/or nitrogen. As shown in figure 3, a GC typically consists of a gas tank providing inert carrier gas, the flow of which is controlled by a flow controller connected with the column oven. The column oven is a chamber is a metal casing wherein a column made of glass or metal is connected at one side with the carrier gas tubing and sample injection chamber and to the detector at the other end.

The column oven provides temperature for the fugitive diffusion of gas inside the column containing microscopic layer either of polymer or liquid as solid bed. The separated compounds are detected by the detector. Detectors may be of various types whereas; flame ionization and electron capture detectors are commonly used in most of the researches of life sciences and environmental sciences.

COLUMN OF GC

The heart of a gas chromatography is the column performance of which sets limit to the separation attainable and determines time of analysis. Two main type of columns viz., 1) packed column, first introduced by Martin in 1952 and 2) capillary packed column introduced by Halasz in 1968 are used these days. Pressure factors such as column inlet pressure, required speed of sample introduced, required sample size (detection limit) are largely determined by properties of the selected type of column.

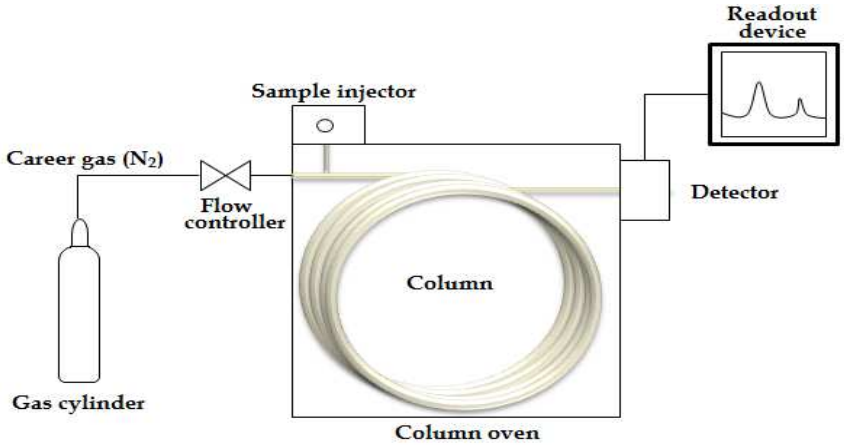


Figure 6.1. Schematic of a gas chromatograph

Factors determining gas chromatographic performance are directly linked to the column efficiency. These factors are:

(a) Analysis time: The residence time of the peak maxima in a column, the retention time t_R can be calculated by

$$\boldsymbol{t_R = t_0\,(r + k)}$$

Where t_0 is gas hold up time, r is the retention time of an unreacted component such as air, k is the capacity ratio. k can be derived from the partition coefficient K_d by multiplying the volumetric ration of stationary to mobile phase in the column.

$$k = K_d\,(\frac{V_{liquid}}{V_{gas}})$$

(b) Efficiency: Efficiency **is** expressed as the plates (n), the degree of band broadening a solute undergoes in a given column. This can be derived from the chromatogram:

$$n = \frac{t_{R^2}}{\sigma 2}$$

Where σ is the standard deviation (time units) is equal to half width at 0.607 of the height of a Gaussian peak.

The plate number (n) or the peak (bands) is related to the length of column and height H equivalent of a theoretical plate. H = L/n

H is further dependent on the linear velocity (*u)* of the carrier gas as in Figure 4.

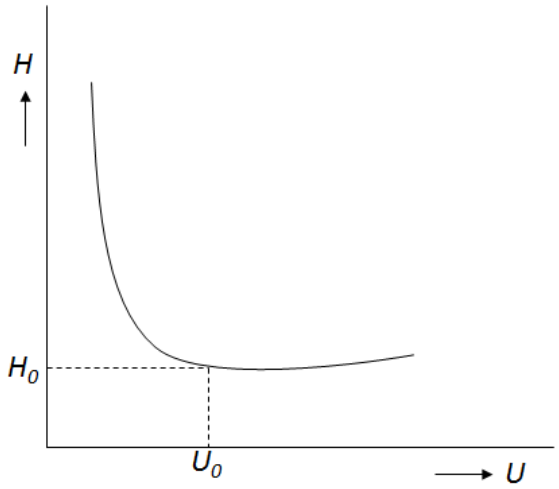


Figure 6.2 Dependency of Plate Height (H) and linear velocity (U) of carrier gas

(c) Resolution: (R) express the degree of separation of two components leaving the column shortly after each other

$$R = \frac{\left(t_{R_2} - t_{R_2}\right)}{\sigma_1}$$

Taking the values of t_R:

$$R = \frac{(\propto -1)k}{(k+I)} \sqrt{3}$$

Where $\propto$ is relative of two components

The separation is almost complete at R = 6.0 and just acceptable for R = 4.0.

