RESEARCH METHODOLOGY

Prof. (Mrs.) Shraddha Bhome
M. Com., M. Phil. (Gold Medalist),
MBA (Finance), Research Scholar
Shri JJT University, C.S.

Dr. Nishikant Jha
Ph.D., ICWA, PGDBM (MBA), M. Com., DCA, DCP,
International Executive MBA UBI Brussels,
Belgium, Europe,
Assistant Professor in Account & Coordinator for
BAF Thakur College of Science & Commerce,
UGC Recognised, University of Mumbai.
Visiting Faculty for: MBA & PGDBM in Sikkim Manipal University,
CFA & CFP Professional Courses of USA,
CIMA Professional Courses of London,
CA, CS Professional Courses of India,
Ph.D. Guide [Research Supervisor] &
Professor for Research Methodology in JJT University.

Prof. Vinod Chandwani
M. Com., UGC – NET, M. Phil., B.Ed.,

Prof. (Ms.) Seethalakshmi Iyer
M. Com., CS (Inter), UGC – NET,
Assistant Professor.

Dr. (Prof.) Mrs. Swati Desai

Prof. (Mrs.) Archana Prabhudesai
M. Com., M. Phil. (Com), MBA (Marketing), B. Ed.,
Assistant Professor

Suryakant D. Koshti
M. Com., MBA, Ph.D.
Associate Professor and Head, Dept. of Commerce
Sheth NKTT College of Com., Thane (W).

Himalaya Publishing House

MUMBAI ● NEW DELHI ● NAGPUR ● BENGALURU ● HYDERABAD ● CHENNAI ● PUNE ● LUCKNOW
● AHMEDABAD ● ERNAKULAM ● BHUBANESWAR ● INDORE ● KOLKATA ● GUWAHATI

First Edition: 2013

Published by	:	Mrs. Meena Pandey for **Himalaya Publishing House Pvt. Ltd.,** "Ramdoot", Dr. Bhalerao Marg, Girgaon, **Mumbai - 400 004.** Phone: 022-23860170/23863863, Fax: 022-23877178 **E-mail: himpub@vsnl.com; Website: www.himpub.com**
Branch Offices	:	
New Delhi	:	"Pooja Apartments", 4-B, Murari Lal Street, Ansari Road, Darya Ganj, New Delhi - 110 002. Phone: 011-23270392, 23278631; Fax: 011-23256286
Nagpur	:	Kundanlal Chandak Industrial Estate, Ghat Road, Nagpur - 440 018. Phone: 0712-2738731, 3296733; Telefax: 0712-2721216
Bengaluru	:	No. 16/1 (Old 12/1), 1st Floor, Next to Hotel Highlands, Madhava Nagar, Race Course Road, Bengaluru - 560 001. Phone: 080-22286611, 22385461, 4113 8821, 22281541
Hyderabad	:	No. 3-4-184, Lingampally, Besides Raghavendra Swamy Matham, Kachiguda, Hyderabad - 500 027. Phone: 040-27560041, 27550139
Chennai	:	8/2 Madley 2nd street, T. Nagar, Chennai - 600 017. Mobile: 09320490962
Pune	:	First Floor, "Laksha" Apartment, No. 527, Mehunpura, Shaniwarpeth (Near Prabhat Theatre), Pune - 411 030. Phone: 020-24496323/24496333; Mobile: 09370579333
Lucknow	:	House No 731, Shekhupura Colony, Near B.D. Convent School, Aliganj, Lucknow - 226 022. Mobile: 09307501549
Ahmedabad	:	114, "SHAIL", 1st Floor, Opp. Madhu Sudan House, C.G. Road, Navrang Pura, Ahmedabad - 380 009. Phone: 079-26560126; Mobile: 09377088847
Ernakulam	:	39/104 A, Lakshmi Apartment, Karikkamuri Cross Rd., Ernakulam, Cochin - 622011, Kerala. Phone: 0484-2378012, 2378016; Mobile: 09387122121
Bhubaneswar	:	5 Station Square, Bhubaneswar - 751 001 (Odisha). Phone: 0674-2532129, Mobile: 09338746007
Indore	:	Kesardeep Avenue Extension, 73, Narayan Bagh, Flat No. 302, IIIrd Floor, Near Humpty Dumpty School, Indore - 452 007 (M.P.). Mobile: 09303399304
Kolkata	:	108/4, Beliaghata Main Road, Near ID Hospital, Opp. SBI Bank, Kolkata - 700 010, Phone: 033-32449649, Mobile: 7439040301
Guwahati	:	House No. 15, Behind Pragjyotish College, Near Sharma Printing Press, P.O. Bharalumukh, Guwahati - 781009, (Assam). Mobile: 09883055590, 08486355289, 7439040301
DTP by	:	HPH, Editorial Office, Bhandup (Asha)
Printed at	:	On behalf of HPH.

PREFACE

Syllabus

Credit Based Semester and Grading System w.e.f. the Academic year 2013-14

M.Com. (Semester III)

RESEARCH METHODOLOGY IN COMMERCE

Course Objectives

1. To understand research and research process
2. To acquaint students with identifying problems for research and develop research strategies
3. To familiarize students with the techniques of data collection, analysis of data and interpretation

Module I – Introduction to Research **(15 Lectures)**

Nature of Scientific Inquiry, Scientific Methods, Induction and Deduction. Hypothesis and Theory and their Interpretation. Nature and Scope of Social Research, Need for Multi-disciplinary and Inter-disciplinary Research

Module II – Defining Research Problems and Research Design **(10 Lectures)**

Planning of Research. Selection of a Problem for Research. Sample Size

Research Design. Important Aspects of Research Design

Module III – Methods and Techniques of Data Collection **(15 Lectures)**

Methods of Data Collection. Sources of Data Collection – Use of Secondary Data and Methods of Collecting Primary Data. Observation and Interviews. Questionnaires and Schedules

Question Paper Pattern

III-Semester End Examination

Marks: 60 Duration: 2 hours

Note: 1. All Questions are Compulsory with internal choice

2. Figures to the right indicate marks

Q.1	Answer Any One from the Following (Out of Two)	Module – I	(20 marks)
Q.2	Answer Any One from the Following (Out of Two)	Module – II	(20 marks)
Q.3	Answer Any One from the Following (Out of Two)	Module – III	(20 marks)

CONTENTS

Semester III

Semester IV

Chapter 1

Introduction to Research

Structure:

1.1 WHAT IS RESEARCH?

Research is undertaken within most professions. More than a set of skills, it is a way of thinking: examining critically the various aspects of professional work. It is a habit of questioning what you do, and a systematic examination of the observed information to find answers with a view to instituting appropriate changes for a more effective professional service.

When you say that you are undertaking a research study, to find answers to a question, you are implying that the process:

1. Is being undertaken within a framework of a set of philosophies (approaches);
2. Uses procedures, methods and techniques that have been tested for their validity and reliability;
3. Is designed to be unbiased and objective.

The word research is composed of two syllables, *re* and *search.*

- **re** is a prefix meaning again, anew or over again
- **search** is a verb meaning to examine closely and carefully, to test and try, or to probe.

Together they form a noun describing a careful, systematic, patient study and investigation in some field of knowledge, undertaken to establish facts or principles.

Research is a structured enquiry that utilizes acceptable scientific methodology to solve problems and create new knowledge that is generally applicable. Scientific methods consist of systematic observation, classification and interpretation of data.

Research means "Search for Knowledge". It aims at discovering the truth. It is the search for knowledge through objective and systematic method of finding solution to problems. It is carried on both for discovering new facts and verification of old ones. Therefore, research is a process of systematic and in-depth study or search of any particular topic, subject or area of investigation backed by collection, computation, presentation and interpretation of relevant data.

Research is, thus, an original contribution to the existing stock of knowledge making for its advancement. It is the pursuit of truth with the help of study, observation, comparison and experiment. In short, the search for knowledge through objective and systematic method of finding solution to a problem is research.

1.2 NATURE OR CHARACTERISTICS OF RESEARCH

Research is a process of collecting, analyzing and interpreting information to answer questions. But to qualify as research, the process must have certain characteristics: it must, as far as possible, be controlled, rigorous, systematic, valid and verifiable, empirical and critical.

1. **Controlled** - in real life there are many factors that affect an outcome. The concept of control implies that, in exploring causality in relation to two variables (factors), you set up your study in a way that minimizes the effects of other factors affecting the relationship. This can be achieved to a large extent in the physical sciences (cookery, bakery), as most of the research is done in a laboratory. However, in the social sciences (Hospitality and Tourism) it is extremely difficult as research is carried out on issues related to human beings living in society, where such controls are not possible. Therefore in Hospitality and Tourism, as you cannot control external factors, you attempt to quantify their impact.
2. **Valid and verifiable** - this concept implies that whatever you conclude on the basis of your findings is correct and can be verified by you and others.
3. **Empirical** - this means that any conclusions drawn are based upon hard evidence gathered from information collected from real life experiences or observations.
4. **Critical** - critical scrutiny of the procedures used and the methods employed is crucial to a research enquiry. The process of investigation must be foolproof and free from drawbacks. The process adopted and the procedures used must be able to withstand critical scrutiny.
5. **Rigorous** - you must be scrupulous in ensuring that the procedures followed to find answers to questions are *relevant, appropriate and justified.* Again, the degree of rigor varies markedly between the physical and social sciences and within the social sciences.
6. **Systematic** - this implies that the procedure adopted to undertake an investigation follow a certain logical sequence. The different steps cannot be taken in a haphazard way. Some procedures must follow others.

1.3 AIMS OF RESEARCH

The purpose of research is to discover answers to questions through the application of scientific procedures. The main aim of research is to find out the truth which is hidden and which has not been discovered as yet. Though each research study has its own specific purpose, we may think of research objectives as falling into a number of following broad groupings:

1. To gain familiarity with a phenomenon or to achieve new insights into it (Studies with this object in view are termed as exploratory or formulative research studies);
2. To portray accurately the characteristics of a particular individual, situation or a group (Studies with this object in view are known as descriptive research studies);
3. To determine the frequency with which something occurs or with which it is associated with something else (Studies with this object in view are known as diagnostic research studies);

4. To test a hypothesis of a causal relationship between variables (Such studies are known as hypothesis-testing research studies)

1.4 TYPES OF RESEARCH

The basic types of research are as follows:

(i) Descriptive vs. Analytical: Descriptive research includes surveys and fact-finding enquiries of different kinds. The major purpose of descriptive research is description of the state of affairs as it exists at present. In social science and business research we quite often use the term *ex post facto* research for descriptive research studies. The main characteristic of this method is that the researcher has no control over the variables; he can only report what has happened or what is happening. Most *ex post facto* research projects are used for descriptive studies in which the researcher seeks to measure such items as, for example, frequency of shopping, preferences of people, or similar data. *Ex post facto* studies also include attempts by researchers to discover causes even when they cannot control the variables. The methods of research utilized in descriptive research are survey methods of all kinds, including comparative and correlational methods. In analytical research, on the other hand, the researcher has to use facts or information already available, and analyze these to make a critical evaluation of the material.

(ii) Applied vs. Fundamental: Research can either be applied (or action) research or fundamental (to basic or pure) research. Applied research aims at finding a solution for an immediate problem facing a society or an industrial/business organisation, whereas fundamental research is mainly concerned with generalisations and with the formulation of a theory. "Gathering knowledge for knowledge's sake is termed 'pure' or 'basic' research." Research concerning some natural phenomenon or relating to pure mathematics are examples of fundamental research. Similarly, research studies, concerning human behavior carried on with view to make generalizations about human behavior, are also examples of fundamental research, but research aimed at certain conclusions facing a concrete social or business problem is an example of applied research. Research to identify social, economic or political trends that may affect a particular institution or copy research or the marketing research are examples of applied research. Thus, the central aim of applied research is to discover a solution for some pressing practical problems. Whereas basic research is directed towards finding information that has a broad base of applications and thus, adds to the already existing organized body of scientific knowledge.

(iii) Quantitative vs. Qualitative: Quantitative research is based on the measurement of quantity or amount. It is applicable to phenomena that can be expressed in terms of quantity. Qualitative research, on the other hand, is concerned with qualitative phenomenon, i.e., phenomena relating to or involving quality or kind. For instance, when we are interested in investigating the reasons for human behavior, we quite often talk of 'Motivation Research', an important type of qualitative research. This type of research aims at discovering the underlying motives and desires, using indepth interviews for the purpose. Other techniques of such research are word association tests, sentence completion tests, story completion tests and similar other projective techniques. Attitude or opinion research, i.e., research designed to find out how people feel or what they think about a particular subject or institution is also qualitative research. Qualitative research is specially important in the behavioral sciences where the aim is to discover the underlying motives of human behavior. Through such research we can analyse the various factors which motivate people

to behave in a particular manner or which make people like or dislike a particular thing. It may be stated, however, that to apply qualitative research in practice is relatively a difficult job and therefore, while doing such research, one should seek guidance from experimental psychologists.

(iv) **Conceptual vs. Empirical**: Conceptual research is related to some abstract idea(s) or theory. It is generally used by philosophers and thinkers to develop new concepts or to reinterpret existing ones. On the other hand, empirical research relies an experience or observation alone, often without due regard for system and theory. It is databased research, coming up with conclusions which are capable of being verified by observation or experiment. We can also call it as experimental type of research, in such a research it is necessary to get at facts firsthand, at their source, and actively to go about doing certain things to stimulate the production of desired information. In such *a research, the researcher* must first provide himself with a working hypothesis or guess as to the probable results. He then works to get enough facts (data) to prove or disprove his hypothesis. He then sets up experimental designs which he thinks will manipulate the persons or the materials concerned so as to bring forth the desired information. Such research is thus characterised by the experimenter's control over the variables under study and his deliberate manipulation of one of them to study its effects. Empirical research is appropriate when proof is sought that certain variables affect other variables in some way. Evidence gathered through experiments or empirical studies is today considered to be the most powerful support possible for a given hypothesis

(v) **Some Other Types of Research:** All other types of research are variations of one or more of the above stated approaches, based on either the purpose of research, or the time required to accomplish research, on the environment in which research is done, or on the basis of some other similar factor.

(a) ***One Time Research:*** From the point of view of time, we can think of research either as one-time research or longitudinal research. In the former case the research is confined to a single time-period, whereas in the latter case the research is carried on over several time-periods.

(b) ***Laboratory Research:*** Research can be *field-setting research or laboratory research or simulation research*, depending upon the environment in which it is to be carried out. Research can as well be understood as *clinical or diagnostic research*. Such research follows case-study methods or in-depth approaches to reach the basic causal relations. Such studies usually go deep into the causes of things or events that interest us, using very small samples and very deep probing data gathering devices.

(c) ***Exploratory Research:*** The research may be *exploratory* or it may be formalized. The objective of exploratory research is the development of hypotheses rather than their testing, whereas formalized research studies are those with substantial structure and with specific hypotheses to be tested.

(d) ***Historical Research:*** *Historical research* is that which utilizes historical sources like documents, remains, etc., to study events or ideas of the past, including the philosophy of persons and groups at any remote point of time.

(e) ***Conclusion-oriented Research:*** Research can also be classified as *conclusion-oriented* and decision-oriented. While doing conclusion-oriented research, a researcher is free to pick up a problem, redesign the enquiry as he proceeds and is prepared to conceptualize as he wishes. Decision-oriented research is always for the need of a decision maker and the researcher in this case is not free to embark upon research

according to his own inclination. Operations research is an example of decision-oriented research since it is a scientific method of providing executive departments with a quantitative basis for decisions regarding operations under their control.

1.5 PROCESS OF RESEARCH

However, the following order concerning various steps provides a useful procedural guideline regarding the research process:

(1) Formulating the research problem;
(2) Extensive literature survey;
(3) Developing the hypothesis;
(4) Preparing the research design;
(5) Determining sample design;
(6) Collecting the data;
(7) Execution of the project;
(8) Analysis of data;
(9) Hypothesis testing;
(10) Generalisations and interpretation, and
(11) Preparation of the report or presentation of the results, i.e., formal write-up of conclusions reached.

1. Formulating the Research Problem

There are two types of research problems, *viz.*, those which relate to states of nature and those which relate to relationships between variables. At the very outset the researcher must single out the problem he wants to study, i.e., he must decide the general area of interest or aspect of a subject matter that he would like to inquire into. Initially the problem may be stated in a broad general way and then the ambiguities, if any, relating to the problem be resolved.

Then, the feasibility of a particular solution has to be considered before a working formulation of the problem can be set up. The formulation of a general topic into a specific research problem, thus, constitutes the first step in a scientific enquiry.

Essentially two steps are involved in formulating the research problem, *viz.*, understanding the problem thoroughly, and rephrasing the same into meaningful terms from an analytical point of view. The best way of understanding the problem is to discuss it with one's own colleagues or with those having some expertise in the matter. In an academic institution the researcher can seek the help from a guide who is usually an experienced man and has several research problems in mind.

Often, the guide puts forth the problem in general terms and it is up to the researcher to narrow it down and phrase the problem in operational terms. In private business units or in governmental organisations, the problem is usually earmarked by the administrative agencies with which the researcher can discuss as to how the problem originally came about and what considerations are involved in its possible solutions.

2. Extensive Literature Survey

Once the problem is formulated, a brief summary of it should be written down. It is compulsory for a research worker writing a thesis for a Ph.D. degree to write a synopsis of the topic and submit it to the necessary Committee or the Research Board for approval.

At this juncture the researcher should undertake extensive literature survey connected with the problem. For this purpose, the abstracting and indexing journals and published or unpublished bibliographies are the first place to go to. Academic journals, conference proceedings, government reports, books, etc., must be tapped depending on the nature of the problem. In this process, it should be remembered that one source will lead to another.

The earlier studies, if any, which are similar to the study in hand, should be carefully studied. A good library will be a great help to the researcher at this stage.

3. Development of Working Hypotheses

After extensive literature survey, researcher should state in clear terms the working hypothesis or hypotheses. Working hypothesis is tentative assumption made in order to draw out and test its logical or empirical consequences. As such the manner in which research hypotheses are developed is particularly important since they provide the focal point for research. They also affect the manner in which tests must be conducted in the analysis of data and indirectly the quality of data which is required for the analysis. In most types of research, the development of working hypothesis plays an important role.

Hypothesis should be very specific and limited to the piece of research in hand because it has to be tested. The role of the hypothesis is to guide the researcher by delimiting the area of research and to keep him on the right track. It sharpens his thinking and focuses attention on the more important facets of the problem. It also indicates the type of data required and the type of methods of data analysis to be used.

How does one go about developing working hypotheses? The answer is by using the following approach:

(a) Discussions with colleagues and experts about the problem, its origin and the objectives in seeking a solution;
(b) Examination of data and records, if available, concerning the problem for possible trends, peculiarities and other clues;
(c) Review of similar studies in the area or of the studies on similar problems; and
(d) Exploratory personal investigation which involves original field interviews on a limited scale with interested parties and individuals with a view to secure greater insight into the practical aspects of the problem.

Thus, working hypotheses arise as a result of *a priori* thinking about the subject, examination of the available data and material including related studies and the counsel of experts and interested parties. Working hypotheses are more useful when stated in precise and clearly defined terms.

It may as well be remembered that occasionally we may encounter a problem where we do not need working hypotheses, especially in the case of exploratory or formulative researches which do not aim at testing the hypothesis. But as a general rule, specification of working hypotheses is another basic step of the research process in most research problems.

4. Preparing the Research Design

The research problem having been formulated in clear-cut terms, the researcher will be required to prepare a research design, i.e., he will have to state the conceptual structure within which research would be conducted. The preparation of such a design facilitates research to be as efficient as possible yielding maximal information. In other words, the function of research design is to provide for the collection of relevant evidence with minimal expenditure of effort, time and money. But how all these can be achieved depends mainly on the research purpose. Research purposes may be grouped into four categories, *viz.*,

(i) Exploration,
(ii) Description,
(iii) Diagnosis,
(iv) Experimentation.

A flexible research design which provides opportunity for considering many different aspects of a problem is considered appropriate if the purpose of the research study is that of exploration. But when the purpose happens to be an accurate description of a situation or of an association between variables, the suitable design will be one that minimises bias and maximises the reliability of the data collected and analysed.

5. Determining Sample Design

All the items under consideration in any field of inquiry constitute a 'universe' or 'population'. A complete enumeration of all the items in the 'population' is known as a census inquiry. It can be presumed that in such an inquiry when all the items are covered no element of chance is left and highest accuracy is obtained. But in practice this may not be true.

Even the slightest element of bias in such an inquiry will get larger and larger as the number of observations increases. Moreover, there is no way of checking the element of bias or its extent except through a resurvey or use of sample checks. Besides, this type of inquiry involves a great deal of time, money and energy. Not only this, census inquiry is not possible in practice under many circumstances.

For instance, blood testing is done only on sample basis. Hence, quite often we select only a few items from the universe for our study purposes. The items so selected constitute what is technically called a sample. The researcher must decide the way of selecting a sample or what is popularly known as the sample design.

6. Collecting the Data

In dealing with any real life problem it is often found that data at hand are inadequate, and hence, it becomes necessary to collect data that are appropriate. There are several ways of collecting the appropriate data which differ considerably in context of money costs, time and other resources at the disposal of the researcher.

Primary data can be collected either through experiment or through survey. If the researcher conducts an experiment, he observes some quantitative measurements, or the data, with the help of which he examines the truth contained in his hypothesis.

7. Execution of the Project

Execution of the project is a very important step in the research process. If the execution of the project proceeds on correct lines, the data to be collected would be adequate and dependable. The researcher should see that the project is executed in a systematic manner and in time. If the survey is to be conducted by means of structured questionnaires, data can be readily machine-processed. In such a situation, questions as well as the possible answers may be coded. If the data are to be collected through interviewers, arrangements should be made for proper selection and training of the interviewers.

The training may be given with the help of instruction manuals which clearly explains the job of the interviewers at each step. Occasional field checks should be made to ensure that the interviewers are doing their assigned job sincerely and efficiently. A careful watch should be kept for unanticipated factors in order to keep the survey as much realistic as possible. This, in other words, means that steps should be taken to ensure that the survey is under statistical control so that the collected information is in accordance with the pre-defined standard of accuracy. If some of the respondents do not cooperate, some suitable methods should be designed to tackle this problem. One method of dealing with the non-response problem is to make a list of the non-respondents and take a small sub-sample of them, and then with the help of experts, vigorous efforts can be made for securing response.

8. Analysis of Data

After the data have been collected, the researcher turns to the task of analysing them. The analysis of data requires a number of closely related operations such as establishment of categories, the application of these categories to raw data through coding, tabulation and then drawing statistical inferences.

The unwieldy data should necessarily be condensed into a few manageable groups and tables for further analysis. Thus, researcher should classify the raw data into some purposeful and usable categories.

Coding operation is usually done at this stage through which the categories of data are transformed into symbols that may be tabulated and counted. *Editing* is the procedure that improves the quality of the data for coding. With coding the stage is ready for tabulation.

Tabulation is a part of the technical procedure wherein the classified data are put in the form of tables. The mechanical devices can be made use of at this juncture. A great deal of data, especially in large inquiries, is tabulated by computers. Computers not only save time but also make it possible to study large number of variables affecting a problem simultaneously.

Analysis work after tabulation is generally based on the computation of various percentages, coefficients, etc., by applying various well defined statistical formulae. In the process of analysis, relationships or differences supporting or conflicting with original or new hypotheses should be subjected to tests of significance to determine with what validity data can be said to indicate any conclusion(s).

9. Hypothesis Testing

After analysing the data as stated above, the researcher is in a position to test the hypotheses, if any, he had formulated earlier. Do the facts support the hypotheses or they happen to be contrary? This is the usual question which should be answered while testing hypotheses.

Various tests, such as Chi square test, *t*-test, *F*-test, have been developed by statisticians for the purpose. The hypotheses may be tested through the use of one or more of such tests, depending upon the nature and object of research inquiry. Hypothesis testing will result in either accepting the hypothesis or in rejecting it. If the researcher had no hypotheses to start with, generalisations established on the basis of data may be stated as hypotheses to be tested by subsequent researches in times to come.

10. Generalisations and Interpretation

If a hypothesis is tested and upheld several times, it may be possible for the researcher to arrive at generalisation, i.e., to build a theory. As a matter of fact, the real value of research lies in its ability to arrive at certain generalisations. If the researcher had no hypothesis to start with, he might seek to explain his findings on the basis of some theory. It is known as interpretation. The process of interpretation may quite often trigger off new questions which in turn may lead to further researches.

11. Preparation of the Report or the Thesis

Finally, the researcher has to prepare the report of what has been done by him. Writing of report must be done with great care keeping in view the following:

1.The layout of the report should be as follows:

(i) The preliminary pages;

(ii) The main text,

(iii) The end matter.

In its preliminary pages the report should carry title and date followed by acknowledgements and foreword. Then there should be a table of contents followed by a list of tables and list of graphs and charts, if any, given in the report.

The main text of the report should have the following parts:

(a) Introduction: It should contain a clear statement of the objective of the research and an explanation of the methodology adopted in accomplishing the research. The scope of the study along with various limitations should as well be stated in this part.

(b) Summary of findings: After introduction there would appear a statement of findings and recommendations in non-technical language. If the findings are extensive, they should be summarised.

(c) Main report: The main body of the report should be presented in logical sequence and broken-down into readily identifiable sections.

(d) Conclusion: Towards the end of the main text, researcher should again put down the results of his research clearly and precisely. In fact, it is the final summing up.

At the end of the report, appendices should be enlisted in respect of all technical data. Bibliography, i.e., list of books, journals, reports, etc., consulted, should also be given in the end. Index should also be given specially in a published research report.

2. Report should be written in a concise and objective style in simple language avoiding vague expressions such as ‘it seems,’ ‘there may be’, and the like.
3. Charts and illustrations in the main report should be used only if they present the information more clearly and forcibly.

4. Calculated ‘confidence limits’ must be mentioned and the various constraints experienced in conducting research operations may as well be stated.

1.6 PURPOSE OF RESEARCH

1. Progress and Good Life

The purpose of all research is progress and good life. Progress results if the space of ignorance is occupied by knowledge and wisdom. The latter are the results of good research. Knowledge and wisdom drive the mankind to live an orderly good life.

2. Development of Scientific Attitude

One of the purposes of research is to develop scientific attitude. Scientific attitude is one that asks ‘Why’ and ‘How’ and answers are found. This ‘Know-why’ and ‘Know-how’ attitude nurtures talents and such intellectual talents are the great assets of society.

3. Creativity and Innovativeness

One of the purposes of research is encouragement to creativity and innovation. New products, new processes and new uses are the means through which the world goes dynamic. A dynamic world is not possible without newness introduced every now and then in every walk of life. And this is possible only through creativity and innovation. Research kindles the creativity and innovative instincts of people and thus experiments on the possibility of new things instead of waiting for the accidental and slow experience path to creativity and innovation.

4. Testing Hypothesis and Establishing Theories

A very important purpose of research is testing of hypothesis and establishing theories. As was already pointed out knowledge is power. That knowledge comes from testing hypotheses and establishing new theories. Proven hypotheses become theories.

5. Prediction and Control

Applied research has a great say in prediction and control in almost all walks of human endeavor Prediction is jumping into the future and the theories constitute the launch pad. Control looks for deviation between actual happening and predicted happening. In the process, the theories get reevaluated and redefined.

6. Purposive Development

Development = Growth + Change, Growth is uni-scaled while change is multi-scaled. In the natural process development does take place through trial and error through casual observations, through actual exposure and the like. But this is evolutionary and time consuming. Revolutionary development takes fourth through discontinuous change. Research is the seed of such dichotomous change or even disruptive change which contributes to purposive development.

7. Problem Solving

The purpose of any research is problem solving. What is a problem? Problem is deprivation or depreciation of something. Knowledge deprivation, efficiency deprivation, productivity depreciation, etc., exist. How can these be solved? Research into the forces that cause deprivation and measures to contain them from causing deprivation is needed. Thus, problem solving is a great purpose of research.

8. Schematic Evaluation

Research is also carried out to systematically evaluate a process or practice of an organisation to know its strengths and weaknesses so that areas for improvement process can be identified.

9. Impact Analysis

Research is undertaken to assess the impact of certain measures or change introduced on relevant variables. Impact studies are useful for biological, social, business, economic and other areas of decision making.

10. Methodological Improvement

Another purpose of research is improving research methodology itself. Developments in the field of measurement and scaling are immense. Whether these can be appropriately used in the case of particular research areas? To answer the question research needs to be done. Validation, revalidation and devalidation of methodological aspects thus constitute good piece of research. And this is one of the purposes of research. In fact, any research has a responsibility towards contribution to methodological enrichment.

1.7 QUESTIONS

1. Define Research and Explain it's Significance.
2. Explain the Nature of Research.
3. What do you mean by Research Methodology? Explain the process of Research.
4. Explain briefly different types of Research.

● ● ●

Chapter
2

Scientific Method

Structure:

2.1 INTRODUCTION TO SCIENTIFIC METHOD

To be termed scientific, a method of inquiry must be based on empirical and measurable evidence subject to specific principles of reasoning. The "scientific method" attempts to minimize the influence of the researchers' bias on the outcome of an experiment. The researcher may have a preference for one outcome or another, and it is important that this preference not bias the results or their interpretation. Sometimes "common sense" and "logic" tempt us into believing that no test is needed. Another common mistake is to ignore or rule out data which do not support the hypothesis.

The scientific method is the process by which scientists, collectively and over time, endeavour to construct an accurate (that is, reliable, consistent and non-arbitrary) representation of the world.

Recognizing that personal and cultural beliefs influence both our perceptions and our interpretations of natural phenomena, we aim through the use of standard procedures and criteria to minimize those influences when developing a theory. As a famous scientist once said, "Smart people (like smart lawyers) can come up with very good explanations for mistaken points of view." In summary, the scientific method attempts to minimize the influence of bias or prejudice in the experimenter when testing an hypothesis or a theory.

Clover and Basely define "scientific method is a systematic step by step procedure following the logical process of reasoning."

Difference between Scientific and non-scientific Method

The scientific approach to knowledge is empirical. The empirical approach emphasizes direct observation and experimentation as a way of answering questions. Intuition can play a role in idea formation, but eventually the scientist is guided by what direct observation and experimentation reveal to be true. Their findings are often counterintuitive.

When observing phenomena a scientist likes to exert a specific level of control. When utilizing control, scientists investigate the effects of various factors one by one. A key goal for the scientist is to

gain a clearer picture of those factors that actually produce a phenomenon. It has been suggested that tight control is the key feature of science. Non-scientific approaches to knowledge are often made unsystematically and with little care. The non-scientific approach does not attempt to control many factors that could affect the events they are observing (don't hold conditions constant). This lack of control makes it difficult to determine cause-and-effect relationships (too many confounds, unintended independent variable).

How can two people witness the same event but see different things? This often occurs due to personal biases and subjective impressions. These characteristics are common traits among non-scientists. Their reports often go beyond what has just been observed and involve speculation. Scientific reporting attempts to be objective and unbiased. One way to lessen the chance of biased reporting is checking to see if other independent observers report the same findings. Even when using this checkpoint the possibility of bias is still present. Following strict guidelines to prevent bias reporting decreases the chances of it occurring. Although I would say 100% unbiased reports rarely, if ever, occur.

It is not unusual for people in everyday conversation to discuss concepts they really don't understand. Many subjects are discussed on a routine basis even though neither party knows exactly what the subject means. They may have an idea of what they are discussing (even though their ideas may be totally opposite). Although they cannot precisely define the concepts they are talking about. The scientist attaches an operational definition (a definition based on the set of operations that produced the thing defined) to concepts. An example of an operational definition follows: hunger a physiological need for food; the consequence of food deprivation. Once an operational definition has been established communication can move forward.

2.2 IMPORTANT CHARACTERISTICS OF SCIENTIFIC METHOD

1. Empirical

Scientific method is concerned with the realities that are observable through "sensory experiences." Itgenerates knowledge which is verifiable by experience or observation. Some of the realities could bedirectly observed, like the number of students present in the class and how many of them are male and how many female. The same students have attitudes, values, motivations, aspirations, and commitments. These are also realities which cannot be observed directly, but the researchers have designed ways toobserve these indirectly. Any reality that cannot be put to "sensory experience" directly or indirectly (existence of heaven, the Day of Judgment, life hereafter, God's rewards for good deeds) does not fallwithin the domain of scientific method.

2. Verifiable

Observations made through scientific method are to be verified again by using the senses to confirm orrefute the previous findings. Such confirmations may have to be made by the same researcher or others. We will place more faith and credence in those findings and conclusions if similar findings emerge on the basis of data collected by other researchers using the same methods. To the extent that it does happen (i.e. the results are replicated or repeated) we will gain confidence in the scientific nature of our research. Replicability, in this way, is an important characteristic of scientific method. Hence revelations and intuitions are out of the domain of scientific method.

3. Cumulative

Prior to the start of any study the researchers try to scan through the literature and see that their study is not a repetition in ignorance. Instead of reinventing the wheel the researchers take stock of the existing body of knowledge and try to build on it. Also the researchers do not leave their research findings intoscattered bits and pieces. Facts and figures are to be provided with language and thereby inferences drawn. The results are to be organized and systematized. Nevertheless, we don't want to leave ourstudies as standalone. A linkage between the present and the previous body of knowledge has to beestablished, and that is how the knowledge accumulates. Every new crop of babies does not have to start from a scratch; the existing body of knowledge provides a huge foundation on which there searchers build on and hence the knowledge keeps on growing.

4. Deterministic

Science is based on the assumption that all events have antecedent causes that are subject toidentification and logical understanding. For the scientist, nothing "just happens" – it happens for areason. The scientific researchers try to explain the emerging phenomenon by identifying its causes. Of the identified causes which ones can be the most important? For example, in the 2006 BA/BSC examination of the Mumbai University 67 per cent of the students failed. What could be the determinants of such a mass failure of students? The researcher may try to explain this phenomenon and come up with variety of reasons which may pertain to students, teachers, administration, curriculum, books, examination system, and so on. Looking into such a large number of reasons may be highly cumbersome model for problem solution. It might be appropriate to tell, of all these factors which one is the most important, the second most important, the third most important, which two in combination are the most important. The researcher tries to narrow down the number of reasons in such a way that some action could be taken. Therefore, the achievement of a meaningful, rather than an elaborate and cumbersome, model for problem solution becomes a critical issue in research. That is parsimony which implies the explanation with the minimum number of variables that are responsible for an undesirable situation.

