Department of Distance Education

Punjabi University, Patiala

Class : Master of Library and Information Science

Semester : 2

Paper : MLIS 205 **Unit : II**

(Research Methodology and Statistical Techniques)

Medium : English

Lesson No.

Department website : www.pbidde.org

LESSON NO. 2.1 | **AUTHOR : DR. AMRITPAL KAUR**

EXPERIMENTAL METHOD

(Definition; Steps, Experimental Validity - internal and external, Limitations, Applicability in the Field of Library and Information Science)

Structure

7.0 Objectives

This lesson will make the learner familiar with the experimental method and its applicability in the field of library and information science.

7.1 Introduction

The experimental method is oriented to future, new programmes, new policies and new procedure. It has proved to be an essential and powerful tool in leading man

towards progress. It is the problem solving approach that attempts to follow more closely some of the canons of research in physical and biological sciences. Experimentation, whether conducted in the laboratory, class room or other field is an attempt to control all essential factors except a single variable, which is manipulated or changed with a view to determine and measure the effect of its operation. An experiment is designed to test hypotheses, theories, principles; identify casual relationship between variables; determine solution to problems and establish some kind of truth.

7.2 Definition

In the words of Beveridge, "An experiment usually consist in making an event occur under known conditions where as many extraneous influences as possible are eliminated and close observation is possible so that relationships between phenomena can be revealed.

Greenwood suggests, "an experiment is the proof of a hypothesis which seeks to hook up two factors into a casual relationship through the study of contrasting situations which have been controlled on all factors except the one of interest, the later being either the hypothetical cause or the hypothetical effect."

According to Fastinger, "The essence of an experiment may be described as observing the effect on a dependent variable of the manipulation of an independent variable."

The basic factor in case of experimental method is thus, the control over the subject of study and manipulation of the independent variable to study its effect upon the dependent variables.

7.3 Steps in Experimental Method

The major steps in experimental method are as under:

7.3.1 Selecting and determining the problem
7.3.2 Review of related literature
7.3.3 Specification of variables
7.3.4 Formulation of research hypotheses
7.3.5 Planning the experiment
7.3.6 Conducting the experiment
7.3.7 Analysing and interpreting the outcomes
7.3.8 Drawing up the results
7.3.9 Reporting the results

7.3.1 Selecting and determining the problem

First of all a problem is identified and it is determined whether or not the experimental approach is appropriate for the solution of the problem.

7.3.2 Review of related Literature

A review of the literature related to similar problem is done.

7.3.3 Specifications of variables

It consists of specifying

(a) Independent variables
(b) Dependent variables
(c) Potential intervening variables

7.3.4 Formulation of research hypotheses

Formulation of research hypothesis helps the researcher to decide which type of data to be collected & what design of the experiment to be followed to test the hypotheses.

7.3.5 Planning the experiment

This includes

(a) Determining the method of experimentation
(b) Place of experiment
(c) Duration of the experiment
(d) Determining the materials of the experiment
(e) Conducting pilot study
(f) Selecting subjects or groups

7.3.6 Conducting the experiment

It involves collection of data by means of predetermined measures; introduction of experimental and control conditions. While the experiment is going on, periodic verification is done to determine whether or not the integrity of experiment is being maintained. After post-test measures have been implemented, the experiment comes to an end.

7.3.7 Analysing and interpreting the outcomes

The outcomes are analysed and interpreted to test the hypotheses with the help of statistical techniques.

7.3.8 Drawing up the results

Care must be taken to restrict the results to the conditions actually present in the experiment. The results of the study must be restricted to the population actually investigated and care must be taken not to over generalize the results.

7.3.9 Reporting the results

The study should be reported in such a way that the reader can make a judgment as to its adequacy.

7.4 Experimental Validity - Internal and External

Once an investigator has conducted an experiment, the results of the experiment - the experimental data must be evaluated. Evaluation involves a number of concepts, including internal as well as external validity.

The internal validity is checked by answering to the following questions:

(a) Are variables sufficiently controlled?
(b) Has randomization been employed throughout?

External validity is associated with questions concerning the extent to which a sample is representative of the target population. Can the results of an experiment be generalised to the real world? It is normally easier to justify such generalisation for field experiments than for laboratory experiments because of the artificiality of the latter. Overall experimental findings are evaluated in terms of the reliability of the data, the scientific importance of the results and the extent to which the data can be generalised.

7.5 Problems and Limitations

Experimental method is not free from problems and limitations especially when

it is applied in social sciences. These may be mentioned as follows:

7.5.1 Difficulty of Co-operation

Human beings who form the subject matter of social experiment are not easy to be manipulated. They work according to their free will. Their willing cooperation with the experimenter is necessary. Generally it is believed that co-operation may be forth coming if the parties concerned find their interest involved. This is the condition which is not always presented and lack of interest makes active co-operation difficult.

7.5.2 Difficulty of Setting

In this method, setting has to be created and it is not always possible in social research. Many difficulties arise in setting the experiment. If this experiment has to be conducted in the natural setting, there is a problem of the place in this wide world. Even if we have located the place, it is not always possible that owner of the setting may permit the experiment to be conducted.

7.5.3 Difficulty of Control

For successful experiment it is necessary that all other causative factors except the one under study must be strictly unchanged and the particular variables may be varied gradually. This is very difficult in social sciences where it is mainly created by complexity of social phenomena, its dynamic nature and independence of human behaviour.

7.5.4 Errors of Measurement:

It is not possible in practical life to measure values of the variables used for estimation without error.

7.6 Application of Experimental Method to Library & Information Science

In the field of library & information science, experiments can be used to test new techniques for developing, maintaining and utilizing library collections, to identify ill-defined or previously unobserved library or information phenomena and to explore conditions under which certain phenomena in library and information science occur.

A subject of an experiment is the basic unit on which an experiment is performed. In library science, the subjects are frequently persons - for example, patrons, librarians or students. The fact that human beings are frequently the subjects of social science experimentation implies a basic difficulty in applying the experimental approach to social science research in general and librarianship in particular. Humans cannot be manipulated as easily as white rats, for moral, ethical and legal reasons. Hence many problems in library & information science are not amenable to the experimental approach.

In addition to subjects, a true experiment design involves a control group on which experiments are not performed. The purpose of the control group is to serve as a comparison. Treatments are applied to the experimental group but not to the control group, and the results are observed. In medical research, the administering of drugs to patients (subjects) or the use of a surgical technique are examples of treatments. In librarianship, a treatment that is being tested might be a system of indexing, a mode of instruction, a type of catalogue organization, or a method of book selection, among others. But the complete equivalence between experimental and control groups can seldom be obtained in research in library and information science. Because with human subjects there are

always numerous experimental and hereditary factors that cannot normally be known. Thus, at best, equivalence can only be an approximate concept.

Theoretically, an ideal experiment may be defined as one in which all factors that might affect the outcome of the experiment (the dependent variable) are controlled by the experimentor. This ideal situation can be approximated in the natural sciences; however it is extremely difficult even to approximate in the social sciences, because of the complexity of the subject matter. Not only is the identification of all possible variables impossible, but even if it were not, the problem of how to control them would remain. The concept of randomization is employed as an attempt to overcome these difficulties. In essence, randomization involves the random assignment of subjects from the target population to treatment groups. By assigning subjects the experimental and control groups at random, the experimentor reduces the possibility that the results will be due to conditions other than those associated with the manipulation of the independent variables. Moreover, random assignment is a requirement of statistical tests of significance of experimental results.

The extent to which an experiment accomplishes its goal is sometimes referred to as its internal validity. In other words, the investigator must pose and answer the following question: Does the experiment adequately test the treatment of interest or is it methodologically unsound? Internal validity implies freedom from bias and assurance that the observed effect on the dependent variable is actually due to the independent variable (the treatment). Thus the internal validity of an experiment involves the concepts of randomization and control.

In library and information science, it is difficult to devise and conduct an experiment due to the reason that the situation keeps on changing. Study of human behaviour (users and staff) is an area of main concern. But human behaviour is not only unpredictable but also uncontrollable. Thus it is not possible to have experiments in library and information science with the same precision and accuracy as in natural sciences. In library and information science user is the crucial factor. In dealing with complex human beings, it is unlikely that all variables can be successfully controlled, leading to results which are less precise. We may take the example of 'use of library catalogue by reference staff'. Use is going to be affected by variables such as competence in using the catalogue, enthusiasm of the staff members etc.

A review of library and information science literature reveals that experimental approach is the least used one. There is a need for using the experimental approach for better results in library and information science as this approach is considered the most useful one.

7.7 Summary

This lesson discusses experimental method in detail - its definition, steps, experimental validity, limitations and applicability in the field of library and information science.

7.8 Key Concepts

Causative : Acting as the cause of something

Equivalence : A thing, amount, word etc. that is equivalent to something else
Randomization : To use a method in an experiment, a piece of research etc. that gives every item an equal chance of being considered
Variable : A situation, number or quantity that can vary.

7.9 References

1. Busha, H. Charles and Harter, Stephen P. Research Methods in Librarianship. New York: Academic Press, 1980.
2. Goldhor, Herbert. An Introduction to Scientific Research in Librarianship. Urbana: University of Illinois, 1972.
3. Goode, William J. and Hatt, Paul K. Methods in Social Research. New York: McGraw Hill, 1985.
4. Kothari, C.R. Research Methodology: Methods and Techniques, New delhi: Anmol Publications, 1985.
5. Krishan Kumar. Research Methods in Library and Information Science. New Delhi: Har-Anand, 1992.

7.10 Self-Check Exercise

Note: Write the answers in the space given below each question and check your answers with the answers given at the end.

1. What is an experimental method of research?

_____
_____
_____
_____
_____

2. Discuss experimental validity in brief.

_____
_____
_____
_____
_____

3. Mention the areas of library and information science in which experiments can be conducted.

_____
_____
_____
_____
_____

7.11 Answers to self - check exercise

1. Experimental method is the problem solving approach that attempts to follow more closely some of the canons of research in physical and biological sciences. Experimentation, whether conducted in the laboratory or field is an attempt to

control all essential factors except a single variable, which is manipulated or changed with a view to determine and measure the effect of its operation.

2. To evaluate the results of the experiment, internal as well as external validity is checked. The internal validity implies freedom from bias and assurance that the observed effects on the dependent variable is actually due to the independent variable. External validity is associated with questions concerning the extent to which a sample is representative of the target population.Overall experimental findings are evaluated in terms of the results and the extent to which the data can be generalised.
3. In the field of library and information science, experiments can be used to test new technique for developing, maintaining and utilizing library collections; to identify ill-defined or previously unobserved library or information phenomena and to explore conditions under which certain phenomena in library and information science occur.

LESSON NO. 2.2 | **AUTHOR : DR. AMRITPAL KAUR**

COMPARATIVE METHOD

(Definition, Steps, Limitations, Applicability in the Field of Library and Information Science)

Structure

8.0 Objectives

This lesson will make the learner familiar with the comparative method and its applicability in the field of library & information science.

8.1 Introduction

Comparative method, as the name suggests is based on comparison. It consists of the sociologists assembling data about several societies or social contexts and comparing them with each other with the intention of explaining the cause of any variation. It forms the basis of modern scientific procedure. It attempts to support the thesis of organic development according to the fixed stage. It is applied, generally, to the theory of biological evolution.

This method has been called the natural experiment or sometimes the quasi experiment. The sociologist collects information about different societies or social contexts as they are found in the real world and identifies the similarities and differences between

them. In other words, instead of setting up situations either in the laboratory or in the field, the researcher studies what is already going on.

This method is based on the theory of famous French Sociologist Auguste Comte that the society as a whole should be studied in comparison to one another. In order to study the characteristics of the man and the human society and various social characteristics, he made certain suggestions about the methodology. He produced his 'law of three stages' i.e. primitive, developing and developed of societal development which maintained that all societies past, present and future necessary passed through the same stages of development. But Comte's methodology lacked the use of scientific hypothesis and laid greater stress on philosophical basis of human development. It was only in 19th century that Comte's comparative study and Darwin's evolutionary approach joined together and gave scientific basis to the method. The credit for making the comparative method really scientific and useful for social research goes to French Sociologist Emile Durkhelm. He not only put forward its utility but gave a scientific look to its form. He presented this method in his famous book 'The Rules of Sociological Method'.

8.2 Definition

J. Gould and W. L. Kolb in Dictionary of the Social Sciences define comparative method as:

"Comparative method is a general term denoting the procedures which by clarifying the resemblances and differences displayed by phenomena deemed, on various criteria, to be comparable aim at eliciting and classifying (a) casual factors in the emergence of such phenomena (b) patterns of inter-relation both within and between such phenomena."

The comparative method is a method of logical analysis. It is based on three of Mill's methods of inductive inquiry:

(a) The method of agreement
(b) The method of difference
(c) The direct method of difference

These methods use all available and pertinent data concerning the preconditions of a particular phenomena to determine its causes by examining the similarities and differences between relevant instances.

8.3 Problems that are studied through Comparative Method

Comparative method cannot be said to be useful only to a particular type of problem. It can be adopted to the study of various types of problems. Normally it is applied to the following types of problems:

(a) Present or synchronic problems
(b) Historical or diachronic problems

The comparative method is applied to various branches of social sciences such as Anthropology, Education, Sociology, Library & Information Science etc. Through this method, various aspects of the problem whether the problem is of the present day or past are studied.

8.4 Approaches adopted in Caomparative Method of Social Research :

In it generally the following approaches are adopted :

8.4.1 Statistical and descriptive approach

8.4.2 Historical approach

8.4.3 Philosophical approach

8.4.4 Sociological and anthropological approach

8.4.5 Problem approach

8.4.6 Area study approach

8.4.1 Statistical and Descriptive Approach

In this form of approach, on one hand comparative study is made by describing the characteristics of different methods or comparing characteristics on the basis of statistics. This method is not considered useful.

8.4.2 Historical Approach

In this method, on the basis of post, comparative study of different problems is made. This approach is useful for the study of historical problems.

8.4.3 Philosophical Approach

In this approach, the education or social set up of a country is studied in the background of; it philosophy, values and customs. This approach is possible to study the problems which has one religion or nation but in a multi-national, multi - religious country, this approach is not possible because more than one value influence various social problems.

8.4.4 Sociological and Anthropological Approach

According to this approach, the educational set up and other social institutions are compared on the basis of various factors, culture or cultural factors. It is very useful method of study.

8.4.5 Problem Approach

According to this approach, there are two type of laws in every society : sociological and narrative. It is possible to find out sociological laws through scientific study but narrative laws can be studied only through philosophical approach. This method was used for the first time by Dr. Brain Holems for the study of the social education problems. Alongwith two types of laws, he also laid stress on natural geographical laws in the comparative study.

8.4.6 Area Study Approach :

In it, attempt is made to study the education and other social problems of a particular area on the basis of various factors and then the same approach is applied in other area. After collecting data, comparison is being done between the data of different areas. This method demands good knowledge about the language and other conditions of those areas.

8.5 Steps of Comparative Method

The steps of comparative method are as follows:

8.5.1 Selection of the problem and area of study

8.5.2 Selection of the units of comparison

8.5.3 Collection of data, tabulation and formulation of hypotheses

8.5.4 Comparison of the facts

8.5.5 Verification of hypotheses

8.5.6 Preparation of report and inviting comments on it

8.5.1 Selection of the problem and area of study

Firstly, selection of the problem and the area under which the study is to be made has to be done. The selection of the problem is governed by the resources and limitations of the collection of data. Selection of area include a small geographical area or a big area like country.