(d) Speed of sample: The capacity of the column, which determines the most practical parameter, the sample size. Quantity of sample can be introduced in a GC either as a narrow band of high concentration or as a plug of correspondingly lower concentration.

If the sample is fed into the column as a band of high concentration the K_d will remain concentration dependent and each part of solute band will move at different speed resulting in the asymmetric peaks. If the sample is fed as plug of lower concentration, it goes under dilution with carrier gas as soon as it enters the column. This causes additional symmetric band

broadening. The variance σ^2 (sec^2) of an eluting peak may be considered to be composed of the variance of the sample at the inlet σ_I^2 and the variance σ_0^2 due to column processes.

$$\sigma^2 = \sigma_1^2 + \sigma_0^2$$

Best resolution can be obtained at $\sigma_I^2 = 0$ (infinitely narrow sample band).

Carrier gas selection

The selection of the carrier gas depends upon the type of analyte, column temperature and detectors. For e.g., helium is used in discharge ionization detectors (DID), nitrogen is used in electron capture detector (ECD) and in flame ionization detector (FID) (in FID hydrogen is used as fuel gas). In general helium, nitrogen, argon, hydrogen and air are used as carrier gases in chromatography.

Detectors

All though advancements have been made in the detector technology and currently number of detectors viz., ECD, FID, DID (as above), PID (photo-ionization detector), PDD (pulse discharge ionization detector), MSD (mass selective detector) etc. are available for the analysis of variety of environmental and medical samples. However; ECD and FID are the most employed detectors worldwide discussed here in details.

Electron Capture Detector (ECD)

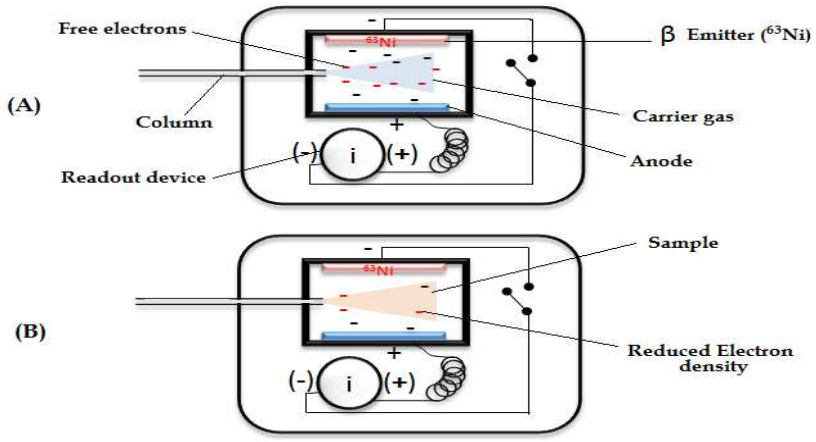


Figure 6.3. Schematic of Electron capture detector (A) high electron density in carrier gas (B) reduced electron density in sample + carrier gas

Electron capture detector (ECD) is a device used to analyze halogenated, nitriles and nitro compounds and is particularly important for the environmental, pharmaceutical, and forensic studies. ECD was first invented by James Lovelock in 1957. The functioning of ECD is based on a radioactive beta particle (electron) emitter ^{63}Ni (~ 10 millicurie or 370 mBq) which remains in a metal foil holding inside a chamber. ^{63}Ni emits electron that move towards anode so as to make the circuit complete resulting in to the generation of a current. Nitrogen is commonly used for analysis by ECD because of its low excitation energy enabling more electron density in the detector chamber generating current of high magnitude. When the sample analyte enters along with the carrier gas, the molecules of analyte absorb / capture some of the electrons and reduce the current. while a background signal of current generated by the carrier gas is also distinctly visualize in a chromatogram. *The reduction in the current due to the analyte is directly proportional to the analyte concentration.*