5. Ethical and Ideological Neutrality

The conclusions drawn through interpretation of the results to data analysis should be objective; that is, they should be based on the facts of the findings derived from actual data, and not on our own subjective or emotional values. For instance, if we had a hypothesis that stated that greater participation indecision making will increase organizational commitment, and this was not supported by the results, it makes no sense if the researcher continues to argue that increased opportunities for employee participation would still help. Such an argument would be based, not on the factual, data based research findings, but on the subjective opinion of the researcher. If this was the conviction of the researcher all along, then there was no need to do the research in the first place. Researchers are human beings, having individual ideologies, religious affiliations, cultural differences which can influence the research findings. Any interference of their personal likings and dis-likings in their research can contaminate the purity of the data, which ultimately can affect the predictions made by the researcher. Therefore, one of the important characteristics of scientific method is to follow the principle of objectivity, uphold neutrality, and present the results in an unbiased manner.

2.3 INDUCTION

From the Oxford English Dictionary (OED); to induce (in relation to science and logic) means "to derive by reasoning, to lead to something as a conclusion, or inference, to suggest or imply," and induction "as the process of inferring a general law or principle from observation of particular instances." Another version is the "adducing (pulling together) of a number of separate facts, particulars, etc. especially for the purpose of proving a general statement."

E. Mayr in his Growth of Biologic Thought offers this definition: "inductivism claims that (we) can arrive at objective unbiased conclusions only by…recording, measuring, and describing what we encounter without any root hypothesis…."

Meaning

It is one of the scientific methods. It follows the logical reasoning process. It is a process of reasoning whereby the researcher arrives at universal generalizations from particular facts. In other words, this method involves studying several individual cases and drawing a generalization. Thus, it involves two elements, i.e., observation and generalization. Conclusions drawn from induction and tentative inferences and they are subject to further confirmation based on more evidence.

Essential Conditions

According to Clover and Basely, four conditions are essential for valid induction. These are:

1. **Accuracy:** Observations must be correctly performed and recorded, and data collected should be accurate. Errors in observations, experiments or interviews and faulty recording of the information can affect the conclusions drawn.
2. **Representative:** Observations must cover representative cases drawn from a specific universe. For instance, to conduct a survey on brand loyalty among young females for a particular product category, then there should be proper representation of young females comprising of college going females, young working females, graduates, non-graduates, and other representations. Appropriate sampling technique can be used for this purpose, such as stratified random sampling.
3. **Appropriate Sample Size:** Observations must cover an adequate number of cases. The sample size must be large enough to make it representative so as to obtain reliable results.
4. **Proper Conclusions:** Conclusions must be confined to inferences drawn from the findings. The conclusions must be drawn only after proper analysis of the collected data relating to a particular study. They should not be generalized to apply to cases not covered in the sample. For instance, if the study is conducted to know the brand loyalty of young females for a particular product, then the conclusions drawn from such a study cannot be applied to young males, or elderly ladies, and so on, as they are not covered in the sample.

Merits

- Induction provides the universal premise and is helpful in finding out the material truth.
- It points out the relativity of generalizations. It indicates that a particular generalization is valid in certain situation.
- It is scientific in nature. Some of the important laws or principles of physical and social sciences have been developed through the use of inductive approach.

- Inductive generalizations are precise and accurate as they are based on observed facts and realistic foundations.

Demerits

- The collection of data for induction is a complex job. This method requires high degree of competence on the part of the researcher.
- Induction is not useful without deduction. Without deduction, induction produces only a mass of unrelated facts.
- It is a time consuming and expensive process; as a lot of time, effort and money is required to collect and analyze a large volume of data to arrive at generalization.

2.4 DEDUCTION

The OED definition of to deduce is "to show or hold a thing to be derived from etc…" or "to draw as a conclusion from something known or assumed, to infer"; deduction thus is "inference by reasoning from generals to particulars," or "the process of deducing from something known or assumed…"

Deductive reasoning is a basic form of valid reasoning. Deductive reasoning, or deduction, starts out with a general statements, or hypothesis, and examines the possibilities to reach a specific, logical conclusion. The scientific method uses deduction to test hypotheses and theories.

In deductive reasoning, if something is true of a class of things in general, it is also true for all members of that class. For example, "All men are mortal. Harold is a man. Therefore, Harold is mortal." For deductive reasoning to be sound, the hypothesis must be correct. It is assumed that the premises, "All men are mortal" and "Harold is a man" are true. Therefore, the conclusion is logical and true.

It's possible to come to a logical conclusion even if the generalization is not true. If the generalization is wrong, the conclusion may be logical, but it may also be untrue. For example, the argument, "All bald men are grandfathers. Harold is bald. Therefore, Harold is a grandfather," is valid logically but it is untrue because the original statement is false.

Essential conditions

The following are the essential conditions for valid deduction:

- The general premise or assumption must be correct. IF the general premise is correct, then the conclusion drawn can be right.
- The general premise must be applied only to the cases that properly come under it.

Merits

- Deduction is relatively simple and less time consuming as compared to inductive method.
- It is precise and accurate in generalization as it makes use of logic and mathematical tools of analysis.
- It social sciences, where there is limited scope for experimentation, this method becomes the only method for the development of generalizations.

Demerits

- It may result in inaccurate generalizations, if the premise are incorrect or partially correct.
- This method is abstract. Therefore, if a large amount of abstraction were used in generalizations, then such generalizations would be useless.

2.5 WHAT ARE THE SCOPE OF SCIENTIFIC METHODS?

Social science research has a vast scope in respect of areas of application. The social science research can be useful in a number of areas such as:

1. **Economic Planning:** Social science research can be of immense use in economic planning in a given society. Economy planning requires basic data on the various aspects of our society and economy, resource endowment and the needs, hopes and problems of the people, etc. Economic planning is undertaken to achieve certain objectives such as:
   - To bring about regional development.
   - To make optimum use of available resources.
   - To bring out self-reliance.
   - To generate employment, etc.

A systematic research provides the required data for planning and developing various schemes or programmes such as employment generation programmes, rural development programmes, etc.

2. **Control over Social Phenomena:** Through social science research, first-hand information can be obtained in respect of the working of institutions and organisation, which in turn provides greater power of control over the social phenomena. The social science research has practical implications for formal and informal styles of managing, organisation structures, and introduction of changes in the organisation.
3. **Social Welfare:** Social research can be used to collect the required data on different aspects of social life in a given society, so as to develop social welfare programmes. For instance, in a developing country like India, there are various social welfare problems such as low literacy, law and order problems due to caste, religion, and other conflicts, social evils like child marriages, abuse of women, and so on. Therefore, to overcome social problems the Government and other organizations can collect relevant data through a systematic research, and accordingly develop various social welfare programmes, such as family welfare campaigns, literacy programmes, women and children welfare programmes, etc.
4. **Helps to Solve Problems:** Research can be undertaken to find solutions to solve specific problems. For instance, an organization may initiate research to find solution to the problem of declining sales of their products in the market. An educational institution can undertake research to find out the causes of low attendance or poor results. A government organisation may undertake research to solve the problem or to ascertain the impact of slums on the quality of life in a particular city, and such other research activities. The research enables to find appropriate solutions to specific problems which in turn helps to improve the quality of performance in various organizations or institutions.
5. **Verifies and Tests Existing Laws:** Research may be undertaken to verify and test existing laws or theories. Such verification and testing of existing theories helps to improve the knowledge and ability to handle situations and events. This is true when the existing theories may not be sufficient or relevant to handle certain situations and events, and

therefore, through research, improvements or modifications can be made in the existing laws or theories.

6. **Develops New Tools and Theories:** Research helps to develop new tools, concepts and theories for a better study of an unknown phenomenon. For this purpose, exploratory research is undertaken to achieve new insights into such phenomenon.
7. **Helps to Predict Events:** Research may be undertaken to predict future course of events. For instance, research may be undertaken to find out the impact of growing unemployment of educated youth on the social life of the society in future. The findings of such research would not only indicate the possible impact, but would also make the concerned authorities to take appropriate measures to reduce unemployment, to reduce the growth of population and to overcome the negative consequences, as and when they take place.
8. **Extends Knowledge:** Researchers undertake research to extend the existing knowledge in physical sciences (such as physics, chemistry, mathematics, etc). as well as in social sciences (like sociology, management, psychology) etc. The knowledge can be enhanced by undertaking research in general and by fundamental research in particular.

2.6 DISTINCTION BETWEEN INDUCTION AND DEDUCTION

1. **Generalizations:** In induction, one arrives at universal generalization from particular facts. In deduction, one deduces generalizations from universal to particular facts.
2. **Material Truth:** Induction is concerned with the establishment of the material truth of universal propositions. Deduction is not concerned with the material truth of the premise.
3. **Certainty of Conclusions:** The conclusions of the inductive method are only probable and not always certain. The deductive method provides conclusion that are certain. This is because in induction method, conclusion is not implied in the premise, whereas, the conclusion in deductive method follows from the premise logically or it is implied in the premise.
4. **Observed Facts:** Induction is concerned with discovering facts and relations between them. Observed facts provide the basis for induction. The propositions from which deductions are made are assumed. In deductive method the observed facts are not relevant.
5. **Conclusion and Premise:** In the induction method, the conclusion goes beyond the premise or the contents of the data. The conclusion is more general than the premise. In deduction method, the conclusion only seeks to discover what is in the premise. It does not go beyond premise. The conclusion in deduction is never more general than the premise.

2.7 QUESTIONS

1. Explain the need of Scientific Method?
2. Explain the Scope of Scientific Method?
3. What do your mean by Scientific Method? Distinguish between induction and deduction?
4. Explain in brief Induction as a Scientific Method?

● ● ●

Chapter
3

Social Science Research

Structure:

3.1 INTRODUCTION

Like physical, biological and technological research, social research is one of the major fields of research. Social research includes social sciences, humanities and languages. Social research studies man and his institutions. It is largely empirical. It is bound to be interdisciplinary as the social phenomenon are not isolated.

In the modern context, every society is facing with serious social, economic, political and financial problems. The systematic and intelligent solution to these problems can be obtained through research.

Social research or social science research is the systematic method of discovering new facts, verifying old facts, their sequences, interrelationships, cause and effect explanation that governs them. It is the manipulation of things, concepts and symbols, with a view to extend new knowledge or verify old knowledge.

3.2 DEFINITIONS

(1) According to Prof. Rummel, "Social Science research is devoted to a study of mankind in his social environment and is concerned with improving his understanding of social orders, groups, instructions and ethics."

(2) According to Prof. M.H. Gopal, "Social Science is the scientific analysis of the nature and trends of social phenomenon or groups or in general of human behaviour so as to formulate broad principles and scientific concepts."

3.3 NATURE OF SOCIAL SCIENCE RESEARCH

(1) Social science research deals with the social phenomenon. It studies the behaviour of human beings. It covers the study of economic, political, educational, administrative and related aspects of social life.

(2) Social science research aims at discovery of new facts and verifications of old ones.

(3) Social science research tries to establish causal relationship between various human activities. It discovers the rules or laws so that they can be used in the guided growth of human society.

(4) Social science research involves the use of scientific methodology. It implies the development of new scientific tools, concepts, theories which would facilitate reliable and valid study of human behaviour.

(5) Social science research assists in the understanding of evolution of new theories. Every research highlights some broad principles, establishes some scientific truth, analyses their sequences.

(6) In social science research, human beings are the sources of data. So, their attitudes, environments, honesty, personal values and biases can change the research decisions.

(7) Social science research is utilitarian in nature. Here, the primary goal is to understand social life and thereby, gain greater knowledge which will be helpful for his project or to control over social behaviour.

(8) The purpose of social science research can be academic. It means to acquire first hand knowledge about the society. To know and understand the laws, to initimate knowledge of human society are the purposes of social science research.

3.4 NEED OF INTERDISCIPLINARY APPROACH IN COMMERCE

Social science research needs interdisciplinary approach in handling problems or situations. There is a need to blend various social sciences to deal with problems, i.e., social problems in a particular society. The interdisciplinary approach facilitates better understanding and management of the complex problems or situations facing a particular society.

The need for multidisciplinary and interdisciplinary approach in social science research can be explained below:

(1) Limitations of individual social science:

There are limitations of individual social sciences. Sociology as a discipline would alone not help to solve the problem of poverty. The problem of poverty has to be seen from the angle of economics, politics, etc.

(2) Interrelationships among social sciences:

There is a need for interrelationships among social sciences. For instance, to solve the problem of unemployment in rural areas, the Govt. may adopt certain employment schemes or projects. Such decision can be taken from economic point of view or social point of view or even from politics point of view.

Thus, there should be proper mix of two related social sciences.

(3) Need for objectivity in research findings:

There are certain problems or issues in social sciences, which can be dealt only through multidisciplinary or interdisciplinary approach. The problem must be studied with discipline. Then only it would affect the objectivity of conclusions drawn from such study.

Thus, overall to solve complexity of social phenomenon and for getting objectivity in research findings, there is a need to adopt interdisciplinary as well as multidisciplinary approach in research.

3.5 UTILITY OF SOCIAL SCIENCE RESEARCH

(1) Social Control: With the help of social science research, one can go for planning. A control over society is possible only when one has complete knowledge of the organisation and working of society and its various institutions. All these can be achieved only through a scientific study of society.

(2) Social Cohesion: The study of society creates better understanding and social cohesion between different groups. It reveals the underlying unity in the midst of disparity and creates a feeling of oneness.

(3) Social Welfare: Social welfare can be achieved through social research. Such research helps us to judge the magnitude of social evils and thus, take necessary steps to remove them.

(4) Social Prediction: Social science research helps the society in making predictions and goals for the future. It becomes possible because they have necessary data at their disposal. Statistical methods can be used for prediction.

(5) Social Growth: Social science research aims at opening new visitors of growth; knowledge, wisdom. The future path of social progress is conditioned by our knowledge of ourselves and other people. Social science research helps in guiding the trend of social growth on proper lines.

3.6 QUESTIONS

1. Explain the need for multidisciplinary and Interdisciplinary approach in research.
2. Explain the Significance of Social Science research.
3. Explain the nature of Social Science Research?
4. Write a note on Multidisciplinary approach in commerce.

● ● ●

Chapter 4

Planning of Research

Structure:

4.1 ESSENTIALS OF A GOOD RESEARCH PROBLEM

1. Question Mark?

The research problem can be in a declarative or in a question form. We recommend you to formulate your research problem as a question. This gives you (and the reader) something to hold on to during the rest of your thesis because it is simple: there is a question and in the text you look for an answer.

2. Possibility to Respond

Some questions are impossible to answer in a scientific way, for example: 'how beautiful is the color yellow'. We don't have the scientifically justified instruments to answer this question. It must also be possible to answer the question in a practical way so it must be researchable, meaning you have to be able to collect evidence that will answer the question.

3. Attainability

The problem must be one that can be solved during the amount of time you have. So it can't be too broad (ex: 'How can we have world peace?'). But it also can't be too narrow (ex: How does my neighbor think about Indians?').

4. Open Question

The research problem should be an open question. That means it cannot be answered by "yes" or "no". But also with open questions you should watch out for the possibility of a shallow answer.

5. Unmistakability

Your research problem must be clear and there has to be only one way to interpret it. For example: The question 'What do Indians think about the West?' is un mistakable because it is not clear what is meant by 'the West', it can be a lot of things.

6. Punctuality

The problem must be clearly specified. For example: Don't write 'How can prejudices against Americans be combated?' if you mean: 'How can prejudices that live among Indian students for Americans be combated?'

7. Brevity

Although your research problem should be as punctual and specific as possible, not all fencings must be placed in your research problem.

R.S. Woodworth defines problem as "a situation for which we have no ready or successful response by instinct or by previously acquired habit we have to find the answer."

According to John Dewey- "the need of clearing up confusion, of straightening out an ambiguity, of overcoming obstacles, of covering the gap between things as they are and as they may be when transformed is a problem." Thus, research is an enquiry geared to the solution of problem. Hence, the first step in any research is to make the problem concrete and explicit. A researcher should indentify some aspect of the topic which can be formulated into specific research questions. These should be capable of investigation with resources available to a researcher.

4.2 SOURCES OF RESEARCH PROBLEMS

A. Research Problem from Expert

The simplest source of a problem to solve is to have it given to you as a class assignment, as a directed research project, or as a task while you are an apprentice in someone's lab. You are told what problem to research and how to do it. This is probably an ideal way to assure that your first research topic is a good one.

Example: Students in Experimental Psychology were assigned the task of finding out if social attention made their roommate study more. They were told to measure the amount of time their roommate studied on day during which they expressed interest in their roommate's course material as compared to days when they refrained from talking about academic topics.

B. Research Problem from Folklore

Common beliefs, common sense, or proverbs could be right but on the other hand, they could also be wrong. You must verify that they are true before considering them as a source of knowledge. It is possible that some unverified beliefs have the roots of a better idea and therefore would be a worthy research topic. It is critical to note, however, that the task of research is not to simply validate or invalidate common sense but rather to come to understand nature.

Example: It's commonly believed that studying within the two hours preceding a test will decrease test scores. To research this belief a randomly selected half of a class was told to study immediately before taking a test while the other half was prohibited from studying before the test. This

research was intended to determine whether or not studying immediately before a test decreased the points earned.

C. Research Problem from Insight

Sometimes people research an issue simply because it occurred to them and it seemed important. The systematic development of the idea is lacking. This is "intuitive" or good guess research. It is risky because you may not be able to get other researchers to understand why the research is important. It is fun because you get to do what interests you at the moment. Alternatively, it could be the application of a general rule of thumb or guessing that a new problem is actually a well-understood function in disguise.

Example: While feeling especially competent after explaining course material to three friends you realize that orally presenting material may help test performance. You conducted a study in which material was orally presented before the test on a random half of the occasions. The research was based on your insightful realization that oral presentation may increase test performance.

D. Research Problem from Informal Discussion

This is a research problem that some discussion group feels is interesting. Discussion among friends can often spark our interest in a problem or provides us with the reinforces for pursuing a question.

Example: After telling a group of friends about your success with oral presentations on test taking, the group talks about it for awhile and becomes interested in the possibility of the subject becoming confused as well as doing better as a result of feedback from the listeners. The group provides you with the idea and the excitement to do research on how students can affect the accuracy of a teacher's understanding.

E. Research Problem from Knowledge of Techniques and Apparatus

This is the selection of a research topic based on your special knowledge outside the field. A technique or apparatus with which you are familiar can offer the potential for a major advance in the field of psychology. Sometimes we realize that we can apply a new technique or apparatus to an area to which it has not yet been applied. Because we are specially qualified to succeed, solving the problem can be especially gratifying.

Example: You may know about microelectronics and be good at detailed work. You find out that many researchers are anxious to discover the migration patterns of butterflies so you mount an integrated circuit transmitter on a butterfly and thereby trace the behavior of the free ranging butterfly.

F. Research Problem from Reading the Literature

These are research problems which capture your interest while reading. While reading you will often wonder why, or will disagree, or will realize that you have a better idea than the original author.

Example: While you were reading about jet lag and its effects on sleep the first night, you realize that the author failed to control for light cycle. You try stretching either the light period or stretching the dark period to make up the phase shift. You implement this by changing the cabin illumination period on various trans-Atlantic flights, and monitoring the passengers sleep for the next three days.

G. Research Problem Deduced from Paradigms or Theories

Researchers who propose theoretical accounts for phenomena cannot think through every possible ramification. As you come to understand a theory, potential errors or extensions become apparent. This type of research tests the implications of theories to confirm or reject them. This is classic deductive "normal" science. Using the object in the lake from the first chapter as an example – this would be deducing "if it is an steam shovel under there, then we should find a long row of high spots coming out of one end." You then test that prediction by probing around trying to find a boom.

If response strength approaches asymptotic response strength on each reinforced trial, then presenting a compound stimulus of asymptotically conditioned stimuli should result in a response decrement on subsequent tests with isolated stimuli.

4.3 FACTORS AFFECTING SELECTION OF RESEARCH PROBLEM

A problem for the purpose of study must be selected only after considering certain factors or criteria. The criteria can be broadly grouped into two groups:

(a) Internal factors
(b) External factors

(I) Internal Factors

The internal factors include personal interest of the researcher, competence of the researcher, and the resources available.

(a) Researcher's Interest: The researcher should be deeply interested in the problem. He should be determined to find solution to the problem. He must have the perseverance to arrive at effective solution. In other words, there should be complete devotion and dedication on the part of the researcher to deal with the research work.

(b) Researcher's Competence: It is not enough for the researcher to have dedication to research work, but there should be proper application or competence on the part of the researcher. The researcher should have the knowledge, ability and skills to handle the research activity.

(c) Availability of Resources: A more significant internal factor affecting problem selection is the availability of resources. The researcher should have adequate time and money to handle the research activity.

It is advisable on the part of the researcher to estimate properly the amount of funds that would be required to undertake the research work. If the funds are not available with the researcher, then he should make arrangement to obtain the funds from external sources, if possible. If sufficient funds were not available then it would not be possible to undertake the research work effectively. The researcher should not just see the availability of funds, but also must conduct a cost benefit analysis of undertaking such research activity. If the costs outweigh the benefits of the research, he may not proceed with such research problem or activity.

Apart from the availability of the funds, the researcher must also look into the availability of the time to undertake the research work. The researcher should complete the research work within a reasonable length of time. The time required for research work depends upon mainly on the nature of the problem. Therefore, proper time estimates must be undertaken. He may list out the time required for each and every research activity. He may focus on that research activity which consumes the maximum possible time, so that activity gets completed on time or ahead of time. In no case, the

researcher should cause unnecessary delays, which not only delay the research work, but at times, the entire research activity may be of no use, if there is delay in executing and completing the research work.

(II) External Factors

(a) Quality of Research Problem: The research problem should of substance. There is no sense to study a research problem if it would not serve any purpose. Also one should not waste one's time and efforts on a problem studied thoroughly by others. However, one may repeat an earlier study to validate the findings to different situations.

(b) Availability of **Facilities:** Research requires certain facilities such as appropriate library facilities, data processing facilities, etc. Therefore, the researcher must consider the availability of external facilities to make the research study effective.

(c) Social Relevance: The research should be socially relevant. It should serve some purpose to the society or to the organization that conducts the research. The factors of social relevance is especially important in the case of higher-level academic research and sponsored research.

(d) Research Personnel: At times, research activity requires a good deal of research personnel to undertake field interviews and other related activities. Therefore, the researchers should consider the availability of competent research personnel to assist him in research work. If necessary, the research personnel need to be provided with necessary guidance and training to undertake the research activities effectively.

(e) Urgency and Importance of the Problem: Business organizations and other institutions face a number of problems. They should list out the problems in the order of their urgency and importance. The more urgent and important problems should be researched first, as they are vital to the success of the organization and/or such problems if solved at the earliest may help to avert major crisis. Therefore, the researcher should give priority to more urgent and important problems. The less urgent and important decisions may be researched later.

(f) Feasibility of the Research: Most of all, the researcher should consider the feasibility of the research. He should find well in advance whether or not he would be able to:

- Collect the relevant data from the right sources.
- Obtain proper responses from the respondents.
- Get the required cooperation and guidance from concerned authorities in providing data or accessing the records. There are some organizations, such as private organizations, do not easily allow researchers to access their records.
- Complete the research activity within the available time.

4.4 HYPOTHESIS

Hypothesis is usually considered as the principal instrument in research. Its main function is to suggest new experiments and observations. In fact, many experiments are carried out with the deliberate object of testing hypotheses. Decision makers often face situations wherein they are interested in testing hypotheses on the basis of available information and then take decisions on the basis of such testing. In social science, where direct knowledge of population parameter(s) is rare, hypotheses testing is the often used strategy for deciding whether a sample data offer such support for a hypothesis that generalization can be made. Thus hypothesis testing enables us to make probability

statements about population parameter(s). The hypothesis may not be proved absolutely. But in practice it is accepted if it has withstood a critical testing. Before we explain how hypotheses are tested through different tests meant for the purpose, it will be appropriate to explain clearly the meaning of a hypothesis and the related concepts for better understanding of the hypothesis testing techniques.

Ordinarily, when one talks about hypothesis, one simply means a mere assumption or some supposition to be proved or disproved. But for a researcher hypothesis is a formal question that he intends to resolve. Thus a hypothesis may be defined as a proposition or a set of proposition set forth as an explanation for the occurrence of some specified group of phenomena either asserted merely as a provisional conjecture to guide some investigation or accepted as highly probable in the light of established facts. Quite often a research hypothesis is a predictive statement, capable of being tested by scientific methods, that relates an independent variable to some dependent variable. For example, consider statement like the following ones.

"Students who receive counseling will show a greater increase in creativity than students not receiving counseling" or

"The automobile A is performing as well as automobile B."

These are hypotheses capable of being objectively verified and tested. Thus, we may conclude that a hypothesis statement is what we are looking for and it is a proposition which can be put to a test to determine its validity.

4.5 FEATURES OF GOOD HYPOTHESIS

Hypothesis must possess the following characteristics:

(i) Hypothesis should be clear and precise. If the hypothesis is not clear and precise, the inferences drawn on its basis cannot be taken as reliable.

(ii) Hypothesis should be capable of being tested. Some prior study may be done by researcher in order to make hypothesis a testable one. A hypothesis "is testable if other deductions can be made from it which, in turn, can be confirmed or disproved by observation."

(iii) Hypothesis should state relationship between variables, if it happens to be a relational hypothesis.

(iv) Hypothesis should be limited in scope and must be specific. A researcher must remember that narrower hypotheses are generally more testable and he should develop such hypotheses.

(v) Hypothesis should be stated as far as possible in most simple terms so that the same is easily understandable by all concerned. But one must remember that simplicity of hypothesis has nothing to do with its significance.

(vi) Hypothesis should be consistent with most known facts i.e. it must be consistent with a substantial body of established facts. In other words, it should be one which judges accept as being the most likely.

(vii) The hypotheses selected should be amenable to testing within a reasonable time. The researcher should not select a problem which involves hypotheses that are not agreeable to testing within a reasonable and specified time. He must know that there are problems that cannot be solved for a long time to come. These are problems of immense difficulty that cannot be profitably studied because of the lack of essential techniques or measures.

(viii) Hypothesis must explain the facts that gave rise to the need for explanation. This means that by using the hypothesis plus other known and accepted generalizations, one should be able to deduce the original problem condition. Thus hypothesis must actually explain what it claims to explain, it should have empirical reference.

4.6 TYPES OF HYPOTHESIS

Social researchers have to work with many kinds of hypotheses. Hence, they can be classified in several ways.

(I) If hypotheses are classified in relation to their functions, we have two types:-

1. **Descriptive hypotheses:** These are propositions which describe the characteristics of a variable, e.g., income, expenditure, hours of work, size of public sector undertakings, etc. The variable may be an object, person, organization, situation or event. In these cases, the hypotheses asserts a particular characteristic. Statements about the rate of inflation during a given period, the number of unemployed in a certain region, the rate of growth of agricultural products in a certain state, the income level of certain class of people, etc., describe particular characteristics.
2. **Relational hypotheses:** These are propositions which describe the relationship between two variables. This type of hypotheses state that something is greater or less than something or one variable occurs in a certain proposition of time of another variable. For example, families with small income spend large proportion of their income on necessities, increase in standard of living tends to reduce in income helps in increasing expenditure, etc.

These statement suggests that the relationship between two variables can be direct or inverse, i.e., it can be either positive or negative. Some of these statements also indicate causal relationships, i.e., change in one variable is a cause or effect of a change in the other variable. The variable that causes the change is called independent variable, the other variable or the affected variable is called the dependent variable. The researcher must be careful in determining, which is the cause and which is the effect.

(II) Another approach is to classify hypotheses into three types:

1. **Working Hypotheses:** Hypotheses are formulated while planning the study of a problem. They may not be specific in the initial stages. In such cases, they are called "working hypotheses." Such hypotheses are subject to change or modification in the course of investigation.
2. **Null Hypotheses:** These are hypothetical statements and they deny what is stated in working hypotheses. They do not, and are not ever expected to, exist in reality. We discuss these type of hypotheses subsequently in the section on testing of hypothesis. There is some justification for using null hypotheses. They ensure detachment and objectivity in testing the hypothesis. Moreever, null hypotheses are more exact statements like "education does not increase the earning capacity of an individual" is a null hypothesis.
3. **Statistical Hypotheses:** These are statements about a statistical population. The statements are derived from a sample drawn from a given population. Statistical hypothesis are quantitative in nature as they are numerically measureable. These hypotheses can be hypotheses of difference or association, i.e., we can formulate them as null hypotheses or causal hypotheses.

(III) Classification of hypotheses on the basis of their level of abstraction is regarded as especially useful. Goode and Hatt have identified three broad levels of abstractions.

1. At the lowest level, we have simple description which gives rise to commonsense hypotheses. They state existence of certain empirical uniformities and hence expect verification of commonsense propositions, e.g., professionally qualified males marry at a late age, or a worker, in a small establishment lacks motivation. It is often said that such statements are not useful as hypotheses, since they merely state what everyone knows may be mistaken. Secondly, what everyone knows is not put in precise terms or is integrated in scientific terms. Commonsense statements are often a mixture of clichés and moral judgments. This is particularly true in social sciences. Scientists have to transform these statements and test them. It requires (a) removal of value judgments from the statements. (b) clarification of terms and (c) application of validity tests. According to Goode and Hatt, "what everybody knows is not known until it has been tested." Hence, such empirical generalizations play and important role in the growth of science.
2. At a relatively higher level of abstraction, we have logical derivations which give rise to complex hypotheses. These aim at testing the existence of logically derived relationships between empirical uniformities. They are purposeful distortions of empirical reality. Hence, they are also called "Ideal types." Such hypotheses try to create tools and problems for further research in complex areas of investigations.
3. The category of hypothesis at the highest level of abstraction is concerned with the relation of analytic variables. Hence, they are called analytical hypotheses. These are statements about how changes in one property will affect another property, e.g., statements about relation between level of education and migration, or level of income and social mobility are some such abstractions. These abstractions are highly sophisticated mode of formulation. They contribute to the development of brilliant abstract theories. However, this does not mean that these types of hypotheses is superior or better than the other types. Each type of hypotheses has its own importance. Use of a particular type of hypotheses depends on the nature of investigation.

4.7 ROLE OF HYPOTHESIS

In social science research, hypothesis serves several important functions.

1. A hypothesis guides the direction of study - A hypothesis guides the direction of study or investigation. It states what we are looking for.
2. Its purpose is to include in the investigation - Its purpose is to include in the investigation all available and pertinent data either to prove or disprove the hypothesis.
3. Research becomes unfocussed or random - Research becomes unfocussed or random without a hypothesis and useless data may be collected in the hope that important data is not omitted.
4. Hypothesis specifies the sources of data - Thus, a hypothesis specifies the sources of data, which shall be studied and in what context they shall be studied.
5. Needs and prevents a blind search - It also determines the data needs, and prevents a blind search.

6. Study a given problem - A hypothesis can suggest the type of research which is likely to be appropriate to study a given problem.
7. Appropriate technique of data analysis - It determines the most appropriate technique of data analysis.
8. Various hypotheses relating to a stated theory - A hypotheses can contribute to the development of theory by testing various hypotheses relating to a stated theory. It is also likely, in some cases, that a hypotheses helps in constructing a theory.

4.8 SOURCES OF HYPOTHESES

Meaning

Hypotheses may be developed from a variety of sources. Some of them are as follows:-

1. **Observation:** Hypotheses can be derived from observation. Relation between production, cost and output of goods or relationship between price variation and demand are hypothesised from observation.
2. **Culture:** A very important and major source of hypotheses is the culture or the socio-economic background in which a researcher has grown. Hypotheses regarding relationships between caste and family size, or income level and education level depend on the socio-economic background. In India, caste system plays an important role in determining socio-economic status, the same may not be the case of some other country.
3. **Analogies:** They are often a source of meaningful hypotheses, e.g., the hypotheses that similar human types or activities may be found occupying the same territory has come from plant ecology. Analogy is very suggestive. But one has to be careful in adopting models from other disciplines. Economic theory has adopted a few models from physics also.
4. **Theory:** Theory is an extremely fertile seed bed of hypotheses. A theory represents what is known and logical deductions from the theory lead to new hypotheses, which must be true if the theory is true, e.g., various hypotheses are derived from the theory based on profit maximization as the aim of a private enterprise. New hypotheses may be derived from the established theory by method of logical induction or logical deduction.
5. **Findings of other studies:** Hypotheses may also be developed from the findings of other studies. This can happen when a study is repeated under different circumstances or different time periods or for a different type of population. The findings of an exploratory studies may be formulated as hypotheses for other structured studies which aim at testing a hypothesis, e.g., the concept of trickle down effect of economic growth, later on becomes a testable hypotheses.
6. **Level knowledge:** An important source of hypotheses is the state of knowledge of any particular science. Hypotheses can be deducted from existing formal theories. If the hypotheses are rejected, the theory can be modified. If formal theories do not exist or are scarce which happens in case of new science, hypotheses are generated from formal conceptual framework. This leads to the growth of theory. The growth of statistical theory of sampling as also the development theories of economic growth illustrate this point. In either case, the hypotheses are related to the conceptual theoretical level.
7. **Continuity of research:** Continuous research in a field is itself an important source of hypotheses. The rejection of hypotheses leads to the formulation of new ones. These new

hypotheses explain the relationships between variables in the subsequent studies on the same subject.

In short, an ideal source of fruitful and relevant hypotheses is a fusion of two elements

(i) Past experience and

(ii) Imagination in the disciplined mind of the scientist.

4.9 QUESTIONS

1. What is Research problem? What are the sources of research problem?
2. What is hypothesis? What are the sources of hypothesis?
3. What is the Role of hypothesis?
4. Explain the factors affecting selection of research problem?

Chapter 5

Sampling

Structure:

5.1 IMPORTANT TERMS RELATED TO SAMPLING

(1) Population

In a statistical investigation the interest usually lies in the assessment of the general magnitude and the study of variation with respect to one or more characteristics relating to individuals belonging to a group. This group of individuals under study is called as population or universe.

So, population is defined as, "an aggregate of objects, animate or inanimate under study." It may be finite or infinite.

(2) Sample

A finite subset of statistical individuals in a population is called sample.

It is quite often used in our day to day life.

For example: Assessing the quality of foodgrain by taking a handful of it from the bag.

5.2 WHAT IS SAMPLING?

A Sampling is a part of the total population. It can be an individual element or a group of elements selected from the population. Although it is a subset, it is representative of the population and suitable for research in terms of cost, convenience and time. The sample group can be selected based on a probability or a non-probability approach. A sample usually consists of various units of the population. The size of the sample is represented by 'n'.

Sampling is the act, process, or technique of selecting a representative part of a population for the purpose of determining the characteristics of the whole population. In other words, the process of selecting a sample from a population using special sampling techniques called sampling. It should be ensured in the sampling process itself that the sample selected is representative of the population.

Though the sampling is not new but sampling theory has been developed recently. People know or not but they have been using the sampling technique in their day to day life.