8.5.2 Selection of the units of the comparisons

After selection of the problem and the area, the units on the basis of which the comparison is to made has to be selected. It has to be found out whether the comparison is to be made on the basis of certain themes, complexes or patterns of certain culture or the entire culture is to be treated as a unit.

8.5.3 Collection of data, tabulation and formulation of hypothesis

Once the units of comparison have been selected, the collection of data and its processing has to be done. The collection and processing of data lead to formulation of hypothesis. The process can be other way also. A hypothesis may be formulated and on the basis of the hypothesis collection of data and its processing may be done. Collection of data may be done through primary as well as secondary sources. Proper notes should also be prepared. Once the data is collected, it should be classified on the basis of characteristics and the qualities but before the processing or the tabulation is done, the reliability of the data should be ascertained.

8.5.4 Comparison of the facts

Once the facts have been collected about the different areas they should be compared. The comparison should be done on the basis of hypothesis.

8.5.5 Verification of hypothesis

The hypothesis is verified on the basis of the data collected. It should be found out as to what extent the hypothesis has been found to be correct.

8.5.6 Preparation of the report and inviting comments on it

After the completion of the above steps, the process of writing and preparation of the report should be started. The report should be discussed in detail with the comparison of facts and the verification of hypothesis. The report should be made attractive with the help of diagrams, tables etc.

Even when the report is prepared, it cannot be said that it is perfectly correct. Many of the weaknesses are left over in the report. In order to make the report correct, it should be laid open to the scholars for comments. If the comments are useful and authentic the report should be modified accordingly.

8.6 Limitations

There are certain limitations of comparative method. These are:

I The comparative method appears to be an easy method of study but actually it is very difficult. No conclusion can be drawn very easily from the data and details collected by this method. Varied conclusions can be drawn in a study of one and

the same group, community and circumstances and if these are different, the results are bound to differ.

II The success of this method depends on the formulation of correct hypothesis. Collection of data and their processing shall not be of much use if the hypothesis is not correct.

III This method involves scientific knowledge, capacity to collect the data in correct manner, knowledge of the principles and the area of study etc. Very few persons can possess this capacity and so this method cannot be said to be very much in vogue.

8.7 Application of Comparative Method to Library & Information Science

In the field of library & information science, some insightful investigations of comparative nature have been undertaken. These studies have encompassed comparisons of the practices and study of library science in nations and regions of the world. They have focused on such topics as education for librarianship, political control of libraries, library resources and services, publishing in relation to libraries, patterns of communications, cross cultural influences in librarianship and other related topics.

J. Periam Danton, a pioneer in the area of comparative library science studies defines comparative librarianship as follows:

"the analysis of libraries, library system, some aspects of librarianship, or library problems in two or more national, cultural, or societal environments in terms of sociopolitical, economic, cultural, ideological, and historical contexts. This analysis is for the purpose of understanding similarities and differences, and for determining explanations of the differences, with the ultimate aim of trying to arrive at valid generalizations and principles."

According to Louis Shores, two research approaches can be taken in comparative librarianship - geographic comparisons and subject comparisons. In his view comparative librarianship can uncover "neglected approaches to important technical library problems and can suggest a critical role for librarianship."

Sources of information for comparative studies in library science include the following: direct observation, historical records, statistical compilations, government documents, library science professional publications, and a variety of other recorded information that can be used to make qualitative and quantitative comparisons.

Comparative librarianship is a descriptive research approach; thus several methods of investigation can be employed, including case studies, library surveys, survey research, and in some instances historical studies of recent past developments. In 1974 Danton indicated that relatively a few studies of a truly comparative nature had been undertaken in the field of librarianship. In characterizing a genuine comparative study, Danton noted that it should be cross societal, compare observed phenomena, explain differences, draw conclusions and establish principles.

In the mid 1960s interest in international and comparative librarianship led to the establishment of an International Library Center in U.S.A. The purpose of this centre is to support the study of international and comparative librarianship through the

provision of information, training and research opportunities. This centre also serves as a clearing house for data about libraries, documentation and book production in the United States and in numerous foreign countries.

Among the most useful publications containing information and guidelines for the conduct of comparative studies in the field are: John F. Harvey's Comparative and International Library Science; J. Periam Danton's Dimensions of Comparative Librarianship; Miler M. Jackson's Comparative and International Librarianship and Simrova and Maker's A Handbook of Comparative Librarianship.

To conclude we can say that comparative method is applicable to those problems in library and information science which go beyond description.

8.8 Summary

This lesson discusses the comparative method in detail - its definition, steps, limitations and applicability in the field of Library & Information Science.

8.9 Key Concepts

Diachronic	:	Refers to be problems existing at different points of times
Hypothesis	:	An idea or explanation of something that is based on a few known facts but that has not yet been proved to be true or correct
Synchronic	:	Refers to the problems existing at one point of time

8.10 References

1. Busha, H. Charles and Harter, Stephen P. Research Methods in Librarianship. New York: Academic Press, 1980.
2. Danton, J. Periam. Dimensions of Comparative Librarianship. Chicago: ALA, 1973.
3. Dorothy, G. Collings. Comparative Librarianship. In Encyclopaedia of Library and Information Science. 5 (1971): 492.
4. Foskett, D.J. Review of the Dimensions of Comparative Librarianship by J. Periam Danton. Library Quarterly. 44 (July 1974): 265.
5. Niessen, Marfred et. al. International Comparative Research. New York: Oxford, 1984.

8.11 Self-Check Exercise

Note: Write the answers in the space given below each question and check your answers with the answers given at the end.

1. Name the three stages of societal development given by Auguste Comte?

_____
_____
_____
_____
_____

2. Define comparative method.

_____
_____
_____
_____

3. State the problems that are studied through comparative method.

4. Name the steps of comparative method.

8.12 Answers to Self - Check Exercise

1. 1. Primitive
   2. Developing
   3. Developed
2. Comparative method is a general term denoting the procedures which by clarifying the resemblances and differences displayed by phenomena deemed, on various criteria, to be comparable aim at eliciting and classifying (a) casual factors in the emergence of such phenomena (b) patterns of inter relation both within and between such phenomena.
3. Comparative method is applied to the following types of problems
   (a) Present or synchronic problems
   (b) Historical or diachronic problems
4. 1. Selection of the problem and area of study
   2. Selection of the units of comparison
   3. Collection of data, tabulation and formulation of hypotheses
   4. Comparison of the facts
   5. Verification of hypotheses
   6. Preparation of report and inviting comments on it

LESSON NO.: 2.3 | **DR. AMRITPAL KAUR**

STATISTICAL TECHNIQUES

(Definition, Steps, Limitations; Applicability in the Field of Library and Information Science)

Structure

9.0 Objectives

This lesson will make the learner familiar with the statistical techniques and their application in the field of library and information science.

9.1 Introduction

Statistical techniques are a mechanical process especially designed to facilitate the condensation and analysis of the large body of quantitative data. The aim of statistical techniques is to facilitate comparison, study relationships between the two phenomena and to interpret the complicated data for the purpose of analysis.

Statistical techniques are applicable to a subject where enquiry and investigation are essential to arrive at the truth. Scientifically anlaysed data are the foundation of sound policy formulation. The use of statistical techniques can be helpful in enlarging human experience and knowledge. These are required for handling the multitudes of numerical results in such a way that the significant relations between properties can be studied.

9.2 Definition

The term statistics (singular) refers to a body of methods and techniques for analyzing numerical facts or data. It is a shortened form of statistical methodology.

The term statistics (plural) refers to a collection of numerical facts or data, such as number of books added in a library in 2005, number of members in a library etc.

In the present lesson, statistics has been used and defined in the singular.

According to Encyclopedia Americana, "Statistics refers to science and art of obtaining and analysing quantitative data with a view to make sound inference in the face of uncertainty."

According to Encyclopedia Britannica, "Statistics is a mathematical discipline concerned with the study of mass of quantitative data of any kind."

According to Kendall, "Statistics is the branch of scientific method which deals with data obtained by counting or measuring properties of population or natural phenomenon."

9.3 Steps/ Stages in Statistical Investigation

Statistical investigation involves the following five steps:

9.3.1 Collection of data

9.3.2 Organisation of collected data

9.3.3 Presentation of organised data

9.3.4 Analysis of data

9.3.5 Interpretation of data

9.3.1 Collection of data

In order to have a reasonable standard of accuracy, data must be collected in a systematic manner. The data should be reliable, purposeful and adequate. The data are observed, counted and analysed in a systematic manner with reference to the problem.

Data is of two kinds i.e. primary and secondary. Primary data is the one which the investigator originates for the purpose of the inquiry being pursued. Secondary data refers to those data which has already been collected by some one else.

The methods for collection of primary data are: (i) Questionnaire (ii) Interview Schedule (iii) Direct personal interview (iv) Indirect oral investigation (v) Information from correspondents. For secondary data, both published and unpublished sources are used.

The techniques employed for the collection of original data are the following:

(i) Census method

(ii) Sampling method

In a census method, the data is collected from every unit of the population, which may constitute the subject matter of the study. In the sampling method only a selected number of units of population or a part of the population are studied. From the result for the sample, inferences are drawn for the whole population.

There are three types of samples:

(i) Random or probability sample

(ii) Purposive sample

(iii) Mixed sample

A random sample is the one where each item in the universe has an equal chance of being included in the sample. This reduces bias. In case the population are homogenous, with regard to characteristics of our interest then random sample is expected to produce satisfactory results. Lottery method, tables of random number and systematic sampling are methods of obtaining a random sample.

In purposive sampling, the choice of the sample depends exclusively upon the discretion of the investigator, using some purposive principle. The investigator, using his judgment, can include only those items in the sample which he thinks are the most typical of the universe with regard to the characteristics being studied by him. In this method, some criterion of selection is decided upon and the items are selected accordingly. It is also known as judgment sampling.

Mixed sampling is a mixture of random sampling and purposive sampling. The following methods are used for getting mixed samples:

- Stratified sampling
- Multistage sampling
- Variable sample fraction

9.3.2 Organization of Collected Data

Organization of data involves two related processes i.e. classification of data and tabulation of data. Classification of data is the process of arranging data with some purpose with respect to some attributes or variables. Tabulation of data refers to orderly arrangement of data in the form of columns and rows. A table summarizes data in a simple and convenient form.

9.3.3 Presentation of Data

After classification and tabulation, the data is presented either with the help of diagrams or graphs. Diagrams and graphs are attractive and provide an overall picture leaving an impact on the mind. These also facilitate comparison. These are helpful in presenting facts to a layman.

The various types of diagrams are: bar diagrams, pie diagrams, pictographs and cartograms.

The various types of graphs are: time series graphs, histogram, frequency polygon, ogive or cumulative frequency curves.

9.3.4 Analysis of Data

After presentation, the data is analysed with the help of statistical tools like dispersion, average, correlation, association etc. The analysis involves mechanical processes, and it requires the knowledge of the application of statistical techniques. In analysis, the significant facts are abstracted from the huge body of data.

Basic steps in data analysis are:

- Categorizing data
- Coding data
- Calculating the appropriate statistics (such as descriptive statistics and inferential statistics)

Categorizing of data involves actual assigning of data to different categories. Coding means converting the raw data or responses to numerical codes, whereby they can be tabulated. If original responses are numerical, then they need not be assigned new codes.

Once data is ready to be analysed, then one can use descriptive statistics, or inferential statistics, or both.

Descriptive statistics simply summarizes and describes the data. It can be used to perform the following functions:

- Display of characteristics of the cases or individuals regarding the variable(s) to be

measured, by means of pictorial representation (graphs, charts, and tables)

- Measures of central tendency including the mean (average score), the median, (middle score), and the mode (most frequently occurring score).
- Measures of dispersion including mean deviation, standard deviation and the variance.
- Measures of relationship between or among the different variables in the data.
- Differences between two or more groups of individuals.
- Inferential statistics is commonly used for the following functions:
- to predict or estimate population parameters or characteristics from random sample statistics.
- to test hypotheses using tests of statistical significance to determine if observed differences between groups are real or merely due to chance.

9.3.5 Interpretation of Data

Interpretation consists of drawing conclusions from the data collected and analysed. It is a difficult job. An investigator must possess a high degree of skill and experience, to be able to do interpretation effectively. Interpretation must be done with reference to the studied data.

9.4 Advantages and Uses of Statistics

Following are some of the advantages and uses of statistics:

- Makes comparison between present and past data possible
- certain important correlation and associations of attributes can be found
- explains facts by revealing quantitative uniformities and relation between facts
- enables us to form a probable inference by calculation of chances and estimation of probabilities
- describes facts through convenient presentation of facts and data.
- leads to economy and high degree of flexibility
- useful for study of management, economics etc. Hence useful to bankers, statesmen, planners, speculators and researchers.
- Its approach is quantitative and hence definite and reliable.

9.5 Limitations of Statistics

There are certain limitations of statistics. These are as under:

1. Statistics deals only with quantitative data. Hence it is applicable to those problems which can be represented in terms of quantitative expressions.
2. Statistical numbers supply information about the characteristics of the group, but they do not speak any thing about any individual items
3. Statistical averages cannot represent an invariable relation in a group.
4. Co-efficients of correlations are not free from defects. Pearson's definition states that any two groups may be investigated for knowing the extent of correlation but the groups may be quite independent of each other.
5. When sampling is not fair, adequate or representative it may lead to dangerous fallacies.
6. Sometimes absolute numbers are used in comparing two groups, and percentages are neglected. This becomes a source of fallacy. If percentages are calculated, perhaps one would find a different story, in many cases, than what the numbers say.
7. Findings based on incomplete data are misleading.

8. Sometimes, statistics are misused intentionally.
9. Statistics neither proves anything, nor disapproves anything. It is merely a tool, which if used properly would benefit the society but if misused, it can lead to more harm than good.
10. Inspite of the limitations, statistical techniques are widely used in every branch of learning and enquiry.

9.6 Application of Statistical Techniques to the Field of Library & Information Science

In the field of library and information science, the use of statistical techniques to collect, organize, present, analyse and interpret data bears fruitful results. Statistical techniques help the librarians and library researchers to make correct interpretations of the collected data. In the process of conducting research, librarians compile and collect many different types of numerical data. Some of these data originate from the recorded observations of daily, routine library operations and others are generated in the form of responses to the various questionnaires, tests and other instruments that have been devised to obtain information regarding library phenomena. Knowledge of the techniques of quantitative analysis can aid librarians in making correct interpretations of these data.

The use of statistics can allow librarians to obtain maximum amount of information from their research efforts. The use of statistical methodology help the librarians to

- test hypotheses
- compute means and other measures of central tendencies
- make prediction
- determine the reliability and validity of instruments and measurements
- generalize conclusions from sample data population
- present research data in graphical and tabular formats
- calculate the validity of research
- determine the significance of the difference between the performance of two groups
- evaluate data in the annual report of a library
- take decisions regarding use of a core collection
- take decisions regarding search of an indexing and abstracting service for a particular query
- write a paper dealing with certain aspects of library and information science.