The flame ionization detector (FID) is the industry standard method of measuring hydrocarbons (HC) concentrations. The sample gas is introduced in to a hydrogen flame inside the flame chamber. Any hydrocarbon in the sample will produce ions when they are burnt. Ions are detected using a metal collector (high voltage ion collector) which is based on a high DC voltage. The current across this collector is thus proportional to the rate of ionization which in turn depends upon the concentration of HC in the sample gas. Remember FID use hydrogen as fuel gas (Fig.6). If hydrogen flow is on and column is connected to the detector inlet fitting, hydrogen gas can flow in to the column oven and create an explosion hazard. Hydrogen flow inside the FID should be standardized and must not exceed 30 mL/min, while the carrier gas (N_2) flow should be 25 mL/min.

The analytical efficiency of a GC depends upon the appropriate column choice. There is variety of basic types of column (as discussed earlier in this section) present in the market out of which wall coasted open tubular (WCOT) and porous layer open tubular (PLOT) columns are widely used. PLOT columns are of three viz., 1) molecular sieve; 2) Divinylbenzene (DVB) and 3)

Alumina (Al_2O_3) types are quite common for wide range of hydrocarbon analysis. The accuracy in the sample analysis by a GC has been developed to an advanced level, and gas chromatography mass spectrophotometer (GCMS) is one of them. The detection limit increases to ppt (parts per trillion) level. Since GCMS is more sophisticated and an expensive device, extra care is required while analyzing the samples.

HIGH PERFORMANCE LIQUID CHROMATOGRAPHY (HPLC)

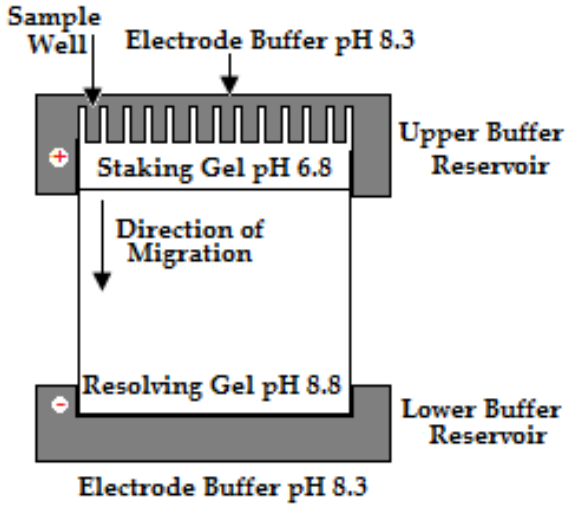
Gas chromatography required the volatilization of the sample, however; HPLC is the method to identify and quantify the analyte that cannot be converted into the gas phase. The unique retention time of the analyte in to the column and its characteristic peak as calibrated with standard is the principle of HPLC analysis. The time for a substance to pass through the column, termed the retention time which is related to the identity of the compound. Quantitative information is obtained from the area or height of the peak produced by the detector.

By the choice of appropriate solute (mobile phase) usually methanol and acetonitrile and the column (packed solid material) efficient separation can be achieved. In an HPLC the sample (20µl) is injected which is then carried by the solutes. The separation of analyte depends upon the interaction of the analyte and mobile phase with the stationary phase. High pressure are required to force a liquid through a tightly packed column filled with small particle material and the availability of high pressure solvent delivery systems is directly responsible for the "high performance".

The detectors used for the identification and quantification in HPLC are complex, in fact no single universal detector system has yet been developed. More than one detector having unique detection property can be used in a single HPLC in series. The most commonly employed detectors include *bulk property detector* (compare over all change in physical property of the mobile phase with and without an eluting solute); *solute property detectors* (respond to a physical property of the solute that is not exhibited by the pure mobile phase. These detectors are 100 times sensitive and detect sample on a nano-gram level; and *ultra-violet photometers* (most commonly used detector).

