Research Methodology

(MBA 3rd Sem/M.Com 4th Sem/ All P.G Diploma 1st Sem)

UTKAL UNIVERSITY

NAAC ACCREDITED A+
DDCE
Education for all

DIRECTORATE OF DISTANCE AND CONTINUING EDUCATION
Utkal University, Bhubaneswar - 7, Odisha
E-mail: helpline@ddceutkal.ac.in
Website: www.ddceutakal.ac.in

Research Methodology

ISBN : 978-93-5299-542-4

Dr. S. K. Podder
Haritha M.

First Edition : November 2019

No. of Copies : 1001

Printed and Published by:
Mrs. Meena Pandey
Himalaya Publishing House Pvt. Ltd.,
"Ramdoot", Dr. Bhalerao Marg, Girgaon, Mumbai - 400 004.
Phone: 022-23860170 & 23863863, Fax: 022-23877178
Email: himpub@bharatmail.co.in; **Website:** www.himpub.com

This SIM has been prepared exclusively for Directorate of Distance and Continuing Education (DDCE), Utkal University. It conforms to the syllabi and contents as approved by Utkal University.

DIRECTORATE OF DISTANCE AND CONTINUING EDUCATION
UTKAL UNIVERSITY : VANI VIHAR
BHUBANESWAR – 751 007

From the Director's Desk

The Directorate of Distance and Continuing Education, originally established as the University Evening College way back in 1962, has travelled a long way in the last 56 years. **'EDUCATION FOR ALL'** is our motto. Increasingly, the Open and Distance Learning institutions are aspiring to provide education for anyone, anytime and anywhere. DDCE, Utkal University has been constantly striving to rise up to the challenges of Open Distance Learning system. Nearly one lakh students have passed through the portals of this great temple of learning. We may not have numerous great tales of outstanding academic achievements but we have great tales of success in life, of recovering lost opportunities, tremendous satisfaction in life, turning points in career and those who feel that without us they would not be where they are today. There are also flashes when our students figure in best ten in their honours subjects. Our students must be free from despair and negative attitude. They must be enthusiastic, full of energy and confident of their future. To meet the needs of quality enhancement and to address the quality concerns of our stakeholders over the years, we are switching over to self-instructional material printed courseware. We are sure that students would go beyond the courseware provided by us. We are aware that most of you are working and have also family responsibility. Please remember that only a busy person has time for everything and a lazy person has none. We are sure that you will be able to chalk out a well planned programme to study the courseware. By choosing to pursue a course in distance mode, you have made a commitment for self-improvement and acquiring higher educational qualification. You should rise up to your commitment. Every student must go beyond the standard books and self-instructional course material. You should read number of books and use ICT learning resources like the internet, television and radio programmes, etc. As only limited number of classes will be held, a student should come to the personal contact programme well prepared. The PCP should be used for clarification of doubt and counseling. This can only happen if you read the course material before PCP. You can always mail your feedback on the courseware to us. It is very important that one should discuss the contents of the course materials with other fellow learners.

We wish you happy reading.

DIRECTOR

SYLLABUS

Unit – I

Research: Meaning, Objectives and Importance of Research, Role of Research in Functional Areas: Finance, Marketing, HRD, Research Methodology, Process of Research.

Unit – II

Defining Research Problem: Process of Formulating Hypothesis, Research Design, Sampling Design.

Unit – III

Collecting, Processing and Analysis of Data: Design of Questionnaire, Testing of Hypothesis, Parametric and Non-parametric Tests, T-test, Z-test, Chi-square test.

Unit – IV

Multivariate Analysis Techniques: Multiple Regression Analysis, Discriminated Analysis, Factor Analysis, ANOVA.

Unit – V

Interpretation and Report Writing: Importance and Techniques of Interpretation, Significance of Report Writing, Steps in Writing Report, Layout of the Research Report, Types of Report.

CONTENTS

Chapter 1

RESEARCH

Objectives

The objectives of this lesson are to:

- Objectives of Research
- Importance of Research
- Scope of Research
- Role of Research in Functional Areas: Finance, Marketing, HRD
- Classification of Research
- Research Methodology
- Process of Research

Structure:

Notes

1.1 INTRODUCTION

Research is a way to systematically solve the research problem. It may be understood as a science of studying how research is done scientifically. In it we study the various steps that are generally adopted by a researcher in studying his research problem along with the logic behind them. It is necessary for the researcher to know not only the research methods but also the methodology. Researchers not only need to know how to develop certain indices or tests, how to calculate the mean, the mode, the median or the standard deviation or chi-square, how to apply particular research techniques, but they also need to know which of these methods or techniques, are relevant and which are not, and what would they mean and indicate and why. Researchers also need to understand the assumptions underlying various techniques and they need to know the criteria by which they can decide that certain techniques and procedures will be applicable to certain problems and others will not. All this means that it is necessary for the researcher to design his methodology for his problem as the same may differ from problem to problem. For example, an architect, who designs a building, has to consciously evaluate the basis of his decisions, i.e., he has to evaluate why and on what basis he selects particular size, number and location of doors, windows and ventilators, uses particular materials and not others and the like.

Similarly, in research the scientist has to expose the research decisions to evaluation before they are implemented. He has to specify very clearly and precisely what decisions he selects and why he selects them so that they can be evaluated by others also. From what has been stated above, we can say that research methodology has many dimensions and research methods do constitute a part of the research methodology.

The scope of research methodology is wider than that of research methods. Thus, when we talk of research methodology we not only talk of the research methods but also consider the logic behind the methods we use in the context of our research study and explain why we are using a particular method or technique and why we are not using others so that research results are capable of being evaluated either by the researcher himself or by others. Why a research study has been undertaken, how the research problem has been defined, in what way and why the hypothesis has been formulated, what data have been collected and what particular method has been adopted, why particular technique of analyzing data has been used and a host of similar other questions are usually answered when we talk of research methodology concerning a research problem or study.

1.2 MEANING AND DEFINITIONS OF RESEARCH

Meaning of Research

Research is a systematic investigative process employed to increase or revise current knowledge by discovering new facts.

Research refers to the search for knowledge or as any systematic investigation, with an open mind, to establish novel facts, solve new or existing problems, prove new ideas, or develop new theories. The primary purposes of basic research as opposed to applied research are documentation, discovery, interpretation or the research and development of methods and systems for the advancement of human knowledge. Approaches to research depend on epistemologies, which vary considerably both within and between humanities and sciences.

Definitions of Research

According to ***Kerlinger,*** research can be defined as a "systematic, controlled, empirical and critical investigation of hypothetical propositions about the presumed relations among natural phenomena".

Notes

According to ***Emory,*** research can be defined as “any organized activity designed and carried out to provide information for solving a problem”.

According to ***Martin Shuttle Worth,*** "Research includes any gathering of data, information and facts for the advancement of knowledge."

According to ***Creswell,*** "Research is a process of steps used to collect and analyze information to increase our understanding of a topic or issue".

The ***Merriam-Webster Online Dictionary*** defines research in more detail as "a studious inquiry or examination; especially: investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or laws in the light of new facts, or practical application of such new or revised theories or laws".

According to ***Clifford Woody,*** research can be defined as “defining and redefining problems, formulating hypothesis or suggested solutions, collecting, organizing and evaluating data, making deductions, reaching conclusions and testing the conclusions to determine whether they fit the formulating hypothesis”.

According to ***Black and Champion,*** research can be defined as “obtaining information through empirical observations that can be used for the systematic development of logically related propositions attempting to establish casual relations among variables”.

According to ***Young,*** research can be defined as “a scientific undertaking which, by means of logical and systematic techniques aims to: (i) discover new facts or verify and test old facts, (ii) analyse their sequences, interrelationships and causal explanations. (iii) develop new scientific tools, concepts and theories which would facilitate reliable and valid study of human behaviour”.

The Encyclopedia of Social Sciences defines research as “the manipulation of generalizing to extend, correct or verify knowledge”.

The Merriam-Webster Online Dictionary defines research in more detail as "a studious inquiry or examination; especially: investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or laws in the light of new facts, or practical application of such new or revised theories or laws".

Business Research can be defined as the systematic and objective process of gathering, recording and analyzing data for aid in making business decisions.

Business Research can be defined as a systematic inquiry that provides information to guide business decisions.

1.3 OBJECTIVES OF RESEARCH

The purpose of research is to discover answers 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, may think of research objectives as falling into number of broad grouping:

i) To gain familiarity with a phenomenon or to achieve new insights into it (studies with this object in view are termed as exploratory or formularize research studies.

ii) 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);

iii) 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).

Notes

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

1.4 IMPORTANCE OF RESEARCH

a) Government Policies: Research provides the basis for nearly all government policies in our economic system. *For example,* government's budgets rest in part on an analysis of the needs and desires of the people and on the availability of revenues to meet those needs. The cost of needs has to be equated to probable revenues and this is a field where research is most needed. Through research we can devise alternative policies and can well examine the consequences of each of these alternatives.

b) Allocation of National Resources: Government has to chalk out programmes for dealing with all facets of the country's existence and most of these will be related directly or indirectly to economic conditions. The plight of cultivators, the problems of big and small business and industry, working conditions, trade union activities, the problem of distribution, even the size and nature of defense services are matters requiring research. Thus research is considered necessary with regard to the allocation of nation's resources.

c) Investigation of Economic Structure: Research is necessary for collection of information on the economic and social structure of the nation. Such information indicates what is happening in the economy, and what changes are taking place. Collecting such statistical information involves a variety of research problems.

d) Social Welfare and Progress: Ignorance and lack of knowledge is the root cause of various social mishaps. Communal troubles, religious riots, the misnomer of social, racial superiority are results of ignorance. Through research it is possible to drive away with all these wrong notions. Research is helpful in the welfare and progress of humanity and the society.

1.5 CHARACTERISTICS OF A GOOD RESEARCH

Good research generates dependable data that are derived by professionally conducted practices and that can be used reliably for decision-making. In contrast, poor research is carelessly planned and conducted, resulting in data that a manager can't use to reduce his or her decision-making risks. Good research follows the standards of the scientific method: systematic, empirically based procedures for generating replicable research.

Following are the characteristics of a good research for managerial decisions:

1. Purpose clearly defined

The purpose of the business research, the problem involved or the decision to be made should be clearly defined and sharply delineated in terms as unambiguous as possible. Getting this in writing is valuable even in instances where the same person serves as researcher and decision-maker. The statement of the decision problem should include its scope, its limitations, and the precise meanings of all words and terms significant to the research. Failure of the researcher to do this adequately may raise legitimate doubts in the minds of research report readers as to whether the researcher has sufficient understanding of the problem to make a sound proposal attacking it. This characteristic is comparable to developing a strategic plan for achieving an objective before developing a tactical plan or an action map.

Notes

2. Research process detailed

The research procedures used should be described in sufficient detail to permit another researcher to repeat the research. Except when secrecy is imposed, research reports should reveal with candor the sources of data and the means by which they were obtained. Omission of significant procedural details makes it difficult or impossible to estimate the validity and reliability of the data and justifiably weakens the confidence of the reader in the research itself as well as any recommendations based on the research. This characteristic is comparable to developing a tactical plan.

3. Research design thoroughly planned

The procedural design of the research should be carefully planned to yield results that are as objective as possible. When a sampling of the population is involved, the report should include evidence concerning the degree of representativeness of the sample. A survey of opinions or recollections ought not to be used when more reliable evidence is available from documentary sources or by direct observation. Bibliographic searches should be as thorough and complete as possible. Experiments should have satisfactory controls. Direct observations should be recorded in writing as soon as possible after the event. Efforts should be made to minimize the influence of personal bias in selecting and recording data. This characteristic is comparable to developing detailed action plans for each tactic.

4. High ethical standards applied

Researchers often work independently and have significant latitude in designing and executing research projects. A research design that includes safeguards against causing mental or physical harm to participants and makes data integrity a first priority should be highly valued. Ethical issues in research reflect important moral concerns about the practice of responsible behavior in society. Researchers frequently find themselves precariously balancing the rights of their subjects against the scientific dictates of their chosen method. When this occurs, they have a responsibility to guard the welfare of the participants in the studies and also the organizations to which they belong, their clients, their colleagues, and themselves. Careful consideration must be given to those research situations in which there is a possibility of physical or psychological harm, exploitation, invasion of privacy, and/or loss of dignity. The research need must be weighed against the potential for adverse effects. Typically, you can redesign a study, but sometimes you cannot. The researcher should be prepared for this dilemma.

5. Limitations frankly revealed

The researcher should report, with complete frankness, flaws in procedural design and estimate their effect on the findings. There are very few perfect research designs. Some of the imperfections may have little effect on the validity and reliability of the data; others may invalidate them entirely. A competent researcher should be sensitive to the effects of imperfect design. The researcher's experience in analyzing data should provide a basis for estimating the influence of design flaws. As a decision-maker, you should question the value of research where no limitations are reported.

6. Analysis adequate for decision maker's needs

Analysis of the data should be extensive enough to reveal its significance, what managers call "insights." The methods of analysis used should be appropriate. The extent to which this criterion is met is frequently a good measure of the competence of the researcher. Adequate analysis of the data is the most difficult phase of research for the novice. The validity and reliability of data should be checked carefully. The data should be classified in ways that assist the researcher in reaching pertinent conclusions and clearly reveal the findings that have led to those conclusions. When statistical methods are used, the probability of error should be estimated and the criteria of statistical significance applied.

Notes

7. Findings presented unambiguously

Some evidence of the competence and integrity of the researcher may be found in the report itself. *For example,* language that is restrained, clear and precise; assertions that are carefully drawn and hedged with appropriate reservations; and an apparent effort to achieve maximum objectivity tend to leave a favorable impression of the researcher with the decision-maker. Generalizations that outrun the evidence on which they are based, exaggerations, and unnecessary verbiage tend to leave an unfavorable impression. Such reports are not valuable to managers wading through the minefields of organizational decision-making. Presentation of data should be comprehensive, easily understood by the decision-maker and organized so that the decision-maker can readily locate critical findings.

8. Conclusions justified

Conclusions should be limited to those for which the data provide an adequate basis. Researchers are often tempted to broaden the basis of induction by including personal experiences and their interpretations data not subject to the controls under which the research data were gathered. Equally undesirable is the all-too-frequent practice of drawing conclusions from a study of a limited population and applying them universally. Researchers also may be tempted to rely too heavily on data collected in a prior study and use it in the interpretation of a new study. Such practice sometimes occurs among research specialists who confine their work to clients in a small industry. These actions tend to decrease the objectivity of the research and weaken readers' confidence in the findings. Good researchers always specify the conditions under which their conclusions seem to be valid.

9. Researcher's experience reflected

Greater confidence in the research is warranted if the researcher is experienced, has a good reputation in research, and is a person of integrity. Were it possible for the reader of a research report to obtain sufficient information about the researcher, this criterion perhaps would be one of the best bases for judging the degree of confidence a piece of research warrants and the value of any decision based upon it. For this reason the research report should contain information about the qualifications of the researcher.

Good business research has an inherent value only to the extent that it helps management make better decisions that help achieve organizational goals. Interesting information about consumers, employees, competitors or the environment may be pleasant to have, but its value is limited if the information cannot be applied to a critical decision. If a study does not help management select more effective, more efficient, less risky or more profitable alternatives than otherwise would be the case, its use should be questioned. Alternatively, management may have insufficient resources (time, money or skill) to conduct an appropriate study or may face a low level of risk associated with the decision at hand. In these situations, it is valid to avoid business research and its associated costs in time and money. Business research finds its justification in the contribution it makes to the decision-makers task and to the bottom line.

1.6 SCOPE OF RESEARCH

The scope of research can be summarized as follows:

i) Business Competition

Competition in business is increasing day by day. You need to know about customers, products and industry competition.

Notes

a) ***Customers:*** You need to know about your customers, their needs, their perceptions and future requirements. Research helps you to find out the variables and factors which are significant for increasing customer loyalty and adding new customers.

b) ***Products:*** Research helps you to know consumer need and this in turn is used to develop new product. To decide about pricing, positioning, packaging, branding, sales promotion and other promotional techniques, we need to carry out business research.

c) ***Industry competition:*** You need to know what other companies are doing to increase their market share, factors responsible for increase and decrease of market share, and trends in industry growth.

2. Business Environment

Business environment is the totality of all those factors which affect the business but are not under the control of managers. Economic and non- economic elements of environment include economic system (ownership rights like in capitalism and socialism), economic anatomy (structure of households whether manufacturing, trading or agriculture society), Government legislations, Government policies, movement of policies, velocity of policies, fiscal and monetary policies, ideology of ruling party, social ideology, social values and systems, social structures, etc. Changing environment affect the business. You need to know the trend in environment and factors responsible for change in environment.

3. Maturing of management as a group of disciplines

The quality of theories and models to explain tactical and strategic results in human resources, marketing, operations and finance is improving, providing managers with more knowledge. In turn managers are expected to use these models to specific field they are attached to; business research can help managers to understand these models and their use in specific situation.

4. Explosive growth and influence of the Internet

The explosive growth of company websites, e-commerce and electronic publications brings extensive amounts of new information, but this information does not help us to make decisions. (We shall read in the coming chapters that information is not knowledge). Information need to be processed to arrive at knowledge. This knowledge can be helpful to have competitive advantage.

5. Stakeholders demanding greater influence

Customers, workers, shareholders and the general public demand to be included in company decision-making; armed with extensive information, they are more sensitive to their own self-interests than ever before and more resistant to an organization's stimuli.

6. More global competition

Competition, both global and domestic, is growing and often coming from unexpected sources; many organizations re-focus on primary competencies, while they seek to improve operations by reducing costs and converting customers to advocates.

7. More government intervention

Governments continue to show concern with all aspects of society, becoming increasingly aggressive in protecting various segments of society with various policies. This throws challenges to managers to be alert to various factors which are not under their control. The decisions under such circumstances can be made after the use of managerial and business research tools.

Notes

8. More complex decisions

Managers have more variables to consider in every decision, increasing their need for more and better knowledge and greater insights from that information.

9. Lower-cost data collection

Computers and telecommunications lowered the costs of data collection. Everybody has access to data. Research helps to convert information into knowledge, hence knowledge is easily available, This drastically changes knowledge about consumers at both store and household levels; employees at the position, team and department levels; suppliers and distributors at the transaction division, and company levels; and machines at the part process, and production-run levels resulting into increased competition. Due to heavy competition among companies and also due to easy availability of data, continuous research process is required.

10. Better visualization tools

With the use of technology, it is possible to download at high speed. High-speed downloads of images allow researchers to visualize complex concepts, which enrich measurement capabilities and in turn help consumers to arrive at better decisions of purchasing the products. Researchers can develop new techniques to create images which can be more competitive.

11. Powerful computations

Sophisticated techniques of quantitative analysis are emerging to take advantage of increasingly powerful computing capabilities. Computer advances permit businesses to create and manage a data warehouse, an electronic storehouse where vast arrays of collected, integrated data are ready for mining. The power and ease of use of computers offer us the capability to analyze more data more quickly to deal with complex managerial problems. Yet, the quantity of collected raw data overwhelms users, necessitating a means to manage it. Early efforts to provide a flow of information to managers used a Management Information System (MIS). Now the challenge of database management from an MIS perspective has increased which include removing obstacles like resistance to use, keenness of managers to disclose fully their information needs and decision criteria, decreasing costs of single-user report generation, system design time, increasing adaptation to changing organization structures, and decision relevance (standard versus tailored reports). While routine MIS reports are useful for well-structured problems and those amenable to a standardized set of procedures, data must be more than timely and standardized; reporting must be customized to be truly meaningful to the user.

12. Advanced analytical tools for enhanced insights

Organizations increasingly practice data mining, applying mathematical models to extract meaningful knowledge from volumes of data contained within internal databases. Enormous quantities of research data are reduced to relatively straightforward equations with statistical models. Expert systems and outgrowth of artificial intelligence, and data mining entered the 21st century as important tools for research. Advanced analytical tools are available to answer a variety of research questions. Traditional topics open to modeling market share, price elasticity, the cannibalization of one product's sales by the introduction of another product, the effects on productivity of changing an employee compensation system, to name a few create decision support models that reflect the behaviour of individuals, households, and industries. Programmes that combine modeling and decision support systems evolved in the latter part of the 20th century to provide the most utility to users.

13. New perspectives on established research methodologies

Older tools and methodologies once limited to exploratory research are gaining wider acceptance in dealing with a broader range of managerial problems.

Notes

1.7 PURPOSE OF RESEARCH

The Purpose of Research can be summarized by considering various types of research and their applications:

1. Purposes of Basic Research

Basic research is the research which is done for knowledge enhancement, the research which does not have immediate commercial potential. The research is done for human welfare, animal welfare and plant kingdom welfare. It is called basic, pure, fundamental research. The main motivation is to expand man's knowledge, not to create or invent something. There is no obvious commercial value to the discoveries that result from basic research. Basic research lay down the foundation for the applied research. Dr. G. Smoot says "people cannot foresee the future well enough to predict what is going to develop from the basic research".

2. Purposes of Applied Research

Applied research is designed to solve practical problem of the modern world, rather than to acquire knowledge for knowledge sake. The goal of applied research is to improve the human condition. It focuses on analysis and solving social and real life problems. This research is generally conducted on large scale basis, it is expensive. As such, it often conducted with the support of some funding agency like government, public corporation, World Bank, UNICEF, UGC etc. According to Hunt, "applied research is an investigation for ways of using scientific knowledge to solve practical problems" for example: improve agriculture crop production, treat or cure a specific disease, improve the energy efficiency homes, offices, how can communication among workers in large companies be improved? Applied research can be further classified as problem oriented and problem solving research. Problem oriented research:- research is done by industry apex body for sorting out problems faced by all the companies. WTO does problem oriented research for developing countries, in India Agriculture and Processed Food Export Development Authority (APEDA) conduct regular research for the benefit of agri-industry. Problem solving:-this type of research is done by an individual company for the problem faced by it. Marketing research and market research are the applied research. *For example:* Videocon international conducts research to study customer satisfaction level. In short, the main aim of applied research is to discover some solution for some pressing practical problem.

3. Purposes of Quantitative Research

Quantitative research aims to measure the quantity or amount and compares it with past records and tries to project for future period. In social sciences, "quantitative research refers to the systematic empirical investigation of quantitative properties and phenomena and their relationships". The objective of quantitative research is to develop and employ mathematical models, theories or hypothesis pertaining to phenomena. The process of measurement is central to quantitative research because it provides fundamental connection between empirical observation and mathematical expression of quantitative relationships. Statistics is the most widely used branch of mathematics in quantitative research. Statistical methods are used extensively with in fields such as economics and commerce. Quantitative research involving the use of structured questions, where the response options have been pre-determined and large number of respondents is involved. *Example:* Total sales of soap industry in terms of rupees cores and or quantity in terms of lakhs tones for particular year, say 2008, could be researched, compared with past 5 years and then projection for 2009 could be made.

4. Purposes of Qualitative Research

Qualitative research presents non-quantitative type of analysis. Qualitative research is collecting, analyzing and interpreting data by observing what people do and say. Qualitative research refers to the

Notes

meanings, definitions, characteristics, symbols, metaphors, and description of things. Qualitative research is much more subjective and uses very different methods of collecting information, mainly individual, in-depth interviews and focus groups. The nature of this type of research is exploratory and open ended. Small number of people is interviewed in depth and or a relatively small number of focus groups are conducted. Qualitative research can be further classified in the following type:

i) ***Phenomenology:*** This is a form of research in which the researcher attempts to understand how one or more individuals experience a phenomenon. *Example:* We might interview 20 victims of Bhopal tragedy. The study has roots in philosophical perspectives.

ii) ***Ethnography:*** This type of research focuses on describing the culture of a group of people. A culture is the shared attributes, values, norms, practices, language and material things of a group of people. *Example:* The researcher might decide to go and live with the tribal in Andaman island and study the culture and the educational practices.

iii) ***Case Study:*** This is a form of qualitative research that is focused on providing a detailed account of one or more cases. *Example:* This may study a classroom that was given a new curriculum for technology use.

iv) ***Grounded Theory:*** Grounded theory generates or discover a theory an abstract, analytical scheme of phenomenon. This is an inductive type of research, based or grounded in the observations of data from which it was developed; it uses a variety of data sources, including quantitative data, review of records, interviews, observation and surveys.

v) ***Historical Research:*** Research on past social forces which have shaped the present is historical research. It allows one to discuss past and present events in the context of the present condition and allows one to reflect and provide possible answers to current issues and problems. *Example:* The lending pattern of business in the 19th century.

v) ***Ex-post facto Research:*** Ex-post facto research is an emperical enquiry for situations that have already occured. *For example:* Market failure for any company's product if studied or researched later may be categorized as post facto research.

1.8 RELEVANCE OF RESEARCH

Relevance of Research can be summarized as follows:

1. Research is a Tool for Building Knowledge and Efficient Learning

Research is required not just for students and academics, but for all professionals. It is also important for budding and veteran writers, both offline and online. Among professionals and scribes, finding an interesting topic to discuss and/or to write about should go beyond personal experience. Determining either what the general public may want to know about or what researchers want others to realize or to think about can serve as a reason to do research. Undoubtedly, it is crucial to finding possible cures for diseases, as well as how to prevent them. Thus, research becomes a must to ascertain if one's ideas are supported by previous studies or if these ideas still need proof to be considered as knowledge.

Example: An example of such endeavor is the 2016 study of several psychologists who examined how sleep affects memory reactivation. In "Relearn Faster and Retain Longer: Along With Practice, Sleep Makes Perfect", they "found that interleaving sleep between learning sessions not only reduced the amount of practice needed by half but also ensured much better long-term retention. Sleeping after learning is definitely a good strategy, but sleeping between two learning sessions is a better strategy." This study supports the fact that: "Both repeated practice and sleep improve long-term retention of information". Their findings also emphasize how highly important sleep is to healthy brain function.

Notes

Research by The World Bank in 2006 also underscored sleep as a key factor of efficient learning or the process of gaining optimal learning using few resources. The study reiterated the role of sleep in: (i) protecting and restoring memory, (ii) advanced learning and (iii) enhancing mathematical ability and problem solving. It further noted that "knowledge is better consolidated when people study at the time when they are supposed to be awake rather than, say, late-night sessions." It cited the need for research on "the memory capacity of the poor in low-income countries" to enable teachers in helping underprivileged students learn basic skills.

Said studies on the effects of sleep on the human brain are among the many topics that have already been examined by academics and specialists in various universities and medical institutions. A myriad of research ideas likewise awaits the attention of avid scholars and inquisitive writers. Indeed, research is instrumental in building and improving knowledge, as well as in facilitating learning.

2. Research Means to Understand Various Issues

Television shows and movies secretion with research both on the part of the writer(s) and the actors encourage to understand various issues. Though there are hosts who rely on their researchers, there are also those who exert effort to do their own research. This helps them get information that hired researchers missed, build a good rapport with the interviewee, and conduct a good interview in the process.

For their part, some film and TV actors would take time to interview detectives, boxers, scientists, business owners, criminals and teachers, among others. Others even go through immersion to make them understand the issues of their respective characters better, such as living in jail or in a drug rehabilitation center. Many would read literature, biographies, or journals to have a better view or context of the story.

As what Terry Freedman says in "The Importance of Research for ICT Teachers" (2011): "Research can shed light on issues we didn't even know existed, and can raise questions we hadn't realized even needed asking." Thus, almost all writers of imaginary and non-fictive tales also do research, for doing so helps them create a good story and/or achieve strong credibility as an academic.

3. Research is an Aid to Business Success

Research benefits business. Many successful companies, such as those producing consumer goods or mass-market items, invest in research and development or R&D. Different business industries with science and engineering processes like agriculture, food and beverage, manufacturing, healthcare and pharmaceuticals, computer software, semiconductor, information and communication technology, construction, robotics, aerospace, aviation and energy have high R&D expenditure because it is critical to product innovation and to improving services.

R&D also helps secure a vantage point over competitors. Finding out how to make things happen and what could differentiate them from others that offer similar products and services can raise the company's market value. Certainly, having relevant knowledge in achieving a good commercial image through sound business strategies like investing in R&D can boost its profitability.

4. Research is a Way to Prove Lies and to Support Truths

Several experienced feeling that your mate is having an affair behind your back. Some people would overlook that and say that it's better not to know; others though would take discreet action, hiring detectives to do the work. Doing research to reveal lies or truths involving personal affairs contributes in either making a relationship work or in breaking away from a dysfunctional one. For the monogamous lot, doing research to disprove or prove infidelity is not simply a trust issue, but a right to find out the truth - unless one's intimate partner has already admitted being polyandrous even before the relationship started. When a person dislikes answering relationship-related questions, including her or his whereabouts,

Notes

it is better to see that as a red flag and take baby steps to save yourself from what could become a more serious emotional mess later.

Scientists also deal with research to test the validity and reliability of their claims or those of other scientists'. Their integrity and competence depend on the quality - and not just quantity - of their research. Nonetheless, not everything scientists come up with get accepted or learned by everyone, especially when factors like religion, state suppression, and access to resources and social services (e.g., education and adequate health programs) either feed the poor majority with lies or deter them from knowing truths to preserve the status quo.

Professional and credible journalists undertake thorough research to establish the veracity of their stories.

Fact-checking to know the truth is integral to the process of research, for it is fueled by an inquisitive and critical mind. Murray, Social News and UGC Hub (2016) suggest that before news readers share information on social media, they need to assess the integrity of the news source and check for similar news on legitimate media outlets. Genuine journalists do not rely on imagination for their news reports nor do they avoid doing research. They eschew propaganda and have no intention of misleading the public. They are messengers of truth, not lies.

5. Research Means to Find, Measure and Seize Opportunities

Research helps people nurture their potential and achieve goals through various opportunities. These can be in the form of securing employment, scholarships, training grants, project funding, business collaboration and budget traveling, among others.

For those looking for a job or for greener pastures, research is necessary. Through this process, not only will the unemployed increase their chances of finding potential employers either through job posting sites or employment agencies, but it can inform them if work opportunities are legitimate. Without research, the gullible, yet hopeful jobseeker or migrant worker may fall prey to unscrupulous headhunters who might be involved in illegal recruitment and/or human trafficking.

After finding a free or low-cost academic course or skills development training, students and professionals can assess their eligibility and know about application requirements and deadlines. Such an opportunity could hone their skills and knowledge, as well as enable them to build new connections.

Doing research also benefit civil society and its members. Funding for projects and research initiatives has been a top concern for those who want to address social issues. However, not all funding organizations accept neither proposals year-long nor are they interested in solving many social problems. Thus, it is necessary to research for agencies that match the objectives of individuals and non-profits involved in advocacy or programs that seek social change.

A wannabe business owner can likewise meet potential investors through research. He/She can examine their profiles and they can do the same. A good fit in terms of vision, mission, goals and work ethic, as well as the capital needed to launch the business is critical to making the opportunity succeed for both.

Some hobbies and interests are expensive to pursue. One of these is traveling. For budget-conscious tourists, searching for airfare and hotel promos, discount rides, and cheap markets is certainly a must to maximize the value of their money.

Seizing opportunities can broaden one's social network, raise one's awareness, or secure the support one direly needs to start a project or a business. Indeed, research contributes to a person's ability to make life-changing decisions. It encourages self-growth, participation in worthwhile causes, and living productively.

6. A Seed to Love Reading, Writing, Analyzing, and Sharing Valuable Information *Notes*

Research entails both reading and writing. These two literacy functions help enable computation and comprehension. Without these skills, it is less likely for anyone to appreciate and get involved in research. Reading opens the mind to a vast horizon of knowledge, while writing helps a reader use her/his own perspective and transform this into a more concrete idea that s/he understands.

Apart from reading and writing, listening and speaking are also integral in conducting research. Interviews, attending knowledge-generating events, and casual talks with anyone certainly aid in formulating research topics. They can also facilitate the critical thinking process. Listening to experts discuss the merits of their studies helps the listener to analyze a certain issue and write about such analysis. With the wide array of ideas available, scholars and non-scholars involved in research are able to share information with a larger audience. Some view this process as ego-boosting, while others see it as a means to stimulate interest and encourage further studies about certain issues or situations.

7. Nourishment and Exercise for the Mind

Curiosity may kill not just the cat, but the human as well. Yet, it is the same curiosity that fuels the mind to seek for answers. The College Admissions Partners notes how scientific research in particular "helps students develop critical reasoning skills helpful for any field of higher education." Such search or the thinking process is food for the brain, allowing creativity and logic to remain active. It also helps prevent mental illnesses like Alzheimer's.

Several studies have shown that mentally stimulating activities like doing research can contribute to brain health. Margaret Gatz (2005) enumerated research findings that support such position. However, she also noted that there may be other factors involved in averting said mental problem. One of these is intelligence. A study involving 11 year-old pupils in Scotland in 2000, for instance, pointed to intelligence quotient (IQ) scores as "predictive of future dementia risk". Gatz opined that clinical trials are needed and that "conclusions must be based on large samples, followed over a long period of time." She further posited:

Indeed, research and doing research encourage people to explore possibilities, to understand existing issues, and to disclose truths and fabricated ones. Without research, technological advancement and other developments could have remained a fantasy. Reading, writing, observing, analyzing, and social interaction facilitate an inquisitive mind's quest for knowledge, learning, and wisdom. Research serves as a bridge to achieve that goal.

1.9 ROLE OF RESEARCH IN FUNCTIONAL AREAS: FINANCE, MARKETING, HRD

Research related to Finance

The research on Finance deals with aspects of financial decision making for individuals, corporates and financial institutions and its consequences for the behaviour of financial markets. This is elaborated in three lines of research: (1) Financial management and corporate finance; (2) Financial markets; (3) Banking and regulation. All three themes formulated within the programme aim at performing fundamental research that can be translated into practically useful results. The finance department stimulates this dual objective by recruiting a balanced mix between full-time academics and academic practitioners, thus blending the cutting edge of academic and practitioner’s insight.

Notes

Area of Research Related to Finance

1. Financial Management and Corporate Finance

The first line of research addresses issues pertaining to how firms should attract funds, allocate capital, and manage risk. A key component in this approach is asset/liability management (ALM) as the simultaneous management of both sides of a firm's balance sheet. ALM focuses on how firm and institutions can efficaciously use available (financial) policy instruments to create value for stakeholders. ALM is encountered at institutions such as banks, insurance companies, pension funds, housing corporations, etc. The current line of research in ALM in the programme focuses on methodology, where the modeling of risk factors and the optimization of financial decisions form the key components. The key research question is what is driving the optimal decisions under different circumstances. This part of the programme aims to develop further into the direction of new investment opportunities beyond the scope of common modeling tools (credits, hedge funds, insurance risks). Such investment opportunities call for new measures of risk, new empirical models, and new optimization techniques. A second extension concerns the inclusion of derivatives and the management of derivative portfolios from a portfolio perspective.

Though tools are an essential ingredient of ALM analyses, in practice ALM is a process within an organization rather than a mere application of tools. The process usually consists of 3 iterative steps:

(i) Setting and identifying goals, means and restrictions,

(ii) Applying ALM tools to identify optimal ALM strategies given the goals, means and restrictions, and

(iii) An evaluation of optimal strategies and a feedback loop.

There is a high degree of interaction between stages 1 and 3, stemming from the fact that management becomes more aware of its preferences when confronted with measurable outcomes of ALM studies. Such increased awareness leads to a further delineation of goals and means. Due to the strategic aspects of ALM, ALM committees are usually located in the upper layers of an institution's governance structure. This more process orientated perspective to ALM forms an interesting possible extension to the current line of research in ALM. The relevant research questions for such an extension pertain to the strategic aspects of ALM for service organizations, the proper set-up (and success factors) of an ALM process and its communication within the organization, the perception of risk and goals for ALM and the control with related IT aspects.

2. Financial Markets

The research line on financial markets covers issues in market microstructure, asset pricing and market design. The red line between these topics is the behaviour and realization of financial asset prices. Market microstructure research studies the behaviour of price formation in financial markets at the transaction level. Important issues in this area are the flow and speed of information, the number and types of active traders and their optimal policies, transaction costs and the speed of adjustment of prices to new market conditions. These topics are addressed from both a theoretical and empirical perspective. Besides their academic relevance, these topics are also clearly relevant for policy makers in the current setting with rapidly developing trading systems, automation, Internet, and consolidation of alternative trading systems and floors.

At a higher level of aggregation (daily, monthly), the research on asset price behaviour takes a slightly different perspective. The main issue here is whether there is a clear-cut relationship between return and risk of an asset. While formal theoretical models usually state that there should be such a relationship ("no return increase without a corresponding risk increase"), there is ample empirical evidence casting doubt on the theoretical framework. The most important topics addressed by the programme in

Notes

this area concern the time varying nature of the risk-return relationship and its coherence with macroeconomic conditions, the reaction of financial markets to increased information flows (investor relationships), and the behaviour of risk over time (including volatility modeling).

The final issue in the programme's theme on Financial Markets concerns the design of new markets. Such markets might be created to trade alternative sources of risk. One can think of two different settings. One setting is that in which a market is set up for a relatively familiar risk factor. In that case, lessons may be drawn from similar markets for the same risk factor. Typical examples include setting up markets to cover producer price risk (e.g. cocoa) in Sub-Saharan countries. Research questions are: what are the main impediments to an efficient operation of the financial market; how can such impediments be removed, or if not, what second-best solutions can be derived and implemented. A different setting is the birth of new markets for old or new risk factors. One can think of markets for weather risk, catastrophe risk, etc. The scope, depth, behavior, and viability of such markets, as well as their repercussions on existing markets, are relevant topics for research.

3. Banking and Regulation

The banking sector has witnessed a strong degree of consolidation and (re-)regulation. Both the US and Europe have seen a large number of mergers over the past decade, with a focus on large or mega-mergers in the past few years. This has led to a reshaping of the financial landscape with corresponding effects on the role of (central) banks, the relationship between banks and customers, and the stability of the financial system as a whole. Moreover, the introduction of the Euro/EMU and the birth of financial conglomerates have called for a rethinking of the role of central banks and other regulators.

To study the issue on how consolidation has affected the financial landscape, the programme focuses on aspects of relationship lending and on the market's perception of bank mergers. The issue on relationship lending is well suited to check whether consolidation has spread out across all banking segments, or whether some banks are still able to exploit niches in the market. Alternatively, economies of scale and scope may have led to increased competition for banks that specialize in relationship lending, manifesting itself in relaxed conditions for granting loans. Another issue is the market's perception of the effect of bank mergers. Especially given the recent wave of mega-mergers a non-linear (and non-monotonic) relationship might be postulated between merger size and risk premia due to an increase in the number of banks that are deemed Too-Big-To-let-Fail.

The issues of Too-Big-To-let-Fail and the relaxation of loan conditions lead to the question of the stability of and policy effectiveness in the banking system and the financial system as a whole. These questions have become even more important in the wake of a new regulatory environment (the New Basle Capital Accord), the continuing emergence of economic and/or financial crises, and the EMU. The financial research studies the effect of these developments on the financial sector. Some of the major questions concern the effectiveness of the proposed new regulations in ensuring stability, the role of regulators and central banks in the new environment, and the effectiveness of monetary policy in an integrated Europe. As an aside, the regulatory developments also take place outside the banking framework, for example, in the pensions and insurance industry. The financial research also pays attention to stability and regulatory issues in this area.

Research Related to Marketing

The research on Marketing deals with aspects of systematic problem analysis, model building and fact-finding for the purpose of improved decision-making and control in the marketing of goods and services.

The environment for marketing has become extremely dynamic. Without adequate preparation, it is difficult for organizations to survive in such an environment. Research in marketing is one of the most

Notes effective tools that help organizations excel in the marketplace. Obtaining necessary information about customers' tastes and preferences is the key to business success.

Research in marketing provides information about consumers and their reactions to various products, prices, distribution, and promotion strategies. Marketers who collect accurate and relevant information quickly and design their strategies quicker than their competitors are more likely to be successful. Marketing research helps in effective planning and implementation of business decisions by providing accurate, relevant, and timely information. The process of marketing research involves a series of steps that systematically investigate a problem or an opportunity facing the organization.

This investigation starts with problem or opportunity recognition and definition, development of objectives for the research, development of hypothesis, planning the research design, selecting a research method, analyzing the research designs, selecting a sampling procedure, Market Information collection, evaluating and analyzing the Market Information and finally preparing and presenting the research report.

The research process provides a scientific platform, contrary to the traditional intuitive approach of decision making by managers which used to put large amounts of resources of the organization at risk. Organizations in areas such as IT, pharmaceuticals, telecom, manufacturing, transportation, advertising, banking, law, education and even governments utilize marketing research to find solutions to different kinds of decision-making problems. Marketing research is used in new product development, in segmenting markets, in identifying the needs of the customers, in sales forecasting and estimating the market potential of products and services, in analyzing the satisfaction levels of customers, and so on.

Role of Research in Marketing Functions

Marketing Research plays a very significant role in identifying the needs of customers and meeting them in best possible way. The main task of Marketing Research is systematic gathering and analysis of information. Before we proceed further, it is essential to clarify the relationship and difference between Marketing Research and Marketing Information System (MIS). Whatever information is generated by Marketing Research from internal sources, external sources, marketing intelligence agencies-consist the part of MIS. MIS is a set of formalized procedures for generating, analyzing, storing and distributing information to marketing decision makers on an ongoing basis. Marketing Research is essential for strategic market planning and decision making. It helps a firm in identifying what are the market opportunities and constraints, in developing and implementing market strategies, and in evaluating the effectiveness of marketing plans.

Marketing Research is a growing and widely used business activity as the sellers need to know more about their final consumers but are generally widely separated from those consumers. Marketing Research is a necessary link between marketing decision makers and the markets in which they operate. Marketing Research includes various important principles for generating information which is useful to managers. These principles relate to the timeliness and importance of data, the significance of defining objectives cautiously and clearly, and the need to avoid conducting research to support decisions already made.

Marketing research is one of the principal tools for answering questions because it:

i) Links the consumer, customer and public to the market through information used to identify and define marketing.

ii) Generates, refines and evaluates marketing actions.

iii) Monitors marketing performance.

iv) Underlines the understanding of marketing as a process.

How Research can support Marketing Operations? *Notes*

1. ***Systematic and continuous activity/process:*** MR is a continuous process. This is natural as new marketing problems are bound to come from time-to-time in the course of marketing of goods and services. One type of research is not adequate to resolve all marketing problems. Similarly, new research projects will have to be undertaken to solve new marketing problems and challenges.
2. ***Wide and comprehensive in scope:*** Marketing research is wide in scope as it deals with all aspects of marketing of goods and services. Introduction of new products, identification of potential markets, selection of appropriate selling techniques, study of market competition and consumer preferences, introduction of suitable advertising strategy and sales promotion measures, are some areas covered by MR.
3. ***Emphasizes on accurate data collection and critical analysis:*** In marketing research, suitable data should be collected objectively and accurately. The data collected must be reliable. It should be analyzed in a systematic manner. This will provide comprehensive picture of the situation and possible solutions.
4. ***Offers benefits to the company and consumers:*** Marketing research is useful to the sponsoring company. It raises the turnover and profit of the company. It also raises the competitive capacity and creates goodwill in the market. It enables a company to introduce consumer-oriented marketing policies. Consumers also get agreeable goods and more satisfaction due to marketing research activities.
5. ***Commercial equivalent of military intelligence:*** MR is a type of commercial intelligence activity. It facilitates planned activities in the field of marketing. It is similar to military intelligence where systematic study of the situation is made before taking any military action. Marketing research acts as the intelligence tool of management.
6. ***Tool for managerial decisions:*** MR acts as a tool in the hands of management for identifying and analyzing marketing problems and finding out solutions to them. It is an aid to decision-making. It suggests possible solutions for the consideration and selection by managers. Marketing research is an aid to judgment and never a substitute for it.

Area of Research related to Marketing

1. ***Product Research:*** Product means the goods and services which are sold to the consumers. It includes consumer products and industrial products. Product research studies the individual product. It studies the making and marketing of the product. It studies the colour, size, shape, quality, packaging, brand name and price of the product. It also deals with product modification, product innovation, product life cycle, etc. The product is modified (changed) as per the needs and wants of the consumers. Therefore, the product will not fail in the market.
2. ***Consumer Research:*** Consumer is the person who purchases the goods and services. The consumer is the king in the market. Consumer research studies consumer behaviour. It studies the consumers needs, wants, likes, dislikes, attitude, age, sex, income, location; buying motives, etc. This data is used to take decisions about the product, its price, place and promotion.
3. ***Packaging Research:*** Packaging research is a part of product research. It studies the package of the product. It improves the quality of the package. It makes the package more attractive. It makes the package more convenient for the consumers. It reduces the cost of packaging. It selects a suitable method for packaging. It also selects suitable packaging material.
4. ***Pricing Research:*** Pricing Research studies the pricing of the product. It selects a suitable method of pricing. It fixes the price for the product. It compares the company's price with the

Notes

competitor's price. It also fixes the discount and commission which are given to middlemen. It studies the market price trends. It also studies the future price trends.

5. ***Advertising Research:*** Advertising research studies the advertising of the product. It fixes the advertising objectives. It also fixes the advertising budget. It decides about the advertising message, layout, copy, slogan, headline, etc. It selects a suitable media for advertising. It also evaluates the effectiveness of advertising and other sales promotion techniques.
6. ***Sales Research:*** Sales research studies the selling activities of the company. It studies the sales outlets, sales territories, sales forecasting, sales trends, sales methods, effectiveness of the sales force, etc.
7. ***Distribution Research:*** Distribution research studies the channels of distribution. It selects a suitable channel for the product. It fixes the channel objectives. It identifies the channel functions like storage, grading, etc. It evaluates the competitor's channel.
8. ***Policy Research:*** Policy research studies the company's policies. It evaluates the effectiveness of the marketing policies, sales policies, distribution policies, pricing policies, inventory policies, etc. Necessary changes, if any, are made in these policies.
9. ***International Marketing Research:*** International marketing research studies the foreign market. It collects data about consumers from foreign countries. It collects data about the economic and political situation of different countries. It also collects data about the foreign competitors. This data is very useful for the exporters.
10. ***Motivation Research:*** Motivation research studies consumers' buying motives. It studies those factors that motivate consumers to buy a product. It mainly finds out, why the consumers buy the product? It also finds out the causes of consumer behaviour in the market.
11. ***Market Research:*** Market research studies the markets, market competition, market trends, etc. It also does sales forecasting. It estimates the demand for new products. It fixes the sales territories and sales quotas.
12. ***Media Research:*** Media research studies various advertising media. The different advertising media are television (TV), radio, newspapers, magazines, the internet, etc. Media research studies the merits and demerits of each media. It selects a suitable media for advertising. It does media planning. It also studies media cost. It helps in sales promotion and to avoid wastage in advertising.