The librarian should become familiar with statistical techniques and their correct application. The determination of appropriate statistical procedures for a given application is based in part on the origin of the data to be analysed.

There are essentially two different classes of quantitative data based on how the data have been collected. Data that are the result of counting something can take on only a finite number of distinct values. Such data are termed discrete. Library examples of discrete data include the number of male or female borrowers, circulation data, records of reference inquiries and acquisition data.

The second type of data is obtained as a result of measurement. Essentially measurement entails the transformation of the physical characteristics of research subjects into numbers. Measurement involves the application of a scale to the phenomena of interest. A scale is characterized by the use of an arbitrary unit of measurement, against which the

phenomena in question is compared. Thus data obtained by measurement are by their nature approximate whereas data obtained by counting are theoretically exact. Measurement data can normally take on any value from an infinite number of possibilities, between given limits on a scale. Such data are referred to as continuous. Example of continuous data is 'an assessment of the relevance of a piece of library material to a patron'.

S.S. Stevans has identified four types of scales for measuring the physical characteristics of research subjects into numbers. These are nominal, ordinal, intervals and ratio scales.

In the nominal scales, one may determine whether two items are, or are not equal with respect to an unordered set of categories. Examples of nominal scales are classification of library users by age, sex, status, experience etc.

Statistics such as mode, contingency correlation, chi-square can be used for the analysis of such data.

In the ordinal scale, categories posses an inherent order, and a determination of greater than or less than is possible, in addition to a determination of equality. Many library data are ordinal. Theoretically, the median, percentiles and rank-order correlation are permissible for use with data arising from ordinal scales.

Interval scales are based on the metric concept i.e. the distance between adjacent interval scale categories are equal e.g. for establishing a scale of relevance "very relevant" "relevant" marginally relevant" and "not relevant" the numerical values can be assigned to these categories (4,3,2 and 1). These numbers could then be treated as interval data and statistics such as mean and standard deviation can be utilized.

A ratio scale is an interval scale in which an absolute zero point exists - a point at which zero indicates a total absence of that which is being measured. Data obtained by counting can often be regarded as arising from a ratio scale and thus such data can be analysed with ratio scale methods. Examples of data obtained from ratio scales in library and information science include salaries of librarians, daily circulation totals and the number of relevant citations retrieved in a literature search.

Thus while applying statistical techniques, the investigator must keep in mind the origin of data. Uptill recently, the librarians and library researchers have utilized whatever statistical procedure strikes their fancy regardless of the origin of data. It may be hoped that as research in library and information science becomes more sophisticated this situation will improve.

9.7 Summary

This lesson discusses the statistical techniques. Gives the definition of statistics; steps in scientific investigation; advantages and uses of statistics; and limitations. It also discusses the applicability of statistical techniques in the field of library and information science.

9.8 References

1. Bajpai, S.R. Methods of Social Survey and Research. Delhi: Har-Anand, 1982.
2. Busha, H. Charles and Harter, Stephen P. Research Methods in Librarianship. New York: Academic Press, 1980.
3. Krishan Kumar. Research Methods in Library and Information Science. New Delhi:

Har-Anand, 1992.
4. Median, J. Statistical Method. 2nd ed. Aucland: McGraw Hill, 1986.
5. Stevens, S.S. "On the Theory of Scales of Measurement". Science. 103, 1946, 677-680.

9.9 Self-Check Exercise

Note: Write the answers in the space given below each question and check your answers with the answers given at the end.

1. What are statistical techniques?

_____
_____
_____
_____
_____

2. Name the steps/ stages in scientific investigation.

_____
_____
_____
_____
_____

3. Mention the four scales given by S.S. Stevens for measuring the physical characteristics of research subjects into numbers.

_____
_____
_____
_____
_____

9.10 Answers to Self - Check Exercise

1. Statistical techniques are a mechanical process especially designed to facilitate the condensation and analysis of the large body of quantitative data. The aim of statistical techniques is to facilitate comparison, study relationships between the two phenomena and to interpret the complicated data for the purpose of analysis.
2. Statistical investigation involves the following five steps:
   1. Collection of data
   2. Organisation of collected data
   3. Presentation of organised data
   4. Analysis of data
   5. Interpretation of data
3. S.S. Stevans has given the following four scales for measuring the physical characteristics of research subjects into numbers.
   (i) Nominal Scale
   (ii) Ordinal Scale
   (iii) Interval Scale
   (iv) Ratio Scale

Lesson No. 2.4 | **Author : Dr. AMRITPAL KAUR**

Elementary Statistics : Mean, Median and Mode

Structure

10.0 Objectives

This lesson will make the learner familiar with the measures of central tendency i.e. mean, median and mode.

10.1 Introduction

Mean, median and mode are the measures of central tendency or averages of data. These describe the tendency of the items to group around the middle in a frequency distribution of numerical values. Usually central point of a distribution is the point where the concentration of values of items is the greatest and the most frequent on the scale of distribution. Therefore the value of the central tendency of the given data is regarded to be the most representative of the series. It depicts a given set of data on an average depicting the general magnitude of the figures.

There are two main objectives of computing a statistical average. One is to give concise picture of a large group, to describe the series it represents. The other is to afford a basis of comparison with the groups. It follows that an average may be computed for its own sake, or for comparison of two or more sets of the given data.

10.2 Mean

Mean is the simplest and the most widely used of all the measures of central tendency. It is obtained by dividing the sum of the values of all the observations in a series by the number of items constituting the series.

Thus mean of a set of numbers X_1, X_2, X_3 .....X_a. is denoted by $\overline{X}$ and is denoted by

$$\overline{X} = \frac{\sum X}{n}$$

Where

$\overline{X}$ = Mean

$\sum X$ = Sum of items in a series

n = number of items

10.2.1 Merits of Mean

(i) It is rigidly defined.

(ii) It is easy to understand and simple to compute.

(iii) It is possible to calculate the arithmetic average even if some of the details of the data are lacking.

(iv) It is not based on the position in the series.

(v) It gives weight to all items in direct proportion to their size.

10.2.2 Demerits of Mean

(i) It cannot be located through a frequency graph nor obtained by inspection.

(ii) It may lead to fallacious conclusion if the details of the data are not given.

(iii) It can ignore any single item only at the risk of losing its accuracy.

(iv) It cannot be employed in the study of qualitative phenomenon.

(v) It is very much affected by extreme values.

10.2.3 Computing Simple Mean (From ungrouped data)

The process of computing mean in a series of individual observations (i.e. where frequencies are not given) is very simple. It requires only two steps:

i. Add up all the values of the variables

ii. Divide the sum of values by number of items.

e.g. if we want to know the mean of professional persons in the college libraries

of Amritsar, it will be computed as:

College Libraries	No. of Professionals
1st	2
2nd	4
3rd	4
4th	6
5th	10
6th	8
7th	6
8th	10
9th	10
10th	10

The mean will be computed as

$$\overline{X} = \frac{\sum X}{n}$$

$$\overline{X} = \frac{70}{10}$$

$$\overline{X} = 7$$

In the above formula mean is represented by '$\overline{X}$', the summation by the Greek letter 'Σ' the data or values of the variable by 'X' and the number of items by 'n'.

10.2.4 Computing Mean in Discrete Series (From grouped data)

In the discrete series, we have grouped data to deal with i.e. items are grouped according to their size. The steps involved in the computation of mean are:

1. Multiply each item of variable 'x' with respective frequency 'f' which will result in 'fx'.
2. Add up all the frequencies (∑f). Total of the frequencies is the total number of items in a distribution (∑f=N)
3. Divide the ∑fx by ∑f or N to find out mean.

Symbolically it is represented as:

$$\overline{X} = \frac{\sum fx}{N}$$

Where

$\overline{X}$ = Mean

$\sum fx$ = sum of product of variables with respective frequencies

N = sum of the frequencies

Example: Calculate the mean of salaries of professionals in a university library from the given data:

Salaries in '000 Rs (x)	4	6	8	10	15	16
Professionals (f)	5	15	6	7	8	2

The following table shows the required computations

Table 2

Salaries in '000 Rs (x)	Professionals (f)	fx
4	5	20
6	15	90
8	6	48
10	7	70
15	8	120
16	2	32
	N=43	$\sum fx$=380

Applying the formula:

$$\overline{X} = \frac{\sum fx}{N}$$

$$= \frac{380}{43}$$

= 8.84

Thus average salary of the professional is 8840

10.2.5 Computing Mean in Case of Continuous Series (from grouped data)

In a continuous frequency distribution, the value of each individual frequency distribution is unknown. All that is known about the value of each individual frequency is that it is classified within a certain range of values known as class interval. Since the exact value of each frequency is unknown, the assumption is made that mid point of each class interval is the true average of all of the frequencies falling within that class interval e.g. if in a class interval say'0-10' there are 5 frequencies then it is assumed that all these frequencies have a value of (0-10)/2=5 each. Hence mid values of all the class intervals have to be found out, which forms the basis of computing arithmetic mean in a continuous series.

The following procedure is adopted for calculating mean in a continuous

series:

1. Mid values of the class intervals (m) are to be found out.
2. Each mid value is to be multiplied by its respective frequency and the product is denoted by 'fm'.
3. All the products of mid values and frequencies are to be added up (∑fm) Add up the number of frequencies (N or ∑f)
4. Divide the sum of the products (∑fm) by the total frequencies (N)

The result is arithmetic mean

Symbolically :

$$\overline{X} = \frac{\sum fm}{N}$$

Where:

$\overline{X}$ = arithmetic mean

∑fm = total of the frequency of each class multiplied with the mid value of the respective class.

N = total of the frequencies

e.g. Following is the frequency distribution of professionals in a university libraries of India. We are to find out the mean of the professionals

Professionals	0-10	10-20	20-30	30-40	40-50	50-60
University libraries	5	20	40	20	15	5

Table 3

Professionals(X)	University lib. (f)	Mid point of class (m)	fm
0-10	5	5	25
10-20	20	15	300
20-30	40	25	1000
30-40	20	35	700
40-50	15	45	675
50-60	5	55	275
	N=105		∑fx= 2975

Applying the formula:

$$\overline{X} = \frac{\sum fx}{N}$$

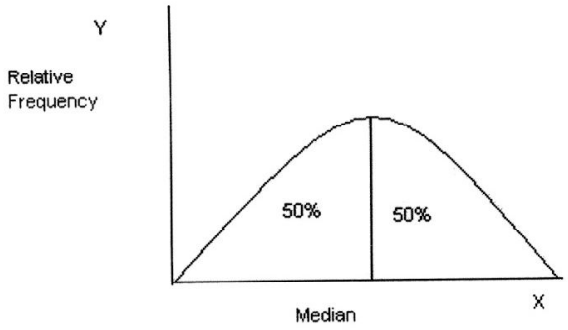
$$= \frac{2975}{105}$$

= 28.30

The average no. of professionals in university libraries is 28.

10.3 Median

The median is another important and widely used statistical average. It has the connotation of the middle most or most central value of a set of numbers. It is a positional measure of a data set. Its value is so located in the frequency distribution that it divides it in half, with 50% of the items below it and 50% of the items above it.



10.3.1 Merits

(i) It is rigidly defined.
(ii) It is easy to understand and simple to complete.
(iii) It is not based on position in the series.
(iv) It gives weight to all items in direct proportion to that their size.

10.3.2 Demerits

(i) It cannot be located through a frequency graph nor obtained by inspection.
(ii) It may lead to fallacious conclusion if the details of the data are not given.
(iii) It can ignore any single item only at the risk of loosing accuracy.
(iv) It is very much affected by extreme values.

10.3.3 Computing median: Individual Series (from ungrouped data)

For computing median in an individual series the following steps are involved:

1. The series is arranged in either ascending or descending order. Both ways the answer will be the same.
2. If n is odd, the median is the middle number.
3. If n is even, the median is the mean of the middle two numbers.

Example: Find the median of the following distribution of N=9 library professionals: 3,5,4,6,7,8,4,3,1

Arranging the professionals in ascending order we get 1,3,3,4,4,5,6,7,8

Since the n is odd, the median is the middle value i.e. '4'. It shows that 4

libraries have less than 4 professional persons and the rest four libraries have more than 4 professional persons. If there had been 10 libraries i.e. even number, then the median should be got by adding the two values in the middle and dividing by two.

e.g. Obtain the value of median from the following data:

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

Median = size of $\left[\frac{n+1}{2}\right]^{th}$ item

$= \frac{11}{2}$

$= 5.5^{th}$ item

size of 5.5th item = $\frac{5^{th} \text{ item} + 6^{th} \text{ item}}{2}$

$= \frac{4+5}{2} = 4.5$

10.3.4 Computing Median: Discrete series (from grouped data)

For computing median in a discrete series, the following steps are involved:

1. Arrange the items in ascending or descending order of their magnitude
2. Compute cumulative frequency by adding their respective frequencies
3. By applying the formula ((n+l)/2) the median item is known.
4. Median is located at the size of the item in which cumulative frequency, the value of ((n+1)2)th item falls.

e.g. Calculate the median no. of professionals in College Libraries from the given data.

No of professionals	10	8	7	4	2	1	0
No of colleges	20	30	40	60	100	30	20

Table 4

No. of Professionals	Colleges acc to frequency		Professionals in ascending order	Cumulated frequency(n)
10	20		0	20
8	30		1	50
7	40		2	150
4	60		4	210
2	100		7	250
1	30		8	280
0	20		10	300

$$M = \frac{N+1}{2} = \frac{300+1}{2} = 150.5^{th}$$

Note : (The number of colleges is arranged according to frequency i.e the ascending order of professionals).

Value equal to on next higher than 150.5 ⇒ 210

No. of professionals corresponding to 210 = 4. Hence, median of No. of Professionals in college libraries is 4.

10.3.5 Computing Median : Continuous Series (From Grouped Data)

To compute median in a continuous series, the following steps are taken:

i. Frequencies are made cumulative

ii. The median item is known by finding $\frac{N}{2}^{th}$ out item.

iii. The median item is located in cumulative frequency column and it lies in that class. The class interval is called median class interval.

iv. The median value is calculated by applying the following formula.

$$M = L + \left[\frac{N/2 - cf}{f}\right]Xc$$

L = Lower limit of the median class

N= Total Frequencies

cf= Cumulative frequency of the class preceeding the median class.

F = frequency of the median class.

C= class interval / width of the median class.

e.g. Following is the frequency distribution of professionals in university libraries of India. We are to find out the median.

University Libraries	5	10	15	20	15	10	15	10
Professionals	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40

Table 5

Professionals	University Libraries	cf
0-5	5	5
5-10	10	15
10-15	15	30
15-20	20	50
20-25	15	65
25-30	10	75
30-35	15	90
35-40	10	100

$\frac{N}{2} = \frac{100}{2} = 50$. Median Class ⇒ [15-20]

$$M = 15 + \frac{50-30}{20} \times 5$$

$$= 15 + \frac{20}{20} \times 5$$

= 15+5

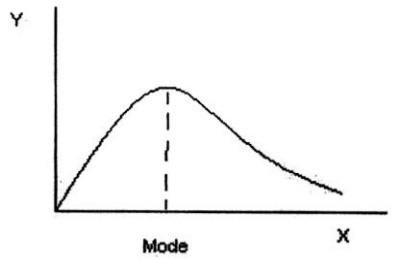
= 20

The median can also be calculated by taking upper limit of median. The formula would be

$$M = U - \frac{N/2 - cf}{f} Xc$$

10.4 Mode:

The mode is defined as the value that occurs most frequently in a statistical distribution e.g. the mode of the series 3, 5, 8, 5, 4, 5, 9, 3 would be 5, since this value occurs more frequently than the others. But if each observation occurs the same number of times, then there is no mode in that distribution if two or more observations occur the same number of times then there is more than one mode and the distribution is multi-modal as against unimodal. If two values occur most frequently then the series is called bimodal and in case of three values trimodal. The following diagram shows the modal value.