For example:

(1) A housewife tests a small quantity of rice or wheat to see whether it is of a good quality and gives the generalised result about the whole rice kept in the bag or vessel. The result arrived at most of the times is 100% correct.

(2) When a doctor wants to examine the blood for investigation of any disease, takes only a few drops of blood of the patient and examines it. The result arrived at most of the times is correct and represent the whole amount of blood in the body of the patient.

In all these examples, by inspecting a few, they simply believe that the samples give a correct idea about the population. Most of our decisions are based upon the examination of few items only, i.e., sample studies. In the words of Croxton and Cowdon, "It may be too expensive or too time consuming to attempt either a complete or a nearly complete coverage in a statistical study. Further to arrive at a valid conclusions, it may not be necessary to enumerate all or nearly then decide to purchase it or not.

(3) **Sample size** – The number of individuals in a sample is called sample size.

(4) **Sampling Distribution –** The number of possible samples of size 'n' that can be drawn from a finite population of size N is N_{C_n} (If N is large or infinite then we can draw a large number of such samples). For each of these samples we can compute a statistic, say 't'… e.g… mean, variance, etc., which will obviously vary from sample to sample. The aggregate of the various values of the statistics under consideration so obtained, (one from each sample) may be grouped into a frequency distribution which is known as the sampling distribution of the statistic. Thus, we can have the sampling distribution of the sample mean $\overline{\chi}$, the sample variance, etc.

Or of a population. We may study a sample drawn from the large population and, if the sample is adequately representative of the populatic we should be able to arrive at valid conclusions.

According to Rosander, "The sample has many advantages over a census or complete enumeration. If carefully designed, the sample is not only considerably cheaper; but may give results which are just accurate and sometimes more accurate than those of a census Hence, a carefully designed sample may actually be better than a poorly planned and executed census.

5.3 ADVANTAGES/MERITS OF SAMPLING

(1) **It saves time:** Sampling method of data collection saves time because fewer items are collected and processed. When the results are urgently required, this method is very useful**.**

(2) **It reduces cost:** Since only a few and selected items are studied in sampling, so there is a reduction in cost of money and reduction in terms of man-hours.

(3) **More reliable results:** Through sampling more reliable results can be obtained because –

(a) There are fewer chances of sampling statistical errors. Though, any sampling error, it is possible to estimate and control the results.

(b) Highly experienced and trained persons can be: employed for scientific processing and analyzing of relatively limited data and they can use their high technical knowledge and get more accurate and reliable results.

(4) It provides more detailed information: As it saves time, money and labour more detailed information can be collected in a sample survey.

(5) Sometimes only method to depend upon: It is observed that sometimes one has to depend only upon sampling method. It happens when the population under study is infinite. In such situation it is only the method to be used.

For example: If someones blood has to be examined, it will become fatal to take all the blood out from the body and study depending upon the total enumeration method.

(6) Administrative convenience: The organization and administration of sample survey are easy for the same time, money and labour reasons.

(7) More scientific: Since the methods used to collect data are based on scientific theory and results obtained can be tested, sampling is a more scientific method to collect data.

5. 4 METHODS OR TECHNIQUES OF SAMPLING

There are two basic approaches to sampling: probabilistic and non-probabilistic sampling. Let us look at the various types of sampling under each category:

1. Probability Sampling

1. Simple random sampling
2. Systematic sampling
3. Stratified sampling
4. Multistage cluster sampling

2. Non-probability Sampling

1. Convenience sampling
2. Quota sampling
3. Judgment sampling
4. Snowball sampling

1. Probability Sampling

A sampling in which every member of the population has a calculable and non-zero probability of being included in the sample is known as probability sampling. Methods of random selection consistent with both the probabilities of inclusion are used in forming estimates from the sample. The probability of selection need not be equal for members of the population. If the purpose of a research is to arrive at conclusions or make predictions affecting the population as a whole, then the choice of a probabilistic sampling approach is desirable.

1. Simple Random Sampling

A sampling process where each element in the target population has an equal chance or probability of inclusion in the sample is known as simple random sampling. For example, if a sample of 15,000 names is to be drawn from the telephone directory, then there is equal chance for each number in the directory to be selected. These numbers (serial no. of names) could be randomly generated by the computer or picked out of a box. These numbers could be later matched with the corresponding names thus fulfilling the list. In small populations random sampling is done without replacement to avoid the instance of a unit being sampled more than once.

The benefits of simple random sampling can be reaped when the target population size is small, homogeneous, sampling frame is clearly defined, and not much information is available regarding the population. It is advantageous in that it is free of classification error, and requires minimum advance knowledge of the population. Two striking features are the elimination of human bias and non-dependency on the availability of the element. It is seldom put into practice because of the application problem associated with it. This sampling method is generally not preferred as it becomes imperative to list every item in the population prior to the sampling and requires constructing a very large sampling frame, resulting in extensive sampling calculations and excessive costs.

2. Systematic Sampling

Systematic sampling involves the selection of every kth element from a sampling frame. Here 'k' represents the skip interval and is calculated using the following formulae.

Skip interval (k) = Population size/Sample size

Often used as a substitute to simple random sample, it involves the selection of units from a list using a skip interval (k) so that every k'th element on the list, following a random start between 1 and k, is included in the sample. For example, if k were to equal 6, and the random start were 2, then the sample would consist of 2^{nd}, 8^{th}, 14^{th}, 20^{th} …….elements of the sampling frame.

It is to be noted here that if the skip interval is not a whole number then it is rounded off to the nearest whole number. This sampling method can be used in industrial operations where the equipment and machinery in the production line are checked for proper functioning as per the specifications. The manufacturer can select every k'th item to ensure consistent quality or for detection of defects. Therefore, he requires the first item to be selected randomly at the starting point and subsequently he can choose every k'th item for evaluation against specifications. It also finds its applicability while questioning people in a sample survey where the interviewer may catch hold of every 10^{th} person entering a particular shop. However, in every case, the researcher has to determine the skip interval and proceed thereafter. In both the cases, it is necessary to select the first item in the population in a random manner and thereafter follow the skip interval. This method is more economical and less time consuming than simple random sampling.

3. Stratified random sampling

Stratification is the process of grouping the members of the population in homogeneous group before sampling. It should be ensured that each element in the population is assigned a particular stratum only. The random sampling is applied within each stratum independently. This often improves the representativeness of the sample by reducing the sampling error.

The number of units drawn for sampling from each stratum depends on the homogeneity of the elements. A smaller sample can be drawn from the known to have the elements with the same value whereas sample can be drawn in much higher proportion from another stratum where the values are known to differ. This is because in the former case the information from the smaller number of respondents can be enumerated to the whole sample stratum. However in the latter case with much variability among the elements the higher elements value will keep the sampling to minimum errors to minimum value. The smaller errors may be due to groups are appreciably represented when strata are combined.

4. Multistage cluster sampling

Clustering involves grouping the population into various clusters and selecting few clusters for study. Cluster sampling is suitable for conducting research studies that cover large geographic area.

Once the cluster is formed the researcher can either go for one stage, two stages, or multistage cluster sampling. In single stage, all the elements from each selected are studied, whereas in two stages, the researchers use random to select few elements from clusters. Multistage sampling involves selecting a sample in two or more successive stages. Here the cluster selected in the first stage can be divided into cluster units.

For example consider the case where a company decides to interview 400 households about the likeability of its new detergent in a metropolitan city. To minimize the resources and time researchers divide the city into separate blocks say 40, each block consist of heterogeneous units. The researcher may opt for the two stage cluster sampling if he finds that individual clusters have little heterogeneity to other clusters. Similarly a multistage cluster sampling involves three or more sampling steps, it differs from stratified sampling that is done in cluster in contrast to elements within strata as is the case in the stratified sampling. Elements are randomly selected from each stratum in case of stratified sampling whereas only selected clusters are studied in cluster sampling.

2. Non-probability Sampling

It involves the selection of units based on factors other than random chance. It is also known as deliberate sampling and purposive sampling. For example, a scheme whereby units are selected purposefully would yield a non-random sample. In a general sense, it is an umbrella term, which includes any sample that does not conform to the requirements of a probability sampling. Convenience sampling, quota sampling, judgment sampling and snowball sampling are few examples of non-probability sampling.

1. Convenience Sampling

The selection of units from the population based on their easy availability and accessibility to the researcher is known as convenience sampling. For example, imagine a Co., that surveys a sample of its employees to know the acceptance for a new flavor of potato chips that it plans to introduce in the market. This type of sampling is a typical example of convenience sampling as the criterion for selecting a sample is convenience and availability. Although this type of research is easy and cost effective, the findings of the sample survey cannot be generalized to the entire population, as the sample is not representative. As there is no set criterion for selecting the sample, there is a scope for research being influenced by the bias of the researcher. As in the above ex, the researcher may conduct a sample survey involving its own employees to find whether the market, would accept the product.

2. Quota Sampling

In quota sampling, the entire population is segmented into mutually exclusive groups. The number of respondents (quota) that are to be drawn from each of several categories is specified in advance and the final selection of respondents is left to the interviewer who proceeds until the quota for each category is filled. Quota sampling finds extensive use in commercial research where the main objective is to ensure that the sample represents in relative proportion, the people in the various categories in the population, such as gender, age group, social class, ethnicity and region of residence. For example, if a researcher wants to segment the entire population based on gender, then he would have two categories of respondents, that is, males and females. If he plans to collect a sample of 30, he may allot a quota of 15 for male and 15 for female respondents. Therefore, the researcher will stop administering the questionnaire to females after he interviews the 15^{th} female respondent, that is, when the quota of 15 females is filled.

Quota sampling is subject to interviewer bias that may result in:

1. The quota reflecting the population in terms of superficial characteristics.
2. The researcher selecting the respondents based on availability rather than on their suitability to the study.

3. Judgment Sampling

The selection of a unit, from the population based on the judgment of an experienced researcher, is known as judgment or purposive sampling. Here, the sample units are selected based on population's parameters. It is often noticed that companies frequently select certain preferred cities during test marketing their products. This is because they consider the population of that particular city to be representative of the total population of the country. The same is the case with the selection of specific shopping malls that according to the researcher's judgment attract a reasonable number of customers from different sections of the society. Polling results predicted on television is also a result of judgment sampling. Researchers select those districts that have voting patterns close to the overall state or country in the previous year. The judgment of the researcher is based on the assumption that the past voting trends of selected sample districts are still representative of the political behavior of the state's population. For example, certain companies test market their new product launches in cities like Mumbai and Bangalore, because the profile of these cities is representative of the total Indian population.

4. Snowball Sampling

Sampling procedures that involve the selection of additional respondents are known as snowball sampling. This sampling technique is used against low incidence or rare populations. Sampling is a big problem in this case, as the defined population from which the sample can be drawn is not available. Therefore, the process sampling depends on the chain system of referrals. Suppose, SG sports Ltd., a manufacturer of sports equipment plans to survey 100 senior players through its new website for getting their feedback on the quality of its products.

However, keeping track of such senior senior squash players can be very difficult, as their presence may be very rare or low. Therefore, it collects the details of the first 200 visitors to its website, to list if any of them is a squash player or knows a squash player. If the visitor is a squash player, then he is requested to refer the names of at least 3 other players known to him. The referred names of the squash players are then called upon for further referrals and this goes on until the sample size of 100 adult players is reached. Although small sample sizes and low costs are the clear advantages of snowball sampling, bias is one of its disadvantages. The referral names obtained from those sampled in the initial stages may be similar to those initially sampled. Therefore, the sample may not represent a cross-section of the total population. It may also happen that visitors to the site or interviewers may refuse to disclose the names of those whom they know.

5.5 STEPS IN SAMPLE DESIGN

1. Defining the Target Population:

Defining the population of interest, for business research, is the first step in sampling process. In general, target population is defined in terms of element, sampling unit, extent, and time frame. The definition should be in line with the objectives of the research study. For example, if a kitchen appliances firm wants to conduct a survey to ascertain the demand for its micro ovens, it may define the population as 'all women above the age of 20 who cook (assuming that very few men cook)'. However this definition is too broad and will include every household in the country, in the population that is to be covered by the survey. Therefore the definition can be further refined and defined at the

sampling unit level, that, all women above the age 20, who cook and whose monthly household income exceeds Rs.20,000. This reduces the target population size and makes the research more focused. The population definition can be refined further by specifying the area from where the researcher has to draw his sample, that is, households located in Hyderabad.

A well defined population reduces the probability of including the respondents who do not fit the research objective of the company. For example, if the population is defined as all women above the age of 20, the researcher may end up taking the opinions of a large number of women who cannot afford to buy a micro oven.

2. Specifying the Sampling Frame:

Once the definition of the population is clear a researcher should decide on the sampling frame. A sampling frame is the list of elements from which the sample may be drawn. Continuing with the micro oven example, an ideal sampling frame would be a database that contains all the households that have a monthly income above Rs.20,000. However, in practice it is difficult to get an exhaustive sampling frame that exactly fits the requirements of a particular research. In general, researchers use easily available sampling frames like telephone directories and lists of credit card and mobile phone users. Various private players provide databases developed along various demographic and economic variables. Sometimes, maps and aerial pictures are also used as sampling frames. Whatever may be the case, an ideal sampling frame is one that entire population and lists the names of its elements only once.

A sampling frame error pops up when the sampling frame does not accurately represent the total population or when some elements of the population are missing another drawback in the sampling frame is over–representation. A telephone directory can be over represented by names/household that have two or more connections.

3. Specifying the Sampling Unit:

A sampling unit is a basic unit that contains a single element or a group of elements of the population to be sampled. In this case, a household becomes a sampling unit and all women above the age of 20 years living in that particular house become the sampling elements. If it is possible to identify the exact target audience of the business research, every individual element would be a sampling unit. This would present a case of primary sampling unit. However, a convenient and better means of sampling would be to select households as the sampling unit and interview all females above 20 years, who cook. This would present a case of secondary sampling unit.

4. Selection of the Sampling Method:

The sampling method outlines the way in which the sample units are to be selected. The choice of the sampling method is influenced by the objectives of the business research, availability of financial resources, time constraints, and the nature of the problem to be investigated. All sampling methods can be grouped under two distinct heads, that is, probability and non-probability sampling.

5. Determination of Sample Size:

The sample size plays a crucial role in the sampling process. There are various ways of classifying the techniques used in determining the sample size. A couple of those hold primary importance and are worth mentioning are whether the technique deals with fixed or sequential sampling and whether its logic is based on traditional or Bayesian methods. In non-probability sampling procedures, the allocation of budget, thumb rules and number of subgroups to be analyzed, importance of the decision, number of variables, nature of analysis, incidence rates, and completion rates play a major role in sample size determination. In the case of probability sampling, however, formulas are used to calculate the sample size after the levels of acceptable error and level of

confidence are specified. The details of the various techniques used to determine the sample size will be explained at the end of the chapter.

6. Specifying the Sampling Plan:

In this step, the specifications and decisions regarding the implementation of the research process are outlined. Suppose, blocks in a city are the sampling units and the households are the sampling elements. This step outlines the *modus operandi* of the sampling plan in identifying houses based on specified characteristics. It includes issues like how is the interviewer going to take a systematic sample of the houses. What should the interviewer do when a house is vacant? What is the recontact procedure for respondents who were unavailable? All these and many other questions need to be answered for the smooth functioning of the research process. These are guidelines that would help the researcher in every step of the process. As the interviewers and their co-workers will be on field duty of most of the time, a proper specification of the sampling plans would make their work easy and they would not have to revert to their seniors when faced with operational problems.

7. Selecting the Sample:

This is the final step in the sampling process, where the actual selection of the sample elements is carried out. At this stage, it is necessary that the interviewers stick to the rules outlined for the smooth implementation of the business research. This step involves implementing the sampling plan to select a sample required for the survey.

5.6 PRINCIPLES/ESSENTIALS OF SAMPLING

In order to reach to the clear conclusion, sampling should possess following essentials:

(1) It must be representative – The sample selected should possess the similar characteristics of the original universe from which it has been drawn.

(2) Homogeneity – Selected samples from the universe should have similar nature and should not have any difference when compared with the universe.

(3) Adequate samples: In order to have a more reliable and representative results, a good number of items are to be included in the sample.

(4) Optimization: All efforts should be made to get maximum results both in terms of cost as well as efficiency.

If size of the sample is larger, there is better efficiency and at the same time cost is more. A proper size of sample is maintained in order to have optimized results in terms of cost and efficiency.

5.7 PROCESS OF SAMPLE SURVEY

Following are the principal steps in the process of sample survey.

(1) Determining objectives of survey

The first step is to determine the objectives study in clear and concrete terms. It is generally found that the researcher many time, is not clear in mind regarding the purpose of the survey.

It is necessary to correlate the objectives with the available resources, i.e., men, money, machine and the time. It will help the researcher to achieve the desired goal in the determined time span with optimum use of resources.

(2) Decision about population

In this step, the population, i.e., the total number of objects (animate or inanimate) from which sample should be selected, is defined in clear and exact terms.

For example: A researcher wants to survey the brand preferences of households regarding detergent powder in Thane area of the city Mumbai.

In this example, a household is the sampling unit. The total of all households in Thane area is the population. Here, it is necessary to define the concept of population in an unambiguious manner.

In certain cases, no lists of population and no information about its nature are available, it is difficult to use a probability sampling method.

(3) Development of sampling frame and sampling units

After developing the population of the research problem, the sample frame is to be considered. Sampling frame is the list of population elements from which the sample is drawn. Ideally it should be a complete and correct list of population elements only.

The defined population must be capable of division into sampling units for purpose of sample selection. The sampling units must be able to cover the entire population and they must be distinct, correct, unambiguous and non–overlapping.

(4) Determination of data

The statistical data is the base of research work. It should be in line with the objectives of the survey. The researcher should focus on collection of the exact, relevant data related to the research problem.

Primary and secondary data sources should be analysed and the practical method should be developed. Wastages of valuable resources can be avoided by collecting exact information.

(5) Developing the sources of Collection of Primary data

After deciding the type of data is to be collected, researcher should focus on development of questionnaire. Questionnaire or schedules are useful to collect the primary data related to research problem.

The questionnaire should be concise, concrete, complete, clear and non-offending proper guidelines should be provided for filling up the questionnaire or schedule.

(6) Deciding about data collection method

The researcher gathers the data mainly from two sources, i.e.,

(a) Primary data source
(b) Secondary data source

(a) Primary data source includes -

(i) Observation
(ii) Interviewing
(iii) Mail survey
(iv) Experimention
(v) Simulation
(vi) Projective Technique

(b) Secondary data source includes -

(i) Reports of government departments
(ii) RBI Bulletins

(iii) Annual Reports on currency and finance published by RBI

(iv) Publications of international organisations like IMF

Any method of data collection should be aimed at -

(i) Accuracy

(ii) Cost factor

(iii) Time bound

(iv) Objective-oriented Approach

(7) Tackling the problem of Non-respondents

Many times, the entire data which is to be collected, is not available to the researcher. It happens mainly in primary data collection.

For example

The selected respondent hesitates or avoids giving certain information

Such non-respondents should be tackled with proper care to achieve valid and exact conclusions. Proper arrangements should be planned to deal with those who do not share the required data. The reasons behind it should be recorded by the researcher.

(8) Development of Proper Sampling Design

It is a critical task for the investigator. The success of the research work is dependent upon sample size, population parameters, sampling frame etc.

A researcher should focus on selection of the best sampling design which will achieve the desired objectives.

(9) Arrangements of field work

It many social science research problems field work is the foundation of the research study. To make the survey successful, there is a need of trained, experienced enumerators and supervisory staff.

(10) Examining at small level

To avoid mistakes, loopholes and wastage of resources, it is advisable to examine the functionality of questionnaire or schedule on a small-scale. It makes the process correct exact as well as effective.

It helps to disclose certain problems and mistakes that will otherwise create a big problem on a large-scale survey.

(11) Analysis and Interpretation of data

As soon as the data is collected, the researcher starts with the work of analysis of it. The raw data is of no use. It must be processed in proper manner. It can be done in following ways.

(a) Editing of data

(b) Classification of data

(c) Codification of data

(d) Tabulation of data

(All are explained in the Chapter Analysis of Data)

The classified and tabulated data passes through the stage of critical analysis, appropriate formulae should be used to arrive at final estimates of the required information. Proper care should be taken to keep the procedures error free.

This step also includes report writing. Report writing is the final step in the process of research work. Through report, the researcher presents the conclusions in detailed form. It is a very skillful, challenging task.

(12) Post-research Task

After completion of research work, it is the duty of the researcher to think upon the areas of weakness of the research. The mistakes occured, the error committed can serve as a potential guide for improved future sampling.

Any completed sample survey may be used as a base by the organisers for future reference in understanding and clearing of mistake in the execution of survey.

5.8 QUESTIONS

1. Define sampling? Discuss the Techniques of sampling
2. List out the steps in sample design.
3. What are the merits and limitations of sampling
4. Explain the process of sample survey.

● ● ●

Chapter
6

Research Design

Structure:

6.1 INTRODUCTION

Designing of the research is done mainly to solve the problem of getting the various stages of the research under control. This control factor is very important for the researcher during any of the research operation. Preparation of the design for the research forms a very critical stage in the process of carrying out some research work or a research project.

Research Design in general terms can be referred to as the scheme of work to be done or performed by a researcher during the various stages of a research project. With the help of the research design, one can very easily handle and operate research work as research design acts as a working plan, which is made by a researcher even before he starts working on his research project. By this, researcher gets a great help and guidance in achieving his aims and goals.

According to Russell Ackoff, "research design is the process of making decisions before a situation arises in which the decision has to be carried out. It is actually a process of deliberate anticipation directed towards bringing an unexpected situation under control."

6.2 MEANING AND DEFINITIONS

1. According to Trochim (2005), research design "provides the glue that holds the research project together. A design is used to structure the research, to show how all of the major parts of the research project work together to try and address the central research questions." The research design is like a recipe. Just as a recipe provides a list of ingredients and the instructions for preparing a dish, the research design provides the components and the plan for successfully carrying out the study.
2. According to Claire Seltiz, Research Design is a catalogue of the various facts relating to the formulation of a research effort. It is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure.
3. According to Paul E. Green and Tull, a Research Design is the specification of methods and procedures for acquiring the information needed. It is the overall operational pattern or

framework, of the project that stipulates what information is to be collected from which sources by what procedures. If it is a good design, it will ensure that the information obtained is relevant to the research questions and that it was collected by objective and economical procedures.

Research Design is the plan, structure and strategy of investigation conceived so as to obtain answers to research questions and to control variance. The plan is the overall scheme or programme of research. It includes an outline of what the investigator will do from writing the hypotheses and their operational implications to the final analysis of the data. To structure the research is to outline the scheme and paradigm of the operations of the variables strategy. It includes the methods to be used to gather and analyse the data. In other words, strategy implies how the research objectives will be reached and how the problems encountered in the research will be tackled. Like an architect prepares a blueprint before he approves a construction – in the same way researcher makes or prepares a plan or a schedule of his own study before he starts his research work. This helps the researcher to save time and also save some of his crucial resources. This plan or blueprint of study is referred to as the research design. A detailed outline of how an investigation will take place.

6.3 ESSENTIALS OF GOOD RESEARCH DESIGN

(**1**) **Reliability:** In general, reliability is concerned with the question of whether the results of a study are repeatable. It is an indication of the ability of a system to perform and maintain its functions consistently in routine circumstances as well as hostile or unexpected circumstances. Reliability is particularly important in quantitative research and may refer to:

- The statistical reliability of a set of data.
- The experimental reliability of an experiment.
- Data reliability, a property of some disk arrays in computer storage.
- Reliability engineering ensures a system will be reliable when operated in a specified manner.
- Reliability theory, as a theoretical concept, to explain biological aging and species longevity.
- Reliability (computer networking) is a category used to describe protocols.

(2) Replication: It is sometimes necessary for researchers to replicate (*i.e.* reproduce or duplicate) the findings of others; in order for this to happen, a study must be replicable. A study must be replicable in order that the reliability of a measure or a concept can be determined. Replications should not be confused with repeated measurements which refer literally to taking several measurements of a single occurrence of a specific phenomenon.

(3) Validity: Validity is concerned with the integrity of the conclusions that are generated from a piece of research. A valid measure is one which is measuring what it is supposed to measure. A valid measure must be reliable, but a reliable measure need not be valid. Validity refers to obtaining results that accurately reflect the concept being measured and it implies reliability (consistency). The main types of validity that are typically distinguished include:

- Measurement (or construct) validity, *e.g.*, does an IQ test really measure variations in intelligence?
- Internal validity, *e.g.*, if we suggest that $\underline{x}$ causes $\underline{y}$, can we be sure that it is $\underline{x}$ that is responsible for the variation in $\underline{y}$ and not something else?

- External validity, *e.g.*, can the results of a study be generalised beyond the specific research content?
- Ecological validity, *e.g.*, are social scientific findings applicable in people's everyday, natural social settings?

6.4 STEPS OF RESEARCH DESIGN

Following are the steps in research design:

1. **The Problem**: The first step involves the proper selection and then carefully defining the problem. By this researcher will be enabled to know about what he has to search, but it should be kept in mind that the problems selected should not be unmanageable in nature and should also not be based on desires.
2. **Objective of the Study**: The objective should be very clear in the mind of the researcher as this will lead to the clarity of design and proper response from the respondents.
3. **Nature of the Study**: The research design should be very much in relation with the nature of the study, which is to be carried out.
4. **Data Sources**: The various sources of the data or the information should be very clearly stated by the researcher.
5. **Techniques of Data Collection:** For the collection of the required information, it sometimes becomes very necessary to use some special techniques.
6. **Social Cultural Context**: Research design based on the social cultural concept is prepared in order to avoid the various study variations.
7. **Geographical Limit**: This step becomes a necessity at this point of time as with the help of this step, research linked to the hypothesis applies only to certain number of social groups.
8. **Basis of Selection**: Selecting a proper sample acts as a very important and critical Step and this is done with the help of some mechanics like drawing a random stratified, deliberate, double cluster or quota sample etc.
9. **Data Analysis:** Analysis of data is a process of inspecting, cleaning, transforming, and modeling data with the goal of highlighting useful information, suggesting conclusions and support decision making.
10. **Data Interpretation:** Data interpretation can be defined as "the application of statistical procedures to analyze specific observed or assumed facts from a particular study".
11. **Conclusions and Recommendations:** Conclusion means a position or opinion or judgment reached after consideration. On the basis of the research findings the conclusion needs to be drawn and suitable recommendations should be made to help improve the research problem.

6.5 EVALUATION OF RESEARCH DESIGN

The research design must be good. The question of good design is related to the purpose or objective of the research problem and also with the nature of it the problem to be studied.

A good design is often characterized by features like flexibility, appropriateness, economical and so on. Generally, the design which minimizes bias and maximizes the reliability of the data collected and analysed is considered a good design. The design which gives the smallest experimental error is supposed to be the best design in many investigations. Similarly, a design which yields maximal

information and provides an opportunity for considering many different aspects of a problem is considered most appropriate and efficient design in respect of many research problems. A design may be quite suitable in one case, but may be found wanting in one respect or the other in the context of some other research problem.

The fundamental questions in evaluating a research design pertain to the precision, reliability and relevance of the data and their analysis. Before actually carrying out research, it is better if the researcher evaluates his research design. This can be achieved if he verifies the following aspects for their explicitness.

- How relevant are the objectives?
- How relevant are the hypotheses?
- How explicit are the hypotheses?
- Have the problems and hypotheses been stated in operational terms scientifically?
- Has the research plan been presented in detail so that its logic is apparent?
- How scientific is data collection tool?
- How scientific is method of data collection?
- How precise are the observations?
- Can other investigators repeat the observations?
- Do the data actually satisfy the demands of the problem, i.e., do they actually demonstrate the conclusion?
- Does the research design ensure a comparison that is not subject to the alternate interpretations?
- Are the statistical designs appropriate?

6.6 QUESTIONS

1. Explain the Essentials of a good research design.
2. Explain the important steps involved in a research design?
3. Explain the process of Evaluation of a good research design.
4. Explain the significance of a good research design.

● ● ●

Chapter 7

Methods of Data Collection

Structure:

7.1 WHAT IS DATA?

Data can exist in a variety of forms as numbers or text on pieces of paper, as bits and bytes stored in electronic memory, or as facts stored in a person's mind. Strictly speaking, data is the plural of datum, a single piece of information. In practice, however, people use data as both the singular and plural form of the word.

Data that is accurate and timely, specific and organized for a purpose, presented within a context that gives it meaning and relevance, and can lead to an increase in understanding and decrease in uncertainty.

7.2 WHAT IS INFORMATION?

Information in raw or unorganized forms (such as alphabets, numbers, or symbols) that refers to, or represent, conditions, ideas or objects. Information is valuable because it can affect behaviour, a decision or an outcome. For example, if a manager is told his/her company's net profit decreased in the past month, he/she may use this information as a reason to cut their financial spending. A piece of information is considered valueless if, after receiving it, things remain unchanged.

Data v/s Information

Data is raw, unorganized facts that need to be processed. Data can be something simple and seemingly random and useless until it is organized. Whereas when data is processed, organized, structured or presented in a given context so as to make it useful, is called Information.

Each student's test score is one piece of data. Whereas the classes, average score or the school's average score is the information that can be concluded from the given data.

The terms data, information and knowledge are frequently used for overlapping concepts. The main difference is in the level of abstraction being considered. Data is the lowest level of abstraction, information is the next level, and finally, knowledge is the highest level among all three. Data on its own carries no meaning. For data to become information, it must be interpreted and take on a meaning. For example, the height of Mt. Everest is generally considered as "data", a book on Mt. Everest geological characteristics may be considered as "information", and a report containing practical information on the best way to reach Mt. Everest's peak may be considered as "knowledge".

7.3 TYPES OF DATA

1. SECONDARY DATA

Secondary data are data that are taken from research works already done by somebody and used for the purpose of the research data collection. The reason why secondary data are being increasingly used in research is that published statistics are now available covering diverse fields so that an investigator finds required data readily available to him in many cases. For certain studies like stock price behavior, interest and exchange rate scenario, etc., only secondary data are used.

2. PRIMARY DATA

Primary data means original data that has been collected specially for the purpose in mind. It means someone collected data from the original source firsthand. Data collected in this way is called primary data. Primary data is the data observed or collected directly from first-hand experience. Primary data is collected by the researcher himself. Primary data is original research data in its raw form, without any analysis or processing. This data provides a wealth of information for researchers. This data can contain results from empirical testing, transcripts of interviews and surveys, and recorded observations. A person conducting a study on mice, for example, would have primary data like test results from blood and urine analysis, along with detailed observations of the mice on a day to day basis.

7.4 SOURCES OF PRIMARY DATA

People can distinguish primary data from other kinds of data by the fact that it is directly collected and presented without commentary. Secondary data consists of things like research papers based on the data. The major disadvantage of primary data is the sheer volume of information. People would need to read through pages and pages of information to extract usable data. In data processing, researchers use statistics and other tools to present the data in a more accessible format, turning raw results into meaningful statements like “20% of study participants reported feeling nauseous.”

7.5 SOURCES OF PRIMARY DATA

1. Experiments

Experiments requires an artificial or natural setting in which one has to perform logical study to collect data. It is more suitable for medicine, nutrition and psychological studies. In experiment the experimenter has to keep control over the influence of any external variable on the results.

2. Survey

Surveying means contacting for getting certain information. **Survey method** is a method of collecting data for research purpose. There are personal surveys, mail surveys, telephone surveys and Internet surveys. Survey may be sample survey or census survey.

1. Personal Survey

Personal survey involves meeting personally every number who has to be surveyed. The features of this method of data collection are as follows.

- The number of respondents that can be contacted is not very high, as the time taken to contact the respondent, and the time spent on the interview itself is very high in relative terms.
- When the time available for research in large, the personal method is used.
- The cost involved is highest in the personal method since it requires field interviewers as well as their conveyance/travelling costs. Also, if a person is not available he may have to be contacted again and again.
- The accuracy obtained is very high, as the right persons are contacted and if there is difficulty in their understanding certain questions the interviewer can take care of it. Also, if the interviewer feels that the respondent is not furnishing the correct facts, by observing he can make his own interpretations, record the responses for better results. The responses rate is high compared to the mail survey, making the accuracy of the results better.
- When a large geographical area is to be covered and the time and cost constraints are high, personal survey method is not resorted to. However, if it is an ongoing syndicated research or census surveys, such in time and costs have to be necessarily borne for the sake of better coverage and accuracy.
- This method would require the agency to have a good infrastructure of data collection, in terms of field force, its supervision and control.
- When the literacy levels are low and the respondent would find it difficult to fill up the questionnaire on his own, this method is the best alternative available.
- A very length questionnaire under a structured survey is difficult to administer personally, with inaccuracy creeping in on account of the monotonous nature and fatigue effect on the part of the interviewer. At times, in certain socio-economic studies when this is unavoidable, the number of interviewers is made larger and certain incentives may be given to the respondents to extend their cooperation in filling up the set of questions, asked.
- The availability of skilled interviewers can reduce the interviewer bias on account of recording incorrect responses of the fatigue effects.
- When the questions require spontaneous answers, this is the best method, However, if the questions are of a personal nature or require too much thought on the part of the respondent he may feel embarrassed or make up the answers without thinking. However, in case of non-structured and non-disguised techniques like the in-depth interviews such probing is called for.
- The interviewer may have the tendency to contact some other person; similar to the respondent to complete his quota of respondents. This affects the accuracy of results thus necessitating a tight control on field work.
- A complete list of the respondents would be required to draw a representative sample. However, the interviewer has at times to use his own discretion and access respondents with similar demography characteristic in case of non-availability.

2. Mail Survey Method

Mail survey involves contacting the respondents through post. A questionnaire is dispatched to elicit views. The features of this method are as follows:

- A large number of respondents can be contacted using the available database of addresses.
- When the time available is again fairly large, and respondents are very distantly located this method is preferred.
- The cost involved is not very high and it is mainly just the cost of mailing which is often very cheap.
- The accuracy obtained is not very high since the response rate is not very high, not more than 20%. Also, the right respondent may not have filled up the questionnaire. Again, the chances of interpreting the questions incorrectly by the respondents are high, resulting in wrong responses which may be inconsistent with the responses in the rest of the questionnaire.
- The mail method most suited to contact people scattered over large geographical areas when the time available for contact is fairly enough An updated database of respondent addresses would go a long way in obtaining a good response rate. With the availability of fax, internet and courier facilities, time can be saved.
- The size of field force required is almost negligible and the agency could function with few data entry operators to take care of the data collected.
- The mail method can be used to contact only people who can read and write and definitely has limitations for social research at the slum level or rural levels.
- This is the method best suited when the questionnaire is very length. The respondents does not have to fill it up in one stroke. He also has time to think about the questions and answer them. Errors on account of fatigue or monotonous nature of the dialogue with the interviewer are not likely to creep in.
- The interviewer bias is not of any consequence in this case.
- The questions which require spontaneous answers would not lend themselves suitable to the mail survey. However, personal questions or those involving certain thought processes are best suited to the mail questionnaire method.
- The questionnaires are out of the control of the agency. The person who fills up the questionnaire may not be the desired respondent but someone in the same office or household.
- The mailing would not solve the purpose if the addresses of the respondents have not been updated. The existence of a mailing list is essential.