6.4. ELECTROPHORESIS

6.4.1 ELECTROPHORESIS



Electrophoresis is an analytical method frequently used in molecular biology and life sciences. It is applied for the separation and characterization of proteins, nucleic acids and subcellular-sized particles like viruses and small organelles. The working principle of electrophoresis is that the charged particles migrate towards their corresponding electrode in an applied electrical field. If conducted in solution, samples are separated according to their *surface net charge density*. The most frequent applications, however, use gels (polyacrylamide, agarose) as a support medium. The presence of such a matrix adds a sieving effect so that particles can be characterized by both charge and size. Protein electrophoresis is often performed in the presence of a charged detergent like sodium dodecyl sulfate (SDS) which usually equalizes the surface charge and, therefore, allows for the determination of protein sizes on a single gel. Additives are not necessary for nucleic acids which have a similar surface charge irrespective of their size. The resistance entrusted by the electrophoretic units also play significant role in producing migration related variations. The resistance of an electrophoresis unit depends on its size, gel thickness, amount of buffer, buffer conductivity and temperature. This resistance will normally decrease in time due to a slowly increasing temperature. Electrophoresis units which have a resistance below the minimum load resistance of a power supply will trigger an alarm.

For a comprehensive study and to achieve better separations, one should have knowledge of basics of electrophoresis. This section deals with such information.

Isoelectric point (pI)

The isoelectric point is the pH at which a particular molecule or surface carries no net electrical charge. Biological amphoteric molecules such as proteins contain both acidic and basic functional groups. Amino acids which make up proteins may be positive, negative, neutral or polar in nature, and together give a protein its overall charge. At a pH below their pI, proteins carry a net positive charge; above their pI they carry a net negative charge. Proteins can thus be separated according to their isoelectric point (overall charge) on a polyacrylamide gel using a technique called isoelectric focusing.

Isoelectric focusing (IEF)

Isoelectric focusing is an electrophoretic method in which proteins are separated on the basis of their pIs. It makes use of the property of proteins that their net charges are determined by the pH of their local environments. Proteins are positively charged in solutions at pH values below their pI and negatively charged above their isoelectric points. Thus at pH values below the pI of a particular protein, it will migrate towards the cathode during electrophoresis. While moving towards the electrode (cathode or anode) proteins lose some protons and their net charge drops and pH changes, this retard the protein movement, finally it stops to move as the pH become equals to the pI. Protein diffuses in to a pH higher than its pI, the protein will become negatively charged and will be driven towards the anode. In this way, proteins condense or focus, into sharp bands in the pH gradient at their individual pI values. Narrow pH range and high applied voltage gives excellent resolution in IEF.

Buffers of Electrophoresis

A buffer should be chosen with a pKa that is very close to the desired pH, preferably within a half point. The buffer will have the greatest capacity both to absorb and/or release protons with the acid and the base form well represented in solution. It should be noted that pKa is not constant for all conditions but is a function of the total ionic strength and the temperature, so the stoichiometry should be modeled after actual running conditions. In general, TAE (tris acetic acid with EDTA) and TBE (tris borate with EDTA) buffers are used in electrophoresis technique. TAE buffer provides optimal resolution of fragments >4 kb in length, while for 0.1 to

3 kb fragments, TBE buffer should be selected. TBE has both a higher buffering capacity and lower conductivity than TAE and therefore should be used for high-voltage electrophoresis. Additionally, TBE buffer generates less heat than TAE at an equivalent voltage and does not allow a significant pH drift.

Gel Concentration

Agarose is a polysaccharide extracted from seaweed. It is typically used at concentrations of 0.5 to 3%. The higher the agarose concentration the "stiffer" the gel. Agarose gels are extremely easy to prepare. It is also non-toxic.Agarose gels have a large range of separation, but relatively low resolving power.Polyacrylamide is a cross-linked polymer of acrylamide. The length of the polymer chains is dictated by the concentration of acrylamide used, which is typically between 3.5 and 20%. Polyacrylamide gels are significantly more complex to prepare than agarose gels. Because oxygen inhibits the polymerization process, they must be poured between glass plates. Polyacrylamide gels have a rather small range of separation, but very high resolving power.

6.5. CRYOPRESERVATION

Cryopreservation or Liquid Nitrogen

Cryopreservation is the method in which very low temperature is used to preserve living cell, organs, microorganism, sperm etc for longer period. Polge *et al* in 1949 was credited for cryopreservation of the first mammalian cell i.e. spermatozoa. This technique is devised to keep the cells genetically stable, viable and metabolic inert. The disadvantage of low temperature preserving cell, tissue, organ and microorganism was formation of ice crystal which eventually disrupts the cell membrane resulting the death of cell. The main objective of cryopreservation is to replace water with other material that will not form ice crystals to overcome the freezing there is addition of antifreeze agent such as Glycerol and DMSO (Dimethyl Sulphoxide).