The value of the variable at which the curve reaches a maximum is called the mode. It is the value around which the items tends to be most heavily concentrated.

10.4.1 Merits of Mode :

(i) Mode is readily comprehensible and easily calculated.

(ii) It is the best representative of data.

(iii) It is not at all affected by extreme value.

(iv) The value of mode can also be determined graphically.

(v) It is usually an actual value of an important part of the series.

10.4.2 Demerits of Mode :

(i) It is not based on all observations.

(ii) It is not capable of further mathematical manipulation.
(iii) Mode is effected to a great extent by sampling fluctuations.
(iv) Choice of grouping has great influence on the value of mode.

10.4.3 Computing Mode :- Individual Series (From Ungrouped Data) For determining mode, count the number of times the various values repeat themselves. The value which occurs the maximum number of times is the modal value. The more often the modal value appears relatively, the more valuable the measure is an average to represent data. e.g let us find the mode of the following data. The following is the number of persons in different libraries:

40,30,30,40,32,40,35

Array: 30, 30, 32, 3 5, 40, 40, 40

Mode 40

10.4.4 Computing Mode: Discrete Series (From grouped data)

In a discrete series, modal class can be identified by the simple examination of the data set or by the grouping.

By Examination:

The mode can be determined by looking to that value of the variable around which the items are most heavily concentrated e.g Observe the following data:

Professional persons	10	8	7	4	2	1	1
No of colleges	20	30	40	30	100	30	20

From the above data we can clearly say that the modal value is 2 i.e. 100 colleges have 2 professional persons. By grouping

Grouping is done as given below:

1. In the first column, the highest frequency is noted and marked.
2. In the second column, frequencies are grouped in twos, starting from the top. Their totals are found out and the highest total is marked.
3. In the third column, frequencies are again grouped into twos but the first frequency is left out i.e this time grouping will start from the second frequency. Their totals are found out and the highest total is marked.
4. In the fourth column, frequencies are grouped in threes, starting from the top. Their totals are found out and the highest total is marked.
5. In the fifth column, frequency is grouped again in three but the grouping will start from the second frequency. The totals are done and the highest total is marked.
6. In the sixth column, leaving the first and second frequency grouping will be done in threes. Their total will be found out and

the highest total is marked.

After preparing the grouping table, an analysis table is prepared. From this analysis table, the item size having the largest number of appearance in the highest grouping class is taken as mode.

The procedure of preparing grouping table and analysis table shall be clear from the following example:

Table 6

Grouping Table

Professional personals	Col1	Col2	Col3	Col4	Col5	Col6
0	20					
1	30	50	(130)			
2	(100)			(150)	(160)	
4	30	(130)	70			(170)
7	40					
8	30	70	50	100	90	
10	20					

Analysis table: value of variables

Table 7

Col.	Frequency→	0	1	2	4 7	8	10
1				1			
2				1	1		
3			1	1			
4		1	1	1			
5			1	1	1		
6							
Total		1	3	(5)	2		

The above table shows that 2 has the most frequent value of variable. Since it has occurred the maximum number of times i.e. 5, the modal value is 2.

10.4.3. Computing Mode: Continuous Series (From grouped data)

There are two methods of computing mode from a grouped data set(1) Calculation method (2) Graphic method

10.4.3.4 Calculation Method :

The steps in the computation of mode from a continuous frequency distribution are:

1. Determine the modal class interval.
2. Determine the value of mode by applying the following formula

$$L + \left[\frac{f^{x} - fi}{2f - f1 - f2} \right] Xc$$

where : L is the lower limit of the modal class
f is the frequency of the modal class
f1 is the frequency of the group before the modal class
f2 is the frequency of the group after the modal class
c is the class interval of the modal class.

Example: Calculate mode from the following data set

GROUPING TABLE

Professional personals	No. of Lib.	Col2	Col3	Col4	Col5	Col6
0-5	5	17				
5-10	12		24	29		
10-15	12				44	
15-20	(20)	32	(40)			(52)
20-25	(20)			(60)		
25-30	(20)	(40)				
30-35	15		35		(55)	45
35-40	10	25				

ANALYSIS TABLE

	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Col 1				1	1	1		
Col 2					1	1		
Col 3				1	1			
Col 4				1	1	1		
Col 5					1	1	1	
Col 6			1	1	1			
Total	0	0	1	4	6	4	1	0

Modal Class = 20 -25

$$\text{Mode} = 20 + \left[\frac{20-20}{40-0-0}\right] \times 5 = 20 + 0 = 20$$

Table 8

Professional persons	No. of University libraries
0-5	5
5-10	10
10-15	15
15-20	20
20-25	15
25-30	10
30-35	15
35-40	10

$$\text{Mode} = 15 + \frac{20-15}{2x20-15-15} X5$$

= 5+5/10x5

= 15+1/2X5

= 15+5/2

= 15+2.5

= 17.5

Mode when class intervals are unequal:

The formula for calculating the value of mode is applicable only where there are equal class intervals. If the class intervals are unequal, then we must make them equal before we start computing the value of mode. The frequencies should be adjusted on the assumption that they are equally distributed throughout the class.

Example: Calculate mode from the following data set

Professional persons	0-5	5-15	15-30	30-35	35-40
University libraries	5	24	60	15	10

The above data will be adjusted as under:

Table 9

Professional persons	University libraries
0-5	5
5-10	12
10-15	12
15-20	20
20-25	20
25-30	20
30-35	15
35-40	10

Mode: $5 + \frac{20-12}{2x20-12-20} X5$

= 15+8/8X5

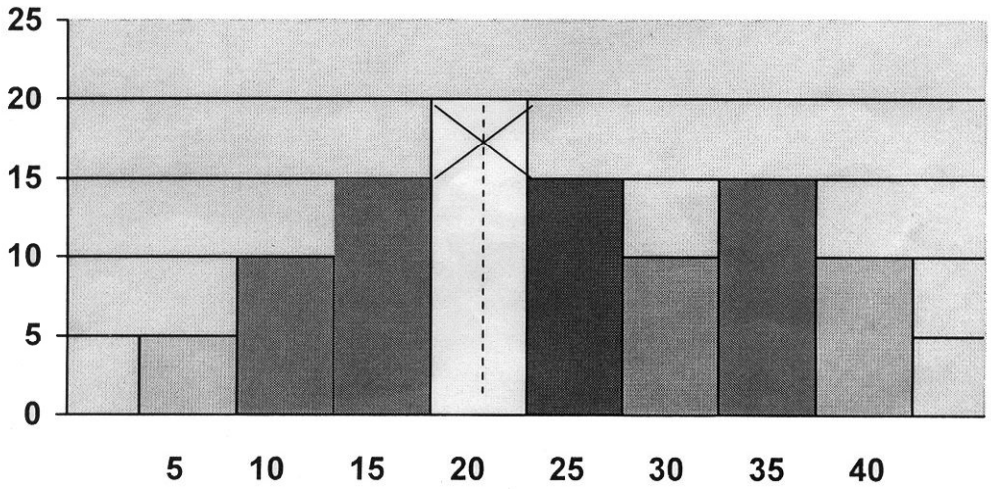
= 15+5

= 20

10.4.3.2 Graphic Method

The steps involved in this process are:

1. Draw a histogram of the given distribution
2. The highest rectangle will be of the modal class
3. Draw a line from the top right hand corner of the modal class rectangle to the point, where top of the next adjacent rectangle to the left touches it. Draw a corresponding line on the opposite diagonal from the top left hand corner to the rectangle on the right.
4. From the point of intersection of these lines, draw a perpendicular on the X-axis and read the modal value:

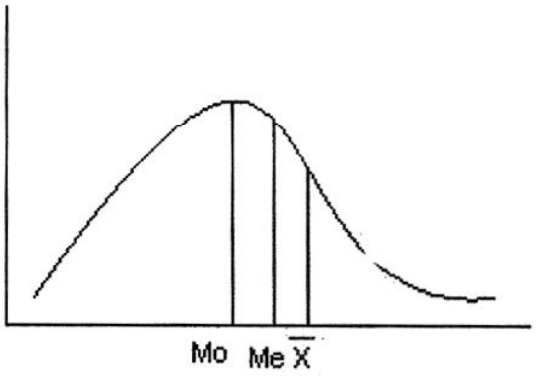


Graphic method of determining mode can be used when there is one class containing the highest frequency. If two or more classes have the same highest frequency, mode cannot be determined graphically.

10.5 Relationship among Mean, Median and Mode:

A distribution in which the values of mean, median and mode coincide is known as symmetrical distribution. But when the values of mean, median and mode are not equal, the distribution is known as asymmetrical or skewed. In

moderately skewed or asymmetrical distribution a very important relationship exists among mean median and mode. In such distributions, the distance between the mean and the median is about one third the distance between the mean and the mode as will be clear from the following diagram:



Karl pearson has explained this relationship as follows:

Mode = mean-3 (mean-median)

Or Mode = 3median -2mean

And median = Mode + 2/3 (mean-mode)

If we know any of the two values out of the three we can compute the third from these relationships

10.6 Summary

This lesson discusses the measures of central tendency or averages of data i.e mean, median and mode in detail. The formulas for computing mean, median and mode in individual, discrete and continuous series have been given. Various examples have been solved to make the learner aware of how to calculate the mean, median and mode from a given data set.

10.7 Expected Question:

1. Describe measures of Central Tendency
2. What do you mean by mean, median and mode?

10.8 References :

1. Busha, C.H and Harter, S.P Research Methods in Librarianship. New York: Academic Press, 1980.
2. Garrett, H.E and Woodworth, R.S. Statistics in Psychology arid Education. Bombay: Vakils, Feffer and Simons Ltd, 1981
3. George W.S. and Cochran, W.G. Statistical Methods. 6th ed Ames:

Iowa State University Press, 1988.

4. Simpson, IS Basic Statistics for Librarians 3rd ed, London: Library Association, 1988.

10.9 Self Check Exercise :

Note : Write the answers in the space given below each question and check your answers with the answers given at the end.

1. Define Mean
2. Calculate simple mean from the following data:

College libraries(N)	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
No of professionals(X)	2	4	6	6	10	8	6	10	10	10

3. Define Median
4. Find the median from the following frequency distribution of professionals in university libraries of India.

University libraries	5	10	15	20	15	10	15	10
Professionals	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40

5. Define Mode and find the mode of following distribution of professional persons in different libraries: 40,30,30,40,32,40,35.

10.10 Answers to Self-Check Exercise:

1. Mean is the measure of central tendency. It is obtained by dividing the sum of the values of all observations in a series by the number of items constituting the series. The mean of a set of numbers X1, X2, X3, X4,X5 .....X9 is denoted by $\overline{X}$

and is defined as: $\overline{X} = \frac{\sum X}{N}$

where

$\overline{X}$ = Mean

$\sum X$ = sum of items in a series

n = Number of items

2.

College libraries (N)	No of professionals (X)
1st	2
2nd	4
3rd	6
4th	6
5th	10
6th	8
7th	6
8th	10
9th	10
10th	10
10	70

$$\overline{X} = \frac{\sum X}{n}$$

= 70/10

=7

3. The median is measure of central tendency. It has the conotation of the middle most or most central value of a set of numbers. Its value is so located in the frequency distribution that it divides it in half, with 50% of the items below it and 50% of the items above it.

4. The formula for calculating median in a continuous series is as under :

$$M = L + \left[\frac{N/2 - cf}{f}\right]Xc$$

where :

L = lower limit of the median class

N = Total frequencies

cf = cumulative frequency of the class preceding the median class.

f = frequency of the median class

c= class interval

Professionals	University librarian	cf
0-5	5	5
5-10	10	15
10-15	15	30
15-20	20	50
20-25	15	65
25-30	10	75
30-35	15	90
35-40	10	100

M = 100/2 =50

$$= 15 + \frac{50-30}{20} x5$$

$$= 15 + \frac{20}{20} x5$$

= 15+5

= 20

5. The mode is defined as the value that occurs most frequently in a statistical distribution. The mode of the series 40,30,30,40,32,40,35 will be 40, since the value occurs more frequently than the others.

10.11 Further Readings :

1. Gupta, S. P. Statistical Methods. New Delhi : Sultan Chand & Sons, Ed 39th ed. 2010.
2. Sardana, J. L. and Sehgal, R. L. Statistical Methods for librarians. New Delhi : Ess Ess Publications, 1982.

Lesson No. 2.5 **Author : Dr. AMRITPAL KAUR**

Elementary Statistics : Mean Deviation, Standard Deviation, Percentage, Ratio, Frequency

Structure

11.0 Objectives

This lesson will make the learner familiar with the following:

i. Measures of Dispersion-Mean deviation and Standard deviation
ii. Percentage
iii. Ratio
iv. Frequency and Frequency Distribution

11.1 Introduction :

Mean deviation and standard deviation are measures of dispersion. Dispersion is the spread of the data in a distribution i.e. the extent to which the observations are scattered. The degree of scatter of an individual observation is measured be measuring the distance between the average and individual observation of the given date set. More distant the individual observations are from the average, higher is the spread of the data.

Percentage is a fraction with its denominator as 100. A ratio is a comparison of one magnitude with another as a multiple or as a fraction. Frequency is a number which tells that how many times does a particular data appear in a given set of data.

11.2 Mean Deviation

Mean deviation is the measure of dispersion. It is simply the mean distance of the difference between the distribution mean and each datum where each difference is expressed as a positive number.

11.2.1 Computation of Mean Deviation for Ungrouped Data

The formula for computing mean deviation for ungrouped data is

$$MD = \frac{\sum |D|}{n}$$

Where IDI denotes deviation from median

Steps:

(i) Compute the median of the series

(ii) Take the deviations of items from medium ignoring + - signs.

(iii) Obtain the total of these deviations i.e $\sum D$

(iv) Divide the total obtained in step (iii) by the number of observations

e.g.

No of Professional Persons	Mean Deviation
10	10
15	5
20	0
25	5
30	10
100	30

$$\text{Mean} = \frac{100}{5} = 20$$

$$MD = \frac{30}{5} = 6$$

11.2.2 Computation of Mean Deviation for Grouped Data

The formula for computing mean deviation from grouped data is

$$MD = \frac{\sum |D|}{N}$$

Steps:

(i) Calculate the median of the series.

(ii) Take the deviation of the items from median ignoring + - signs.

(iii) Multiply these deviations by the respective frequencies and obtain the total

(iv) Divide the total obtained in step (ii) by the number of observations. This gives us the value of mew deviation.