Applications of Research in Marketing

Marketing research is the gathering, recording, and analyzing of Market Information that relates to a specific problem in marketing products or services. While this definition implies a systematic approach to marketing, marketing research is often performed as a reaction to a problem that occurs. Marketing research efforts, therefore, often are undertaken for specific projects that have set beginning and ending points.

1. ***Market and Economic Analysis:*** Market analysis involves analyzing market-segment factors to determine the market potential of a given product or service. The marketing researcher gathers Market Information and analyzes the factors that affect possible sales in a given market segment. The economic analysis is also used by marketing research departments to determine:
   - How actively a company should market in a given market segment?
   - How much money it should invest in marketing to that segment?
   - How much it may have to produce to fulfill the needs of the market segment?

Economic analysis often involves economic forecasting, which analyzes and attempts to forecast developing market trends and demands. *Notes*

2. ***Marketing Research for new product:*** Marketing research departments conduct product research for a variety of reasons, including:
   - Measuring potential acceptance of new products
   - Finding improvements or additions for existing products
   - Making changes or improvements in product packaging
   - Determining acceptability of a product over a competitor's product

   When a new product is being developed, marketing research departments will often use product concept testing to see how customers might react to the new product. Typically, before a business invests in the development of a prototype for a new or improved product, it will have its marketing researchers verbally describe or visually depict the prospective product to a group of potential customers in the target market.
3. ***Customer Satisfaction Research:*** Customer satisfaction research is that area of marketing research which focuses on customers' perceptions with their shopping or purchase experience.

   Many firms are interested in understanding what their customers thought about their shopping or purchase experience, because finding new customers is generally more costly and difficult than servicing existing or repeat customers.

   Many people are familiar with "business to customer" (B2C) or retail-level research, but there are also many "business to business" (B2B) or wholesale-level projects commissioned as well.

Research Related to HRD

Research on Human Resource Development deals with the aspects of improving the knowledge, ability, skills and other talents of their employees. It is the integrated use of training, organization, and career development efforts to improve individual, group, and organizational effectiveness.

Human Resource Development (HRD) research is a process of developing skills, competencies, knowledge, and attitudes of people in an organization. The people become human resource only when they are competent to perform organizational activities. Research on HRD ensures that the organization has such competent human resource to achieve its desired goals and objectives. HRD imparts the required knowledge and skill in them through an effective arrangement of training and development programs. HRD is an integral part of Human Resource Management (HRM) which is more concerned with training and development, career planning and development and the organization development. The organization has to understand the dynamics of HR and attempt to cope with changing the situation in order to deploy its HR effectively and efficiently. And HRD helps to reach this target. Hence, HRD is a conscious and proactive approach applied by employers which seek to capacitate employees through training and development to give their maximum to the organization and to fully use their potential to develop themselves.

The Role of Research in Human Resource Development

a) It is the result of advancing knowledge created in the past.

b) It is designed to solve particular existing problems that are likely to be profitable or solve problems of immediate concern.

c) Research is the basic foundation for a successful endeavor.

d) Organizations are messy entities. Studying people within the organization is challenging. Studying the external economic forces and their impact on an organization therefore adds another challenge.

Notes

1.10 CLASSIFICATION OF RESEARCH

1) Descriptive Research

Descriptive research seeks to provide an accurate description of observations of a phenomenon. It is a fact finding investigation with adequate interpretation. It is the simplest type of research which focuses on particular aspects or dimensions of the problem studied. It is designed to gather descriptive information and provide information for formulating more sophisticated studies. The objective of descriptive research is to map the terrain of a specific phenomenon. A descriptive study identifies relevant variables but does not aim at testing hypothesis. It applies simple statistical techniques like averages and percentages. In social science and business research the term "Ex Post Facto Research" is used for descriptive research studies.

A study of this type could start with questions such as: 'What similarities or contrasts exist between A and B? where A and B are different departments in the same organisation, different regional operations of the same firm or different companies in the same industry. Such descriptive comparisons can produce useful insights and lead to hypothesis-formation. Descriptive studies are valuable in providing facts needed for planning social action programmes. *Example:* A detailed set of data on the profile of clients would be an example of this type of research. By understanding the customer better, sales and marketing management will be able to take better decisions on new product development.

2) Exploratory Research

Exploratory research is a preliminary study of an unfamiliar problem about which the researcher has little or no knowledge. It involves a literature search or conducting focus group interviews. The exploration of new phenomena in this way may help the researcher's need for better understanding, may test the feasibility of a more extensive study or determine the best methods to be used in a subsequent study. For these reasons, exploratory research is broad in focus and rarely provides definite answers to specific research issues.

Exploratory means which are not known to us before but has existence. Just if anything discover or unearth or unveil that thing then it will be exploratory research. Exploratory research not only include the things about which man cannot think before but also include the things which are already has been described by someone but you are describing it from different angle or different view point.

The objective of exploratory research is to identify key issues and key variables. For example, one outcome might be a better system of measurement for a specific variable. If the researcher defines the study as exploratory research, then there is a need to clearly define the objectives. e.g.: An example in the business environment might be an exploratory study of a new management technique in order to brief a management team. This would be a vital first step before deciding whether to embrace the technique.

Purposes of Exploratory Research

The purpose of an exploratory study may be:

(i) To generate new ideas.

(ii) To increase the researcher's familiarity with the problem.

(iii) To make a precise formulation of the problem.

(iv) To gather information for clarifying concepts.

(v) To determine whether it is feasible to attempt the study.

Notes

3) Applied Research

Applied research refers to scientific study and research that seeks to solve practical problems. Applied research is used to find solutions to everyday problems, cure illness, and develop innovative technologies.

It is designed to solve practical problems of the modern world, rather than to acquire knowledge for knowledge sake. It is also known as action research. The goal of applied research is to improve the human condition. It focuses on analysis and solving social and real life problems. This research is generally conducted on a large scale basis. As such, it is often conducted with the support of some financing agency like government, public corporations, World Bank, UNICEF, UGC etc. *According to Hunt,* "Applied research is an investigation for ways of using scientific knowledge to solve practical problems". *For example:* Improve agriculture crop production, treat or cure a specific disease, improve energy efficiency homes and offices, improving communication among workers in large companies.

Applied research can be further classified as problem oriented and problem solving research.

a) ***Problem oriented research:*** Research is done by industry apex body for sorting out problems faced by all the companies. *For example,* WTO does problem oriented research for developing countries, In India, agriculture and processed food export development authority (APEDA) conduct regular research for the benefit of Agri-industry.

b) ***Problem solving research:*** This type of research is done by an individual company for the problem faced by it. Marketing research and market research are typical problem solving research. *For example,* Videocon international conducts research to study customer satisfaction level. In short, the main aim of applied research is to discover some solution for some pressing practical problem.

4) Fundamental Research/Pure Research

Pure research advances fundamental knowledge about the human world. It focuses on refuting or supporting theories that explain how this world operates, what makes things happen, why social relations are a certain way and why society changes. Pure research is the source of most new scientific ideas and ways of thinking about the world. It can be exploratory, descriptive or explanatory.

Pure research generates new ideas, principles and theories, which may not be immediately utilized; though are the foundations of modern progress and development in different fields. Today's computers could not exist without the pure research in mathematics conducted over a century ago, for which there was no known practical application at that time.

It is concerned with generalizations and formulation of theories. It involves developing and testing theories and hypotheses that are intellectually challenging to the researcher but may or may not have practical application at the present time or in the future. "Gathering Knowledge for Knowledge's sake" is termed as pure research. Research concerning some natural phenomenon or relating to pure mathematics are examples of pure or fundamental research. The knowledge produced through pure research is sought in order to add to the existing body of scientific knowledge.

5) Quantitative Research

Quantitative research is generally associated with the positivist/post-positivist paradigm. It usually involves collecting and converting data into numerical form so that statistical calculations can be made and conclusions drawn.

Quantitative research is based on measurement of quantity or amount. It aims to measure the quantity or amount and compares it with past records and tries to project for future periods. In social sciences, "Quantitative research refers to the systematic empirical investigation of quantitative properties

Notes

and phenomena and their relationships". The process of measurement is central to quantitative research because it provides fundamental connection between empirical observation and mathematical expression of quantitative relationships. Statistics is the most widely used branch of mathematics in quantitative research. Statistical methods are used extensively in fields such as economics and commerce. Quantitative research involves the use of structured questions, where the response options have been pre-determined and large number of respondents are involved. *For example,* Total sales of soap industry in terms of rupees and quantity in terms of for a particular year, say 2012, could be researched, compared with past 5 years and then projection for 2013 could be made.

6) Qualitative Research

Qualitative research involves looking in-depth at non-numerical data. This is a method of inquiry employed in many different academic disciplines, traditionally in the social sciences, but also in market research and further contexts.

Qualitative research is concerned with qualitative phenomenon.It presents non-quantitative type of analysis. Qualitative research is collecting, analyzing and interpreting data by observing what people do and say. Qualitative research refers to the meanings, definitions, characteristics, symbols, metaphors, and description of things. Qualitative research is much more subjective and uses very different methods of collecting information, mainly individual, in-depth interviews and focus groups. The nature of this type of research is exploratory and open ended. Small numbers of people are interviewed in depth or a relatively small number of focus groups are conducted. Other techniques include word association tests, sentence completion tests and other projective techniques. Qualitative research is specially important in behavioral sciences.

Qualitative research can be further classified as:

a) Phenomenology: A form of research in which the researcher attempts to understand how one or more individuals experience a phenomenon. Example: Researcher might interview 20 victims of Bhopal gas tragedy.

b) Ethnography: This type of research focuses on describing the culture of a group of people. A culture is the shared attributes, values, norms, practices, language and material things of a group of people. Example: Researcher might decide to go and live with the tribals in Andaman island and study the culture and the educational practices.

c) Case study: It is a form of qualitative research that is focused on providing a detailed account of one or more cases.

7) Conceptual Research

Conceptual research is involves investigation of thoughts and ideas and developing new ideas or interpreting the old ones based on logical reasoning. A conceptual framework is used in research to outline possible courses of action or to present a preferred approach to an idea or thought. For example, the philosopher Isaiah Berlin used the "hedgehogs" versus "foxes" approach; a "hedgehog" might approach the world in terms of a single organizing principle; a "fox" might pursue multiple conflicting goals simultaneously. Alternatively, an empiricist might approach a subject by direct examination, whereas an intuitionist might simply intuit what's next.

It is that related to some abstract idea or theory. It is generally used by philosophers and thinkers to develop new concepts or to reinterpret existing ones. Conceptual analysis is the preferred method of analysis in social sciences and philosophy. Here, a researcher breaks down a theorem or concept into its constituent parts to gain a better understanding of the deeper philosophical issue concerning the theorem. Though this method of analysis has gained popularity, there are sharp critiques of the method. However, most agree that conceptual analysis is a useful method of analysis but should be used in conjunction with other methods of analysis to produce better results.

Notes

8) Empirical Research

It relies on experience or involves observation alone, often without due regard for system and theory. It is data-based research, with analyses coming up with conclusions, which are capable of being verified by observation or experiment. Such research is thus, characterized 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. Research design varies by field and by the questions being investigated. Many researchers combine qualitative and quantitative forms of analysis to better answer questions which cannot be studied in laboratory settings, particularly in the social sciences and in education.

9) Experimental Research

Experimental research is designed to assess the effects of particular variables on a phenomenon by keeping the other variables constant or controlled. It aims at determining whether and in what manner variables are related to each other. The factor which is influenced by other factors is called a dependent variable and the other factors which influence it are known as independent variables. It is conducted mostly in laboratories in the context of basic research. The principle advantage of experimental designs is that it provides the opportunity to identify cause and effect relationships. Non-experimental research (case studies, surveys, correlation studies) is non-manipulative observational research usually conducted in natural settings. While laboratory controlled experimental studies tend to be higher in internal validity, non-experimental studies tend to be higher in external validity. One major limitation of experimental research is that studies are typically conducted in contrived or artificial laboratory settings. Results may not generalize or extrapolate to external settings. Two exceptions to this rule are natural experiments and field experiments. Natural experiments document and compare the behaviors of subjects before and after some natural event; e.g., floods, tornadoes, hurricanes. Field experiments involve manipulating conditions in the natural setting for the purpose of determining their influence on behavior. The field experiment is unique as it tends to be moderately high on both external and internal validity.

In experimental research, the investigator manipulates conditions for the purpose of determining their effect on behavior. Subjects should be unaware of their membership in an experimental group so that they don't act differently (Hawthorne Effect). In the simplest experimental design, investigators administer a placebo to the control group and a treatment to the experimental group. Experimental designs vary in terms of subjects' assignments to different groups, whether subjects were pre-tested, whether different treatments were administered to different groups, and the number of variables being investigated. Experiments are typically structured in terms of independent, organism, and dependent variables. The independent variable is a manipulated environmental stimulus dimension, the organism-variable is some dimension (e.g., sex, race) of more or less stable characteristics of the organism, and the dependent variable is a behavioral dimension that reflects the influence of the independent and organism-variables. The general objective in experimental research is to define the relationship between the antecedent (independent and organism) variables and the consequent (dependent) variables.

10) Historical Research

It refers to the induction of principles through research into the past and social forces which have shaped the present. It is the process of systematically studying past records with a view to reconstruct the origin and development of an institution or a movement or a system and discovering trends in the past. It is not a mere accumulation of facts and dates or even a description of past events. It is a flowing, dynamic account of past events which involves an interpretation of these events in an attempt to recapture the nuances, personalities and ideas that influenced these events. One of the goals of historical research

Notes is to communicate an understanding of past events. It is a difficult task as it must often depend on inference and logical analysis of recorded data and indirect evidences rather than upon direct observation.

Significance of Historical Research

The following gives five important reasons for conducting historical research:

1. To uncover the unknown (i.e., some historical events are not recorded).
2. To answer questions (i.e., there are many questions about our past that we not only want to know but can profit from knowing).
3. To identify the relationship that the past has to the present (i.e., knowing about the past can frequently give a better perspective of current events).
4. To record and evaluate the accomplishments of individuals, agencies or institutions.
5. To assist in understanding the culture in which we live (e.g., education is a part of our history and our culture).

1.11 APPROACHES TO RESEARCH

I. Quantitative Approach

Quantitative research is generally associated with the positivist/post positivist paradigm. It usually involves collecting and converting data into numerical form so that statistical calculations can be made and conclusions drawn.

Process

Researchers will have one or more hypotheses. These are the questions that they want to address which include predictions about possible relationships between the things they want to investigate (variables). In order to find answers to these questions, the researchers will also have various instruments and materials (e.g. paper or computer tests, observation check lists etc.) and a clearly defined plan of action. Data is collected by various means following a strict procedure and prepared for statistical analysis. Nowadays, this is carried out with the aid of sophisticated statistical computer packages. The analysis enables the researchers to determine to what extent there is a relationship between two or more variables. This could be a simple association (e.g. people who exercise on a daily basis have lower blood pressure) or a causal relationship (e.g. daily exercise actually leads to lower blood pressure). Statistical analysis permits researchers to discover complex causal relationships and to determine to what extent one variable influences another.

The results of statistical analyses are presented in journals in a standard format. For people who are not familiar with scientific research jargon, the discussion sections at the end of articles in peer reviewed journals usually describe the results of the study and explain the implications of the findings in straightforward terms.

Principles

Objectivity is very important in quantitative research. Consequently, researchers take great care to avoid their own presence, behaviour or attitude affecting the results. They also critically examine their methods and conclusions for any possible bias. Researchers go to great lengths to ensure that they are really measuring what they claim to be measuring. For example, if the study is about whether background music has a positive impact on restlessness in residents in a nursing home, the researchers must be clear about what kind of music to include, the volume of the music, what they mean by restlessness, how to measure restlessness and what is considered a positive impact. This must all be considered, prepared

Notes

and controlled in advance. External factors, which might affect the results, must also be controlled for. In the above example, it would be important to make sure that the introduction of the music was not accompanied by other changes as it might be the other factor which produces the results. Some possible contributing factors cannot always be ruled out but should be acknowledged by the researchers. The main emphasis of quantitative research is on deductive reasoning which tends to move from the general to the specific. This is sometimes referred to as a top down approach. The validity of conclusions is shown to be dependent on one or more premises being valid. If the premises of an argument are inaccurate, then the argument is inaccurate. However, most studies also include an element of inductive reasoning at some stage of the research.

Researchers rarely have access to all the members of a particular group for example, all people with dementia, healthcare professionals. However, they are usually interested in being able to make inferences from their study about these larger groups. For this reason, it is important that the people involved in the study are a representative sample of the wider population/group. However, the extent to which generalizations are possible depends to a certain extent on the number of people involved in the study, how they were selected and whether they are representative of the wider group. For example, generalizations about psychiatrists should be based on a study involving psychiatrists and not one based on psychology students. In most cases, random samples are preferred but sometimes researchers might want to ensure that they include a certain number of people with specific characteristics and this would not be possible using random sampling methods. Generalizability of the results is not limited to groups of people but also to situations. It is presumed that the results of a laboratory experiment reflect the real life situation which the study seeks to clarify.

II. Qualitative Approach

Qualitative research is the approach usually associated with the social constructivist paradigm which emphasizes the socially constructed nature of reality. It is about recording, analysing and attempting to uncover the deeper meaning and significance of human behaviour and experience, including contradictory beliefs, behaviours and emotions. Researchers are interested in gaining a rich and complex understanding of people's experience and not in obtaining information which can be generalized to other larger groups.

Process

The approach adopted by qualitative researchers tends to be inductive which means that they develop a theory or look for a pattern of meaning on the basis of the data that they have collected. This involves a move from the specific to the general and is sometimes called a bottom-up approach. However, most research projects also involve a certain degree of deductive reasoning.

Qualitative researchers do not base their research on pre-determined hypotheses. Nevertheless, they clearly identify a problem or topic that they want to explore and may be guided by a theoretical lens - a kind of overarching theory which provides a framework for their investigation. The approach to data collection and analysis is methodical but allows for greater flexibility than in quantitative research. Data is collected in textual form on the basis of observation and interaction with the participants e.g. through participant observation, in-depth interviews and focus groups. It is not converted into numerical form and is not statistically analysed. Data collection may be carried out in several stages rather than once and for all. The researchers may even adapt the process mid-way, deciding to address additional issues or dropping questions which are not appropriate on the basis of what they learn during the process. In some cases, the researchers will interview or observe a set number of people. In other cases, the process of data collection and analysis may continue until the researchers find that no new issues are emerging.

Notes

Principles

Researchers will tend to use methods which give participants a certain degree of freedom and permit spontaneity rather than forcing them to select from a set of pre-determined responses which none might be appropriate or accurately describe the participant's thoughts, feelings, attitudes and to try to create the right atmosphere to enable people to express themselves. This may mean adopting a less formal and less rigid approach than that used in quantitative research. It is believed that people are constantly trying to attribute meaning to their experience. Therefore, it would make no sense to limit the study to the researcher's view or understanding of the situation and expect to learn something new about the experience of the participants. Consequently, the methods used may be more open-ended, less narrow and more exploratory (particularly when very little is known about a particular subject). The researchers are free to go beyond the initial response that the participant gives and to ask why, how, in what way etc. In this way, subsequent questions can be tailored to the responses just given.

Qualitative research often involves a smaller number of participants. This may be because the methods used such as in-depth interviews are time and labour intensive but also because a large number of people are not needed for the purposes of statistical analysis or to make generalizations from the results. The smaller number of people typically involved in qualitative research studies and the greater degree of flexibility does not make the study in any way "less scientific" than a typical quantitative study involving more subjects and carried out in a much more rigid manner. The objectives of the two types of research and their underlying philosophical assumptions are simply different.

III. Pragmatic Approach (Mixed Methods)

The pragmatic approach to science involves using the method which appears best suited to the research problem and not getting caught up in philosophical debates about which is the best approach. Pragmatic researchers therefore grant themselves the freedom to use any of the methods, techniques and procedures typically associated with quantitative or qualitative research. They recognise that every method has its limitations and that the different approaches can be complementary. They may also use different techniques at the same time or one after the other. *For example,* they might start with face-to-face interviews with several people or have a focus group and then use the findings to construct a questionnaire to measure attitudes in a large scale sample with the aim of carrying out statistical analysis. Depending on which measures have been used, the data collected is analysed in the appropriate manner. However, it is sometimes possible to transform qualitative data into quantitative data and vice versa although transforming quantitative data into qualitative data is not very common.

Being able to mix different approaches has the advantages of enabling triangulation. Triangulation is a common feature of mixed methods studies. It involves

(i) The use of a variety of data sources (data triangulation)

(ii) The use of several different researchers (investigator triangulation)

(iii) The use of multiple perspectives to interpret the results (theory triangulation)

(iv) The use of multiple methods to study a research problem (methodological triangulation)

In some studies, qualitative and quantitative methods are used simultaneously. In others, first one approach is used and then the next, with the second part of the study perhaps expanding on the results of the first. For example, a qualitative study involving in-depth interviews or focus group discussions might serve to obtain information which will then be used to contribute towards the development of an experimental measure or attitude scale, the results of which will be analysed statistically.

IV. Advocacy/Participatory Approach (Emancipatory Method)

To some degree, researchers adopting an advocacy/participatory approach feel that the approaches to research described so far do not respond to the needs or situation of people from marginalised or

Notes

vulnerable groups. As they aim to bring about positive change in the lives of the research subjects, their approach is sometimes described as emancipatory. It is not a neutral stance. The researchers are likely to have a political agenda and to try to give the groups they are studying a voice. As they want their research to directly or indirectly result in some kind of reform, it is important that they involve the group being studied in the research, preferably at all stages, so as to avoid further marginalising them.

The researchers may adopt a less neutral position than that which is usually required in scientific research. This might involve interacting informally or even living amongst the research participants. The findings of the research might be reported in more personal terms, often using the precise words of the research participants. Whilst this type of research could by criticised for not being objective, it should be noted that for some groups of people or for certain situations, it is necessary as otherwise the thoughts, feelings or behaviour of the various members of the group could not be accessed or fully understood. For this reason, researchers are sometimes members of the group they are studying or have something in common with the members of the group.

1.12 ADVANTAGES AND LIMITATIONS OF RESEARCH

Advantages of Research

Various advantages of Research are:

(i) ***Research addresses the target audience:*** Research helps to address the target audience. The organization asking for the research has the complete control on the process and the research is streamlines as far as its objectives and scope is concerned. Researching company can be asked to concentrate their efforts to find data regarding specific market rather than concentration on mass market.

(ii) ***Research helps to identify the problems:*** The problem should be clearly defined and sharply delineated. The statement of the decision problem should include its scope, limitations and precise specifications of areas significant to research.

(iii) ***Research assists for data interpretation:*** The collected data can be examined and interpreted by the marketers depending on their needs rather than relying on the interpretation made by collectors of secondary data.

(iv) ***Research ensures the data proprietary:*** Collector of primary data is the owner of that information and he need not share it with other companies and competitors. This gives an edge over competitors replying on secondary data.

(v) ***Research considers the objectivity:*** Research concentrates on identifying and working towards common objectives. Objective is the sense that it must answer the research questions. This necessitates the formulation of a proper hypothesis; otherwise there may be lack of congruence between the research questions and the hypothesis.

(vi) ***Research helps to find the ambiguously:*** Generalizations that outrun the evidence on which the researchers are based tend to leave an unfavorable impression. Such reports are not valuable to managers for business decision making. Presentation should be comprehensive, easily understood and organized. Language should be restrained, clear and precise when findings are presented.

Limitations of Research

1) ***Lack of Training:*** The lack of scientific training in the methodology of research is a great handicap for researchers in our country. There is a paucity of competent researchers in our country.

Notes

2) ***Lack of confidence:*** The business houses are often reluctant to supply the needed information to research because of fear of misuse of information.

3) ***Repetition:*** Research studies overlapping one another are undertaken quite often for want of adequate information.

4) ***Lack of Interaction:*** There is insufficient interaction between the university research department, on the one hand and business establishments, government departments and research institutions, on the other.

5) ***Absence of Code of Conduct:*** There does not exist a code of conduct for researchers and inter-University and inter-departmental rivalries are also quite common.

6) ***Lack of Resources:*** For conducting quality research adequate funds are not provided.

7) ***Lack of Co-ordination:*** There exists lack of co-ordination among various agencies responsible for conducting research.

8) ***Problem of Conceptualization:*** Many a time problems of conceptualization and problems relating to the process of data collection and related things crop up resulting in wastage of resources.

1.13 RESEARCH METHODOLOGY

Research methodology is a way to systematically solve the research problem. It may be understood as a science of studying how research is done scientifically. In it we study the various steps that are generally adopted by a researcher in studying his research problem along with the logic behind them. It is necessary for the researcher to know not only the research methods/techniques but also the methodology. Researchers not only need to know how to develop certain indices or tests, how to calculate the mean, the mode, the median or the standard deviation or chi-square, how to apply particular research techniques, but they also need to know which of these methods or techniques, are relevant and which are not, and what would they mean and indicate and why. Researchers also need to understand the assumptions underlying various techniques and they need to know the criteria by which they can decide that certain techniques and procedures will be applicable to certain problems and others will not. All this means that it is necessary for the researcher to design his methodology for his problem as the same may differ from problem to problem. For example, an architect, who designs a building, has to consciously evaluate the basis of his decisions, i.e., he has to evaluate why and on what basis he selects particular size, number and location of doors, windows and ventilators, uses particular materials and not others and the like. Similarly, in research the scientist has to expose the research decisions to evaluation before they are implemented. He has to specify very clearly and precisely what decisions he selects and why he selects them so that they can be evaluated by others also. From what has been stated above, we can say that research methodology has many dimensions and research methods do constitute a part of the research methodology. The scope of research methodology is wider than that of research methods. Thus, when we talk of research methodology we not only talk of the research methods but also consider the logic behind the methods we use in the context of our research study and explain why we are using a particular method or technique and why we are not using others so that research results are capable of being evaluated either by the researcher himself or by others. Why a research study has been undertaken, how the research problem has been defined, in what way and why the hypothesis has been formulated, what data have been collected and what particular method has been adopted, why particular technique of analysing data has been used and a host of similar other questions are usually answered when we talk of research methodology concerning a research problem or study.

Notes

1.14 PROCESS OF RESEARCH

The research process consists of the following distinctive interrelated phases:

Step: 1 Defining the Research problem

Step: 2 Review of Literature

Step: 3 Formulation of Hypothesis

Step: 4 Developing the Research Design

Step: 5 Data Collection

Step: 6 Data Analysis and Interpretation

Step: 7 Research Reporting.

Research process involves execution of a series of phases towards accomplishment of the objectives of research. Each phase in the research process need not be carried out in a sequential process. Some the phases can be carried out simultaneously. One should remember that the various steps involved in research are not mutually exclusive; nor they are separate and distinct. They do not necessarily follow each other in any specific order and the researcher has to be constantly anticipating at each step in the research process the requirements of the subsequent steps. However, the idea of sequence will be useful for developing and carrying out research study in a systematic manner.

Step - 1 Defining the Research Problem

A problem need not necessarily mean that something is wrong in the current situation which needs to be rectified immediately. It simply indicates an issue for which finding a solution could help to improve an existing situation. Problem can be defined as any situation where a gap exists between the actual and the desired state. Problem statement or problem definition refers to a clear, precise and succinct statement of question or issue that is to be investigated with the goal of finding an answer or solution.

Components of Research Problem

The components of research problem are as suggested by R. L. Ackoff in the "Design of Social Research" is elaborated below:

There must be an individual or a group which has some difficulty with problem:

(i) There must be some objective(s) to be attained at.

(ii) There must be alternative means or course of action for obtaining the objectives

(iii) There must be some doubt in the minds of a researcher with regard to the selection of alternatives.

(iv) There must be some environment to which the difficulty pertains.

Criteria for Selecting the Research Problem

The following criteria can be kept in the minds of researchers in selecting the research problem:

(i) Subjects on which the research is carried on amply should not be normally chosen as there will not be a new dimension to reveal.

(ii) Too narrow or too vague problems should be avoided.

(iii) The researcher should be familiar with the subject chosen for research. The researcher should have enough knowledge, qualification and training in the selected problem area.

(iv) The resources needed to solve the problem in terms of time, money, efforts, manpower requirement should be taken into account before embarking on a problem.

Notes

(v) The subject of research should be familiar and feasible so that related research material or sources of research can be obtained easily.

(vi) The selection of a problem must be preceded by a preliminary study.

Research problems trigger the research process. Defining the research problem is a critical activity. A thorough understanding of research problem is a must for achieving success in the research endeavor. Defining the research problem begins with identifying the basic dilemma that prompts the research. It can be further developed by progressively breaking down the original dilemma into more specific and focus oriented objectives.

Five steps could be envisaged:

(1) Identifying the broad problem area

(2) Literature review

(3) Identifying the research question

(4) Refining the research question

(5) Developing investigative questions.

Step - 2 Review of Literature

Literature survey is the review of published and unpublished work from secondary sources in the area of interest to the researcher. The purpose of conducting literature survey at this stage is:

(i) To document the studies relevant to the problem identified for research.

(ii) To ensure that no variable that has been taken up in the past related studies is ignored.

(iii) To avoid conducting similar type of study and thereby stopping the researcher from investing his resources in terms of time and effort in a research venture which is already solved.

(iv) To provide a good frame work and a solid foundation to proceed further in the investigation.

(v) To have a comprehensive theoretical framework from which hypothesis can be developed for testing.

(vi) To enable to develop the problem statement in a precise and clear manner.

(vii) To enhance the testability and replicability of the findings of the current research.

(viii) To understand the research gap.

(ix) To stimulate the researcher to carry out the work.

(x) To confirm the appropriateness of procedure by referring to similar studies conducted in the past.

(xi) To trace inconsistencies, contradictions and consistencies.

(xii) To clear conceptualization.

(xiii) To familiarize with methodology, research tools and statistical analysis.

The literature review needs to be performed on the variables identified through the interview process.

It comprises of three steps viz., (i) Identifying the sources

(ii) Gathering relevant information

(iii) Writing up the Literature review.

Notes

(i) Identifying the Sources

The data can be obtained from the library by going through books, journals, newspapers, magazines, conference proceedings, doctoral dissertation, thesis, government publications and other reports. The development of information technology has led to many online databases like Prowess, EBSCO etc and the interlinking of libraries has led to a myriad of information in the hands of the researcher with the click of the mouse. Computerized databases include bibliographies, abstract and full text of articles. Bibliographic databases display only the bibliographic citations i.e., the name of the author, the title of the article / journal, source of publication, year, volume and page numbers. The abstract databases in addition to the above said information provides an abstract or summary of the article. The full-text databases as the name suggests, enables to download the full text of the article.

(ii) Gathering Relevant Information

The articles gathered either from books, journal or online sources could as such act as a reservoir of information. These sources could lead to further information through the citation and references used. The list of journals and references referred in the articles could lead us further to the source of information. Also, during the course of reading the articles, the researcher can get insight into new variables or new avenues hitherto unexplored.

(iii) Presenting the Literature Review

The literature should be presented in a clear and logical manner citing the author, year of study, objectives of the research, major findings and implications. The researcher should present the literature in a chronological order and in a coherent manner. There are several methods of citing references in the literature.

What is Hypothesis?

Ordinarily, when one talks about hyppthesis, one simply means 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 forth as an explation for the occurance of some specified group of phenomena either asserted merely as a provisional conjencture to guide some investigation or accepted as highly probable in the light of established facts. Quite often a research hypothesis is a predective statement, capable of being tested by scientific methods, that related an independent variable to some dependent variables. *For example,* consider statements like the following ones. “Students who receive counselling will show a greater increase in creativity than students not receiving Counselling “Or” The automobile A is performing as well as automobile B”

Step - 3 Formulation of Hypothesis

A proposition that can be verified to determine its reality is a hypothesis. Therefore one can say that a hypothesis is a verifiable counterpart of a proposition. A hypothesis may be defined as a logically conjectured relationship between two or more variables, expressed in the form of a testable statement. Relationship is proposed by using a strong logical argumentation. This logical relationship may be part of theoretical framework of the study.

Step - 4 Developing the Research Design

A Research design is the specification of methods and procedures for acquiring the information needed to structure or to solve problems. It is an overall operational pattern or framework of the project that stipulates the information to be collected, the sources from which information can be collected the procedures for collection of information. In other words the researcher should consider: (i) the design technique (ii) the type of data (iii) the sampling methodology and procedures (iv) the schedule and the budget.

Notes A good research design ensures that the information obtained is relevant to the research problem in an objective and economical manner. The research design can be described as a master plan or model or blueprint for the conduct of investigation.

Step - 5 Collection of Data

The data gathering phase begins with the pilot testing. It is done to detect the weakness in the research design, questionnaire/interview schedule and provides proxy data for selection of probability sample. The pilot testing should stimulate the procedure and protocols designed for data collection. If the study is to be conducted by email then the pilot questionnaire should be email. The size of the pilot group may range normally from 25 to 100 respondents who need not be statistically selected. There are a number of variations of pilot testing. Some of them may be restricted to data collection only. One form is 'pretesting' where the responses are collected from colleagues, respondents surrogates or actual respondents for the main purpose of refining the questionnaire. Based on the pilot testing the questionnaire may be redesigned, rephrased and improved. Pretesting may be repeated many times to refine questions or procedures.

Data is the facts presented to the researcher from the study environment. Data can be gathered from a singe location or from all over the world based on the research objectives and the resource allocation. The data collection method ranges from observation, questionnaires, laboratory notes and other modern instruments and devices. Data can be characterized by their abstractness, verifiability, elusiveness and closeness to the phenomenon. As abstractions, data are more metaphorical than real. When sensory experiences consistently produce the same result then the data is said to be trustworthy as they are verified. Data capturing is elusive, complicated by the speed at which events occur and the time-bound nature of observation. Data reflect their truthfulness measured by the degree of closeness to the phenomena. Secondary data has at least one level of interpretation inserted between the event and its recording. Primary data are close to the truth. Data collected need to be edited for ensuring consistency and to locate omissions. In case of survey method editing reduces errors in the recording, improves legibility and clarifies unclear and inappropriate responses. Edited data are then converted into analyzable form. Computers can be used to find missing data, validate data, edit and code so that further analysis can be carried out in a valid manner.

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. But in the case of a survey, data can be collected by any one or more of the following ways:

(i) ***By observation:*** This method implies the collection of information by way of investigator's own observations, without interviewing the respondents. The information obtained relates to what is currently happening and is not complicated by either the past behaviour or future intentions or attitudes of respondents. This method is no doubt an expensive method and the information provided by this method is also very limited. As such this method is not suitable in inquiries where large samples are concerned.

(ii) ***Through personal interview:*** The investigator follows a rigid procedure and seeks answers to a set of pre-conceived questions through personal interviews. This method of collecting data is usually carried out in a structured way where output depends upon the ability of the interviewer to a large extent.

(iii) ***Through Telephone Interviews:*** This method of collecting information involves contacting the respondents on telephone itself. This is not a very widely used method but it plays an important role in industrial surveys in developed regions, particularly, when the survey has to be accomplished in a very limited time.

Notes

(iv) ***By mailing of questionnaires:*** The researcher and the respondents do come in contact with each other if this method of survey is adopted. Questionnaires are mailed to the respondents with a request to return after completing the same. It is the most extensively used method in various economic and business surveys. Before applying this method, usually a Pilot Study for testing the questionnaire is conduced which reveals the weaknesses, if any, of the questionnaire. Questionnaire to be used must be prepared very carefully so that it may prove to be effective in collecting the relevant information.

(v) ***Through schedules:*** Under this method the enumerators are appointed and given training. They are provided with schedules containing relevant questions. These enumerators go to respondents with these schedules. Data are collected by filling up the schedules by enumerators on the basis of replies given by respondents. Much depends upon the capability of enumerators so far as this method is concerned. Some occasional field checks on the work of the enumerators may ensure sincere work.

The researcher should select one of these methods of collecting the data taking into consideration the nature of investigation, objective and scope of the inquiry, financial resources, available time and the desired degree of accuracy. Though he should pay attention to all these factors but much depends upon the ability and experience of the researcher. In this context Dr A.L. Bowley very aptly remarks that "In collection of statistical data ,commonsense is the chief requisite and experience the chief teacher".

Step - 6 Data Analysis and Interpretation

Research is conducted for the purpose of acquiring information. Raw data as such does not provide information. Further analysis needs to be done to obtain information out of data. Data analysis involves application of statistical techniques for reducing accumulated data to a manageable size leading to summaries. Responses acquired by way of administering questionnaires should be subjected to analysis so as to ascertain the behaviour of variables, the relationship between variables etc. Analysis should be focused to find answers to research questions/hypothesis. Various statistical softwares are available to make the job of data analysis easier,. However, interpretation needs to be made with expertise as the recommendations are based on them.

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 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, specially 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). 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.

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 research.

Notes

Step - 7 Research Reporting

It is only through reports the researcher communicates about the research work, findings and recommendations to the outside world. The report has to be prepared in the style that will be understood by the target audience. The reports may be communicated by way of written documents or in an oral manner, through letters or through telephone calls or a combination of all. The type of report varies depending on the type of research, length of report and the purpose. The researcher should take care to see that the report addresses all the objectives of research in a lucid manner. The report should be adapted to the needs of the target audience and care must be taken to use appropriate words in projecting the interpretation, recommendations and conclusion. A report should contain an executive summary consisting of synopsis of problem, findings and recommendations. It should speak about the background of the study, the statement of the problem, literature summary, methods and procedures, findings, recommendations and conclusion

The layout of the report should be as follows:

(i) The preliminary pages;

(ii) The main text and

(iii) The end matter.

In its preliminary pages the report should carry title and date followed by acknowledgment 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. 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. Charts and illustrations in the main report should be used only if they present the information more clearly and forcibly. Calculated 'confidence limits' must be mentioned and the various constraints experienced in conducting research operations may as well be stated.

1.15 SUMMARY

Research is a systematic investigative process employed to increase or revise current knowledge by discovering new facts.

Research can be defined as "defining and redefining problems, formulating hypothesis or suggested solutions, collecting, organizing and evaluating data, making deductions, reaching conclusions and testing the conclusions to determine whether they fit the formulating hypothesis.

Notes

Basic research is the research which is done for knowledge enhancement, the research which does not have immediate commercial potential. The research is done for human welfare, animal welfare and plant kingdom welfare. It is called basic, pure, fundamental research.

Descriptive research seeks to provide an accurate description of observations of a phenomenon. It is a fact finding investigation with adequate interpretation. It is the simplest type of research which focuses on particular aspects or dimensions of the problem studied. It is designed to gather descriptive information and provide information for formulating more sophisticated studies.

Exploratory research is a preliminary study of an unfamiliar problem about which the researcher has little or no knowledge. It involves a literature search or conducting focus group interviews.

Applied research is designed to solve practical problems of the modern world, rather than to acquire knowledge for knowledge sake. It is also known as action research. The goal of applied research is to improve the human condition.

Pure research advances fundamental knowledge about the human world. It focuses on refuting or supporting theories that explain how this world operates, what makes things happen, why social relations are a certain way and why society changes. Pure research is the source of most new scientific ideas and ways of thinking about the world. It can be exploratory, descriptive or explanatory.

Quantitative research is based on measurement of quantity or amount. It aims to measure the quantity or amount and compares it with past records and tries to project for future periods.

Qualitative research is concerned with qualitative phenomenon.It presents non-quantitative type of analysis. Qualitative research is collecting, analyzing and interpreting data by observing what people do and say. Qualitative research refers to the meanings, definitions, characteristics, symbols, metaphors and description of things.

Conceptual research is involves investigation of thoughts and ideas and developing new ideas or interpreting the old ones based on logical reasoning. A conceptual framework is used in research to outline possible courses of action or to present a preferred approach to an idea or thought.

Experimental research is designed to assess the effects of particular variables on a phenomenon by keeping the other variables constant or controlled. It aims at determining whether and in what manner variables are related to each other.

Deductive research is a logical process in which a conclusion is based on the concordance of multiple premises that are generally assumed to be true. Deductive research can be explained by the means of hypotheses, which can be derived from the propositions of the theory.

Inductive research works the opposite way, moving from specific observations to broader generalizations and theories. This is sometimes called a “bottom up” approach. The researcher begins with specific observations and measures, begins to then detect patterns and regularities, formulate some tentative hypotheses to explore and finally ends up developing some general conclusions or theories.

Historical research refers to the induction of principles through research into the past and social forces which have shaped the present. It is the process of systematically studying past records with a view to reconstruct the origin and development of an institution or a movement or a system and discovering trends in the past.

Research methodology is a way to systematically solve the research problem. It may be understood as a science of studying how research is done scientifically. In it we study the various steps that are generally adopted by a researcher in studying his research problem along with the logic behind them. It is necessary for the researcher to know not only the research methods/techniques but also the methodology.

Notes

1.16 SELF ASSESSMENT QUESTIONS

1. What are the objectives of research?
2. What is research? Explain its importance and limitations.
3. State the characteristics of research.
4. Discuss the scope of research.
5. Analyse descriptive approach to research.
6. Explain the various types of research.
7. Discuss the role of research in functional areas: Finance, Marketing and HRD.
8. Write short note on Research Methodology.
9. Explain the steps involved in research process.
10. What are the advantages and limitations of research?

*****

Chapter

2

RESEARCH PROBLEM

Objectives

The objectives of this lesson are to:

- Components of Research Problem
- Features of Research Problems
- Process of Formulating Hypothesis
- Concepts of Research Design
- Charateristics of Research Design
- Components of Research Design
- Concepts of Sampling
- Sampling Techniques
- Charateristics of a good Sampling Design
- Elements of Sampling Design

Structure:

Notes

2.1 INTRODUCTION

A research problem is a statement regarding an area of concern, a circumstance to be improved upon, a difficulty to be eliminated or a troubling question that exists in scholarly literature, in theory or in practice that point to the need for meaningful understanding and deliberate investigation. In some social science disciplines the research problem is typically posed in the form of one or more questions. A research problem does not state how to do something, offer a vague or broad proposition or present a value question.

Meaning of Research Problem

Research problem refers to the situation where a gap exists between the actual and the desired state. The problem can be generated either by an initiating idea or by a perceived problem area.

Example:

Investigation of 'rhythmic patterns in settlement planning' is the product of an idea that there are such things as rhythmic patterns in settlement plans, even if no one has detected them before. This kind of idea will then need to be formulated more precisely in order to develop it into a researchable problem. We are surrounded by problems connected with society, the built environment, education etc., many of which can readily be perceived.

2.2 DEFINING RESEARCH PROBLEM

A problem need not necessarily mean that something is wrong in the current situation which needs to be rectified immediately. It simply indicates an issue for which finding a solution could help to improve an existing situation. Problem can be defined as any situation where a gap exists between the actual and the desired state. Problem statement or problem definition refers to a clear, precise and succinct statement of question or issue that is to be investigated with the goal of finding an answer or solution.

Notes

Components of Research Problem

The components of research problem are as suggested by R. L. Ackoff in the "Design of Social Research" is elaborated below:

There must be an individual or a group which has some difficulty with problem:

(i) There must be some objective(s) to be attained at.

(ii) There must be alternative means or course of action for obtaining the objectives.

(iii) There must be some doubt in the minds of a researcher with regard to the selection of alternatives.

(iv) There must be some environment to which the difficulty pertains.

Criteria for Selecting the Research Problem

The following criteria can be kept in the minds of researchers in selecting the research problem:

(i) Subjects on which the research is carried on amply should not be normally chosen as there will not be a new dimension to reveal.

(ii) Too narrow or too vague problems should be avoided.

(iii) The researcher should be familiar with the subject chosen for research. The researcher should have enough knowledge, qualification and training in the selected problem area.

(iv) The resources needed to solve the problem in terms of time, money, efforts, manpower requirement should be taken into account before embarking on a problem.

(v) The subject of research should be familiar and feasible so that related research material or sources of research can be obtained easily.

(vi) The selection of a problem must be preceded by a preliminary study.

Research problems trigger the research process. Defining the research problem is a critical activity. A thorough understanding of research problem is a must for achieving success in the research endeavor. Defining the research problem begins with identifying the basic dilemma that prompts the research. It can be further developed by progressively breaking down the original dilemma into more specific and focus oriented objectives.

Five steps could be envisaged:

1. Identifying the broad problem area.
2. Literature review.
3. Identifying the research question.
4. Refining the research question.
5. Developing investigative questions.

2.3 COMPONENTS OF RESEARCH PROBLEM

The components of research problem are as suggested by R. L. Ackoff in the "Design of Social Research" is elaborated below:

(i) There must be an individual or a group which has some difficulty with problem.

(ii) There must be some objective(s) to be attained at.

(iii) There must be alternative means or course of action for obtaining the objectives.

Notes

(iv) There must be some doubt in the minds of a researcher with regard to the selection of alternatives.

(v) There must be some environment to which the difficulty pertains.

2.4 FEATURES OF RESEARCH PROBLEMS

Various important features of research problems are:

(i) It should be of great interest to the researcher. Researcher shall have to spend many months investigating the problem. A lively interest in the subject will be an invaluable incentive to persevere.

(ii) The problem should be significant. It is not worth time and effort investigating a trivial problem or repeating work that has already been done elsewhere.

(iii) The problem should be delineated. Consider the time researcher have to complete the work and the depth to which the problem will be addressed. Researcher can cover a wide field only superficially, and the more researchers restrict the field, the more detailed the study can be. Researcher should also consider the cost of necessary travel and other expenses.

(iv) Researcher should be able to obtain the information required. Researcher cannot carry out research if researcher fail to collect the relevant information needed to tackle researcher problem, either because researcher lack access to documents or other sources, and/or because researcher have not obtained the cooperation of individuals or organizations essential to the research.

(v) Researcher should be able to draw conclusions related to the problem. The point of asking a question is to find an answer. The problem should be one to which the research can offer some solution, or at least the elimination of some false 'solutions'.