Cryopreservation media generally consists of a base medium, protein source, and a cryopreservative. The cry0preservative both protects the cells from mechanical and physical

stress and reduces the water content within the cells, thus minimizing the formation of cell-lysing ice crystals. The protein source, often fetal bovine serum (FBS), also protects the cells from the stress associated with the freeze-thaw process. Cells are frozen slowly, at 1C/minute, using programmable coolers or by techniques outlined below. Generally, the optimum cell density to freeze per 1mL of cell suspension depends on the type of cell. Mammalian cells are usually frozen between 10^6 cells/mL to 10^7 cells/mL. The cryopreservation media may differ slightly for adherent and suspension cell types.

The following is procedure for cryopreservation of cells

1. Expand culture to allow for adequate cell density for the desired volume to freeze. The cell culture to be cryopreserved must be in the log phase of the growth cycle (approximately 2-4 days after subculturing). Determine the cell count for number of viable cells and the total cell concentration.
2. Centrifuge cells at approximately 200 to 400 x g for 10 minutes, allowing the cells to form a pellet. During centrifugation, determine the amount of freezing media to prepare. For example, if using 1 mL cryovials, divide the total cell concentration by the desired cell density. *Example: A $4x10^7$ cell suspension will yield a total of ten 1 mL alliquots at $4x10^6$ cells per alliquot.Prepare 10 mL of freezing medium to easily suspend the pellet at the correct cell density.*
3. Prepare the necessary volume of cryopreservation media (determined above) using the following guidelines:
4. Resuspend cell pellet(s) using the cryopreservation media, triturating to ensure a single-cell suspension with as few cell clumps as possible.
5. Dispense into the desired number of vials for cryopreservation.
6. Immediately transfer the vials to a freezer with a minimum temperature of -20 ^oC for one hour.
7. Transfer the vials to a -80^oC freezer for 24 hours. Alternatively, a dry ice/methanol slurry using or other insulated box with a cover or lid may be used if a -80^oC freezer is not available.
8. After 24 hours at -80^oC, cells may be transferred to a liquid nitrogen storage (-196 ^oC)

6.6. SUMMARY

As the technology and techniques are developing day by day, which made the analysis and detection faster than earlier conventional method used. This chapter deals with some modern technique which is now used in research laboratories. Chromatography is chemical procedure to find out and to separate the contents of a mixture of two or more chemicals. Column chromatography is a separation technique in which the stationary bed is within a tube (glass, plastic or metal). The stationary phase comprised of various polar or non-polar porous materials. The filled column is packed by a liquid mobile phase carrying the mixture to be separated through the column. Gas chromatography is a column chromatography where the mobile phase is a carrier gas, usually inert gasses such as helium and/or nitrogen. As shown in figure 4.2, a GC typically consists of a gas tank providing inert carrier gas, the flow of which is controlled by a flow controller connected with the column oven. Gas chromatography required the volatilization of the sample, however; HPLC is the method to identify and quantify the analyte that cannot be converted into the gas phase. The unique retention time of the analyte in to the column and its characteristic peak as calibrated with standard is the principle of HPLC analysis. Electrophoresis is an analytical method frequently used in molecular biology and life sciences. It is applied for the separation and characterization of proteins, nucleic acids and subcellular-sized particles like viruses and small organelles. Cryopreservation is the method in which very low temperature is used to preserve living cell, organs, microorganism, sperm etc. for longer period.

6.7. GLOSSARY

Absorption: In chromatography, absorption signifies the process by which a solute partitions into a liquid-like stationary phase.

Mobile Phase: The eluate moving through the column. In gas chromatography (GC) this will be a gas, and in liquid chromatography (LC) a liquid.

Stationary Phase: The substance that remains in one place in the column. In GC this will be a liquid of high-viscosity, which clings to the inner walls of the column; in LC it will be some sort of packing, either solid or gel-based.

Eluate: The mobile phase exiting a column.

Eluent: The mobile phase entering a column**.**

Elution: The passage of the mobile phase through the column to transport solutes.