Table 2

Prof. Persons	Univ. Libs.	Mid Point	Step (M-45) Deviation	MD	X|D|	fX
0-10	10	5	4	40	30	300
10-20	20	15	3	60	20	400
20-30	30	25	2	60	10	300
30-40	50	35	1	50	0	0
40-50	30	45	0	0	10	300
50-60	20	55	1	+20	20	400
60-70	10	65	2	+20	30	300
	170			170		2000

$$\text{Mean } (\bar{x}) = A + \frac{\sum f\ D}{N} \times i$$

$$45 - \frac{170}{170} x10$$

45-10 =35

$$\text{Variation} = \frac{\sum fx}{N}$$

$$= \frac{2000}{170}$$

= 11.76

11.2.3 Merits and Limitations of Mean Deviation

Merits

(i) As compared to other calculated measures of dispersion, mean deviation is easy to Calculate and understand.

(ii) It takes into account all the items of the series and and hence it is affected by every value in the distribution.

(iii) It is less affected by the extreme items.

(iv) It shows the significance of an average in the distribution.

Limitations :

(i) The greatest drawback of this method is that algebric signs are ignored while taking the deviations of the items e.g if from 20, 50 is deducted we write 30 and not –30. This is mathematically wrong and makes the method non-algebric.

(ii) It is not a well defined measure as it can be computed by an average. Mean deviation calculated from various averages will not be the same.

11.3 Standard Deviation

Standard deviation is widely used measure of studying dispersion. It is also known as root mean square deviation for the reason that it is the square root of the mean of the square deviation from the arithmetic mean. Standard deviation is denoted by the small Greek letter σ (read as sigma).

Standard deviation measures the absolute dispersion of variability of a distribution. A small standard deviation means a high degree of uniformity of the observations as well as homogeneity of a series; a large standard deviation means just the opposite. Hence standard deviation is extremely useful in judging the representativeness of the mean.

11.3.1. Comparision between SD and MD : Both these measures of dispersion are based on each and every item of the distribution. But they differ in the following respects.

	Standard Deviation	Mean Deviation
1.	These are calculated from arithmetic mean.	M. D. are calculated from Mean, mode, median.
2.	Algebric signs are not ignored, so it is mathematically sound.	Here only absblute values are considered, thus it takes mathematical properties.
3.	It is based on the square root of the average of the square deviation.	It is based on simple average of the sum of absolute deviation.
4.	It is simple to calculate.	It is complex to calculate.
5.	The algebric signs have to be ignored only values of deviations are taken.	Since the deviations are squared, the plus and minus signs need not be omitted.

11.3.2 Computation of Standard Deviation : Ungrouped Data

The formula for computing standard deviation from the ungrouped data is as under:

$$\sigma = \sqrt{\frac{\sum X^2}{n}}$$

Steps :

(i) Calculate the actual mean of the series i.e. $\bar{X}$

(ii) Take the deviation of the items from the mean i.e. find $(x - \bar{x})$. Denote these deviations by the x

(iii) Square these deviations and obtain the total $\sum X^2$

(iv) Divide $\sum X^2$ by the total number of observations i.e. n and extract the square root. This gives us the value of standard deviation.

e.g.

Table 3

Professional Persons	$x = X - \bar{X}$	x^2
10	- 20	400
15	- 15	225
20	- 10	100
25	- 5	25
30	0	0
35	5	25
40	10	100
45	15	225
50	20	400
270		1500

$$\bar{X} = \frac{270}{9} = 30$$

$$\sigma = \sqrt{\frac{\sum x^2}{n}}$$

$$\sigma = \sqrt{\frac{1500}{9}}$$

$$= \sqrt{166.66}$$

$$= 12.9$$

In the above example, first of all the mean of professional persons is taken i.e. 30. Then the mean is subtracted from each individual item. i.e. x (e.g. 10-30 = 20). These deviations are squared (X^2) to get their sum i.e. $\sum X$. $\sum X^2$ is divided by n to take square root and the final result is standard deviation.

11.3.3 Computation of Standard Deviation for Grouped Data.

The formula for computation of standard deviation for grouped data is as under:

$$SD = \sqrt{\frac{\sum fd^2}{n}}$$

Where f is frequency and d is the deviation of interval mid point from mean.

11.3.2.1 Computation of Standard Deviation for Grouped Data

Steps (i) Take the deviations of the items from the mean. Denote these deviations by d

(ii) Square these deviations to get d^2 and multiply by the respective frequencies to get fd^2

(iii) Add all the items i.e. $\sum fd^2$; take the average i.e. $\left(\frac{\sum fd^2}{n}\right)$ and then take square root to get

$$SD = \sqrt{\frac{\sum fd^2}{n}}$$

Computation of standard deviation from grouped data:

Table 4

X	(f)	m	fm	α = m – 35	d^2	fd^2
0-10	5	5	25	-30	900	4500
10-20	10	15	150	-20	400	4000
20-30	15	25	375	-10	100	1500
30-40	20	35	700	0	0	0
40-50	15	45	675	10	100	1500
50-60	10	55	550	20	400	4000
60-70	5	65	325	30	900	4500
Total	80 = N		2800			20000

$\overline{X}$ = Arithmetic mean = $\frac{\sum fm}{N}$ = 2800/80 =35

$$S.D. = \sigma = \sqrt{\frac{20000}{80}}$$

$$= \sqrt{250}$$

=15.81

11.3.4. Uses of Standard Deviation

1. Standard deviation helps in ascertaining the degree of variation or uniformity in two or more data sets.

2. It helps to gauge the representativeness of the mean. If we have two or more comparable distributions with the same mean, then the distribution with the smallest standard deviation has the most representative mean.

11.4. Percentage

Percent means out of hundred or per hundred. The symbol % is used for percent. Thus seven percent = 7% = 7 out of hundred.

Percentage is a fraction with its denominator as 100. The computation of percentages is generally straightforward and consists merely of dividing the number in each category by the total number. e.g. If a student gets 38 marks in a paper of maximum marks 50. Then his percentage score Will be :

(38/50) X100 = 76%

11.5 Ratio

A ratio is a comparison of one magnitude with another as a multiple or as a fraction. The main purpose of ratio is to simplify the numbers used in certain comparisons. If we compare the number of male employees with the number of female employees in a library, we may express the comparison in absolute numbers as 150 male employees to 50 female employees. As a fraction this becomes 150/50 or sometimes for typographical convenience 150/50. This may also be stated as a ratio of 150:50 or 3:1. This later form for expressing a ratio is called a proportion.

It is often useful to express ratios with 100 as the base (or 10 or 1000 or still others). Thus we prefer 300/100 or 300:100 than 150/50 or 150:50; all four forms are mathematically equal but the ratio to 100 are perhaps more easily grasped and compared. A special and common example of such a ratio is percentage.

A ratio between two magnitudes usually shown over a period of time is called a rate if the magnitudes are qualitatively different even though expressesed in the same units. Thus an interest rate of 4% on a corporation's bonds means that for every 100 Rs of Principal invested in these bonds an interest of Rs 4 a year is paid.

Thus ratios may take the form of fractions, proportions, and percentages or rates.

11.6 Frequency

Frequency is a number, which tells that how many times a particular data appear in a given set of data e.g. Consider the following set of data 1,2,3,2,1,5,4,3,2,1,2. In the given set of data, 1 appears 3 times therefore frequency of 1 is 3. Similarly 2 appears four times, therefore the frequency of 2 is four and so on.

11.6.1 Frequency Distribution

A frequency distribution is the simple tally of scores or values of characteristics which have been taken from any collection of elements. The frequency permits the researcher to see at a glance how certain scores taken from these elements are distributed.

e.g In a group of 10 persons with their ages varying from 25 to 44 years a frequency distribution maybe found as follows.

Table 6

Age classification of 10 persons

Age category	Corresponding number of persons
25-29	2
30-24	3
35-39	3
40-49	2

The above table gives quick glance at the age group of all persons in the group where variable 'X' is ranging from 25 to 44 years, which is depicted in an array of ascending order i.e. from lowest value of variable to the highest value, divided in four class intervals 25-29,30-34,35-39 and 40-44. The class limits involved are 25,29,30,34,35,39,40 and 44. The size of a class interval is 5.

11.6.2 Cumulative frequency and Cumulative frequency Table

The cumulative frequency of a class interval is the sum of frequency of all classes up to that class (including the frequency of that particular class) The cumulative frequency table for example given above is as follows:

Table 7

Age category	Frequency	Cumulative Frequency
25-29	2	2
30-34	3	2+3=5
35-39	3	2+3+3=8
40-49	2	2+3+3+2=10

11.6.3. Frequency Distribution Graphs

A frequency distribution can be presented graphically in any of the

following ways.

11.6.3.1. Histogram

11.6.3.2. Frequency polygon

11.6.3.3. Bar Graphs

11.6.3.4. Pie Charts

11.6.3.1. Histogram

It is the most popular and widely used in practice for presenting a frequency distribution graphically. It is a set of vertically bars whose areas are proportional to the frequencies represented.

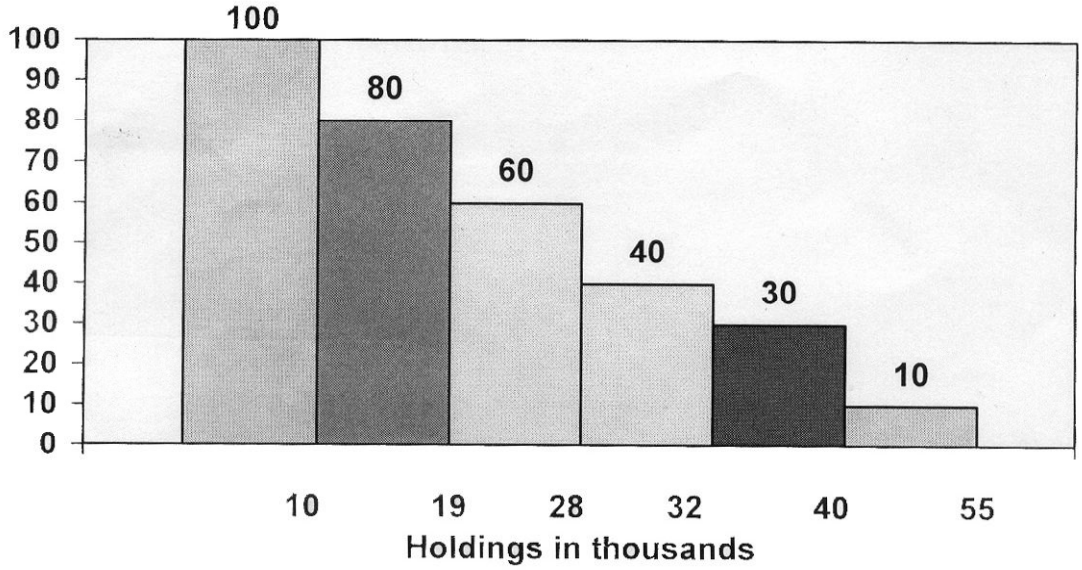
While constructing histogram, the variable is always taken on the X axis and the frequency depending on it on the Y axis.

Histogram for the equal class interval.

When class intervals are equal, the height of rectangles will be proportional to the frequencies e.g. we have to conduct a survey of libraries of Punjab. We take six classes of public libraries and they have holdings as shown in the following table

Variables	Frequencies
1000-10,000	100
10,000-19,000	80
19000-28000	60
28000-37,000	40
37000-46000	30
46000-55000	10

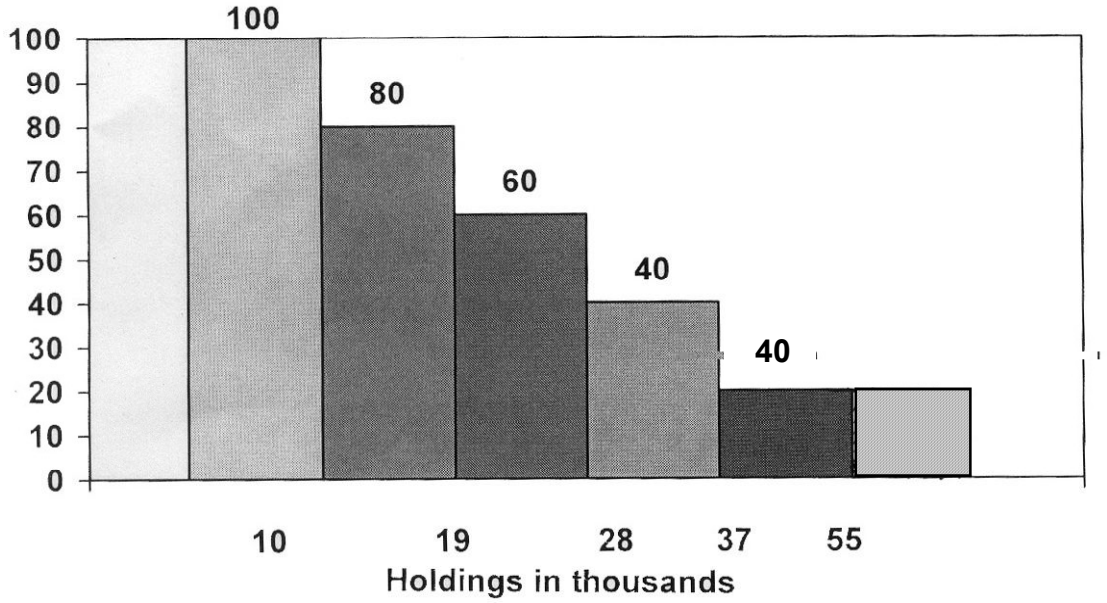
Graph 1



Histogram for unequal interval

When class intervals are unequal, the frequencies must be adjusted before constructing the histogram. We take that class which has the lowest class interval and adjust the frequencies of other classes in the following manner. If the class interval is twice as wide as the one having lowest class interval, we divide the height of its rectangles by two, if it is three times more, we divide the height of its ractangles by the three etc. i.e. the height will the proportional to the ratio of the frequencies to the width of the classes e.g.

Variables	Frequencies
1000-10,000	100
10,000-19,000	80
19000-28000	60
28000-37,000	40
37000-55000	40



11.6.3.2 Frequency Polygon

A frequency polygon is a graph of frequency distribution. It has more than four sides. It is particularly effective in comparing two or more frequency distributions. There are two ways in which a frequency polygon may be constructed.

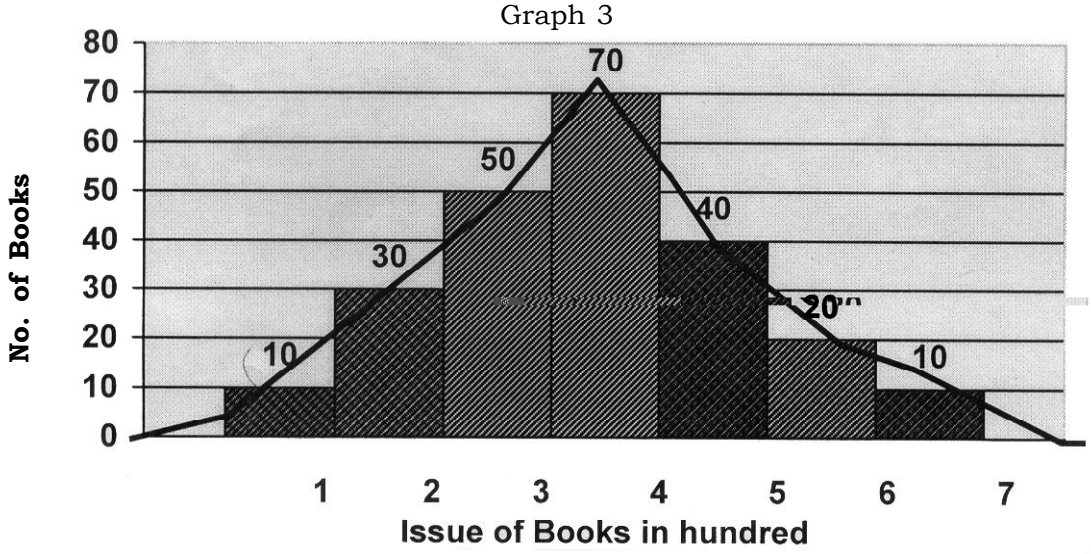
1. We may draw a histogram of the given data and then join by straight lines the mid points of the upper horizontal side of each rectangle with the adjacent ones. The figure so formed is called frequency polygon.
2. Another method is to take the mid points of various class intervals

and then plot the frequency corresponding to each point and join all these points by straight lines.