3. Telephone Survey Method

In telephone survey voice contact is directly established with the respondents. The features of this method are as follows:

- The number of respondents who can be contacted is fairly large, as the time to contact them is less than that for a personal interview.
- The cost involved is moderately high as skilled telephone operators need to be employed. Also if the respondent is not available he needs to be contacted more than once.
- The accuracy obtained is fairly high in this case as the response rate is comparable to the personally administered questions. In certain cases, it might be higher also. The skill of the

telephone interviewer makes the respondent at ease and comfortable to answer the questions. Also, certain questions, which need explanations can also be posed through the telephone.

- The telephone method can be used for respondents having the ability to communicate and express themselves. As such, only a certain class of respondents lend themselves useful to such methods.
- The length of the questionnaire has to be extremely short in this case. The issue which is addressed through the questions has to be focused one so that, less time is required to be spent over the telephone to explain the purpose of the research, etc. The telephone method would thus have a limited use.
- The skill of the operator/interviewer largely is responsible for the size of bias in recording the respondents incorrectly.
- The questions requiring spontaneous answers or of a personal nature can be administered over the telephone but those which require thinking would be difficult to take care of using this method.

4. **Internet Survey Method**

This is the world of connectivity through Internet. Internet survey involves using Internet for survey. The superiority of this survey is that it has no limitations of geography. The questionnaire may be put through a websites, forums, blogs, wikis, mail-groups, etc. The survey may be advertised through Internet or other mode inviting the attention of the prospective respondents. They may send their responses again over the Internet. This is the most modern type of survey and has great potentials. It is not prohibitively expensive.

3. Questionnaire

It is most commonly used survey method. It is list of questions, that is, either open-ended or closed-ended for which respondent gives an answer. It can be conducted via telephone, mail, face to face or other methods. A questionnaire is a form prepared and distributed to respondents to secure responses to certain questions. It is a device for securing answers to questions by using a form which the respondent fills by himself. It is a systematic compilation of questions that are submitted to a sample drawn from the population from which information is desired. It is an important instrument in normative survey research, being used to gather information from widely scattered sources. The questionnaire procedure normally comes into use where one cannot readily see personally all the people from whom one desires responses or where there is no particular reason to see them personally.

Purposes of questionnaire in research are twofold:

- To collect information from the respondents who are scattered in a vast area and
- To achieve success in collecting reliable and dependable data.

Types of Questionnaire

There are diverse forms of questionnaire used in research. These are discussed briefly here.

1. **Structured and Non-structured Questionnaires**: The structured questionnaire contains definite, concrete and direct questions, whereas non-structured questionnaire may consist of partially completed questions or statements. A non-structured questionnaire is often used as the interview guide, which is non-directive. The interviewer possesses only a blueprint of the enquiries and he is largely free to arrange the form or statements of the questions. The

enquiries framed in a general form beforehand are given a specific form during the actual process of interview.

2. **Closed Form and Open Form:** The questions that call for short or check responses are known as restricted or closed form type. This provides for making a yes or no, a short response, or checking an item from a list of given responses. It restricts the choice of response for the respondent. He has simply to select a response out of supplied responses and has not to frame his response in his own way. It is easy to fill out, takes less time, keeps the respondent on the subject, is relatively more objective, more acceptable and convenient to the respondent, and is fairly easy to tabulate and analyze. The open form, open-end or unrestricted type questionnaire calls for a free response in the respondent's own words. The respondent – frames and supplies his own response. No clues are provided. It probably provides for greater depth of response. The subject reveals his mind, gives his frame of reference and possibly the reasons for his responses. This type of item is sometimes difficult to interpret, tabulate and summarize in the research report. When the respondent is allowed freedom of response his expressions, may take any unique direction which may not find any uniformity with earlier responses.
3. **The Mixed Questionnaire:** The mixed questionnaire consists of both closed-end and open-end type questions. For social research, this method is very useful. Many questionnaires include both open and closed type items. Each type has its specific merits and limitations and the research worker has to decide which type is more likely to supply the information he wants.
4. **Fact and Opinion Questionnaires**: Questionnaire are also classified as: (1) Questionnaire of fact, which requires certain information of facts from the respondent without any reference to his opinion or attitude about them, and (2) Questionnaire of opinion and attitude in which the informant's opinion, attitude or preference regarding some phenomena is sought.
5. **Pictorial and Verbal Questionnaires**: In the pictorial questionnaire, pictures are used to promote interest in answering questions. It is used extensively in studies of social attitudes and prejudices in children or illiterate persons. In a pictorial questionnaire, the selected alternative answers in the form of pictures are given and the respondent is required to tick the concerned picture. This questionnaire may be very useful for collecting data in a developing country like India, specially from the rural masses who are mostly illiterate and less knowledgeable. The serious limitation of this questionnaire is that it is lengthy in form. Also it is highly expensive. Verbal questionnaire uses words and numbers only. It is the usual form meant for literate respondents.

In the questionnaire technique, great reliance is placed on the respondent's verbal report for data on the stimulus experiences in which he is exposed and for knowledge of his behaviour. The questionnaire is effective only when the respondent is able or willing to express his reactions clearly. A good questionnaire can elicit cooperation of the respondent to get frank answers on almost any subject, even such personal matters as sex and income. Thus, it is clear that the respondent can judge the study only by what he can see. The questionnaire, by its very nature, is an impersonal technique and it is several pieces of paper appeals/persuades the respondent that he ought to participate.

4. Interview

Interview is face to face conversation with respondents. It is slow, expensive and takes people away from regular work. Interviewer can not only record the statements the interviewee speaks but

also can observe the body language or non verbal communication such as face-pulling, shrugging, hand gestures that add further meaning to spoken words. **Interview** is one of the popular **methods of research data collection.** The term interview can be dissected into two terms as, 'inter' and 'view'. The essence of interview is that one mind tries to read the other. The interviewer tries to assess the interviewed in terms of the aspects studied or issues analyzed.

Types of Interview used in Research

There are different types of interviews used in the research data collection. An interview is either structured or unstructured, depending upon whether a formal questionnaire has been formulated and the questions asked in a prearranged order or not. An interview is also either direct or indirect as a result of whether the purposes of the questions asked are plainly stated or intentionally disguised. Cross-classifying these two characteristics provides four different types of interviews. That is, an interview may be

1. structured and direct,
2. unstructured and direct,
3. structured and indirect, or
4. unstructured and indirect.

Types (1) and (2) are basically objective types; (3) and (4) are subjective types.

1. **Structured-Direct Interview**: The usual type of interview conducted during a consumer survey to obtain descriptive information is one using a formal questionnaire consisting of non-disguised questions, a questionnaire designed to "get the facts". If the marketing search manager of a television set manufacturer wants to find out how many and what kinds of people prefer various styles of television cabinets, for example, he may have a set of questions drawn up that asks for these facts directly. Assuming that personal interviewing is being used, each interviewer will be instructed to ask the questions in the order given on the questionnaire and to ask only those questions. The resulting interviews will be structured-direct in nature.
2. **Unstructured-Direct Interview**: In the unstructured-direct method of interviewing, the interviewer is given only general instructions on the type of information desired. He is left to ask the necessary direct questions to obtain this information, using the warding and the order that seems most appropriate in the context of each interview. Unstructured-direct interviews are often used in exploratory studies. Many research projects that use a formal questionnaire for the final interviews go through an exploratory phase in which respondents are contacted and unstructured interviews are held. These interviews are useful in obtaining a clearer understanding of the problem and determining what areas should be investigated.
3. **Structured-indirect interview**: In the case of structured indirect interview the questions are pre-decided and arranged in a structured way. However the purpose of the study is not revealed.
4. **Unstructured-indirect interview**: In the case of unstructured indirect interview the questions aren't pre-decided and neither the purpose of the study made known explicitly.

There are **other types of interviews**, like focus-group interview, depth interview, etc. All these are dealt here.

(i) Focus-Group Interviews: Perhaps the best-known and most widely used type of indirect interview is the one conducted with a focus group. A focus-group interview is one in which a group of people jointly participate in an unstructured-indirect interview. The group,

usually consisting of 8 to 12 people, is generally selected purposively to include persons who have a common background or similar buying or use experience that relates to the problem to be researched. The interviewer, moderator, as he or she is more often called, attempts to focus the discussion on the problem areas in a relaxed, nondirected manner. The objective is to foster involvement and interaction among the group members during the interview will lead to spontaneous discussion and the disclosure of attitudes, opinions, information on present or prospective buying and use behavior.

(ii) **Focused Interviews:** This is a semi-structured interview where the investigator attempts to focus the discussion on the actual effects of a given experience to which the respondents have been exposed. It takes place with the respondents known to have involved in a particular experience, e.g., seeing a particular film, viewing a particular program on TV., involved in a train/bus accident, etc. The situation is analyzed prior to the interview. An interview guide specifying topics relating to the research hypothesis is used. The interview is focused on the subjective experiences of the respondent, i.e., his attitudes and emotional responses regarding the situation under study. The focused interview permits the interviewer to obtain details of personal reactions, specific emotions and the like. The merits of using this type of interview is that, it's free from the inflexibility of formal methods, yet gives the interview a set form and insures adequate coverage of all the relevant topics. The respondent is asked for certain information, yet he has plenty of opportunity to present his views. The interviewer is also free to choose the sequence of questions and determine the extent of probing.

(iii) **The Third Person Technique**: The simplest way of obtaining information through indirect questioning of a respondent is to ask for the view of a neighbor, an (unnamed) associate, or some other person whose views on the subject at hand might reasonably be known. This permits the respondent to project his own views with no feeling of social pressure to give an "acceptable" answer.

(iv) **The Depth Interview**: There is substantial use of the unstructured, informal interview in marketing research to explore the underlying predispositions, needs, desires, feelings, and emotions of the consumer toward products and services. This method of interviewing is referred to as a "depth interview". The depth interview in marketing research may consist of either direct or indirect questions, or some combination of the two. The skilled interviewer will generally employ both types of questions, A direct, free answer question such as "What are the major reasons why you bought your iPhone? Might well be followed up, for example, with an indirect question such as "Why do you think people who own smart phones bought them?" By following leads and cues provided by respondents, phrasing questions to continue the flow and pattern of the conversation and to maintain the rapport established, the competent interviewer can explore and probe the underlying motivations of the respondent.

(v) **The Personal Interview**: As the name implies, the personal interview consists of an interviewer asking questions of one or more respondents in a face to face situation. The interviewer's role is to get in touch with the respondent(s), ask the desired questions, and to record the answers obtained. The recording of the information obtained may be done either during or after the interview. In either case, it is a part of the interviewer's responsibility to ensure that the content of the answers is clear and unambiguous and that it has been recorded correctly.

(vi) **The Telephone Interview**: Telephone interviews are sometimes used in lieu of personal interviews, especially when the information must be collected quickly and inexpensively

and the amount of information required is limited. The telephone interview is well suited to such research problems as determining "coincidental" viewing of television or listening to radio programmes. In this type of study, calls are placed to a sample of telephone subscribers during the time the programme is on the air. The person receiving the call is simply asked "Are you now watching television?" and, if so, "What programme are you watching?" Other questions such as "How often do you watch this programme?" "Who sponsors this programme?" and the like may also be asked. The result is a rapid and inexpensive measurement of audience level. Either a structured or an unstructured interview may be held. Since the amount of information sought is usually well defined, non-confidential in nature, and limited in amount, virtually all telephone interviews are structured in nature. This medium does not lend itself well to indirect interviews and has not been used for this purpose.

5. **Observation:**

Observation involves three processes, i.e.,

(i) Sensation
(ii) Attention
(iii) Perception.

Sensation is gained through the sense organs which depend upon the physical alertness of the observer. The sense organs are receptive to stimuli and get attracted leading to the first stage in observation. Then comes attention or concentration which is largely a matter of commitment and will-power. Adequate training and experience can make it almost a matter of habit. The third is perception which comprises the interpretation of sensory reports. Thus, sensation merely reports the facts as observed but perception enables the mind to recognize the facts.

Through this process, observation serves the purpose of

(i) studying collective behavior and complex social situations.
(ii) following up of individual units composing the situations.
(iii) understanding the whole and the parts in their interrelations.
(iv) getting the out of the way details of the situation.

Types of Observation

There are different types of observation. The important ones are listed below:

1. **Casual and Scientific Observation**

   An observation may be either casual or scientific. Casual observation occurs without any previous preparation. It is a matter of chance that the right thing is observed at the right time and in the right place. Scientific observation, on the other hand, is carried out with due preparations and is done with the help of right tools of measurement experienced enumerators and under able guidance. Scientific observations yield thorough and accurate data.

2. **Simple and Systematic Observation**

   An observation may be either Simple or Systematic. Simple Observation is found in almost all research studies, at least in the initial stages of exploration. Its practice is not very standardized. It befits the heuristic nature of exploratory research. Participant studies are also usually classified as simple observation because participant roles do not permit systematic observation. Systematic observation, on the other hand, employs standardized

procedures, training of observers, schedules for recording and other devices to control the observer sometimes even the subject. Clearly some systematization is valuable in research observation, but the situation often limits what can be done. A systematic observation is a scientific observation too.

3. Subjective and Objective Observation

An observation may be either Subjective or Objective. In every act of observation there are two components namely, the object (or what is observed) and the subject (or the observer). It may be that sometimes one may have to observe one's own immediate experience. That is called Subjective Observation or Self-observation or introspection. Prejudices and biases are generally parts of subjective observation. Many data of psychological interest are gathered by the method of subjective observation. To avoid such prejudices, the observer takes stock of him and discovers what prejudices and biases will prevent impartial study and disinterested points of view. Persistent self-observation and criticism by others may ultimately overcome prejudice and biases. Such introspection may have another social value, i.e., it sensitizes the observer to the problems of others and creates sympathetic insight which facilitates, at least to some degree, the understanding of people's behavior in similar circumstances and similar cultural contexts. The net result is impartial subjective observation. When the observer is an entity apart from the thing observed, the observation of this type is objective.

4. Factual and Inferential Observation

Observation may be either factual or inferential. In factual observation things or phenomena observed with naked eyes are reported. In inferential observation behavior or psychological aspects are observed.

5. Direct and Indirect Observation

Observation may be either Direct or Indirect. In the case of direct observation the observer is physically present and personally monitors what takes place. This approach is very flexible of events and behavior as they occur. He is also free to shift places, change the focus of the observation, concentrate on unexpected events if they should occur. In indirect observation recording is done by mechanical, photographic or electronic means. For example a special motion picture camera which takes one frame every second is mounted in a department of a large store to study customer and employee movement.

6. Behavioral and Non-behavioral Observations

Observation may be either behavioral or non-behavioral. As pointed earlier the concept of observation involves not only watching but also listening and reading. Thus, observation includes the full range of monitoring behavioral and non-behavioral activities and conditions. Non-verbal analysis, linguistic analysis, extra-linguistic analysis and spatial analysis are the four major categories of behavioral observational study of persons. Record analysis, physical condition analysis and physical process analysis are the three major categories of non-behavioral study of persons. Non-verbal behavioral observation includes observation of body movement, motor expressions and even exchanged glances. Body movement, is an indicator of interest or boredom, anger or pleasure in a certain environment. Motor expressions such as facial movements can be observed as a sign of emotional studies. For instance, eye-blink rates are studied as indicators of interest in advertising messages. Finally, exchanged glances might be of interest in studies of interpersonal behavior Linguistic behavior is a second frequently used from of behavioral observation. One simple type, familiar to most students, is the tally of 'ahs' (or other annoying words or sounds) that a professor emits during a class.

7.6 SOURCES OF SECONDARY DATA

It is data that has been already collected by and readily available from other sources. When we use statistical methods with primary data from another purpose we are using secondary data. It means one purpose's primary data is another purpose's secondary data. So secondary data is data that is being reused. Such data are cheaper and more quickly obtainable than primary data.

Secondary data are data that are taken from research works already done by somebody and used for the purpose of the research data collection. The reason why secondary data are being increasingly used in research is that published statistics are now available covering diverse fields so that an investigator finds required data readily available to him in many cases. For certain studies like stock price behavior, interest and exchange rate scenario, etc., only secondary data are used.

Sources of secondary data

(1) Published printed sources: There are varieties of published printed sources. Their credibility depends on many factors like the writer, the publishing company, time and date of publication. New sources are preferred and old sources should be avoided as new technology and researchers bring new facts into light.

(2) Books: Books are available today on any topic that you want to research. Their use starts even before you have selected the topic. After the topic selection, books provide insights on how much work has already been done on the same topic and you can prepare the literature review.

(3) Journals/Periodicals: They are becoming more important as far as data collection is concerned. The reason is that journals provide up to date information which at times books cannot and secondly they can give information on very specific topic.

(4) Magazines and newspapers: Magazines are also effective but not very reliable sources. Newspapers on the other hand are more reliable.

(**5) Published electronic sources:**

As Internet is becoming more advance, fast and reachable to the masses, it has been seen that much information that is not available in printed form is available on internet. In the past the credibility of internet was questionable but today almost every journal and book are available online.

1. **e-Journals:** e-Journals are more commonly available than printed journals. Latest journals are difficult to retrieve without subscription but if there is e-library then one can view any journal any time.
2. **General Websites:** They are generally not very reliable hence their content should be checked for their reliability.
3. **Weblogs:** Weblogs are becoming common. They are actually diaries written by different people. These are as reliable as personal written diaries.

(6) Unpublished personal records:

Some unpublished data may also be useful in some cases.

1. **Diaries:** Diaries are personal records and rarely available. But if you are conducting descriptive research then they might be very useful. The Anne Franks diary is the most famous example of this. That diary contained the most accurate records of Nazi wars.
2. **Letters:** Like diaries letters are also rich sources of data.

3. **Government records:** Government records are very important for marketing, management and social researches. Examples are census data, health records, and educational institute's records.

7.7 PRIMARY DATA V/S SECONDARY DATA

1. Primary data is data which has been collected by you, which is more reliable and up to date. Secondary data has been collected from a secondary source (Other people, business etc.) so it may not be valid or up to date.
2. "Primary data" are data collected for the need at hand. "Secondary data" are data that were collected for another reason but is being re-purposed to address the need at hand.
3. When describing the expertise of data analysts, it is not uncommon to distinguish between primary and secondary data analytics. Primary data analytics involves the ability to analyze data for the purpose by which it has been collected. Secondary data analytics involves identifying "secondary data sources" to solve a new problem and then the ability to re-purpose that data.
4. **Primary data** is a data which is created for the first time and there is no previous source available. **Secondary data** is a readily available data like data from trade directories, statistics from websites etc. In Dissertation, literature review is done through secondary data which includes the contents such as theories, models, compilation, research findings by some other scholar etc.

7.8 APPROPRIATENESS OF METHODS OF DATA COLLECTION

The choice of appropriate data collection methods should be based on the research questions, design, sample, and the possible data sources. The technique used for data collection should gather information that will allow the research questions to be answered, take into account the characteristics of the sample, and provide information that is linked to each intended learning outcome.

7.9 ADVANTAGES AND DISADVANTAGES OF SECONDARY DATA

Following merits are usually claimed for using secondary data source.

1. **Provides an insight into total situation:** The purpose of use of available materials is to explore the nature of the data and the subjects to get an insight into the total situation. While looking for the data required by the researcher he may uncover many more available data than are often assumed to exists and hence, contributes significantly to the unfolding of hidden information.
2. **Helps in the formulation of hypothesis:** The use of documentary sources sometimes, helps in the formulation of research hypothesis. While an investigator may have one or two hypotheses which he might have deduced from theory, the study of available materials may suggest further hypotheses. If a research idea or hypotheses can be formulated in such a manner that the available recorded material bears on the question, the use of such material becomes possible.
3. **Helps in testing the hypotheses:** The available records may also help in testing the hypothesis.
4. **Provides supplementary information:** Available documents may be used to supplement or to check information gathered specifically for the purposes of a given investigation. For

example, if one has drawn in random sample of a small group in order to interview individuals, the accuracy of one's sample could be checked by comparing socio-economic data of the sample, like income, education standard, caste, family size, etc., with the same data of the most recent census or with available data in local Government offices.

The following are the demerits of using secondary data source for research purpose.

1. **Collected for a specific purpose:** Data are often collected with a specific purpose in mind, a purpose that may produce deliberate or unintentional bias. Thus, secondary sources must be evaluated carefully. The fact that secondary data were collected originally for particular purposes may produce other problems. Category definitions, particular measures or treatment effects may not be the most appropriate for the purpose at hand.
2. **Old data:** Secondary data are by definition, old data. Thus, the data may not be particularly timely for same purposes.
3. **Aggregation of data in inappropriate unit:** Seldom are secondary data available at the individual observation level. This means that the data are aggregated in some form, and the unit of aggregation may be inappropriate for a particular purpose.
4. **Authenticity:** The authenticity of same secondary sources of data is doubtful.
5. **Context change:** Secondary data refer to a given situation. As situations change, the data lose their contextual validity.

7.10 QUESTIONS

1. What are the Sources of collecting Secondary and primary data?
2. Explain methods of data collection?
3. Distinguish between primary and Secondary data?
4. Explain the advantages and disadvantages of Secondary data?
5. Explain the appropriateness of methods of data collection.

● ● ●

Chapter 8

Processing and Analysis of Data

Structure:

8.1 INTRODUCTION

A researcher collects huge amount of data with the help of different methods and tools. Data collection is an important stage in the research process. But mere collection of data does not help the researcher in any manner. It is necessary to process the data, to derive the desired results from it.

Hence, data processing is an important or even more important stage in the research work. The processing of data includes different stages like editing. Coding, classification and tabulation of data. Data processing is a connecting link between the data collection stage and the stage of analysis and interpretation.

8.2 STAGES OF PROCESSING OF DATA

The data processing is done in following four stages:

(1) Editing
(2) Coding
(3) Classification
(4) Tabulation

(1) Editing of Data

It is the basic stage in the data processing task.

Definition

(a) Editing refers to, "A process of checking to detect and correct errors and omissions."
(b) Editing is a process of examining the collected raw data to detect errors, inconsistencies and omissions if any, and to correct these wherever possible.

Types/Levels of Data Editing

(i) Field Editing

It includes reviewing the reporting forms by the investigator for completing or translating what the other has written in abbreviated form at the time of interviewing of respondents.

This form of editing deals with handling the errors or complexities made by the enumerator while collecting the data from respondent. It might be difficult to understand the data for researcher.

It is necessary to avoid guess work or recording by fluke, which may affect the reliability of the research work.

(ii) Central Editing

This type of editing is done at central place when all the questionnaires or forms are returned by the enumerators to the researcher.

The editing should be done according to the size of the research work. It means an individual can take care of editing in case of small study or a small group of persons can be appointed to edit a large study.

The work of editing is done in the following manner:

(1) Correction of errors such as a wrong entry
(2) Recording of the missing or inappropriate replies by referring the information recorded in the schedule.
(3) Deletion of incorrect replies from the schedule or questionnaire.

The data editing is also done at two levels.

(1) Data editing at the time of recording of data.

It includes following things to be done with proper care.

(i) Wheather the filters agree or are the data inconsistent.
(ii) Whether "missing values" been set to standardised values, which are same for all research questions.
(iii) Whether variable descriptions been specified or not.

At this stage, all editing and cleaning steps are documented, which helps to redefine the variables or modify the requirements in future.

(2) Data editing at the time of Analysis of data.

Data editing has a great role before starting the work of analysis. It ensures that the data is complete in all respects for further analysis. Following are few questions that can be had by a researcher for editing data before the beginning of analysis task.

(i) Is the coding frame complete?
(ii) Is the documentary material sufficient?
(iii) Is the number of cases correct?
(iv) Are there differences between questionnaire, coding frame and data?
(v) Are there any undefined or duplicate cases?

The success of the **Editing of data** is based upon following 3 Cs.

(a) Correctness:

Editing of data should focus not only on checking for omissions, but also for accuracy of each recorded reply.

The collected data should be free from following:

(i) Fake responses
(ii) Inconsistencies
(iii) Errors and omissions

A random check process, the cross verification will help to achieve the objective of correctness.

(b) Completeness:

To achieve the desired objective, the questionnaire or schedules must be complete in all aspects.

If there are any omissious, the researcher can fill the answers by refering the information given by respondents at some other places.

If the information is vital and has been found to be incomplete, then the researcher can contact the respondent and gather the requisite data again.

(c) Commonness:

It refers to uniformity or homogeneity. It aims at maintaining the consistency in recording throughout the questionnaire or schedule. The researcher should focus on avoiding the lack of uniformity in interpretation of questions and instructions by the data recorders.

(2) Coding of Data

In the stage of data collection huge data is collected by the researcher through questionnaire and schedule. It is necessary to assign certain numbers, signs to such data.

Definition

"The process of assigning some symbols, either alphabetical or numerals or both to the answers is known as coding."

The recording of data is done on the basis of this coding scheme. The responses colleted in a data sheet varies, sometimes the response could be the choice among a multiple response, sometimes the response could be in terms of values or it may be in alphanumeric form coding helps to bring uniformity which serves towards data analysis. When codification is done, it is necessary to keep a log of the codes alloted to the observations. This code sheet will helps in the identification of variables and the basis for such codification.

Advantages of Codification of Data

(1) It helps to reduce vast data into small signs or symbols.
(2) It aims at saving of time and space.
(3) Codification makes the data handling in soother way.
(4) The task of classification, analysis and interpretation becomes easy and simple with the help of codification.
(4) Codification or coding gives a concise or contracted form to data.

Types of Coding

(a) Numeric Coding:

It includes use of different numbers to indicate different data.

For example

'0' can be used for teenagers

'1' can be used for adults

'2' can be used for senior-citizens.

This will give a concise form to the vast data and it can be exhibited in a concise and convenient manner.

(b) Alphabetic Coding:

A mere tabulation or frequency count or graphical representation if the variables may be given an alphabetic coding.

(c) Zero Coding:

This method should be assigned with proper care. In many cases, the zero coding is used to record "no response" from the respondents.

But while utilizing it care should be taken for non allotment of "Zero" as a code for any other response.

The coding sheet needs to be prepared carefully if the data recording is done by someone else other than the researcher.

(3) Classification of Data

After the stage of editing and coding of data the next stage in the data processing is classification of data.

Definition

"The process of arranging the data in groups and classes according to resemblance and similarities which is technically called as classification."

In classification, units having common characteristics are placed in a class and in this fashion the whole data is deviled into a number of classes.

Unless and until the data is classified, it is unable to express or derive any conclusion. Classification gives the value of certainty to the data. It is a process of arranging the data according to the points of similarities or dissimilarities.

Objectives of Classification

(1) To condense the entire data is such a manner that the main features can be easily noticed.
(2) To compare the different attributes of variables.
(3) To prepare data which can be presented in fabular form
(4) To highlight the significant features of the data at a glance.

Types or Bases of Classification

(1) Qualitative Classification:

When the data is classified according to some attributes which are not capable of measurement. Such classification is known as qualitative classification.

For e.g. – Classification such as

(i) literate-illiterate,

(ii) Primary, secondary, higher secondary education,

(iii) Rich class, middle class, poor class etc.

(2) Quantitative Classification:

When the data is classified according to some characteristics that can be measured, it is called as quantitative classification.

For e.g. – Age, height, income, production, etc.

Quantitative data may be further classified into one or two types namely, discrete or continuous.

(i) Discrete data refers to quantitative data that is limited to certain numerical values of a variable.

For e.g. – Number of students in a college.

(ii) Continuous data considers all values of the variable. If considers integers.

For e.g. – Data relating to temperature, height, distance, etc.

(3) Geographical Classification

When the data is classification according to area or region, it is called as geographical classification.

For e.g. – Country, state, district, taluka, village, etc.

(4) Chronological classification

When the data is classified according to the time of its occurrence

For e.g. – Days, weeks, months, years, etc.

Essentials of a Good Classification

(1) It should be concise, precise and exact.

(2) The classification must be based upon uniformity

(3) It should be clear and unambiguous

(4) The classification should be flexible to accommodate changes if necessary

(5) It should be correct or accurate and should not include errors, omissions

(6) The bases of classification should be uniform and constant

(7) It must be capable of further analysis of data

(8) The classification should not be too small or too large in nature

(9) It should be favourable or conducive to the objectives of research work.

(4) Tabulation of Data

After the classification of data, they are presented in the form of tables so that the qualifies and characteristics implicit in them may be explicit. Classification and tabulation are interrelated and interdependent.

The data which is classified must be presented in the form of tables, then only it can be used in the research work.

Meaning and Definition

(a) Tabulation simply means presenting a data through tables. It is after the stage of classification in the process of statistical investigation.

(b) "Tabulation is an orderly arrangement of data in columns and rows."

(c) Tabulation means exhibiting or presenting the data in vertical and horizontal columns and rows. It includes preparing a chart or table of the data according to the attributes.

Tabulation can be done manually or through the computer. The decision depends upon different factors like, type of study, cost considerations, time pressures and the availability of software packages. Manual tabulation is useful for small and simple studies.

Tabulation is used for summarisation and condesation of data. It helps in analysis of relationships, trends and other summarisation of the given data.

Tabulation may be simple or complex in nature. Simple tabulation focuses on one-way tables. Whereas complex tabulation usually results in two way tables or three way tables. In tabulation, every row and column is assigned a suitable title, to present the data according to the convenience of research study.

Characteristics/Essentials of Tabulation/Table

(1) Each and every table should have an appropriate concise and exact title. This helps to recognise the data without reference to the text. Title should be given at the top of the table, some researchers prefer to put it below the table.

(2) Every table should carry distinct number to have an easy reference.

(3) Each table should have captions and stubs captions are the headings given to the columns whereas stubs are the row headings. Both the captions and stubs must be clear and brief.

(4) The table should help to conduct comparative study of the data presented in it.

(5) Every table should be simple, easy, neat to understand.

(6) The unit of measurements used must be indicated properly.

(7) Sources of the data in the table must be indicated at the bottom of the table. (If taken from any secondary source)

(8) The table should have thick or double dividing lines, in order to exhibit, different attributes, features or characteristics.

(9) Explanatory footnotes, if required, then they should be given exactly at the bottom of the table along with reference symbol.

(10) The table should not contain any abbreviation.

(11) The table should not be too large or too small. It must be proper to serve the need of the research work.

(12) The table should have a serial number, this helps for easy availability of it.

(13) All the colums and rows in the table should be serially numbered and should be given a concise and clear title.

(14) The table should be as logical, accurate, simple as possible.

(15) Arrangement of the colums in a table is based upon chronological, geographical or in alphabetical order.

(16) A rough framework of table helps before finalising the statistical table.

Types of Tabulation/Statistical Table

(1) Simple and Complex Tabulation

When a single attribute is taken into account while exhibiting the data, it is called as simple tabulation. It is always in small size.

For example:

(i) Tabulation which shows the division of data into two or more categories and gives information about one or more sets of inter related questions it is called as complex tabulation.

(ii) Simple tabulation- If a table exhibits rainfall in the Mumbai city in different years, it will be known as simple tabulation.

Example of complex tabulation – Preparation of census report.

(2) General Tabulation and Particular Tabulation

This tabulation does not aim at any particular single objective. General tabulation focuses to present or to show all available information in a systematic manner.

It is a broad-based type of tabulation. It is useful for comparative study of the available attributes.

For example: Statistical tables published by government, Reserve Bank of India Bulletins, etc. They only focus or show the facts but are not useful for any specific situation. They are mainly used by researchers to draw their own conclusions related to the research work.

Particular tabulation focuses on some particular objective. These tables are small in size serving mainly to the statical analysis or comparison of particular issues. They are also known as summary tables. They are mainly used in research reports.

(3) One - Way Table

It considers only one quantity or characteristic.

For example - Distribution of students on the basis of subjects of their study. Construction of one way table is easy and simple.

An example of one way Table:

Table 1.1 One Way Table

Education Level of Residents in a City

Sr. No	Levels of Education	Number of Citizen
1	Primary	
2	Secondary	
3	Higher Secondary	
4	Graduate	
5	Postgraduate	

6	Professional	
	Total	

(4) Two Way Table

It gives information about two interrelated characteristics of a particular phenomenon.

For example – The number of citizens given in previous table is divided on the basis of sex, the table would become a two way level as now it gives information about two characteristics, i.e.,

(i) Level of Education

(ii) Sex-wise distribution of citizens

Table 1.2 Two Way Table

Education Level of Residents in a City

Sr. No.	Level of Education	Number of males	number of Females	Total
1	Primary			
2	Secondary			
3	Higher Secondary			
4	Graduate			
5	Postgraduate			
6	Professional			
	Total			

In this table, citizens are classified on the basis of their educational level and sex.

(5) Three Way Table

It is prepared to focus on three interrelated characteristics of a given data.

Table 1.3 Three Way Table

Educational Level of the Residents in a City

(Sex-wise and on the basis of residence)

Number of residents										
Sr. No	Level of Education	Males			Females			Total		
		Rural	Urban	Total	Rural	Urban	Total	Rural	Urban	Total
1	Primary									
2	Secondary									
3	Higher Secondary									
4	Graduate									
5	Post-graduate									
6	Professional									
	Total									

(6) Manifold Table

If is prepared to cover number of interrelated characteristics related to main heads or subheads.

Such tables may be four way, five-way or more than that all such tables are called manifold tables.

Table 1.4 Manifold Table

Educational level of the residents of the city

(On the basis of sex, area of residence and martial)

Residence	Educational Level	No. of Residents								
		Male			Female			Total		
		Married	Unmarried	Total	Married	Unmarried	Total	Married	Unmarried	Total
Rural	1									
	2									
	3									
	4									
	5									
	6									
Urban	1									
	2									
	3									
	4									
	5									
	6									
Total	1									
	2									
	3									
	4									
	5									
	6									

The above table gives information about large number of interrelated questions.

Advantages of Tabulation

(i) It avoids the repeat heading and duplicate items.

(ii) It helps to correlate the terms and makes the comparison easier.

(iii) It saves time, space and energy of researcher.

8.3 ANALYSIS OF DATA

Analysis of data is considered as an important stage in the research work. Analysis is placing the collected data in some order or format so that the data acquire a meaning. Raw data become information only when they are placed in a meaningful form. Interpretation of the anlysed data involves drawing conclusions from the gathered data.