(vi) Researcher should be able to state the problem clearly and concisely. A precise, well thought out and fully articulated sentence, understandable by anyone, should normally clearly be able to explain just what the problem is.

2.5 CRITERIA FOR SELECTING THE RESEARCH PROBLEM

The following criteria can be kept in the minds of researchers in selecting the research problem:

(i) Subjects on which the research is carried on amply should not be normally chosen as there will not be a new dimension to reveal.

(ii) Too narrow or too vague problems should be avoided.

(iii) The researcher should be familiar with the subject chosen for research. The researcher should have enough knowledge, qualification and training in the selected problem area.

(iv) The resources needed to solve the problem in terms of time, money, efforts, manpower requirement should be taken into account before embarking on a problem.

(v) The subject of research should be familiar and feasible so that related research material or sources of research can be obtained easily.

(vi) The selection of a problem must be preceded by a preliminary study.

Research problems trigger the research process. Defining the research problem is a critical activity. A thorough understanding of research problem is a must for achieving success in the research endeavor. Defining the research problem begins with identifying the basic dilemma that prompts the research. It

can be further developed by progressively breaking down the original dilemma into more specific and focus oriented objectives. *Notes*

Five steps could be envisaged:

1. Identifying the broad problem area
2. Literature review
3. Identifying the research question
4. Refining the research question
5. Developing investigative questions.

2.6 SOURCES OF PROBLEMS FOR RESEARCH

Indentifying a problem to study can be challenging, not because there's a lack of issues that could be investigated, but due to pursuing a goal of formulating an academically relevant and researchable problem that is unique and does not simply duplicate the work of others. To facilitate how you might select a problem from which to build a research study, consider these sources of inspiration:

(i) Deductions from Theory

This relates to deductions made from social philosophy or generalizations embodied in life in society that the researcher is familiar with. These deductions from human behavior are then fitted within an empirical frame of reference through research. From a theory, the researcher can formulate a research problem or hypothesis stating the expected findings in certain empirical situations. The research asks the question: "What relationship between variables will be observed if theory aptly summarizes the state of affairs?" One can then design and carry out a systematic investigation to assess whether empirical data confirm or reject the hypothesis, and hence, the theory.

(ii) Interdisciplinary Perspectives

Identifying a problem that forms the basis for a research study can come from academic movements and scholarship originating in disciplines outside of your primary area of study. A review of pertinent literature should include examining research from related disciplines that can reveal new avenues of exploration and analysis. An interdisciplinary approach to selecting a research problem offers an opportunity to construct a more comprehensive understanding of a very complex issue that any single discipline may be able to provide.

(iii) Interviewing Practitioners

The identification of research problems about particular topics can arise from formal or informal discussions with practitioners who provide insight into new directions for future research and how to make research findings more relevant to practice. Discussions with experts in the field, such as, teachers, social workers, health care providers, lawyers, business leaders, etc., offers the chance to identify practical, "real world" problems that may be understudied or ignored within academic circles. This approach also provides some practical knowledge which may help in the process of designing and conducting your study.

(iv) Personal Experience

The everyday experiences can give rise to worthwhile problems for investigation. Think critically about your own experiences and/or frustrations with an issue facing society, your community, your neighborhood, your family, or your personal life. This can be derived, for example, from deliberate

Notes observations of certain relationships for which there is no clear explanation or witnessing an event that appears harmful to a person or group or that is out of the ordinary.

(v) Relevant Literature

The selection of a research problem can be derived from an extensive and thorough review of pertinent research associated with your overall area of interest. This may reveal where gaps exist in our understanding of a topic. Research may be conducted to: 1) fill such gaps in knowledge; 2) evaluate if the methodologies employed in prior studies can be adapted to solve other problems; or 3) determine if a similar study could be conducted in a different subject area or applied to different study sample. Also, authors frequently conclude their studies by mentioning implications for further research; this can also be a valuable source of new problems to investigate.

2.7 HYPOTHESIS

Meaning of Hypothesis

Hypothesis testing refers to the formal procedures used by statisticians to accept or reject statistical hypotheses. It is an assumption about a population parameter. This assumption may or may not be true.

The best way to determine whether a statistical hypothesis is true would be to examine the entire population. Since that is often impractical, researchers typically examine a random sample from the population. If sample data are not consistent with the statistical hypothesis, the hypothesis is rejected.

In doing so, one has to take the help of certain assumptions or hypothetical values about the characteristics of the population if some such information is available. Such hypothesis about the population is termed as statistical hypothesis and the hypothesis is tested on the basis of sample values. The procedure enables one to decide on a certain hypothesis and test its significance. "A claim or hypothesis about the population parameters is known as Null Hypothesis and is written as, H_0."

This hypothesis is then tested with available evidence and a decision is made whether to accept this hypothesis or reject it. If this hypothesis is rejected, then we accept the alternate hypothesis. This hypothesis is written as H_1.

For testing hypothesis or test of significance we use both parametric tests and nonparametric or distribution free tests. Parametric tests assume within properties of the population, from which we draw samples. Such assumptions may be about population parameters, sample size etc. In case of non-parametric tests, we do not make such assumptions. Here we assume only nominal or ordinal data.

Definitions of Hypothesis

According to ***Kerlinger (1956),*** "A hypothesis is a conjectural statement of the relation between two or more variables".

According to ***Eric Rogers (1966),*** "Hypotheses are single tentative guesses, good hunches assumed for use in devising theory or planning experiments intended to be given a direct experimental test when possible".

According to ***Creswell (1994),*** "Hypothesis is a formal statement that presents the expected relationship between an independent and dependent variable."

Notes

2.8 CHARACTERISTICS OF GOOD HYPOTHESIS

Characteristics of a good Hypothesis can be summarized as follows:

(i) Simple to Understand

A hypothesis should be so dabble to every layman, P.V young says, "A hypothesis would be simple, if a researcher has more in sight towards the problem". W-ocean stated that, "A hypothesis should be as sharp as razor's blade". So, a good hypothesis must be simple and have no complexity.

(ii) Conceptually clear

A hypothesis must be conceptually clear. It should be clear from ambiguous information's. The terminology used in it must be clear and acceptable to everyone.

(iii) Testability

A good hypothesis should be tested empirically. It should be stated and formulated after verification and deep observation. Thus testability is the primary feature of a good hypothesis.

(iv) Relevant to Problem

If a hypothesis is relevant to a particular problem, it would be considered as good one. A hypothesis is guidance for the identification and solution of the problem, so it must be accordance to the problem.

(v) Power of Prediction

One of the valuable attribute of a good hypothesis is to predict for future. It not only clears the present problematic situation but also predict for the future that what would be happened in the coming time. So, hypothesis is a best guide of research activity due to power of prediction.

(vi) Closest to observable things

A hypothesis must have close contact with observable things. It does not believe on air castles but it is based on observation. Those things and objects which we cannot observe, for that hypothesis cannot be formulated. The verification of a hypothesis is based on observable things.

(vii) Specific Problem

It should be formulated for a particular and specific problem. It should not include generalization. If generalization exists, then a hypothesis cannot reach to the correct conclusions.

(viii) Relevant to available Techniques

Hypothesis must be relevant to the techniques which is available for testing. A researcher must know about the workable techniques before formulating a hypothesis.

(ix) Fruitful for new Discoveries

It should be able to provide new suggestions and ways of knowledge. It must create new discoveries of knowledge J.S. Mill, one of the eminent researcher says that "Hypothesis is the best source of new knowledge it creates new ways of discoveries".

(x) Consistency and Harmony

Internal harmony and consistency is a major characteristic of good hypothesis. It should be out of contradictions and conflicts. There must be a close relationship between variables which one is dependent on other.

Notes

2.9 TYPES OF HYPOTHESIS

(i) Descriptive Hypothesis

Descriptive hypothesis contains only one variable thereby it is also called as univariate hypothesis. Descriptive hypotheses typically state the existence, size, form or distribution of some variable. The first hypothesis contains only one variable. It only shows the distribution of the level of commitment among the officers of the organization which is higher than average. Such a hypothesis is an example of a Descriptive Hypothesis. Researchers usually uses research questions rather than descriptive hypothesis. *For example,* What is the level of commitment of officers in your organization?

(ii) Relational Hypothesis

These are the propositions that describe a relationship between two variables. The relationship could be non-directional or directional, positive or negative, causal or simply correlation. While stating the relationship between the two variables, if the terms of positive, negative, more than or less than is used then such hypotheses are directional because the direction of the relationship between the variables has been indicated. These hypotheses are relational as well as directional. The directional hypothesis is the one in which the direction of the relationship has been specified. The relationship may be very strong but whether it is positive or negative has not been postulated.

(iii) Correlational Hypothesis

These state merely that the variables occur together in some specified manner without implying that one causes the other. Such weak claims are often made when we believe that there are more basic causal forces that affect both variables.

For example, Level of job commitment of officers is positively associated with their level of efficiency. Here we do not make any claim that one variable causes the other to change. That will be possible only if we have control on all other factors that could influence our dependent variable.

(iv) Explanatory (Causal) Hypothesis

This Implies the existence of or a change in, one variable causes or leads to a change in the other variable. This brings in the notions of independent and the dependent variables. Cause means to "help make happen." So the independent variable may not be the sole reason for the existence of or change in the dependent variable. The researcher may have to identify the other possible causes and control their effect in case the causal effect of independent variable has to be determined on the dependent variable. This may be possible in an experimental design of research.

(v) Null Hypothesis

It is used for testing the hypothesis formulated by the researcher. Researchers treat evidence that supports a hypothesis differently from the evidence that opposes it. They give negative evidence more importance than to the positive one. It is because the negative evidence tarnishes the hypothesis. It shows that the predictions made by the hypothesis are wrong. The null hypothesis simply states that there is no relationship between the variables or the relationship between the variables is "zero."

It does not take into consideration the direction of association, which may be a second step in testing the hypothesis. First we look whether or not there is an association then we go for the direction of association and the strength of association. Experts recommend that we test our hypothesis indirectly by testing the null hypothesis. In case we have any credibility in our hypothesis then the research data should reject the null hypothesis. Rejection of the null hypothesis leads to the acceptance of the alternative hypothesis.

Notes

(vi) Alternative Hypothesis

The alternative (to the null) hypothesis simply states that there is a relationship between the variables under study. In our example it could be: There is a relationship between the level of job commitment and the level of efficiency. Not only there is an association between the two variables under study but also the relationship is perfect which is indicated by the number "1". There is a relationship between the level of job commitment of officers and their level of efficiency.

2.10 SOURCE OF HYPOTHESIS

The various sources of hypotheses may be:

(i) Previous Study

Previous study is also a source of developing a concrete hypothesis. If a researcher uses previous knowledge about a phenomenon for a particular place, then another researcher followed his techniques and formulates his own.

(ii) Personal Experience

On the basis of his personal experience he uses his mind and suggests some points for the eradication of a social problem through developing a good hypothesis. Greater the researcher experience lead to higher degree of formation.

(iii) Imagination and Thinking

Creative thinking and imagination of a researcher sometimes help in formulating a good hypothesis. Personal ideas and the thinking capabilities of a researcher would lead to greater number of hypothesis formulation as well as control over the problem.

(iv) Observation

In consideration and undertaking a research problem, observation is necessary. The collection of previous facts and current facts related to the problem lead to the formulation of a good hypothesis.

(v) Scientific Theory

Theory is capable in explaining all the facts relating to the problem. Scientific theory is a fertile source of hypothesis formulation. The theory which is used by a researcher may satisfy the needs of making it, because theory explains the known facts.

(vi) Culture

Culture is the accumulation of ways of behaving and adoption in a particular place and time. While formulating a hypothesis for a problem, culture should be studied. If we want to study trends towards female education in a particular area, for this purpose we will study, traditions, family system, Norms, Values, region and education system of that area.

2.11 PROCESS OF FORMULATING HYPOTHESIS

Formulating a hypothesis helps by defining an initial explanation to be tested in the research process. A proposition that can be verified to determine its reality is a hypothesis. Therefore one can say that a hypothesis is a verifiable counterpart of a proposition. A hypothesis may be defined as a logically conjectured relationship between two or more variables, expressed in the form of a testable statement. Relationship

Notes is proposed by using a strong logical argumentation. This logical relationship may be part of theoretical framework of the study.

Both quantitative and qualitative research involves formulating a hypothesis to address the research problem. Hypotheses that suggest a causal relationship involve at least one independent variable and at least one dependent variable; in other words, one variable which is presumed to affect the other. An independent variable is one whose value is manipulated by the researcher or experimenter. A dependent variable is a variable whose values are presumed to change as a result of changes in the independent variable.

Formulation of hypothesis is an assumption or suggested explanation about how two or more variables are related. It is a crucial step in the scientific method and, therefore, a vital aspect of all scientific research. There are no definitive guidelines for the production of new hypotheses. The history of science is filled with stories of scientists claiming a flash of inspiration or a hunch, which then motivated them to look for evidence to support or refute the idea.

1. State the null hypothesis as well as the alternate hypothesis

For example, let us assume the population mean = 50 and set up the hypothesis i = 50. This is called the null hypothesis and is denoted as;

Null hypothesis, $H_0 : \mu = 50$

Alternative hypothesis $H_1 : \mu = 50$

or $\mu > 50$

$\mu < 50$

2. Establish a level of significance

The level of significance signifies the probability of committing Type 1 error á and is generally taken as equal to 0.05. Sometimes, the value a is established as 0.01, but it is at the discretion of the investigator to select its value, depending upon the sensitivity of the study. To illustrate per cent level of significance indicates that a researcher is willing to take 5 per cent risk of rejecting the Null Hypothesis when it happens to be true.

3. Choosing a suitable test statistic

Now the researcher would choose amongst the various tests (i.e. z, t, χ^2 and f-tests). Actually, for the purpose of rejecting or accepting the null hypothesis, a suitable statistics called 'test statistics' is chosen. This means that H_0 is assumed to be really true. Obviously due to sampling fluctuations, the observed value of the statistic based on random sample will differ from the expected value. If the difference is large enough, one suspects the validity of the assumption and rejects the null hypothesis (H_0). On the other hand, if the difference may be assumed due to sampling fluctuation, the null hypothesis (H_0) is accepted.

4. Defining the critical rejection regions and making calculations for test statistics

If we select the value of a = Level of significance = 0.05, and use the standard normal distribution (z-test) as our test statistic for testing the population parameter u, then the value of the difference between the assumption of null hypothesis (assumed value of the population parameter) and the value obtained by the analysis of the sample results is not expected to be more than than 1.96 σ at $\alpha = 0.05$.

Notes

2.12 ERRORS IN HYPOTHESIS

In statistical hypothesis testing, type I and type II errors are incorrect rejection of a true null hypothesis or failure to reject a false null hypothesis, respectively. More simply stated, a type I error is detecting an effect that is not present, while a type II error is failing to detect an effect that is present. The terms "type I error" and "type II error" are often used interchangeably with the general notion of false positives and false negatives in binary classification, such as medical testing, but narrowly speaking refer specifically to statistical hypothesis testing.

T*wo types of errors can result from a hypothesis test:*

i) ***Type I error:*** A Type I error occurs when the researcher rejects a null hypothesis when it is true. The probability of committing a Type I error is called the significance level. This probability is also called alpha, and is often denoted by á.

ii) ***Type II error:*** A Type II error occurs when the researcher fails to reject a null hypothesis that is false. The probability of committing a Type II error is called Beta, and is often denoted by â. The probability of not committing a Type II error is called the Power of the test.

2.13 RESEARCH DESIGN

A research design is a master plan specifying the methods and procedures for collecting and analyzing the data. Research design is a blue print of action. It involves a series of rational decision making choices regarding the purpose of the study, its scope, its location, the type of investigation, the extent to which it is controlled and manipulated by the researcher, the time aspects, the collection, measurement and analysis of data. It is a plan and structure to obtain answers to the research questions. A research design 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. It aids the researcher in the allocation of resources in a well-defined manner. A research design helps to decide upon issues like what, when, where, how much, by what means etc. with regard to an enquiry or a research study. The more sophisticated and rigorous the research design is, the greater is the outcome of the research.

Broadly it is composed of different elements like: the purpose of the study, the unit of analysis, time dimension, mode of observation, sampling design, observation tools, data processing and data analysis. Specifically, the research design highlights decisions which include:

(i) The nature of the study.

(ii) The purpose of the study.

(iii) The location where the study would be conducted.

(iv) The nature of data required.

(v) From where the required data can be collected.

(vi) What time period the study would cover.

(vii) The type of sample design that would be used.

(viii) The techniques of data collection that would be used.

(ix) The methods of data analysis that would be adopted.

(x) The manner in which the report would be prepared.

Notes *In view of the stated research design decisions, the overall research design may be divided into the following:*

(a) The sampling design that deals with the method of selecting items to be observed for the selected study.

(b) The observational design that relates to the conditions under which the observations are to be made.

(c) The statistical design that concerns with the question of how many items are to be observed and how the information and data gathered are to be analysed.

(d) The operational design that deals with the techniques by which the procedures specified in the sampling, statistical and observational designs can be carried out.

2.14 MEANING OF RESEARCH DESIGN

Research design is a plan of action indicating the specific steps that are necessary to provide answers to those questions, test the hypotheses and thereby achieve the research purpose that helps choose among the decision alternatives to solve the management problem or capitalize on the market opportunity.

2.15 DEFINITIONS OF RESEARCH DESIGN

According to ***David J. Luck and Ronald S. Rubin,*** "A research design is the determination and statement of the general research approach or strategy adopted/or the particular project. It is the heart of planning. If the design adheres to the research objective, it will ensure that the client's needs will be served."

According to ***Kerlinger,*** "Research design is the plan, structure and strategy of investigation conceived so as to obtain answers to research questions and to control variance."

According to ***Green and Tull,*** "A research design is the specification of methods and procedures for acquiring the information needed. It is the over-all operational pattern or framework of the project that stipulates what information is to be collected from which source by what procedures."

2.16 FEATURES OF RESEARCH DESIGN

i) ***Objectivity:*** It refers to the findings related to the method of data collection and scoring of the responses. The research design should permit the measuring instrument which is fairly objective in which every observer or judge scoring the performance must precisely give the same report. In other words, the objectivity of the procedure may be judged by the degree of agreement between the final scores assigned to different individuals by more than one independent observer. This ensures the objectivity of the collected data which shall be capable of analysis and drawing generalizations.

ii) ***Reliability:*** Reliability refers to consistency throughout a series of measurements. *For example*: if a respondent gives out a response to a particular item, he is expected to give the same response to that item even if he is asked repeatedly. If he is changing his response to the same item, the consistency will be lost. So the researcher should frame the items in a questionnaire in such a way that it provides consistency or reliability.

Notes

iii) ***Validity:*** Any measuring device or instrument is said to be valid when it measures what it is expected to measure. *For example:* an intelligence test conducted for measuring the I.Q should measure only the intelligence and nothing else and the questionnaire shall be framed accordingly.

iv) ***Generalizability:*** It means how best the data collected from the samples can be utilized for drawing certain generalizations applicable to a large group from which sample is drawn. Thus a research design helps an investigator to generalize his findings provided he has taken due care in defining the population, selecting the sample, deriving appropriate statistical analysis etc. while preparing the research design.

2.17 CHARACTERISTICS OF A GOOD RESEARCH DESIGN

A good research design often possesses the qualities of being flexible, suitable, efficient, economical and so on. Generally, a research design which minimizes bias and maximizes the reliability of the data collected and analysed is considered a good design.

Following are certain characteristic elements of a good research design:

a) ***Simplicity:*** It should be simple and understandable.

b) ***Economical:*** It must be economical. The technique selected must be cost effective and less time-consuming.

c) ***Reliability:*** It should give the smallest experimental error. This should have the minimum bias and have the reliability of data collected and analysed.

d) ***Workability:*** It must be workable. It should be pragmatic and practicable.

e) ***Flexibility:*** It must be flexible enough to permit the consideration of many different aspects of a phenomenon.

f) ***Accuracy:*** It must lead to accurate results.

2.18 NATURE OF RESEARCH DESIGN

A research design is a plan for comprehensive data collection to answer research questions and/or test research hypotheses. It consists of detailed prescriptions for solving problems from either a scientific or humanistic perspective.

This is a systematic plan to study a scientific problem. The design of a study defines the study type (descriptive, correlation, semi-experimental, experimental, review, meta-analytic) and sub-type (e.g., descriptive-longitudinal case study), research question, hypotheses, independent and dependent variables, experimental design and if applicable, data collection methods and a statistical analysis plan. Research design is the framework that has been created to seek answers to research questions.

Explanatory research focuses on why questions. *For example,* it is one thing to describe the crime rate in a country, to examine trends over time or to compare the rates in different countries. It is quite a different thing to develop explanations about why the crime rate is as high as it is why some types of crime are increasing or why the rate is higher in some countries than in others. The way in which researchers develop research designs is fundamentally affected by whether the research question is descriptive or explanatory. It affects what information is collected.

Research design is different from the method by which data are collected. Many research methods texts confuse research designs with methods. It is not uncommon to see research design treated as a mode of data collection rather than as a logical structure of the inquiry. But there is nothing intrinsic

Notes about any research design that requires a particular method of data collection. Although cross-sectional surveys are frequently equated with questionnaires and case studies are often equated with participant observation, data for any design can be collected with any data collection method.

Exploratory research is conducted when the researcher does not know how and why certain phenomenon occurs. Here, the hypothetical solutions or actions are explored and evaluated by the decision-maker, *example* evaluation of quality of service of a bank/hotel/airline. Here, the quality cannot be accessed directly as tangible features are not available.

Descriptive research is undertaken when the researcher desires to know the characteristics of certain groups such as age, gender, occupation, income or education. The objective of descriptive research is to answer the "who, what, when, where and how" of the subject under study/investigation.

2.19 CONCEPTS OF RESEARCH DESIGN

Some of the important concepts relating to Research Design are discussed below:

1. Dependent and Independent Variables

A magnitude that varies is known as a variable. The concept may assume different quantitative values like height, weight, income etc. Qualitative variables are not quantifiable in a strict sense. However, the qualitative phenomena may also be quantified in terms of the presence or absence of the attribute considered. The phenomena that assume different values quantitatively even in decimal points are known as 'continuous variables'. But all variables need not be continuous. Values that can be expressed only in integer values are called 'non-continuous variables'. In statistical terms, they are also known as 'discrete variables'. *For example,* age is a continuous variable, whereas the number of children is a non-continuous variable. When changes in one variable depend upon the changes in other variable or variables, it is known as a dependent or endogenous variable and the variables that cause the changes in the dependent variable are known as the independent or explanatory or exogenous variables. *For example,* if demand depends upon price, then demand is a dependent variable, while price is the independent variable. And, if more variables determine demand, like income and price of the substitute commodity, then demand also depends upon them in addition to the price of original commodity. In other words, demand is a dependent variable which is determined by the independent variables like price of the original commodity, income and price of substitutes.

2. Extraneous Variables

The independent variables which are not directly related to the purpose of the study but affect the dependent variable are known as extraneous variables. For instance, assume that a researcher wants to test the hypothesis that there is a relationship between employee's performance and their self-concepts, in which case the latter is an independent variable and the former, a dependent variable. In this context, intelligence may also influence the employee's performance. However, since it is not directly related to the purpose of the study undertaken by the researcher, it would be known as an extraneous variable. The influence caused by the extraneous variable(s) on the dependent variable is technically called the 'experimental error'. Therefore, a research study should always be framed in such a manner that the influence of extraneous variables on the dependent variable/s is completely controlled, and the influence of independent variable/s is clearly evident.

3. Control

One of the most important features of a good research design is to minimize the effect of extraneous variable(s). Technically, the term 'control' is used when a researcher design the study in such a manner that it minimizes the effects of extraneous variables. The term 'control' is used in experimental research to reflect the restraints used in experimental conditions.

Notes

4. Confounded Relationship

The relationship between the dependent and independent variables is said to be confounded by an extraneous variable, when the dependent variable is not free from its effects.

5. Research Hypothesis

When a prediction or a hypothesized relationship is tested by adopting scientific methods, it is known as research hypothesis. The research hypothesis is a predictive statement which relates to a dependent variable and an independent variable. Generally, a research hypothesis must consist of at least one dependent variable and one independent variable. Whereas, the relationships that are assumed but not to be tested are predictive statements that are not to be objectively verified, thus are not classified as research hypotheses.

6. Experimental and Non-experimental Hypothesis Testing Research

When the objective of a research is to test a research hypothesis, it is known as hypothesis-testing research. Such research may be in the nature of experimental design or non-experimental design. The research in which the independent variable is manipulated is known as 'experimental hypothesis-testing research', whereas the research in which the independent variable is not manipulated is termed as 'non-experimental hypothesis-testing research'. *For example,* assume that a researcher wants to examine whether family responsibilities influences the employee absentism by calculating the coefficient of correlation between the two variables. Such an *example* is known as a non-experimental hypothesis-testing research, because the independent variable - family responsibilities is not manipulated here. Again assume that the researcher randomly selects 150 female employees from a group of students who are very regular and then classifies them into two sub-groups by randomly including 75 in Group A, who have people to take household responsibilities and 75 in group B, who do not have any helping hand. Assume that at the end of the study, the researcher conducts a test on each group in order to examine the effects of supporting people to take household of the responsibility on the absentism of the employees. Such a study is an example of experimental hypothesis-testing research, because in this particular study the independent variable regular earnings of the parents have been manipulated.

7. Experimental and Control Groups

When a group is exposed to usual conditions in an experimental hypothesis-testing research, it is known as 'control group'. On the other hand, when the group is exposed to certain new or special condition, it is known as an 'experimental group'. In the afore-mentioned example, Group A can be called as control group and Group B as experimental group. If both the groups, A and B are exposed to some special feature, then both the groups may be called as 'experimental groups'. A research design may include only the experimental group or both the experimental and control groups together.

8. Treatments

Treatments refer to the different conditions to which the experimental and control groups are subject to. In the example considered, the two treatments are the employees with people support and take responsibility at home and those with no helping hand. Likewise, if a research study attempts to examine through an experiment the comparative effect of three different types of fertilizers on the yield of rice crop, then the three types of fertilizers would be treated as the three treatments.

9. Experiment

Experiment refers to the process of verifying the truth of a statistical hypothesis relating to a given research problem. For instance, an experiment may be conducted to examine the yield of a certain new variety of rice crop developed. Further, Experiments may be categorized into two types, namely, 'absolute experiment' and 'comparative experiment'. If a researcher wishes to determine the impact of a chemical

Notes fertilizer on the yield of a particular variety of rice crop, then it is known as absolute experiment. Meanwhile, if the researcher wishes to determine the impact of chemical fertilizer as compared to the impact of bio-fertilizer, then the experiment is known as a comparative experiment.

10. Experimental Units

Experimental Units refer to the pre-determined plots, characteristics or the blocks, to which different treatments are applied. It is worth mentioning here that such experimental units must be selected with great caution.

2.20 PROCESS OF RESEARCH DESIGN PREPARATION

Stated in simple terms, a research design is a plan of action, a plan for collecting and analyzing data in an economic, efficient and relevant manner. Whatever be the nature of design, the following steps are generally followed:

Step - 1: Selection and Definition of a problem: The problem selected for study should be defined clearly in operational terms so that researcher knows positively what facts he is looking for and that is relevant to the study.

Step - 2: Sources of Data: Once the problem is selected, it is the duty of the researcher to state clearly the various sources of information such as library, personal documents, field work, a particular residential group etc.

Step - 3: Nature of Study: The research design should be expressed in relation to the nature of study to be undertaken. The choice of the statistical, experimental or comparative type of study should be made at this stage so that the following steps in planning may have relevance to the proposed problem.

Step - 4: Object of Study: Whether the design aims at theoretical understanding or presupposes a welfare notion must be explicit at this point. Stating the object of the study helps not only in clarity of the design but also in a sincere response from the respondents.

Step - 5: Social-Cultural Context: The research design must be set in the social-cultural context. *For example:* In a study of the fertility rate of people among backward class, the context of the so-called backward class of people and the conceptual reference must be made clear. Unless the meaning of the term is clearly defined there tends to be a large variation in the study because the term backward could have religious, economic and political connotations.

Step - 6: Temporal context: The geographical limit of the design should also be referred to at this stage as research related to the hypothesis is applicable to particular social group only.

Step - 7: Dimension: It is physically impossible to analyze the data collected from a large universe. Hence the selection of an adequate and representative sample is a by-word in any research.

Step - 8: Basis of Selection: The mechanics of drawing a random, stratified and purposive, double cluster or quota sample when followed carefully will produce a scientifically valid sample in an unbiased manner.

Step -9: Technique of Data Collection: Relevant to the study design a suitable technique has to be adopted for the collection of required data. The relative merit of observation, interview and questionnaire, when studied together will help in the choice of suitable technique. Once the collection of data is complete, analysis, coding and presentation of the report naturally follow.

Notes

2.21 TYPES OF RESEARCH DESIGN

The different types of Research Design are:

I. Exploratory Research Design

Exploratory research is conducted when the researcher does not know how and why certain phenomenon occurs. Here, the hypothetical solutions or actions are explored and evaluated by the decision-maker, e.g. evaluation of quality of service of a bank/hotel/airline. Here, the quality cannot be accessed directly as tangible features are not available.

It is appropriate when the research objective is to provide insights into:

(i) Identifying the problems or opportunities.

(ii) Defining the problem more precisely.

(iii) Gaining deeper insights into the variables operating in a situation.

(iv) Identifying relevant courses of action.

(v) Establishing priorities regarding the potential significance of problems and opportunities.

(vi) Gaining additional insights before an approach can be developed.

(vii) Gathering information on the problems associated with doing conclusive research.

Exploratory research could also be used in conjunction with other research. Since it is used as a first step in the research process, defining the problem, other designs will be used later as steps to solve the problem. For instance, it could be used in situations when a firm finds the going gets tough in terms of sales volume, the researcher may use exploratory research to develop probable explanations. Analysis of data generated using exploratory research is essentially abstraction and generalization. The exploratory research design is best characterized by its flexibility and versatility. This is so, because of the absence of the non-imperativeness of a structure in its design.

Exploratory Research is used:

(i) To define the problem more precisely

(ii) To identify relevant courses of action i.e. find the most likely alternatives, which are then turned into hypotheses.

(iii) Isolate key variables and relationships for further examinations.

(iv) Gain insights for developing an approach to a problem.

(v) Establish priorities for further research.

II. Conclusive Research Design

Conclusive Research Design is typically more formal and structured than exploratory research. It is based on large representative samples and the market information obtained is subjected to quantitative analysis. Conclusive Research is designed to assist the decision maker in determining, evaluating and selecting the best course of action to take in a given situation.

It involves providing information on evaluation of alternative courses of action and selecting one from among a number available to the researcher. Conclusive research is again classified as:

(i) Descriptive research and

(ii) Causal research.

Notes

(i) Descriptive Research

Descriptive research is undertaken when the researcher desires to know the characteristics of certain groups such as age, gender, occupation, income or education. The objective of descriptive research is to answer the "who, what, when, where and how" of the subject under study/investigation. It is simple to understand as the name itself suggests that it involves describing something, *for example:*

(a) Market conditions.

(b) Characteristics or functions.

(c) Estimating the percentage of customers in a particular group exhibiting the same purchase behaviour.

(d) Perceptions of product characteristics and

(e) To predict the pattern of behaviourial characteristics of consumers.

Majority of research studies are descriptive studies. As research studies involve investigating the customers/consumers, collection of data includes interrogating the respondents in the market and data available from secondary data sources. However, it cannot be concluded that descriptive studies should be simply a fact-gathering process. Descriptive study deals with the respondents in the market and hence, extreme caution has to be exercised in developing this study. Much planning should be done, objectives should be clearer than in exploratory studies. In descriptive research, the data is collected for a specific and definite purpose and involves analysis and interpretation by the researcher. The major difference between exploratory and descriptive research is that descriptive research is characterised by the formulation of specific objectives. The success of descriptive studies depends on the degree to which a specific hypothesis acts as a guide. Descriptive studies restrict flexibility and versatility as compared to exploratory research. It involves a higher degree of formal design specifying the methods for selecting the sources of information and for collecting data from those sources. Formal design is required in order to ensure that the description covers all phases desired. It is also required to restrain collection of unnecessary data. Descriptive studies require a clear specification of the who, when, where, what, why and how. While designing a descriptive research, the researcher should also have sufficient knowledge on the nature and type of statistical techniques he/she is going to use. This will greatly help to have the right design in place. Mostly descriptive studies are conducted using questionnaire, structured interviews and observations. The results of descriptive studies are directly used for marketing decisions.

Uses of Descriptive Research

Descriptive research is conducted for the following reasons:

1. To describe the characteristics of relevant groups, such as consumers, salespeople, or organizations or market areas.
2. To estimate the percentage of units in a specified population exhibiting a certain behavior *example:* the percentage of heavy users of prestigious department stores who also patronize discount department stores.
3. To determine the “perceptions of product characteristics. *For example:* how do households perceive the various department stores in terms of salient factors of the choice criteria?
4. To determine the degree to which marketing variables are associated. *For example:* to what extent is shopping at department stores related to eating out?
5. To collect demographic information of consumers/users of a product under study.
6. To discover the relationship between certain variables, *example:* sale of toothpaste among rural population and urban population or rate of savings among low, middle and higher income groups.

Descriptive studies are again classified into two types: *Notes*

(a) Longitudinal

(b) Cross sectional

(a) Longitudinal Research relies on panel data and panel methods. It involves fixing a panel consisting of fixed sample of subjects that are measured repeatedly. The panel members are those who have agreed to provide information at specific intervals over an extended period. *For example,* data obtained from panels formed to provide information on market shares are based on an extended period of time but also allow the researcher to examine changes in market share over time. New members may be included in the panel as and when there is a dropout of the existing members or to maintain representativeness.

(b) Cross-sectional Research is the most predominantly and frequently used descriptive research design in marketing. It involves a sample of elements from the population of interest. The sample elements are measured on a number of characteristics. Cross-sectional study is a study involving a sample of elements from the population of interest at a single point of time. It is a study concerned with a sample of elements from a given population.

(ii) Causal Research

Casual research design is the third type of research design. As the name indicates, casual design investigates the cause and effect relationship between two or more variables. This design measures the extent of relationship between the variables. Casual research designs attempt to specify the nature of functional relationship between two or more variables.

It is used to obtain evidence of cause-and-effect relationships which is otherwise known as the independent-dependent relationship or the predictive relationship. Causal research requires a strong degree of planning on the design as its success depends on the structure of the design. This is an important type of research useful for marketers as this allows marketers to base their decisions on assumed causal relationships. Causal research is done in the following situations;

(a) To identify which variables are the cause and which are the effect. In statistical terms causal variables are called independent variables and effectual variables are called dependent variables.

(b) To determine the nature of the relationship between the causal variables and the effect to be predicted.

The casual research design is based on reasoning. The designs for casual research can be divided into three categories: (a) Historical, (b) Survey and (c) Experimental.

III. Experimental Research Design

Experimental research studies generally require testing of hypothesis for causal relationship amongst the variables. Naturally, these types of research studies require procedures that should not only reduce the bias but also lead to inferences about causality. This leads to necessity for experimental designs. Experimental design develops a framework of experiments based on thumb rule or statistical procedures.

Need for Experimental Research Design

To achieve the goal of process optimization, to prevent, or to minimize the occurrence of defective product, a thorough understanding of the process behavior under different sets of process conditions is needed. Planning an experiment so that conformation relevant to the problem on hand will be made available is known as: "Designing and Experiment". Experience has shown that if the data collection is properly planned, organized, summarized and interpreted using statistical principles, one will be able to draw valid and meaningful conclusions from the results. The design of experiment was found to be an excellent tool of effecting engineering development, quality improvement, process optimization as well

Notes as cost reduction. In general, planned experimentation is necessary to distinguish between critical factors effect and need to be controlled within the narrow limits and non critical factors which are insignificant and do not require close control as well as to identify the optimum levels of the critical factors so as to achieve significantly improved performance.

Basic Principles of Experimental Research Design

i) ***Principle of Replication:*** Under this principle emphasis is on doing the same experiment more than once. Researcher applies each treatment in many experimental units instead of one. By doing so he increases the statistical accuracy. *For example,* It can get a more precise effect of the mean effect of any factor.

ii) ***Principle of Randomization:*** The principle of randomization provides researcher protection against the effect of extraneous factor, when he undertakes any experiment. It provides the freedom of designing and planning the experiment in such a fashion that variations, caused by extraneous factors can all be combined together and termed as chance. The basic idea is to compare all treatment effects within a block of experimental material by eliminating environmental effects. Randomization procedure is done with the help of random number table by the following steps:

a) Open the page of the table randomly.

b) Select the column of numbers on that page randomly.

c) Numbers in that column will be used in order to determine the order or rows of the columns to be chosen.

d) Extra numbers will be omitted.

Types of Experimental Designs

Experimental design is the basic framework or structure of an experiment on which the whole research work is focused. There are two broad classification of experimental designs: formal experimental designs and informal experimental designs. The formal experimental designs offer the researcher more control and use of precise statistical procedures for analysis of the study where as informal experimental designs normally use less sophisticated form of statistical procedures for analysis. *The important experimental designs are as follows:*

1. Informal Experimental Designs

a) Before-and-without Control Design

In such an experimental design, a set of single test group is selected and the dependent variable is measured prior to application of a specific treatment. Subsequently treatment is introduced and dependent variable is again measured. Therefore the interpretation would be that treatment produced the delta (Ä) difference in the outcome of dependent variable. An example of this can be say to observe the level of bacteria in a public swimming pool, prior and after the chlorination treatment. The main difficulty in such a design is that there could be other extraneous variations while the treatment is being introduced. If we continue with the above example, it can so happen that while chlorination treatment is being applied there is a rain fall, which adds air borne bacteria with rain water into the swimming pool.

b) After-only with Control Design

In this type of experimental design, two areas viz, test area and control area, are selected. In such a design, the treatment is applied only to the test area. The dependent variable is measured in both the areas at the same time. This leads to possible elimination of extraneous variations. The impact of treatment is assessed by subtracting the value of dependent variable in the control area from the value obtained in the test area.

Notes

For example, there are two adjacent fields of a farmer of equal size. In one field, fertilizer is put and in the other field no fertilizer is applied. After one month, the growth of crop is measured in both the fields. So, it can be deduced that, fertilizer leads to increase by 3 cm if the average height of crop is 12 cm in test field and 9 cm in control field. Other extraneous factors such as water, rain fall, and climatic conditions are common to both. Therefore, it can be said that this experiment design is superior to before and after, without control design.

c) Before-and-after with Control Design

This design in a way, is an improvement on the first design and also combines control features of the second design. In this experimental design two areas are selected and dependent variable is measured in both for common time period prior to the treatment. Then, the treatment is applied only in the test area and the dependent variable is measured again in both the test and control areas for an identical time period after the introduction of treatment. The impact of treatment is determined by subtracting the delta change in the dependent variable obtained in the control area from the delta change achieved in the dependent variable in the test area. This design is superior to earlier two design because not only it avoids the extraneous variations but also the variations of non-comparability of the test and control areas.

2. Formal Experimental Designs

a) Completely Randomized Design (CR Design)

This type of design involves the principle of replication and principle of randomization. In a sense this is the easiest possible experimental design and therefore the procedure of analysis is also simpler. The basic characteristics of a completely randomized design is that subjects are randomly assigned to experimental treatments. *For example,* if we have 8 patients and we wish to give medication to four, on the basis of treatment A and other four under treatment B the Randomization process provides the possible opportunity that the group of four patients be selected from a set of eight and being treated by treatment A and treatment B. Analysis procedure required to analyze such design is called one way analysis of variance. This design provides the greatest number of degrees of freedom to the error. Normally this design is used when experimental areas are homogenous. Strictly speaking when all possible variation due to uncontrollable experimental factors is included under chance variation, the design of experiment is known as completely randomized design.

Advantages of Completely Randomized Design

This design has following advantages:

a) Complete flexibility is possible. The number of replications can be varied at will from treatment to treatment. It is possible to utilize all the experimental data.

b) Statistical analysis is easy even if number of replications are not same for all treatments.

c) The analysis remains simple even when results from some units or treatments are rejected. The relative loss of information due to such rejection is smallest compared with any other design.

b) Randomized Block Design (RB Design)

This is the most familiar and a very important design among all experimental designs. Apart from completely randomized design, it is the simplest design to construct and analysis is known as randomized block design. The term randomized block emanated from agronomic research wherein several variables or treatments are applied to different blocks of land to study the effect of replication on experimental effort, such as, yield of different types of sugarcane by using variable amounts of water to irrigate the fields. However, difference in sugarcane yield may not be attributed only to the different strains of sugar

Notes cane but also to difference in fertility of soil in the various blocks of lands. To remove the block effect; randomization is obtained by providing treatments at random to blocks of land. In such cases, blocks are formed in a way and each contains as many plots as there are treatments to be experimented with. And one plot from each is randomly selected for each treatment. The scheme is easily understood by looking at it, as a field planning of an agronomic experiment. The randomized block design is widely used in many types of business research experiments. For example to determine the difference in output of various types of machines, we may be able to isolate the effect due to difference in efficiencies of works by assigning machines at random to randomly selected workers. The underlying idea in this kind of experiment is to compare the effect of all treatments within a block of experimental set up by eliminating possible environmental effects. By comparing mean square of treatments by the means square of remainder it can be determined by F test whether the treatments have any effect, regardless of the fact of possibility of a significant variation from block to block.

Advantages of Completely Randomized Experimental Design

Such a design offers the following major advantages:

a) It is very easy to plan out the design.

b) It provides a great degree of flexibility because any number of factors, types and replications may be used.

c) Analysis of such a design through statistical methods is rather simple. This is so, even in cases when a number of replications for each factor type or if the experimental errors are not similar from type to type of this factor.

d) Even when data are missing or rejected, the method or analysis is quite simple in completely randomized block design. The loss of information due to missing data is limited as compared to any other experimental design.

Major drawback of this design is that it is suited when the number of treatments is small and experimental conditions are homogenous. When the number of treatments is larger, it is possible to select designs which are more efficient than the completely randomized design. Therefore, randomized designs are rarely used for field experiments where numbers of treatments are relatively larger.

c) Latin Square Design (LS Design)

This experimental design also emerged out of agronomic experimentations and is extensively used where there is a need to eliminate the trend of soil fertility in two directions simultaneously. In such a design data is classified in rows and columns according to different treatments and varieties and is organized in the form of a square which is called a Latin Space.

The genesis of the term "Latin Square" came from a mathematical puzzle that was devised many years before such experiments came into being. In such a design, since there have to be as many replications as are treatments, the domain of experiment is divided into slots organized in a square in a manner that they are as many slots in each row as there are in each column. This number is also same as the number of treatments. These slots are then assigned to various treatments in a manner for each treatments occurs only once in each row and only once in each columns. This can be organized in a large number of ways. However, particular way in which any particular layout is done must be determined randomly.

The major advantages of Latin Square Experimental Design over other such designs are:

a) The two way stratification of latin square design leads a better control of the variation than the completely randomized design or the randomised block design.

b) The two way stratification leads to elimination of variation which often results in a small error mean square.

Notes

c) By and large, analysis is still simple, however, it may be slightly more complex than an analysis for randomised block design

d) Analysis remains relatively simple with latin square design even if some of the data are missing. There are procedures available to analyse latin squares in cases one wishes to omit one or more treatments, rows or columns.

A major drawback of latin square design may be that number of the treatments must be equal to the number of treatments of the rows and columns. Also when number of treatment is more than seven, latin square design hardly is ever utilized, due to complexity in number of permutations and combinations.

d) Factorial design

In recent times with a view to improve rational foundation of a scientific experimentation, the factorial design has proved to be one of the useful developments. Factorial experiments allow the researcher to evaluate the combined effect of two or more variables when used simultaneously. It is considered that information obtained from.

Factorial experiments is more complete than that which is obtained from a set of single factor experiments. This is due to that fact that factorial experiments allow the evaluation of interaction effects. An interaction effect is generally attributed in two or more combination of variables over and above those that can be predicted from the variables if considered alone.

Major reasons for including several factors in one experiment are:

a) Understanding the overall effect of the factors economically by conducting one single experiment of moderate size.

b) To enlarge basis of inference on a single factor by testing it under graded conditions of other factors.

c) Find out the manner in which the effect of factors interacts with one another. These may not be entirely independent but emphasis can be made to vary with a degree of experimentation.

Limitations of Experimental Designs

Following are the difficulties faced by a researcher in case of experimental designs:

1. ***Problems in experiment setting:*** Generally it is not easy to determine the conditions under which experiments should be set up. In case of scientific experiments laboratory conditions may be established but this may not be possible in case of special science experiments.
2. ***Problems in getting cooperation:*** In case of business and social research, obtaining cooperation from people who form the subject of experimentation is not easy. Human subjects at times work according to their free will. A lack of interest also at times makes cooperation impossible.
3. ***Difficulties in establishing control:*** Control at times in an experimental situation is more complex in compension to the case of complex business and socio-economic research is lost since it is very difficult to get complete knowledge of various factors influencing the experiments.
4. ***Problems of consciousness:*** In case of business experimental design, experimental subject is rather fluid and possesses a consciousness which limits the degree of experimentation.

2.22 COMPONENTS OF RESEARCH DESIGN

The primary intent behind the research design is to help avoid the situation in which the evidence doesn't address the primary research questions. A research design is concerned with a logical problem and not a logistical problem. Five major components of research design are:

Notes

1. Research study's questions
2. Study propositions
3. Unit(s) of analysis
4. Linking data to propositions
5. Interpreting a study's findings

1. Research study's questions

This first component suggests the type of the question-in terms of "who," "what," "where," "how," and "why"-provides an crucial clue concerning the most relevant research method to be used. Use three stages: In the first, make use of the literature to narrow your interest to a key topic or two. In the 2nd, take a look at closely-even dissect-a few key studies on your topic of interest. Identify the questions in those few studies and whether they conclude with new questions for future research. In the last phase, examine another group of scientific studies on the same topic. They might provide support for your potential questions or even suggest means of sharpening them.

2. Study propositions

Each proposition directs focus on something which needs to be examined within the scope of study. Only if you are forced to state some propositions will you move in the right direction. *For example,* you may think that businesses collaborate as they gain mutual benefits. This proposition, apart from highlighting a crucial theoretical issue (that other incentives for collaboration don't exist or are unimportant), also starts to tell you where you can search for related proof (to define and determine the extent of certain advantages to each business).