Table 10

Issue of books	Mid Point	Libraries
0-100	50	10
100-200	150	30
200-300	250	50
300-400	350	70
400-500	450	40
500-600	550	20
600-700	650	10

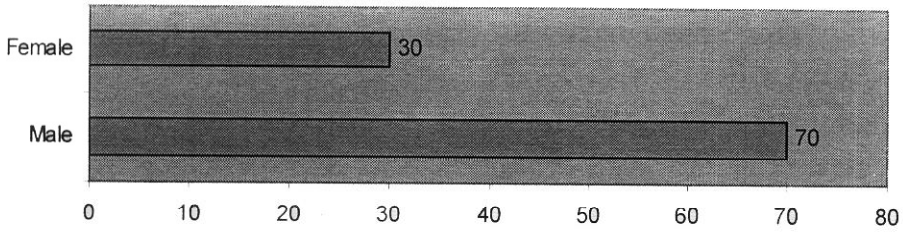
Graph 3



It is quite simple to make. One straight-line joints the mid points. Two false classes are made on each side. Their frequency is supposed to be zero. Area covered by polygon is shadowed.

11.6.3.3 Bar Graphs

Bar graphs are horizontal bars representing the degree to which certain characteristics may exist among any collection of elements. The following figure is a bar graph showing the percentage of male female ratio in a library.



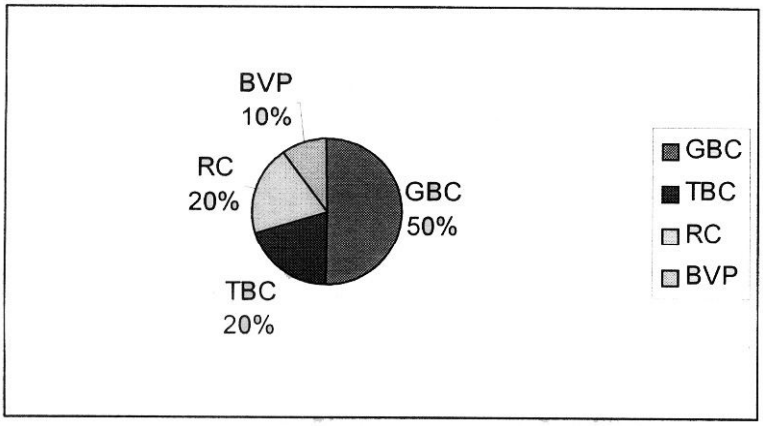
Bar graphs are very easy to construct. We do not need to be concerned about the width of the bars as we do with the histograms.

11.6.3.4 Pie Charts

Pie charts are circular graphs which are divided into sectors representing fractions of the total circle.

e.g. Various types of collection in a library can be shown by a pie chart as under:

General Book collection (GBC) = 50%
Text Book Collection (TBC) = 20%
Reference Collection (RC) = 20%
Back volumes of Periodicals (BVP) = 10%



The main advantage of pie charts is that they offer a rapid appraisal of the data. They serve to make a discussion of findings illustrate and meaningful. Their only limitation is that they cannot easily accommodate too many divisions of the data.

11.7. Summary

This lesson discusses the measures of dispersion i.e. mean deviation and standard deviation, percentage, ratio, frequency and frequency distribution in great detail. This lesson also includes some of the primary graphic techniques used by the social researchers to categorise and visualize some of the information which they collect.

11.8 Expected Questions

1. What do you mean by mean deviation. Compute mean deviation for ungrouped as well as grouped data giving suitable examples.
2. What do you mean by standard deviation. Compute standard deviation for ungrouped as well as grouped data giving suitable examples.
3. What is frequency. Discuss frequency distribution giving suitable examples.

11.9 References

1. Prusha, C. H. and Harter, S.P. Research Methods in Librarianship. New York: Academic Press, 1980.
2. Garrett, M.E and Woodworth, R.S. Statistics in Psychology and Education, Bombay: Vakils, Feffer and Simons Ltd, 1981
3. Simson, J.S. Basic Statistics for Librarians. 3rd ed. London, Library Association, 1988
4. Gupta, S.P. Statistical Methods, 39th ed.-New Delhi : S. Chandism, 2010.

11.10 Self Check Exercise

Note: Write the answers in the space given below each question and check your answers with the answers given at the end.

1. Define mean deviation.
2. Write the formula for computing standard deviation for grouped data.
3. What is percentage and how is it calculated ?
4. What is ratio?
5. What do you mean by frequency?

1. Mean deviation is the measure of dispersion. It is simply the means distance of the difference between the distribution mean and each datum where each difference is expressed as a positive number.
2. The formula for computing standard deviation for grouped data is as under

$$SD = \sigma = \sqrt{\frac{\sum fd^2}{n}}$$

3. Percentage is a fraction with its denominator as 100. The computation of percentages is generally straightforward and consists merely of dividing the number in each category by the total number e.g. if a student gets 38 marks in a paper of maximum marks 50, then his percentage score will be.

$$\frac{38}{50} \times 100$$

4. A ratio is a comparison of one magnitude with another as a multiple or as a fraction. e.g. if in a library, the number of male employees is 120 and female 40, then the ratio will be 120 : 40 or 3: 1
5. Frequency is a number which tells that how many times a particular data appear in a given set of data e.g. in the following set of data. 1,2,3,2,1,5,4,3,2,1,2,
   1 appears 3 times, therefore frequency of 1 is 3. Similarly 2 appears 4 times, therefore frequency of 2 is 4 and so on.

LESSON NO. 2.6 | **AUTHOR : DR. AMRITPAL KAUR**

Computerized Data Analysis : Description, Analysis and Interpretation : Use of SPSS

Structure

12.0 Objectives

This lesson will make the learner familiar with data analysis techniques and use of statistical package for social sciences (SPSS)

12.1 Introduction

Prior to early 1960's, analysis of data collected in large scale research projects might take weeks or even months to accomplish because it used to be done manually. Today, the speed of the computer allows the same analysis to be performed in minutes. Therefore, projects can now be undertaken which might not have been considered four decades ago because of the enormity of the associated data analysis problem. Although the drudgergy associated with computation has been effectively removed from data analysis by the computer, the analysis is left with the creative managerial decision relating to the choice of statistical technique and interpretation of results

12.2 Data Analysis

In most social research, the data analysis involves three major steps as given below:

- Cleaning and organizing the data for analysis (Data Preparation)
- Describing the data (Descriptive Statistics)
- Hypotheses testing (Inferential Statistics)

12.2.1 Data Preparation

Data preparation involves checking the data for accuracy, coding the data, entering the data into the computer, transforming the data and statistically adjusting the data.

12.2.1.1 Checking for Data

In any research project, the data is received from a number of different following sources :

- Mail surveys returns
- Coded interview data
- Pretest or posttest data
- Observational data

As soon as data is received, it is screened for accuracy The following questions should be raised as part of initial data screening:

- Are the responses legible / readable?
- Are all important questions answered?
- Are the responses complete?
- Is all relevant information included?

Assuring that the data collection process does not contribute inaccuracies will help assure the overall quality of subsequent analysis.

12.2.12 Coding the Data

Coding for computer analysis generally consists of assigning a code number to each answer category, So that the answer may be stored on computer cards. It is much easier to store and retrieve numbers than it is letters or words i.e. rather than punch a 'Yes' or a 'No' response on a computer card, it is much simpler and takes less space to assign each answer a number e.g. 'yes' equals 1 and 'no' equals 2 and simply punch the appropriate number on the card.

In every research project, generally a printed codebook that describes the data and indicates where and how it can be accessed is generated. The codebook should include the following items for each variable,

- Variable name
- Variable description
- Variable format (number, data, text)
- Instrument/method of collection

- Data collected
- Respondent or group
- Variable location (in database)
- Notes

12.2.1.3 Logging/Entering the Data

Though there are a wide variety of ways to enter the data into the computer for, analysis, the easiest one is to just type the data in directly using standard statistical programmes, such as SPSS, SAS, datadesk. To use one of these programmes, investigators need only record their data in machine readable form e.g. on punched cards or magnetic tape and to instruct the operating system associated with the computer system being used in (a) a description of data (b) the input device to be utilized and (c) the statistical analyses that will be called upon.

12.2.1.4 Data

Once the data has been entered, it is transformed into variables that are usable in the analysis. There are a wide variety of transformations that the investigator might perform. Some of the most common are:

- If the analysis programme does not treat blank values as missing, the investigator need to designate specific values to represent missing values. For instance, the investigator might use a value of-99 to indicate that the item is missing.
- On scales and surveys, sometimes reversal items are used to help reduce the possibility of a response set. When the investigator analyses the data, he wants all scores for scale items to be in the same direction where high and low scores mean the same thing. In these cases, he has to reverse the rating for some of the scale items. Once the individual scale items are transformed, these are added or averaged across individual items to get a total score for the scale.
- Sometimes the variables are collapsed into categories.

12.2.2 Describing the Data

Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures, together with sample graphic analysis, they form the basis of virtually every quantitative analysis of the data.

12.2.3 Interpreting the Data

There are four interrelated components that influence the conclusions the investigator might reach from a statistical test in a research project. These are as under:

- Sample size, or the number of units
- Effect size
- Alpha level (a) (or significance level)
- Power

Given values for any three of these components, it is possible to compute the value of the fourth. For instance, the investigator wants to determine what a reasonable sample size would be for a study. If he could make reasonable estimates of the effect size, alpha level and power, it would be simple to compute the sample size.

Some of these components will be more manipulable than others depending on the circumstances of the project. For example, if the project is an evaluation of an educational programme with a specific number of available consumers, the sample size is set or predetermined.

12.2.4 Hypothesis Testing

Statistics that are used to infer the truth or falsification of a hypothesis are called inferential statistics. All statistical conclusions involve constructing two mutually exclusive hypotheses - null and alternative. The null hypothesis is symbolized by Ho and the alternative by Ha. Together the hypotheses describe all possible outcomes with respect to the inference. The central decision involves determining which hypothesis to accept and which to reject. For instance, in the typical case, the null hypothesis might be:

Ho : Program Effect = 0

While the alternative might be

Ha: Program Effect < > 0

The null hypothesis is so termed because it usually refers to the "no difference" or "no effect" case. Usually in social research we expect that our treatments and programmes will make a difference. So typically, one theory is described in the alternative hypothesis.

12.2.4.1 Chi-square (χ^2)

The most commonly used test of significance for independence for tables containing nominal and ordinal values is X^2. Imagine that we conducted a survey to ask about relationship between education and income and received the hypothetical data as shown in the following table

Table: 1

Income	Education	
	High (College)	Low (High School or less)
High	40	10
Low	10	40
Total	50	50

To support our hypothesis we have to show that a person's education makes a difference when predicting his or her income.

Table 1 shows that education and income are not independent but are related. Eighty percent (40 out of 50) of people with high education also have high income, while only 20% of those with low education also have high income. Thus we would

guess that the relationship is non zero. A good way to approach the problem is to determine what independence, or a zero relationship would look like in the table and to compare our data with independence to see how much the two tables differ:

Table 2 A

Income	Education	
	High (College)	Low (High School or less)
High	25	25
Low	25	25
Total	50	50

Table 2 shows that 50% of our sample with high education have high income, and 50% have low income and 50% with low education also have high income and 50% low income. This shows that there is no relationship between the variables. Education does not affect income.

The values in above table are known as expected values, since they are the one that we would expect in the case of independence. In a sense these values are only hypothetical in that they are computed but not found in the data collected. They merely serve as a criterion for computing the data actually collected during research. The values actually collected are shown in table 1 and are known as observed values To see whether the results we found (table 1) are significantly different from independence (Table 2) we compare the observed data with the expected data. The distribution of the difference between observed and expected values have been found to approximate the X^2 distribution as indicated by the formula.

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

Where 'O' indicates the observed value and E the expected value. The larger the summed difference between observed and expected values as a ratio of expected values, the larger the value of Chi-square and the greater the likelihood that the relationship is statistically different from zero. To see whether our relationship is significant we need only compare each respective cell of table I with table 2 as follows:

$$\chi^2 = \frac{(40-25)^2}{25} + \frac{(10-25)^2}{25} + \frac{(10-25)^2}{25} + \frac{(40-25)^2}{25}$$

$$= \frac{(15)^2}{25} + \frac{(15)^2}{25} + \frac{(15)^2}{25} + \frac{(15)^2}{25}$$

$$= \frac{225}{25} + \frac{225}{25} + \frac{225}{25} + \frac{225}{25}$$

$$= \frac{900}{25} = 36$$

Most statistics texts provides a table of X^2 values that allow us to determine whether our value is statistically significant. The value of 36 is far above 7.88, Since X^2 is basically only a probability measure of faith we can have that our relationship is nonzero. It, too, is affected by sample size. The table requires us to know the number of degrees. The formula for computing degrees of freedom (d. f) for contingency table is (R - 1)* (C- 1), where R = number of rows in the table and C = number of columns. In the 2*2 case, d f=(2-1) or 1. X^2 table tells us that for 1 d. f. a X^2 value of 7.88 is significant at the .005 level meaning that if we reject the null hypothesis of independence we will be wrong only five times out of 1000. Our value of 36 is far above 7.88. So we reject the hypothesis and conclude that the relationship between education and income in table 2 is statistically significant at the P=.005 level.

12.2.4.2 Correlation

The correlation is one of the most common and most useful statistics. A correlation is a single number that decides the degree of relationship between two variables. Several techniques have been devised to represent the relationship between two variables. The numerical values rendered by such statistical techniques are called association coefficients, and generally they can range in value from – 1.00 to +1. 00. The coefficient of association + 1.00 means that there is a positive relation between the two variables and -1,00 means negative relation between them.

The pearson r or product-movement correlation coefficient is the best known parametric measure of association. It is appropriate when both variables are measured at an interval level.

Suppose an investigator wants to determine the extent to which a relationship exists between size of income and the number of years of education the individual has completed. Table 3 shows the income of individuals in hundreds alongwith their respective years of education. To facilitate the computatrion of r, these columns have been labeled X and Y respectively, standing for income and education. Third and fourth columns have been labeled X^2 and Y^2. These columns simply contain the squares of each value in columns X and Y, The last column contains the product of each X by its respective Y value.