Definition – "Analysis of data means a critical examination of the assembled and grouped data for studying the characteristics of the object under study and for determining the pattern of relationships among the variables relating to it."

The most crucial aspects of research are the analysis and interpretation of data. It is considered to be highy skilled and technical job. It is to be carried out only by the researcher or under his close supervision. For analysis of data the researcher should require thorough knowledge of data, judgement, skills, ability of generalisation and familiarity with the background objectives and hypothesis of the study.

According to Willinson and Bhandarkar, "Analysis of data involves a number of closely related operations that are performed with the purpose of summarising the collected data and organizing these in such a manner that they will yield answer to the research questions or suggest hypothesis or questions."

Analysis aims at determining certain indices or measures along with searching for pattern of relationship that exists among the data group.

Objectives of Analysis/Purpose of Analysis

1. **Summary Form –** Statistical Analysis summarizes large man of data into understand able and meaningful form. This is the role of **descriptive statistics**. The reduction of data facilitates further analysis.
2. **Exact –** Analysis helps to make exact descriptions possible.

   For example: when we say that the employment level of people in a district is very high, the description is not exact; but when stistical measures like percentage of people employed in agriculture, industry or service sector are available, the description becomes exact.
3. **Identification of Causal factors –** Statistical analysis helps to identify the causal factors underlying complex phenomena. What are the factors that determine a variable like labour productivity or academic performance of the students? Answers to such questions can be identified from statistical multivariate analysis.

Process of Analysis

The following 3 main steps are to be performed, to start with the work of analysis.

1. **Review:** An intensive review must be made of all the data collection for project with reference to the research study's objectives and hypothesis.

   It is necessary to make the data conceptualised and tabulated for the aggregate group and provide a general statistical or quantitative description.
2. **Use of different modes of consolidation of data:** After tabulation of statistical data different modes of consolidating data for many seperate access is done. It is used to provide statistical descriptions.
3. **Examining the related phenomena:** All these data are examined in the context of related phenomeana so as to improve the evaluation of the findings.
4. **Use of qualitative attributes:** A variety of qualitative, non-statistical measures are used to supplement the quantitative description wherever necessary.
5. **Drawing of reliable inferences:** Statistical analysis aids the drawling of reliable inferences from observational data. Data are collected and analyzed in order to predit or make inferences about situations that not have been measured in full.

   What would be the growth rate of National. Income during coming year? What would be the probable demand for a particular product in the coming year? Questions of this kind require predictions of future states to be made on the basis of current knowledge. The statistical analysis helps to make such inferences.
6. **Helps to make estimations:** Statistical analysis helps to make estimations or generalisations from sample survey. Samples are based on probability samples which may give good estimates of particular population parameters.

7. **Assessing the significance of specific sample results:** This analysis is useful for assessing the significance of specific sample result under assumed population conditions. This type of analysis is called as **hypothesis testing.**

Approaches to Statistical Analysis

1. **Descriptive Analysis**

   This involves construction of statistical distributions and calculations of sample measures like averages, percentages and measure of dispersion for describing the features of the research queries.

   For example: If the respondents are individuals, then the according to descriptive analysis their distribution by gender, age, social class, income level, etc., should be shown.

2. **Comparison**

   According to this approach, the statistical analysis involves comparison of two or more distributions or two or more subgroups within a distribution.

   Here the use of measures of dispersion which include the range, the standard deviation and the coefficient of variation. It helps to compare the relative wideness of spread in any two or more frequency distribution.

   Other methods like ratios, proportions and percentages are also used for comparing distribution or subgroups.

3. **Study of Nature of Relationship among Variables**

   To know the effect of one or more independent variables on the dependent variable co-efficient of correlation, partial and multiple/correlation and regression analysis can be used.

   It also includes various parametric and non-parametric tests. Factor analysis is also an important way to find out the factors that explain relationship between a sets of variables.

Types of Statistical Analysis

(A) Descriptive Analysis

This is largely the study of distribution of one variable. This analysis is called one dimensional analysis. This analysis shows the benchmark data and measures the state or condition at any particular time. This analysis is the suitable type for bivariate and multivarlate analysis.

Sucha study provides profiles of companies, work group, persons and other subjects. This type of analysis may be in respect of one variable (undimensional), two variables (bivariate analysis) or more than two variables (multivariable analysis).

For example

The management of a public company wanted to study the pattern of monthly salary of its employers. The frequency distribution of the employees' salary is shown in the following table.

Frequency distribution of monthly salary of employers

Monthly Salary (Rs)	No. of Employes
1000 – 4999	45
5000 – 9999	25
10000 – 29999	16
30000 – 49999	14
50000 – 69999	10

70000 – 89999	4
90000 and above	2

Generally the method of descriptive analysis is used in historical research. Following are the important techniques which are used in "Descriptive Analysis."

(1) Bivariate Analysis

It is an analysis which is concerned with two variables. It places the collected data in to tabular form so that the real meaning of these data can be obtained. The starting point of a Bivariate analysis is to develop simple dimensional data. Then put the data into two or more categories. This technique change according to survey.

(2) Sequential Analysis

In this the data is presented in a table with one factor at a time and analysed thereafter.

3) Multivariate Analysis

It is the collection of methods for analysing in which dependent variable is represented in terms of several independent number of observations which are available to define such relationship.

It considers the various relationships among variables. It includes

(a) Multivariate Analysis of Variance -

This technique is appropriate when there is a single variable, which is a function of a number of other explanatory variables.

(b) Multi Regression Analysis –

This technique is appropriate when there is a single variable, which is a function of a number of other explanatory variables.

The main object of this analysis is, to predit the variability of the dependent variable based on its covariance with all the independent variables. The main advantage of multiple regression is that it allows to utilise more of the information available to us to estimate the dependent variable resulting in more accuracy in determining relationship.

(c) Canonical Correlation Analysis –

In this technique, simultaneous prediction is made to predict a set of criterion variables from the joint co-variance with a set of explanatory variable. This technique can be successively employed in case of both measurable and non-measurable variables for their joint co-variance with a set of independent variable.

(d) Multi Discriminate Analysis

It is considered as an important technique of classifying individuals or objects into one or two or more mutually exclusive, and exhaustive group on the basis of a set of independent variable. The discriminate analysis provides a predictive equation. This equation measures the relative importance of each variable.

(B) Causal Analysis

It involves the study of how one or more variable affect changes in another variable. It is a study of functional relationship existing between two or more variables.

Causal analysis is also termed as regression analysis. It is more important in experimental researches. It explains now one variable affects another. The causal analysis is done with the help of statistical tools.

(C) Correlative Analysis

It is concerned with two or more variables to determine correlation between two or more variables.

It studies the joint variation of two or more variables to determine the amount of correlation between two or more variables. In order to study, understand and control the relationships between the variable, correlation analysis is more important.

(D) Inferential Analysis

It is concerned with the various tests of significance for testing hypothesis to determine which validity data can be said to reveal some conclusion. It is also concerned with the estimation of population values. On the basis of inferential analysis, task of interpretation, i.e., the task of drawing inferences and conclusions is done.

8.4 INTERPRETATION OF DATA

After the collection and analysis of data, the researcher has to interpret the results from the analysis that he has done. This has to be done very carefully, otherwise misleading conclusions may be drawn and the whole purpose of doing research may get vitiated. It is only through interpretation that the researcher can expose relations and processes that underlie his findings.

Interpretation means drawing inference from the collected facts after the analytical study. Definition of According C. William Emory "Interpretation has two major aspects namely establishing continuity in research through linking the results of a given study with those of another and the establishment of some relationship with the collected data."

It is the device through which the factors, that seem to explain what has been observed by researcher in the course of the study can be better understood. Interpretation provides a theoretical conception which can serve as a guide for further analysis."

Need and Importance of Interpretation

(1) The interpretation helps the researcher to think on what his findings are and why: He can make others aware of the real importance of his research.

(2) Interpretation helps the researcher to understand the abstract principle behind his own findings. He can link his findings with those of other studies having the same abstract principle through interpretation.

(3) Interpretation opens new avenues of intelletual adventure and stimulates the quest for more knowledge through further research.

(4) Interpretation establishes the explanatory concepts that can serve as a guide for future research studies.

(5) Interpretation involved in the transition from explanatory to experimental research, because the interpretation of research study often results into forming hypothesis for new experimental research.

Errors of Interpretation

Errors may arise in the process off interpretation of data. Hence it is essential that the researcher should be very careful in the interpretation of data. Following are the reasons for errors of interpretation.

(i) False Generalisation

When the researcher draws conclusions about the whole by only studying a part of it, such type of mistakes arise. It is not always necessary that whatever is true of the particular part must be true of the whole also.

To make valid generalisation about the whole, if it is essential to know the movements recorded in various parts of the universe became there is poosibility that in one part, the movement may be in reverse direction.

It is essential to know what is the cause and what is the effect before drawing conclusions. Unless it is realised that it is due to wrong questions the findings will have wrong answers.

(ii) Wrong Interpretation of the Statistical Measures

Though the average is considered as a representative of the universe, it does not mean that it may be true for all items (the universe) sometimes interpretations are misleading and incorrect. The mistakes in the interpretation of data may also arise because of wrong interpretation of statistical measures such as mean, median, percentage, etc.

Effects of Wrong Interpretation

(1) Wrong Interpretation. Promotes distrust about statistics.
(2) The people may lose faith and confidence in statistics or look at them with suspicion.
(3) In many cases, good, bad and indifferent statistics get mixed up and lead to confusion.

Precautions in interpretation

(1) The use of inadequate statistical techniques, inadequate sample size and faulty calculations make the result spurious.
(2) The data should be adequately large and unbiased and should reflect good homogeny.
(3) Researcher must never lose sight of the fact that his task is to make sensitive observations of relevant assurances and to identify the hidden factors.
(4) The interpretation task needs more skill on the part of researcher. The quality of research does not depend on how much data has been collected but on how the researcher has interpreted it.

8.5 TRANSCRIPTION OF DATA

Transcription is a specific kind of data entry that means turning oral language into written form. In research process, when the observations collected by the researcher are not very large the simple inferences can be drawn from the observations and it is transferred to a data sheet. It is a summary of all responses on all observations from a research instrument.

The main aim of transcription process is to help in the process of presentation of all responses and observations on data sheets, which can help the researcher to arrive at preliminary conclusions as to the nature of sample collected, etc. Hence, transcription is a connecting link between data coding and data tabulation.

Methods of Transcription

(1) Manual Transcription

When the sample size is manageable, the researcher need not use any computerization process to analyse the data. In this situation, researcher could prefer a manual transcription and analysis of responses.

It is suitable only when the number of responses in the research work are very less. Which may be 10 responses; and the number of observations collected are within 100. A transcription sheet with 100 × 50 row/columns can be easily managed by researcher manually.

Hence, when the number of responses and number of observations are less, manual transcription is useful.

(2) Computerised Transcription

When the no. of observations and no. of responses are more, computerised transcription is useful.

This type of transcription is done using a data base package such as spreadsheets, text files or other databases.*

Important Concepts in Manual Transcription

(1) Use of Long Worksheets –

Generally it is used for manual transcription it requires' quality paper, preferably chart sheets, thick enough to last several usage.

These worksheets are normally ruled both horizontally and vertically, allowing a response to be written in the boxes. If one sheet is not sufficient, the researcher may use multiple ruled sheets to accomodate all observations.

The worksheets can be used for preparing summary tables or can be subject to further analysis of data.

Transcription can be made as and when edited instruments are ready for processing. Once all schedules have been transcribed, the frequency tables can be constructed straight from the worksheet.

8.6 QUESTIONS

1. What is editing? Discuss the types of editing
2. Define and Explain the meaning and types of classification.
3. List out the types of Tabulation
4. Discuss the process of Analysis
5. Enumerate different types of analysis
6. Explain Importance of Errors of interpretation

● ● ●

Chapter
9

Preparation of Report

Structure:

9.1 WHAT IS RESEARCH REPORT?

Research report is considered a major component of the research study for the research task remains incomplete till the report has been presented and/or written. As a matter of fact even the most brilliant hypothesis, highly well designed and conducted research study, and the most striking generalizations and findings are of little value unless they are effectively communicated to others.

The purpose of research is not well served unless the findings are made known to others. Research results must invariably enter the general store of knowledge. All this explains the significance of writing research report. There are people who do not consider writing of report as an integral part of the research process.

But the general opinion is in favour of treating the presentation of research results or the writing of report as part and parcel of the research project. Writing of report is the last step in a research study and requires a set of skills somewhat different from those called for in respect of the earlier stages of research. This task should be accomplished by the researcher with utmost care; he may seek the assistance and guidance of experts for the purpose.

9.2 TYPES OF REPORT

Research reports vary greatly in length and type. In each individual case, both the length and the form are largely dictated by the problems at hand. For instance, business firms prefer reports in the letter form, just one or two pages in length. Banks, insurance organisations and financial institutions are generally fond of the short balance-sheet type of tabulation for their annual reports to their

customers and shareholders. Mathematicians prefer to write the results of their investigations in the form of algebraic notations.

The reports can be prepared by governmental bureaus, special commissions, and similar other organisations are generally very comprehensive reports on the issues involved. Such reports are usually considered as important research products. Similarly, Ph.D. theses and dissertations are also a form of report writing, usually completed by students in academic institutions.

9.3 CONTENTS OF REPORT

The researcher must keep in mind that his research report must contain following aspects:

1. Purpose of study
2. Significance of his study or statement of the problem
3. Review of literature
4. Methodology
5. Interpretation of data
6. Conclusions and suggestions
7. Bibliography
8. Appendices

These can be discussed in detail as under:

(1) Purpose of Study

Research is one direction-oriented study. He should discuss the problem of his study. He must give background of the problem. He must lay down his hypothesis of the study. Hypothesis is the statement indicating the nature of the problem. He should be able to collect data, analyze it and prove the hypothesis. The importance of the problem for the advancement of knowledge or removal of some evil may also be explained. He must use review of literature or the data from secondary source for explaining the statement of the problems.

(2) Significance of Study

Research is re-search and hence the researcher may highlight the earlier research in new manner or establish new theory. He must refer earlier research work and distinguish his own research from earlier work. He must explain how his research is different and how his research topic is different and how his research topic is important. In a statement of his problem, he must be able to explain in brief the historical account of the topic and way in which he can make and attempt. In his study to conduct the research on his topic.

(3) Review of Literature

Research is a continuous process. He cannot avoid earlier research work. He must start with earlier work. He should note down all such research work, published in books, journals or unpublished thesis. He will get guidelines for his research from taking a review of literature. He should collect information in respect of earlier research work. He should enlist them in the manner given below:

1. Author/Researcher
2. Title of research /Name of book

3. Publisher
4. Year of publication
5. Objectives of his study
6. Conclusion/Suggestions

Then he can compare this information with his study to show separate identity of his study. He must be honest to point out similarities and differences of his study from earlier research work.

(4) Methodology

It is related to collection of data. There are two sources for collecting data; primary and secondary. Primary data is original and collected in field work, either through questionnaire/interviews. The secondary data relied on library work. Such primary data are collected by sampling method. The procedure for selecting the sample must be mentioned. The methodology must give various aspects of the problem that are studied for valid generalization about the phenomena. The scales of measurement must be explained along with different concepts used in the study.

While conducting a research based on field work, the procedural things like definition of universe, preparation of source list must be given. We use case study method, historical research, etc. He must make it clear as to which method is used in his research work. When questionnaire is prepared, a copy of it must be given in appendix.

(5) Interpretation of Data

Mainly the data collected from primary source need to be interpreted in a systematic manner. The tabulation must be completed to draw conclusions. All the questions are not useful for report writing. One has to select them or club them according to hypothesis or objectives of study.

(6) Conclusions/Suggestions

Data analysis forms the crux of the problem. The information collected in field work is useful to draw conclusions of study. In relation with the objectives of study the analysis of data may lead the researcher to pin point his suggestions. This is the most important part of study. The conclusions must be based on logical and statistical reasoning. The report should contain not only the generalization of inference but also the basis on which the inferences are drawn. All sorts of proofs, numerical and logical, must be given in support of any theory that has been advanced. He should point out the limitations of his study.

(7) Bibliography

The list of references must be arranged in alphabetical order and be presented in appendix. The books should be given in first section and articles are in second section and research projects in the third. The pattern of bibliography is considered convenient and satisfactory from the point of view of reader.

(8) Appendices

The general information in tabular form which is not directly used in the analysis of data but which is useful to understand the background of study can be given in appendix.

9.4 LAYOUT OF THE RESEARCH REPORT

There is scientific method for the layout of the research report. The layout of the report means as to what the research report should contain. The contents of the research report are noted below:

1. Preliminary Page
2. Main Text
3. End Matter

(1) Preliminary Pages

These must be title of the research topic and data. There must be preface or foreword to the research work. It should be followed by table of contents. The list of tables, maps should be given.

(2) Main Text

It provides the complete outline of research report along with all details. The title page is reported in the main text. Details of text are given continuously as divided in different chapters.

(a) Introduction
(b) Statement of the problem
(c) The analysis of data
(d) The implications drawn from the results
(e) The summary

(a) Introduction

Its purpose is to introduce the research topic to readers. It must cover statement of the problem, hypotheses, objectives of study, review of literature, and the methodology to cover primary and secondary data, limitations of study and chapter scheme. Some may give in brief in the first chapter the introduction of the research project highlighting the importance of study. This is followed by research methodology in separate chapter.

The methodology should point out the method of study, the research design and method of data collection.

(b) Statement of the Problem

This is crux of his research. It highlights main theme of his study. It must be in non technical language. It should be in simple manner so ordinary reader may follow it. The social research must be made available to common man. The research in agricultural problems must be easy for farmers to read it.

(c) Analysis of Data

Data so collected should be presented in systematic manner and with its help, conclusions can be drawn. This helps to test the hypothesis. Data analysis must be made to confirm the objectives of the study.

(d) Implications of Data

The results based on the analysis of data must be valid. This is the main body of research. It contains statistical summaries and analysis of data. There should be logical sequence in the analysis of data. The primary data may lead to establish the results. He must have separate chapter on conclusions and recommendations. The conclusions must be based on data analysis. The conclusions must be such which may lead to generalization and its applicability in similar circumstances. The conditions of research work limiting its scope for generalization must be made clear by the researcher.

(e) Summary

This is conclusive part of study. It makes the reader to understand by reading summary the knowledge of the research work. This is also a synopsis of study.

(3) End Matter

It covers relevant appendices covering general information, the concepts and bibliography. The index may also be added to the report.

9.5 PRINCIPLES OF REPORT WRITING

Following are some important principles for writing a good research report:

- **Make small sentences:** Reading begins to get strenuous when sentences used in the research report average more than 25 words.
- **Vary sentence length:** In using short sentences do not let the work become choppy. Sentences of considerable length are all right provided, they are balanced with enough short sentences.
- **Use simple words:** The researcher is advised to use simple words in his research report.
- **Use familiar words:** It is better to use familiar words in a research report.
- **Avoid unnecessary words:** The use of unnecessary words tire a reader and fog up the writing.
- **Write to express not to impress:** The best way to impress the reader of report is to express what you have to say clearly and directly.
- **Write as you talk:** The researcher should make his report writing as though it is his speech.
- **Keep as many active verbs as possible:** Use of active verbs puts life into report writing.
- **Tie in with reader's experience:** Always write research reports with a particular reader in mind. Relate what you have to tell him about your research report. This is the way to have the reader understand your report.
- **Make the report short and sweet:** A short report makes reading interesting and sweet. Short report should not mean short-cut report.

9.6 STEPS IN REPORT WRITING

It is the critical stage and hence, it requires patience. These is no mechanical formulate to present a report, though there are certain steps to be followed while writing a research report. The usual steps in report writing can be indicated in the following manner:

- Logical analysis of subject matter.
- Preparation of final outline.
- Preparation of rough draft.
- Rewriting and polishing.
- Preparation of final bibliography.
- Writing the final draft.

It is pertinent to follow these steps and hence, it is essential to understand these steps thoroughly.

(a) Logical Analysis of Subject Matter

When a researcher thinks of doing a research, he must select subject and topic of his research work. The subject must be of his own interest and there must be scope for further research. Such can be selected and developed logically or chronologically. He must find out mental connections and associations by way of analysis to finalize his subject. Logical treatment often consists in developing from the simple possible to the most complex strictures. He can use the deductive method or inductive method in his research work. Secondly, the alternative in selecting research subject is to use chronological method. In this method, he should concentrate on the connection or sequence in time or occurrence. The directions for doing or making something usually follow the chronological method. In this method, he should concentrate on the connection or sequence in time or occurrence. The directions for doing or making something usually follow the chronological order.

(b) Preparation of Final Outline

Outlines are the framework upon which long written works are constructed. It is an aid to logical organization of the material and remainder of the points to be stressed in the report. He should rely on review of literature. The earlier research works can provide basic information as well as thinking to the researcher to pursue his subject.

(c) Preparation of Rough Draft

The purpose of the report is to convey to the interested persons the whole result of the study in sufficient detail and so arranged as to enable each reader to comprehend the data and so determine for himself the validity of conclusions. Taking into account this purpose of research, the research report writing has its own significance. The researcher has already collected primary data and secondary data. He has also set his objectives of the study. Taking into account the objectives of his study, he should make an attempt to prepare a draft report on the basis of analysis of the data. He should prepare a procedure to be followed in report writing. He must mention the limitations of his study. He may analyze data systematically with the help of statistical methods to arrive at the conclusions. The research is fact finding study which may lead the researcher to point out suggestions or recommendations.

(d) Rewriting and Polishing the Rough Draft

Research is a continuous process. Research is not essay writing. Researcher must consider the data, write down his findings, reconsider them, and rewrite. This careful revision makes the difference between a mediocre and a good license of writing. The researcher must concentrate on weakness in the logical development or presentation. He should check the consistency in his presentation. He must be

aware that his report writing must be of definite pattern. He must also take atmost care of the language of writing a report.

(e) Bibliography

This helps the researcher to collect secondary source of the data. This is also useful to review the earlier research work. He should prepare the bibliography from the beginning of his research work. While selecting a topic or subject of research, he must refer books, journals, research projects and list the important documents in systematic manner. The bibliography must be in proper form. The researcher must have separate cards, indicating following details, readily available with him, so that he can make a note of it while he refers to a book/journal/research report.

The bibliography must be included in the appendix of his research report. It must be exhaustive to cover all types of works the researcher has used. It must be arranged alphabetically. He can divide it in different sections, such as books in first section, journals in second, research reports in third, etc. Generally the prescribed form for preparation of bibliography is as given below:

The book must be noted in following manner:

(1) Name of Author (Surname first).
(2) Title of book.
(3) Publisher's name, place and data of publication.
(4) Number of volumes.

The article can be mentioned in following manner:

(1) Name of author (surname first)
(2) Title of article (in quotation mark)
(3) Name of periodical (underline it)
(4) The volume or volume and number
(5) Data of issue
(6) The pagination

(f) Final Report

The final report must be written in a concise and objective style and in simple language. The researcher should avoid expressions in his report, such as "it seems", "there may be" and like ones. He should avoid abstract terminology and technical jargon. He may refer to usual and common experiences to illustrate his point. The report writing is an art. No two researchers may have common style of report writing. But it must be interesting for a common man to add to his knowledge.

However report on scientific subject may have most technical presentation. The scientists may be familiar with technical concepts and they may find it valuable if such report is mostly technical in form.

9.7 STEPS INVOLVED IN DRAFTING A RESEARCH REPORT

A research report must be well drafted so that it is seriously taken by others and all that have to be said are well said. Reader orientation, purpose orientation, time orientation, technology orientation, etc., are all needed. The steps in writing a research report are presented below.

1. Organization of Thought

Organization of thought as to how the report be presented is the most fundamental starting point in the journey of preparation a research report. Ideas come before the mind eye. These are thought over again and a frame of presentation is planned. This plan does not in itself constitute style, but it is the foundation.

2. Acquaintance with the Research

Full acquaintance with research is needed. This is facilitated by notes. It is desirable to make notes on separate cards or slips called form-facet. Accuracy and adequacy are required – The second facet of mastery over notes consists in the investigator's complete control over the data, called study facet, i.e., understanding each fact by itself and in terms of others and of the researcher's own thoughts; the notes to be compared; criticized and revalued in order to enable the investigator to direct and organize the data in his own way and perhaps differently from what others have done.

3. First Draft

The first draft concentrates on substance, i.e., fullness of facts. All the facts of value are to be brought together. In addition to fullness, accuracy of the facts incorporated into the text becomes necessary. Another requirement is that there should be balance, proportion and development in facts. Importance is to be given to the comprehensiveness of the report but not to the language and form. For writing the first draft the researcher should have control over his notes and should think continuously over the problem. In a way, the first draft is the most important of the different stages in reporting. This report may have to be re-written a number of times and still it continues to remain only a working draft. There are three purposes in writing the first draft, *viz.*, to weave the material together for making clear connections, to assure the investigator himself of a satisfactory organizations and fullness of the facts, and to avoid blank paper fright that may be present in every young researcher.

4. Second Draft

After a lapse of sometime from the completion of first draft, the revision is made for writing the second draft. While drafting the second one, the researcher should concentrate largely on form and language. The researcher should give the first draft, at this stage, a shape so that it can be readable, clear and lucid. Considerable trimming or editing will have to be done to make the writing precise, concise and brief. Finally, at the second draft stage, critical evaluation will have to be made of all that has been written-facts, findings, conclusions and recommendations. To make the report readable and effective, the language plays major part.

5. Third Draft

The final stage in drafting is the preparation of final report. It concentrates mainly on the finish and final touches, i.e., on documentation and polish to make the report weighty, authoritative, convincing and attractive. Documentation indicates the references to the sources, other previous and current work and view, additional data and discussion and suggested further reading on the specific problem as handled by the researcher. In other words, it indicates the thoroughness of the investigation and on the other a guide to further work. A good research paper depends not only upon the amount of reading or notes taken or upon the form of presentation but also upon the accurate and thorough recording of the investigation.

9.8 DOCUMENTATION

Documentation is the process of collecting and extracting the documents which are relevant to research.

Documents may be classified into:

- Personal documents;
- Company documents;
- Consultants' report and published materials; and
- Public documents.

Personal documents are those that are written by or on behalf of individuals. They may include autobiographies, biographies, diaries, memoirs, letters, observations and inscriptions, which are primarily written for the use and satisfaction of individuals and which can be utilized for research purposes. Personal documents play a very vital role in research, when certain information is not available from any other sources; however, such documents are subject to the difficulties of availability, reliability and validity of inferences.

Company documents are the most essential types of documents is management research, annual reports, statements of income and expenditure and balance sheets, files and records, policy statements, resolutions, minutes of board of directors, general bodies and executive conferences, performance records and evaluation files, specific forecasting and evaluation reports, directors' reports, etc. Many of these documents are reliable ones, though they are subject to the problems of availability and adequacy. Such documents can be published or unpublished.

Consultant's published materials consist of report of professional consultants, records of commodity boards, chambers of commerce, FICCI, manufacturers' associations and industry associations.

Public documents are documents, both published and unpublished, of government organizations and documents of public interest. They include government records and files, draft outlines of five-year plans, consultative committee reports, finance commission reports, special enquiry commission reports, Company Law Board reports, MRTP Commission reports, reports and files of the Registrar of Companies, the Registrar of Firms, the Ministry of Commerce and Industry, etc., report of population census, National Sample Survey and such other government research institutions. Such documents are valuable if they are reliable and suitable for a particular study if they can be obtained. Documentation is one of the most important needs of any management researcher at the primary sate of his research.

9.9 FOOTNOTES

Footnotes are meant to give complete bibliographical references and to provide the reader with information to enable him consult sources independently. These may be placed at the foot of a page or at the end of every chapter. When footnotes are given at the foot of the page, they are to be separated from the text by a fifteen space solid line drawn from the left margin and one double space below the last line of the text. If given at the end of a chapter, a centered heading 'FOOTNOTES' is necessary. Reference to footnotes is made by the use of superscripts i.e., numerals rose by one-half space. Footnotes should be numbered consecutively through a chapter, right after the statement. However, footnotes should not be resorted to common place statements or ordinary facts. They must be restricted to direct quotations, original ideas, statements, definitions, illustrations and diagrams. If a quotation found in one source that is obtained from another source, is used, both the sources should be listed in

the footnote. Footnotes should be given in single line spacing with the first line only indented. A double space should separate successive footnotes.

In this context, certain conventions in footnoting should be kept in mind. While writing the author's name, the first name or initials should be written first. Titles of completed works, such as books and reports should be underlined. Titles of articles should be enclosed by double quotation marks. The publisher's name, year or date of publication and the exact page(s) of the source of reference should also be given in this order.

9.10 BIBLIOGRAPHY

The 'works cited' form of bibliography is preferable over the 'sources consulted'. Every book, thesis, article, documents which has been cited should be included in the list of 'works cited'. The bibliography should follow a logical arrangement in alphabetical order. In report of current practice is to have one comprehensive listing - not to divide into books, journals, newspapers, official papers, documents and manuscripts.

The author(s) name, the title of the work, date of publication, name of the publisher and the place of publication be mentioned.

For articles, the volume number and inclusive pages be also given, the author's initials or surname should follow the name.

When there are three or more authors of a particular work, the co-authors may be referred alphabetically. If there be more than one work by the same author, the author's name should be listed only once; subsequently a line will substitute his name.

This bibliographical listing should not be numbered. It should be given only at the end of the thesis.

9.11 QUESTIONS

1. Explain what is a research report? And types of Research report?
2. Explain the contents of research report
3. Explain the principles in report writing
4. Explain the steps involved in research report writing

● ● ●

Chapter
10

Quantitative Tools

Structure:

10.1 MEASURES OF CENTRAL TENDENCY

What are the measures of Central Tendency?

Measures of Central Tendency are nothing but **statistical averages**. It tell us the value about which items have a tendency to cluster. It is representative of the mass of data. It is useful in comparing different distributions. The average have a general tendency to lie at center and hence they, are termed as **'measures of central tendency'**. The requisites of measures of central tendency include it's simplicity in definition, easiness in computation, capability of further algebraic treatments, sampling stability and non-influence by extreme observations.

Types: The measures of central tendencies or averages can be classified as

(i) Algebraic averages and

(ii) Positional averages.

Algebraic averages require algebraic formula to compute. While Positional averages can be located from graphs. Algebraic averages cannot be obtained from graph.

Amongst Mean, Mode, Median, The **Mean** falls in **Algebraic** average category. While **Median** and **Mode** are **Positional** averages.

1. Arithmetic mean

Mean is the simplest measures of central tendency and is widely used. It is used in summarizing the essential features of a series and enables data to be compared. It is easy to define and simple to understand. It is based on all the observations and hence, treated as a good representative of the distribution. It has a sampling stability and also capability of further algebraic treatment. Its only limitation is it **cannot** be obtained for '**open end**' class interval distribution. Also it is duly affected by extreme observations. Sometimes it gives absurd results. It may be the value which is not part of the distribution. Specially in Economics and Social Studies where direct quantitative measurements are possible, Mean is the better average than others. Simple Mean indicates Arithmetic Mean.

Defination: Let X be the variable taking values x_1, x_2,…..x_n, the Arithmetic mean of X is given by $\frac{\Sigma X}{n}$

For frequency distribution, ungrouped as well as grouped the arithmetic mean is calculated by $\frac{\Sigma fX}{\Sigma f}$. For Grouped frequency distribution, x denote class marks, i.e., mid values of the class intervals.

2. Geometric mean

Geometric Mean is defined as n^{th} root of the product of the values of n times.

Its application is in determination of average per cent of change. Whenever ratios, percentages are to be averaged, Geometric Mean can be used. Generally in construction of Index numbers, Geometric Mean is often used.

Occasionally a frequency distribution is encounted that is skewed to right, but if logarithmsX values are used with the class intervals of logs constant, the curve becomes symmetrical. In such situation the Geometric Mean may be appropriate.

Geometric Mean can also be used in averaging the rate of change.

3. Harmonic mean

Harmonic Mean is defined as the reciprocal of the average values of items of a series. It has limited applications particularly where time and rate are involved. It is used in case like time and motion study where time is variable and distance constant.

4. Median

Median is the value of the middle item of the series where the series is arranged in ascending or descending order. Median is used only in the context of qualitative phenomenon for example in estimating intelligence. Median is not useful where items need to be assigned relative importance and weights. It is not frequently used in sampling statistics.

There are two specific situations where the **median serves as a valuable alternative to the mean.** These occur when

(1) There are a few extreme scores in the distribution and

(2) Some scores have undetermined values.

In psychology, this often occurs in learning experiments where you are measuring the number of errors or amount of time required for an individual to solve a particular problem. Generally in Open-end class interval frequency distribution mean is unable to compute and hence in such cases median will be preferred.

5. Mode

Mode is the value which occurs most frequently in the distribution. It is easy to compute and it can be used with any scale of measurement. The fact that mode can be used in any scale of measurement made, i.e., flexible when scores are measured in a nominal scale it is impossible to calculate either mean or median so mode is used to describe central tendency. Mode describes the typical or most representative academic major for the sample. Because the mode identifies the most typical value/case, it often produces a more sensible measure of central tendency.

Thus comparing mean, median and mode it is noted that the mean is the commonly used average, taking into consideration all the observations. It can be good representative of the distribution. The goal of central tendency is to find a single value that best represents the distribution. Besides being a good representative the mean has added average of being a good measure for purpose of inferential statistics. Specifically, whenever you take a sample from a population the sample mean will give a good indication of the value of the population mean. Also mean satisfies majority requisites of an ideal average so mean is the superior amongst all. But there are certain situations where mean cannot be computed then median or mode can be used.

10.2 MEASURES OF DISPERSION

What is the meaning and need of Dispersion?

An average can represent a distribution only as a best single representative. There are some situations where averages fail to compare the distributions. Consider the following case.

Four candidates Sanchit, Saurabh, Shivani, Sayali, score marks in three tests as follows.

Test	Sanchit	Saurabh	Shivani	Sayali
I	80	95	80	98
II	75	80	80	92
III	85	65	80	50

On the basis of average if candidates are compared, the conclusion is that all four are equal or same as far as scores are concerned, since the average score for everyone is 80. If studied minutely, we see that Shivani is most consistent. Sanchit comes next to Saurabh and then Sayali. So here there is need to study scatter of the values from average and it is defined as dispersion. Thus Dispersion means scatter or spread of individual values from its average in the distribution. The measures which are used to measure dispersion are known as measures of dispersion.

Requisites of Good Measures of Dispersion

(1) It should be easy to understand and calculate.

(2) It should be based on all the observations.

(3) It should not be affected much by sampling fluctuations.

(4) It should not be affected much by extreme observations.

(5) It should be capable for further algebraic/mathematical treatment.