3. Unit of analysis

It is associated with the fundamental problem of defining what the "case" is-a problem which has affected many researchers at the beginning of case studies. Take example of clinical patients. In this situation, an individual is being studied and the individual is the key unit of analysis. Information regarding the appropriate individual will be collected, and several such individuals could be part of a multiple-case study. You would need study questions and propositions to help find out the appropriate information to be collected relating to this individual or individuals. Without such questions and propositions, you could be lured to cover "everything" with regards to the individual(s), which is not possible.

4. Linking data to propositions

Methods of linking data to propositions are pattern matching, explanation building, time-series analysis, logic models and cross-case synthesis. The actual analyses will demand that you merge or compute your study data as a direct reflection of your initial study propositions.

5. Interpreting a study's findings

A statistical analysis determines if the results of the study support the hypothesis. A number of statistical tests, for example T-tests (that determine if two groups are statistically distinct from one another), Chi-square tests (where data are compared to an anticipated outcome) and one-way analysis of variance (provides for the comparison of multiple groups), are carried out according to the type of data, number and types of variables and data categories. Statistical analysis offer some explicit criteria for interpretations. *For example,* by convention, social science views a p level of less than .05 to indicate that observed differences were "statistically important." On the other hand, much case study analysis is not going to depend on the use of statistics and so focuses on other methods of thinking about such criteria.

Notes

2.23 SAMPLING

Sampling is an important concept which is practiced in every activity. Sampling involves selecting a relatively small number of elements from a large defined group of elements and expecting that the information gathered from the small group will allow judgments to be made about the large group. The basic idea of sampling is that by selecting some of the elements in a population, the conclusion about the entire population is drawn. Sampling is used when conducting census is impossible or unreasonable. In a census method a researcher collects primary data from every member of a defined target population. It is not always possible or necessary to collect data from every unit of the population. The researcher can resort to sample survey to find answers to the research questions. However, they can do more harm than good if the data is not collected from the people, events or objects that can provide correct answers to the problem. The process of selecting the right individuals, objects or events for the purpose of the study is known as sampling

Meaning of Sampling

Sampling is defined as the selection of some part of an aggregate or totality on the basis of which a judgment or inference about the aggregate or totality is made. Sampling is the process of learning about the population on the basis of a sample drawn from it.

Purpose of Sampling

There are several reasons for sampling. They are explained below:

(i) ***Lower cost:*** The cost of conducting a study based on a sample is much lesser than the cost of conducting the census study.

(ii) ***Greater accuracy of results:*** It is generally argued that the quality of a study is often better with sampling data than with a census. Research findings also substantiate this opinion.

(iii) ***Greater speed of data collection:*** Speed of execution of data collection is higher with the sample. It also reduces the time between the recognition of a need for information and the availability of that information.

(iv) ***Availability of population element:*** Some situations require sampling. When the breaking strength of materials is to be tested, it has to be destroyed. A census method cannot be resorted as it would mean complete destruction of all materials. Sampling is the only process possible if the population is infinite.

Essentials of Sampling

In order to reach a clear conclusion, the sampling should possess the 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 result, 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 the size of the sample is larger, there is better efficiency and at the same time the cost is more. A proper size of sample is maintained in order to have optimized results in terms of cost and efficiency.

Notes

Features of Sampling

The sampling technique has the following good features of value and significance:

1. ***Economy:*** Sampling technique brings about cost control of a research project as it requires much less physical resources as well as time than the census technique.
2. ***Reliability:*** In sampling technique, if due diligence is exercised in the choice of sample unit and if the research topic is homogenous then the sample survey can have almost the same reliability as that of census survey.
3. ***Detailed Study:*** An intensive and detailed study of sample units can be done since their number is fairly small. Also multiple approaches can be applied to a sample for an intensive analysis.
4. ***Scientific Base:*** As mentioned earlier this technique is of scientific nature as the underlined theory is based on principle of statistics.
5. ***Greater Suitability in most Situations:*** It has a wide applicability in most situations as the examination of few sample units normally suffices.
6. ***Accuracy:*** The accuracy is determined by the extent to which bias is eliminated from the sampling. When the sample elements are drawn properly some sample elements underestimates the population values being studied and others overestimate them.

Steps in Developing a Sampling Plan

A number of concepts, procedures and decisions must be considered by a researcher in order to successfully gather raw data from a relatively small group of people which in turn can be used to generalize or make predications about all the elements in a larger target population.

The following are the logical steps involved in the sample execution:

1) Define the target population

The first task of a researcher is to determine and identify the complete group of people or objects that should be included in the study. With the statement of the problem and the objectives of the study acting as guideline the target population should be identified on the basis of descriptors that represent the characteristics features of element that make the target population's frame. These elements become the prospective sampling unit from which a sample will be drawn. A clear understanding of the target population will enable the researcher to successfully draw a representative sample.

2) Select the data collection method

Based on the problem definition, the data requirements and the research objectives, the researcher should select a data collection method for collecting the required data from the target population elements. The method of data collection guides the researcher in identifying and securing the necessary sampling frame for conducting the research.

3) Identify the sampling frames needed

The researcher should identify and assemble a list of eligible sampling units. The list should contain enough information about each prospective sampling unit so as to enable the researcher to contact them. Drawing an incomplete frame decreases the likelihood of drawing a representative sample.

4) Select the appropriate sampling method

The researcher can choose between probability and non-probability sampling methods. Using a probability sampling method will always yield better and more accurate information about the target population's parameters than the non-probability sampling methods. Seven factors should be considered in deciding the appropriateness of the sampling method viz., research objectives, degree of desired

accuracy, availability of resources, time frame, advanced knowledge of the target population, scope of the research and perceived statistical analysis needs. *Notes*

5) Determine necessary sample sizes and overall contact rates

The sample size is decided based on the precision required from the sample estimates, time and money available to collect the required data. While determining the sample size due consideration should be given to the variability of the population characteristic under investigation, the level of confidence desired in the estimates and the degree of the precision desired in estimating the population characteristic. The number of prospective units to be contacted to ensure that the estimated sample size is obtained and the additional cost involved should be considered. The researcher should calculate the reachable rates, overall incidence rate and expected completion rates associated with the sampling situation.

6) Creating an operating plan for selecting sampling units

The actual procedure to be used in contacting each of the prospective respondents selected to form the sample should be clearly laid out. The instruction should be clearly written so that interviewers know what exactly should be done and the procedure to be followed in case of problems encountered, in contacting the prospective respondents.

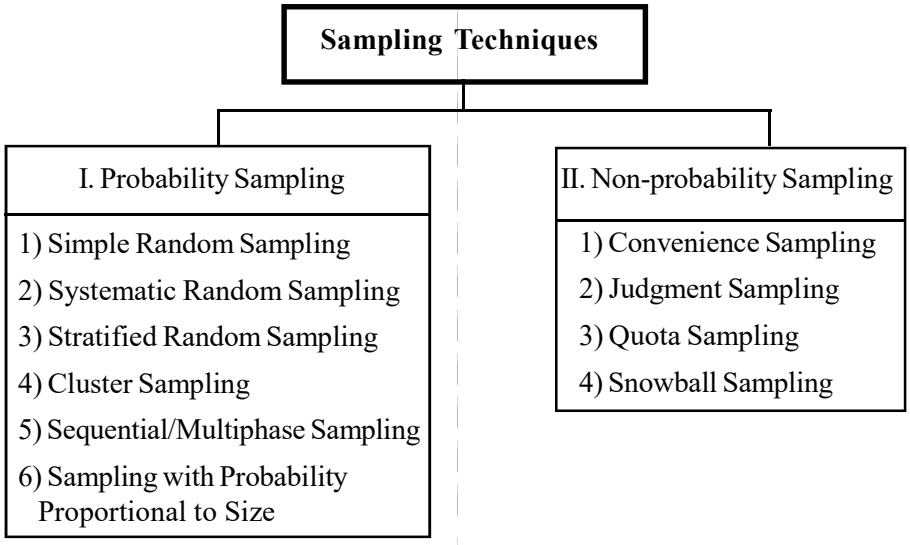
7) Executing the operational plan

The sample respondents are met and actual data collection activities are executed in this stage. Consistency and control should be maintained at this stage.

2.24 SAMPLING TECHNIQUES

The sampling design can be broadly grouped on two basis viz., representation and element selection. Representation refers to the selection of members on a probability or by other means. Element selection refers to the manner in which the elements are selected individually and directly from the population. If each element is drawn individually from the population at large, it is an unrestricted sample. Restricted sampling is where additional controls are imposed, in other words it covers all other forms of sampling.

The classification of sampling design on the basis of representation and element selection is shown below:



Notes

I. Probability Sampling

Probability sampling is where each sampling unit in the defined target population has a known non-zero probability of being selected in the sample. The actual probability of selection for each sampling unit may or may not be equal depending on the type of probability sampling design used. Specific rules for selecting members from the operational population are made to ensure unbiased selection of the sampling units and proper sample representation of the defined target population. The results obtained by using probability sampling designs can be generalized to the target population within a specified margin of error.

Probability samples are characterised by the fact that, the sampling units are selected by chance. In such a case, each member of the population has a known, non-zero probability of being selected. However, it may not be true that all samples would have the same probability of selection, but it is possible to say the probability of selecting any particular sample of a given size. It is possible that one can calculate the probability that any given population element would be included in the sample. This requires a precise definition of the target population as well as the sampling frame. Probability sampling techniques differ in terms of sampling efficiency which is a concept that refers to trade off between sampling cost and precision. Precision refers to the level of uncertainty about the characteristics being measured. Precision is inversely related to sampling errors but directly related to cost. The greater the precision, the greater the cost and there should be a tradeoff between sampling cost and precision. The researcher is required to design the most efficient sampling design in order to increase the efficiency of the sampling.

The different types of probability sampling designs are discussed below:

1) Simple Random Sampling

The following are the implications of random sampling:

(i) It provides each element in the population an equal probability chance of being chosen in the sample, with all choices being independent of one another and

(ii) It offers each possible sample combination an equal probability opportunity of being selected.

In the unrestricted probability sampling design every element in the population has a known, equal non-zero chance of being selected as a subject. *For example,* if 10 employees (n = 10) are to be selected from 30 employees (N = 30), the researcher can write the name of each employee in a piece of paper and select them on a random basis. Each employee will have an equal known probability of selection for a sample. The same is expressed in terms of the following formula:

Probability of selection = Size of sample / Size of population

Each employee would have a 10/30 or .333 chance of being randomly selected in a drawn sample. When the defined target population consists of a larger number of sampling units, a more sophisticated method can be used to randomly draw the necessary sample. A table of random numbers can be used for this purpose. The table of random numbers contains a list of randomly generated numbers. The numbers can be randomly generated through the computer programs also. Using the random numbers the sample can be selected.

Advantages and Disadvantages

The simple random sampling technique can be easily understood and the survey result can be generalized to the defined target population with a pre specified margin of error. It also enables the researcher to gain unbiased estimates of the population's characteristics. The method guarantees that every sampling unit of the population has a known and equal chance of being selected, irrespective of the actual size of the sample resulting in a valid representation of the defined target population.

Notes

The major drawback of the simple random sampling is the difficulty of obtaining complete, current and accurate listing of the target population elements. Simple random sampling process requires all sampling units to be identified which would be cumbersome and expensive in case of a large population. Hence, this method is most suitable for a small population.

2) Systematic Random Sampling

The systematic random sampling design is similar to simple random sampling but requires that the defined target population should be selected in some way. It involves drawing every nth element in the population starting with a randomly chosen element between 1 and n. In other words individual sampling units are selected according their position using a skip interval. The skip interval is determined by dividing the sample size into population size. *For example,* if the researcher wants a sample of 100 to be drawn from a defined target population of 1000, the skip interval would be 10(1000/100). Once the skip interval is calculated, the researcher would randomly select a starting point and take every 10th until the entire target population is proceeded through. The steps to be followed in a systematic sampling method are enumerated below:

(i) Total number of elements in the population should be identified

(ii) The sampling ratio is to be calculated (n = total population size divided by size of the desired sample)

(iii) A sample can be drawn by choosing every nth entry

Two important considerations in using the systematic random sampling are:

(i) It is important that the natural order of the defined target population list be unrelated to the characteristic being studied.

(ii) Skip interval should not correspond to the systematic change in the target population.

Advantages and Disadvantages

The major advantage is its simplicity and flexibility. In case of systematic sampling there is no need to number the entries in a large personnel file before drawing a sample. The availability of lists and shorter time required to draw a sample compared to random sampling makes systematic sampling an attractive, economical method for researchers.

The greatest weakness of systematic random sampling is the potential for the hidden patterns in the data that are not found by the researcher. This could result in a sample not truly representative of the target population. Another difficulty is that the researcher must know exactly how many sampling units make up the defined target population. In situations where the target population is extremely large or unknown, identifying the true number of units is difficult and the estimates may not be accurate.

3) Stratified Random Sampling

Stratified random sampling requires the separation of defined target population into different groups called strata and the selection of sample from each stratum. Stratified random sampling is very useful when the divisions of target population are skewed or when extremes are present in the probability distribution of the target population elements of interest. The goal in stratification is to minimize the variability within each stratum and maximize the difference between strata. The ideal stratification would be based on the primary variable under study. Researchers often have several important variables about which they want to draw conclusions. A reasonable approach is to identify some basis for stratification that correlates well with other major variables. It might be a single variable like age, income etc. or a compound variable like on the basis of income and gender. Stratification leads to segmenting the population into smaller, more homogeneous sets of elements. In order to ensure that the sample maintains the required precision in terms of representing the total population, representative

Notes samples must be drawn from each of the smaller population groups.

There are three reasons as to why a researcher chooses a stratified random sample:

(i) To increase the sample's statistical efficiency

(ii) To provide adequate data for analyzing various sub populations

(iii) To enable different research methods and procedures to be used in different strata.

Drawing a stratified random sampling involves the following steps:

1. Determine the variables to use for stratification
2. Select proportionate or disproportionate stratification
3. Divide the target population into homogeneous subgroups or strata
4. Select random samples from each stratum
5. Combine the samples from each stratum into a single sample of the target population.

There are two common methods for deriving samples from the strata viz., proportionate and disproportionate. In proportionate stratified sampling, each stratum is properly represented so the sample drawn from it is proportionate to the stratum's share of the total population. The larger strata are sampled more because they make up a larger percentage of the target population. This approach is more popular than any other stratified sampling procedures due to the following reasons:

(i) It has higher statistical efficiency than the simple random sample

(ii) It is much easier to carry out than other stratifying methods

(iii) It provides a self-weighing sample i.e., the population mean or proportion can be estimated simply by calculating the mean or proportion of all sample cases.

In disproportionate stratified sampling, the sample size selected from each stratum is independent of that stratum's proportion of the total defined target population. This approach is used when stratification of the target population produces sample sizes that contradict their relative importance to the study. An alternative of disproportionate stratified method is optimal allocation. In this method, consideration is given to the relative size of the stratum as well as the variability within the stratum to determine the necessary sample size of each stratum. The logic underlying the optimal allocation is that the greater the homogeneity of the prospective sampling units within a particular stratum, the fewer the units that would have to be selected to estimate the true population parameter accurately for that subgroup. This method is also opted for in situation where it is easier, simpler and less expensive to collect data from one or more strata than from others. Stratified random sampling provides several advantages viz., the assurance of representativeness in the sample, the opportunity to study each stratum and make relative comparisons between strata and the ability to make estimates for the target population with the expectation of greater precision or less error.

4) Cluster Sampling

Cluster sampling is a probability sampling method in which the sampling units are divided into mutually exclusive and collectively exhaustive subpopulation called clusters. Each cluster is assumed to be the representative of the heterogeneity of the target population. Groups of elements that would have heterogeneity among the members within each group are chosen for study in cluster sampling. Several groups with intragroup heterogeneity and intergroup homogeneity are found. A random sampling of the clusters or groups is done and information is gathered from each of the members in the randomly chosen clusters. Cluster sampling offers more of heterogeneity within groups and more homogeneity among the groups.

Notes

Single Stage and Multistage Cluster Sampling

In single stage cluster sampling, the population is divided into convenient clusters and required number of clusters are randomly chosen as sample subjects. Each element in each of the randomly chosen cluster is investigated in the study. Cluster sampling can also be done in several stages which is known as multistage cluster sampling. For example: To study the banking behaviour of customers in a national survey, cluster sampling can be used to select the urban, semi-urban and rural geographical locations of the study. At the next stage, particular areas in each of the location would be chosen. At the third stage, the banks within each area would be chosen. Thus multi-stage sampling involves a probability sampling of the primary sampling units; from each of the primary units, a probability sampling of the secondary sampling units is drawn; a third level of probability sampling is done from each of these secondary units, and so on until the final stage of breakdown for the sample units are arrived at, where every member of the unit will be a sample.

Area Sampling

Area sampling is a form of cluster sampling in which the clusters are formed by geographic designations. For example, state, district, city, town etc., Area sampling is a form of cluster sampling in which any geographic unit with identifiable boundaries can be used. Area sampling is less expensive than most other probability designs and is not dependent on population frame. A city map showing blocks of the city would be adequate information to allow a researcher to take a sample of the blocks and obtain data from the residents therein.

Advantages and Disadvantages of Cluster Sampling

The cluster sampling method is widely used due to its overall cost-effectiveness and feasibility of implementation. In many situations the only reliable sampling unit frame available to researchers and representative of the defined target population, is one that describes and lists clusters. The list of geographical regions, telephone exchanges, or blocks of residential dwelling can normally be easily compiled than the list of all the individual sampling units making up the target population. Clustering method is a cost efficient way of sampling and collecting raw data from a defined target population.

One major drawback of clustering method is the tendency of the cluster to be homogeneous. The greater the homogeneity of the cluster, the less precise will be the sample estimate in representing the target population parameters. The conditions of intra-cluster heterogeneity and inter-cluster homogeneity are often not met. For these reasons this method is not practiced often.

5) Sequential/Multiphase Sampling

This is also called Double Sampling. Double sampling is opted when further information is needed from a subset of groups from which some information has already been collected for the same study. It is called as double sampling because initially a sample is used in the study to collect some preliminary information of interest and later a sub-sample of this primary sample is used to examine the matter in more detail The process includes collecting data from a sample using a previously defined technique. Based on this information, a sub sample is selected for further study. It is more convenient and economical to collect some information by sampling and then use this information as the basis for selecting a sub sample for further study.

6) Sampling with Probability Proportional to Size

When the case of cluster sampling units does not have exactly or approximately the same number of elements, it is better for the researcher to adopt a random selection process, where the probability of inclusion of each cluster in the sample tends to be proportional to the size of the cluster. For this, the number of elements in each cluster has to be listed, irrespective of the method used for ordering it. Then the researcher should systematically pick the required number of elements from the cumulative totals.

Notes The actual numbers thus chosen would not however reflect the individual elements, but would indicate as to which cluster and how many from them are to be chosen by using simple random sampling or systematic sampling. The outcome of such sampling is equivalent to that of simple random sample. This method is also less cumbersome and is also relatively less expensive.

II. Non-Probability Sampling

In non probability sampling method, the elements in the population do not have any probabilities attached to being chosen as sample subjects. This means that the findings of the study cannot be generalized to the population. However, at times the researcher may be less concerned about generalizability and the purpose may be just to obtain some preliminary information in a quick and inexpensive way. Sometimes when the population size is unknown, then non probability sampling would be the only way to obtain data. Some non-probability sampling techniques may be more dependable than others and could often lead to important information with regard to the population.

Non-probability sampling does not involve random selection. It involves personal judgement of the researcher rather than chance to select sample elements. Sometimes this judgement is imposed by the researcher, while in other cases the selection of population elements to be included is left to the individual field workers. The decision maker may also contribute to including a particular individual in the sampling frame. Evidently, non probability sampling does not include elements selected probabilistically and hence, leaves a degree of sampling error associated with the sample. Sampling error is the degree to which a sample might differ from the population. Therefore, while inferring to the population, results could not be reported plus or minus the sampling error. In non-probability sampling, the degree to which the sample differs from the population remains

Unknown However, we cannot come to a conclusion that sampling error is an inherent of non probability sample. on-probability samples also yield good estimates of the population characteristics. Since, inclusion of the elements in the sample are not determined in a probabilistic way, the estimates obtained are not statistically projectable to the population.

1) Convenience Sampling

Non-probability samples that are unrestricted are called convenient sampling. Convenience sampling refers to the collection of information from members of population who are conveniently available to provide it. Researchers or field workers have the freedom to choose as samples whomever they find, thus it is named as convenience. It is mostly used during the exploratory phase of a research project and it is the best way of getting some basic information quickly and efficiently. The assumption is that the target population is homogeneous and the individuals selected as samples are similar to the overall defined target population with regard to the characteristics being studied. However, in reality there is no way to accurately assess the representativeness of the sample. Due to the self selection and voluntary nature of participation in data collection process the researcher should give due consideration to the non-response error.

Advantages and Disadvantages

Convenient sampling allows a large number of respondents to be interviewed in a relatively short time. This is one of the main reasons for using convenient sampling in the early stages of research. However the major drawback is that the use of convenience samples in the development phases of constructs and scale measurements can have a serious negative impact on the overall reliability and validity of those measures and instruments used to collect raw data. Another major drawback is that the raw data and results are not generalizable to the defined target population with any measure of precision. It is not possible to measure the representativeness of the sample, because sampling error estimates cannot be accurately determined.

Notes

2) Judgment Sampling

Judgment sampling is a non-probability sampling method in which participants are selected according to an experienced individual's belief that they will meet the requirements of the study. The researcher selects sample members who conform to some criterion. It is appropriate in the early stages of an exploratory study and involves the choice of subjects who are most advantageously placed or in the best position to provide the information required. This is used when a limited number or category of people have the information that are being sought. The underlying assumption is that the researcher's belief that the opinions of a group of perceived experts on the topic of interest are representative of the entire target population.

Advantages and Disadvantages

If the judgment of the researcher or expert is correct then the sample generated from the judgment sampling will be much better than one generated by convenience sampling. However, as in the case of all non-probability sampling methods, the representativeness of the sample cannot be measured. The raw data and information collected through judgment sampling provides only a preliminary insight.

3) Quota Sampling

The quota sampling method involves the selection of prospective participants according to pre specified quotas regarding either the demographic characteristics (gender, age, education, income, occupation etc.,) specific attitudes (satisfied, neutral, dissatisfied) or specific behaviours (regular, occasional, rare user of product). The purpose of quota sampling is to provide an assurance that pre specified subgroups of the defined target population are represented on pertinent sampling factors that are determined by the researcher. It ensures that certain groups are adequately represented in the study through the assignment of the quota.

Advantages and Disadvantages

The greatest advantage of quota sampling is that the sample generated contains specific subgroups in the proportion desired by researchers. In those research projects that require interviews the use of quotas ensures that the appropriate subgroups are identified and included in the survey. The quota sampling method may eliminate or reduce selection bias.

An inherent limitation of quota sampling is that the success of the study will be dependent on subjective decisions made by the researchers. As a non-probability method, it is incapable of measuring true representativeness of the sample or accuracy of the estimate obtained. Therefore, attempts to generalize the data results beyond those respondents who were sampled and interviewed become very questionable and may misrepresent the given target population.

4) Snowball Sampling

Snowball sampling is a non-probability sampling method in which a set of respondents are chosen who help the researcher to identify additional respondents to be included in the study. This method of sampling is also called as referral sampling because one respondent refers other potential respondents. This method involves probability and non-probability methods. The initial respondents are chosen by a random method and the subsequent respondents are chosen by non-probability methods. Snowball sampling is typically used in research situations where the defined target population is very small and unique and compiling a complete list of sampling units is a nearly impossible task. This technique is widely used in academic research. While the traditional probability and other non-probability sampling methods would normally require an extreme search effort to qualify a sufficient number of prospective respondents, the snowball method would yield better result at a much lower cost. The researcher has to identify and interview one qualified respondent and then solicit his help to identify other respondents with similar characteristics.

Notes

Advantages and Disadvantages

Snowball sampling enables to identify and select prospective respondents who are small in number, hard to reach and uniquely defined target population. It is most useful in qualitative research practices. Reduced sample size and costs are the primary advantage

of this sampling method. The major drawback is that the chance of bias is higher. If there is a significant difference between people who are identified through snowball sampling and others who are not then, it may give rise to problems. The results cannot be generalized to members of larger defined target population.

2.25 SAMPLING DESIGN

A sample design is a definite plan for obtaining a sample from a given population Sample constitutes a certain portion of the population or universe. Sampling design refers to the technique or the procedure the researcher adopts for selecting items for the sample from the population or universe. A sample design helps to decide the number of items to be included in the sample, i.e., the size of the sample. The sample design should be determined prior to data collection. There are different kinds of sample designs which a researcher can choose. Some of them are relatively more precise and easier to adopt than the others. A researcher should prepare or select a sample design, which must be reliable and suitable for the research study proposed to be undertaken.

Every research study requires the selection of some kind of sample. It is the life blood of research. Any research study aims to obtain information about the characteristics or parameters of a population. A population is the aggregate of all the elements that share some common set of characteristics and that comprise the universe for the purpose of the research problem. In other words, population is defined as the totality of all cases that conform to some designated specifications. The specification helps the researcher to define the elements that ought to be included and to be excluded. Sometimes, groups that are of, interest to the researcher may be significantly smaller allowing the researcher to collect data from all the elements of population. Collection of data from the entire population is referred to as census study. A census involves a complete enumeration of the elements of a population. Collecting data from the aggregate of all the elements in case of, the number of elements being larger, would sometimes render the researcher incur huge costs and time. Hence, Sampling is the process of selecting units (e.g., people, organizations) from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen. While deciding on the sampling, the researcher should clearly define the target population without allowing any kind of ambiguity and inconsistency on the boundary of the aggregate set of respondents.

2.26 CHARACTERISTICS OF A GOOD SAMPLE DESIGN

The following are the characteristic features of a good sample design:

(a) The sample design should yield a truly representative sample.

(b) The sample design should be such that it results in small sampling error.

(c) The sample design should be viable in the context of budgetary constraints of the research study.

(d) The sample design should be such that the systematic bias can be controlled.

(e) The sample must be such that the results of the sample study would be applicable, in general, to the universe at a reasonable level of confidence.

Notes

2.27 ELEMENTS OF SAMPLING DESIGN

A researcher should take into consideration the following aspects while developing a sample design:

(i) Type of Universe

The first step involved in developing sample design is to clearly define the number of cases, technically known as the Universe, to be studied. A universe may be finite or infinite. In a finite universe the number of items is certain, whereas in the case of an infinite universe the number of items is infinite (i.e., there is no idea about the total number of items). *For example,* while the population of a city or the number of workers in a factory comprise finite universes, the number of stars in the sky or throwing of a dice represent infinite universe.

(ii) Sampling Unit

Prior to selecting a sample a decision has to be made about the sampling unit. A sampling unit may be a geographical area like a state, district, village etc. or a social unit like a family, religious community, school, etc. or it may also be an individual. At times, the researcher would have to choose one or more of such units for his/her study.

(iii) Source List

Source list is also known as the 'sampling frame', from which the sample is to be selected. The source list consists of names of all the items of a universe. The researcher has to prepare a source list when it is not available. The source list must be reliable, comprehensive, correct and appropriate. It is important that the source list should be as representative of the population as possible.

(iv) Size of the Sample

Size of the sample refers to the number of items to be chosen from the universe to form a sample. The size of sample must be optimum. An optimum sample may be defined as the one that satisfies the requirements of representativeness, flexibility, efficiency, and reliability. While deciding the size of sample a researcher should determine the desired precision and the acceptable confidence level for the estimate. The size of the population variance should be considered, because in the case of a larger variance generally a larger sample is required. The size of the population should be considered, as it also limits the sample size. The parameters of interest in a research study should also be considered, while deciding the sample size. Besides, costs or budgetary constraint also plays a crucial role in deciding the sample size.

(a) ***Parameters of Interest:*** The specific population parameters of interest should also be considered while determining the sample design. *For example,* the researcher may want to make an estimate of the proportion of persons with certain characteristics in the population, or may be interested in knowing some average regarding the population. The population may also consist of important sub-groups about whom the researcher would like to make estimates. All such factors have strong impact on the sample design the researcher selects.

(b) ***Budgetary Constraint:*** From the practical point of view, cost considerations exercise a major influence on the decisions related to not only the sample size, but also on the type of sample selected. Thus, budgetary constraint could also lead to the adoption of a non-probability sample design.

(c) ***Sampling Procedure:*** Finally, the researcher should decide the type of sample or the technique to be adopted for selecting the items for a sample. This technique or procedure itself may represent the sample design. There are different sample designs from which a researcher should select one for his/her study. It is clear that the researcher should select that design which, for a given sample size and budget constraint, involves a smaller error.

Notes

2.28 DETERMINATION OF APPROPRIATE SAMPLING DESIGN

Determining an appropriate sampling design is a challenging issue and has greater implications on the application of the research findings. Apart from considering the theoretical components, sampling issues, advantages and drawbacks of different sampling techniques, the decision should take into consideration the following factors:

1. Research Objectives

A clear understanding of the statement of the problem and the objectives will provide the initial guidelines for determining the appropriate sampling design. If the research objectives include the need to generalize the findings of the research study, then a probability sampling method should be opted rather than a non-probability sampling method. In addition the type of research viz., exploratory or descriptive will also influence the type of the sampling design.

2. Scope of the Research

The scope of the research project is local, regional, national or international has an implication on the choice of the sampling method. The geographical proximity of the defined target population elements will influence not only the researcher's ability to compile needed list of sampling units, but also the selection design. When the target population is equally distributed geographically a cluster sampling method may become more attractive than other available methods. If the geographical area to be covered is more extensive then complex sampling method should be adopted to ensure proper representation of the target population.

3. Availability of Resources

The researchers command over the financial and human resources should be considered in deciding the sampling method. If the financial and human resource availability are limited, some of the more time-consuming, complex probability sampling methods cannot be selected for the study.

4. Time Frame

The researcher who has to meet a short deadline will be more likely to select a simple, less time consuming sampling method rather than a more complex and accurate method.

5. Advanced Knowledge of the Target Population

If the complete lists of the entire population elements are not available to the researcher, the possibility of the probability sampling method is ruled out. It may dictate that a preliminary study be conducted to generate information to build a sampling frame for the study. The researcher must gain a strong understanding of the key descriptor factors that make up the true members of any target population.

6. Degree of Accuracy

The degree of accuracy required or the level of tolerance for error may vary from one study to another. If the researcher wants to make predictions or inferences about the 'true' position of all members of the defined target population, then some type of probability sampling method should be selected. If the researcher aims to solely identify and obtain preliminary insights into the defined target population, non-probability methods might prove to be more appropriate.

7. Perceived Statistical Analysis needs

The need for statistical projections or estimates based on the sample results is to be considered. Only probability sampling techniques allow the researcher to adequately use statistical analysis for estimates beyond the sample respondents. Though the statistical method can be applied on the non-

probability samples of people and objects, the researcher's ability to accurately generalize the results and findings to the larger defined target population is technically inappropriate and questionable. The researcher should also decide on the appropriateness of sample size as it has a direct impact on the data quality, statistical precision and generalization of findings.

Notes

2.29 SUMMARY

Research problem refers to the situation where a gap exists between the actual and the desired state. The problem can be generated either by an initiating idea or by a perceived problem area.

Hypothesis testing refers to the formal procedures used by statisticians to accept or reject statistical hypotheses. It is an assumption about a population parameter. This assumption may or may not be true.

Descriptive hypothesis contains only one variable thereby it is also called as univariate hypothesis. Descriptive hypotheses typically state the existence, size, form or distribution of some variable.

Hypothesis test is a method of making decisions using data from a scientific study. In statistics, a result is called statistically significant if it has been predicted as unlikely to have occurred by chance alone, according to a pre-determined threshold probability, the significance level.

Research design is a plan of action indicating the specific steps that are necessary to provide answers to those questions, test the hypotheses and thereby achieve the research purpose that helps choose among the decision alternatives to solve the management problem or capitalize on the market opportunity.

Casual research design is the third type of research design. As the name indicates, casual design investigates the cause and effect relationship between two or more variables. This design measures the extent of relationship between the variables. Casual research designs attempt to specify the nature of functional relationship between two or more variables.

Experimental research studies generally require testing of hypothesis for causal relationship amongst the variables. Naturally, these types of research studies require procedures that should not only reduce the bias but also lead to inferences about causality.

Sampling is defined as the selection of some part of an aggregate or totality on the basis of which a judgment or inference about the aggregate or totality is made. Sampling is the process of learning about the population on the basis of a sample drawn from it.

Probability sampling is where each sampling unit in the defined target population has a known non-zero probability of being selected in the sample. The actual probability of selection for each sampling unit may or may not be equal depending on the type of probability sampling design used.

A sample design is a definite plan for obtaining a sample from a given population Sample constitutes a certain portion of the population or universe. Sampling design refers to the technique or the procedure the researcher adopts for selecting items for the sample from the population or universe.

2.30 SELF ASSESSMENT QUESTIONS

1. What is research problem? Discuss selection of a research problem.
2. Discuss formulation of a research problem.
3. What do you mean by hypothesis? Explain the process of formulation of hypothesis.
4. Explain need for research design.

Notes

5. What is research design? Explain the nature and importance of research design.
6. What are the essential features of good research design?
7. Discuss various components of a research design?
8. Explain the content of research design.
9. Discuss various types of research design.
10. List the factors affecting choice of research design.
11. Discuss suitability collection of exploratory research.
12. Explain different types of descriptive research.
13. Explain the types of experimental design.
14. Discuss advantages and limitations of research design.
15. What is sampling? Discuss various merits and demerits of sampling.
16. Explain the various steps involved in sampling process.
17. Explain various technique of sampling.
18. What is sampling design? Explain the charateristics of a good sample design.
19. Explain the elements of sample design.
20. Discuss the determination of appropriate sampling design.

*****

Chapter

3

COLLECTION, PROCESSING AND ANALYSIS OF DATA

Objectives

The objectives of this lesson are to:

- Concepts of Data
- Collection of Data
- Methods of Data Collection
- Analysis of Data
- Design of Questionnaire
- Testing of Hypothesis
- Parametric and Non-parametric Tests
- T-test and Z-test
- Chi-square test

Structure:

Notes

3.1 DATA

Meaning of Data

Data is the facts in raw or unorganized form such as alphabets, numbers or symbols that refer to or represent conditions, ideas or objects. This represents facts and statistics which are collected together for reference or analysis.

Characteristics of Data

In order that numerical description may be called data, they must possess the following characteristics:

i) ***Data is aggregate of facts:*** *For example,* single unconnected figures can not be used to study the characteristics of a business activity.

ii) ***Data is affected to a large extent by multiplicity of factors:*** *For example,* in business environment the observations recorded are affected by a number of factors (controllable and uncontrollable).

iii) ***Data is estimated according to reasonable standard of accuracy:*** *For example,* in the measurement of length one may measure correct upto 0.01 of a cm., the quality of the product is estimated by certain tests on small samples drawn from big lots of products.

iv) ***Data is collected in a systematic manner for a predetermined objective:*** Facts collected in a haphazard manner and without a complete awareness of the objective will be confusing and can not be made the basis of valid conclusions. *For example,* collected data on price serves no purpose unless one knows whether he wants to collect data on wholesale or retail prices and what are the relevant commodities under considerations.

v) ***Data must be related to one another:*** The data collected should be comparable, otherwise these can not be placed in relation to each other, *example:* data on the yield of crop and quality of soil are related but the crop yields cannot have any relation with the data on the health of the people.

vi) ***Data must be numerically expressed:*** That is, any facts to be called data must be numerically or quantitatively expressed. Qualitative characteristics such as beauty, intelligence etc. are called attributes and must be scaled to express in numeric terms.

Sources of Data

Data sources can be broadly categorized into three types viz., primary, secondary and tertiary.

1. Primary Data Sources

Primary data refers to information gathered firsthand by the researcher for the specific purpose of the study. It is raw data without interpretation and represents the personal or official opinion or position. Primary sources are most authoritative since the information is not filtered or tampered. Some examples of the sources of primary data are individuals, focus groups, panel of respondents. Data collection from individuals can be made through interviews, observation etc.

2. Secondary Data Sources

Secondary data refers to the information gathered from already existing sources. Secondary data may be either published or unpublished data. The published data are available in the following forms:

(i) Publications of central, state and local governments.

(ii) Publications of foreign governments, international bodies and their subsidiary organizations.

(iii) Technical and trade journals.

Notes

(iv) Books, magazines and newspapers.

(v) Reports and publications of various business and industrial associations, stock exchanges, banks and other financial institutions.

(vi) Reports prepared by research scholars, universities, economists in different fields.

(vii) Public records and statistics, historical documents and other sources of published information.

(viii) Online and real time databases etc.

The unpublished sources include the company records or archives, diaries, letters, biographies and autobiographies and other public/private organizations.

3. Tertiary Sources

Tertiary sources are an interpretation of a secondary source. It is generally represented by index, bibliographies, dictionaries, encyclopedias, handbooks, directories and other finding aids like the internet search engines.

3.2 COLLECTION OF DATA

The data collection technique is different for different types of research design. There are predominantly two types of data: (i) the primary data and (ii) the secondary data. Primary data is one a researcher collects for a specific purpose of investigating the research problem at hand. Secondary data are ones that have not been collected for the immediate study at hand but for purposes other than the problem at hand. Both types of data offer specific advantages and disadvantages. Secondary data offer cost and time economies to the researcher as they already exist in various forms in the company or in the market.

3.3 METHODS OF DATA COLLECTION

Data collection method is an integral part of the research design. There are various methods of data collection, each method has its own advantages and disadvantages. Selection of an appropriate method of data collection may enhance the value of research and at the same time the wrong choice may lead to questionable research findings. Data collection methods include interviews, Self-administered questionnaires, observations and other methods.

A. Methods of Collecting Primary Data

Primary data means the data that have been collected originally for the first time. In other words, primary data may be the outcome of an original statistical enquiry, measurement of facts or a count that is undertaken for the first time. For instance data of population census is primary. Primary data being fresh from the fields of investigation is very often referred to as raw data. In the collection of primary data, a good deal of time, money and energy are required.

Primary data may be obtained by applying any of the following methods:

1. Observation Methods.
2. Direct Personal Interviews.
3. Indirect oral interviews.
4. Information from correspondents.
5. Questionnaire methods.
6. Schedule methods.

Notes

1. Observation Methods

Observation is the most commonly used data collection method in many of the studies relating to behavioral sciences. Observation enables to collect data without asking questions from the respondents. The respondents can be observed in the natural work environment or in lab settings and their activities and behaviors of interest can be recorded. In conducting research, casual examination without purpose cannot be called as observation. Observation becomes a scientific tool for data collection, if it is conducted specifically to answer a research question. It should be systematically planned and executed using proper controls and should provide a reliable and valid account of what has happened.

This is another type of method used when the researcher feels that survey type of methods may not be so relevant in data collection. In subjective issues, respondents need to be observed rather than asked lest biases and prejudices happen in their response. Observation method may be either structured or unstructured. Structured observation method involves having a set of items to be observed and how the measurements are to be recorded. In unstructured observation, the observer monitors all aspects of the phenomena that seem relevant to the problem at hand. In this context, the observer may have an open mind to study the persons or object. Observation is a method that employs vision as its main means of data collection. It implies the use of eyes rather than of ears and the voice. It is accurate watching and noting of phenomena as they occur with regard to the cause and effect or mutual relations. It is watching other persons' behavior as it actually happens without controlling it. *For example,* watching bonded labourer's life, or treatment of widows and their drudgery at home, provide graphic description of their social life and sufferings. Observation is also defined as "a planned methodical watching that involves constraints to improve accuracy".

Types of Observation

Observation can be grouped under the following categories:

(i) Type of activity under observation

Observation includes monitoring both behavioral and non-behavioral activities and conditions. Behavioral observation includes nonverbal analysis, linguistic analysis, extra linguistic analysis and spatial analysis. Non-verbal analysis includes body movements, motor expressions and exchanged glances. Body movement indicates interest, boredom, anger or pleasure. Motor expression includes facial movements, blink of eye and exchanged glances. Linguistic behaviour includes the number of repeated words used by persons in a conversation. It also includes the type of interaction process that occurs, between two persons or in small groups.

(ii) Directness of the observation

Based on the directness of observation, it can be grouped as direct or indirect. Direct observation happens when the observer is physically present and monitors while the event is taking place. This is highly flexible as the observer can decide what to observe, how much time to spend on observation of an aspect, when to shift focus etc. The observer may feel bored or frustrated by constantly being on the watch and may tend to loose focus. This might reduce the accuracy and completeness of the observation. Another weakness is that the observer may be overloaded when the events takes place quickly which cannot be kept track of or recorded. Observation carried out using mechanical, photographic or electronic means are grouped under indirect observation. *For example,* the uses of video cameras, pupilometric devices etc to capture the behaviour of consumers are grouped under indirect observation. Indirect observation can be carried out in an unbiased manner. Further, loss of information due to boredom, fatigue, overloading etc is avoided. However, the indirect observation is less flexible as they may be programmed earlier.

Notes

(iii) Concealment

This categorization is based on whether the participant is aware of the observer's presence. The presence of observer may cause the participant to behave in a different manner which might arrest the very purpose of observation. If the activity in which the participants are involved is highly absorbing then there is a high chance that the participant may remain unaffected by the presence of the observer. However, the potential bias due to the presence of the observer cannot be totally ruled out. In order to rule out the bias in behaviour the observers may conceal themselves from the object being observed using some mechanical means. For example, one way mirror, camera, microphone etc. However, this has to be carefully evaluated on the basis of ethical grounds. Partial concealment is where the presence of the observer is not concealed but his objectives or interest is not revealed. In order to evaluate the performance of a sales person, a sales manager may be present when the salesman is dealing with the customer.

(iv) Participation

The presence of the observer and his involvement in the research setting is called participant observation. He plays the role of observer as well as the participant. The participants may or may not know about the same. The observer should be more efficient as he has to play a dual role. Non-participant observation occurs when the observer collects the data without becoming an integral part of the research setting. The observer merely observes the activities, records them and tabulates them in a systematic manner. This type of observation requires the observer to be physically present in the research setting for a extended period of time which makes it a time consuming task.

(v) Definiteness of Structure

The observation can be grouped as structured and unstructured observation. Clear definition of various aspects of observation viz., the units to be observed, method of recording, extent of accuracy needed, conditions of observation and selection of pertinent data of observation etc are the characteristic of structured observation. Structured observation is appropriate in case of descriptive studies. If the observation is conducted without the above characteristics defined in advance, it is termed as unstructured observation. This method of observation is usually followed in exploratory studies.

(vi) Extent of Control

The observation can be carried out in controlled or uncontrolled settings. Uncontrolled observation is carried out in a natural setting. No attempt is made to use precision instruments. The main aim of using this method is to get a spontaneous picture of reality. It provides naturalness and completeness to observation. However, it may lead to subjective interpretation and over confidence that the observer knows more about the observed phenomena than the actual. It is usually used in exploratory research. Controlled observation takes place according to a definite predetermined plan. It involves experimental procedure and involves the use of precision instruments to record the observation. The observation is usually carried out in a standardized and accurate manner leading to certain assured degree of generalization.

Merits of Observation Method

a) ***Common method:*** The method of observation is common to all the disciplines of research.

b) ***Simplicity:*** The method is very simple to use.

c) ***Realistic:*** Since observation is based on actual and first hand experience, its data are more realistic than the data of those techniques which are indirect and secondary source of information.

d) ***Formulation of hypothesis:*** In all the business operations, the method of observation is used as the basis of formulating hypothesis, regarding business research problem.

Notes

e) ***Verification:*** For verification of hypothesis, again we depend upon observation. Therefore, it can be said that the problem presents itself and resolves itself through observation method.

f) ***Greater reliability of conclusions:*** The conclusions of observations are more reliable than non-observation conclusions, because they are based on first hand perception by the eyes and can be verified by any one by visual perception.

Demerits of Observation Method

a) ***Some events cannot be objects of observation:*** There are certain events which are microscopic, indefinite and may not occupy any definite space or occur at a definite time and can not be noticed for observation purposes. *For example,* it is not possible to observe emotions and sentimental factors, likes and dislikes etc.

b) ***Illusory observation:*** Since we have to depend upon our eyes for observation, we can never be sure if what we are observing is the same as it appears to our eyes, Eyes are prone to deception. It is well known that eyes see a mirage in desert at noon.

c) ***Self-consciousness in the observed:*** In observation method, the atmosphere tends to become artificial and this leads to a sense of self consciousness among the individuals who are being observed. This hampers their naturalness in behaviour and thus the purpose of observation which is to know the behaviour of individuals under normal conditions get defeated.

d) ***Subjective explanation:*** The final results of observation depend upon, the interpretation and understanding of the observer, the defects of subjectivity in the explanation creep in description of the observed and deductions from it. *For example,* if we see a man coming out of a wine shop, quite drunk, and he starts firing at random, we may believe that liquor induces irrational violence in a man, which may not be the case always.

e) ***Slowness of Investigation:*** The slowness of observation methods lead to disheartening, disinterest among both observer and observed.

f) ***Expensive methodology:*** Being a long drawn process, the technique of observation is expensive.

g) ***Inadequacy:*** The full answer cannot be obtained by observation alone, observation must be supplemented by other methods of study.

2. Direct Personal Interviews

Face to Face contact is made with the informants under this method of collecting data. The interviewer asks them questions pertaining to the survey and collects the desired information.

There are many merits and demerits of this method, which are discussed as under:

Merits:

1. Most often respondents are happy to pass on the information required from them when contacted personally and thus response is encouraging.
2. The information collected through this method is normally more accurate because interviewer can clear doubts of the informants about certain questions and thus obtain correct information. In case the interviewer apprehends that the informant is not giving accurate information, he may cross-examine him and thereby try to obtain the information.
3. This method also provides the scope for getting supplementary information from the informant, because while interviewing it is possible to ask some supplementary questions which may be of greater use later.
4. There might be some questions which the interviewer would find difficult to ask directly, but with some tactfulness, he can mingle such questions with others and get the desired information.