Table 3
Relation between Income and Education

Individual	Income (In Hundreds) X	Education (Years) Y	X^2	Y^2	XY
1	45	20	2025	400	900
2	63	19	3669	361	1197
3	36	16	1296	1256	576
4	52	20	2704	400	1040
5	29	12	841	144	348
6	33	14	1089	196	462
7	48	16	2304	256	768
8	55	18	3025	324	990
9	72	20	5184	400	1440
10	66	22	4356	484	1452
N=10	ΣX=499	ΣY=177	ΣX^2=26793	ΣY^2=3221	ΣXY=9173

The following formula is used to compute Pearson r for such data

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N(\Sigma X^2) - (\Sigma X^2)][N(\Sigma Y^2) - (\Sigma Y^2)]}}$$

Substituting the values in Table 3 for symbols in the formula, we have

$$r = \frac{(10)\ (9173) - (499)(177)}{\sqrt{[(10)(29796) - (499)^2][(10)(3221) - (177)^2]}}$$

$$r = \frac{91730 - 88323}{\sqrt{(267{,}930 - 249{,}001)(32210 - 31{,}329)}}$$

$$r = \frac{3407}{\sqrt{(18{,}929)(88)}}$$

$$r = \frac{3407}{4083.63} = 0.83$$

The interpretation here would be that there is a high degree of positive association between one's income and years of education.

The other popular measures of correlation include the Phi co-efficient (also

called the four fold point co-efficient) for two nominal variables of two categories each; gamma and its fourfold equivalent Q, for ordinal variable ; and sho(Spearman rank order correlation) for ranked ordinal variables.

Most researchers do not compute Pearson's or any other commonly used statistics by hand, but rely upon so called packaged computer programmes, such as Statistical Package for Social Sciences (SPSS). Such programmes require only that the researcher have a clean deck of computer data cards and that he or she punch a few control cards containing such details as the statistic to be used, the columns that the variables are in and the number of cases. After the results are received, the researcher need only know how to interpret the statistics.

12.3 Statistical Package for Social Sciences (SPSS)

SPSS is a computerized statistical package for Social Scientists. It contains pre- written programmes for statistical analysis. The users are to indicate to the system which of these they wish to employ, characteristics of the data on which the analysis are to be conducted, and the data themselves. Punched cards for communicating this information to SPSS are called control cards. More than 75 different types of control cards may be used with SPSS. The SPSS conventions used to construct the control cards constitute a language, with its own syntax and vocabulary. These language rules must be followed by the analyst who runs an SPSS programme.

Table 4 illustrates the ease Table 4 with which an SPSS programme can be written.

Table 4

SPSS Program to Conduct Correlation and Regression Analysis.

File Name	SCATLIB
Variable list	VARI.VAR2
Input medium	CARD
N of cases	30
Input format	Freefield
Scatter gram	VAR 2 with VAR 1
Option	7
Statistics	All
Read input data (data)	
Finish	

This program performs a correlation and regression analysis on a set of bivariate data. Each line in the program is a separate SPSS statement, or instruction, and accomplishes a particular function. The correct order of the statement is important to maintain as well. The function of each statement of the program is illustrated in Table 5 :

Table 5
Explanation of Statements in a SPSS Program

SPSS statement		Explanation
File name	SCATLIB	Identifies the user-provided name of the data file. In this case the name of the file is SCATLIB. (Required if reference to the file is to be made at a later time)
Variable list	VAR1. VAR2	Names variables lo be reffered to in the program Variable names invented by the user In this case VAR1 and VAR2, (Required)
Input medium	CARD	Specific input medium, in this case, punched card. (Required)
N of cases	30	Specifies number of cases in data file, in this case, 30. (Conditionally required)
Input format	Freefield	Specifies the format of the data on the input medium In this case, no format is specified. (Conditionally required)
Scatter gram	VAR2 with VAR1	A procedure card, specifying the particular statistical procedure to be followed. In this case, the procedure card requires that a scatter gram be plotted with (VAR1-VAR2) (at least one procedure card is required)
Options	All	Specifies the statistics (associated with the procedure) that are desired. In this case regression and correlation coefficients are computed, as well as other statistics, and a t-test for significance is conducted.
Read input data (data)		Instruct system to read input data. (Required) (user -supplied data Cards)
Finish		Signals end of program (Required)

The preceding SPSS program illustrates that a fairly sophisticated set of analysis can be utilized in a short –10 statements - and easily written set of instructions. This is generally true with SPSS, Unfortunately, this ease of access and use implies at least two kinds of misuse of SPSS programs.

The first type of misuse involves a tendency to allow consideration of previous and potential hypotheses, theories, concepts and ideas to be supplemented by "fishing expeditions" on the data. Because potential relationship among variables can be tested so easily with SPSS and other statistical software packages, one is tempted to consider wholesale examinations of possible relationships, whether or not pertinent theory suggests that the relationships might exists. In the long run, library and other social science research will suffer from this practice, unless this is resisted.

The second type of misuse arises from a lack of understanding of invoked statistical procedures. In SPSS and other packages, one may use a given statistical procedure without having the slightest understanding either of the procedure or the information (or analysis) produced.

The SPSS manual attempts to avoid this type of misuse by including a useful prose description of each statistical procedure and examples of its potential use.

12.3.1 Main Features of SPSS

1. **Options :** SPSS provides three main options to input the data for analysis and computation. These are :
   - **(a) Prompt Option :** This command makes easy to input and execute command in quick and easy way.
   - **(b) Review editing :** This option provides option to review the text and make modifications and changes as per the need.
   - **(c) Menu help system :** The menu help system provides the menu of commands so that these commands are easily selected from the commands for execution and it also provides help to the user to operate the software.
2. **Command Used in SPSS :** In the SPSS, the commands which are used for execution are simple English statements like, File, Edit, Save, Review, Open, Exit, View.
3. **Multiple Approach System :** In SPSS there is a capacity to fulfil the multiple approach. In reserach it is easy to compare and analyse the variables like univariable, bivariate and multivariate variables.
4. **User's Guide :** It provides user's guide. It makes the use of SPSS easy by the layman. There is no need to have technical background to operate on it.
5. **Windows Version :** The latest edition of SPSS is Window Version. It is easy to operate became every one has the basic knowledge to use windows. The commands Open, file, Exit, Save i.e. window commands are used.

Thus SPSS is the most important software which helps in the analysis of computerised data.

12.4 Summary

This lesson discusses the analysis, description, interpretation of data and statistical hypotheses testing. Also discusses X^2 test of significance and correlation. Statistical Package for Social Sciences (SPSS) has been discussed in great detail.

12.5 Expected Questions

1. What do you mean by data analysis ? Discuss the use of SPSS for computerized data analysis?
2. Discuss in detail the use of SPSS for computerized data analysis?

12.6 Self-Check Exercise

Note: Write the answers in the space given below each question and check your answers with the answers given at the end.

1. Name the steps involved in data analysis in any social research project.
2. What do you mean by coding for computer analysis ?
3. Name the components that influence the conclusions in any social research project.
4. Write the formula for computing the Pearson 'r' for interval variables.
5. What is SPSS ?

12.7 Answers to Self- Check Exercise

1. The data analysis involves three major steps as given below:
   - Cleaning and organizing the data for analysis.
   - Describing the data.
   - Hypotheses testing.
2. Coding for computer analysis generally consists of assigning a code number to each answer category, so that the answers may be stored on computer cards.
3. The components that influence the conclusion the investigator might reach from a statistical test in a research project are as under :
   - Sample size, or the number of units .
   - Effect Size.
   - Alpha level (a) (or significance level).
   - Power.
4. The formula for computing the Pearson 'r' for interval variables is as under :

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N(\Sigma X^2) - (\Sigma X^2)][N(\Sigma Y^2) - (\Sigma Y^2)]}}$$

5. SPSS is a computerised statistical package for social scientists. It contains pre-written programmes for statistical analysis.

12.8 Further Readings :

1. Gupta, S. P. Statistical Methods. New Delhi, Sultan Chand & Sons, Ed 39th 2010.
2. Sardana, J. L. and Sehgal, R. L. Statistical Mehtods for Librarians, New

Delhi: Ess Ess Publications, 1982.

3. Busha, C. H. and Harter, S. P. Research Methods in Librarianship, New York : Academic Press, 1980.
4. Champion, D. J. Basic Statistics for Social Research. USA: Chandler Pub Co., 1970.
5. Simpson, 1. S. Basic Statistics for Librarian. 3rd ed. London : Library Association, 1988.
6. Kothari (CR) : Research Methodology : Methods & Techniques Ed2. 2008.

Lesson No. 2.7 | **Author : Dr. AMRITPAL KAUR**

Research Reporting: Structure, Style, Contents. Guidelines for Research Reporting

Structure

13.0 Objectives

This lesson will make the learner familiar with the

i. Structure, contents and style of research reporting.
ii. Guidelines for research reporting
iii. Style of writing
iv. Citations as per MLA & APA Manuals; and
v. Methods of research evaluation

l3.1 Introduction

Research reporting is an important facet of the research process, and scholars have a scientific obligation to effectively communicate views about completed investigations. It is at the reporting that the researcher reports the findings of the study and draws the conclusions out of his findings.

Research reports are detailed and accurate accounts of the conduct of disciplined studies accomplished to solve problems or to reveal new knowledge. These can be useful if these are well prepared, thorough and disseminated in an appropriate publication vehicle. Research reports must be complete, well

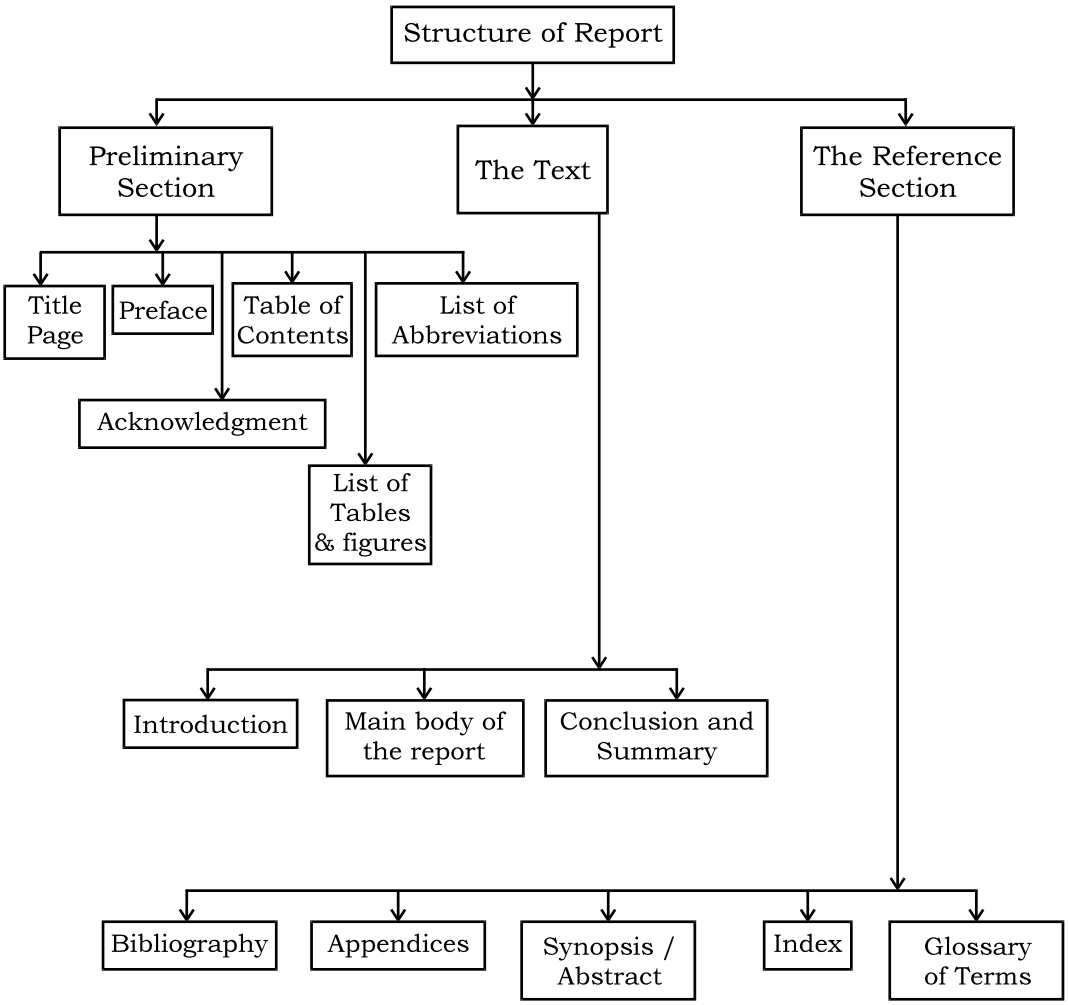
organized and carefully written if their contents are to contribute to librarianship's store of knowledge.

A research report is meant for several categories of audiences i.e. scientist's, practioners, policy formulators, funding agents and general public. The reports should focus on the target audience and include the material of interest to other audience.

13.2 Structure and Contents of a Research Report

Generally a research report consists of three parts:

1. The preliminary section
2. The Text
3. The Reference material



13.2.1 The Preliminary Section

The preliminary section usually consists of

I. Title page
II. Preface
III. Acknowledgement
VI. Table of Contents
IV. List of Tables and Figures
V. List of Abbreviations

I. Title page: The title must be descriptive and appropriate to the problem proposed to be investigated.

II. Preface: Preface may include reasons why in the first place, the topic was selected by the researcher. It disseminates information about history, scope and methodology.

III. Acknowledgement: Acknowledgement is written to thank those who have helped the researcher during the course of study.

IV. Table of Contents: The purpose of table of contents is to provide an outline of the contents of the report. It contains a list of titles of chapters and their appropriate Roman numerals.

V. List of Tables and figures: There are separate pages for list of tables and figures. Arabic numerals are usually used for identifying tables, figures, etc.

VI. List of Abbreviations: It contains list of abbreviations used throughout the report.

13.2.2 The Text

The text usually consists of:

I. Introduction
II. Main body of the report
III. Summary and conclusion

I. Introduction

3 Statement of the problem, specific questions to be answered or hypotheses to be tested.

- Review of Literature: a review of the pertinent past work and contradictions, pit falls and failings of the earlier work mainly to substantiate the need for another research study.
- Significance of and justification for the present study
- Coverage: Full information about universe of the study, sampling frame and units of observation and sampling size.
- Conceptual framework: Definitions or special meanings of all the important terms so as to enable the reader to understand the concepts underlying the development of the investigation.
- Preview of the scheme of the chapters in the main body of the thesis and their is interrelationship.

II. Main body of the report

This is the heart of the research report and probably the largest section of the report. It contains all the arguments, ideas, concepts, interpretations and findings. The data should be described fully and analyzed in detail and the evidence resulting from the analysis should be presented.

III. Summary and conclusion

It consists of the summary, conclusions, generalizations, suggestions and recommendations.

13.2.3 The Reference Material

The various varieties of reference material are as under:

I. Bibliography

It is a list of sources, either published or unpublished, consulted in the preparation of the report during the course of research.

II. Appendices

An appendix is used for additional or, supplementary material which has not found place in the main text. Appendix includes original data schedules, questionnaires and interview forms, copies of cover letter used in the study, long explanatory notes to the text etc.