Measures of dispersion can be either Absolute or Relative.

Absolute measures are with respect to given distribution and hence are expressed in corresponding units of measurements. While Relative measures are free from any measurement units. They are pure numbers.

To compare different distributions which differ in units of measurement always relative measures are used. The relative measures generally are referred by ‘coefficient’

1. Range

Range is the difference between two extreme observations. It is the simplest measure of dispersion.

Thus Range = H-S

Where H: Highest- Value in the data

S: Smallest- Value in the data

The utility of range is that it gives an idea of variability very quickly. But it affected very greatly by sampling fluctuations. Range is mostly used as a rough measure of variability.

2. Quartile deviation

$$\text{Quartile Deviation or Semi Inter Quartile Range} = \frac{Q_3 - Q_1}{2}$$

Where Q_1 and Q_3 are first and third quartiles.

Quartiles are the partition values which divide the distribution into Four equal parts, when data is arranged in order. There are3 Quartiles in all denoted by Q_1, Q_2 and Q_3.

Q_1 is known as lower quartile and divide the distribution such that 25% observations have value less than Q_1 and 75% observations have value above Q_1. Q_2 is known as second quartile and divide the distribution such that 50% observations have value less than Q_2 and 50% observations have value above Q_2. In other words Q_2 is nothing but Median of the distribution.

Q_3 is known as upper quartile and divide the distribution such that 75% observations have value less than Q_3 and 25% observations have value above Q_3.

Q_1 and Q_3 provide limits for central 50% observations.

The computations of Quartiles is similar to that of Median with minor changes.

The other partition values are Deciles (D_1, D_2,..... D_9) and

Percentiles (P_1, P_2,……..P_{99}). The interpretations and computational procedures for Deciles and Percentiles are similar to that of Median. Deciles divide the distribution into 10 equal parts, while Percentiles divide the distribution into 100 equal parts. Fifth DecileD_5, Fiftieth Percentile P_{50} are median of the distribution. Tenth Percentile is D_1. 25th Percentile is Q_1. and so on.

Partition values can be used to determine limits for desired percentage of central observations. For example Q_1, and Q_3 provide limits for central 25% observations, P_{20}, or D_2 and P_{80} or D_8 provide limits for central 60% observations.

3. Coefficient of quartile deviation

$$\text{Coeff. of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

If open-end class intervals are given the suitable measure is Quartile Deviation.

4. Mean deviation

Mean deviation about an average 'A' is defined as arithmetic mean of absolute deviations from an average 'A'. A can be mean/median/mode. Accordingly there are three mean deviations. Mean Deviation about median is most commonly used measure.

5. Coefficient of mean deviation

Coefficient of mean deviation is the relative measure and is defined as the ratio of mean deviation to an average 'A'.

$$\text{Thus Coefficient of mean deviation} = \frac{\text{Mean Deviation about 'A'}}{A}$$

6. Standard deviation

As mean is in Central Tendency we have standard deviation or variance in dispersion.

The standard deviation concept is very important in further analysis. It is defined as root mean squared deviation.

$$\sigma = \sqrt{\frac{1}{N}\Sigma\left(X = \overline{X}\right)^2} \qquad \text{OR} \qquad + \sqrt{\frac{\Sigma f\left(\mathrm{X} - \overline{\mathrm{X}}\right)^2}{\Sigma f}}$$

For Raw data and Frequency distribution respectively.

Variance $= \sigma^2$

7. Coefficient of Variation

Coefficient of Variation is very important measure used to compare different distributions on the basis of consistency homogeneity, variability, uniformity.

$$\text{Coeff. of Variation} = \frac{\sigma}{\overline{\mathrm{X}}} \times 100$$

Less CV indicates more consistency, more uniformity, more stability while More CV indicates more variability, more heterogeneity.

Following is the list of Absolute and Relative measures of dispersion with their formulas.

Absolute Measures of Dispersion	**Relative Measures of Dispersion**
1. Range = H-S Where H: Higest- Value in the data S: Smallest- Value in the data	$\text{Coeff. of Range} = \frac{\mathrm{H} - \mathrm{S}}{\mathrm{H} + \mathrm{S}}$
2. Quartile deviation or Semi Inter Quartile Range $= \frac{\mathrm{Q}_3 - \mathrm{Q}_1}{2}$	$\text{Coeff. of Q.D} = \frac{\mathrm{Q}_3 - \mathrm{Q}_1}{\mathrm{Q}_3 - \mathrm{Q}_1}$
3. Mean Deviation about an average	Coefficient of Mean Deviation
4. Standard Deviation σ OR Variance σ^2	Coefficient of Variation (CV) $CV = \frac{\sigma}{\overline{\mathrm{X}}} \times 100$

8. Lorenz curve

Definition

In 1905 Max Lorenz, an American economist developed a graphical representation of 'wealth distribution'. On the graph, a straight diagonal line represents perfect equality of wealth distribution; the Lorenz curve lies beneath it, showing the reality of wealth distribution. The difference between the straight line and the curved line is the amount of inequality of wealth distribution, a figure described by the Gini coefficient.

The Lorenz curve is a measure of the distribution of wealth (or income or other factors) in a society. The x value of the curve corresponds to a percentile of the population ordered according to the characteristic in question. The y value of the curve represents that portion of the total value of the

characteristic in question held by people no wealthier than the x-valued percentile of the population. Thus, the value (0.7, 0.3) means that the bottom 70% of the population owns 30% of the total wealth in society. The Gini coefficient is the (shaded) area between the Lorenz curve that would exist in a perfectly egalitarian society (the dashed line) and the Lorenz curve that does exist, divided by the area under the Lorenz curve that would exist in a perfectly egalitarian society. The coefficients are thus normalized to run from zero in a perfectly egalitarian society, to one in a society in which the wealthiest person held all the wealth.

Creating a Lorenz Curve in Excel

1. **Enter original data for case study**

Region	% Population	% Income
A	15	25
B	35	35
C	30	15
D	15	5
E	5	20

2. **Add a column for Income/Population**

Region	% Population	% Income	Income/population
A	15	25	1.666666667
B	35	35	1
C	30	15	0.5
D	15	5	0.333333333
E	5	20	4

3. **Sort by column % Income / % Population**

Region	%Population	% Income	Income/population
D	15	5	0.333333333
C	30	15	0.5
B	35	35	1
A	15	25	1.666666667
E	5	20	4

4. **Add columns for Cumulative % Population and Cumulative % Income**

Region	% Population	% Income	Income/Population	C % Population	C % Income
D	15	5	0.333333333	15	5
C	30	15	0.5	45	20
B	35	35	1	80	55
A	15	25	1.666666667	95	80
E	5	20	4	100	100

5. **Add a row for zeroes**

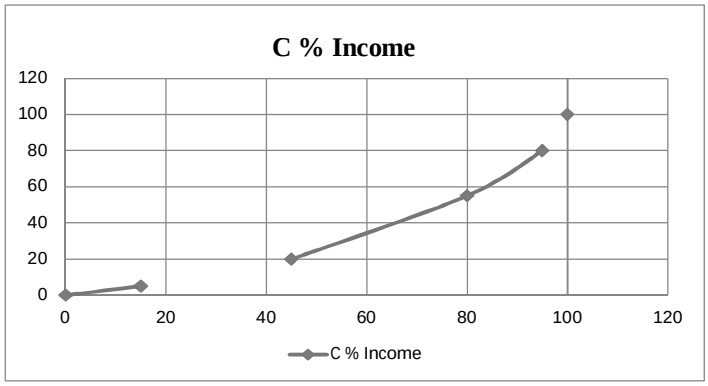
Region	% Population	% Income	Income/Population	C % Population	C % Income
				0	0
D	15	5	0.333333333	15	5
C	30	15	0.5	45	20
B	35	35	1	80	55
A	15	25	1.666666667	95	80
E	5	20	4	100	100

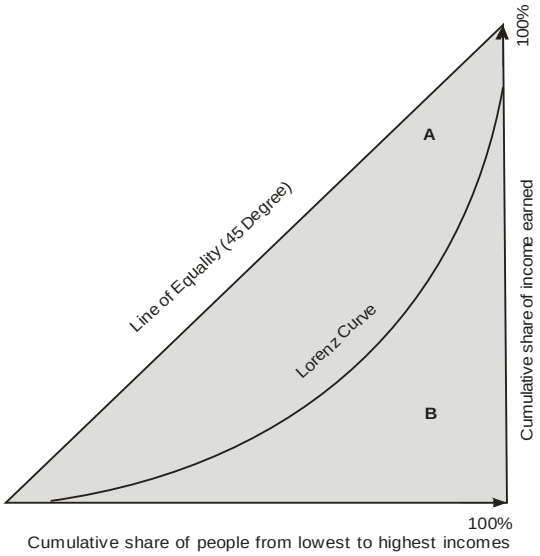
6. **Using the cumulative % columns, insert a scatter plot with data points connected by smoothed lines.**
7. **Format the axes to end at 100. (Right-click on x-axis and select Format Axis. Do the same for the y-axis.)**
8. **Label the axes and add a chart title. (Right-click on the chart and select Chart Options>Titles.)**
9. **Label the legend line. (Right-click on the chart and select Source Data. Enter a name such as "Case Study".)**
10. **Add the equality line to the chart.**
    a. Enter new data into the spreadsheet. You can add as many points in the line asyou want as long as you include at least (0, 0) and (100,100).

X - Values	Y - Values
0	0
100	100

    b. Then right-click on the chart and select Source Data. Press the "Add" button. For the name, type "Equality". For the x values, select the x-value cells. For the y values, select the y-value cells.

The rough sketch will be as follows





10.3 CORRELATION ANALYSIS

Correlation Coefficient and Coefficient of Determination

The measures we studied in the above sections are related to univariate study only.

In practice, we come across a large number of problems involving the use of two or more than two variables. The variables are said to be correlated if change in one variable causes change in others either in same direction or in opposite direction. The degree of relationship between the variables under consideration is measured through the correlation analysis. The measure of correlation is called the correlation coefficient.

Correlation analysis attempts to determine the "degree of relationship between the variables"

Thus the correlation is a statistical device which helps us in analyzing the covariation of two or more variables.

The problem of analyzing the relation between different series should be broken into three steps:

(1) Determining whether a relation exists and if it does measuring it.

(2) Testing whether it is significant.

(3) Establishing the cause and effect relation if any.

Correlation is described or classified in different ways. Some of the important ways are as follows.

(i) Positive or Negative

(ii) Simple or Multiple/Partial

(iii) Linear or Non-linear.

Methods of studying Correlation:

(a) Scatter Diagram Method.

(b) Karl Pearson's Coefficient of Correlation. (r)

(c) Concurrent Deviation Method.

(d) Spearman's Rank Correlation Coefficient. (R)

Karl Pearson's Correlation Coefficient is also known as Product Moment Correlation Coefficient. It is based on the following assumptions:

- The variables are related by linear relationship.
- The two variables under study are affected by a large number of independent causes so as to form a normal distribution.
- There is cause and effect relationship between the forces affecting the distribution of the items in the two series.

Amongst the mathematical methods used for measuring the degree of relationship, Karl Pearson's method is most popular. The correlation coefficient summarizes in one figure not only the degree of correlation but also the direction positive or negative.

Interpretation of Coefficient of Correlation

Coefficient of Correlation lies between -1 and +1 (both inclusive)

When r = -1 it implies perfect negative linear relationship between the variables.

When r = +1 it implies perfect positive linear relationship between the variables.

When r = 0 it implies no relationship between the variables, i.e., variables are uncorrelated.

The value of r closer to 0 indicates weak relationship or weak association between the variables. While r closer to -1 or +1 indicates strong association amongst the variables.

The full interpretation of r depends upon circumstances one of which is the size of the sample.

The probable error of the coefficient of correlation helps to determine the reliability of the value coefficient.

The probable error $P.E._r = 0.6745 (1- r^2) / \sqrt{N}$

Where N is number of pairs of observations and r is coefficient of correlation.

- If the value of r is less than probable error there is no evidence of correlation, i.e., the value of r is not at all significant.
- If the value of r is more than six times the probable error, the coefficient of correlation is practically certain, i.e., the value of r is significant.
- The (r – P.E., r + P.E.) provide the limits for population correlation coefficient expected to lie.

The measure of probable error can be properly used only when the following conditions exists.

(1) The data must approximate a normal frequency curve.

(2) The statistical measure for which the P.E. is computed must have been calculated from a sample.

(3) The sample must have been selected in an unbiased manner and the individual items must be independent.

COEFFICIENT OF DETERMINATION AND NON-DETERMINATION

If r is coefficient of correlation then r^2 is coefficient of determination. It is used to measure explained variation. It is defined as the ratio of the explained variance to the total variance.

Coefficient of determination = (Explained variation) / Total Variance.

Coefficient of non – determination = 1 - coefficient of determination.

$= 1 – r^2$

Coefficient of non-determination is denoted by K^2. And square root of it, i.e., K is known as coefficient of alienation.

According to Tuttle **"the coefficient of correlation has been grossly overrated and is used entirely too much. Its square, the coefficient of determination is much more useful to measure the linear covariation of two variables. The reader should develop the habit of squaring every correlation coefficient he finds cited or stated before coming to any conclusion about the extent of the linear relationship between the two correlated variables."**

Consider the following table of r and r^2

R	.90	.80	.70	**.60**	.50	.40	**.30**	.20	.10
R^2	.81	.64	.49	**.36**	.25	.16	**.09**	.04	.01

Following things are to be noted.

(1) As r decreases from 1 to minimum., r^2 decreases more rapidly.

(2) The value of r = 0.707 implies r^2 = 0.499849, i.e., half variance in Y is due to X

(3) If r = 0.6, 36% of total variation is explained, while if r = 0.3 only 9% of total variation is explained.

(4) Value of r^2 is always positive. It cannot tell whether the relationship between the two variables is positive or negative. For that purpose r is to be computed.

10.4 REGRESSION ANALYSIS

Regression Analysis

Regression technique is a tool to isolate the casual relationship between the variables. Several regression models are available to test and establish a statistically satisfactory fit between the dependent variable and a specific range of independent variables. Forecasts are made by substituting in values of the independent variables in the equation and then computing the dependent variables. These methods are useful for long-term forecasting are relatively more sophisticated and expensive to use.

Regression analysis is one of the scientific techniques used for predictions. According to M.M. Blair, "Regression Analysis is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data.

Regression analysis is confined to the study of only two variables at a time is termed as simple regression. The regression analysis for studying more than two variables at a time is known as multiple regression.

In this analysis there are two types of variables as Dependent variables and Independent variables.

The variables whose value is influenced or is to be predicted is called dependent variable.

The variable which influences the values or is used for prediction is called independent variable.

In regression analysis, independent variable is also known as regressor or explanatory while the dependent variable is known as regressed or explained variable.

The estimation are done with the help of equations known as regression equations. The regression equations gives in accordance with the Principle of Least Squares which consists in minimizing the sum of the squares of the deviations between and the given observed values of the variables and their corresponding estimated values given by the line of best fit.

10.5 QUESTIONS

1. Explain the various measures of central tendency.
2. Explain the measures of dispersion.
3. Explain regression analysis is brief.
4. Explain in detail correlation analysis?

Chapter
11

Tools for Testing of Hypothesis

Structure:

11.1 INTRODUCTION

Hypothesis is one of the important aspects in any research study. Hypothesis is any statement or assertion made about population parameters. The purpose of hypothesis testing is to be determine the accuracy of hypothesis due to the fact that data is collected through sampling method and not complete enumeration method.

The accuracy of hypothesis is evaluated by determining the statistical likelihood that the data reveal true difference and not random sampling error.

There are two approaches to hypothesis testing as

(i) Classical or Sampling Theory approach and

(ii) Bayesian approach.

Classical or Sampling theory approach is most widely used in research applications. This approach represents an objective view of probability in which the decision making rests totally on an analysis of available sampling data. A hypothesis is established, it is rejected or accepted based on the sample data collected.

Bayesian statistics are an extension of the classical approach. Here also sampling data is used for decision making, but here research goes beyond it to consider all other available information. The additional information consists of subjective probability estimates stated in terms of degree of belief. These subjective estimates are based on general experience than on specific data collected. They are expressed as a prior distribution that can be revised after the sample information is gathered. The revised estimate known as posterior distribution information and so on various decision rules are established, cost and other estimates can be introduced, and these element are used to judge decision alternative hypothesis testing procedure.

In classical tests of significance two kinds of hypothesis are used – The Null hypothesis and an Alternative hypothesis.

Null Hypothesis: It is a statement that no difference exists between the parameter and the statistic. The **"No Difference"** type hypothesis is termed as Null hypothesis and denoted by H_0.

Alternative Hypothesis: It is the logical opposite of the Null hypothesis. It is denoted by H_1 or H_A. The alternative hypothesis may take several forms depending on the objective of researcher. It

may be of the "nSot equal to" or "greater than" or "less than" type. And these types will be used be decide whether the underlying test is two tailed or one tailed.

If H_1 or H_A is "Not equal to" types ($\neq$). The underlying test is **two tailed** or **two sided** or **non directional test**.

Otherwise the test is **One tailed** or **One sided** or **directional.**

H_0 and H_1/H_A are complementary to each other. If H_0 is rejected means H_1 is accepted and *vice versa.*

Based on sample results H_0 may be accepted or rejected. And H_0 may be True or False in legal or true sense. Thus, it will arise following four situations.

H_0 / Decision	True	False
Accept H_0	Correct Decision	**Type II Error**
Reject H_0	**Type I Error**	Correct Decision

The error committed in rejecting true Hypothesis is termed as Type I Error. The probability of community Type I Error is denoted by $\propto$ and known as **level of significance**. The standard values of $\propto$ are 5% and 1%

$\propto$ = P [Type I Error]

= P [Reject H_0/H_0 is True].

In Quality control Type I Error is termed as **Producer's Risk.**

The error committed in accepting false hypothesis is termed as Type II Error. Probability of community Type II Error is denoted by ß.

ß = P [Type II Error]

= P [Accept H_0/H_0 is false]

|- ß is known as **power of test.**

Some Basic Concepts/Definitions Related to Estimation

→ Sampling theory is the study of relationships between a population and samples drawn from the population.

Sampling theory helps us to determine whether the differences between two samples are actually due to chance variation or whether they are really significant.

(1) Parameter: It is a statistical measure based on all the units of a **population**. For example, population mean, population standard deviation, proportion of defectives in population, etc.

(2) Statistic: It is also a statistical measure that based is on all units selected in

Sample: For example, sample mean, sample standard deviation, etc.

Consider the case of selecting 100 houses from the city of Mumbai to study effect of Internet on children. Let us assume that Mumbai city consists of 50,000 houses having Internet connection. Here 50,000 is the population size and 100 is sample size selected from these 50,000.

Now any statistical measures say average age of user, standard deviation or variance of age of user, if these are obtained/calculated from all 50,000 users, it will be '**Parameter**' and if calculated from selected 100 houses then it will be **sample statistic** or simply statistic.

Since the units selected in two or more samples drawn from a population are not the same, the value of statistic varies from sample to sample. But the parameter always remains constant. This variation in the value of statistic is called sampling fluctuation.

A parameter has no sampling fluctuation.

Sampling Distribution of a Statistic

A sampling distribution is a theoretical distribution that express the functional relation between each of the distinct values of the sample statistic and the corresponding probability for all the different possible samples of size n from the same population.

In the other words the frequency distribution or probability distribution of a sample statistic is called sampling distribution of statistic. For such distribution standard deviation, etc., the characteristics mean and standard deviation of the distribution are very important and plays important role in theory of estimation.

The mean of sampling distribution, i.e., an expectation of statistic fit is equal to value of the parameter then it is known as an **unbiasedness** property of the statistic.

The standard deviation of sampling distribution is termed as

Standard Error of an Estimate

It is used as a tool in tests of hypothesis. It gives an idea about the reliability and the precision of a sample. It helps in determining the limits within which the parameters are expected to lie.

It is possible to draw valid conclusions about the population parameter from each sampling distribution.

Type of Estimation: It is difficult to obtain population parameters in many studies, in such cases it is essential to estimate them as accurately as possible.

In estimation there are two types of estimates – Point estimate and Interval estimate.

(A) Point Estimation: A Point estimate is a single value that is used to estimate the unknown parameter, e.g., sample mean is a Point estimate population mean

(B) Interval Estimation: In this type instead of obtaining a single value as an estimate, pair of values are obtained and is used to estimate an interval or range within which parameter lies with certain confidence (probability). Such interval is known as **confidence interval** and the two values are known as **confidence limits**. The probability or confidence generally in terms of percentages are 95%, or 99%. Higher the probability, higher is the confidence. Standard error plays very important role in determining confidence limits and hence confidence interval.

Difference between estimator and estimate

Any sample statistic which is used to estimate a population parameter is called as **estimator.**

An estimate is a specific observed value of statistic. An estimate is formed by taking a sample and computing the value taken by the estimator in that sample.

Properties or Characteristics of Good Estimator

(a) Unbiasedness: Let T denote an estimator and θ denotes parameter. Thus T can be sample mean/proportion/standard deviation, etc., and θ can be population mean/population proportion/population standard deviation, etc. T is said to be an unbiased estimate of θ if E (T) = θ. In other words if on an average T is same as θ, It is said to be unbiased estimate.

(b) Consistency: If V(T) → 0 as n →∝, i.e., as sample size increases, variance approaches to zero which shows spread or dispersion diminishes as sample size becomes large then T is said to be consistent estimator, i.e., as sample size increases the difference between T and θ should be smaller and smaller.

(c) Efficiency: Efficiency is measured by variance. The estimator with smallest variance is an efficient estimator.

(d) Sufficiency: A sufficient statistic is an estimator that utilizes all the information a sample contains about the parameter to be estimated.

Among all the estimators sample mean $\overline{X}$ and sample proportion *P* are sufficient statistics for population mean μ and population proportion P.

Also $\overline{X}$ and P possesses all above four properties.

Method of maximum likelihood provides estimators with the desired properties.

Statistical Testing Procedure

A step by step procedure is as follows:

(1) State the Null Hypothesis.

(2) State the Alternative Hypothesis and decide One Tailed or Two Tailed Tests.

(3) Select the desired level of significance. The most common level is 0.05 and .01 The exact level to choose is largely determined by two much ∝ risk one is willing to accept and the effect that this choice has on ß risk. The larger the ∝, lower is the ß.

(4) Choose the statistical test. To test a hypothesis one must choose an appropriate statistical test. There are various criteria to choose a test. One is the power efficiency, Nature of population, method of sampling, type of measurement scale used and so on.

(5) Obtain the critical test value. (Table Values) for specified level of significance.

(6) Compute the value of Test statistic.

(7) Decision: If calculated Value < Table Value H_0 is accepted at specified level of significance otherwise rejected.

(8) Interpret the results. Draw conclusions.

Test of Significance: Generally there are two classes of significance tests as Parametric and Non-parametric.

Parametric test are more powerful because their data are derived from **interval** and **ratio** measurements.

Non-parametric tests are used to test hypothesis with **nominal** and **ordinal** data.

11.2 PARAMETRIC TESTS

Assumptions for Parametric Tests

(1) The observation must be independent mean, the selection of any one case should not affect the chances for any other case to be included in the sample.
(2) The observations should be drawn from normally distributed populations.
(3) These populations should have equal variance.
(4) The measurement scales should be interval or ratio so that arithmetic operation can be used with them.

The researcher is responsible for reviewing the assumptions pertinent to the chosen test. Performing diagnostic checks on the data allows the researcher to select the most appropriate test.

The **Z-test** or **t-test** is used to determine the statistical significance between a sampling distribution of mean and a parameter.

Z-test is a **large sample** test. Generally, if sample size exceeds 30 it is said to be large sample. Otherwise small sample distribution. This is because of lack of information about the population standard deviation.

When sample size approaches 120, the sample standard deviation becomes a very good estimate of population standard deviation.

Beyond 120, the **t** and **z** distribution are virtually identical.

For the characteristic like average and proportion Z or t distribution based tests are most appropriate tests.

1. Z-Test/ Z-Distribution

(I) One sample Mean test

To test H_0: $\mu = \mu o$

H_1: $\neq \mu o$ or ($\mu < \mu o$ or $\mu > \mu o$)

Here μ denote population mean and μo denote specified value of population mean.

The available data includes $\overline{X}$ (sample mean), σ (population standard deviation) or s (sample standard deviation) and sample size 'n'

The test statistic

$$Z = \left| \frac{\overline{X} - \mu o}{\sigma / \sqrt{n}} \right| \quad \text{or} \quad \left| \frac{\overline{X} - \mu o}{s / \sqrt{n}} \right|$$

The critical values of Z depends upon

(i) level of significance 5% or 1%
(ii) Two tailed or one tailed test (sign in H_1 $\neq$ or $<, >$)

The following table gives critical values

Level of Significance / Type	5%	1%
Two Tailed (H_1 has $\neq$ sign)	1.96	2.58
One Tailed (H_1 has > or < sign)	1.64	2.33

Criteria for Decision

If $|Z| \leq$ critical value

Then H_0 is accepted at specified level of significance.

Otherwise H_0 is rejected.

(II) Two sample mean test

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$ or ($\mu_1 < \mu_2$ or $\mu_1 > \mu_2$)

Available data: For two samples, their sample sizes n_1, $n_2$with mean

$\overline{X}_1$, $\overline{X}_2$ and population standard deviations (may or may not) or sample standard deviations.

The computation of test statistic Z is done as follows:

Case I: If both the samples are drawn from same population with standard deviation σ

$$Z = \left| \frac{\overline{X}_{1-} \overline{X}_2}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \right|$$

Case II: Population standard deviation unknown. Let s_1^2, s_2^2 denotes sample variances.

$$S_1^2 = \frac{\Sigma \left(\overline{X}_1 - \overline{X}_2 \right)^2}{n_1 - 1} \qquad S_2^2 = \frac{\Sigma \left(\overline{X}_2 - \overline{X}_2 \right)^2}{n_2 - 1}$$

Then work out

$$S^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2}$$

$$Z = \left| \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right|$$

Case III: If two samples are drawn from two different populations with standard deviation σ_1 and σ_2

$$Z = \left| \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right|$$

Rest procedure is same as above

(III) One Sample Proportion

H_0: $P = P_0$

H_1: $\neq$ $P \neq P_0$ (or $P < P_0$ or $P > P_0$)

P denotes desired proportion. P_0 is specified value

$$Z = \left| \frac{P - P_0}{\sqrt{\frac{P_0 \, Q_0}{n}}} \right|$$

Where P = sample proportion

$Q_0 = 1 - P_0$

(IV) Two Sample Proportion test

H_0: $P_1 = P_2$

H_1: $P_1 \neq P_0$ (or $P_1 > P_2$ or $P_1 < P_2$)

The test statistic is

(a) If population are heterogeneous

$$Z = \left| \frac{P_1 - P_2}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} \right|$$

Where P_1 P_2 are sample proportions

$q_1 = 1 - P_1$, $q_2 = 1 - P_2$

(b) If proportions are similar with respect to given attribute, the best estimate of population proportion is obtained as

$$P_0 = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$q0 = 1 - P_0$ and

$$Z = \left| \frac{P_1 - P_2}{P_0 q_0 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right|$$

In all above four tests the probability distribution of test statistic is Normal distribution.

So let us discuss about **Normal distribution**.

1. Normal Distribution

This is the most widely used probability distribution. This is applicable for continuous random variable.

A random variable means a real valued function defined over a sample space. For every value of random variable there is associated probability.

If a random variable takes only integer values, it is known as **discrete random variable**. If a random variable assumes any value within range it is known as **continuous random variable**.

From the most widely used probability distributions, Normal distribution is for continuous random variable and Binomial, Poisson distributions are for discrete random variables.

The probability distribution of random variable is either a tabular form or a functional form showing probabilities distributed over various values of random variable such that individual probabilities lies between 0 and 1 and sum or Total probability is 1(unity)

For discrete probability distributions the tabular form or functional form referred as probability mass function (p.m.f.) and for continuous random variables the function is referred as probability density function (p.d.f.)

To write p.m.f. or p.d.f. we require parameter(s) of the distribution.

Parameters specify the distribution completely.

For Normal distribution, there are two parameters mean 'μ' and Standard deviation σ (or variance σ^2)

Let the random variable X is said to follow Normal distribution with parameters $\mu$and σ (or σ^2) Then it's p.d.f. is given by

$$f(\chi) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

$-\infty < \chi < \infty$

$\int_{\alpha}^{\alpha} f(\chi) = 1$

And it is written as X~ N (μ, σ)

For a normal distribution if mean is '0' and standard deviation is 1 then that variable or variate is known as Standard Normal Variate (SNV). It is generally denoted by t or Z

Thus t or Z ~ N (0,1).

The p.d.f of Z (or t) is

$$f(Z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}Z^2}$$

$-\infty<Z<\infty$

$$\int_{-\alpha}^{\alpha} f(Z)dz=1$$

The graph of f (z) is known as Standard Normal curve.

Properties of Normal distribution/Normal curve

(1) The p.d.f. is given by

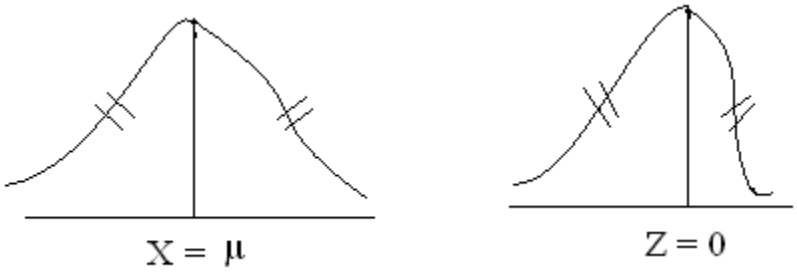
$$f(\chi)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{1}{2}\left(\frac{X-M}{\sigma}\right)^2}$$

$-\infty<\chi<\infty$

$\sigma > 0$

(2) For Normal distribution mean = median = mode = μ

(3) Normal curve is a bell shaped symmetric curve
Symmetric about X = μ or Z = 0



(4) The total area under the curve is unity i.e.

$$\int_{-\alpha}^{\alpha} f(\chi)dx=\int_{-\alpha}^{\alpha} f(Z)dz=1$$

The area under the curve and probability of X or Z between the two values is same concept.
Thus

$$\int_{a}^{b} f(Z)dz=P[a\le z\le b]$$

(5) The normal curve is an asymptotic curve, i.e., two tails of the curve do not touch X axis but remain parallel to X axis.

(6) Following are some standard areas under the normal curve.
   (i) Area between X = μ ± σ is 0.6826 or 68.26%
   (ii) Area between X = μ ± 2σ is 0.9545 or 95.45%
   (iii) Area between X = μ ± 3σ is 0.9973 or 99.73%

(7) Under standard normal curve the areas are as follows:
   (i) Area between Z = ± 1 is 0.6826
   (ii) Area between Z = ± 2 is 0.9545
   (iii) Area between Z = ± 3 is 0.9973

(8) Any normal variate with mean μ and standard deviation σ can be converted into corresponding SNV as

$$Z=\frac{X-\mu}{\sigma}$$

2. Student's t- test/t-distribution

We have so far discussed large sample test Z test for $n \geq 30$.

If $n \geq 30$, population is normally distributed but if standard deviation is not known then sampling distribution of the mean is also normally distributed. But in case of small samples similar to Normal Probability distribution there is another probability distribution known as Student's t-distribution can be applied.

Student's t-distribution is one of the important continuous probability distribution introduced by W.S. Gosset under the pet name 'student' and it is used for testing of hypothesis on small samples.

For small sample tests the concept of 'degrees of freedom' is introduced.

The degrees of freedom is a number which tells us how many of the values may be independently or freely chosen. So as the conditions are satisfied. There is a rule to set degree of freedom as if n is the sample size and one parameter is specified then the degree of freedom is n – 1. If two parameters are specified then n – 2 and so on. If there are two samples of $n_1$and n_2 as sizes for specified means of two population then degree of freedom will be $(n_1 – 1) + (n_2 – 1) = n_1 + n_2$-2.

The probability distribution (p.d.f) of the random variables following t distribution with degree of freedom n – 1 is as follows:

$$f(t)=K\left(1+\frac{t^2}{v}\right)^{-(v+1)/2}$$

$-\infty < t > \infty$

Where *K* is constant.

Properties of t-distribution

(i) t–distribution is symmetrical about the line t = 0.
(ii) t–distribution is asymptotic to the t – axis.
(iii) The shape of t-curve changes with degree of freedom, i.e., sample size.
(iv) t–distribution has a greater dispersion than standard normal distribution.
(v) It is unimodal with Mean = Median = Mode.

Uses of t–test

When

(i) Sample size n is small, i.e., < 30
(ii) The variance of the population is not known.
(iii) The sample is a random one.
(iv) The population is normal.

T-test is used

(i) To test for a specified mean
(ii) To test for equality of two means of two independent samples drawn from two normal population, standard deviation of the population being unknown and
(iii) To test the significance of difference between the means of paired data.

The procedure for tests of significance is same as Z test or general procedure. Here critical value is obtained for required degree of freedom at specified level of significance.

Calculated value of test statistic is compared with critical value

If t calculated ≤ t table (critical value)

H_0 is accepted. Other wise H_0 is rejected.

The formula for test statistic, corresponding degree of freedom are tabulated below.