Notes

He can twist the questions keeping in mind the informant's reaction. Precisely, a delicate situation can usually he handled more effectively by a personal interview than by other survey techniques.

5. The interviewer can adjust the language according to the status and educational level of the person interviewed, and thereby can avoid inconvenience and misinterpretation on the part of the informant.

Demerits:

1. This method can prove to be expensive if the number of informants is large and the area is widely spread.
2. There is a greater chance of personal bias and prejudice under this method as compared to other methods.
3. The interviewers have to be thoroughly trained and experienced; otherwise they may not be able to obtain the desired information. Untrained or poorly trained interviewers may spoil the entire work.
4. This method is more time consuming as compared to others. This is because interviews can be held only at the convenience of the informants. Thus, if information is to be obtained from the working members of households, interviews will have to be held in the evening or on week end.

Telephonic Interviews

Interviewing through telephones enables to gain the following advantages:

1. Conducting interview through telephone enables to reduce the cost. The cost reduction arises due to reduction in traveling and administrative expenses involved.
2. In training and supervision. It is enough to train less number of interviewer since the interview is conducted through telephone. Coverage per person through telephone will be more than the face to face interviews.
3. Telephonic interview enables to screen and cover large population spread over a wide geographical location. It enables to have a much more representative sample.
4. Computer administered telephone surveys can also be conducted where the computer can replace the interviewer. A computer calls the phone number, conducts the interview and place data into a file for later tabulation.
5. The interviewer's bias caused by physical appearance, body language and actions are reduced by using telephones. The respondent may feel more relaxed, comfortable and unhesitant to reveal information as face to face contact is not present.
6. Unlike face to face interview where the respondent may avoid contact with the researcher, the contact rate is higher in telephonic interviews as the respondent has to pick up the ringing phone. However, the use of caller identification facility may reduce the contact rate.

The following drawbacks arise out of telephonic interviews:

1. Though the penetration rate of telephones is increasing in India, still there is a vast population without telephone facility. Also the number of users with only cell phone connection is increasing. Their numbers are not listed and reaching them would be difficult.
2. The random sample identified through telephone directories may be sometimes not available in the number given or may be malfunctioning.
3. The length or duration for which the telephonic interview can be conducted is limited. Ten minutes interview is considered as ideal. However sometimes the interview may extend to more than an hour also.

Notes

4. It is difficult or impossible to use maps, illustration, visual aids , measurement scale techniques in the telephonic interview. The researcher cannot depend more on the visualization techniques.
5. The interview can be terminated by the respondent as easily as the contact could be made. Also the level of interest and rapport in the telephonic interview is much lesser when compared to face to face interviews
6. The challenging and distracting physical environment either at home or office may reflect on the quality of data collection and may also result in refusal to participate in the interviews.

3. Indirect Oral Interviews

Under this method of data collection, the investigator contacts third parties generally called 'witnesses' who are capable of supplying necessary information. This method is generally adopted when the information to be obtained is of a complex nature and informants are not inclined to respond if approached directly. *For example,* when the researcher is trying to obtain data on drug addiction or the habit of taking liquor, there is high probability that the addicted person will not provide the desired data and hence will disturb the whole research process. In this situation taking the help of such persons or agencies or the neighbors who know them well becomes necessary. Since these people know the person well, they can provide the desired data. Enquiry Committees and Commissions appointed by the Government generally adopt this method to get people's views and all possible details of the facts related to the enquiry.

Though this method is very popular, its correctness depends upon a number of factors which are discussed below:

(i) The person or persons or agency whose help is solicited must be of proven integrity; otherwise any bias or prejudice on their part will not bring the correct information and the whole process of research will become useless.

(ii) The ability of the interviewers to draw information from witnesses by means of appropriate questions and cross-examination.

(iii) It might happen that because of bribery, nepotism or certain other reasons those who are collecting the information give it such a twist that correct conclusions are not arrived

Therefore for the success of this method it is necessary that the evidence of one person alone is not relied upon. Views from other persons and related agencies should also be ascertained to find the real position .Utmost care must be exercised in the selection of these persons because it is on their views that the final conclusions are reached.

4. Information from Corresponents

The investigator appoints local agents or correspondents in different places to collect information under this method. These correspondents collect and transmit the information to the central office where data are processed. This method is generally adopted by news paper agencies. Correspondents who are posted at different places supply information relating to such events as accidents, riots, strikes, etc., to the head office. The correspondents are generally paid staff or sometimes they may be honorary correspondents also. This method is also adopted generally by the government departments in such cases where regular information is to be collected from a wide area. For example, in the construction of a wholesale price index numbers regular information is obtained from correspondents appointed in different areas. The biggest advantage of this method is that it is cheap and appropriate for extensive investigation. But a word of caution is that it may not always ensure accurate results because of the personal prejudice and bias of the correspondents. As stated earlier, this method is suitable and adopted in those cases where the information is to be obtained at regular intervals from a wide area.

Notes

5. Questionnaire Methods

A questionnaire is defined as a formalised schedule for collecting data from respondents. It may be called as a schedule, interview form or measuring instrument. Measurement error is a serious problem in questionnaire construction. The broad objective of a questionnaire includes one without measurement errors.

6. Schedule Method

A Schedule contains a set of questions which are asked and filled by an interviewer in a face to face situation with a respondent. It is a standardized device or tool of observation to collect the data in an objective manner. In this method the interviewer puts certain questions and the respondent furnishes certain answers and the interviewer records them as in a research instrument called schedule.

Objectives of Schedule

The main objectives of the schedule are as follows:

i) ***Delimitation of the topic:*** A schedule is always about a definite item of research study. It's subject is a single subject item rather than the research subject in general. The schedule delimits and specifies the subject of inquiry.

ii) ***Aids memory:*** It is not possible for the interviewer to keep in mind or memorize all the information that he has collected from different respondents. If no standardized tool is available he might put different questions to different persons and thereby get confused when he has to analyse and tabulate the data. Schedule acts as an "Aide Memoire".

iii) ***Aid to Classification and Analysis:*** Another objective of schedule is to tabulate and analyse the data collected in a scientific manner. Through schedules, the researcher can collect the matter in a homogeneous manner.

Types of Schedule

Schedules that are used in business research are classified as under:

i) ***Observation Schedule:*** The schedules that are used for observation are known as observation schedule. In these schedules observer records the activities and responses of a worker or a group under specific conditions. The main purpose of the observation schedule is to verify information.

ii) ***Rating Schedule:*** In the fields of business guidance, psychological research and social research the rating schedules are used to assess the attitude, opinions, preferences, inhibitions and other like elements, where the value and the trend of the above mentioned qualities are measured on a rating scale.

iii) ***Document Schedule:*** The schedule of this type are used to obtain data regarding return evidence and case histories from autobiography, diary, case histories or final records of governments etc. It is a good method for collecting exploratory data or preparing source list.

iv) ***Interview Schedule:*** In an Interview Schedule an interviewer presents the questions of the schedule to the interviewee and records their responses on blank spaces.

Characteristics of a good Schedule

The following are the essentials or characteristics of a good schedule:

1. ***Accurate Communication:*** It means that the questions that are given in a schedule should be such that the respondent is able to understand them in the light in which they are asked. For accurate communication, the questions should be of the following types:

Notes

a) ***Questions Should be Interlinked:*** It means that if information about different aspects the questions asked should be such that their answers may present a compact picture of the information.

b) ***Suggestive Questions:*** The questions should be suggestive. There should be questions on each topic, but the questions should be so designed that the respondent may be encouraged to give the correct answers.

2. ***Accurate Response:*** It means that the schedule should be such that the required information may be easily secured. For this the interviewer has to prepare the schedule in a scientific manner and also make efforts to inspire the respondent to give answers. For this the following steps should be taken.

a) The size of the schedule should not be too lengthy.

b) The questions of the schedule should be clearly worded and be unambiguous.

c) The questions should be free from subjective evaluation.

d) Information sought should be capable of being tabulated.

Design of Schedule

Questions to be included in the schedule:

The basic thing while framing schedule is that those questions should be included which reflects the nature of study and problem. Normally the questions that are included in the schedule should have the following characteristics:

i) Questions should be short, clearly worded, simple for the respondents to answer.

ii) Questions should have a direct bearing on the problem.

iii) Questions should be such that the information that is collected through them can be subject to processing and tabulation.

iv) The questions should be interrelated and they should be such that cross checking may be possible.

v) Questions should be free from personal bias.

vi) Questions should be standardized and precise terms should be used.

vii) Questions should be thorough and they should be such that the respondents have to take minimum effort to answer them. If the questions are cumbersome and require too much pontification on the part of the respondents they shall not invite accurate and easy replies.

Organization of Schedule

Once the schedule has been properly and scientifically framed, the process of interview starts. It is through the interview that schedule is completed and the data collected. For this purpose, the following steps have to be taken.

1) ***Selection of the Respondents:*** The first thing that has to be done after the framing of the schedule is to select the proper type and number of the informants or respondents. Generally sampling method is employed in the use of the schedule. The sample selected should be perfectly representative. The sample having been selected, their names and addresses should be legibly and correctly noted. This would enable the field workers to approach them.

2) ***Selection, Training and Job of the Field Worker:*** In schedule method, it is the field workers who carry on the interview and collect data. Since there is a dearth of field workers; they have to be selected according to the requirements and characteristics of the study. They also have

Notes

to be trained accordingly. Apart from the training, they should possess certain basic characteristics to conduct the study properly. The field worker has to possess the following characteristics:

a) Honesty and integrity

b) Initiative and tactfulness

c) Patience

d) Unbiased and scientific outlook

e) Interest in research area

f) Knowledgeable about the subject of study

g) Trained in techniques and methods of study

3) ***Interview and Correct replies:*** In schedule method, the success very much depends upon the results of the interview. If the field worker has taken a successful interview, there is every likelihood that he has collected the correct information. It requires the following things:

   a) Correct Approach: It means that the field worker should approach the respondent in such a manner that he may get the right input from him. Generally he should be approached when he is not busy or through some such contact that he may not refuse to provide the information.

   b) Proper Response: Proper response is the result of a proper approach. Apart from it th proper response depends upon other factors also. For this, the field worker should be able to convince the respondent.

   c) Correct Reply: The field worker collects his data on the basis of the answers given by the respondent. It involves two factors, one is the correctness of the schedule, second the proper approach to the respondent. For proper response and correct reply, the researcher should use probing questions, but without hurting the feelings of the respondents.

4) ***Testing the Validity of the Results:*** When the schedule has been completed and returned by the field worker to the researcher, it should be subjected to certain tests so that it may be verified if the data collected is accurate or not. It can be done through various ways. The investigator may himself select certain respondents and interview them again. In case the reply is different, then what has actually been recorded in the schedule, should be either rejected or subjected to a study again. If there is slight variation, the validity should not be doubted.

Suitability of Schedule Method

This method is generally employed in following situations:

a) The field of investigation is wide.

b) Where the researcher/investigator requires quick results at low cost.

c) Where the respondents are educated.

d) Where trained and educated investigators are available.

Merits of Schedule Method

The main merits or advantages of this method are listed below:

(i) It can be adopted in those cases where informants are illiterate.

(ii) There is the scope of non-response as the enumerators go personally to obtain the information.

(iii) The information received is more reliable as the accuracy of statements can be checked by supplementary questions wherever necessary.

Notes

Demerits of Schedule Method

(i) In comparison to other methods of collecting primary data this method is quite costly as enumerators are generally paid persons.

(ii) The success of the method depends largely upon the training imparted to the enumerators.

(iii) Interviewing requires skill, experience and training. Many statisticians have the tendency to neglect this extremely important part of the data collecting process and this result in bad interviews. Without good interviewing most of the information collected is of doubtful value.

(iv) When questions are asked by a number of different interviewers, it is possible that variations in the personalities of the interviewers will cause variation in the answers obtained. This variation will not be obvious. Hence, every effort must be made to remove as much of variation as possible due to different interviewers.

B. Methods of Collecting Secondary Data

Secondary data are those data which have already been collected and analyzed by some earlier agency for its own use and later the same data are used by a different agency. According to W. A. Neiswanger, "A primary source is a publication in which the data are published by the same authority which gathered and analyzed them. A secondary source is a publication, reporting the data which was gathered by other authorities and for which others are responsible."

Secondary data are the data that are in actual existence in accessible records, having been already collected and treated statistically by the persons maintaining the records. In other words, secondary data are the data that have been already collected, presented tabulated, treated with necessary statistical techniques and conclusions have been drawn. Therefore, collecting secondary data doesn't mean doing some original enumeration but it merely means obtaining data that have already been collected by some agencies, reliable persons, government departments, research workers, dependable organisations etc. Secondary data are easily obtainable from reliable records, books, government publications and journals.

When once primary data have been originally collected, moulded by statisticians or statistical machinery then it becomes secondary in the hands of all other persons who may be desirous of handling it for their own purpose or studies. It follows, therefore that primary and secondary data are demarcated separately and that the distinction between them is of degree only.

Sources of Secondary Data

The various sources of secondary data can be divided into two broad categories:

1. Published sources and

2. Unpublished sources.

1. Published Sources

The governmental, international and local agencies publish statistical data, and chief among them are explained below:

(a) International Publications: There are some international institutions and bodies like I.M.F, I.B.R.D, I.C.A.F.E and U.N.O who publish regular and occasional reports on economic and statistical matters.

(b) Official publications of Central and State Governments: Several departments of the Central and State Governments regularly publish reports on a number of subjects. They gather additional information. Some of the important publications are: The Reserve Bank of India Bulletin, Census of India, Statistical Abstracts of States, Agricultural Statistics of India, Indian Trade Journal, etc.

Notes

(c) ***Semi-official publications:*** Semi-Government institutions like Municipal Corporations, District Boards, Panchayats, etc. publish reports relating to different matters of public concern.

(d) ***Publications of Research Institutions:*** Indian Statistical Institute (I.S.I), Indian Council of Agricultural Research (I.C.A.R), Indian Agricultural Statistics Research Institute (I.A.S.R.I), etc. publish the findings of their research programs..

(e) Publications of various Commercial and Financial Institutions

(f) Reports of various Committees and Commissions appointed by the Government as the Raj Committee's Report on Agricultural Taxation, Wanchoo Committee's Report on Taxation and Black Money, etc. are also important sources of secondary data.

(g) ***Journals and News Papers:*** Journals and News Papers are very important and powerful source of secondary data. Current and important materials on statistics and socio-economic problems can be obtained from journals and newspapers like Economic Times, Commerce, Capital, Indian Finance, Monthly Statistics of trade etc.

2. Unpublished Sources

Unpublished data can be obtained from many unpublished sources like records maintained by various government and private offices, the theses of the numerous research scholars in the universities or institutions etc.

Benefits of Secondary Data

Various benefits of secondary data are:

1. Time Saving

The first advantage of using secondary data has always been the saving of time. Not enough with this, in the so called Internet Era, this fact is more than evident. In the past, secondary data collection used to require many hours of tracking on the long libraries corridors. New technology has revolutionized this world. The process has been simplified. Precise information may be obtained via search engines. All worth library has digitized its collection so that students and researchers may perform more advance searches.

2. Accessibility

In the past, secondary data was often confined to libraries or particular institutions. Internet has especially been revolutionary in this sense. Having a internet connection is frequently the only requirement to access. A simple click is sometimes more than enough to obtain vast amount of information. The problem, nevertheless, is now being able to see whether the data is valid.

3. Saving of money

Strongly connected to the previous advantages is the saving of money. In general, it is much less expensive than other ways of collecting data. One may analyze larger data sets like those collected by government surveys with no additional cost.

4. Feasibility

Feasibility is of both longitudinal and international comparative studies. Continuous or regular surveys such as government censuses or official registers are especially good for such research purposes.

5. Generating new insights from previous analyses

Reanalyzing data can also lead to unexpected new discoveries. Returning to the previous example, the World Values Survey Association usually publishes the so called World Values Survey Books. They are a collection of publications based on data from the World Values Surveys. Since the database used

Notes may be accessible for outsider, you can analyze the data and come up with new relevant conclusions or simply verify and confirm previous results.

Drawbacks of Secondary Data

Various drawbacks of secondary data are:

1. Bias

Many documents used in research were not originally intended for research purposes. The various goals and purposes for which documents are written can bias them in various ways. For example, personal documents such as confessional articles or autobiographies are often written by famous people or people who had some unusual experience such as having been a witness to a specific event. While often providing a unique and valuable research data, these documents usually are written for the purpose of making money. Thus they tend to exaggerate and even fabricate to make good story. They also tend to include those events that make the author look good and exclude those that cast him or her in a negative light.

2. Selective Survival

Since documents are usually written on paper, they do not withstand the elements well unless care is taken to preserve them. Thus while documents written by famous people are likely to be preserved, day-to-day documents such as letters and diaries written by common people tend either to be destroyed or to be placed in storage and thus become inaccessible. It is relatively rare for common documents that are not about some events of immediate interest to the researcher (e.g., suicide) and not about famous occurrence or by some famous person to be gathered together in a public repository that is accessible to researchers.

3. Incompleteness

Many documents provide incomplete account to the researcher who has had no prior experience with or knowledge of the events or behavior discussed. A problem with many personal documents such as letters and diaries is that they were not written for research purposes but were designed to be private or even secret. Both these kinds of documents often assume specific knowledge that researcher unfamiliar with certain events will not possess. Diaries are probably the worst in this respect, since they are usually written to be read only by the author and can consist more of "soul searching" and confession than of description. Letters tend to be little more complete, since they are addressed to a second person. Since many letters assume a great amount of prior information on the part of the reader.

4. Lack of availability of documents

In addition to the bias, incompleteness and selective survival of documents, there are many areas of study for which no documents are available. In many cases information simply was never recorded. In other cases it was recorded but the documents remain secret or classified or have been destroyed.

5. Sampling bias

One of the problems of bias occurs because persons of lower educational or income levels are less likely to be represented in the sampling frames. The problem of sampling bias by educational level is more acute for document study than for survey research. It is a safe generalization that a poorly educated people are much less likely than well educated people to write documents.

6. Limited to verbal behavior

By definition, documents provide information only about respondent's verbal behavior, and provide no direct information on the respondent's nonverbal behavior, either that of the document's author or other characters in the document.

Notes

3.4 PROCESSING OF DATA

The various stages of data analysis process are given below:

Stage-1: Data cleaning

Data cleaning is an important procedure during which the data are inspected, and erroneous data are if necessary, preferable and possible corrected. Data cleaning can be done during the stage of data entry. If this is done, it is important that no subjective decisions are made. It should always be possible to undo any data set alterations. Therefore, it is important not to throw information away at any stage in the data cleaning phase. All information should be saved (i.e., when altering variables, both the original values and the new values should be kept, either in a duplicate data set or under a different variable name) and all alterations to the data set should carefully and clearly documented, for instance in a syntax or a log.

Stage-2: Initial data analysis

The most important distinction between the initial data analysis phase and the main analysis phase, is that during initial data analysis one refrains from any analysis that are aimed at answering the original research question. The initial data analysis phase is guided by the following four questions

Stage-3: Check the quality of data

The quality of the data should be checked as early as possible. Data quality can be assessed in several ways, using different types of analyses: frequency counts, descriptive statistics (mean, standard deviation, and median), normality (skewness, kurtosis, frequency histograms, normal probability plots), associations (correlations, scatter plots). Other initial data quality checks are:

i) Checks on data cleaning have decisions influenced the distribution of the variables? The distribution of the variables before data cleaning is compared to the distribution of the variables after data cleaning to see whether data cleaning has had unwanted effects on the data.

ii) Analysis of missing observations is there many missing values, and are the values missing at random? The missing observations in the data are analyzed to see whether more than 25% of the values are missing, whether they are missing at random (MAR) and whether some form of imputation is needed.

iii) Analysis of extreme observations outlying observations in the data are analyzed to see if they seem to disturb the distribution.

iv) Comparison and correction of differences in coding schemes variables are compared with coding schemes of variables external to the data set and possibly corrected if coding schemes are not comparable.

Stage-4: Measurement of Quality

The quality of the measurement instruments should only be checked during the initial data analysis phase when this is not the focus or research question of the study. One should check whether structure of measurement instruments corresponds to structure reported in the literature.

Stage-5: Initial transformations

After assessing the quality of the data and of the measurements, one might decide to impute missing data or to perform initial transformations of one or more variables, although this can also be done during the main analysis phase.

Notes

Stage-6: Characteristics of data sample

In any report or article, the structure of the sample must be accurately described. It is especially important to exactly determine the structure of the sample (and specifically the size of the subgroups) when subgroup analyses will be performed during the main analysis phase. The characteristics of the data sample can be assessed by looking at:

i) Basic statistics of variables

ii) Scatter plots

iii) Correlations

iv) Cross-tabulations

Stage-7: Final stage of the initial data analysis

During the final stage, the findings of the initial data analysis are documented, and necessary, preferable, and possible corrective actions are taken. Also, the original plan for the main data analyses can and should be specified in more detail and/or rewritten.

3.5 ANALYSIS OF DATA

Data analysis is a body of methods that help to describe facts, detect patterns, develop explanations and test hypotheses. It is used in all of the sciences. It is used in business, in administration and in policy. The numerical results provided by a data analysis are usually simple. It finds the number that describes a typical value and it finds differences among numbers. Data analysis finds averages, like the average income or the average temperature and it find differences like the difference in income from group to group or the differences in average temperature from year to year. Fundamentally, the numericalanswers provided by data analysis are that simple. But data analysis is not about numbers it uses them. Data analysis is about the world, asking, always asking, "How does it work?" And that's where data analysis gets tricky.

Meaning of 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 supporting decision making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science and social science domains.

3.6 TYPES OF DATA ANALYSIS

1. Univariate Data Analysis
2. Bivariate Data Analysis
3. Multivariate Data Analysis

1. Univariate Data Analysis

Univariate analysis is the simplest form of quantitative (statistical) analysis. The analysis is carried out with the description of a single variable and its attributes of the applicable unit of analysis. For example, if the variable age was the subject of the analysis, the researcher would look at how many subjects fall into a given age attribute categories.

Notes

Univariate analysis contrasts with bivariate analysis the analysis of two variables simultaneously or multivariate analysis the analysis of multiple variables simultaneously. Univariate analysis is also used primarily for descriptive purposes, while bivariate and multivariate analysis is geared more towards explanatory purposes. Univariate analysis is commonly used in the first stages of research, in analyzing the data at hand, before being supplemented by more advance, inferential bivariate or multivariate analysis.

A basic way of presenting univariate data is to create a frequency distribution of the individual cases, which involves presenting the number of attributes of the variable studied for each case observed in the sample. This can be done in a table format, with a bar chart or a similar form of graphical representation.

2. Bivariate Data Analysis

Bivariate data is data that has two variables. The quantities from these two variables are often represented using a scatter plot. This is done so that the relationship (if any) between the variables is easily seen.

Dependent and Independent Variables

In some instances of bivariate data, it is determined that one variable influences or determines the second variable and the terms dependent and independent variables are used to distinguish between the two types of variables.

Correlations occur between the two variables or data sets. These are determined as strong or weak correlations and are rated on a scale of 0-1.1 being a perfect correlation and 0.1 being a weak correlation.

Analysis of Bivariate Data

In the analysis of bivariate data, one typically either compares summary statistics of each of the variable quantities or uses regression analysis to find a more direct relationship between the data.

3. Multivariate Data Analysis

Multivariate analysis (MVA) is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical outcome variable at a time. In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest. Uses for multivariate analysis include:

i) Design for capability (also known as capability-based design).

ii) Inverse design, where any variable can be treated as an independent variable.

iii) Analysis of Alternatives (AoA), the selection of concepts to fulfill a customer need.

iv) Analysis of concepts with respect to changing scenarios.

v) Identification of critical design drivers and correlations across hierarchical levels.

3.7 QUESTIONNAIRE

Questionnaire is a list of questions or statements pertaining to an issue or program. It is used for studying the opinions of people. It is commonly used in opinion polls. People are asked to express their responses to the listed or reactions to the listed statements. Specifically, the objectives of a questionnaire are as follows:

a) It must translate the information needed into a set of specific questions that the respondents can and will answer.

Notes

b) The questions should measure what they are supposed to measure.

c) It must stimulate the respondents to participate in the data collection process. The respondents should be adequately motivated by the virtual construct of the questionnaire.

d) It should not carry ambiguous statements that confuses the respondents.

Most of the research studies carried out for solving business problems require the researcher to depend on primary data. The researcher should collect data through questionnaires/ interview schedules and process the same so as to provide solution to the identified problem. A questionnaire is a formalized framework consisting of a set of questions and scales designed to generate primary raw data. It is a preformulated written set of questions to which the respondents record their answers. The answers are mostly chosen by a respondent from within the closely defined alternatives.

The questionnaires can be administered personally, mailed to the respondents or electronically distributed:

A) Personally Administered Questionnaire

If the study is confined to a local area, the questionnaires can be collected by personally administering the same. The main advantage is that the researcher can collect all the completed responses within a short period of time. The researcher has an opportunity to introduce the research topic and motivate the respondents to offer frank answers. Any doubts that the respondents have on any questions is clarified on the spot. Administering the questionnaire to a large number of respondents at a time would save time and expenses and also ensure quick collection of data as against personal interviewing.

Hence, wherever possible group administration of questionnaire should be opted for depending on the sample frame work. The major drawback will be the reluctance of organizations to give time to conduct surveys among groups of employees.

B) Mail Questionnaire

Under this method, a list of questions pertaining to the survey which is known as 'Questionnaire' is prepared and sent to the various informants by post. Sometimes the researcher himself too contacts the respondents and gets the responses related to various questions in the questionnaire. The questionnaire contains questions and provides space for answers. A request is made to the informants through a covering letter to fill up the questionnaire and send it back within a specified time.

The questionnaire studies can be classified on the basis of:

(i) The degree to which the questionnaire is formalized or structured.

(ii) The disguise or lack of disguise of the questionnaire and

(iii) The communication method used.

When no formal questionnaire is used, interviewers adapt their questioning to each interview as it progresses. They might even try to elicit responses by indirect methods, such as showing pictures on which the respondent comments. When a researcher follows a prescribed sequence of questions, it is referred to as structured study. On the other hand, when no prescribed sequence of questions exists, the study is non-structured. When questionnaires are constructed in such a way that the objective is clear to the respondents then these questionnaires are known as non-disguised; on the other hand, when the objective is not clear, the questionnaire is a disguised one. On the basis of these two classifications, four types of studies can he distinguished:

(i) Non-disguised structured, (ii) Non-disguised non-structured,

(iii) Disguised structured and (iv) Disguised non-structured.

Types of Questionnaire *Notes*

The questionnaire may be of following types:

a) Structured Questionnaire: Structured Questionnaires are those in which a question is presented to the respondents with fixed response categories.

b) Un-structured questionnaire: Here every question is not necessarily presented to the respondent in the same wording and does not have fixed responses. Respondents are free to answer the question the way they like.

c) Mixed Questionnaire: This is a questionnaire which is neither completely structured nor un-structured. It consists of both the types of questions.

Components of a Questionnaire

A questionnaire consists typically of five sections. They are:

a) Identification data

b) Request for cooperation

c) Instruction

d) Information sought

e) Classification of data

Merits of Questionnaire Method

This method is an indirect method of data collection. It has certain advantages as compared to other methods.

i) Economical: In comparison to other methods of data collection (observation methods, case study, interview etc.) the mailed questionnaire method is the cheapest and quickest method. The cost in this method is only that of getting the questionnaire prepared and the postage expense. There is no need to visit the respondents personally or continue the study over a long period.

ii) Less skill of administration: The questionnaire method requires less skill to administer than an interview, observation or case study method of data collection.

iii) Research in wide area: If the informants or the respondents are scattered in large geographical areas, the Questionnaire method is the only means of research. The other methods of data collection such as schedule, interview or observation method do not prove to be successful. Even after spending large amount of money, it may not be possible to collect the information quickly but through questionnaire method, large areas can be covered. Some times certain agencies also co-operate in the task of dispatches or sending of the questionnaire to the informants.

iv) Time Saving: Besides saving money, questionnaire method saves time. Simultaneously hundreds of persons are approached through it whereas if they are to be interviewed it may take a long time.

v) More reliable in special cases: This is a method of collecting data in an objective manner through standardized impersonal questions. The respondents give free, frank and reliable information. Moreover the informants or respondents are free to give information as and when they want. Because of this freedom, the information that is provided is more dependable and reliable.

Notes

vi) ***Free from external influence:*** In questionnaire method, informants or respondents are free from external influences, as researcher is not present. They provide reliable, valid and meaningful information based on his knowledge, views and attitudes.

vii) ***Suitable for special type of responses:*** The information about certain problems can be best obtained through this method. *For example,* the research about marital relations, dreams etc. can easily be obtained by keeping the name of respondents anonymous.

viii) ***Less errors:*** Chances of errors are very low, because the supply of information is given by the respondent himself.

ix) ***Originality:*** The informants are directly involved in the supply of information, so the method is more original.

x) ***Uniformity:*** The impersonal nature of questionnaires ensure uniformity from one measurement situation to another.

xi) ***Collection of information relevant to the objective:*** Through this method, the questionnaires are framed according to the objective, hence data collection is also accordingly to that objective.

Demerits of Questionnaire Method

The method has the following disadvantages/limitations:

i) ***Lack of interest:*** Lack of interest on the part of respondents is very common. The respondents get disinterested due to large number of questions.

ii) ***Incomplete response:*** Some respondents give answers which are so brief that the full meaning is incomprehensible.

iii) ***Useless in-depth research problems:*** If a problem requires deep and long study, it cannot be studied through this method.

iv) ***Inelastic:*** This method is very rigid since no alteration may be introduced.

v) ***Prejudices and biases of the researcher influences the questions:*** Since researcher frames the questions his personal views, prejudices and biases influence the questions instead of becoming objective and impersonal, he becomes biased and prejudiced.

vi) ***Poor response and lack of reality:*** All the informants do not give answers or do not fill the questionnaire. There is a large percentage of those who do not send back the questionnaire. This makes the study unreliable.

vii) ***The incompleteness of the form of questionnaire:*** Sometimes the questionnaire is itself incomplete and some of the important aspects about which the information is required are not given, hence data collected is neither reliable nor helpful for the study.

viii) ***Lack of personal contact:*** There is no provision in this method for coming face to face with the respondent. This may result in manipulation of replies by the respondents.

3.8 DESIGN OF QUESTIONNAIRE

Guidelines for Questionnaire Design:

A good questionnaire accomplishes the research objectives. The logical sequences of the steps involved in the development of a good questionnaire are discussed below:

Step-1: Deciding the Information to be Collected

Step-2: Formulating the Questions

Step-3: Decide on the Wordings of the Questions and Layout of the Questionnaire *Notes*

Step-4: Pretesting the Questionnaire

Step-1: Deciding the Information to be Collected

The researcher should have a clear idea of exactly what information is to be collected from each respondent. Lack of clarity will lead to collection of irrelevant and incomplete information which does not contribute towards the research purpose. The situation will diminish the value of the study.

Clarity can be facilitated by:

1. Clear research objectives that will provide an insight into the kind of information needed, the hypotheses and the scope of the research.
2. Exploratory research will reveal the variables to be explored and will enable to understand the point of view of the respondents.
3. Experience with similar studies.
4. Pretesting the preliminary version of the questionnaire.

In deciding the content of the questionnaire the following guiding factors should be considered: The question may be asked to get information regarding objective or subjective variables or both.

In the case of objective variables like age, gender, income etc a single direct question can be asked. However, if the question is regarding subjective variable for e.g., regarding attitude, feeling, satisfaction etc then the questions should tap the dimensions and elements of the concept concerned.

- The researcher should challenge each questions in terms of its contribution towards providing an answer for the objectives. Questions which merely contribute interesting information and not towards the fulfillment of the objectives should be avoided. The researcher should learn the art of getting more information with fewer questions.
- The question should have a proper scope and should cover the issue. The questions asked should reveal all that is needed to know. Questions are considered to be ineffective if they do not provide the right information that is needed.
- The question should ask precisely what is needed. For e.g., if the researcher needs to know the 'family income' of the respondent but the question is asked regarding 'income' then it may mean income and not family income. Unambiguous words can be used so that clarity can be ensured.
- The question asked by the researcher may be contributing towards the theme and may be precise but it may not be possible for the respondent to answer the same adequately. The respondent may require time to think and answer certain questions. Sometimes the respondent may not be able to give an accurate answer due to his inability to recall things from memory.

Step-2: Formulating the Questions

Before formulating the questions a decision has to be made by the researcher regarding the degree of freedom to be given to the respondents in answering the questions. The various types of questions that can be included in a questionnaire are discussed below:

1. Open-Ended Versus Closed Questions

Unstructured questions or open-ended questions allow respondents to reply to the questions in own words. It enables the respondent to answer in any way he chooses.

Predetermined responses are not given to aid the respondent. For example a question asking the respondent to list five factors which made him to choose a particular investment proposal. This type of

Notes

question requires more thinking and effort on the part of respondents. In most cases an interviewer is required to prompt the response by asking probing questions. If correctly administered the open ended question can provide the researcher with a rich array of information.

Structured or closed-ended question in contrast provides a set of predetermined responses and the respondents is required to choose among the same. This question reduces the amount of thinking and effort required by the respondent. Instead of asking the respondent to list five factors, the questionnaire may provide a set of 10 to 15 factors and ask the respondent to rank the first five among the list, in the order of their preference. All items in the questionnaire using nominal, ordinal or Likert or ratio scale are considered closed. The closed-ended questions enable the researcher to code the responses easily for the purpose of carrying out subsequent analysis. Care should be exercised in making the alternatives provided as mutually exclusive and collectively exhaustive. Even a well – delineated category in closed question may make the respondent feel confined and he may be willing to provide additional comments. The researcher can tackle this issue by substantiating the closed-ended questionnaire with a final open ended question.

2. Dichotomous Questions

Two alternatives are suggested in dichotomous questions. The choices presented should be mutually exclusive i.e. the respondent should choose either of the answers only. At the same time the given choices should be collectively exhaustive.

3. Multiple Choice Questions

Multiple choices offer more than one alternative answer and from which the respondent can make a single choice. The list of answers provided should be collectively exhaustive. The alternatives provided should represent different aspects of the same conceptual dimension. The multiple choice question usually generates nominal data. When the choices are numbers, the response structure will produce at least interval and sometimes ratio data.

4. Checklist Questions

Checklist questions are used when the researcher wants the respondent to give multiple responses to a single question. For e.g., the factors leading to the choice of a particular brand laptop. The same information can be obtained from the respondent using a series of dichotomous selection questions, one for each factor. However, it would be time and space consuming. Checklists are more efficient.

5. Ranking Questions

Ranking question is used when the response regarding the relative order of the alternatives are important. For e.g., the check list question regarding the factors leading to the choice of laptop will only provide the factors considered but not the order of importance. The ranking question will lead the respondent to rank the most important factor as '1' the next important as '2' and so on.

6. Positively and Negatively worded Questions

The questionnaire should include both positively and negatively worded questions. If all the questions are positively worded then the respondent will tend to mechanically circle all the points toward one end of the scale. A respondent who is interested in completing the questionnaire soon will tend to circle all the questions to one end. The researcher can keep a respondent more alert by including both positive and negative worded questions. The use of double negatives and excessive use of words such as 'not', 'only' etc., should be avoided in the negatively worded question as they will tend to confuse the respondents.

Notes

7. Double-Barreled Questions

A question that leads to different possible responses to its sub-parts is called a double-barreled question. Such questions should be avoided by way of breaking the questions into two or more parts. For example, the question – do you like the flavour and the taste of the soft drink?. The question may lead to an ambiguous reply. It should be broken into two questions addressing flavour and taste separately so as to obtain unambiguous response. The type of question dealt below should be carefully avoided or used with caution by the researcher.

8. Ambiguous Questions

The question may not be double-barreled but still it may lead to ambiguity. For e.g., if the researcher involved in the study of job satisfaction asks the respondent to rate the level of satisfaction, the respondent may be confused as to whether the question is addressing satisfaction related to work environment, salary, team spirit or overall satisfaction. The question should not give rise to ambiguous response and bias.

9. Memory Related Questions

If the questions require respondents to recall experiences from a distance past that are very hazy in their memory, then the answers to such questions might be biased.

10. Leading/Loaded Questions

Questions should not be asked in such a way that the respondents are forced or directed to respond in a manner that he would not have, under normal situations where all possible alternatives are given. Questions should not prompt the respondents to answer in the way the researcher wants it answered. For example, "Don't you think that salary is the main reason for software employees to quit the job"?. Questions which are emotionally charging the respondents are called as loaded questions. Such questions would lead to bias in response and should be avoided.

Step-3: Decide on the Wordings of the Questions & Layout of the Questionnaire

The basic component of a questionnaire is the words. The researcher should be careful in considering the words to be used in creating the questions and scales for collecting raw data from respondents. The words used can influence respondent's reaction to the question. Even a small change in the words can affect the respondent's answers, but it is difficult to know in advance whether or not a change in wording will have an effect.The wording used in the questionnaire and the language used should be appropriate and understandable by the respondents.

Certain guidelines in deciding the wordings of the questionnaire are given below:

- The vocabulary should be simple, direct and familiar to all respondents. If the wordings jargons used or the language is not understood by the respondent, then it may lead to wrong or biased answers. The wording and language should be selected keeping in mind the educational level of the respondents, the terms used in the culture and the frames of reference of the respondents.
- The words used should not give rise to ambiguity or vagueness. This problem arises because of not giving the respondent an adequate frame of reference, in time and space for interpreting the question. Words such as 'often', 'usually' lack an appropriate time referent leading the respondents to choose their own which will lead to answers not comparable. Similarly, appropriate space or location is not often specified. *For example,* the question "Mention your place of origin" Does it elicit response as the district, state or country?
- Double-barreled questions should be avoided. The respondent may agree with one part of the question but not the other. For example, are you satisfied with the salary and increments given? The question should be broken or else it would lead to confusion and incorrect answers The

Notes

instructions provided to answer the question should not be confusing to the respondent. The questions should be directed more towards measuring the respondent's knowledge or interest in the subject.

- The questions asked should be applicable to all the respondents. Otherwise it will make a respondent to answer a question though they don't qualify to do so or may lack an opinion. *For example,* which other airways have you traveled before? This situation can be avoided by asking a qualifying or filter question and limit further questioning to those who qualify.
- Simple short questions should be asked instead of long ones. Researcher should see that a question or a statement in the questionnaire should be worded as minimum as possible.
- Questions should not be asked in such a manner that it will elicit socially desirable response. *For example,* "Do you think that physically challenged people should be given more weightage in employment opportunities"? Irrespective of the true feelings of respondents a socially desirable answer would be provided.

Sequencing and Layout Decisions

The order in which the questions are to be presented can encourage or discourage commitment and promote or hinder the development of researcher- respondent rapport. The sequence of questions asked in the questionnaire should lead the respondents from questions of general nature to specific nature. It should start with relatively easy questions which does not involve much thinking and should progress to difficult questions. This facilitates easy and smooth progress of the respondents through the various items in the questionnaire. Care should be taken to see that the positively and negatively worded questions addressing the same issue or concept are not placed continuously.

For example: I am satisfied with the working environment

I am not satisfied with the working environment

If the above questions appear in the same order it will appear meaningless to the respondent. The two questions should be placed in different places of the questionnaire. The way in which questions are sequenced would introduce bias in the response which is frequently referred to as the ordering effects. Randomly placing the questions in the questionnaire would reduce bias in the response, however, it is not attempted as it would lead to difficulty in categorizing, coding and analyzing the responses.

Layout of the Questionnaire

The appearance of the questionnaire is as important as its content. A neat, properly aligned and attractive questionnaire with a good introduction, instructions and well sequenced questions and response alternatives will make things easier for the respondents to answer. These aspects are explained below:

- In the Introduction section, the researcher can disclose his identity and communicate the purpose of the research. It is also used to motivate the respondents to answer the questions by conveying the importance of the research work and by specifying the importance of contribution from the respondent. The researcher should also ensure the confidentiality of the information provided. The introduction section should end with a courteous note, thanking the respondent for the time devoted to respond to the survey.
- The questions should be organized in a logical manner and numbered sequentially under appropriate sections. Proper instructions should be provided to complete the questions in an unambiguous manner. The questions should be neatly assigned so as to enable the respondent to read and answer the same without difficulty. The questionnaire should be designed in such a way that the respondent spends only minimum time and effort in completing the same.
- Questions relating to the personal profile of the respondents viz., name, gender, age, education, income, marital status etc., can appear in the beginning or at the end of the questionnaire. The

Notes

questions should provide a range of response options rather than seeking an exact figure. The personal profile related questions asked at the end may have a greater chance of response because the respondent would have gone through other questions which would have convinced him about the legitimacy and genuineness of the questions framed. This would make them more amenable to reveal the personal information. Some researchers feel that asking personal data in the beginning would enable the respondent to psychologically identify themselves with the questionnaire and enhance the commitment to respond.

- The open ended questions should be put at the end so the respondent may find it easy to comment on the various aspects.
- The questionnaire should end with an expression of sincere thanks to the respondent for spending their valuable time and effort. The researcher can also include a courteous note, reminding the respondents to check if all the items have been completed properly.

Step-4: Pre-testing the Questionnaire

The purpose of a pretest is to ensure that the questionnaire meets the researcher's expectations in terms of the information to be obtained. The objective of the pretest is to identify and correct the deficiencies in the questionnaire. It may lead to revising questions many times. It involves the use of a small number of respondents to test the appropriateness of the questions. 15 respondents are sufficient for a short and straightforward questionnaire, whereas 25 may be needed in case of a long and complex questionnaire with many branches and multiple options. Feedback is obtained from the respondents involved in the pretest on the general reaction to the questionnaire and regarding the effort involved in completing the questionnaire. Any difficulty or ambiguity can be identified and rectified before administering the questionnaire to a large number of respondents. This helps to rectify any mistakes in time and enables to reduce the biases.

Various type of pre testing can be carried out ranging from informal reviews by colleagues to creating conditions similar to the final study. Some types are discussed below:

a) ***Researcher Pretesting:*** It is conducted in the initial stages so as to build more structure in to the test. Fellow researchers can be involved. Many suggestions and discussions may take place leading to a refined questionnaire

b) ***Participant Pretesting:*** It involves testing the questionnaire in the field by involving the participants or participant surrogates. Surrogates are those individuals with characteristics and backgrounds similar to the desired participants.

c) ***Collaborative Pretesting:*** It can be conducted by the researcher where the researcher informs or alerts the participants of their involvement in the preliminary test of the questionnaire. This makes the participants as the collaborators in the process of refinement of the questionnaire. A detailed probing of the parts of the question, including the words and phrases is carried out.

d) ***Non-Collaborative Pretesting:*** In this type of pretesting the researcher does not inform the participant that the activity is a pretest. However, the probing of the questionnaire is done.

The pretest is conducted for the following reasons:

- The most important purpose for pretesting is to know whether the meaning of the questions is understood in the manner in which it is intended to. This problem may arise because, the respondent may not be familiar with certain words which will result in distortion of the meaning of the question. The respondent is likely to modify a difficult question in a way that makes it easier for him to respond. Flow of the questionnaire should be tested to know whether the transition from one topic to another is natural, logical and ensures a coherent flow.

Notes

- Many questionnaires have instructions on what question to skip, depending on the answer to a previous question. The skip pattern must be clearly laid out. In this context a questionnaire is like a road map with signs. Researchers who have been involved with the questionnaire design may not spot any inconsistencies or ambiguities as they are highly involved in the task. Pretesting will ensure the correct layout of the questionnaire.
- The length of the questionnaire is pretested as a lengthy questionnaire will often lead to fatigue among the respondents, interview break-off and refusal if the respondents know in advance the expected length.
- Task difficulty should also be identified through pretesting. The respondent maybe confused if the question requires that a respondent make connections or put together information in an unfamiliar way. For example, questions related to annual income. It involves calculation by the respondent. Instead the researcher can get monthly income and calculate the annual income on his own.
- Ability to capture and maintain the interest of the respondent throughout the entire questionnaire is a major challenge. The extent to which this is successful should be pretested
- Testing the items for an acceptable level of variation in the target population is one of the common goals of pretesting. The researcher should lookout for items showing greater variability.

Finally the pretest analysis should return to the first step in the design process. Each question should be reviewed again and again regarding its contribution to objectives of the study, leading to other steps. The last step in the process may be another pretest, if major changes are needed again.

3.9 TESTING OF HYPOTHESIS

Hypothesis test is a method of making decisions using data from a scientific study. In statistics, a result is called statistically significant if it has been predicted as unlikely to have occurred by chance alone, according to a pre-determined threshold probability, the significance level. The phrase "test of significance" was coined by statistician Ronald Fisher. These tests are used in determining what outcomes of a study would lead to a rejection of the null hypothesis for a pre-specified level of significance; this can help to decide whether results contain enough information to cast doubt on conventional wisdom, given that conventional wisdom has been used to establish the null hypothesis. The critical region of a hypothesis test is the set of all outcomes which cause the null hypothesis to be rejected in favor of the alternative hypothesis. Statistical hypothesis testing is sometimes called confirmatory data analysis, in contrast to exploratory data analysis, which may not have pre-specified hypotheses. Statistical hypothesis testing is a key technique of frequents inference.

Statistical hypothesis tests define a procedure that controls (fixes) the probability of incorrectly deciding that a default position (null hypothesis) is incorrect based on how likely it would be for a set of observations to occur if the null hypothesis were true. Note that this probability of making an incorrect decision is not the probability that the null hypothesis is true, nor whether any specific alternative hypothesis is true. This contrasts with other possible techniques of decision theory in which the null and alternative hypothesis are treated on a more equal basis. One naive Bayesian approach to hypothesis testing is to base decisions on the posterior probability, but this fails when comparing point and continuous hypotheses. Other approaches to decision making, such as Bayesian decision theory, attempt to balance the consequences of incorrect decisions across all possibilities, rather than concentrating on a single null hypothesis. A number of other approaches to reaching a decision based on data are available via decision theory and optimal decisions, some of which have desirable properties, yet hypothesis testing is a dominant approach to data analysis in many fields of science. Extensions to the theory of hypothesis testing

include the study of the power of tests, which refers to the probability of correctly rejecting the null hypothesis when a given state of nature exists. Such considerations can be used for the purpose of sample size determination prior to the collection of data.