III. Synopsis / Abstract

It is also attached to the thesis. It should not be too long. Its limits either in the number of words or pages should be stipulated.

IV. Index

The index disseminates information to readers, researchers, scholars and scientists about specific material. It is usually arranged alphabetically.

V. Glossary of terms

The Glossary of terms depicts the definition about the particular topics and the meaning of abbreviations.

13.3 Guidelines for Research Reporting

1. The report should be long enough to cover the subject but short enough to maintain interest.
2. The layout should be well thought out and must be appropriate and in accordance with the objectives of the research problem.
3. The presentation must fit to the special interest and need of the audience for whom it is intended.
4. Abstract terminology should be avoided.
5. The report must be attractive in appearance, neat, clean whether typed or printed.
6. Include whatever information may be necessary for the reader to critically access your findings and the conclusions you draw from them.
7. whenever possible relate your findings to those of other studies and to broader problems related to the one earlier studied.

8. Ideas, concepts, and words of other authors must be formally acknowledged, otherwise it leads to plagiarism.
9. If possible include information about procedures, special problems and other topics that would be of particular interest to other researchers in the area. Put them in one or more appendices if it is not appropriate to include them in the text.
10. Use footnotes for clarifying explanations that would not fit into the text.
11. If the report is intended for the use in an action program, give some specifications of how the findings may be used and what their limitations are.
12. Aim at originality and independent thinking while writing the research report.
13. Bibliography is included at the end of the report. It is an essential part of report writing.

13.4 Style of Writing

The basic qualities of good scientific writing are accuracy and clarity. While writing the report, the following things should be kept in mind.

1. Sentences must be as simple as possible. Writing long paragraphs should be avoided. Appropriate subheads should be provided wherever necessary.
2. More care in using -terminology is needed. Scientific terms must be adequately defined and used consistently. The target audience and their knowledge of technical terms should be considered.
3. Adequate attention should be paid to the correct use of spellings and grammar.
4. As far as possible present tense should be used. When past events are referred, these may be reported entirely in the past tense.
5. Direct and positive sentences should be used. Round about constructions using long, technical or unusual words or phrases should be avoided
6. All chapters and sections, subsections, tables and charts should be labeled adequately. The system of headings and subheadings should be simple
7. Use footnotes sparingly, label serially and give them either at the bottom of the respective pages or at the end.
8. The report must go through three drafts -first, second and final to make it authoritative.

Apart from the above, a standard unabridged dictionary, subject encyclopedias, text books of grammar, and a manual for report writing should be on the table when writing a report.

13.5 Style Manuals - MIA & APA

The use of style manuals in citing references ensures efficiency and uniform system of citing the information. Today, a number of referencing styles are available. Some well known among them are: Harvard, American

Psychological Association (APA), Modern Language Association (MLA), Chicago Manual etc. Out of these, the styles prescribed by MLA and APA are outlined in the following sections:

While writing, citations must be given at the following two places:

i. Within the running text section: and

ii. In the list of references

Both of these types of citations are given below:

(A) Citations in the Text

Single Author

MLA	APA
(Marcuse 197)	(Marcuse, 1975 p. 197)

Comments: MLA uses surname and pages within a bracket but APA includes the date of publication as well as the abbreviation p. before the page number.

Two or upto three authors

MLA	APA
(Cheek and Buss 332)	(Cheek and Buss, 1981, p. 332)
(Oilman, King and Porter 217)	(Gilman, King and Porter, 1993 p. 217)

Comments: MLA uses surnames of two/three authors followed by pages within a bracket whereas APA uses surnames of two/three authors followed by year of publication and pages within a bracket.

Four or more authors

MLA	APA
(Quirk et al. 210/Quirk,	(Quirk et al., 1985, p. 210/Quirk,
Greenbaum, Leech and Sartvik)	Greenbaum, Leech and Sartvik)

Comments: Both the styles prescribe first author's surname followed by et al. or surnames of all the authors.

Note: Chapters in a book as well as conference proceedings papers are cited on the analogy of simple book. In case of editorial works, the abbreviations ed. and comp. etc. are not included in the Text References but they appear in References list entries given at the end.

Citations in the Reference List

Reference list is the complete list of sources used in writing. It is also referred to

as works cited, literature cited, sources cited, references etc. It is given at the end of the work, usually in alphabetical order with full bibliographical details of all sources used in the work.

Citing Books and other Non-Periodical Publications in the Reference list:

Single Author

MLA

Gibaldi, Joseph. MLA Handbook for Writers of Research papers. 6th ed. New Delhi: Affiliated East-West Press, 2004.

APA

Gibaldi, Joseph(2004). MLA handbook for writers of research papers. 6th ed. New Delhi: Affiliated East-West Press.

Comments: In MLA, the title is underlined whereas in APA, it is italicized. In MLA every important word of the title is capitalized and the publication date is placed at the end whereas in APA, only the first word of the tide is capitalized and the date (in parentheses) immediately follows the name of the author.

Two Authors

MLA

Chowdhury, G. G. and Sudatta Chowdhury : Searching CD ROM and Information Sources. London: Library Association, 1992.

APA

Chowdhury, G. G. and Chowdhury. Sudatta(1992). Searching CD Rom and information sources. London: Library Association.

Comments: In MLA only first author is inverted other author is given in its natural order, whereas in APA the names of both the authors are inverted.

Three Authors

MLA

Leech, Geoffry, Margaret Deucher and Robert Hoogenroad. English Grammar for Today: A New Introduction. London: Macmillan, 1982.

APA

Leech,Geonry, Deucher, Margaret and Hoogenroad, Robert(1982). *English grammar for today:* A new introduction. London: Macmillan.

Four Authors

As per MLA and APA styles, give either the first name only followed by et

al. Quirk, Rendolph, et. al. or all name in full in the order in which they appear on the title page: Quirk, Randolph, Sidney Greenabaum, Geoffrey Leech, and Jan Savartvik (MLA) Quirk, Randolph, Greenabaum, Sidney, Leech Geoffrey, and Savartvik, Jan (APA).

Editorial Work

To cite an anthology or compilation that appears under the name of editor/ compiler, the entry begins with the name of editor or compiler followed by a comma and the abbreviation ed. or comp. e.g.

Sardana. J. L.. ed.

Spafford, Peter, comp.

A work in an Anthology/Book

To cite a work in an anthology or a book in MLA style the following information is added.

Author, title and if relevant translator of the part of the book being cited. The title of the piece is enclosed in quotation marks.

e.g. Hansberg, Lorraine: A Raisin in the Sun Black Theatre: A Twentieth Century Collection of the World of its Best Playwrights. Ed. Lindsay Patterson. New York Dodd, 1971.221-76

Conference Papers and Proceedings

MLA and APA treat the published proceedings of a conference like a book, but add the pertinent information about the conference.

MLA

Freed, Barbra F. ed. Foreign Language Acquisition Research and the Classroom, Proc. of Consortium for Lang. Teaching and Learning Conf., Oct. 1989, U of Pennsylvinia. Lexington: Heath, 1991.

APA

Freed, Barbra F. ed. (1991) Foreign language acquisition research and the classroon Proceedings of Consortium for Language Teaching and Learning Conference, Oct 1989, University of Pennsylvania. Lexington: Heath.

An Article in a Printed Journal

To cite an article in a printed journal in the reference list, the following elements are generally included:

- Author's name
- Title of the article
- Name of the periodical
- Series number or name, if relevant
- Volume number
- Issue number, if needed
- Date of publication
- Page numbers of the article,

MLA

Mann, Susan. "Myths of Asian Womanhood." Journal of Asian Studies 59 (2000): 835- 62.

Natrajan, M. "Digital Preservation of Cultural Heritage and its Implication." Journal of Library and Information Science 29.1-2 (2004)" 14-25.

APA

VandenBos, G., Knapp, S. & Doe, J. (2001). Role of reference elements in the selection of resources by psychology undergraduates. Journal of Bibliographic Research, 5, 117-123.

Internet Sources / E-citations

Citations in the Text

Citations for Internet sources in the text are given just like printed documents.

Citations in the Reference List.

In the Reference List, an Internet source should contain the following information:

- Author's name if given. In case editor, compiler or translator is there, cite that followed by appropriate abbreviation
- Title of the work
- Name of the editor, compiler, or translator, if required
- Date of electronic publication
- Date of accessing the source
- Network address or URL

Online Book

MLA

Keats, John. Poetical Works. 1884. Bartleby.com: Great Books Online. Ed. Steven Van Leeuwen. 2002. 5 may 2002 <http://www.bartleby.com/126/>.

APA

Chou, L. McClintock, R. Moretti, F., Nix, D. H. (1993). Technology and education: New wine in new bottles: Choosing past and imagining educational futures: Retrived August 24, 2004 from http://www.ilt.columbia.edu/ publications/paper/ newwine 1 .html

A part of an Online Book

MLA

Keats, John. "Ode on a Gracian Urn." Poetical Works. 1884. Bartleby.com: Great Books Online. Ed. Steven Van Leeuwen. 2002.5niay 2002 <http:// www.bartiety.com/ 126/41.html>.

Article in an Online Journal

MLA

Dane, Grabrielle. "Reading Ophelia's madness." Exemplaria 10.2 (1998). 22 June 200S <http://web.english:ufl.edu/english/exemplaria/danefram.html>

APA

Fredrickson, B. L. (2000, March 7). Cultivating positive emotions to optimize health and well being. Prevention and Treatment. 3. Article OOOla. Retrieved November 20.2000 from http: //journals.apa.org/prevention/ volume3/pre003000la.html.

13.6 Methods of Research Evaluation

Although research is scientific, it also involves creativity, intuition and expertise Hence every research project provides an opportunity for learning and the researcher should critically evaluate the entire project to obtain new insights and knowledge. Tb critical analysis of many aspects of another researcher's report helps researchers it develop competency in their own research and reporting skills. An evaluation done objectively by an expert can be helpful in understanding the real worth of a study. In fact a research study fulfils its purpose only to the extent the evaluation considers it sc Even though it is a difficult task, an understanding of the evaluation exterior can be helpful to the researcher himself. This can serve as self evaluation and help to improve the study further.

What constitutes a good research report is not agreed upon uniformly by research advisors and research project report examining committees. However, the following questions are suggested as a possible digest of suggestions agreed upon by a number of authors.

Title

- Is the title clear, concise and correctly worded
- Is it descriptive of and appropriate to the problem under investigation.

Problem

- Is the problem significant enough?
- Is it clearly and concisely stated?
- Is it broken down into specific questions and issues?
- Is it properly delimited?
- Is the analysis of the problem logically sound?
- Is the solution to the problem likely to prove a good contribution to knowledge?
- Are specific questions raised: Hypotheses clearly stated?
- Are assumptions, limitations and delimitations stated?
- Is review of related literature covered?

Research Design of the Study

- Is the research design described in detail?
- Is it adequate? Are there any alternatives available?
- Is the description of design understandable so that it can be replicated with ease ?

Collection of data

- Are the population and samples described?
- Are relevant variables recognized?
- Are appropriate controls provided?
- Are data gathering instruments appropriate and explained adequately?
- Is the kind of data chosen for the study adequate enough for the solution to the problem?

Presentation of Data

- Is appropriate use made of tables and figures to present pertinent data?
- Are the methods of presentation of data accurate and effective?

Analysis of Data

- Are the techniques and methods of analysis appropriate to the problem and applied adequately and effectively?
- Is the analysis of data adequate?
- Are the results of the analysis presented clearly?
- Are validity and reliability established?
- Is the analysis objective and unbiased?

Interpretation of Data

- Are the inferences based on data seem to be sound?

Summary and Conclusions

- Does the conclusion of the study really answer the research questions or issues raised in the introduction of the study?
- Are findings and conclusions justified by the data presented and analyzed?
- Are conclusions based on evidence only?
- Are recommendations made judiciously?
- Whether topics for further research included / suggested?
- Can the conclusions be generalized to a larger population and incorporated in the existing theory?

Form and Style

- Are table of contents, list of tables, list of figures, bibliography, appendices, etc.are properly placed?
- Are proper spacing, punctuation, spelling, margin, pagination, etc. followed?
- Is it free from typographical errors?
- Are headings, footnotes, quotations, tables etc. properly arranged?
- Is the style of writing precise; simple and direct?

Generally an evaluator is expected to have in his mind the following criteria for evalution:

(i) Social relevancy of the theme

(ii) Scope of new contribution
(iii) Actual contribution of the study
(iv) Soundness of methodology of mastery of the subject matter
(v) Survey of past literature
(vi) Volume and size of the work
(vii) Language format and presentation

13.7 Summary

This lesson describes in detail the structure, contents and style of writing the research report. The use of style manuals - MLA and APA in citing references within the running text and in the list of references has been discussed with illustrations. It also throws light on criterion of research evaluation.

13.8 Expected Questions

1. Discuss the structure and contents of a research report in detail.
2. Describe in detail the guidelines for research reporting.
3. Discuss MLA or APA Style Manual for citing references in the running text and in the list of references.
4. Discuss in detail the criteria for evaluating research project / report.

13.9 References

1. Kothari (CR) : Research Methodology : Methods & Techniques Ed 2. 2008.
2. Krishan Kumar. Research Methods in Library and Information Science, Delhi: Har Anand,1992.
3. Malhotra, N. K. Marketing Research: An Applied Orientation. 4th ed. Delhi: Pearson Education, 2004.
4. Totterndell, Anne. The Library and Information Work, London : Library Association, 2001.
5. Chattopadhyan, D.K. Research Methodology 2nded 2006.
6. Sharma, Jai Narain : Research Methodology New Delhi : Deep & Deep 2009.

13.10 Self-Check Exercise

Note: Write the answers in the space given below each question and check your answers with the answers given at the end.

1. Name the parts of a research report
2. List five important guidelines for research reporting.
3. Cite the following book in the reference list as per MLA and APA Style Manuals.

The Global Information Society
By
William J. Martin
England
Aslib Gower
1995

4. List the questions which should be raised to evaluate the appropriateness of the problem chosen for research?

13.11 Answers to Self- Check Exercise

1. A research report consists of three parts:
   1. The preliminary section
   2. The text
   3. The reference material
2. Five important guidelines for research reporting are as under:
   1. The report must be attractive in appearance, neat, clean whether typed or printed.
   2. The layout should be well thought out and must be appropriate and in accordance with the objectives of the research problem.
   3. The presentation must fit to the special interest and need of the audience for whom it is intended.
   4. Aim at originality and independent thinking while writing the research report.
   5. Bibliography is included at the end of the report.
3. MLA

   Martin, William J. Global Information Society. England : Aslib Gower, 1995.

   APA

   Martin, William J. (1995). Global Information Society. England: Aslib Gower
4. The questions which should be raised to evaluate the appropriateness of the problem are as under:
   - Is the problem significant enough?
   - Is it clearly and concisely stated?
   - Is it broken down into specific questions and issues?
   - Is it properly delimited ?
   - Is the analysis of the problem logically sound?
   - Is the solution to the problem likely to prove a good contribution to knowledge?
   - Are specific questions raised: Hypotheses clearly stated?
   - Are assumptions, limitation and delimitations stated?
   - Is review of related literature covered ?

Type Setting : Deptt. of Distance Education, Punjabi University, Patiala