(I) To test specified mean

$H_0: \mu = \mu_0$

$$t = \left| \frac{\overline{X} - \mu_0}{S/\sqrt{n-1}} \right|$$

d.f. = n – 1

s is sample standard deviation

(II) To test equality of two means (Two sample)

$n_1 < 30$, $n_2 < 30$

$H_0: \mu_1 = \mu_2$

σ_1 σ_2 unknown

Available information : $\overline{X}_1$, $\overline{X}_2$ (Two sample means) n_1, n_2, s_1^2, s_2^2 (sample variances)

Test statistic

$$t = \left| \frac{\overline{X}_1 - \overline{X}_2}{S.E_{\overline{X}_1 - \overline{X}_2}} \right|$$

Where

$$S.E_{\overline{X}_1 - \overline{X}_2} = S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

And

$$S = \sqrt{\frac{n_1 s_1^{\,2} + n_2 s_2^{\,2}}{n_1 + n_2 - 2}}$$

dof = $n_1 + n_2 - 2$

If observations for two samples are available then steps to calculate $\overline{X}_1, \overline{X}_2$ and S.E are as follows:

Let X_{11}, X_{12}, X_{13}…..X $1n_1$ are observations of sample I

X_{21}, X_{22}, X_{23} ….X$2n_2$ are observations of sample II

(i) Obtain $\sum X_1, \sum X_2$

(ii) $\overline{X}_1 = \frac{\sum X_1}{n_1}$ $\overline{X}_2 = \frac{\sum X_2}{n_2}$

(iii) Obtain $X_1 - \overline{X}_1$, $(X_1 - \overline{X}_1)^2$, $X_2 - \overline{X}_2$ $(X_2 - \overline{X}_2)^2$ columns

(iv) Obtain $\sum (X1 - \overline{X}_1)^2$, $\sum (X2 - \overline{X}_2)2$

(v) Obtain $S^2 = \frac{\sum (X_1 - \overline{X}_1)^2 + \sum (X_2 - \overline{X}_2)^2}{n_1 + n_2 - 2}$

(vi) Test statistic $t = \left| \frac{\overline{X}_1 - \overline{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right|$

(iii) Paired t test

$H_0: \mu_1 = \mu_2$ i.e. There is no difference between the averages before and after

Available information: n pairs of observations (Xi, Yi) i = 1, 2….n

dof = n – 1

Steps to calculate test statistic

(I) Obtain di = Xi – Yi fori= 1, 2….n

(II) Obtain $\bar{d} = \frac{\sum di}{n}$

(III) Obtain di – $\bar{d}$, (di – $\bar{d}$)^2

(IV) Obtain $\sum$ (di – $\bar{d}$)^2

(V) $S^2 = \frac{\sum (di - \bar{d})}{n-1} = \frac{1}{n-1}\left[\sum d^2 - \frac{(\sum d^2)}{n} \right]$

(VI) Test statistic $t = \left| \frac{\bar{d}}{S/\sqrt{n}} \right|$

11.3 CHI-SQUARE TEST

The chi-square test is an important test amongst the several tests of significance developed by statisticians. Chi-square, symbolically written as χ^2 (Pronounced as Ki-square), is a statistical measure used in the context of sampling analysis for comparing a variance to a theoretical variance.

As a non-parametric test, it "can be used to determine if categorical data shows dependency or the two classifications are independent. It can also be used to make comparisons between theoretical populations and actual data when categories are used." Thus, the chi-square test is applicable in large number of problems.

The test is, in fact, a technique through the use of which it is possible for all researchers to

(i) test the goodness of fit;
(ii) Test the significance of association between two attributes, and
(iii) Test the homogeneity or the significance of population variance.

1. Chi- Square Distribution

Chi- Square Distribution is one of the important continuous probability distribution. It was first formulated by Helmertand then by Karl Pearson.

If X is a normal variate with mean μ and standard deviation σ then

$Z = \frac{X - \mu}{\sigma}$ Is SNV with mean zero and standard deviation 1.

And

$Z^2 = \left(\frac{X - \mu}{\sigma}\right)^2$ Is a Chi-square variate with 1 degree of freedom. It is denoted by χ^2

If X_1, X_2 ….. X_n are n independent normal variates with means μ_1, μ_2 …. μ_n and standard deviation σ_1, σ_2…. σ_n respectively then

$X^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu_i}{\sigma_i}\right)$ Is a Chi-square variate with n degree of freedom.

The p.d.f. of Chi-square distribution with ν = (n- 1) degree of freedom

is given by

$f\left(X^2\right) = K.\left(X^2\right)\frac{1}{2}^{(\nu-2)}{}_{e^{-}} {}^{X^2/2}$

$0 < \chi^2 < \infty$

Where K is constant and ν= n-1 is degree of freedom.

And the variable following this distribution is known as χ^2variate.

Properties of χ^2-Distribution

(1) Mean = degree of freedom = ν
Mode = ν -2

Standard deviation $\sqrt{2v}$

(2) For n >1 χ^2probability curve is positively skewed extended towards right from 0 to $\propto$.
As degree of freedom increases, skewness decreases and curve tends to symmetry.
Thus as $n \to \propto$, Chi-square $\to$ Normal Distribution.

(3) Sum of two Chi-square variates is also a Chi-square variate with degree of freedom as sum of degrees of freedom.

(4) If X_1, $X_{2....}X_n$ be a random sample from a normal population with mean μ and standard deviation σ then

$$\sum_{i=1}^{n}\left(X_i-\mu\right)^2/\sigma^2$$

Follows χ^2 distribution with ndof.

$$\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2/\sigma^2$$

Follows χ^2 distribution withdof (n-1)

This distribution test is irrespective of sample size. (For large as well as small samples)

Application of χ^2 tests

(i) To test the homogeneity or the significance of population variance, i.e., to test specified value of variance or standard deviation of population

H_0:$\sigma^{2=}\sigma_0^2$ or $\sigma = \sigma_0$

(ii) To test goodness of fit

(iii) To test the significance of association between two attributes, i.e. ,to test independence of attributes.

(1) Test for specified population variance

H_0:$\sigma^{2=}\sigma_0^2$

H_0:$\sigma^2 \neq \sigma_0^2$ or $\sigma^2 > \sigma_0^2$

Test statistic

$$X^2=\frac{\sum\left(X-\bar{X}\right)^2}{\sigma_0^{\,2}}=\frac{ns^2}{\sigma_0^{\,2}}$$

d.f = n – 1

If χ2cal>. χ^2 table, we reject H_0 otherwise we accept H_0.

(2) Test for goodness of fit.

This is very popular test. This test is used to decide whether the discrepancy between theory and the experiment is significant or not, i.e., to test whether the difference between the theoretical and observed values can be attributed to chance or not.

H0: There is no significant difference between the observed values and the corresponding expected (theoretical) values.

H_1: The difference is significant Test statistic

$$X^2=\sum_{i=1}^{n}\frac{\left(O_i-E_i\right)^2}{E_i}$$

$$= \sum_{i=1}^{n} \frac{O_i^2}{E_i} - N$$

Where

n $\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i = N$

dof = n – 1

χ^2 ca l > χ^2 table H_0 is rejected otherwise H_0 is accepted.

Conditions of validity

Above test can be used only when

(i) The total frequency N > 50 (large)

(ii) The observations are independent

(iii) The constraints on the cell frequencies if any are linear

(iv) No theoretical frequency is less than 5. If it is so use pooling technique so that frequency > 5.Either add frequency with preceding or succeeding frequency. Accordingly degree of freedom will be adjusted.

(3) Test for independence of Attributes.

Let the observations be classified according to two attributes and the frequencies O_i in different categories are shown in two way table called as contingency table. And we have to test whether the two attributes are independent or not.

Under the Null Hypothesis H_0 that the two attributes are independent, the expected frequency for cell is calculated as

$$Frequencies\ of\ any\ cell = \frac{Row\ total \times Column\ Total}{Grand Total}$$

Thus expected frequencies for all cells are calculated.

Test Statistic

$$X^2 = \sum_{j=1}^{n} \sum_{i=1}^{m} \frac{(Oij - Eij)^2}{Eij}$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{m} \frac{Oi^2}{Ei} - N$$

Where

$$\sum_j \sum_i Oij = \sum_j \sum_i Eij = N \,.\, Eij = \frac{Ri \times Cj}{N}$$

There are m rows and n columns m x n contingency table.

R_i denote i^{th} row total.

C_j denote j^{th} column total.

E_{ij} = Expected frequency of $(i, j)^{th}$ cell

N is grand total

dof = (m - 1) (n – 1)

If χ^2 cal< χ^2 table we say that H_0 is accepted, i.e., the attributes are independent.

Otherwise H_0 is rejected, i.e., The attributes are **not** independent.

2 x 2 Contingency Table and Yates Correction.

Let A and B denote existence of two attributes, A□' B□ denote not A and not B. respectively.

a, b, c, d are cell frequencies as under

	A	Not A A□	Total
B	a	B	a + b
Not B B□	c	d	c + d
Total	a + c	b + d	a + b + c + d = N

Here

$$X^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

And dof = 1

In 2 x 2 contingency table if any of the cell frequency is less than 5, Yates correction can be applied.

$$X^2\,corrected = \frac{N\left[\left|ad-bc\right|^{-\frac{N}{2}}\right]^2}{(a+b)(c+d)(a+c)(b+d)}$$

dof = 1.

2. ANOVA

ANalysis**O**f**VA**riance abbreviated as ANOVA is an extremely useful technique concerning researches in almost all the fields. This technique is used when multiple sample cases are involved. Using ANOVA one can draw inference about whether the samples have been drawn from populations having the same mean.

The ANOVA technique is important in all those situations where we want to compare more than two populations such as

(i) in comparing the yield of crop from several varieties of seeds.

(ii) the gasoline mileage of say five automobiles.

(iii) the smoking habits of seven groups of university students and so on.

The essence of ANOVA is that the **total amount of variation** in a set of data is broken down into two types that amount which can be **attributed to chance** and that amount which can be **attributed to specified causes**.

There may be variation between samples and also within sample items. ANOVA consists in splitting the variance for analytical purpose.

ANOVA is method of analyzing the variance to which a response is subject into its various components corresponding to various sources of variation.

Through ANOVA technique one can investigate only number of factors which are hypothesized or said to influence the dependent variable.

ANOVA can also be used to investigate the differences amongst various categories within each of these factors which may have a large number of possible values.

If only one factor is taken and the differences amongst its various categories are investigated, it is said that **One way ANOVA** is used.

If two factors are taken simultaneously for investigation then it is **Two way ANOVA** and so on.

In two or more way ANOVA the interaction (i.e., interrelation between two independent variables or factors) if any between two independent variables affecting a dependent variable can also be studied.

Use of ANOVA/Application of ANOVA

(1) To test whether various varieties of seeds or fertilizers or soils differ significantly.
(2) To study the difference in various types of feed prepared for a particular class of animals.
(3) To test whether there is any significant difference between various types of drugs manufactured for curing a disease.
(4) To do analysis of performance of various salesmen, i.e., to test whether their performances differ significantly or not.

Thus basically ANOVA is a technique to test differences among the means of populations by examining the amount of variation within each samples, relative to the amount of variation between the samples

The test statistic used is F ratio based on ratio of variances And has ‘ F’ distribution.

3. Non-parametric Tests

Non-parametric tests have fewer and less stringent assumptions. They do not specify normally distributed populations or homogeneity of variance. Non-parametric tests are the only ones useable with **nominal** data. They are the only technically correct tests that are sometimes employed in this case. Non-parametric tests are also easy to use for **interval** and **ratio** data. These are easy to use and understand.

Parametric tests have greater efficiency when their use is appropriate but even in such cases non parametric test often achieve an efficiency upto 95%.

Chi-square can be used as a non-parametric statistic which is used frequently for cross-tabulation or contingency tables. It’s applications include testing for differences between proportions in populations and testing for independence.

Non-parametric tests are also known as Distribution-free tests.

Some important Non-parametric test with their application area.

(1) Test concerning some single value for the given data (**One sample sign test**)

(2) Test concerning no difference among any two or more sets of data (**Two samples sign test, Fisher – Irwin Test, Rank Sum Test)**

(3) Test of hypothesis of a relationship between variables (Rank correlation, Kendalls Coefficient and other tests of dependence)

(4) Test of a hypothesis concerning variations in the given data (**similar to ANOVA, Kruskal-Wallis Test)**

(5) Tests of randomness of a sample based on theory of runs (**one sample run test)**

(6) Tests of hypothesis to determine if categorical data shows dependency (**Chi-square test for independence of attributes)**

4. SOLVED EXAMPLES

Ex.1. A company has the head office at Kolkata and a branch at Mumbai. The personnel director wanted to know if the workers at the two places would like the introduction of a new plan of work and a survey was conducted for this purpose. Out of a sample of 500 workers at Kolkata 62% favoured the new plan. At Mumbai out of a sample of 400 workers 41% were against the new plan. Is there any significant difference between the two groups in their attitude towards the new plan at 5% level?

Sol. Let $P_1$and P_2 be the population proportions in Kolkata and Mumbai respectively who favour the new plan.

Let the Null Hypothesis be H_0:P_1= P_2

The alternative Hypothesis is H_1:$P_1 \neq P_2$

$$We have\ \ p=\frac{n_1 p_1+n_2 p_2}{n_1+n_2} and\ S.E.of\ (p_1-p_2)=\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$Here n_1=500,\ p_1=0.62, n_2=400,\ p_2=\left(\frac{100-41}{100}\right)=0.59$$

$$\therefore\ p=\frac{500\times 0.62+400\times 0.59}{500+400}=\frac{546}{900}=0.607, q=1-p=0.393$$

$$S.E.of\ (p_1-p_2)=\sqrt{0.607\times 0.393\left(\frac{1}{500}+\frac{1}{400}\right)}$$

$$=\sqrt{\frac{0.607\times 0.393\times 9}{2000}}$$

$$=\sqrt{0.00107}$$

$$=0.0327$$

Assuming that $H_0$is true, the Null Hypothesis at 5% level of significance and conclude that there is no significant difference between the two group in their attitude towards the new plan.

Ex.2 A manufacturer of ball point pens claims that a certain pen he manufactures has a mean writing life of 400 pages with a standard deviation of 20 pages. A purchasing agent selects a sample of

100 pens and puts them for test. The mean writing life for the sample was 390 pages. Should the purchasing agent reject the manufacturer's claim at 5% level.

Sol. Let the null hypothesis H_0 be that the mean writing life of ball pens is 400 pages.

Alternative hypothesis = The mean writing life of ball pens is not 400 pages.

$$Z = \frac{\bar{\chi} - \mu}{SE_{\bar{\chi}}} = \frac{390 - 400}{\sigma / \sqrt{n}} = \frac{-10}{20} \times \sqrt{100} = -5$$

The tabulated value of z at 5% level of significance is 1.96

Since the calculated value is more than the tabulated value at 5% level the claim of the manufacturer is rejected. The purchasing agent should reject the manufacturer's claim, that the mean writing life of pens is 400 pages.

Ex.3 A manufacturer claimed that at least 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at a significance level of

(i) 0.05,
(ii) 0.01.

Sol. Null Hypothesis is that the proportion of equipments conformed to specification is 95% i.e., So H_0 (P = 0.95). Alternative hypothesis is that it is less than 95%. H_1: (P< 0.95). Now equipment found to be not faulty = 200 – 18 = 182.

The proportion of equipments conforming to specifications, i.e., observed value

$$= \frac{182}{200} = 0.91$$

Assuming H_0 to be true, expected value = 0.95

$$S.E. = \sqrt{\frac{0.95 \times 0.05}{200}} = 0.0154$$

$$Z = \frac{0.91 - 0.95}{0.0154} = -\frac{0.04}{0.0154} = -2.597 = -2.60$$

(Approx)

(i) Z = (-2.60) is less than – 1.645 therefore at 5% level, claim is not justified.
(ii) Z = (-2.60) is less than – 2.33 therefore at 1% level, claims not justified.

Note: Since we are interested to check only the lower proportion, one tailed test has been considered.

Ex. 4 In random samples of 600 and 1000 men from two cities, 400 and 600 men are found to be literate. Do the data indicate (at 5% level of significance) that the populations are significantly different in the percentage of literacy?

Sol. Null Hypothesis is H_0: ($P_1 = P_2$); Alternative hypothesis H_1: ($P_1 \neq P_2$)

© $Here\ n_1 = 600, p_1 = \frac{400}{600} = \frac{4}{6}$

$$n_2 = 1000, P_2 = \frac{600}{1000} = \frac{6}{10}; P_1 - P_2 = \frac{4}{6} - \frac{6}{10} = \frac{1}{15}$$

If H_0 is true, the best estimate of the value of p is given by

$$P = \frac{\frac{4}{6} \cdot 600 + \frac{6}{10} \cdot 1000}{600 + 1000} = \frac{1000}{1600} = \frac{10}{16}, q = 1 - \frac{10}{16} = \frac{6}{16}$$

$$S.E.(of\ P_1 - P_2) = \sqrt{\frac{10}{16} \cdot \frac{6}{16}\left(\frac{1}{600} + \frac{1}{100}\right)} = \frac{1}{40}$$

$$Z = \frac{P_1 - P_2}{S.E._{P_1 - P_2}} = \frac{1}{15} \times 40 = 2.67$$

This value of z is greater than 1.96 (at 5% level), so it is significant and we conclude that the difference between the two proportions in percentage of literacy is significant.

Ex. 5 In an infantile paralysis epidemic 500 persons contracted the disease. 300 received no serum treatment and of them 75 became paralysed. Of those who received serum treatment 65 became paralysed. Was serum treatment effective?

Sol. We have the null hypothesis H_0 that the serum treatment is not effective, i.e., $P_1 = P_2$ and Alternative hypothesis H_1: $P_1 < P_2$

Number of persons who received the serum = 500 – 300 = 200

Number of persons who did not receive the serum = 300

The proportion of persons who became paralysed after receiving the serum

$$= \frac{65}{200} = 0.325 = P_1(say)$$

The proportion of persons who became paralysed without receiving the serum

$$= \frac{75}{300} = 0.25 = P_2(say)$$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{65 + 75}{500} = 0.28, q = 1 - 0.28 = 0.72$$

$$S.E.(p_1 - p_2) = \sqrt{0.28 \times 0.72\left(\frac{1}{200} + \frac{1}{300}\right)} = 0.041$$

$$Z = \frac{p_1 - p_2}{S.E.} = \frac{0.325 - 0.25}{0.041} = 1.83$$

At 5% level of significance the tabulated value of Z is 1.64 which is more than the calculated value.

Hence at 5% level of significance the null hypothesis is accepted, i.e., there is no difference in the proportion of persons getting paralysed with or without serum treatment, i.e., the serum treatment was not effective.

Ex.6 In order to make a survey of the buying habits, two markets A and B are chosen at two different parts of a city. 400 women shoppers are chosen at random in market A. Their average weekly expenditure on food is found to be Rs. 250 with a S.D. of Rs. 40. These figures are 220 and Rs. 55 respectively in the markets B where also 400 women shoppers are chosen at random. Test at 1% level of significance whether the average weekly food expenditure of the two populations of shoppers are equal.

Sol.	Given	Sample I	Sample II
	Sample Size	$n_1 = 400$	$n_2 = 400$
	Sample Mean	$\bar{x}_1$ = Rs.250	$\bar{x}_2 = 220$
	Sample S.D	σ_1 = Rs.40	σ_2 = Rs.55

Let μ_1= Mean of first population, And μ_2 = Mean of second population

Null Hypothesis H_0:($\mu_1 = \mu_2$) : Alternative Hypothesis H_1 : ($\mu_1 \neq \mu_2$)

Assuming H_0 to be true,

$$S.E\,of\ \bar{\chi}_1 - \bar{\chi}_2 = \sqrt{\frac{\sigma^2{}_1}{n_1} + \frac{\sigma^2{}_2}{n_2}} = \sqrt{\frac{1600}{400} + \frac{3025}{400}} = 3.4$$

$$Z = \frac{\bar{\chi}_1 - \bar{\chi}_2}{S.E} = \frac{250 - 220}{3.4} = \frac{30}{3.4} = 8.82$$

This value of Z is greater than 2.58 at 1% level of significance, so it lies in food rejection region. So we reject the null hypothesis and conclude that the said average expenditure of two populations of shoppers are not equal.

Ex7. A margarine firm has invited 200 men and women to see if they can distinguish margarine from butter. It is found that 120 of the women, but only 108 of the men can. Investigate whether there is any evidence of sex difference in taste discrimination.

Sol. We have Null Hypothesis, H_0 that there is no evidence of sex difference in taste discrimination, i.e., $p_1 = p_2$

Proportion of women who can distinguish maragarine from butter.

$$= \frac{120}{200} = \frac{3}{5} = 0.6 = p_1$$

Proportion of men who can distinguish maragarine from butter

$$= \frac{108}{200} = 0.54 = p_2$$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{120 + 108}{200 + 200} = \frac{228}{400} = 0.57 = p$$

$$S.E.(of\ p_1 - p_2) = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_1}\right)} = \sqrt{0.57 \times (1 - 0.57)\left(\frac{1}{200} + \frac{1}{200}\right)}$$

$$=\sqrt{(0.57\times 0.43\times 0.01)}=0.0495$$

$$Z=\frac{p_1-p_2}{S.E._{(p_1-p_2)}}=\frac{06-0.54}{0.0495}=1.21$$

The value of Z for two tailed test at 1% level of significance is 2.58 and at 5% level of significance is 1.96.

Since the calculated value is less than the tabulated value, we accept our null hypothesis. Hence there is no evidence of sex difference in taste discrimination.

Ex.8 Random samples drawn from two places the following data relating to the heights of adult males:

	Place A	Place B
Mean height (inches)	68.50	68.58
Standard Deviation (in inches)	2.5	3.0
Number of items in sample	1200	1500

Test, at 5% level, that the mean height is the same for adults in the two places. (Table value of Z of 5% level for two tailed test is 1.96)

Sol. We set $H_0:\mu_1=\mu_2$

$H_1:\mu_1\neq\mu_2$

The standard error of the difference between the number of the samples is given by

$$S.E=\sqrt{\frac{\sigma^2{}_1}{n_1}+\frac{\sigma^2{}_2}{n_2}}=\sqrt{\frac{(2.5)^2}{1200}+\frac{(3.0)^2}{1500}}$$

$$=\sqrt{0.0052+0.006}=0.1058$$

$$Z=\frac{Differenceof\ means}{S.E.of\ means}$$

$$=\frac{68.58-68.50}{0.1058}$$

$$=\frac{0.08}{0.1058}=0.756$$

Computed value of Z being less than the table value, we cannot reject the null hypothesis, and so the mean height for adults in the two places.

Ex.9 You are given the gain in weights (lbs) of cows fed on two diets of X and Y

GAIN IN WEIGHT (lbs)									
Diet X	25	32	30	32	24	14	32		
Diet Y	24	34	22	30	42	31	30	32	35

Test, at 5% level, whether the two diets differ as regards their effect on mean increases in weight. (tabulated value of 't' for 15 degrees of freedom at 5% = 1.753)

Sol. Let us take null hypothesis that the two diets X and Y do not differ significantly as regards their effect on increase in weight. Appling t-test of difference of means:

$$f = \frac{\overline{X} - \overline{Y}}{S} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$\overline{X} = \frac{25+32+30+32+24+14+32}{7} = 27$$

$$= \frac{\Sigma Y}{n} = \frac{\Sigma Y}{10} = \frac{320}{10} \Rightarrow \overline{Y} = 32$$

X	(x – 27) = dx $\overline{X} = 27$	(X – 27)2	Y	(y – 32) dy $\overline{Y} = 32$	(Y – 32)2
25	-2	4	24	-8	64
32	+5	25	34	+2	4
30	+3	9	22	-10	100
32	+5	25	30	-2	4
24	-3	9	42	+10	100
14	-13	169	31	-1	1
32	+5	25	40	+8	64
			30	-2	4
			32	0	0
			35	+3	9
ΣX = 189	Σ (x – 27) =Σdx = 0	Σ(X – 27)2 = 266	ΣY = 320	Σ(y – 32) = Σdy = 0	Σ(Y – 32)2 = 350

$$S = \sqrt{\frac{\Sigma\left(X - \overline{X}\right)^2 + \Sigma\left(Y - \overline{Y}\right)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{266 + 350}{7 + 10 - 2}} = 6.408$$

$$\textit{Common variance} = \frac{n_1 S^2{}_1 + n_2 S^2{}_2}{n_1 + n_2 - 2} = \frac{\left(X - \overline{X}\right)^2 + \left(Y - \overline{Y}\right)^2}{n_1 + n_2 - 2}$$

$$t = \frac{27 - 32}{6.408}\sqrt{\frac{7 \times 10}{7 \times 10}} = \frac{-5}{6.408} \times 2.029 = -1.583 = 1.583$$

[Absolute value]

d.o.f. = $n_1 + n_2 - 2 = 7 + 10 - 2 = 15$

For 15 d.o.f. $t_{0.05} = 1.753$. The calculated value of | t | is less than the table value. The null hypothesis is accepted. Hence the two diets do not differ significantly with regarded to their effect on mean increases in weight.

Ex.10 The following data show the cost per square foot of floor area connecting randomly selected 7 schools and 5 office blocks from those completed during the period 1984 to 1989.

Building Type	**Cost Per Square Foot (Rs.)**						
Schools	28	31	26	27	23	38	37
Office blocks	37	42	34	37	35		

Do the data support the hypothesis that the cost per square foot for office blocks was greater than that for schools? Test at 5% level of significance using ‘t’test.

Sol. Let us the null hypothesis that the cost per square foot for office blocks was not greater than that for schools. Applying t-test of difference of means.$H_0: \mu_{sc} = \mu_{off}$

$H_1: \mu_{sch} = \mu_{off.}$

S. No	X_1	$(X_1-\bar{X}_1)$	$(X_1-\bar{X}_1)^2$	S. No.	X_2	$(X_2-\bar{X}_2)$	$(X_2-\bar{X}_2)^2$
1	28	-2	4	1	37	0	0
2	31	+1	1	2	42	5	25
3	26	-4	16	3	34	-3	9
4	27	-3	9	4	37	0	0
5	23	-7	49	5	35	2	4
6	38	+8	64				
7	37	+7	49				
N_1 – 7	$\Sigma X_1 = 210$	$\Sigma(X_1-\bar{X}_1)=0$	$\Sigma(X_1-\bar{X}_1)^2 = 192$	n_2 = 5	$\Sigma X_2 = 185$	$\Sigma(X_2-\bar{X}_2)=0$	$\Sigma(X_2-\bar{X}_2)^2 = 38$

$$t=\frac{(\bar{X}_1-\bar{X}_2)}{S}\sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

$$\bar{X}_1=\frac{\Sigma X_1}{n_1}=\frac{210}{7}=30;\bar{X}_2=\frac{\Sigma X_2}{n_2}=\frac{185}{5}=37$$

$$S=\sqrt{\frac{\Sigma(X_1-\bar{X}_1)^2+\Sigma(X_2-\bar{X}_2)^2}{n_1+n_2-2}}=\sqrt{\frac{192+38}{7+5-2}}=4.796$$

$$\therefore t=\frac{30-37}{4.796}\sqrt{\frac{7\times 5}{7+5}}=\frac{7\times 1.708}{4.796}=2.491$$

Degree of freedom = $n_1 + n_2 - 2 = 10$

For 10 degrees of freedom, the calculated value of t at 5% level of significance for one tailed (right tail) test i.e., $t_{0.05}$ = 1.812. The calculated value of t is greater than the table value. The hypothesis is rejected. The cost per squares foot for office blocks was greater than that for schools.

Ex.11 A soap manufacturing company was distributing a particular brand of soap through a large number of retail shops. Before a heavy advertisement compaign the mean sales per week per shop was 140 dozens. After the compaign, a sample of 26 shops was taken and the mean sales was found to be 147 with standard deviation 16. What conclusion do you draw on the impact of advertisement on sales? Use 5% significance level.

Sol. We set up the hypothesis

H_0:μ = 140 i.e., the compaing is not effective

H_0:$\mu \neq 140$

It is given that S = 16, n = 26, $\bar{x}$ = 147

The unbiased estimate of the S.E. of the mean is given by

$$S.E._{\bar{X}}=\frac{S}{\sqrt{n-1}}=\frac{16}{5}=3.2 \text{ where } S \text{ is the } S.D. \text{ of the sample}$$

$$Now\, t=\frac{|\bar{X}-\mu|}{S.E._{\bar{X}}}=\frac{|147-140|}{3.2}=\frac{7}{3.2}=2.19$$

From the table, for 25 degrees of freedom $t_{0.05} = 1.708$ [for one tailed test]*

Since computed (or calculated) value of $t > t_{0.05}$, we reject the null hypothesis. i.e, advertisement may be considered to have changed the average sales volume or we can say the compaign had impact on sales.

Ex.12 Two salesmen A and B are working in a certain district. From a sample survey conducted by the Head Office. The following results were obtained. State whether there is any significant difference in the average sales between the two salesmen:

	A	B
No. of Sales	$20 = n_1$	$18 = n_2$
Average Sales in Rs.	170	205
Samples s.d. in Rs.	$20 = S_1$	$25 = S_2$

Sol. We set the null and alternative hypothesis as follows:

$H_0: \mu_1 = \mu_2$ i.e., there is no difference in the average slaes

$H_1: \mu_1 \neq \mu_2$

$$Unbiased\ estimate\ S^2\ of\ the\ common\ variance = \frac{n_1 S^2_1 + n_2 S^2_2}{n_1 + n_2 - 2}$$

$$= \frac{20 \times 400 + 18 \times 625}{36} = 534.72 \Rightarrow s = \sqrt{534.72} = 23.12$$

$$t = \frac{(\overline{X}_1 - \overline{X}_2)}{S\sqrt{\frac{n_1 n_2}{n_1 + n_2}}} = \frac{170 - 205}{23.12\sqrt{\frac{1}{20} + \frac{1}{18}}} = -4.66 \; or \; |t| = 4.66$$

*Here one tailed test is used because under normal circumstances, the sales can be expected to increase as a result of compaign. However if we use two tailed test then $t_{0.025} = 2.06$. Even then we conclude that the compaign was effective. But one tailed test is more suitable here.

Since the calculated value of t is much greater than 3, the null hypothesis is rejected, and we conclude that the average sales of the two salesmen are significantly different.

Ex.13 Two type of batteries are tested for their length of life and the following data are obtained.

	No. of samples	**Mean life in hours**	**Variance**
Type A	9	600	121
Type B	8	640	144

Is there a significant difference in the two mean? Value of t for 15 degrees of freedom at 5% level is 2.131.

Sol. The null and alternative hypotheses are, $H_0: \mu_1 = \mu_2$i.e. the two type of batteries an identical i.e., statistically there is no difference between their mean lives.

$H_1: \mu_1 \neq \mu_2$ i.e. the two type of batteries are different with regard to their mean life.

An unbiased estimate of the common population s.d. is given by

$$S_p = \sqrt{\frac{(n_1 - 1)s^2_1 + (n2 - 1)s^2_2}{n_1 + n_2 - 2}}$$

S^2_1, S^2_2 and n_1, n_2 being respective sample variances, and the corresponding sample sizes.

Here $n_1 = 9$, $S^2_1 = 121$, $n_2 = 8$, $S^2_2 = 144$

$$S_p = \sqrt{\frac{(9-1)\times 121 + (8-1)\times 144}{9+8-2}}$$

$$= \sqrt{\frac{8\times 121 + 7\times 144}{15}} = \sqrt{\frac{968\times 1008}{15}}$$

$$= \sqrt{\frac{1976}{15}} = \sqrt{1341.733} = 11.47$$

The standard error of the difference between the two means is given by

In some books the formula variance is

$$S = \frac{n_1 S^2_1 n_2 S^2_2}{n_1 + n_2 - 2}$$

In the above formula unbiased estimate of S.D. has been assumed.

$$S.E = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 11.47 \times \sqrt{\frac{1}{9} + \frac{1}{8}} = 11.47 \times \sqrt{\frac{17}{12}}$$

$$= 11.47 \times \sqrt{0.236} = 11.47 \times 0.486 = 5.57$$

Now

$$t = \frac{\bar{\chi}_1 - \bar{\chi}_2}{S.E}$$

where $\bar{x}_1$, and $\bar{x}_2$ are respectively the means of the first and the second sample.

$$= \frac{600-640}{5.57} = -\frac{40}{5.57} = -7.18 \therefore |t| = 7.18$$

Degrees of freedom = 9 + 8 – 2 = 15.

Table value of t for d.o.f at 5% level of significance [two tails test] = 2.131

Since computed value of | t | is more than the table value, the difference between the means is significant.

Ex.14. IQ test was administered to 5 persons before and after they were trained. The results are given below:

Candidates	I	II	III	IV	V
IQ before Training	110	120	123	132	125
IQ after Training	120	118	125	136	121

Test whether there is any change in IQ after the training programme.

[$t_{0.05}$ (4) = 4.6]

Sol. Let us apply paired t test.

Let the null hypothesis be $H_0:\mu_1 = \mu_2$, there is no significant effect of the training.

The alternative hypothesis is $H_1:\mu_1 \neq \mu_2$, i.e., the IQ before training is less than the IQ after training.

Candidates	IQ before Training (x)	IQ after Training (y)	d = y – x	d^2
I	110	120	10	100
II	120	118	-2	4
III	123	125	2	4
IV	132	136	4	16
V	125	121	-4	16
Total			$\Sigma d = 10$	$\Sigma d^2 = 140$

$$\therefore \bar{d} = \frac{\Sigma d}{n} = \frac{10}{5} = 2$$

$$t = \frac{\bar{d} n \sqrt{n-1}}{\sqrt{n\Sigma d^2 - (\Sigma d^2)}}$$

$$= \frac{2 \times 5\sqrt{4}}{\sqrt{5} \times 140 - 100} = \frac{2}{\sqrt{6}} = 0.82$$

And d.f. = n – 1 = 5 – 1 = 4

Thus t = 0.82 < 4.6 at 1% level with 4 d.f. (Two Tailed)

Since the calculated value of t < the tabulated value with 4 d.f. at 1% level, we accept H_0 at 1 % level and conclude that there is no significant change in IQ after the training programme.

Ex.15. The sales data of an item in six shops before and after a special promotional compaign are as under:

Shops	A	B	C	D	E	F
Before Compaign	53	28	31	48	50	42
After Compaign	58	29	30	55	56	45

Can the compaign be judged to be a success? Test 5% level of significance.

Sol. The compaign will be success if there is a significant increase in the average sales after compaign. In this case we have to consider the significance on one side only i.e., increase in sales.

Shops	SALES		d = X1 – X2	
	Before Compaign	After Compaign	d	d2
A	53	58	-5	25
B	28	29	-1	1
C	31	30	1	1
D	48	55	-7	49
E	50	56	-6	36
F	42	45	-3	9
n = 6			$\Sigma d = -21$	$\Sigma d2 = 121$

Null Hypothesis $H_0:\mu_1 = \mu_2$ i.e., there is no difference in the average sales before and after the compaign.

Alternative hypothesis $H_1: \mu_1 < \mu_2$ i.e., average sales have improved after the compaign.

$$\bar{d} = \frac{21}{6} = 3.5, \sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{121}{6} - (3.5)^2} = 2.8136$$

$$Unbiased\ estimate\ of\ S.E. = \frac{\sigma}{\sqrt{n-1}}$$

$$t = \frac{\bar{d}}{SE} = \frac{3.5}{1.258} = 2.78$$

The tabulated value of t at 5 d.o.f. at 5% level of significance [one tailed test] is 2.015.

The computed value of t = 2.78 being more than the table value, we reject $H_0$and conclude the sales compaign has been a success.