Notes

3.10 PARAMETRIC AND NON-PARAMETRIC TESTS

Parametric Tests

Parametric statistics is a branch of statistics which assumes that sample data comes from a population that follows a probability distribution based on a fixed set of parameters. Most well-known elementary statistical methods are parametric. Conversely a non-parametric model differs precisely in that the parameter set (or feature set in machine learning) is not fixed and can increase or even decrease if new relevant information is collected.

Since a parametric model relies on a fixed parameter set, it assumes more about a given population than non-parametric methods do. When the assumptions are correct, parametric methods will produce more accurate and precise estimates than non-parametric methods, i.e. have more statistical power. However, as more is assumed by parametric methods, when the assumptions are not correct they have a greater chance of failing and for this reason are not robust statistical methods. On the other hand, parametric formulae are often simpler to write down and faster to compute. For this reason their simplicity can make up for their lack of robustness, especially if care is taken to examine diagnostic statistics.

Non-parametric Tests

Non parametric statistics refer to a statistical method in which the data is not required to fit a normal distribution. Nonparametric statistics uses data that is often ordinal, meaning it does not rely on numbers but rather a ranking or order of sorts.

Non parametric statistics refer to a statistical method in which the data is not required to fit a normal distribution. Non parametric statistics uses data that is often ordinal, meaning it does not rely on numbers but rather a ranking or order of sorts. *For example,* a survey conveying consumer preferences ranging from like to dislike would be considered ordinal data.

In statistics, parametric statistics includes parameters such as the mean, median, standard deviation, variance etc. This form of statistics uses the observed data to estimate parameters of the distribution. Under parametric statistics, data is assumed to fit a normal distribution with unknown parameters μ (population mean) and s2 (population variance), which are then estimated using the sample mean and sample variance. *For example,* a researcher that wants an estimate of the number of babies in North America born with brown eyes in 2017 may decide to take a sample of 1,50,000 babies and run an analysis on the data set. The measurement that s/he derives will be used as an estimate of the entire population of babies with brown eyes born in 2017.

Non parametric statistics does not assume that data is drawn from a normal distribution. Instead, the shape of the distribution is estimated under this form of statistical measurement. While there are many situations in which a normal distribution can be assumed, there are also some scenarios in which it will not be possible to determine whether the data will be normally distributed. *For example,* consider a researcher who wants to know whether going to bed early or late is linked to how frequently one falls ill. Assuming the sample is chosen randomly from the population, the sample size distribution of illness frequency can be assumed to be normal. However, an experiment that measures the resistance of the human body to a strain of bacteria cannot be assumed have a normal distribution. This is because a randomly selected sample data may be resistance to the strain. On the other hand, if the researcher considers factors such as genetic make-up and ethnicity, he may find that a sample size selected using these characteristics may not be resistant to the strain. Hence, one cannot assume a normal distribution.

Notes

Nonparametric statistics includes nonparametric descriptive statistics, statistical models, inference, and statistical tests. The model structure of nonparametric models is not specified a priori but is instead determined from data. The term 'non-parametric' is not meant to imply that such models completely lack parameters but that the number and nature of the parameters are flexible and not fixed in advance. A histogram is an example of a nonparametric estimate of a probability distribution.

Nonparametric statistics makes no assumption about the sample size or whether the observed data is quantitative. This method is useful when the data has no clear numerical interpretation and is best to use with data that has a ranking of sorts. *For example,* a personality assessment test may have a ranking of its metrics set as strongly disagree, disagree, indifferent, agree and strongly agree. In this case, non-parametric methods should be used.

Nonparametric statistics have gained appreciation due to their ease of use. As the need for parameters is relieved, the data becomes more applicable to a larger variety of tests. This type of statistics can be used without the mean, sample size, standard deviation or the estimation of any other related parameters when none of that information is available. Since nonparametric statistics makes fewer assumptions about the sample data, its application is wider in scope than parametric statistics.

In cases where parametric testing is more appropriate, nonparametric methods will be less efficient. This is because the results obtained from nonparametric statistics have a lower degree of confidence than if the results were obtained using parametric statistics.

3.11 T-TEST

A statistical examination of two population means. A two-sample t-test examines whether two samples are different and is commonly used when the variances of two normal distributions are unknown and when an experiment uses a small sample size.

Formula:
$$t = \frac{X - \mu}{\frac{S}{\sqrt{N}}}$$

Where, is the sample mean, Ä is a specified value to be tested, s is the sample standard deviation, and n is the size of the sample. Look up the significance level of the z-value in the standard normal table.

When the standard deviation of the sample is substituted for the standard deviation of the population, the statistic does not have a normal distribution; it has what is called the t-distribution. Because there is a different t-distribution for each sample size, it is not practical to list a separate area of the curve table for each one. Instead, critical t-values for common alpha levels (0.10, 0.05, 0.01, and so forth) are usually given in a single table for a range of sample sizes. For very large samples, the t-distribution approximates the standard normal (z) distribution. In practice, it is best to use t-distributions any time the population standard deviation is not known.

Values in the t-table are not actually listed by sample size but by degrees of freedom (df). The number of degrees of freedom for a problem involving the t-distribution for sample size n is simply n – 1 for a one-sample mean problem.

Uses of T Test

Among the most frequently used t-tests are:

i) A one-sample location test of whether the mean of a normally distributed population has a value specified in a null hypothesis.

Notes

ii) A two sample location test of the null hypothesis that the means of two normally distributed populations are equal. All such tests are usually called Student's t-tests, though strictly speaking that name should only be used if the variances of the two populations are also assumed to be equal; the form of the test used when this assumption is dropped is sometimes called Welch's t-test. These tests are often referred to as "unpaired" or "independent samples" t-tests, as they are typically applied when the statistical units underlying the two samples being compared are non-overlapping.

iii) A test of the null hypothesis that the difference between two responses measured on the same statistical unit has a mean value of zero. For example, suppose we measure the size of a cancer patient's tumor before and after a treatment. If the treatment is effective, we expect the tumor size for many of the patients to be smaller following the treatment. This is often referred to as the "paired" or "repeated measures" t-test: A test of whether the slope of a regression line differs significantly from 0.

Assumptions

Most *t*-test statistics have the form $T = \frac{Z}{S}$, where *Z* and *s* are functions of the data. Typically, *Z* is designed to be sensitive to the alternative hypothesis (i.e. its magnitude tends to be larger when the alternative hypothesis is true), whereas *s* is a scaling parameter that allows the distribution of *T* to be determined.

As an example, in the one-sample *t*-test $Z = \frac{\overline{X}}{\frac{\sigma}{\sqrt{n}}}$, where $\overline{X}$ is the sample mean of the data, *n* is the sample size, and ó is the population standard deviation of the data; *S* in the one-sample *t*-test is $\hat{\sigma} / \sigma$, where $\hat{\sigma}$ is the sample standard deviation.

The assumptions underlying a t-test are that:

i) Z follows a standard normal distribution under the null hypothesis

ii) ps2 follows a $\chi 2$ distribution with p degrees of freedom under the null hypothesis, where p is a positive constant

iii) Z and S are independent.

Unpaired and paired two-sample t-tests

Two-sample t-tests for a difference in mean can be either unpaired or paired. Paired t-tests are a form of blocking, and have greater power than unpaired tests when the paired units are similar with respect to "noise factors" that are independent of membership in the two groups being compared. In a different context, paired t-tests can be used to reduce the effects of confounding factors in an observational study.

Unpaired

The unpaired or "independent samples" t-test is used when two separate sets of independent and identically distributed samples are obtained, one from each of the two populations being compared. For example, suppose we are evaluating the effect of a medical treatment and we enroll 100 subjects into our study, and then randomize 50 subjects to the treatment group and 50 subjects to the control group. In this case, we have two independent samples and would use the unpaired form of the t-test. The randomization is not essential here if we contacted 100 people by phone and obtained each person's age and gender, and then used a two-sample t-test to see whether the mean ages differ by gender, this would also be an independent samples t-test, even though the data are observational.

Notes

Paired

Dependent samples (or "paired") t-tests typically consist of a sample of matched pairs of similar units or one group of units that has been tested twice (a "repeated measures" t-test). A typical example of the repeated measures t-test would be where subjects are tested prior to a treatment, say for high blood pressure, and the same subjects are tested again after treatment with a blood-pressure lowering medication.

A dependent t-test based on a "matched-pairs sample" results from an unpaired sample that is subsequently used to form a paired sample, by using additional variables that were measured along with the variable of interest. The matching is carried out by identifying pairs of values consisting of one observation from each of the two samples, where the pair is similar in terms of other measured variables. This approach is often used in observational studies to reduce or eliminate the effects of confounding factors.

Calculations

Explicit expressions that can be used to carry out various t-tests are given below. In each case, the formula for a test statistic that either exactly follows or closely approximates a t-distribution under the null hypothesis is given. Also, the appropriate degrees of freedom are given in each case. Each of these statistics can be used to carry out either a one-tailed test or a two-tailed test.

Once a t value is determined, a p-value can be found using a table of values from Student's t-distribution. If the calculated p-value is below the threshold chosen for statistical significance (usually the 0.10, the 0.05 or 0.01 level), the null hypothesis is rejected in favor of the alternative hypothesis.

One-sample* t-*test

In testing the null hypothesis that the population means is equal to a specified value μ_0 one uses the statistic.

$$t = \frac{\bar{x} - \mu o}{s/\sqrt{n}}$$

Where is the sample mean, S is the sample standard deviation of the sample and n is the sample size. The degrees of freedom used in this test is n - 1.

Slope of a regression line

Suppose one is fitting the model -

$$Y_i = \alpha + \beta x_i + \varepsilon_i,$$

where x_i, $i = 1, ..., n$ are known, α and β are unknown, and ε_i are independent identically normally distributed random errors with expected value 0 and unknown variance σ^2, and Y_i, $i = 1, ..., n$ are observed. It is desired to test the null hypothesis that the slope β is equal to some specified value β_0 (often taken to be 0, in which case the hypothesis is that x and y are unrelated).

Illustration - 1

A machine is designed to produce insulated washers with an average thickness of 0.025cms. A random sample of 10 washers was found to have an average thickness of 0.024 cms and a standard deviation of 0.002 cms. Test the significance of the deviation (take the tabulated value of t for 9 d.f at 0.05 level as 2.262).

Solution: *Notes*

Let the null hypothesis be that average thickness of washer is 0.025.

$H_0 : \mu_{HO} = 0.025$

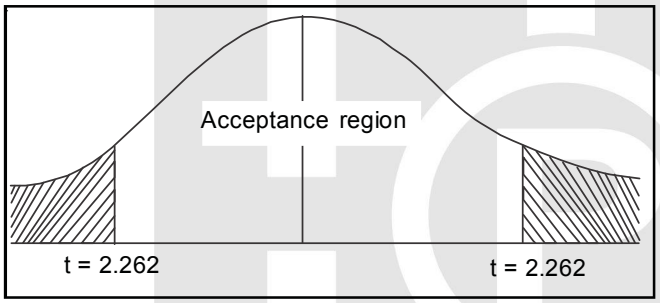
$H_a : \mu_{HO} \neq 0.025$

$$t = \frac{\overline{X} - \mu_{HO}}{\frac{\sigma_s}{\sqrt{n}}}$$

Given $\overline{X} = 0.024$; $\sigma_s = 0.002$; $n = 10$; $\mu_{HO} = 0.025$

As the sample size is small and population variation not known, 't' distribution is used.

$$t = \frac{\overline{X} - \mu_{HO}}{\frac{\sigma_s}{\sqrt{n}}} = \frac{0.024 - 0.025}{\frac{0.002}{\sqrt{10}}} = -1.58$$



Degree of freedom d.f = n – 1 = 10 – 1 = 9.

Value of 't' from table for df = 9 and $\alpha = 0.05$ is 2.262

Calculated value of |t| = 1.58

Table value of t = 2.262

Since calculated value of 't' falls in the acceptance region, H_o is accepted. We can conclude that average thickness of washer is 0.025 cms.

Illustration - 2

The intelligence quotients (IQs) of 16 students from one area of a city showed a mean of 107 and a standard deviation of 10, while the IQs of 14 students from another area of the city showed a mean of 112 and a standard deviation of 8. Is there a significant difference between the IQs of the two groups at 0.01 level of significance?

Solution:

$n_1 = 16$ $\overline{x}_1 = 107$ $\sigma_1 = 10$

$n_2 = 14$ $\overline{x}_2 = 112$ $\sigma_2 = 8$

The number of sample is 16 + 14 = 30

$\therefore$ Small sample test (t-test)

$H_0 : \overline{x}_1 = \overline{x}_2$

$H_1 : \overline{x}_1 \neq \overline{x}_2$

LOS $\alpha = 1\% = 0.01$

Notes

Test statistic $t = \frac{\bar{x}_1 - \bar{x}_2}{SE}$

$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{10^2}{16} + \frac{10^2}{14}} = \sqrt{625 + 4.57} = 3.2893$$

$$\therefore t = \frac{\bar{x}_1 - \bar{x}_2}{SE} = \frac{107 - 112}{3.2893} = \frac{-5}{3.2893} = -1.5200 \; |t| = 1.520$$

Table value $t_\alpha = 2.462$

$t < t_\alpha \Rightarrow$ Accept H_0

There is no significant difference between the 2 groups wrt 10

Illustration - 3

The tea stall near the railway station at Yesvantpura has been having average sale of 500 tea cups per day. Because of the development of bus stand nearby, it expects its sales to increase. During the first 12 days after the start of the bus stand, the daily sales were as under:

550, 570, 490, 615, 505, 580, 570, 460, 600, 580, 530, 526. On the basis of sample information, can one conclude that the tea stall's sales have increased? (take $\alpha = 0.05$).

Solution:

Let, the null hypothesis be that the average sales of the tea stall is 500 cups per day.

$H_o: \mu = 500$

$H_a: \mu > 500$ (as we have to conclude that sales have increased)

As the sample size is small and population variance not known, 't' distribution is used.

Given: $\mu = 500$; $n = 12$; σ_s – to be computed

$$t = \frac{\bar{X} - \mu_{HO}}{\frac{\sigma_s}{\sqrt{n}}}$$

where, $\sigma_s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}$

Day	Sales/day X_i	$X_i - \bar{X}$ ($\bar{X} = 548$)	$(X_i - \bar{X})^2$
1	550	2	4
2	570	22	484
3	490	–58	3364
4	615	67	4489
5	505	–43	1849
6	580	32	1024

7	570	22	484	*Notes*
8	460	– 88	7744	
9	600	52	2704	
10	580	32	1024	
11	530	– 18	324	
12	526	– 22	484	

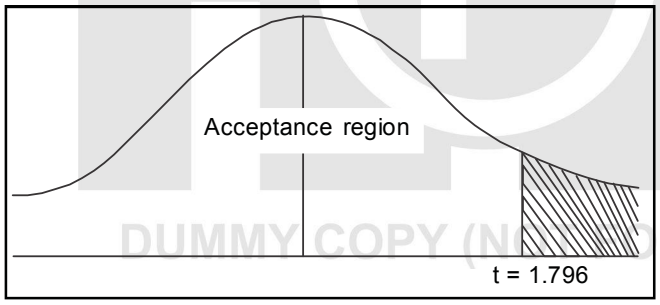
Total $\sum X_i = 6576$ $\sum (X_i - \overline{X})^2 = 23978$

$$\overline{X} = \frac{\sum X_i}{n} = \frac{6576}{12} = 548$$

$$\sigma_s = \frac{\sqrt{\sum (X_i - \overline{X})^2}}{n-1} = \frac{\sqrt{23978}}{12-1} = 46.68$$

$$\therefore \quad t = \frac{548 - 500}{\frac{46.68}{\sqrt{12}}} = \frac{48}{13.49} = 3.558$$

Degree of freedom d.f = n – 1 = 12 – 1 = 11



The rejection region for d.f = 11 and $\alpha = 0.05$ is 't' > 1.796, for a right tailed test.

Calculated t = 3.558

Since calculated value of t falls in the rejection region, H_0 is rejected and H_o is accepted. We can conclude that the sample data indicate tea stall sales have increased.

Illustration - 4

A random sample of size 16 has 53 as mean, the sum of the square of the deviation taken from mean is 135. Can this sample be regarded as taken from population having 56 as mean? Test at 5% significance level.

Solution:

Let, the null hypothesis be that hypothetical population mean is 56.

$H_o : \mu = 56$

$H_a : \mu \neq 56$

As the sample size is small and population variance not known 't' distribution is used.

Notes

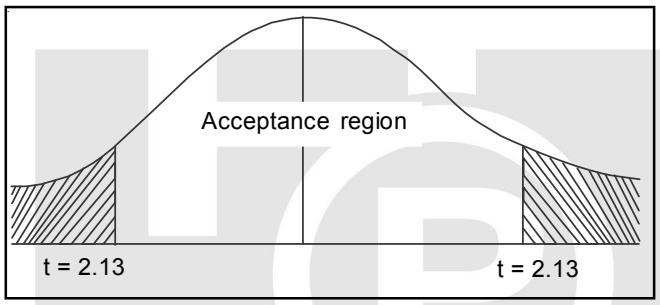
$$t = \frac{\overline{X} - \mu_{HO}}{\frac{\sigma_s}{\sqrt{n}}}$$

Given: $\overline{X} = 53$; $\mu_{HO} = 56$; $n = 16$; $\sum (X - \overline{X})^2 = 135$

$$\sigma_s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n-1}} = \sqrt{\frac{135}{16-1}} = 3$$

$$\therefore \quad t = \frac{53 - 56}{\frac{3}{\sqrt{16}}} = -4$$

Degree freedom d.f = n – 1 = 16 – 1 = 15



Rejection region for df = 15 and $\alpha = 0.05$ is $|t| > 2.13$, for a two tailed test.

Calculated value of $|t| = 4$.

Since calculated value falls in the rejection region, H_0 is rejected and H_a is accepted. Hence we can conclude that the sample has not come from the population having 56 as mean.

3.12 Z-TEST

A Z-test is any statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution. Because of the central limit theorem, many test statistics are approximately normally distributed for large samples. For each significance level, the Z-test has a single critical value (for example, 1.96 for 5% two tailed) which makes it more convenient than the Student's t-test which has separate critical values for each sample size. Therefore, many statistical tests can be conveniently performed as approximate Z-tests if the sample size is large or the population variance known. If the population variance is unknown (and therefore has to be estimated from the sample itself) and the sample size is not large, the Student t-test may be more appropriate.

General form

The most general way to obtain a Z-test is to define a numerical test statistic that can be calculated from a collection of data, such that the sampling distribution of the statistic is approximately normal under the null hypothesis. Statistics that are averages of approximately independent data values are generally well-approximated by a normal distribution. An example of a statistic that would not be well-approximated by a normal distribution would be an extreme value such as the sample maximum.

If T is a statistic that is approximately normally distributed under the null hypothesis, the next step in performing a Z-test is to determine the expected value 0 of T under the null hypothesis and

Notes

then obtain an estimate s of the standard deviation of T. Then calculate the standard score Z = (T– 0) / s, from which one-tailed and two-tailed p-values can be calculated as $\Phi(-|Z|)$ and $2\Phi(-|Z|)$, respectively, where Φ is the standard normal cumulative distribution function.

Use in Location Testing

The term Z-test is often used to refer specifically to the one-sample location test comparing the mean of a set of measurements to a given constant. If the observed data X1, ..., Xn are (i) uncorrelated, (ii) have a common mean m and (iii) have a common variance s^2, then the sample average X has mean ì and variance s^2 / n. If our null hypothesis is that the mean value of the population is a given number m 0, it can use X - m0 as a test-statistic, rejecting the null hypothesis if X - m0 is large.

To calculate the standardized statistic Z = (X - m0) / s, we need to either know or have an approximate value for s^2, from which we can calculate $S^2 = s^2 / n$. In some applications, s^2 is known, but this is uncommon. If the sample size is moderate or large, can substitute the sample variance for s^2, giving a plug-in test. The resulting test will not be an exact Z-test since the uncertainty in the sample variance is not accounted for however, it will be a good approximation unless the sample size is small. A t-test can be used to account for the uncertainty in the sample variance when the sample size is small and the data are exactly normal. There is no universal constant at which the sample size is generally considered large enough to justify use of the plug-in test. Typical rules of thumb range from 20 to 50 samples. For larger sample sizes, the t-test procedure gives almost identical p-values as the Z-test procedure.

Conditions

For the Z-test to be applicable, certain conditions must be met:

i) Nuisance parameters should be known, or estimated with high accuracy (an example of a nuisance parameter would be the standard deviation in a one-sample location test). Z-tests focus on a single parameter and treat all other unknown parameters as being fixed at their true values. In practice, due to Slutsky's theorem, "plugging in" consistent estimates of nuisance parameters can be justified. However if the sample size is not large enough for these estimates to be reasonably accurate, the Z-test may not perform well.

ii) The test statistic should follow a normal distribution. Generally, one appeals to the central limit theorem to justify assuming that a test statistic varies normally. There is a great deal of statistical research on the question of when a test statistic varies approximately normally. If the variation of the test statistic is strongly non-normal, a Z-test should not be used.

iii) If estimates of nuisance parameters are plugged in as discussed above, it is important to use estimates appropriate for the way the data were sampled. In the special case of Z-tests for the one or two sample location problem, the usual sample standard deviation is only appropriate if the data were collected as an independent sample.

iv) In some situations, it is possible to devise a test that properly accounts for the variation in plug-in estimates of nuisance parameters. In the case of one and two sample location problems, a t-test does this.

Z-tests other than location tests

Location tests are the most familiar t-tests. Another class of Z-tests arises in maximum likelihood estimation of the parameters in a parametric statistical model. Maximum likelihood estimates are approximately normal under certain conditions, and their asymptotic variance can be calculated in terms of the Fisher information. The maximum likelihood estimate divided by its standard error can be used as a test statistic for the null hypothesis that the population value of the parameter equals zero.

Notes

When using a Z-test for maximum likelihood estimates, it is important to be aware that the normal approximation may be poor if the sample size is not sufficiently large. Although there is no simple, universal rule stating how large the sample size must be to use a Z-test, simulation can give a good idea as to whether a Z-test is appropriate in a given situation.

Z-tests are employed whenever it can be argued that a test statistic follows a normal distribution under the null hypothesis of interest. Many non-parametric test statistics, such as U statistics, are approximately normal for large enough sample sizes, and hence are often performed as Z-tests.

Illustration - 1

Given a sample mean of 83, a sample standard deviation of 12.5 and sample size of 22, test the hypothesis that the value of the population mean is 70 against alternative that it is more then 70. Use the 0.025 significance level.

Solution:

$\bar{x} = 83 \quad \sigma_x = 12.5 \quad n = 22 \quad \mu = 70$

Null Hypothesis $H_0 : \bar{x} = \mu$

Alternate Hypothesis $H_1 : \bar{x} > \mu$ (one failed)

LOS $\alpha = 0.025$

$Z_\alpha = 2.055$

Test statistic:

$$Z = \frac{\bar{x} - \mu}{SE} \quad SE = \frac{\alpha}{\sqrt{x}} = \frac{83 - 70}{2.665} = \frac{12.5}{\sqrt{22}} = 4.878 = 2.665$$

$Z > Z_\alpha$ Reject H_0

Illustration - 2

The mean height of 50 male students who showed above average participation in college athletics was 68.2 inches with a standard deviation of 2.5 inches, while 50 male students who showed no interest in such participation had a mean height of 67.5 inches with a standard deviation of 2.8 inches. Test the hypothesis that male student who participate in college athletics are taller than other male students.

Solution:

$n_1 = 50 \quad \bar{x}_1 = 68.2 \quad s_1 = 2.5$

$n_2 = 50 \quad \bar{x}_2 = 67.5 \quad s_2 = 2.8$

$H_0 : \bar{x}_1 = \bar{x}_2$

$H_1 : \bar{x}_1 > \bar{x}_2$ (one failed test)

LOS $\alpha = 5$

Test statistics $z = \dfrac{\bar{x}_1 - \bar{x}_2}{SE}$

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(2.5)^2 + (2.8)^2}{50}} = 0.5308$$

Notes

$$\therefore z = \frac{\bar{x}_1 - \bar{x}_2}{SE} = \frac{68.2 - 67.5}{0.5308} = 1.3187$$

$Z_\alpha = 1.645$

$Z < Z_\alpha$

Accept H_0

The male students do not differ from the students who do not participate.

Illustration - 3

A sample of 400 boys is found to have a mean height of 67.47". Can it reasonably be regarded as a sample from a large population with mean height 67.39" and standard deviation 1.30"? (Test at 5% significance level).

Solution:

Let, the null hypothesis be that the mean height of the population is equal to 67.39".

i.e., H_0 $\mu_{HO} = 67.39$

H_a $\mu_{HO} \neq 67.39$

The population is infinite, sample is large sample ($\because$ N = 400) and population variance know. So the formula to be used is

$$Z = \frac{\bar{X} - \mu_{HO}}{\frac{\sigma}{\sqrt{n}}}$$

$\bar{X}$ = sample mean = 67 .47

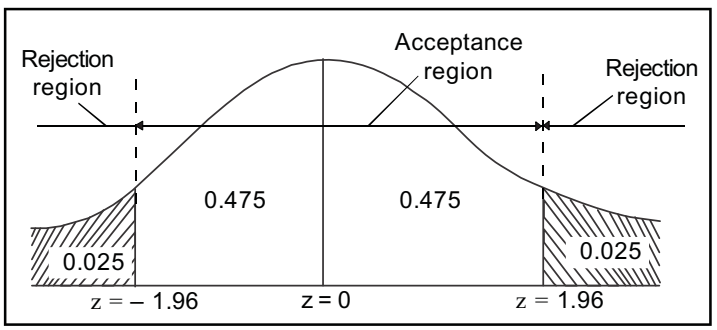
μ_{HO} = population mean = 67 .39

σ = population standard deviation = 1.30

n = sample size = 400

$$\therefore \quad z = \frac{67.47 - 67.39}{\frac{1.30}{\sqrt{400}}} = 1.231$$

As H_a is two sided, applying two tailed test for determining the rejection region at 5% significance level, we get 'z' value from normal table as 1.96 (i.e., for area = $\frac{0.95}{2}$ = 0.475).



Notes

Critical or table vale of z = 1.96

Calculated or observed value of z = 1.231

Since calculated value falls in the acceptance region, H_0 is accepted.

i.e., we may conclude that the given sample with mean height 67.47" can be regarded to have been taken from a population with mean height 67.39" and standard deviation 1 .30" at 5% level of significance.

Illustration - 4

A manufacturer of fluorescent tubes claims that his tubes have a lifetime on an average 2000 burning hours, a sample of 100 tubes was taken at random and tested for burning life. It was found to have a mean life of 1950 hours with a standard deviation of 150 hours, can the claim, of the manufacturer be accepted at 5% level of significance.

Solution:

Let, the null hypothesis be that the mean lifetime of bulbs is 2000 burning hours.

Null hypothesis $H_0 : \mu_{HO} = 2000$

Alternative hypothesis $H_a : \mu_{HO} < 2000$ (prove that mean is less than 2000)

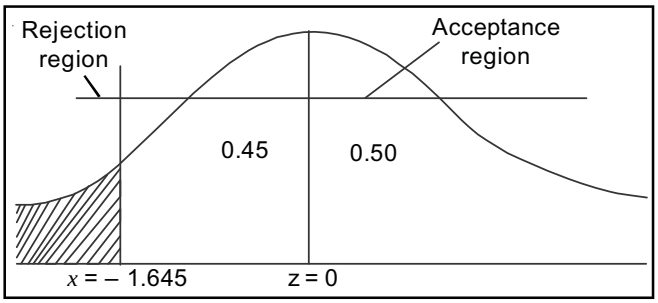
$$z = \frac{\overline{X} - \mu_{HO}}{\frac{\sigma_p}{\sqrt{n}}}$$

Where, σ_p can be replaced by σ_s

Given : $\overline{X} = 1950$; $\mu_{HO} = 2000$;

$\sigma_s = 150$; $n = 100$

$$\therefore \quad z = \frac{1950 - 2000}{\frac{150}{\sqrt{100}}} = \frac{-50}{15} = -3.33$$



It is a left tailed test and the rejection region for 5% level of significance s $Z < -1.645$

Calculated $Z = -3.33$; Table $Z = -1.645$

Since calculated value falls in the rejection region, H_0 is rejected and H_a accepted. Hence the manufacturer's claim cannot be accepted.

Notes

3.13 CHI-SQUARE TEST

χ^2 test is a test that uses the chi-square statistic to test the fit between a theoretical frequency distribution and a frequency distribution of observed data for which each observation may fall into one of several classes.

Conditions of Chi-square (χ^2) Test

A chi-square (χ^2) test can be used when the data satisfies four conditions:

i) There must be two observed sets of data or one observed set of data and one expected set of data (generally, there are n-rows and c-columns of data)

ii) The two sets of data must be based on the same sample size.

iii) Each cell in the data contains the observed or expected count of five or large?

iv) The different cells in a row of column must have categorical variables (male, female or younger than 25 years of age, 25 year of age, older than 40 years of age etc.)

Assumptions of Chi-square Test

The chi-squared test, when used with the standard approximation that a chi-squared distribution is applicable, has the following assumptions:

i) Simple random sample: The sample data is a random sampling from a fixed distribution or population where each member of the population has an equal probability of selection. Variants of the test have been developed for complex samples, such as where the data is weighted.

ii) Sample size (whole table): A sample with a sufficiently large size is assumed. If a chi squared test is conducted on a sample with a smaller size, then the chi squared test will yield an inaccurate inference. The researcher, by using chi squared test on small samples, might end up committing a Type II error.

iii) Expected cell count: Adequate expected cell counts. Some require 5 or more, and others require 10 or more. A common rule is 5 or more in all cells of a 2-by-2 table and 5 or more in 80% of cells in larger tables, but no cells with zero expected count. When this assumption is not met, Yates's correction is applied.

iv) Independence: The observations are always assumed to be independent of each other. This means chi-squared cannot be used to test correlated data (like matched pairs or panel data). In those cases you might want to turn to McNamara's test.

Application areas of Chi-square test

The χ^2 distribution typically looks like a normal distribution, which is skewed to the right with a long tail to the right. It is a continuous distribution with only positive values. It has following applications:

i) To test whether the sample differences among various sample proportions are significant or can they be attributed to chance.

ii) To test the independence of two variables in a contingency table.

iii) To use it as a test of goodness of fit.

Notes

Degrees of Freedom (d.f)

The degree of freedom, abbreviated as d.f, denotes the extent of independence (freedom) enjoyed by a given set of observed frequencies. Degrees of freedom are usually denoted by the letter 'v' of the Greek alphabet.

Suppose, if we are given a set of 'n' observed frequencies which are subjected to 'k' independent constraints (restrictions). Then

Degrees of Freedom = No. of frequencies – No. of independent constraints v = n – k

Formula of Chi-square text:

$$x^2 = \sum\left(\frac{(O_i - E_i)^2}{E_i}\right)$$

Table value of x^2 for d.f and α

$X^2_{cal} < X^2_{table}$, accept H_0

Illustration - 1

Find the value of χ^2 for the following data:

Group	A	B	C	D	E
Observed frequency	9	29	44	15	12
Theoretical frequency	5	24	30	30	16

Solution:

If the frequencies of any group is less than 10, it should be re-grouped by combining with the adjacent group as follows:

Group	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
A + B	38	29	9	81	2.79
C	44	30	14	196	6.53
D	15	30	– 15	225	7.50
E	12	16	– 4	16	1.00
				Total	**17.82**

$$\chi^2 = \sum\left(\frac{(O_i - E_i)^2}{E_i}\right) = 17.82$$

Here, degree of freedom = n – k = 5 – 1 – 2 = 2.

(2 d.f is subtracted for combining A and B).

Illustration - 2

The following data gives the information about no. of students per year who secured distinction in MBA in Osmania University. Are these frequencies in agreement with the belief that distinctions were the same during this 10 years period.

20 10 8 12 15 14 2 6 4 9

Solution: *Notes*

Null hypothesis:

Given, frequencies i.e., no of distinctions per year in MBA are consistent with the belief that distinction were the same during the 10 years period.

i.e., $H_0 : O_i = E_i$

Total number of distinctions = 20 + 10 + 8 + 12 + 15 + 14 + 2 + 6 + 4 + 9 = 100

$$\text{Expected no. of distinctions} = \frac{\text{Total no. of distinctions}}{\text{No. of years}} = \frac{100}{10} = 10$$

$$\chi^2 = \sum\left(\frac{(O_i - E_i)^2}{E_i}\right)$$

Where, O_i observed frequencies

E_i expected frequencies

Table

Year	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
1	20	10	10	100	10.0
2	10	10	0	0	0.0
3	8	10	–2	4	0.4
4	12	10	2	4	0.4
5	15	10	5	25	2.5
6	14	10	4	16	1.6
7	2	10	–8	64	6.4
8	6	10	–4	16	1.6
9	4	10	–6	36	3.6
10	9	10	–1	1	0.1
Total	**100**	**100**	**0**	**266**	**26.6**

$$\chi^2 = \sum\left(\frac{(O_i - E_i)^2}{E_i}\right) = 26.6$$

Thus, the calculated χ^2 value = 26.6

Table value of χ^2 for degree of freedom (9) and level of significance (5%) is 16.92 (degree of freedom = n – 1 = 10 – 1 = 9).

$\chi^2_{cal} = 26.6$ $\chi^2_{table} = 16.92$

Since, $\chi^2_{cal} > \chi^2_{table}$, null hypothesis is not accepted. Hence, conclude that the distinctions are nor uniform (same) over the 10 years period.

Notes

Illustration - 3

Genetic theory states that children having one parent of blood type A and the other blood type B will always be one of the three types A, AB, B and that the proportion of three types will be on an average as 1 : 2 : 1. A report states that out of 300 children having one 'A' parent and 'B' parent, 30% were found to be type 'A', 45% type AB and remainder type B. Test the hypothesis by X^2 test.

Solution:

Null hypothesis:

Theoretical hypothesis of the generate theory is supported by the report

i.e., $O_i = E_i$

Observed frequencies: (30%, 45%, 25%)

Total = 300

Type A: O_1 = 30% of 300 = 90

Type AB: O_2 = 45% of 300 = 135

Type B: O_3 = 25% of 300 = 75

Expected frequencies: (1 : 2 : 1)

Total = 300

Type A: $E_1 = \frac{1}{4} \times 300 = 75$

Type AB: $E_2 = \frac{2}{4} \times 300 = 250$

Type B: $E_3 = \frac{1}{4} \times 300 = 75$

Table

Type	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
A	90	75	15	225	3
AB	135	150	– 15	225	1.5
B	75	75	0	0	0
				Total	**4.5**

Thus, calculated value = 4.5

Degree of freedom = n – 1 = 3 – 1 = 2

Level of significance (α) = 5% (assumed)

The table value of χ^2 for 2d.f and α = 0.05 is 5.99

Since $\chi^2_{cal} < \chi^2_{table}$, null hypothesis is accepted. Which means the report supports the theoretical hypothesis of the genetic theory that on an average type A, AB, B stand in the proportion 1 : 2 : 1.

Illustration - 4

The details of number of male and female children in 800 families having four children each is given below. Test whether the data are consistent with the hypothesis that the binomial law holds and the chance of a male birth is equal to that of a female birth.

Notes

No. of births	Male 3	0 2	1 1	2 0	3	4	Female 4
Frequency		32	178	290	236	64	

Solution:

Null Hypothesis:

Data are consistent with the binomial law of equal probability for male and female births.

Expected frequencies are calculated using binomial probability law formula.

$f(r) = N\ P\ (r) = N \times (n\ C_r\ p^r\ q^{n-r})$

$N = 800, n = 4, p = \frac{1}{2}, q = \frac{1}{2}$

($\because$ Equal male and female birth)

No. of male births	*Expected frequency*
0	$800 \times \left(4\ C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4\right) = 50$
1	$800 \times \left(4\ C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^3\right) = 200$
2	$800 \times \left(4\ C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2\right) = 300$
3	$800 \times \left(4\ C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^1\right) = 200$
4	$800 \times \left(4\ C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^0\right) = 50$

Table

No. of male births	Observed frequency (O)	Expected frequency E	(O – E)	$(O - E)^2$	$\frac{(0 - E)^2}{E}$
0	32	50	– 18	324	6.48
1	178	200	– 22	484	2.42
2	290	300	– 10	100	0.33
3	236	200	36	1296	6.48
4	64	50	14	196	3.92
Total	800	800	0	2400	19.63

Notes

Calculated, $\chi^2 = \sum\left(\frac{(O_i - E_i)^2}{E_i}\right) = 19.63$

Degree of freedom d.f = 5 – 1 = 4

Level of significance (α) = 5% (assumed)

The table vale of χ^2 for d.f = 4, α = 0.05 is 9.49.

Since $\chi^2_{calculated} > \chi^2_{table}$, H_0 is rejected which means hypothesis of equal male and female births is wrong. Hence, the binomial distribution with p = q = $\frac{1}{2}$ is a not good fit to this data.

Illustration - 5

Test whether the marriage adjustment score and the level of education are dependent at α = 0.05.

Level of education	Marriage adjustment score			
	10 -15	26-45	46-70	71-100
Post graduate	30	40	75	20
Under graduate	50	60	50	40
Others	60	30	20	20

Solution:

Let, the null hypothesis be that marriage adjustment score and level of education are independent.

Expected frequency = $\frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$

$E_{ij} = \frac{RT_i \times CT_j}{GT}$

$RT_1 = 30 + 40 + 75 + 20 = 165$

$RT_2 = 50 + 60 + 50 + 40 = 200$

$RT_3 = 60 + 30 + 20 + 20 = 130$

$CT_1 = 30 + 50 + 60 = 140$

$CT_2 = 40 + 60 + 30 + 130$

$CT_3 = 75 + 50 + 20 + = 145$

$CT_4 = 20 + 40 + 20 = 80$

$GT = 165 + 200 + 130$

(or)

$= 140 + 130 + 145 + 80 = 495$

$E_{11} = \frac{165 \times 140}{495} = 46.67$ $\quad$ $E_{12} = \frac{165 \times 130}{495} = 43.33$

$E_{13} = \frac{165 \times 145}{495} = 48.33$ $\quad$ $E_{14} = \frac{165 \times 80}{495} = 26.67$

Notes

$$E_{21} = \frac{200 \times 140}{495} = 56.56 \qquad E_{22} = \frac{200 \times 130}{495} = 52.52$$

$$E_{23} = \frac{200 \times 145}{495} = 58.59 \qquad E_{24} = \frac{200 \times 80}{495} = 32.32$$

$$E_{31} = \frac{130 \times 140}{495} = 36.77 \qquad E_{32} = \frac{130 \times 130}{495} = 34.14$$

$$E_{33} = \frac{130 \times 145}{495} = 38.08 \qquad E_{34} = \frac{130 \times 80}{495} = 21.01$$

$$\chi^2 = \sum\left(\frac{(O_{ij} - E_{ij})^2}{E_{ij}}\right)$$

Where, O_{ij} – Observed frequency

E_{ij} – Expected frequency

Table

Cell	O_{ij}	E_{ij}	$(O_{ij} - E_{ij})$	$(O_{ij} - E_{ij})^2$	$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
E_{11}	30	46.67	– 16.67	277.89	5.95
E_{12}	40	43.33	3.33	11.09	0.26
E_{13}	75	48.33	26.67	711.29	14.72
E_{14}	20	26.67	– 6.67	44.49	1.67
E_{21}	50	56.56	– 6.56	43.03	0.76
E_{22}	60	52.52	7.48	55.95	1.07
E_{23}	50	58.59	– 8.59	73.79	1.26
E_{24}	40	32.32	7.68	58.98	1.82
E_{31}	60	36.77	23.23	539.63	14.68
E_{32}	30	34	– 4.14	17.14	0.50
E_{33}	20	38.08	– 18.08	326.89	8.58
E_{34}	20	21.01	– 1.01	1.02	0.05
				Total	**51.32**

From the table, $\chi^2_{cal} = 51.32$

Degree of freedom = (r – 1) (c - 1) = (3 – 1) (4 – 1) = 2 × 3 = 6

Level of significance (α) = 5%.

Table vale of χ^2 for d.f = 6 and α = 0.05 is 12.592.

Since $\chi^2_{cal} > \chi^2_{table}$, null hypothesis is not accepted which means the marriage adjustment score and level of education are dependent.

Notes

Illustration - 6

Out of a sample of 120 children in a village, 76 were administered a drug for prevention of a particular disease. Out of these 76 children, 24 were attacked by the disease whereas 12 children were not attacked by the disease who were not administered the drug. Prepare a 2 × 2 table showing actual and expected frequencies and use χ^2 test to determine whether or not the new drug was effective. (The value of χ^2 distribution from the table at 1 d.f at 0.5 level is 3.84).

Solution:

Let, the null hypothesis be that drug and prevention of disease are independent i.e., drug is not effective in prevention of disease.

2 × 2 Table

	Attacked	Not attacked	Total
Drug given	24	52	76
Drug not given	32	12	44
Total	56	64	120

$$\text{Expected frequencies} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

$$E_{11} = \frac{76 \times 56}{120} = 35 \qquad E_{12} = \frac{76 \times 64}{120} = 41$$

$$E_{21} = \frac{44 \times 56}{120} = 21 \qquad E_{22} = \frac{44 \times 64}{120} = 23$$

Group	O_{ij}	E_{ij}	$O_{ij} - E_{ij}$	$(O_{ij} - E_{ij})^2$	$\frac{(0_{ij} - E_{ij})^2}{E_{ij}}$
11	24	35	– 11	121	3.46
12	52	41	11	121	2.95
21	32	21	11	121	5.76
22	12	23	– 11	121	5.26
				Total	**17.43**

$$\text{Thus, Calculated } \chi^2 = \sum\left(\frac{(O_{ij} - E_{ij})^2}{E_{ij}}\right) = 17.43$$

Degrees of freedom = (r – 1) (c – 1) = (2 – 1) (2 – 1) = 0

Table value of χ^2 at d.f = 1, α = 0.05 is 3.84. since $\chi^2_{cal} > \chi^2_{table}$, null hypothesis is rejected which means the drug is effective in preventing the disease.

Illustration - 7 *Notes*

A Brand Manager is concerned that her brand's share may be unevenly distributed throughout the country. In a survey in which the country was divided into four Geographic regions, a random sampling of 100 consumers in each region was surveyed with the following results:

Region	NE	NW	SE	SW
Purchase the Brand	40	55	45	50
Do not purchase	60	45	55	50

i) State the null and alternative hypothesis.

ii) At 5% level, Test the Hypothesis.

Solution:

Purchase	40	55	45	50	190
Do not purchase	60	45	55	50	210
	100	100	100	100	400

Expected value $E_{40} \frac{100 \times 190}{400} = 47.5$

$E_{60} \frac{210 \times 100}{400} = 52.5$ $E_{55} \frac{190 \times 100}{400} = 47.5$

$E_{45} \frac{210 \times 100}{400} = 52.5$ $E_{45} \frac{190 \times 100}{400} = 47.5$

$E_{55} \frac{210 \times 100}{400} = 52.5$ $E_{50} \frac{190 \times 100}{400} = 47.5$

$E_{50} \frac{210 \times 100}{400} = 52.5$

Observed	E	O – E	$(O - E)^2$	$(O - E)^2/E$
40	47.5	-7.5	56.25	1.18
60	52.5	7.5	56.25	1.07
45	47.5	-2.5	6.25	0.13
55	52.5	2.5	6.25	0.11
45	47.5	-2.5	6.25	0.13
55	52.5	2.5	6.25	0.11
50	47.5	2.5	6.25	0.13
50	52.5	-2.5	6.25	0.11
				Ψ^2 = **2.96**

Notes

Straight line equation $Y = a + bX$

$\Sigma Y = na + B \Sigma X$ $\qquad$ $\Sigma XY = a \Sigma X + b \Sigma X^2$

$\Rightarrow 7,180 = 6a + 3b$ $\qquad$ $\Rightarrow 6,480 = 3a + 19b$

Solving $a = 1114.1$ $\qquad$ $b = 165.1$

i) $Y = 1114.1 + 165.1X$

ii) Curve

Y	X	X^2	X^3	X^4	XY	X^2Y
50	– 2	4	– 8	16	– 100	200
110	– 1	1	– 1	1	– 100	100
350	0	0	0	0	0	0
1,020	1	1	1	1	1,020	1,020
1,950	2	4	8	16	3,900	7,800
3,710	3	9	27	81	11,130	33,390
7,180	3	19	27	115	15,850	42,510

H_0 : There is no significant difference between the region and purchase of the brand.

H_1 : There is a significant difference.

LOS $\alpha = 5\%$

Test static $\Psi = 2.96$

Table value Ψ^2_α at $(r - 1)(c - 1)$

$\Psi^2_\alpha (3) = 7.115$

$\Psi^2 < \Psi^2_\alpha$

Accept H_0

3.14 SUMMARY

Data is the facts in raw or unorganized form such as alphabets, numbers or symbols that refer to or represent conditions, ideas or objects. This represents facts and statistics which are collected together for reference or analysis.

Primary data refers to information gathered firsthand by the researcher for the specific purpose of the study. It is raw data without interpretation and represents the personal or official opinion or position. Primary sources are most authoritative since the information is not filtered or tampered.

Secondary data refers to the information gathered from already existing sources. Secondary data may be either published or unpublished data.

Tertiary sources are an interpretation of a secondary source. It is generally represented by index, bibliographies, dictionaries, encyclopedias, handbooks, directories and other finding aids like the internet search engines.

A questionnaire is defined as a formalised schedule for collecting data from respondents. It may be called as a schedule, interview form or measuring instrument. Measurement error is a serious problem

Notes

in questionnaire construction. The broad objective of a questionnaire includes one without measurement errors.

Unpublished data can be obtained from many unpublished sources like records maintained by various government and private offices, the theses of the numerous research scholars in the universities or institutions etc.

Univariate analysis is the simplest form of quantitative (statistical) analysis. The analysis is carried out with the description of a single variable and its attributes of the applicable unit of analysis.

Bivariate data is data that has two variables. The quantities from these two variables are often represented using a scatter plot. This is done so that the relationship (if any) between the variables is easily seen.