Flow Rate: The amount of mobile phase that has passed through the column per unit time. The units are millilitres per second (mL/sec) or, more commonly, millilitres per minute (mL/min).

Isoelectric focusing (IEF): Electrophoresis technique that separates proteins according to their isoelectric point (pI)

Isoelectric point (pI): pH value at which a molecule carries no electrical charge, or at which the negative and positive charges are equal.

PAGE: Polyacrylamide gel electrophoresis, a common method of separating proteins

Polyacrylamide gel electrophoresis (PAGE): Electrophoresis technique that uses polyacrylamide as the separation medium

Rf value: Relative distance a protein has traveled compared to the distance traveled by the ion front. This value is used to compare proteins in different lanes and even in different gels. It can be used with standards to generate standard curves, from which the molecular weight or isoelectric point of an unknown may be determined.

Adsorption: The process of retention in which the interactions between the solute and the surface of an adsorbent dominate.

Retention Time: The elapsed time between sample injection and the appearance of the chromatographic peak apex

Isocratic: Chromatographic conditions in which a constant composition eluent is used.

Ion Chromatography: An ion-exchange technique in which low concentrations of organic and inorganic anions or cations are determined using ion-exchangers of low ion-exchange capacity with dilute buffers.

Degassing: The practice of removing dissolved gases from the eluent. It can be achieved by helium sparging, applying vacuum to the eluent, ultrasonification or heating.

Isoelectric Point: The pH point at which a molecule no longer has a net charge.

Effective distribution coefficient: It is defined as the total amount, as distinct from the concentration, of substance present in one phase divided by the total amount present in the other phase.

SDS-PAGE: Separation of molecules by molecular weight in a polyacrylamide gel matrix in the presence of a denaturing detergent, sodium dodecyl sulfate (SDS). SDS denatures polypeptides and binds to proteins at a constant charge-to-mass-ratio. In a sieving polyacrylamide gel, the rate at which the resulting SDS-coated proteins migrate in the gel is relative only to their size and not to their charge or shape.

Running buffer: Buffer that provides the ions for the electrical current in an electrophoresis run. It may also contain denaturing agents. The running buffer provides the trailing ions in discontinuous electrophoresis.

Sample buffer: Buffer in which a sample is suspended prior to loading onto a gel. SDS-PAGE sample buffer typically contains denaturing agents (including reducing agents and SDS), tracking dye, and glycerol.

6.8. SELF ASSESMENT QUESTION

Multiple Choice Questions:

1. The Stationary phase in Reverse phase chromatography is made of
   A. Polar B. Non Polar
   C.Both D. None of the above

2. TLC is based on the principle of
   A. Electrical mobility B. Partition chromatography
   C.Ion exchange D. Adsorption

3. In isocratic method in HPLC the composition of solvent is
   A. Remains Variable B. Remains constant
   C.both A and B D. None of the above

4. In gradient method in HPLC the composition of solvent is
   B. Remains Variable B. Remains constant
   C.both A and B D. None of the above

5. HPLC stands for
   A. High Performance Liquid Chromatography B. High Peak Liquid Chromatography
   .C.High Pressure Liquid Chromatography D. All of the above

6. Antifreeze agent is

A. Glycerol B. DMSO

C.Both A and B D. Methanol

6.9. REFERENCES

1. Harold M. McNair, James M. Miller.2009 Basic Gas Chromatography, 2nd Edition, John Wiley.
2. Elsa Lundanes, Leon Reubsaet, Tyge Greibrokk.2013. Chromatography: Basic Principles, Sample Preparations and Related Methods. John Wiley
3. G. Lunn and N. Schmuff, John Wiley & Sons, 1997, HPLC Methods for Pharmaceutical Analysis.

6.10. TERMINAL QUESTIONS

Q1. Describe the principle and procedure of Thin Layer Chromatography.

Q.2. Write short notes on following

a. Flame Ionisation detectoe

b. Electron capture detector

Q.3. Write the principle, procedure and working of HPLC with suitable diagram.

Q.4. Write the principle, procedure and working of Gas Chromatography with suitable diagram.

Q.5. Write the principle, procedure and working of Electrophoresis.

Q.6. Write short notes on:
A. Isoelectric point
B. Isoelectric focusing (IEF)
C. Cryopreservation