Important Note

Some author use the letter s for standard deviation and calculate the unbiased estimate of S.D. by using the formula

$$S = \sqrt{\frac{\Sigma(X - \bar{\chi})^2}{n-1}} \quad \text{and then} \quad SE = \frac{S}{\sqrt{n}}$$

But the general formula for S.E. for pair t-test remains the same i.e.,

$$S.E = \frac{\sqrt{n\Sigma d^2 - (\Sigma d^2)}}{n\sqrt{n-1}}$$

$$= \frac{\sqrt{n\Sigma d^2 - n(\bar{d})^2}}{n(n-1)}$$

Ex.16 10 Accountants were given intensive coaching and four tests were conducted in a month. The scores of tests 1 and 4 given below:

Serial No. of Accountants	1	2	3	4	5	6	7	8	9	10
Marks in 1st test	50	42	51	42	60	41	70	55	62	38
Marks in 4th test	62	40	61	52	68	51	64	63	72	50

Does the score from the 1 to test 4 show an improvement? Test at 5% level of significance. (The value of t for 9 d.o.f. at 5% level for one tail test is 1.833 and for two tail test is 2.262)

Sol. Let us denote the score of first test with suffix 1 and that of the fourth test with suffix 4 and then taking the null hypothesis that there is no improvement, we can write.

$H_0: \mu_1 = \mu_4$ (i.e., there is no improvement)

$H_1: \mu_4 > \mu_1$ (i.e., the coaching has resulted in improvement)

Since we have matched pairs, we use paired t-test and work out test static 't' given by:

$$t = \bar{d} / S.E. \text{ where } \bar{d} = mean\ of\ d, and\ S.E. = \frac{\sqrt{n\Sigma d^2 - n(\bar{d})^2}}{n(n-1)}$$

Marks in 1st test x_1	**Marks in 4th test x_2**	**Difference d = x_1 - x_2**	**Difference Square d^2**
50	62	+12	144
42	40	-2	4
51	61	+10	100
42	52	+10	100
60	68	+8	64
41	51	+10	100
70	64	-6	36
55	63	+8	64
62	72	+10	100
38	50	+12	144
n = 10		Σ d = 72	Σd^2 = 856

$\bar{d} = 72/10 = 7.2$

$$S.E = \frac{\sqrt{n\Sigma d^2 - \left(\Sigma d^2\right)}}{n\sqrt{n-1}} = \frac{\sqrt{856 - 10\left(7.2\right)^2}}{10\left(10-1\right)} = 1.937$$

$$Hence\ t = \frac{\text{d}}{\text{S.E.}} = \frac{7.2}{1.937} = 3.717$$

Degrees of freedom = n – 1 = 10 – 1 = 9

As H_1 is one sided, we shall apply one tailed test (in the right tail, because H_1 is greater than type for determining the rejection region at 5% level.)

The observed value of t = (3.717), is more than 1.833, and hence, in the rejection region. Accordingly, we reject H_0 (i.e., we accept H_1), and conclude that coaching has improved the standard.

Ex. 17 A certain diet newly introduced to each of the 12 pigs resulted in the following increases of body weight:

6, 3, 8, -2, 3, 0, -1, 1, 6, 0, 5 and 4.

Can you conclude that the diet effective in increasing the weight of the pigs? (given $t_{0.05}$ for 11 d.o.f.= 2.20)

Sol. Let x: 6, 3, 8, -2, 3, 0, -1, 1, 6, 0, 5 , 4 so that $\sum$ x = 33.

$\sum x^2 = 36 + 9 + 64 + 4 + 9 + 0 + 1 + 1 + 36 + 0 + 25 + 16 = 201$

$$\bar{X} = \frac{33}{12} = 2.75; S = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma X}{n}\right)^2} = \sqrt{\frac{201}{12} - \left(2.75\right)^2}$$

H_0:μ_0= 0, i.e., diet is not effective

$$Now\ test\ statistic\ is\ t = \frac{\bar{X}}{S}\sqrt{n-1} = \frac{2.75}{\sqrt{9.1875}}\sqrt{12-1}$$

$$= \frac{2.75 \times 3.3166}{3.0311} = 3.009$$

At 5% level of significance (d.f. = 11), tabulated value of t is 2.20 (two tailed test) and the calculated value is greater than the tabulated value, so we reject $H_0$and conclude that the diet is effective. [In fact it is question of paired t-test].

Note: If we consider single tailed test, the value of t = 1.796. Since we are interested only in the increase in weight we should prefer one tailed test because two tailed test will simply test the change in weight.

Ex.18 The number of car accidents per month in a metropolitan city was found as below:

20, 17, 12, 6, 7, 15, 8, 5, 16 and 14. Use chi- square test to check whether these frequencies are in agreement with the belief that occurrence of accidents was the same during the 10 months period. Test at 5% level of significance.

Sol. We have the null hypothesis, H_0, that the occurrence of accidents was the same during the ten month period.

Total number of accidents in 10 months. = 20 + 17 + 12 + 6 + 7 + 15 + 8 + 5 + 16 + 14 = 120

$$\text{Expected number (Average number of accidents) per month} = \frac{120}{10} = 12$$

$$X^2 = \sum_{i-1}^{10} \left\{ \frac{(Oi - Ei)^2}{Ei} \right\}$$

$$= \frac{(20-12)^2}{12} + \frac{(17-12)^2}{12} + \frac{(12-12)^2}{12} + \frac{(6-12)^2}{12} + \frac{(7-12)^2}{12} + \frac{(15-12)^2}{12} + \frac{(8-12)^2}{12} + \frac{(5-12)^2}{12}$$

$$+ \frac{(16-12)^2}{12} + \frac{(14-12)^2}{12}$$

$$= \frac{64}{12} + \frac{25}{12} + 0 + \frac{36}{12} + \frac{25}{12} + \frac{9}{12} + \frac{16}{12} + \frac{49}{12} + \frac{16}{12} + \frac{4}{12}$$

$$= \frac{244}{12} = 20.33$$

Degree of freedom = 10 – 1 = 9

The tabulated value of X2 for 9 degree of freedom for two tailed test at 5% level of significance is 19.02. Since the calculated value is more than the tabulated value, we reject our null hypothesis and say that the given data do not support the belief that the number of accidents were same during 10 months period.

Ex.19 A sample analysis of examination results of 500 students was made. It was found that 180 students had failed, 170 had secured a third class, 110 were placed in second class and 40 got a first class. Are these figures commensurate with the general examination result which is in the ratio 4:3:2:1 for the various categories respectively? Answer at $\alpha = 0.05$ (Table values of chi-squares at α 0.05 for 3 d.f. and 4 d.f. and 4 d.f. are 5.99, 7.81 and 9.49 respectively).

Sol. We have the null hypothesis H_0, that the result of the examination were commensurate with the general examination result which is in the ratio 4:3:2:1.

Total number of students = 500

Observed number of failed students = 180

$$Expected\,number\,of\ \ failed\ students = \frac{500}{4+3+2+1}\times 4 = 200$$

Observed number of students getting third class = 170

$$Expected\,number\,of\ students\,getting\,third\,class = \frac{500}{10}\times 3 = 150$$

Observed number of students getting second class = 110

$$Expected\,number\,of\ students\,getting\,second\,class = \frac{500}{10}\times 2 = 100$$

Observed number of students getting first class = 40

$$Expected\,number\,of\ students\,getting\ \ first\,class = \frac{500}{10}\times 1 = 50$$

$$\therefore \mathrm{X}^2 = S\frac{(\mathrm{O}i-\mathrm{E}i)^2}{\mathrm{E}i} = \frac{(180-200)^2}{200}+\frac{(170-150)^2}{150}+\frac{(110-100)^2}{100}+\frac{(40-50)^2}{50}$$

= 2 + 2.67 + 1 + 2 = 7.67

Value of X2 at 3d.f. for $\alpha = 0.05$ is given to be 7.81

As the calculated value is less than the tabulated value, we accept our null hypothesis and say that the observed figures are quite commensurate with the general examination result.

Ex.20 The following table shows the distribution of goals in football match:

No.of Goals:	0	1	2	3	4	5	6	7
No.Of Matches	95	158	108	63	40	9	5	2

Fit a Poisson distribution and test the goodness of fit.

Sol. We have null hypothesis H_0, that the Poisson distribution can be fitted to the data.

Total number of matches = 95 + 158 + 108 + 63 + 40 +9 + 5 + 2 = 480

Total number of goals = 95 × 0 + 158 × 1 + 63 ×3 + 40 × 4 + 9 × 5 + 5 × 6 + 2 × 7 = 812

$$\mathrm{A}verage number\,of\ goals\,per\,match = \frac{812}{480} = 1.7\,(approx)$$

The expected frequencies of the Poisson distribution are computed from the expression:

$$\mathrm{E}xpected\ frequency = \mathrm{N}\left(\frac{e^{-m}m^x}{x!}\right), Hence\,\mathrm{N} = 480, m = 1.7$$

$\therefore$ For Poisson distribution we have:

$$\mathrm{E}xpected\,Frequency = 480\frac{e^{-}(1.7)^x}{x!}$$

Where x = 0,1,2,3,4,5,6 and 7.

Working out the successive terms of this distribution, we get the following frequencies (results expressed to the nearest whole number):

No. of Goals	Observed Frequency	Expected Frequency
0	95	88
1	158	150
2	108	126
3	63	72
4	40	30
5	9	10
6	5 (16)	3 (14)
7	2	1

Since no expected frequency should be less than 5, we pooled the last three frequencies.

$$X^2 = \Sigma \frac{(Oi - Ei)^2}{Ei}$$

$$X^2 = \frac{(95-88)^2}{88} + \frac{(158-150)^2}{150} + \frac{(108-126)^2}{126} + \frac{(63-72)^2}{72} + \frac{(40-30)^2}{30} + \frac{(16-14)^2}{14}$$

$X^2 = 0.56 + 0.43 + 2.57 + 1.12 + 3.33 + 0.29 = 8.30$

Then no. of degrees of freedom: 6 – 2 = 4 (Pl. Note)

For 4 degrees of freedom at 5% level of significance, the table value of $X^2 = 9.488$, while the calculated value $X^2 = 8.30$. Since the calculated value is less than the tabulated value, difference between expected and observed frequencies is not significant and can be ignored. Hence the fit is good.

11.4 QUESTIONS

1. Explain various tools available for testing of hypotheses
2. Explain parametric tests?
3. Write a note on chi-square test?

● ● ●

Chapter
12

Forecasting Techniques

Structure:

12.1 FORECASTING TECHNIQUES

1. Introduction

Forecasting is a key element of management decision making. Since the ultimate effectiveness of any decision depends upon a sequence of events following the decision should permit an improved choice over that which would otherwise be made.

Forecasting techniques are useful in

(a) Inventory management to estimate the usage rate for each part in order to determine procurement quantities.
(b) Production planning to forecast unit sales for each item by delivery period for a number of months in the future.

Forecasting is an integral part of planning process. Correct forecasting is not usually possible because of uncertainty which inevitably attaches to the future. By forecasting we only try to minimize the impact of uncertainty. Thus forecasting is only a means of attempting to reduce uncertainty of the future and not of eliminating it.

Forecasting can be both short-term and long-term.

2. Forecasting Techniques

Various techniques are available for forecasting. The choice of a method is generally dictated by data availability and/or by urgency of forecast. Many times forecast are forced to use less reliable method for the required data as the use of more reliable method are not always available. If the use of better techniques is time consuming and forecast are urgently needed, forecast are made on the basis of easy and less reliable techniques.

Following are some commonly used techniques

(1) Historical Analogy Method.
(2) Executive Opinion Method.
(3) Survey Techniques.
(4) Barometric Techniques.
(5) Regression Analysis.

(6) Time Series Analysis.
(7) Exponential Smoothing.
(8) Input-Output Models.

12.2 TIME SERIES ANALYSIS METHOD

A Time Series is a set of observations taken at specified times, usually (but not always) at equal intervals. Thus a set of data depending on time (which may be year, quarter, month, week, days, etc.)

It is an arrangement of numerical data in chronological order. The time series data shows certain definitive patterns which can be meaningfully analysed for purpose of projection.

Examples of Time Series are

(a) The Annual Production of Tea/Coffee in India over the last 20 years;
(b) The Monthly Sales of a Chemical Industry for the last 6 months;
(c) The daily closing price of a share in Bombay Stock Exchange;
(d) Yearly Price or Quantity Index numbers.

Analysis of Time Series helps us to understand the past behaviour of time series data. One can understand the changes that took place in the past. With the knowledge of the past behaviour, it would be possible, within certain limits, to forecast for the probable future variations (or movements) of such data. Thus it helps in planning future operations.

With the help of Time Series Analysis, we can compare the actual performance with the expected performance and analyse the cause of variation.

Analysis of Time Series shows that the observed values of the variable are fluctuating from time to time.

The fluctuations are due to various factors (or forces) like changes in habits and tastes of people, weather conditions, etc. On the action of these forces, the values of the variable are changing with time.

The object of time series analysis is to isolate and ascertain these forces (i.e., the various components).

Time series smoothing is done to minimize the influence of extreme values in the historical data which might have been caused by random factors. The smoothing process brings out the underlying pattern in the time series data.

The underlying pattern may be horizontal or may involve some fluctuations or trend, a steady increase or steady decrease. The techniques of simple moving averages and weighted moving averages are use for smoothing the horizontal patterns. Where there is as evident trends in the time series, least square method is used.

1. Components of Time Series

Fluctuations in a Time Series are mainly due to four basic types of variations.

These four types of movements are called the four components or elements of Time Series.

The four components are:

(1) Secular Trend or Trend (T),
(2) Seasonal Variation (S),

(3) Cyclical Variation or Cyclical Fluctuation (C),

(4) Irregular or Random Movement (I).

The changes in Time Series data are the result of the combined effect of these four components.

(1) Secular Trend or Trend (T)

The trend factor is the long-term underlying movement of time series. It may be a steady state trend or a growing or declining trend.

Trend means general tendency of the variable in increasing or decreasing direction. The upward trend in economics and business time series is due to the increase in population, advances in technology, etc. The decline or downward trend is because of decreasing demand of the product concerned or availability of a better substitute in its place or change in the Government's economic policy, etc. In our practical life, we may notice upward trends in a series concerning population, National Income, Bank-deposits, etc., and downward trends in a series concerning birth and death rates.

(2) Seasonal Variation (S),

The seasonal factor is the periodic pattern in the data during the course of a year. Seasonal Variation is a fluctuations due to season. Season is a period of one year or shorter. It recurs periodically year after year. Due to the presence of seasonal variation, business activities are found to have brisk and slack periods at different parts of the year. The major factors that cause seasonal variations are climatic and weather conditions, customs and habits of people; religious festivals, etc. For example, the demand for electric fans rises in summer and falls in winter. The sales and profits of department stores show a sudden rise before the Durga Puja, Diwali and Christmas. The prices of grain vary between the harvest and the non-harvest seasons.

Although the period of seasonal variations refers to a year in business and economics, it can also be taken as a month, week, day, hour, etc., depending on the type of the data available. For example, seasonal fluctuations can be observed in the sales of a departmental store during 12 months of a year, withdrawals (or deposits) in a bank during the days of a month, number of books issued by a library during the 7 days in a week, temperatures recorded during the 24 hours of a day, etc.

(3) Cyclical Variation or Cyclical Fluctuation (C)

The cyclical factor is the periodic ups and downs in the observations forming into a cycle every few years. Cyclical fluctuation is a long term periodic movement which occurs over a long period of time -usually two or more years. It is Oscillatory in nature, but it is not as regular as seasonal variation. One complete period which is more than a year is called a cycle. Cyclical fluctuations are found to exist in almost all business and economic time series where it is known as business cycle or trade cycle. The ups and downs (or rises and decline) in business, recurring at intervals of times, are the effects of cyclical variations. A business cycle showing the recurrence of the up and down movements of business activity consists of four phases:

(i) Prosperity

(ii) Recession

(iii) Depression and

(iv) Recovery

Each phase changes gradually into the next phase in the order given above until one business cycle is completed. The study of cyclical fluctuations is very useful in framing suitable policies for

avoiding periods of booms and depressions in business activity as both are equally bad for an economy. Depression may cause a complete disaster and shatter the economy.

(4) Irregular or Random Movement (I)

The random factor arises out of erratic events which do not occur frequently. Irregular or Random movements, as the name indicates, are such variations in business activity which are caused by factors of irregular (or erratic) nature. These are purely random and unpredictable. These include all movements not already covered in trend, seasonal variation and cyclical fluctuation. Irregular movements are caused by unforeseen events like floods, wars, earthquakes, strikes, elections, etc. Random movements, also known as Residual variations.

The techniques of moving averages, simple linear or non-linear regression equation can be applied to isolate the above factors and they can be cumulatively analysed to forecast the future values.

Time series may be used for both short-term and long-term forecast. But more useful for short term forecast.

2. Mathematical Models of Time Series

In traditional or classical time series analysis, a Multiplicative relationship between the four components is usually assumed, i.e., any particular observation is considered to be thc: the effects of four components. Symbolically, Y=TxS xCxI where Y = the result of the four components (or original data).

Instead of the multiplicative model, some statisticians may prefer additive model as Y=T+S+C+I in which Y is the sum of the four components

3. Measurement of Trend

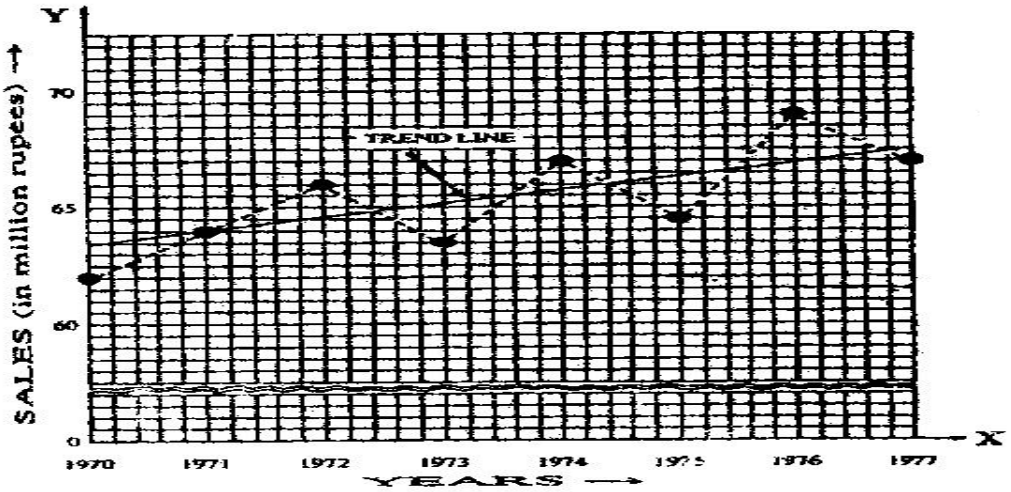
There are four methods for determining trend in time series

Methods of estimation of Trend:

(i) Freehand (or Graphic) Method
(ii) Semi-Average Method
(iii) Moving Average Method
(iv) Method of Least Squares

(i) Free hand Method (or Graphical Method)

Freehand (or., Graphic) method is the simplest method for studying trend, In this method, the actual figures (given data) are first plotted as points on a graph paper showing the time series data along the vertical axis OY and time t along the horizontal axis OX. Then a straight line i.e. a freehand smooth curve of is drawn to fit as closely as possible the plotted points. (To draw the line, leave equal number of points on both sides of it at more or less equal distances), The line so obtained shows the direction of the trend and the vertical distance of this line from OX gives the trend value for each time period.By this methods quick estimate of the trend can be obtained, but this depends too much on who draws the curve.



(ii) Semi-average Method

In the Semi-average Method, the given data is first divided into two parts (preferably equal) and an average (i.e., A.M.) for each part is found. Then these two averages are plotted on a graph paper as point against the mid-points of the time intervals covered by the respective two parts. These two points are joined by a straight line. This straight line is the required trend line and the distances of the line from the horizontal axis OX give the trend values.

Although this method is simple to apply, it may lead to poor results when used indiscriminately. It is applicable only where the trend is linear or approximately linear.

Example 1: Draw a trend line by the Semi-average Method using the following data:

Year :	1973	1974	1975	1976	1977	1978
Production of Steel (in lakh tones) :	253	260	255	266	259	264

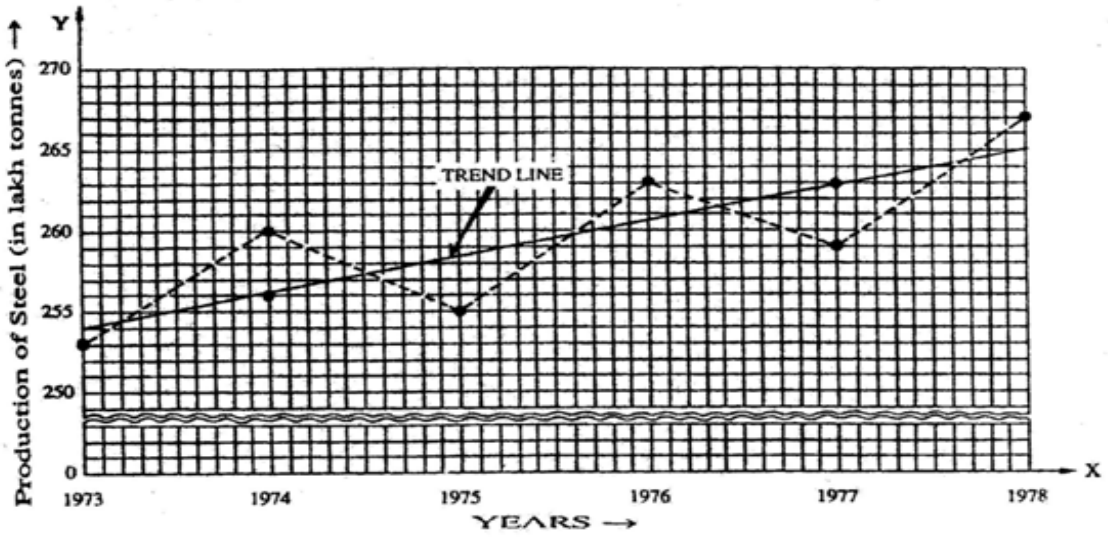
Solution: The average production of Steel for the first three years

$$=\frac{253+260+255}{3}=\frac{768}{3}=256\,lakh\,tonnes$$

And the average production for last three years

$$=\frac{263+259+264}{3}=\frac{786}{3}=262\,lakh\,tonnes$$

Thus we get two points 256 and 262 which are plotted against the respective middle years (mid-pints) 1974 and 1977 of two parts 1973-75 and 1976-78. By joining these two points, the required trend line is obtained. (see fig. given below)



(iii) Moving Average Method

For a given numbers $Y_1, Y_2, Y_3, \ldots$, we define moving totals of order N by the sums

$Y_1 + Y_2 + \ldots Y_N, Y_2 + Y_3 + \ldots + Y_{N+1}, Y_3 + Y_4 + \ldots + Y_{N+2}, \ldots$

And moving averages of order N by the sequence of arithmetic means

$$\frac{Y_1 + Y_2 + \ldots Y_N}{N}, \frac{Y_2 + Y_3 + \ldots + Y_{N+1}}{N}, \frac{Y_3 + Y_4 + \ldots + Y_{N+2}}{N}, \ldots$$

In Moving Average Method, a series of moving averages of specific order is calculated. Starting from the beginning of the given series, an average for a specific number of years for yearly data or a time interval (called period) is calculated and this is placed against the mid-point of the time interval. Keeping the period fixed the process is replaced by dropping the first yearly figure of the given values and adding the figure of the next year, we had not added before. We continue with this till the end of the series is reached.

If the period of moving average is odd, the moving totals and moving averages correspond to the given years of time. But if the period is even, a two-point moving average of the moving averages is to be found for centering them, i.e., for synchronizing the moving averages with the original data (see example 3 (ii) given below).

This method is commonly used for measuring trend. By using moving averages of appropriate orders, cyclical fluctuation, seasonal and irregular movements may be eliminated, leaving only the trend movement.

If the moving averages are strongly affected by extreme values, a weighted moving average with appropriate weights is sometime used.

Advantages and Disadvantages

(i) This method is used to measure trend seasonal, cyclical and irregular fluctuations.

(ii) Moving average method is easy to apply as this method does not involve any difficult calculation.

(iii) If an appropriate period is chosen (i.e., if the period of the moving average coincide with the period of cyclical fluctuations), then these fluctuations are automatically eliminated from the data by using this method.

(iv) The choice of the period of moving averages is made by observing the oscillatory movements in the data and not by the personal judgement of the Statistician.

(v) This method is quite flexible in the sense that when a few more observations are added to the given data, the trend values already obtained will not be affected, only some more trend values will be included in the series.

Limitations (or Disadvantages)

(i) Some trend values at the beginning and at the end of the series cannot be determined.

(ii) It is not easy to determine the period of moving average when the oscillatory movement does not exhibit any regular periodic cycle.

(iii) This method cannot be used to forecast future trend values as the moving averages do not obey any law.

(iv) This method is used to find only linear trend. Non-linear trend values obtained by this method are biased and deviate from the actual trend values.

(v) This method may generate cycles or other movements which were not present in the original data.

Example 2:

(i) Obtain the five-year moving averages for the following series of observations:

Year	1967	1968	1969	1970	1971	1972	1973	1974
Annual Sales (Rs. '000)	3.6	4.3	4.3	3.4	4.4	5.4	3.4	2.4

(ii) Construct also the 4 year centered moving average.

Solution:

(i)

Table 1: Calculations of 5-Year Moving Averages

Year (1)	**Annual Sales (Rs. '000) (2)**	**5- year moving total (3)**	**5- year moving average (Rs. '000) (4)**
1967	3.6	-	-
1968	4.3	-	-
1969	4.3	20.0	4.00
1970	3.4	21.8	4.36
1971	4.4	20.9	4.18
1972	5.4	19.0	3.80
1973	3.4	-	-
1974	2.4	-	-

Note that the first moving total 20.0 of column 3 is the sum of the first 5 values 3.6, 4.3, 4.3, 3.4, 4.4. The second moving total is 4.3 + 4.3 + 3.4 + 4.4 + 5.4 = 21.8 which can also be easily obtained by adding (5.4 – 3.6), i.e., 1.8 with the first moving total. Similarly, the 3rd moving total is 21.8 + (3.4 – 4.3) = 20.9 and so on.

Note: The five year moving averages (or trend values) for the years 1969-1972 are shown in coumn 4. (Note that the moving averages correspond to the given years.) For the other years 1967, 1968 and 1973, 1974, moving averages cannot be determined.

(ii) First Method

Table 2: Calculations of 4-Year Centered Moving Averages

Year (1)	Annual Sales (Rs. '000) (2)	4-year moving total (3)	2-point moving total of Col 3 (Centered) (4)	4-year centered moving average (Rs. '000) (Col 4 ÷ 8)
1967	3.6	…..	…...	…..
1968	4.3	…..	…..	…..
		15.6		
1969	4.3		32.0	4.00
		16.4		
1970	3.4		33.9	4.24
		17.5		
1971	4.4		34.1	4.26
		16.6		
1972	5.4		32.2	4.03
		15.6		
1973	3.4	….	…..	…..
1974	2.4	…..	…..	…..

In the above table 4-year moving totals are shown against the mid-points of the time intervals in Col.3. As the moving totals do not correspond to the given years, 2-point moving total of Col.3 are found in Col.4, for centering them (i.e., for synchronizing them with the original data).

(iii) Second Method

Table 3: Calculations for 4-Year Centered Moving Averages.

Year (1)	Data(Annual Sales Rs. '000) (2)	4-year moving total (3)	4-year moving average (4)	2-year moving total of col. 4 (centered)	4-year centered moving Average (Col. 5 ÷ 2)
1967	3.6	…..	…..	…..	….
1968	4.3	…..	…..	….	….
		15.6	3.9		
1969	4.3			8.0	4.0
		16.4	4.1		
1970	3.4			8.5	4.2
		17.5	4.4		
1971	4.4			8.6	4.3
		16.6	4.2		
1972	5.4			8.1	4.0
		15.6	3.9		
1973	3.4	….	….	….	…..
1974	2.4	…..	….	…..	…..

Example 3: Find the trend for the following series using a three-year weighted moving average with weight 1, 2; 1:

Year :	1	2	3	4	5	6	7
Values :	2	4	5	7	8	10	13

Solution:

Table 4: Calculations OF 3-Year Weighted Moving Average

Year (1)	Values (2)	3-year weighted moving total (3)	3-year weighted moving average (Col. 3 ÷ 4) (4)
1	2	…	…..
2	4	2x1+4x2+5x1 = 15	3.75
3	5	4x1+5x2+7x1 = 21	5.25
4	7	5x1+7x2+8x1 = 27	6.75
5	8	7x1 +8x2 +10x1 = 33	8.25
6	10	8x1+10x2+13x1 = 41	10.25
7	13	….	…..

Col. 4 = Col. 3 ÷ total weight, where total weight = 1 + 2 + 1 = 4.

Example 4: For the following series of observations, verify that the 4-year centered moving average is equivalent to a 5-year weighted moving average with weight 1, 2, 2, 2, 1 respectively:

Year	1	2	3	4	5	6	7	8	9	10	11
Sales (Rs.'000)	2	6	1	5	3	7	2	6	4	8	3

Solution:

Table 5: Calculations of 4-Year Centered Moving Average

Year (1)	**Annual Sales (Rs.'000) (2)**	**4-year moving total (3)**	**2-year moving total of Col. 3 (4)**	**4-year centered moving average (Col. 4 ÷ 8)**
1	2	…..	…..	…..
2	6	…..	……	…...
		14		
3	1		29	3.625
		15		
4	5		31	3.875
		16		
5	3		33	4.125
		17		
6	7		35	4.375
		18		
7	2		37	4.625
		19		
8	6		39	4.875
		20		
9	4		41	5.125
		21		
10	8	…..	…..	…..
11	3	…...	…..	……

Table 6: Calculation of 5-Year Weighted Moving Average

Year (1)	Annual Sales (Rs.'00000) (2)	5-year weighted moving total (3)	5-year weighted moving average (Col. 3 ÷ total wt. 8)
1964	2	…..	…..
1965	6	…..	…..
1966	1	$2 \times 1 + 6 \times 2 + 1 \times 2 + 5 \times 2 + 3 \times 1 = 29$	3.625
1967	5	$6 \times 1 + 1 \times 2 + 5 \times 2 + 3 \times 2 + 7 \times 1 = 31$	3.875
1968	3	$1 \times 1 + 5 \times 2 + 3 \times 2 + 7 \times 2 + 2 \times 1 = 33$	4.125
1969	7	$5 \times 1 + 3 \times 2 + 7 \times 2 + 2 \times 2 + 6 \times 1 = 35$	4.375
1970	2	$3 \times 1 + 7 \times 2 + 2 \times 2 + 6 \times 2 + 4 \times 1 = 37$	4.625
1971	6	$7 \times 1 + 2 \times 2 + 6 \times 2 + 4 \times 2 + 8 \times 1 = 39$	4.875
1972	4	$2 \times 1 + 6 \times 2 + 4 \times 2 + 8 \times 2 + 3 \times 1 = 41$	5.125
1973	8	…..	…..
1974	3	…..	…..

From the last columns of the two Tables 5 and 6, we see that the 4-year centered moving average is equivalent to a 5-year weighted moving average with weight 1, 2, 2, 2, 1 respectively.

(iv) Methods of Least Squares

This method is widely used for the measurement of trend.

Linear Trend

Let (X1, Y1), (X2, Y2), …. , (XN, YN) be N pairs of observations where Yi represents time series and Xi represents time. Suppose the equation of the straight line to be fitted to the time series data by the Method of Least Squares is

$$Y = a + bX \qquad (1)$$

For a given value of X, say X1, the corresponding value of Y obtained from (1) is a + b X1. The difference E1 = Y1 – (a + b X1) or Y1 – a – b X1, which may be positive, negative or zero, is called an error or residual.

Similarly we obtain

$E_2 = Y_2 - a - bX_2, \ldots, E_N = Y_N - a - bX_N.$

By the Principle of Least Squares, the line of the best fit is obtained when the sum of the squares of the differences E_i between the observed values Y_i and the corresponding calculated values a + b X_i, is minimum, i.e. when

$$\sum_{i=1}^{N} E^2{}_i = \sum_{i=1}^{N} (Y_i - a - b X_i)^2$$

Is minimum.

When

$\sum_{i=1}^{N} E^2{}_i$ is minimum, we obtain the normal equations

$$\sum Y = N a + b \sum X \qquad (2)$$

$$\text{And } \sum XY = a \sum X + b \sum X^2 \qquad (3)$$

Solving these two equations, a and b can be determined, and substituting these values of a and b in (1), the required equation of the straight line trend is obtained. From this equation, we can compute the trend values.

If we take the mid-point in time as the origin, the negative values in the first half of the series balance out the positive values in the second half so that $\sum X = 0$. The normal equations (2) and (3) would reduce to

$\sum Y = N a$ and $\sum XY = B\sum X^2$;

$$\therefore a = \frac{\Sigma Y}{N} \text{ and } b = \frac{\Sigma XY}{\Sigma X^2}$$

Example 5: Determined the equation of a straight line which best fits the following data:

Year :	1974	1975	1976	1977	1978
Sales (in Rs. '000) :	35	56	79	80	40

Compute the trend values for all the years from 1974 to 1978.

Solution: Let the equation of the straight-line best of fit, with the origin at the middle year 1976 and unit of X as 1 year, be

$Y = a + bX$ (1)

By the Method of Least squares, the values of a and b are given by

$a = \sum Y / N$ and $b = \sum XY / \sum X^2$ (2)

Hence N = number of years = 5.

Table 7: Calculations for the Line of the Best Fit

Year	**Sales (in Rs. '000)**	**X**	**X^2**	**XY**
1974	35	-2	4	-70
1975	56	-1	1	-56
1976	79	0	0	0
1977	80	1	1	80
1978	40	2	4	80
Total	$290 = \sum Y$	0	$10 = \sum X^2$	$34 = \sum XY$

Using (2), $a = \sum Y / N = 290 / 5 = 58$, and $b = \sum XY / \sum X^2 = 34/10 = 3.4$.

From (1), the required equation of the best fitted straight line is $Y = 58 + 3.4 X$.

Year	**X**	**Trend Values (Y = 58 + 3.4 X)**
1974	-2	$58 + 3.4 \times -2 = 51.2$
1975	-1	$58 + 3.4 \times -1 = 54.6$
1976	0	$58 + 3.4 \times 0 = 58.0$
1977	1	$58 + 3.4 \times 1 = 61.4$
1978	2	$58 + 3.4 \times 2 = 64.8$

NOTE. Unless otherwise specified, we shall assume that the values of Y refer to mid-year values, i.e. as a July, 1. Thus in Example 6, X = 0 corresponds to July, 1 1976, X = -1 to July 1, 1975, X = 1 to July 1, 1977, etc.

12.3 QUESTIONS

1. Explain various forecasting techniques.
2. Explain time series analysis.
3. What are the components of time series analysis method.
4. What is moving average method? Explain it's advantages and disadvantages?