A Schedule contains a set of questions which are asked and filled by an interviewer in a face to face situation with a respondent. It is a standardized device or tool of observation to collect the data in an objective manner. In this method the interviewer puts certain questions and the respondent furnishes certain answers and the interviewer records them as in a research instrument called schedule.

Observation is the most commonly used data collection method in many of the studies relating to behavioral sciences. Observation enables to collect data without asking questions from the respondents. The respondents can be observed in the natural work environment or in lab settings and their activities and behaviors of interest can be recorded.

Hypothesis test is a method of making decisions using data from a scientific study. In statistics, a result is called statistically significant if it has been predicted as unlikely to have occurred by chance alone, according to a pre-determined threshold probability, the significance level.

Parametric statistics is a branch of statistics which assumes that sample data comes from a population that follows a probability distribution based on a fixed set of parameters.

Non parametric statistics refer to a statistical method in which the data is not required to fit a normal distribution. Nonparametric statistics uses data that is often ordinal, meaning it does not rely on numbers but rather a ranking or order of sorts.

A statistical examination of two population means. A two-sample t-test examines whether two samples are different and is commonly used when the variances of two normal distributions are unknown and when an experiment uses a small sample size.

A Z-test is any statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution. Because of the central limit theorem, many test statistics are approximately normally distributed for large samples.

χ^2 test is a test that uses the chi-square statistic to test the fit between a theoretical frequency distribution and a frequency distribution of observed data for which each observation may fall into one of several classes.

3.15 SELF ASSESSMENT QUESTIONS

1. What is Data? Discuss the Collection of Data.
2. Explain the Methods of Data Collection.
3. Discuss the Processing of Data.
4. Explain the Analysis of Data.
5. Give the meaning of Data Analysis. Discuss the Types of Data Analysis.

Notes

6. What is Questionnaire? Explain the Design of Questionnaire.
7. Explain the Testing of Hypothesis.
8. What is Parametric and Non-parametric Tests?
9. What is T-test? Discuss the Unpaired and paired two-sample t-tests.
10. What is Z-test? What are the Conditions of Z-test?
11. What is Chi-square test? Explain the Assumptions and Application areas of Chi-square test.

*****

Chapter

4

MULTIVARIATE ANALYSIS TECHNIQUES

Objectives

The objectives of this lesson are to:

- Concept of Multivariate Data Analysis
- Techniques of Multivariate Analysis
- Multiple Regression Analysis
- Discriminated Analysis
- Factor Analysis
- ANOVA

Structure:

4.1 INTRODUCTION

Multivariate analysis is based in observation and analysis of more than one statistical outcome variable at a time. In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest. The development of multivariate methods emerged to analyze large databases and increasingly complex data. Since the best way to represent the knowledge of reality is the modeling, we should use multivariate statistical methods. Multivariate methods are designed to simultaneously analyze data sets, i.e., the analysis of different variables for each person or object studied. Keep in mind at all times that all variables must be treated accurately reflect the reality of the problem addressed. There are different types of multivariate analysis and each one should be employed according to the type of variables to analyze: dependent, interdependence and structural methods.

Notes

4.2 MULTIVARIATE DATA ANALYSIS

Multivariate analysis (MVA) is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical outcome variable at a time. In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest. Uses for multivariate analysis include:

i) Design for capability (also known as capability-based design).

ii) Inverse design, where any variable can be treated as an independent variable.

iii) Analysis of Alternatives (AoA), the selection of concepts to fulfill a customer need.

iv) Analysis of concepts with respect to changing scenarios.

v) Identification of critical design drivers and correlations across hierarchical levels.

Multivariate analysis can be complicated by the desire to include physics-based analysis to calculate the effects of variables for a hierarchical "system-of-systems." Often, studies that wish to use multivariate analysis are stalled by the dimensionality of the problem. These concerns are often eased through the use of surrogate models, highly accurate approximations of the physics-based code. Since surrogate models take the form of an equation, they can be evaluated very quickly. This becomes an enabler for large-scale MVA studies: while a Monte Carlo simulation across the design space is difficult with physics-based codes, it becomes trivial when evaluating surrogate models, which often take the form of response surface equations.

4.3 MULTIVARIATE ANALYSIS TECHNIQUES

Multivariate analysis techniques which can be conveniently classified into two broad categories viz., dependence methods and interdependence methods. This sort of classification depends upon the question: Are some of the involved variables dependent upon others? If the answer is 'yes', we have dependence methods; but in case the answer is 'no', we have interdependence methods. Two more questions are relevant for understanding the nature of multivariate techniques. Firstly, in case some variables are dependent, the question is how many variables are dependent? The other question is, whether the data are metric or non-metric? This means whether the data are quantitative, collected on interval or ratio scale, or whether the data are qualitative, collected on nominal or ordinal scale. The technique to be used for a given situation depends upon the answers to all these very questions. The category are included techniques like multiple regression analysis, multiple discriminant analysis, multivariate analysis of variance and canonical analysis

4.4 MULTIPLE REGRESSION ANALYSIS

Multiple regression is the most commonly utilized multivariate technique. It examines the relationship between a single metric dependent variable and two or more metric independent variables. The technique relies upon determining the linear relationship with the lowest sum of squared variances; therefore, assumptions of normality, linearity, and equal variance are carefully observed. The beta coefficients (weights) are the marginal impacts of each variable, and the size of the weight can be interpreted directly. Multiple regression is often used as a forecasting tool.

Assumptions:

Regression residuals must be normally distributed.

Notes

A linear relationship is assumed between the dependent variable and the independent variables.

The residuals are homoscedastic and approximately rectangular-shaped.

Absence of multicollinearity is assumed in the model, meaning that the independent variables are not too highly correlated.

At the center of the multiple linear regression analysis is the task of fitting a single line through a scatter plot. More specifically the multiple linear regression fits a line through a multi-dimensional space of data points. The simplest form has one dependent and two independent variables. The dependent variable may also be referred to as the outcome variable or regressand. The independent variables may also be referred to as the predictor variables or regressors.

There are 3 major uses for multiple linear regression analysis. First, it might be used to identify the strength of the effect that the independent variables have on a dependent variable.

Second, it can be used to forecast effects or impacts of changes. That is, multiple linear regression analysis helps us to understand how much will the dependent variable change when we change the independent variables. For instance, a multiple linear regression can tell you how much GPA is expected to increase (or decrease) for every one point increase (or decrease) in IQ.

Third, multiple linear regression analysis predicts trends and future values. The multiple linear regression analysis can be used to get point estimates.

The Multiple Regression Model

In general, the multiple regression equation of Y on X1, X2, …, Xk is given by:

$Y = b0 + b1\ X1 + b2\ X2 + \ldots\ldots\ldots\ldots\ldots + bk\ Xk$

Interpreting Regression Co-efficients

Here b0 is the intercept and b1, b2, b3, …, bk are analogous to the slope in linear regression equation and are also called regression coefficients. They can be interpreted the same way as slope. Thus if bi = 2.5, it would indicates that Y will increase by 2.5 units if Xi increased by 1 unit.

The appropriateness of the multiple regression model as a whole can be tested by the F-test in the ANOVA table. A significant F indicates a linear relationship between Y and at least one of the X's.

1. Discriminant Analysis

Discriminant analysis is the regression based statistical technique that is used in determining the particular classification or group for an item of data or an object belongs to on the basis of its characteristics or essential features. It differs from group building techniques such as cluster analysis in that the classifications or groups to choose from must be known in advance.

Purposes of Discriminant Analysis

Discriminant Analysis undertakes the same task as multiple linear regressions by predicting an outcome. However, multiple linear regressions is limited to cases where the dependent variable on the Y axis is an interval variable so that the combination of predictors will, through the regression equation, produce estimated mean population numerical Y values for given values of weighted combinations of X values.

2. Factor Analysis

A type of analysis used to discern the underlying dimensions or regularity in phenomena. Its general purpose is to summarize the information contained in a large number of variables into a smaller number of factors.

Notes Factor analysis attempts to identify underlying variables or factors, that explain the pattern of correlations within a set of observed variables. Factor analysis is often used in data reduction to identify a small number of factors that explain most of the variance that is observed in a much larger number of manifest variables.

3. Cluster Analysis

A body of techniques with the purpose of classifying individuals or objects into a small number of mutually exclusive groups, ensuring that there will be as much likeness within groups and as much difference among groups as possible.

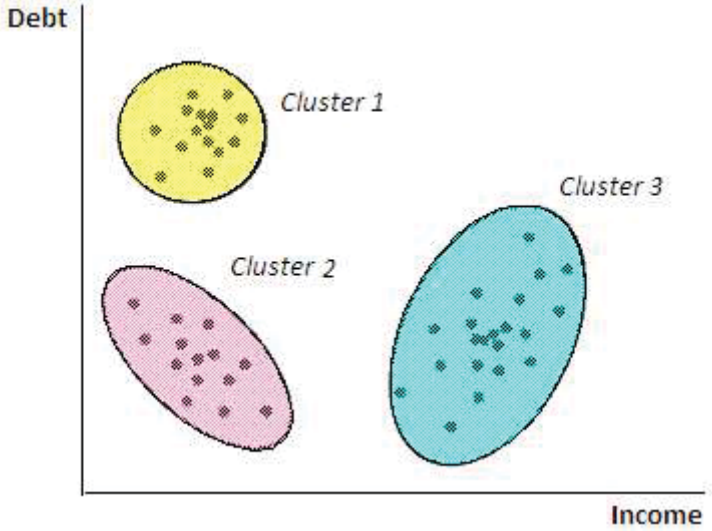
Cluster analysis is a collection of statistical methods which identifies groups of samples that behave similarly or show similar characteristics. In common parlance it is also called look-a-like groups.

Concept of Cluster Analysis

Cluster analysis is a collection of statistical methods, which identifies groups of samples that behave similarly or show similar characteristics. In common parlance it is also called look-a-like groups. The simplest mechanism is to partition the samples using measurements that capture similarity or distance between samples. In this way, clusters and groups are interchangeable words. Often in market research studies, cluster analysis is also referred to as a segmentation method. In neural network concepts, clustering method is called unsupervised learning. Typically in clustering methods, all the samples with in a cluster is considered to be equally belonging to the cluster. If each observation has its unique probability of belonging to a group and the application is interested more about these probabilities than we have to use multinomial models.

Cluster analysis is a class of statistical techniques that can be applied to data that exhibit "natural" groupings. Cluster analysis sorts through the raw data and groups them into clusters. A cluster is a group of relatively homogeneous cases or observations. Objects in a cluster are similar to each other. They are also dissimilar to objects outside the cluster, particularly objects in other clusters.

Explanation



Clustering and segmentation basically partition the database so that each partition or group is similar according to some criteria or metric. Clustering according to similarity is a concept which appears in many disciplines. If a measure of similarity is available there are a number of techniques for forming clusters. Membership of groups can be based on the level of similarity between members and from this the rules of membership can be defined. Another approach is to build set functions that measure some

property of partitions i.e. groups or subsets as functions of some parameter of the partition. This latter approach achieves what is known as optimal partitioning. *Notes*

Many data mining applications make use of clustering according to similarity for example to segment a client/customer base. Clustering according to optimization of set functions is used in data analysis e.g. when setting insurance tariffs the customers can be segmented according to a number of parameters and the optimal tariff segmentation achieved.

Clustering/segmentation in databases are the processes of separating a data set into components that reflect a consistent pattern of behaviour. Once the patterns have been established they can then be used to "deconstruct" data into more understandable subsets and also they provide sub-groups of a population for further analysis or action which is important when dealing with very large databases. *For example,* a database could be used for profile generation for target marketing where previous response to mailing campaigns can be used to generate a profile of people who responded and this can be used to predict response and filter mailing lists to achieve the best response.

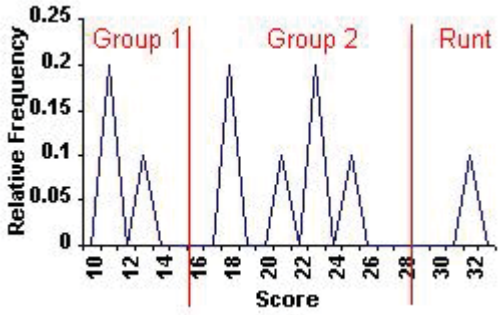
Simple Cluster Analysis

In cases of one or two measures, a visual inspection of the data using a frequency polygon or scatter plot often provides a clear picture of grouping possibilities. *For example,* the following is the data from the "Example Assignment" of the cluster analysis homework assignment.

Example Assignment

	Julie	John	Ryan	Bob	Ted	Kristi	Carol	Alice	Kari	Dave
SCORE	11	11	13	18	18	21	23	23	25	32

The relative frequency polygon appears as follows:



It is fairly clear from this picture that two subgroups, the first including X, Y, and Z and the second including everyone else except describe the data fairly well. When faced with complex multivariate data, such visualization procedures are not available and computer programs assist in assigning objects to groups. The following text describes the logic involved in cluster analysis algorithms.

Steps in Doing a Cluster Analysis

A common approach to doing a cluster analysis is to first create a table of relative similarities or differences between all objects and second to use this information to combine the objects into groups. The table of relative similarities is called a proximities matrix. The method of combining objects into groups is called a clustering algorithm. The idea is to combine objects that are similar to one another into separate groups.

The Proximities Matrix

Cluster analysis starts with a data matrix, where objects are rows and observations are columns. From this beginning, a table is constructed where objects are both rows and columns and the numbers in

Notes the table are measures of similarity or differences between the two observations. *For example,* given the following data matrix:

	X_1	X_2	X_3	X_4	X_5
O_1					
O_2					
O_3					
O_4					

A proximities matrix would appear as follows:

	O_1	O_2	O_3	O_4
O_1				
O_2				
O_3				
O_4				

The difference between a proximities matrix in cluster analysis and a correlation matrix is that a correlation matrix contains similarities between variables (X_1, X_2) while the proximities matrix contains similarities between observations (O_1, O_2).

The researcher has dual problems at this point. The first is a decision about what variables to collect and include in the analysis. Selection of irrelevant measures will not aid in classification. For example, including the number of legs an animal has would not help in differentiating cats and dogs, although it would be very valuable in differentiating between spiders and insects.

The second problem is how to combine multiple measures into a single number, the similarity between the two observations. This is the point where univariate and multivariate cluster analysis separate. Univariate cluster analysis groups are based on a single measure, while multivariate cluster analysis is based on multiple measures.

Univariate Measures

A simpler version of the problem of how to combine multiple measures into a measure of difference between objects is how to combine a single observation into a measure of difference between objects. Consider the following scores on a test for four students:

Student	Score
X	11
Y	11
Z	13
A	18

The proximities matrix for these four students would appear as follows:

	X	Y	Z	A
X				
Y				
Z				
A				

Notes

The entries of this matrix will be described using a capital "D", for distance with a subscript describing which row and column. *For example,* D_{34} would describe the entry in row 3, column 4, or in this case, the intersection of Z and A.

One means of filling in the proximities matrix is to compute the absolute value of the difference between scores. *For example,* the distance, D, between Z and A would be |13-18| or 5. Completing the proximities matrix using the example data would result in the following:

	X	Y	Z	A
X	0	0	2	7
Y	0	0	2	7
Z	2	2	0	5
A	7	7	5	0

A second means of completing the proximities matrix is to use the squared difference between the two measures. Using the example above D_{34}, the distance between Z and A, would be (13-18)2 or 25. This distance measure has the advantage of being consistent with many other statistical measures, such as variance and the least squares criterion and will be used in the examples that follow. The example proximities matrix using squared differences as the distance measure is presented below.

	X	Y	Z	A
X	0	0	4	49
Y	0	0	4	49
Z	4	4	0	25
A	49	49	25	0

Note that both example proximities matrices are symmetrical. Symmetrical means that row and column entries can be interchanged or that the numbers are the same on each half of the matrix defined by a diagonal running from top left to bottom right.

Other distance measures have been proposed and are available with statistical packages. For example, SPSS/WIN provides the following options for distance measures.

Some of these options themselves contain options. *For example,* Minkowski and Customized are really many different possible measures of distance.

Multivariate Measures

When more than one measure is obtained for each observation, then some method of combining the proximities matrices for different measures must be found. Usually the matrices are summed in a combined matrix. *For example:* given the following scores.

	X1	X2
O1	25	11
O2	33	11
O3	34	13
O4	35	18

The two proximities matrices resulting from squared Euclidean distance that result could be summed to produce a combined distance matrix.

Notes

	O1	O2	O3	O4
O1	0	64	81	100
O2	64	0	1	4
O3	81	1	0	1
O4	100	4	1	0

+

	O1	O2	O3	O4
O1	0	0	4	49
O2	0	0	4	49
O3	4	4	0	25
O4	49	49	25	0

=

	O1	O2	O3	O4
O1	0	64	85	149
O2	64	0	5	53
O3	85	5	0	26
O4	149	53	26	0

Note that each corresponding cell is added. With more measures there are more matrices to be added together.

This system works reasonably well if the measures share similar scales. One measure can overwhelm the other if the measures use different scales. Consider the following scores.

	X1	X2
O1	25	11
O2	33	21
O3	34	33
O4	35	48

The two proximities matrices resulting from squared Euclidean distance that result could be summed to produce a combined distance matrix.

	O1	O2	O3	O4
O1	0	64	81	100
O2	64	0	1	4
O3	81	1	0	1
O4	100	4	1	0

+

Notes

	O1	O2	O3	O4
O1	0	100	484	49
O2	100	0	144	729
O3	484	144	0	225
O4	1369	729	225	0

=

	O1	O2	O3	O4
O1	0	164	485	153
O2	164	0	145	733
O3	565	145	0	226
O4	1469	733	226	0

It can be seen that the second measure overwhelms the first in the combined matrix.

For this reason the measures are optionally transformed before they are combined. For example, the previous data matrix might be converted to standard scores before computing the separated distance matrices.

	X1	X2	Z1	Z2
O1	25	11	-1.48	-1.08
O2	33	21	.27	-.45
O3	34	33	.49	.30
O4	35	48	.71	1.24

The two proximities matrices resulting from squared Euclidean distance that result from the standard scores could be summed to produce a combined distance matrix.

	O1	O2	O3	O4
O1	0	3.06	3.88	4.80
O2	3.06	0	.05	.19
O3	3.88	.05	0	.05
O4	4.80	.19	.05	0

+

	O1	O2	O3	O4
O1	0	.40	1.90	5.38
O2	.40	0	.56	2.86
O3	1.9	.56	0	.88
O4	5.38	2.86	.88	0

=

Notes

	O1	O2	O3	O4
O1	0	3.46	5.78	10.18
O2	3.46	0	.61	3.05
O3	5.78	.61	0	.93
O4	10.18	3.05	.93	0

The point is that the choice of whether to transform the data and the choice of distance metric can result in vastly different proximities matrices.

4. Multidimensional Scaling

A statistical technique that measures objects in multidimensional space on the basis of respondents' judgments of the similarity of objects.

5. Multivariate Analysis of Variance (MANOVA)

A statistical technique that provides a simultaneous significance test of mean difference between groups for two or more dependent variables.

4.5 DISCRIMINATED ANALYSIS

Discriminant analysis is the regression based statistical technique that is used in determining the particular classification or group for an item of data or an object belongs to on the basis of its characteristics or essential features. It differs from group building techniques such as cluster analysis in that the classifications or groups to choose from must be known in advance.

Purposes of Discriminant Analysis

Discriminant Analysis undertakes the same task as multiple linear regressions by predicting an outcome. However, multiple linear regressions is limited to cases where the dependent variable on the Y axis is an interval variable so that the combination of predictors will, through the regression equation, produce estimated mean population numerical Y values for given values of weighted combinations of X values.

Assumptions of Discriminant Analysis

The major underlying assumptions of discriminant analysis are:

i) The observations are a random sample.

ii) Each predictor variable is normally distributed.

iii) Each of the allocations for the dependent categories in the initial classification is correctly classified.

iv) There must be at least two groups or categories with each case belonging to only one group so that the groups are mutually exclusive and collectively exhaustive.

v) Each group or category must be well defined, clearly differentiated from any other group(s) and natural. Putting a median split on an attitude scale is not a natural way to form groups. Partitioning quantitative variables is only justifiable if there are easily identifiable gaps at the points of division.

vi) The attribute is used to separate the groups should discriminate quite clearly between the groups so that group or category overlap is clearly non-existent or minimal.

vii) Group sizes of the dependent should not be grossly different and should be at least times the number of independent variables. *Notes*

4.6 FACTOR ANALYSIS

Factor analysis attempts to identify underlying variables, or factors, that explain the pattern of correlations within a set of observed variables. Factor analysis is often used in data reduction to identify a small number of factors that explain most of the variance that is observed in a much larger number of manifest variables. Factor analysis can also be used to generate hypotheses regarding causal mechanisms or to screen variables for subsequent analysis.

Factor analysis is used to analyze large numbers of dependent variables to detect certain aspects of the independent variables affecting those dependent variables without directly analyzing the independent variables. It enables an analyst to reduce the number of elements to be studied and to observe how they are interlinked. Factor analysis techniques are used in constructing factor models.

Factor analysis is based on a model that supposes that correlations between pairs of measured variables can be explained by the connections of the measured variables to a small number of non-measurable but meaningful variables, which are termed factors.

The factor analysis procedure offers a high degree of flexibility:

(i) Seven methods of factor extraction are available.

(ii) Five methods of rotation are available, including direct oblimin and promax for nonorthogonal rotations.

(iii) Three methods of computing factor scores are available and scores can be saved as variables for further analysis.

Meaning of Factor Analysis

Factor analysis is a statistical technique that can uncover relationship patterns underlying hundreds of interacting phenomenon such as changes in interest rates, inflation and/or oil prices.

Objectives of Factor Analysis

The aims of factor analysis are to:

(i) Identify the number of factors;

(ii) Define the factors as functions of the measured variables;

(iii) Study the factors which have been defined.

4.7 ANOVA

Analysis of Variance (ANOVA) is a collection of statistical models and their associated procedures, in which the observed variance in a particular variable is partitioned into components attributable to different sources of variation. In its simplest form ANOVA provides a statistical test of whether or not the means of several groups are all equal, and therefore generalizes t-test to more than two groups. Doing multiple two-sample t-tests would result in an increased chance of committing a type I error. For this reason, ANOVAs are useful in comparing two, three or more means.

An important technique for analyzing the effect of categorical factors on a response is to perform an Analysis of Variance. An ANOVA decomposes the variability in the response variable amongst the different factors. Depending upon the type of analysis, it may be important to determine: (a) which

Notes factors have a significant effect on the response, and/or (b) how much of the variability in the response variable is attributable to each factor.

Statgraphics Centurion provides several procedures for performing an analysis of variance:

1. ***One-Way ANOVA -*** used when there is only a single categorical factor. This is equivalent to comparing multiple groups of data.
2. ***Multifactor ANOVA -*** used when there is more than one categorical factor, arranged in a crossed pattern. When factors are crossed, the levels of one factor appear at more than one level of the other factors.
3. ***Variance Components Analysis -*** used when there are multiple factors, arranged in a hierarchical manner. In such a design, each factor is nested in the factor above it.
4. ***General Linear Models -*** used whenever there are both crossed and nested factors, when some factors are fixed and some are random, and when both categorical and quantitative factors are present.

One-Way ANOVA

A one-way analysis of variance is used when the data are divided into groups according to only one factor. The questions of interest are usually: (a) Is there a significant difference between the groups and (b) If so, which groups are significantly different from which others? Statistical tests are provided to compare group means, group medians, and group standard deviations. When comparing means, multiple range tests are used, the most popular of which is Tukey's HSD procedure. For equal size samples, significant group differences can be determined by examining the means plot and identifying those intervals that do not overlap.

Multifactor ANOVA

When more than one factor is present and the factors are crossed, a multifactor ANOVA is appropriate. Both main effects and interactions between the factors may be estimated. The output includes an ANOVA table and a new graphical ANOVA from the latest edition of Statistics for Experimenters by Box, Hunter and Hunter (Wiley, 2005). In a graphical ANOVA, the points are scaled so that any levels that differ by more than exhibited in the distribution of the residuals are significantly different.

Variance Components Analysis

A Variance Components Analysis is most commonly used to determine the level at which variability is being introduced into a product. A typical experiment might select several batches, several samples from each batch and then run replicates tests on each sample. The goal is to determine the relative percentages of the overall process variability that is being introduced at each level.

Assumptions of ANOVA

The analysis of variance has been studied from several approaches, the most common of which use a linear model that relates the response to the treatments and blocks. Even when the statistical model is nonlinear, it can be approximated by a linear model for which an analysis of variance may be appropriate.

(1) The model is correctly specified.

(2) The εij's are normally distributed.

(3) The εij's have mean zero and a common variance, σ^2.

(4) The εij's are independent across observations.

Notes

With multiple populations, detection of violations of these assumptions requires examining the residuals rather than the Y-values themselves.

Illustration - 1

The following are measurements of performance obtained after training 4 groups by different methods:

Method 1:	17	19	18	15	21	19	16	14
Method 2:	21	23	20	19	19			
Method 3:	20	16	21	17	19	16	16	
Method 4:	13	15	16	17	13	16		

Find out whether there is a significant overall differences between these 4 groups in terms of their performance after training (α = 0.05).

Solution:

Let, the null hypothesis be that different methods of training do not result difference in performance after training.

1	2	3	4
17	21	20	13
19	23	16	15
18	20	21	16
15	19	17	17
21	19	19	13
19		16	16
16		16	
14			

By coding of data (i.e., add, subtract, multiply or divide all observations by a number), can simplify the task. Let us subtract 15 from all observations, we get

1	2	3	4
2	6	5	-2
4	8	1	0
3	5	6	1
0	4	2	2
6	4	4	-2
4		1	1
1		1	
-1			

$T_1 = 19 \quad n_1 = 8$

$T_2 = 27 \quad n_2 = 5$

Notes

$T_3 = 20 \quad n_3 = 7$

$T_4 = 0 \quad n_4 = 6$

$T = 66 \quad N = 26$

Correction factor = $\frac{T^2}{N}$ where T = total of all observations

= no. of all observations

$$\frac{66^2}{26} = 167.54$$

Sum of squares between samples:

$$SSB = \sum \frac{T_j^2}{n_j} - \frac{T^2}{N}$$

$$= \left(\frac{19^2}{8} + \frac{27^2}{5} + \frac{20^2}{7} + \frac{0^2}{6}\right) - 167.54$$

$$= (45.125 + 145.8 + 57.14 + 0) - 167.54 = 80.525$$

Sum of squares withinsamples:

$$SSW = \sum X_{ij}^2 - \frac{\sum T_j^2}{n_2}$$

$$= (2^2 + 4^2 + 0^2 + 6^2 + 4^2 + 1^2 + 1^2 + 6^2 + 8^2 + 5^2 + 4^2 + 4^2 + 5^2 + 1^2 + 6^2 +$$

$$2^2 + 4^2 + 1^2 + 1^2 + 2^2 + 0^2 + 1^2 + 2^2 + 2^2 + 1^2) - \left(\frac{19^2}{8} + \frac{27^2}{5} + \frac{20^2}{7} + \frac{0^2}{6}\right)$$

$$= 338 - 248.07 = 89.93$$

ANOVA Table

Sources of squares (SS)	Sum of Freedom	Degrees of (df) MS = SS/df	Mean square	F-ratio variation
Between samples	80.525	(k – 1) = (4 – 1) = 3	$\frac{80.525}{3} = 26.84$	$\frac{26.84}{4.09} = 6.56$
Within samples	89.930	(n – k) = (26 – 4) = 22	$\frac{89.930}{22} = 4.09$	
Total	170.455	(n – 1) = (26 – 1) = 25		

F-ratio calculated = 6.56

F-ratio from table for $v_1 = 3$ and $v_2 = 22$ at 5% level of significance is 3.05

Since, $F_{calculated} > F_{table}$, to reject the null hypothesis, which means there is a significant overall difference between 4 groups in terms of performance after training.

Illustration - 2 *Notes*

Three methods are used in the production process test. At 5% level of significance test whether the three methods can be considered to be equivalent as far as output are concerned.

Method I	70	72	75	80	53	
Method II	100	110	108	112	120	107
Method III	60	65	57	84	87	73

Solution:

Let the null hypothesis be that there is no significant difference between the three methods.

Method	I	II	III
	70	100	60
	72	110	65
	75	108	57
	80	112	84
	53	120	87
	107	73	
Total	350	657	426

Correction factor $= \frac{T^2}{N}$

where, T - sum of all observations

N - no. of observations

Here, $T_1 = 350 \quad T_2 = 657. \; T_3 = 426, \; T = 1433$

$n_1 = 5 \quad n_2 = 6, \; n_3 = 6 \quad N = 17$

Sum of squares between samples:

$$SSB = \sum \frac{T_j^2}{n_j} - \frac{T^2}{N}$$

$$= \left(\frac{350^2}{5} + \frac{657^2}{6} + \frac{426^2}{6} \right) - \left(\frac{1433^2}{17} \right)$$

$= 24{,}500 + 71{,}941.5 + 30{,}246 - 1{,}20{,}793.5 = 5894$

Sum of squares within samples:

$$SSW = \sum X_{ij}^2 - \frac{\sum T_j^2}{n_2}$$

$$= (70^2 + 72^2 + 75^2 + 80^2 + 83^2 + 100^2 + 110^2 + 108^2 + 112^2 + 120^2 + 107^2 +$$

$$60^2 + 65^2 + 57^2 + 84^2 + 87^2 + 73^2 - \left(\frac{350^2}{5} + \frac{657^2}{6} + \frac{426^2}{6} \right)$$

$= 1{,}32{,}183 - 1{,}26{,}687.50 = 5195.5$

Notes

ANOVA Table

Sources of squares	Sum of (SS) Freedom d.f	Degrees of	Mean square MS	F-ratio variation
Between samples	5894	(k – 1) = (3 – 1) = 2	$\frac{5894}{2} = 2947$	$\frac{2954}{392.54} = 7.51$
Within samples	5495.5	(n – k) = (17 – 3) = 14	$\frac{5495.5}{14} = 392.54$	
	6279.00	(n – 1) = (17 – 1) = 16		

F-ratio calculated = 32.4

F-ratio from table for $v_1 = 2$ and $v_2 = 14$ at 5% level C1 significance = 3.74

Since $F_{calculated} > F_{table}$, reject the null hypothesis which means there is a significant difference between the three methods.

Illustration - 3

The following table gives the monthly sales in rupees (in thousands) of a certain firm in three different states of 4 different salesmen.

	Salesmen			
States	1	2	3	4
A	10	8	8	14
B	14	16	10	8
C	18	12	12	14

Test whether:

i. Sales between salesmen are significant

ii. Sales between states are significant.

Solution:

Two Way ANOVA:

Let, the first null hypothesis be that sales between salesmen are insignificant and second null hypothesis be that sales between states are in significant.

i.e., H_0 (1) : Sales between salesmen are insignificant

H_0 (2) : Sales between states are insignificant'

By coding the data, we can simplify the task. Let us subtract 12 from all the observations and we get:

	Salesmen				Total
	–2	–4	–4	2	–8
State	2	4	–2	–4	0
	6	0	0	2	8
Total	6	0	–6	0	0

Notes

Correction factor:

$$\frac{T^2}{N} = \frac{0^2}{12} = 0$$

Where, T - total of all samples

N - no. of samples

Total sum of squares:

$$SST = \sum X_{ij}^2 - \frac{T_j^2}{N}$$

$$= (2^2 + 2^2 + 6^2 + 4^2 + 4^2 + 0^2 + 4^2 + 2^2 + 0^2 + 2^2 + 4^2 + 2^2) - 0 = 120$$

Sum of squares between columns (i.e., between salesmen):

$$SSC = \sum \frac{T_j^2}{n_j} - \frac{T^2}{N} = \left(\frac{6^2}{3} + \frac{0^2}{3} + \frac{(-6)^2}{3} + \frac{0^2}{3}\right) - 0 = 24$$

Sum of squares between rows (i.e., between states):

$$SSR = \sum \frac{T_i^2}{n_i} - \frac{T^2}{N} = \left(\frac{(-8)^2}{4} + \frac{0^2}{4} + \frac{8^2}{4}\right) - 0 = 32$$

Sum of squares of residual or error:

SS_{res} = SST – (SSC + SSR) = 120 – (24 + 32) = 64

ANOVA Table

Sources of variation	Sum of squares (SS)	Degrees of Freedom	Mean square	F-ratio
Between samples	24	(c – 1) = (4 – 1) = 3	$\frac{24}{3} = 8$	$\frac{10.67}{8} = 1.33$
Between States	32	(r – 1) = (3 – 1) = 2	$\frac{32}{2} = 16$	$\frac{16}{10.67} = 1.50$
Residual or error	64	(c – 1) (r – 1) = (3) (2) = 6	$\frac{64}{6} = 10.67$	
Total	120	(n – 1) = (12 – 1) = 11		

Note:

$$\text{F-ratio} = \frac{\text{Greater variance}}{\text{Smaller variance}}$$

Table values of F at 5% level of significance

$F_{(6, 3)} = 8.94$; $F_{(2, 6)} = 5.14$.

i) Calculated $F_{(6, 3)} = 1.33 <$ Table $F_{(2, 6)} = 8.94$.

Notes Hence, conclude that null hypothesis holds good and there is no significant difference between the salesmen.

ii) Calculated $F_{(2, 6)} = 1.5 <$ Table $F_{(2, 6)} = 5.14$

Hence null hypothesis is accepted and conclude the there is no significant difference between the states

Illustration - 4

The following table shows the lifetimes in hours of samples from three different types of television tables manufactured by a company. Determine whether there is a difference between the three types of significance level of 0.01.

Sample 1	407	411	409	
Sample 2	404	406	408	405	402
Sample 3	410	408	406	408	

Let, X = 406 be the change of scale

						T_i	T_i^2	T_i^2/x
S_1	1	5	3			9	81	27
S_2	–2	0	2	–1	–4	–5	25	5
S_3	4	2	0	2		8	64	16
						T = 12		48

$$CF = \frac{T^2}{N} = \frac{12^2}{12} = 12$$

$$SS = \Sigma\ \Sigma\ x_{ij}^2 - CF = 1^2 + 5^2 + 3^2 + 2^2 + 2^2 + 2^2 + 1^2 + 4^2 + 4^2 + 2^2 + 2^2 - 12$$

$$= 1 + 25 + 9 + 4 + 4 + 4 + 1 + 16 + 16 + 4 + 4 - 12 = 76$$

$$SSR = \Sigma \frac{T_i^2}{n} - CF$$

$$= 48 - 12 = 36$$

SSE = SS – SSR = 76 – 36 = 40

ANOVA Table

SV	SS	df	MS	F ratio
B/w rows	36	2	18	F = 4.0909
Error	40	9	44	

$F_{(2, 5)}$ Table value = 8.02

$\therefore F < F\alpha$

$\Rightarrow$ Accept H_0

i.e., there is no significant differences between the 3 samples.

Illustration - 5 *Notes*

A research company has designed three different systems to clear up oil spills. The following table contains the results, measured by how much surface area (in square meters) is cleared in 1 hour. The data were found by testing each method in several trials. Are the three systems equally effective? Use the 0.05 level of significance.

System A:	55	60	63	56	59	55
System B:	57	53	64	49	62	
System C:	66	52	61	57		

Solution:

Let, us change the origin

$X - 55$

							T_i	T_i^2	T_i^2/n
System A:	0	5	8	1	4	0	18	324	54
System B:	2	–2	9	–6	7		10	100	20
System C:	11	–3	6	2			16	256	64
							44		138

$$CF = \frac{T^2}{N} = \frac{44^2}{15} = \frac{1936}{15} = 129.07$$

$$SS = \Sigma\ \Sigma\ x_{ij}^2 - CF$$

$$= 25 + 64 + 1 + 16 + 4 + 4 + 81 + 36 + 49 + 121 + 9 + 36 + 4 - 129.07$$

$$= 450 - 129.7 = 320.93$$

$$= SSR = \frac{\Sigma T_i^2}{n} - CF$$

$$= 138 - 129.07 = 8.93$$

$$SSE = SS - SSk$$

$$= 320.93 - 8.93 = 312$$

ANOVA Table

SV	SS	df	MS	F ratio
B/w system	8.93	2	4.465	F = 5.823
w/n system	312	12	26	

Table value $F_\alpha = 19.43$

$\Rightarrow F < F_\alpha$

$\Rightarrow$ Accept H_0

i.e., There is no significant difference between the system

Notes

Illustration - 6

The following table shows the yields per acre of four different plant crops grown on lots treated with three different types of fertilizer. Determine at the 0.05 significance level whether there is a difference in yield per acre

i) due to the fertilizers and

ii) due to the crops

Table	Crop I	Crop II	Crop III	Crop IV
Fertilizer A	4.5	6.4	7.2	6.7
Fertilizer B	8.8	7.8	9.6	7.0
Fertilizer C	5.9	6.8	5.7	5.2

Solution:

	I	II	III	IV	T_i	T_i^2	T_i^2/n
A	4.5	6.4	7.2	6.7	24.8	615.04	153.76
B	8.8	7.8	9.6	7.0	33.2	1102.24	275.56
C	5.9	6.8	5.7	5.2	23.6	556.96	139.24
T_j	19.2	21	22.5	18.9	81.6		568.56
T_j^2	368.64	441	506.25	357.21			
$\frac{T_j^2}{n}$	122.88	147	168.75	119.07	557.7		

$$C.F = \frac{T^2}{N} = \frac{(81.6)^2}{12} = \frac{6658.56}{12} = 554.88$$

$$SS = \Sigma\Sigma x_{ij}^2 - CF = 577.96 - 554.88 = 23.08$$

$$SSR = \frac{\Sigma T_i^2}{n} - CF = 568.56 - 554.88 = 13.68$$

$$SSC = \Sigma\frac{T_j^2}{k} - CF = 557.7 - 554.88 = 2.82$$

$$SSE = SS - SSR - SSC = 23.08 - 13.68 - 2.82 = 6.58$$

ANOVA Table

SV	SS	df	MS	
B/w Rows	13.68	2	6.84	$F_1 = 6.218$
B/w Column	2.82	3	0.94	$F_2 = 1.170$
Residual error	6.58	6	1.10	

Notes

$F_{\alpha_1} = 5.14$

$F_{\alpha_2} = 8.94$

$F_1 > F_{\alpha_1}$ $F_2 < F_{\alpha_2}$

$\Rightarrow$ Reject H_0 :

There is a significant difference in yield due to fertilizers and there is no significant difference between the crops.

Illustration - 7

The following data are the out puts per day from three machines when operated by four mechanics.

Mechanics	Machine		
	A	B	C
1	44	48	38
2	37	40	36
3	45	38	32
4	40	44	44

Test whether:

(i) Mean productivity is same for machines.

(ii) Mean productivity is same for mechanics.

Solution:

A 2-way ANOVA technique will enable us to solve and answer the question asked.

Let us take null hypothesis that

i) There is no significant difference between the machines productivity.

ii) There is no significant difference between the mechanics productivity.

Let us code the data by subtracting 40 from all observations to simplify the task.

	Machines			Total
	4	8	– 2	10
Mechanics	– 3	0	– 4	– 7
	5	– 2	– 8	– 5
	0	4	4	8
Total	6	10	– 10	6

Correction factor:

$$\frac{T^2}{N} = \frac{6^2}{12} = 3$$

Where, T - total of all observations

N - No. of observations

Notes

Total sum of squares:

$$SST = \sum X_{ij}^2 - \frac{T_j^2}{N}$$

$$= (4^2 + 8^2 + 2^2 + 3^2 + 0 + 4^2 + 5^2 + 2^2 + 8^2 + 0^2 + 4^2 + 4^2) - 3 = 231$$

Sum of squares between columns (i.e., between machines):

$$SSC = \sum \frac{T_j^2}{n_j} - \frac{T^2}{N} = \left(\frac{6^2}{4} + \frac{10^2}{4} + \frac{(-10)^2}{4}\right) - 3 = 56$$

Sum of squares between rows (i.e., between mechanics)

$$SSR = \sum \left(\frac{T_i^2}{n_i}\right) - \frac{T^2}{N} = \left(\frac{10^2}{3} + \frac{(-7)^2}{3} + \frac{(-5)^2}{3} + \frac{8^2}{3}\right) = 76.33$$

Sum of squares of residual or error:

$$SS_{res} = SST - (SSC + SSR) = 231 - (56 + 76.33) = 98.67$$

ANOVA Table

Sources of squares SS	Sum of freedom d.f	Degree of	Mean squares	MS-ratio variation
Between machines	56	(c – 1) = 2	$\frac{56}{2} = 28$	$\frac{28}{16.45} = 1.7$
Between mechanics	76.33	(r – 1) = 3	$\frac{76.33}{3} = 25.44$	$\frac{25.44}{16.45} = 1.55$
Residual or error	98.67	(c – 1) (r – 1) = 6	$\frac{98.67}{6} = 16.45$	
Total	231	(n – 1) = 11		

Table values of F ratio at 5% level of significance:

$F_{(2, 6)} = 5.14$

$F_{(3, 6)} = 4.76$

(i) Calculated $F_{(2, 6)} = 1.7 <$ Table $F_{(2, 6)} = 5.14$.

Hence, null hypothesis is accepted i.e., there is no significant difference between machines which means the mean productivity is same for machines.

(ii) Calculated $F_{(3, 6)} = 1.55 <$ Table $F_{(3, 6)} = 4.76$.

Hence, null hypothesis is accepted i.e., there is no significant difference between mechanics which means the mean productivity is same for mechanics.

Illustration - 8

Set up an ANOVA table for the following information relating to three drugs testing to judge the effectiveness in reducing blood pressure for three different groups of people.

Notes

Amount of BP reduction in mm of mercury:

Group of people	Drug		
X	Y	Z	
A	14	10	11
	15	9	11
B	12	7	10
	11	8	11
C	10	11	8
	11	11	7

Do the drug act differently? Are the different group of people affected differently? Is the interaction term significant? Table $\alpha = 0.05$.

Solution:

As repeated values are given in the table, this is a case of 2 way ANOVA with interaction.

(Interaction is the measure of inter-relationship among two different classifications).

Let the null hypothesis be that:

i) There is no significant difference between drugs.

ii) There is no significant difference between groups of people.

iii) The interaction term is insignificant

	Drug			Total
Groups of people	14	10	11	70
	15	9	11	
	12	7	10	59
	11	8	11	
	10	11	8	58
	11	11	7	
Total	73	56	58	187

Correction factor:

$$\frac{T^2}{N} = \frac{187^2}{18} = 1942.72$$

Total sum of squares:

$$SSC = \sum X_{ij}^2 - \frac{T^2}{N}$$

$$= (14^2 + 15^2 + 12^2 + 11^2 + 10^2 + 9^2 + 10^2 + 9^2 + 7^2 + 8^2 + 11^2 + 11^2 + 11^2 + 11^2 + 10^2 + 11^2 + 8^2 + 7^2) - 1942.72$$

$$= 76.28$$

Notes

Sum of squares between columns (i.e., between drugs)

$$SST = \sum X_{ij}^2 - \frac{T^2}{N}$$

$$= (14^2 + 15^2 + 12^2 + 11^2 + 10^2 + 9^2 + 10^2 + 9^2 + 7^2 + 8^2 + 11^2 + 11^2 + 11^2 + 11^2 + 10^2 + 11^2 + 8^2 + 7^2) - 1942.72 = 76.28$$

Sum of squares between rows (i.e., between people):

$$SSR = \sum \frac{T_i^2}{n_i} - \frac{T^2}{N} = \left(\frac{70^2}{6} + \frac{59^2}{6} + \frac{58^2}{6}\right) - 1942.72 = 14.78$$

Sum of squares within samples:

$$SSW = \sum \left(X_{ij} - \overline{X}_w\right)^2 \text{ where } \overline{X}_w \text{ – mean within samples}$$

$$= (14 - 14.5)^2 + (15 - 14.5)^2 + (10 - 9.5)^2 + (9 - 9.5)^2 + (11 - 11)^2 + (11 - 11)^2 + (12 - 11.5)^2 + (11 - 11.5)^2 + (7 - 7.5)^2 + (8 - 7.5)^2 + (10 - 10.5)^2 + (11 - 10.5)^2 + (10 - 10.5)^2 + (11 - 10.5)^2 + (11 - 11)^2 + (11 - 11)^2 + (8 - 7.5)^2 + (7 - 7.5)^2$$

$$= 3.50$$

Sum of squares for interaction variation:

SSI = SST – (SSC + SSR + SSW) = 76.28 – (28.77 + (14.78 + 3.50) = 29.23

ANOVA Table

Sources of squares SS	Sum of freedom d.f	Degree of	Mean square Ms	F-ratio variation
Between Drugs	28.77	(c – 1) = 2	$\frac{28.77}{2} = 14.385$	$\frac{14.385}{0.389} = 36.9$
Between groups	14.78	(r – 1) = 2	$\frac{14.78}{2} = 7.390$	$\frac{7.390}{0.389} = 19.0$ of people
Interaction	29.23	17 – 2 – 2 – 9 = 4	$\frac{29.23}{4} = 7.308$	$\frac{7.308}{0.389} = 18.8$
Within samples	3.50	(n – rc) = 9	$\frac{3.5}{9} = 0.389$	(error)
Total			76.28	(n – 1) = 17

Table value of F-ratios at 5% level of significance $F_{(2, 9)} = 4.26$; $F_{(2, 9)} = 3.63$

i) Calculated $F_{(2, 9)} = 36.9 >$ Table $F_{(2, 9)} = 4.26$.

Hence, null hypothesis is rejected which means the drugs act differently.

ii) Calculated $F_{(2, 9)} = 19.0 >$ Table $F_{(2, 9)} = 4.26$.

Hence, null hypothesis is rejected which means the different groups of people are affected differently.

iii) Calculated $F_{(4, 9)} = 18.8 >$ Table $F_{(4, 9)} = 3.63$. *Notes*

Hence, null hypothesis is rejected which means the interaction term is significant.

Illustration - 9

Is the interaction variation significant in case of the following information concerning mileage based on different brands of gasoline and cars?

		Brands of gasoline			
		W	X	Y	Z
Cars	A	13	12	12	11
		11	10	11	13
	B	12	10	11	9
		13	11	12	10
	C	14	11	13	10
		13	10	14	8

Solution:

Correction factor $\frac{T^2}{N}$

T - Sum of all observations

N - No. of all observations.

	W	X	Y	Z	Total
A	13	12	12	11	93
	11	10	11	13	
B	12	10	11	9	88
	13	11	12	10	
C	14	11	13	10	93
	13	10	14	8	
Total	76	64	73	61	274

Here,

$T_1 = 76, T_2 = 64, T_3 = 73, T_4 = 61, T = 274$

$n_1 = 6, n_2 = 6, n_3 = 6, n_4 = 6, N = 24$

Correction factor $= \frac{T^2}{N} = \frac{274^2}{24} = 3128.17$

Total sum of squires:

$$SST = \sum X_{ij}^2 - \frac{T^2}{N}$$

$= (13^2 + 11^2 + 12^2 + 13^2 + 12^2 + 10^2 + 10^2 + 11^2 + 10^2 + 12^2 + 11^2 + 11^2 +$
$12^2 + 13^2 + 14^2 + 11^2 + 13^2 + 9^2 + 10^2 + 10^2 + 8^2) - 3128.17$

$= 3184 - 3128.17 = 55.83$

Notes

Sum of squares between columns (i.e., between bas oline):

$$SSC = \sum \frac{T_j^2}{n_j} - \frac{T^2}{N} = \left(\frac{76^2}{6} + \frac{64^2}{6} + \frac{73^2}{6} + \frac{61^2}{6} \right) - 3128.17$$

$$= 3,153.67 - 3,128.17 = 25.50$$

Sum of squares between rows (i.e., between cars):

$$SSR = \sum \frac{T_i^2}{n_i} - \frac{T^2}{N} = \left(\frac{93^2}{8} + \frac{88^2}{8} + \frac{93^2}{8} \right) - 3,128.17$$

$$= 3130.25 - 3128.17 = 2.08$$

Sum of squares within samples:

$$SSW = \sum \left(X_{ij} - \overline{X}_w \right)^2$$

$= (13 - 12)^2 + (11 - 12)^2 + (12 - 12.5)^2 + (13 - 12.5)^2 + (14 - 13.5)^2 + (12 - 11)^2 + (10 - 11)^2 + (10 - 10.5)^2 + (11 - 10.5)^2 + (10- 10.5)^2 + (12 - 11.5)^2 + (11 - 11.5)^2 + (11 - 11.5)^2 + (12 - 11.5)^2 + (13 - 13.5)^2 + (14 - 13.5)^2 + (11 - 12)^2 + (13 - 12)^2 + (9 - 9.5)^2 + (10 - 9.5)^2 + (10 - 9)^2 + (8 - 9)^2 = 12$

Sum of squares for interaction variation:

$SSI = SST - (SSC + SSR + SSW) = 55.83 - (25.50 + 2.08 + 12) = 16.25$

4.8 SUMMARY

Multivariate analysis is based in observation and analysis of more than one statistical outcome variable at a time. In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest.

Multivariate analysis (MVA) is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical outcome variable at a time. In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest.

Multivariate analysis techniques which can be conveniently classified into two broad categories viz., dependence methods and interdependence methods.

Multiple regression is the most commonly utilized multivariate technique. It examines the relationship between a single metric dependent variable and two or more metric independent variables.

Discriminant analysis is the regression based statistical technique that is used in determining the particular classification or group for an item of data or an object belongs to on the basis of its characteristics or essential features. It differs from group building techniques such as cluster analysis in that the classifications or groups to choose from must be known in advance.

Cluster analysis is a collection of statistical methods which identifies groups of samples that behave similarly or show similar characteristics. In common parlance it is also called look-a-like groups.

A statistical technique that measures objects in multidimensional space on the basis of respondents' judgments of the similarity of objects.

Notes

A statistical technique that provides a simultaneous significance test of mean difference between groups for two or more dependent variables.

Factor analysis attempts to identify underlying variables, or factors, that explain the pattern of correlations within a set of observed variables. Factor analysis is often used in data reduction to identify a small number of factors that explain most of the variance that is observed in a much larger number of manifest variables. Factor analysis can also be used to generate hypotheses regarding causal mechanisms or to screen variables for subsequent analysis.

Analysis of Variance (ANOVA) is a collection of statistical models and their associated procedures, in which the observed variance in a particular variable is partitioned into components attributable to different sources of variation.

4.9 SELF ASSESSMENT QUESTIONS

1. What is Multivariate Data Analysis? Discuss the Techniques of Multivariate Analysis.
2. Explain the Multiple Regression Analysis.
3. What is Discriminated Analysis? Explain the purposes and assumptions of Discriminant Analysis.
4. What do you mean by Factor Analysis? What are its objectives?
5. What is ANOVA? Explain the assumptions of ANOVA.

*****

Chapter

5

INTERPRETATION AND REPORT WRITING

Objectives

The objectives of this lesson are to:

- Importance of Interpretation
- Techniques of Interpretation
- Significance of Report Writing
- Steps in Writing Report
- Layout of the Research Report

Structure:

Notes

5.1 INTRODUCTION

Data analysis is the process of bringing order, structure and meaning to the mass of collected data. It is a messy, ambiguous, time consuming, creative and fascinating process. It does not proceed in a linear fashion; it is not neat. Data analysis is a search for answers about relationships among categories of data.

In the beginning the data is raw in nature but after it is arranged in a certain format or a meaningful order this raw data takes the form of the information. The most critical and essential supporting pillars of the research are the analysis and the interpretation of the data. With the help of the interpretation step one is able to achieve a conclusion from the set of the gathered data. Interpretation has two major aspects namely establishing continuity in the research through linking the results of a given study with those of another and the establishment of some relationship with the collected data. Interpretation can be defined as the device through which the factors, which seem to explain what has been observed by the researcher in the course of the Presentations, can be better understood.

5.2 MEANING OF INTERPRETATION

Interpretation is the process of making sense of numerical data that has been collected, analyzed and presented.

Interpretation is the act of explaining, reframing, or otherwise showing your own understanding of something. Interpretation provides a theoretical conception which can serve as a guide for the further research work.

Interpretation of the data has become a very important and essential process, mainly because of some of the following factors -

1. Enables the researcher to have an in depth knowledge about the abstract principle behind his own findings.
2. The researcher is able to understand his findings and the reasons behind their existence.
3. More understanding and knowledge can be obtained with the help of the further research.
4. Provides a very good guidance in the studies relating to the research work.

5.3 IMPORTANCE OF INTERPRETATION

Data interpretation refers to the implementation of processes through which data is reviewed for the purpose of arriving at an informed conclusion. The interpretation of data assigns a meaning to the information analyzed and determines its signification and implications.

The importance of data interpretation is evident and this is why it needs to be done properly. Data is very likely to arrive from multiple sources and has a tendency to enter the analysis process with haphazard ordering. Data analysis tends to be extremely subjective. That is to say, the nature and goal of interpretation will vary from business to business, likely correlating to the type of data being analyzed. While there are several different types of processes that are implemented based on individual data nature, the two broadest and most common categories are "quantitative analysis" and "qualitative analysis".

The interpretation of data is designed to help people make sense of numerical data that has been collected, analyzed and presented. Having a baseline method for interpreting data will provide your analyst teams a structure and consistent foundation. Indeed, if several departments have different approaches to interpret the same data, while sharing the same goals, some mismatched objectives can

Notes result. Disparate methods will lead to duplicated efforts, inconsistent solutions, wasted energy and inevitably time and money.

Interpretation is essential for the simple reason that the usefulness and utility of research findings lie in proper interpretation. It is being considered a basic component of research process because of the following reasons:

1. It is through interpretation that the researcher can well understand the abstract principle that works beneath his findings. Through this he can link up his findings with those of other studies, having the same abstract principle, and thereby can predict about the concrete world of events. Fresh inquiries can test these predictions later on. This way the continuity in research can be maintained.
2. Interpretation leads to the establishment of explanatory concepts that can serve as a guide for future research studies; it opens new avenues of intellectual adventure and stimulates the quest for more knowledge.
3. Researcher can better appreciate only through interpretation why his findings are what they are and can make others to understand the real significance of his research findings.

Finally interpretation is concerned with relationships within the collected data, partially overlapping analysis. Interpretation also extends beyond the data of the study to include the results of other research, theory and hypotheses. Thus, interpretation 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 and it also provides a theoretical conception which can serve as a guide for further researches.

5.4 TECHNIQUES OF INTERPRETATION

The technique of interpretation often involves the following steps:

1. Researcher must give reasonable explanations of the relations which he has found and he must interpret the lines of relationship in terms of the underlying processes and must try to find out the thread of uniformity that lies under the surface layer of his diversified research findings. In fact, this is the technique of how generalization should be done and concepts be formulated.
2. Extraneous information, if collected during the study, must be considered while interpreting the final results of research study, for it may prove to be a key factor in understanding the problem under consideration.
3. It is advisable, before embarking upon final interpretation, to consult someone having insight into the study and who is frank and honest and will not hesitate to point out omissions and errors in logical argumentation. Such a consultation will result in correct interpretation and, thus, will enhance the utility of research results.
4. Researcher must accomplish the task of interpretation only after considering all relevant factors affecting the problem to avoid false generalization. He must be in no hurry while interpreting results, for quite often the conclusions, which appear to be all right at the beginning, may not at all be accurate.

5.5 MEANING AND DEFINITIONS OF REPORT

Meaning of Report

A report may be defined as the presentation of tangible output of the efforts of the research. A research report starts with the statement of the issue on which the study was focused. It contains the

statement of the procedure adopted the stages covered during the research survey and the findings and conclusions arrived at. In fact, it is the statement and description of the significant facts that are necessary for an understanding of the conclusions drawn.

Notes

Definition of Report

Koontz and O'Donnell define report as, "a documentation in which by the purpose of providing information a specified problem is researched and analyzed and conclusions, thoughts and sometimes references are presented". In a nut shell, a business report is any factual, objective document that serves a business purpose.

5.6 RESEARCH WRITING

Report writing is an integral part of a research process. Research reports are written to communicate to the world at large the results of the research, field work and other activities. Research report is a concrete outcome of the research work undertaken by the researcher. The quality of research is judged by the quality of the writing and how well the importance of the findings are conveyed. Research carried out very scientifically revealing findings of great importance may not be of value if the same is not communicated effectively. In the context of business, the research report assumes importance as it is through reports, management gets information regarding the activities performed at various levels of the organization. The management takes decisions and controls various activities of the business on the basis of information provided through the business reports.

Research report writing is the oral or written presentation of evidence and the findings in such a way that it is readily understood and assessed by the reader and enables him to verify the validity of the conclusions. Research report writing is the culmination of the research investigation. It is at the stage of reporting that the researcher assembles the findings of the study, draws conclusions and evaluates his own findings. Report writing is the end product of research activity. It is highly skilled work; it is an interesting, fascinating, challenging, grueling and sometimes even exasperating experience.

Writing a research report is a technical activity that demands all the skills and patience of the researcher. It requires considerable thought, effort, patience and penetration and an overall approach to the problem, data and analysis. Also needed is firm control over language and great objectivity. A vast amount of planning and preparation is necessary for organising and writing the report. Perfection in a research report is achieved by continuous and persistent thought and creative and intelligent writing. Only hard and patient work on the facts, careful and critical assessment and intelligent planning in organizing the report can facilitate communication. There are no standard criteria for the organization of a report, popular or technical. They depend on each investigation, problem, the novelty or familiarity of the methods, nature and volume of facts, techniques of analysis and so on. No research project is complete without a report. The nature of the report is determined by the project itself and to whom it is addressed. Academic research is expected to produce lengthy reports, or theses, covering all aspects of the research and reporting on them in a precise and rather formal manner. But no matter what the size or formality of the report, it is reasonable to expect it to convey information on a fairly standard set of topics. First, it must say why the work was done, what events led up to it and what other work was found to be relevant. This is usually contained in the introduction, which should also include the precise statement of the objective and aims of the project.

Generally, there should be a section describing what work was done. This should cover the methods used, their selection and any problems experienced in their application. From this it is easy to move on to what was found out, or the results. In turn, these lead on to the conclusions, which are a statement of what the researcher deduced from the results, and then on to the recommendations, which set out what the researcher feels should be the action taken as a result of the conclusions. Writing is not an activity

Notes that can be allocated an odd half-hour whenever it is convenient. It requires sustained concentration. The amount of time needed to make real progress in your writing depends on the way you prefer to work. Most people find that it takes a day to write about 2,000 words. But we all work in different ways. Some people, once they get started, prefer to continue until they drop from exhaustion! Others like to set a strict timetable, devoting three or four hours a day to writing. Whichever category you fall into, make sure you have time for writing allocated in your diary. We have found that it is helpful to have blocks of time where writing can take place on successive days. This ensures a degree of continuity of ideas, which isn't easy to maintain if you keep having to 'think your way back' into your research.

5.7 SIGNIFICANCE OF REPORT WRITING

1. ***Decision Making Tool:*** Today's complex business organizations require thousands of information. A Reports provide the required information a large number of important decisions in business or any other area are taken on the basis of information presented in the reports. This is one of the great importance of report.
2. ***Investigation:*** Whenever there is any problem, a committee or commission or study group investigates the problem to find out the reason behind the problem and present the findings with or without the recommendation in the form of a report. It is another importance of report.
3. ***Evaluation:*** Large scale organizations are engaged in multidimensional activities. It is not possible for a single top executive to keep personal watch on what others are doing. So, the executive depends on reports to evaluate the performance of various departments or units.
4. ***Quick Location:*** There is no denying the fact that business executives need information for quick decision-making. As top executives are found to be busy for various purposes), they need vital sources of information. Such sources can be business reports.
5. ***Development of skill:*** Report writing skill develops the power of designing, organization coordination, judgment and communication.
6. ***Neutral presentation of facts:*** Facts are required to be presented in a neutral way; such presentation is ensured through a report as it investigates, explains and evaluates any fact independently.

7. ***Professional Advancement:*** Report also plays a major role in professional achievement. For promotion to the rank and file position, satisfactory job performance is enough to help a person. But for promotion to high level position, intellectual ability is highly required. Such ability can be expressed through the report submitted to higher authority.
8. ***Proper Control:*** Whether activities are happening according to plan or not is expressed through a report. So, controlling activities are implemented based on the information of a report.
9. ***A managerial Tool:*** Various reports make activities easy for the managers. For planning, organizing, coordinating, motivating and controlling, manager needs help from a report which acts as a source of information.
10. ***Encountering Advance and Complex Situation:*** In a large business organization, there is always some sort of labor problems which may bring complex situations. To tackle that situation, managers take the help of a report.

Notes

5.8 CHARACTERISTICS OF RESEARCH REPORT

The desirable features of a good report are listed below:

(i) A good research report should focus on the purpose of the study and the type of audience.

(ii) It should also have clarity, conciseness and coherence.

(iii) Right emphasis should be placed on the important aspects of the problem identified meaningful organization of paragraphs, sentences and smooth transition from one topic to next should be achieved by ensuring parallelism and specificity.

(iv) The report should be free of technical or statistical jargon if the same is addressed to audiences who may not understand.

(v) Care should be taken to avoid grammatical, spelling and typographical errors.

(vi) Assumptions made by the researcher should be clearly spelled out.

(vii) Operational definitions of words used with specific meaning should be given in the beginning of the report.

(viii) The report should be organized in a meaningful manner so as to enable smooth flow of information.

(xi) Ambiguity, multiple meanings and allusions should be avoided by choosing the right words and sentences.

(x) The report should adhere to the guideline.

5.9 PURPOSE OF REPORTS

Reports are prepared for the following purposes:

(i) A Report enables management to monitor the operations undertaken at various levels and control the same.

(ii) A written report serves as a guideline for future course of action. It enables to plan and organize things in an effective manner.

(iii) Feed back regarding the various aspects, controls and processes implemented in the organization can be obtained through the reports.

(iv) Information regarding specific problems or issues can be obtained by through reports.

(v) Information provided in reports enable decision making.

(vi) A report may also be prepared to convince the reader or to sell an idea. The report in this case would be more detailed and convincing as to how the proposed idea could add to the organization's value or the justification as to why it should be adopted.

(vii) A report may also be prepared to provide several alternative solutions or recommendations so as to compare the pros and cons and select the best course of action.

5.10 ESSENTIALS OF A REPORT

The following are the essentials of a good research report:

a) Style: Reports should be easy to read and understand. The style of the writer should ensure that sentences are succinct and the language used is simple, to the point and avoiding excessive jargon.

Notes

b) ***Layout:*** A good layout enables the reader to follow the report's intentions, and aids the communication process. Sections and paragraphs should be given headings and sub-headings. Bullet points are an option for highlighting important points in the report.

c) ***Accuracy:*** The report should be factually accurate. It should not mislead or misinform those for whom the report is prepared.

d) ***Clarity:*** The report should be clear without ambiguity with simple language used to express views.

e) ***Readability:*** Experts agree that the factors which affect readability the most are:

- Attractive appearance
- Non-technical subject matter
- Clear and direct style
- Short sentences
- Simple words

f) ***Review:*** The researcher should thoroughly review the report multiple times before preparing the final draft.

5.11 PRINCIPLES OF DRAFTING A RESEARCH REPORT

A research report to be useful must be prepared on a sound basis. Whatever is the nature and type of report it is necessary to observe certain fundamental principles of report preparation. They are as follows:

1. ***Purpose:*** A business research report must have a specific purpose which can be translated into more effective business management. Careful statement of purpose helps in the preparation of a well-directed report on which management can take the right decision. Proper consideration of purpose is necessary because: (i) reports form the basis for discussion of facts and recommendations; and (ii) they become the record for that phase of business activity.
2. ***Organization:*** It is necessary that the report should be well-planned and well-organized. The organizational plan of a report usually includes the following:
   a) The purpose, the information that it includes and the method used in collecting the data.
   b) The summary of the conclusions reached and the supporting details.
   c) The problem and focus areas.
   d) The recommendations made for action.
   e) Any Appendix to explain and support the contents and conclusion of the report.
3. ***Brevity:*** The report should be brief. Brevity in reporting is essential because (i) long reports are expensive to prepare, (ii) long reports are difficult to analyse, (iii) long reports are subject to criticism as they show inefficiency, and (iv) long reports tend to highlight irrelevant minor details and thus ignore major issues.
4. ***Clarity:*** It is essential that reports should observe the principle of clarity. For this purpose, it is necessary to use simple language. If new terms are used, they should be properly defined and explained so that confusion does not occur at any stage. Besides, the report will fail to convey its message if it has long and complicated sentences. Simple language is a great asset of any report.

Notes

5. ***Scheduling:*** Reports should be scheduled in a way that they can be prepared without undue burden on the staff and with sufficient time to do a good job. However the time interval between the collection of data and the finished report should not be long. If it is so the report may become obsolete and thus useless by the time it is completed.
6. ***Cost Effectiveness:*** It is necessary to have a cost-benefit analysis of a report. A report should not only cost the minimum but also give the maximum benefit. If the cost of preparation of a report is unusually high and its consequent benefit is low, it would not be worthwhile to prepare such a report.

5.12 BASIS OF REPORTS

Reports can be classified on the basis of source, frequency, target audience, length, subject dealt, function performed and intention.

1. Source

Source refers to the person/persons who initiated the report. Voluntary reports are prepared on own initiative and they require to be more comprehensive. The background of the subject should be more carefully planned. Authorized reports are those which are prepared as a response to a request.

2. Frequency

Routine or periodic reports are submitted on a recurring basis which may be weekly, monthly, daily etc. Some routine reports may be prepared in preprinted computerized form. Due to the routine nature of report it requires only less introduction than the special reports. Special reports are non recurring in nature and they present the results of specific one time studies or investigations.

3. Length

A short report differs from a long report in scope, research and duration. A long report examines the problem in detail and requires more extensive time and effort in preparation. On the contrary a short report may discuss only a module of a problem. A summary is a short report which gives a concise overview of a situation. It highlights the important details but does not include background material examples or specific details. A short report is suitable when the problem is well defined, is of limited scope and has a simple methodology.

4. Intent

Informal reports focuses on the facts and explains or educates the readers. Analytical report is designed to solve a problem by convincing readers that the conclusions and recommendations reached are justified based on the data collected. Information provided plays a supporting role in convincing the reader.

5. Function

The reports may be classified as informative and interpretative on the basis of function performed. Informative reports present facts pertinent to the issue or situation. Common types of informational reports include those for monitoring and controlling operations, statements of policies and procedures, compliance reports and progress reports. It may take the form of an operating or a periodic report. Operating reports provide managers with detailed information regarding all activities like sales, inventory, costs etc., Periodic reports describe the activities in a department during a particular period.

Interpretative also knows as analytical or investigative report analyses the facts and presents recommendations and conclusions. The report presents facts and persuades the reader to accept a

Notes stated decision, action or the recommendations detailed throughout the report. It may take the form of problem solving report providing the background information and analysis about the various options. Trouble shooting reports is a form of problem solving report which discusses the source of the problem, extent of damage done and solutions possible. A feasibility report is a problem solving report that studies proposed options to assess whether all or any one of them is sound.

6. Subject

The reports may be categorized as problem determining, fact finding, performance report, technical report etc. The problem determining report focuses on underlying problem or to ascertain whether a problem actually exists. Technical reports are concerned with presenting data on a specialized subject with or without comments.

7. Legality

Reports may be prepared to meet the government regulations. For example, A compliance report explains what a company is doing to conform to the government regulations. It may be prepared on an annual basis like the income tax returns, annual share holders report etc. Interim compliance reports can also be prepared to monitor and control the licenses granted by the government.

5.13 METHODS OF RESEARCH REPORT WRITING

1. Technical Method of Report Writing

In the technical report the main emphasis is onthe methods employed,assumptions made in the course of the study,the detailed presentation of the findings including their limitations and supporting data.

A general outline of a technical report can be as follows:

1. ***Summary of results:*** A brief review of the main findings just in two or three pages.
2. ***Nature of the study:*** Description of the general objectives of study, formulation of the problem in operational terms, the working hypothesis, the type of analysis and data required, etc.
3. ***Methods employed:*** Specific methods used in the study and their limitations. For instance, in sampling studies we should give details of sample design viz., sample size, sample selection, etc.
4. ***Data:*** Discussion of data collected their sources, characteristics and limitations. If secondary data are used, their suitability to the problem at hand is fully assessed. In case of a survey, the manner in which data were collected should be fully described.
5. ***Analysis of data and presentation of findings:*** The analysis of data and presentation of the findings of the study with supporting data in the form of tables and charts be fully narrated. This, in fact, happens to be the main body of the report usually extending over several chapters.
6. ***Conclusions:*** A detailed summary of the findings and the policy implications drawn from the results be explained.
7. ***Bibliography:*** Bibliography of various sources consulted be prepared and attached.
8. ***Technical appendices:*** Appendices be given for all technical matters relating to questionnaire, mathematical derivations, elaboration on particular technique of analysis and the like ones.
9. ***Index:*** Index must be prepared and be given invariably in the report at the end.

The order presented above only gives a general idea of the nature of a technical report; the order of presentation may not necessarily be the same in all the technical reports. This, in other words, means

Notes

that the presentation may vary in different reports; even the different sections outlined above will not always be the same, nor will all these sections appear in any particular report.

It should, however, be remembered that even in a technical report, simple presentation and ready availability of the findings remain an important consideration and as such the liberal use of charts and diagrams is considered desirable.

2. Professional Method of Report Writing

The professional report is one which gives emphasis on simplicity and attractiveness. The simplification should be sought through clear writing, minimization of technical, particularly mathematical, details and liberal use of charts and diagrams. Attractive layout along with large print, many subheadings, even an occasional cartoon now and then is another characteristic feature of the professional report. Besides, in such a report emphasis is given on practical aspects and policy implications.

1. ***The findings and their implications:*** Emphasis in the report is given on the findings of most practical interest and on the implications of these findings.
2. ***Recommendations for action:*** Recommendations for action on the basis of the findings of the study is made in this section of the report.
3. ***Objective of the study:*** A general review of how the problem arise is presented along with the specific objectives of the project under study.
4. ***Methods employed:*** A brief and non-technical description of the methods and techniques used, including a short review of the data on which the study is based, is given in this part of the report.
5. ***Results:*** This section constitutes the main body of the report wherein the results of the study are presented in clear and non-technical terms with liberal use of all sorts of illustrations such as charts, diagrams and the like ones.
6. ***Technical appendices:*** More detailed information on methods used, forms, etc. is presented in the form of appendices. But the appendices are often not detailed if the report is entirely meant for general public.

There can be several variations of the form in which a popular report can be prepared. The only important thing about such a report is that it gives emphasis on simplicity and policy implications from the operational point of view, avoiding the technical details of all sorts to the extent possible.

5.14 STEPS IN WRITING REPORT

Report writing is a process which should be carried out at various stages. The goal of the writing process is to generate clear, effective documentation so as to enable the audience to act. The writing process is performed in the following three stages:

Step-1: Pre-Writing Stage

Pre-writing stage involves planning the task for writing the reports. It includes collection of all the relevant information and deciding the steps to be followed. It involves three tasks viz., analyzing the situation, investigation and adaptation.

i) Analyzing the Situation

A thorough analysis of the situation should be made to decide whether the situation merits writing report. Sometimes, it may be enough to make a phone call or email or conduct a meeting. If situation warranties writing reports, then the next step is to decide the type of report needed. It may be informational

Notes or an analytical report. In case of informational report the specific purpose of the report should be defined and report type that is appropriate should be selected. For analytical reports, the problem should be defined before stating the purpose of the report.

a) Problem Definition

The problem addressed by a report may be defined by the person who authorizes the report or by the researcher himself. The readers of the report should be convinced about the existence of the problem. This requires persuasive writing method. The problem definition can be made by answering the following issues:

a) What needs to be ascertained?

b) When did the problem start?

c) What is the importance of the issue?

d) Who are involved in the situation?

e) Where is the trouble located?

Problem factoring can also be done which involves breaking down the perceived problem into a series of logical, connected questions that try to identify the cause and effect. Speculating the cause for a problem leads to forming a hypothesis. A hypothesis is a potential explanation that needs to be tested. Dividing the problem and framing the hypothesis based on the available evidence enables to tackle even the most complex situation.

b) Developing the Statement of Purpose

The problem statement enables to define what is going to be investigated whereas the statement of purpose defines, why the report is prepared. The purpose statement can be started with an infinite phrase. *For example,* "To analyse the reasons for fall in the share price". Using an infinite phrase encourages to take control and decide where the starting should be made. The purpose statement should be highly specific and the same should be checked with the person who has authorized the report. The confirmed statement can be used as the basis for developing the preliminary outline of the report.

c) Developing a Preliminary Outline

Preliminary outline establishes the framework for the report preparation. It provides a visual diagram of the report to be prepared the order in which the discussion will take place and the details to be included. The preliminary outline might look different from the final outline of the report, however, the outline guides the research effort and acts as a foundation for organizing and composing the report. Since, outline is only a working draft it will be revised and modified in the further steps. The grammatical parallelism should be ensured among the various items presented at the same level.

d) Preparing the Work Plan

Most of the reports have a firm deadline to be met. A carefully prepared work plan ensures that quality reports are produced on the schedule. If the work plan is prepared for the researcher himself, it can be prepared in an informal manner. However in case of proposal, a detailed work plan should be prepared which becomes the basis for the contract if the proposal is accepted. A formal work plan might include the following elements:

- Statement of the problem which enables to stay focused on the core problem.
- The purpose statement which describes the plan to be accomplished with the report and the boundaries of the work.

Notes

- A description of the product that will arise out of the investigation. Many times the report may be the only outcome.
- A review of the project assignments, schedules and resource requirements indicating who will be responsible for what, when the task will be completed and how much will be the investigation cost.
- Plan for following up after delivering the report should be explained.

ii) Investigating Information

Information should be gathered for writing reports on various perspectives such as the specific company information, trends, issues, product, events, related literature, micro and macro economic perspectives of the problem taken for the study etc. The following tasks should be completed in investigating the information:

a) Identify the right questions.

b) Find and access primary and secondary sources of information.

c) Evaluate and finalize the resources.

d) Process the information.

e) Analyze the data.

f) Interpret the findings.

iii) Adapting the Report

A good relationship with the audience should be maintained in order to ensure that the report is audience centered. A report will be successful, only if it focuses on the audience. The focus on the audience can be maintained by following the criteria given below:

a) The attitude should be followed and the report should answer the audience questions and solve their problems.

b) Emphasize should be given to the positive aspects. If the report recommends a negative action, the facts should be stated and the recommendation should be made positively.

c) Credibility should be established by building audience trust. The trust can be gained by researching the topic from all perspectives and documenting the findings with credible sources.

d) The report should address the audience in a polite manner. The audience's respect should be earned by being courteous, kind and tactful.

e) Bias- free language should be used. Unethical and embarrassing blunders in language related to gender, race, age and disability should be avoided.

f) The style and language of the report should reflect and adapt to the image of the organization.

Selecting the Appropriate Channel and Medium

The right medium should be selected for conveying the report. It may be in the form of oral presentation, e-format, email, letter or a formal written report. Written reports are opted to convey complex lengthy information which needs to be presented in a structural format and is needed for further reference. If immediate feedback is needed, oral reports are appropriate. Electronic reports are stored in electronic media and may be distributed on disk, attached to an email or posted on the website. When compared to paper based reports, electronic reports enable to save cost and space. It also enables faster distribution as well as include multimedia features. The appropriate channel should be chosen based on the requirement of the audience and the researcher.

Notes

Step-2: Writing Stage

Actual composing of report should be preceded by organizing the material collected and arranging the same in a logical order that will meet the audience needs. The format, length, order and structure of the report should be decided before drafting the report.

i) Deciding the Format and Length

Four options are available to format the report viz.,

a) Pre-printed form: The pre-printed form is a fill-in-the-blank type report. These reports are relatively short and deals with only routine information.

b) Memo: It is a short informal report distributed within the organization. It has headings and visual aids and if the length exceeds ten pages it is called as memo report.

c) Letter: It includes the normal parts of a letter and in addition may have headings, footnotes, tables and illustrations. It is commonly used for reports of five or fewer pages that are directed to outsiders.

d) Manuscript: It is commonly used for reports that require a formal approach and may range from a few pages to several hundred pages. The prefatory parts and supplementary parts will have more number of pages as the length of the report increases.

If the report is more of routine nature the flexibility in deciding the format and length is much lesser. The length of the report is often decided by the subject matter and the type of relationship with the audience.

ii) Choosing the Approach

The researcher may choose a direct or indirect approach in writing the report. A direct report starts with the main idea first and thereby saves time and enables easier understanding of the report. The direct approach is used when the audience is more receptive or open-minded. The report starts with the findings, conclusions and recommendations. This method is mostly followed in the business reports. The indirect approach withholds the main idea until the latter part of the report. If the audience is skeptical or hostile then the complete findings and all supporting details should be presented before presenting the findings and conclusions.

iii) Structuring the Reports

Structure of the report deals with the way in which the ideas will be subdivided and developed. The structure of the report depends on its type viz., informational, analytical, investigative etc. The reports may follow topical organization i.e., arranging materials according to one of the following topics:

a) Materials may be organized on the basis of the importance of the subject matter. The most important topic may be presented first and least important at the end of the report If the report is presented on the process, then it should be arranged in sequential order of the process.

b) If events are reported in the study then the same should be reported in the chronological order.

c) If a physical object is discussed in a report ,then the same should be discussed from left to right, top to bottom, outside to inside.

d) If the report is organized on the basis of geographical area occupied it has to be organized on the basis of the regions under study viz., city, district, state or country.

iv) Composing Reports

Once the decision regarding the length, approach and structure is made, composing the first draft can begin. The writing task should start with preparation of a final outline. This would act as a guide to

Notes

the writing process and will also enable to critically evaluate the selection and order of information to be presented in the report. The outline preparation may lead to rephrasing the points and tone of the report. While composing the reports, the researcher should only concentrate on drafting the message and not editing the same which is done at a later stage. While composing the reports the following points should be kept in mind:

a) Formal language should be used in writing reports. Obsolete and pompous language should be avoided. Similarly using big words, trite expressions and overly complicated sentences to impress others should not be attempted.

b) Correct words should be used in report. The words selected should convey the meaning clearly, specifically and dynamically. The words that are familiar to the audience should be chosen. Clinches and jargons can be used only when it is understood by the audiences for whom the report is directed to.

c) Due attention should be paid to the grammatical accuracy of the content delivered as it affects the image of the researcher.

d) The report should concentrate on presenting the facts

e) The arguments for or against any aspect should be constructed in a rational manner

f) Active or passive voice should be used appropriately in composing the reports. Active voice can be used to emphasize the subject and to produce shorter sentences. Passive voice is mostly used in research reports as it is prepared in a formal situation.

g) Consistent time perspective should be ensured in the report i.e., the report should be in past or present tense. The chronological sequence should also be adapted in presenting the events.

h) The reader's perspective of the report might be different from the researcher's perspective. Hence a preview or road map of the report structure should be included. This will clarify the reader regarding the overall organization and flow of report.

Step-3: Post-Writing Stage

A research report will undergo many drafts before finalization. The report is revised many times to ensure the content, organization, style and tone, readability, clarity and conciseness. Post-writing stage involves revision of the report, production and proofreading the same.

i) Revision

Revision takes place during and after preparation of the first draft. It is an ongoing process that occurs throughout the writing process. Revision involves searching for the best way of saying something, probing for right words, rephrasing sentences, reshaping, juggling elements etc. Revision is a never ending process, however, every research report has a deadline and hence schedules should be drawn and met.

a) Evaluating content, organization, style and tone

During the process of evaluating the content the following aspects should be given due attention:

- Accuracy of the information presented.
- Relevance of the facts presented to the concerned audience.
- Completeness of information provided to suit the audience needs.
- Balance between specific and general information.

Notes

While reviewing the organization the following aspects should be considered:

- Logical order in presentation and coverage of all main points to be ensured.
- Assuring that the main theme is given more space and prominence.

More attention should be given to the introduction and conclusion of the report as it has major impact on the audience. The words used should be of right style and tone. The opening statements should be relevant, interesting and enticing the reader to read further. It should establish the subject, purpose and organization of the information in the report. The conclusion should be reviewed to ensure that it summarizes the main idea and leaves the reader with a positive impression.

b) Reviewing for Readability and Scannability

Readability depends on choice of words, sentence length, sentence structure, organization and the physical appearance of the message. The following techniques can be used to ensure readability:

- Variety in sentence structure makes the information presented more appealing to the reader. While long sentences should be avoided, use of too many short sentences should not be attempted. Average sentence length should consist of 20 words or fewer.
- Important ideas can be presented in the forms of list. Lists are effective tools for highlighting and simplifying the information presented. It provides the reader with clues, simplifies the complex subjects, highlights the main point, breaks up the pages visually and ease the skimming process for busy readers
- Heading is a brief title that provides clues to the reader about the content of the section that follows. Heading should be properly used to attract the readers attention and to divide the material into shorter sections.

c) Editing for Clarity and Conciseness

Clarity in information presented should be ensured. Clarity prevents confusion. If the information is presented in a cluttered manner it can be interpreted by the reader in several ways which is not intended by the researcher. The following aspects should be considered to ensure clarity:

- Long sentences should be broken up. Connecting too many clauses with and should be avoided.
- Too many hedging statements should be avoided.
- Parallelism should be ensured among related ideas. It can be achieved by repeating the pattern in words, phrases, clause or entire sentences.
- Long noun sequences should be avoided.

To ensure conciseness, every word in the report should be carefully scrutinized. Words which do not serve any function should be eliminated. Every long word should be replaced with a short word. Conciseness should be ensured by way of deleting unnecessary words and phrases, shortening words and phrases and by eliminating redundancies. Use of computer enables to revise the report in a much faster and efficient manner. Word processor helps to add, delete and move text with functions like cut and paste, search and replace, replace all options etc. Autocorrect feature enables to store words commonly misspelled or mistyped along with correct spelling. History of revisions made can also be fetched by enabling the software options. Three advanced software functions viz., spell checker, thesaurus and grammar checker enables to create an effective report.

ii) Producing the Report

Producing the report involves adding elements such as graphics and designing the page layout to give the report an attractive and contemporary appearance. The appearance of the report meets the eyes of the reader first and plays an important role in creating impression.

Effective design should have the following elements: **Notes**

a) Consistency should be ensured throughout the report in terms of the margins, type face, type size, spacing, paragraph indent, borders, columns etc.

b) Proper balance should be maintained between the text, graphs and spacing.

c) Too much of highlighting, decorative touches and design elements should be avoided.

d) Attention should be paid to details like heading should not be separated from the information, avoiding narrow columns and the like.

e) Variety of design elements such as line justification, typefaces, styles etc can be used to create a professional and interesting report, but it should be kept in mind that too many design elements might confuse the reader.

iii) Proof Reading

While proofreading attention should be devoted to spelling, punctuations and typographical errors. Credibility of the researcher is affected by the attention paid to details, mechanics and form. Researcher should carefully check the grammar usage, language errors, missing material, design errors and typographical errors. Design errors include elements like wrong typeface, wrong type style, misalignment etc. typographical errors include uneven spacing between lines and words, heading at the bottom of a page, incorrect hyphenation, non confirmation with the guidelines provided etc. Attention should also be paid to overall format. Routine documents will have only fewer elements to check. Longer and more complex documents have many components that need checking and more time should be devoted for the same.

5.15 LAYOUT OF THE RESEARCH REPORT

Research report has a set of identifiable components. The components of report should be decided keeping in mind the needs of the audience. It should also focus on the requirements of the audience and the problem identified for the study. Generally, a Research Report consists of three parts: A. Preliminaries B. Text C. Reference Materials.

Each of the main parts may consist of several subsections as shown below:

A. Preliminaries

They do not make a direct contribution to the identified research problem. However, it assists the reader in using the research report. The subsections in preliminaries are discussed below:

i) Letter of Transmittal

A letter of transmittal is required in case of formal relationship between the researcher and audience at whom the report is directed at. It is mostly used in case of carrying out research work for a specific client or for an outside organization. The letter should highlight the authorization for conducting the project and the specific instructions provided to complete the study. It should also state the purpose and scope of the study. The letter of transmittal is not necessary if the report is aimed at authorities within the organization.

ii) Title page

Most of the organizations have their own form of title page for the research report and the same should be complied with. The title page generally has the following information:

- Title of the report.
- Month and year of submission.

Notes

- For whom and by whom the report is submitted.
- If project report is submitted for award of degree, the degree for which the dissertation is submitted for should be listed.
- The best practice is to centre the title of the report on the page in upper case letters. If the title is too long to be centered on one line, an inverted pyramid principle should be followed without splitting word or phrases.

iii) Preface

The preface may include the writer's purpose in conducting the study, a brief resume of the background, scope, purpose, general nature of the research for which the report is prepared and the acknowledgments. A preface can be prepared only after the final form of the report is ready. In the case of dissertation submitted for award of degree the preface is omitted and instead an acknowledgment is added. An Acknowledgment recognizes the persons to whom the writer is indebted for guidance and assistance during the study. It also credits the institution for providing funds to conduct the study and for granting permission to use the facilities. The researcher should acknowledge the assistance provided by all concerned honestly in a simple and tactful manner.

iv) Executive Summary

An executive summary is a brief account of the research study. It is a report in miniature covering all aspects in the body of the report but in a brief manner. It provides an overview of the research problem identified and highlights the important information such as the sampling design, data collection method used, results of data analysis, findings and recommendation. The length of the executive summary will normally be two to three pages. The executive summary is usually written after the completion of the report. Sometimes a synopsis or an abstract may be included instead of the executive summary, however, they are not one and the same. Executive summaries are more comprehensive than a synopsis. Although executive summaries are not designed to replace the report, in some cases it may be the only thing that may be read by the audience. By contrast, a synopsis is only a brief overview of the entire report and may either highlight the main points as they appear in the report or simply inform the reader as to the content of the report. The purpose of synopsis is to entice the audience to read the report.

v) Table of Contents

The table of contents includes the major divisions of the report. It indicates in outline form the topics included in the report. The purpose of a table of contents is to provide an analytical overview of the topics included in the report together with the sequence of presentation. Depending on the length and complexity of the report, the content page may show only the top two or three levels of headings or only the first level headings. Care should be exercised to see that the titles of chapters and captions of subdivisions within chapters correspond exactly with those included in the body of the report. Page numbers for each of the divisions are given. The relationship between major divisions and minor subdivisions should be shown by using capital letters and indentation or by using numeric system.

The table of contents is prepared after the other parts of report have been typed, so that the page numbers can be given. If they are fewer than four visual aids, the same may be listed in the table of contents, but if there are more than four visual aids, a separate list of illustration should be prepared.

Some guidelines for writing table of contents are given below:

(a) The page should be titled as Table of Contents or Contents.

(b) The name of each section should be worded and formatted as it appears in the text.

(c) The table of contents should not be underlined as they may overwhelm the words.

(d) Use only the page number on which the section starts.

(e) The margins should be set such that the page numbers align on the right. *Notes*

(f) Not more than three levels of headings should be given.

(g) The leaders, a series of dots can be used to connect the words to page numbers.

vi) List of Tables

The researcher should prepare a list of tables compiled under the heading "LIST OF TABLES". It should be centered on a separate page by itself. Two spaces below the headings 'Table number', 'Title', and 'Page number' should be given. Table number should be aligned to the left, page number should be aligned to the right and the title should be centered.

vii) List of Illustrations

The list of figures should be prepared in the same form as the list of tables. The page is headed as LIST OF FIGURES. The list includes the Figure number, title of the figure and page number.

B. Text

The text is the most important part of a report as it is in this section that the researcher presents the facts. The researcher should devote the greater part of attention to the careful organization and presentation of his findings or arguments. The text may be organized as introduction, methodology and as many chapters as required for presenting the report.

i) Introduction

The introduction prepares the reader for the report by describing the various parts; background, problem statement and research objectives.

ii) Background

The background information provides a prelude to the reader of the research report. It may be the preliminary results of exploration, the survey or any other source. The secondary data from the literature review could also be highlighted. Previous research, theory or situations that led to the research issue can be discussed. The literature should be organized, integrated and presented in a logical manner. The background includes definitions, assumptions etc. It provides the needed information to understand the remainder of the research report. It contains information pertinent to the management problem or the situation that led to the study. It may be placed before the problem statement.

iii) Problem Statement

The problem statement contains the need for the research project. The problem is usually represented by a management question. It is followed by a more detailed set of objectives.

The guidelines are given below:

(a) It gives basic facts about the problem.

(b) It specifies the causes or origin of the problem.

(c) It explains the significance of the problem.

iv) Research Objectives

The research objectives provide the purpose of the research. The objectives may be research questions and associated investigative questions. In correlational study the hypothesis statements are included. Hypothesis are declarative statements describing the relationship between two or more variables. They state clearly the variables of concern the relationships among them and the target group being studied.

Notes

v) Methodology

The methodology contains the following sections:

(a) The type of the study viz., descriptive, exploratory should be mentioned in the methodology.

(b) The sampling design explains the sample method and sample size.

(c) The data collection method is described in the report.

(d) The tools used for analysis of data should be explained.

vi) Findings and Conclusion

The findings section is generally the longest section of the report. The objective is to explain the data. Wherever needed data should be supplemented with charts and graphs. The conclusion serves the important function of tying together the whole thesis or assignment. The recommendations of the study are also presented in this section. It provides an idea about the corrective actions. In academic research, the suggestions broaden the understanding of the subject area. In applied research, the recommendation includes the guidelines for further managerial actions. Several alternatives may be provided with further justifications. The conclusion should leave the reader with the impression of completeness and of positive gain.

C. Reference Material

The reference material includes bibliography, appendix and index.

i) Bibliography

The bibliography follows the main body of the text and is a separate but integral part of a thesis, preceded by a division sheet or introduced by a centered capitalized heading A bibliography is a list of secondary sources consulted while preparing the report.

ii) Appendix

The appendix contains information of a subordinate, supplementary or highly technical nature that the researcher does not want to place in the body of the report. Each appendix should be clearly separated from the other and should be listed in the table of contents.

The guidelines for preparing appendix are:

- Each appendix item should be referred in the appropriate place in the body of the report. In short reports, the page number numbers may be continued in sequence from the last page of the body.
- In long reports, a separate pagination system can be followed as the appendixes are often identified as Appendix A, Appendix B and so on. The page numbers can be given along with the appropriate letter: A-1, A-2, B-1, B-2.
- The illustrations in the appendix may continue with the sequence started in the body of the report.

iii) Index

The index should be included after bibliography and the appendix. It acts as a good guide to the reader. Index may be prepared both as subject index and author index. The subject index gives the names of the subject-topics or concepts along with the number of pages on which they have appeared or discussed in the report. The author index gives similar information regarding the names of the authors. The index should always be arranged alphabetically. An index is not required for an unpublished thesis or a report. If the findings in the report are subsequently published as a book, monograph or bulletin an index is necessary.

Standardized Elements of a Report *Notes*

The Different Elements of a Report are:

1. Cover and the Title Page
2. Introductory Pages
   a) Foreword
   b) Preface
   c) Acknowledgment
   d) Table of Contents
   e) Lists of tables and illustrations
3. Main Text
   a) Introduction
   b) Statement of findings and recommendations
   c) Results
   d) Indications of the results
   e) Summary
4. Reference Section
   a) Appendices
   b) Bibliography
   c) Glossary (if required)
   d) Index

5.16 TYPES OF REPORT

Generally Reports can be grouped into the following:

1. Information Reports

Information reports are meant to understand the existing situation in terms of parameters of business, economy, technology, market or research scenario. The information reports may also provide the background for subsequent decision reports and research reports.

2. Decision Reports

Decision reports are structured sequentially to reflect comprehensively the iterative thinking process of decision-makers. Decision reports adopt the problem-solving approach. It involves the identification of problems. It is important to correctly identify and define the problem area. To do so the following issues need to be taken into consideration:

i) Situational Analysis.

ii) Cause and Effect.

iii) Major and Minor Issues.

iv) Scope of Decision-Making Domain.

Notes

3. Research Reports

The basic purpose of research reports is to develop the C-BOK (Common Body of Knowledge) further in the area of the researched subject. They provide the pointers for new hypothesis and more rigorous methods for deeper investigation into the research area.

The components are as follows:

i) Survey of literature to identify gaps in C-BOK
ii) Nature and scope of the research study
iii) Significance and usage of the research study
iv) Hypothesis to be tested
v) Data collection methodology
vi) Design of experiments
vii) Analysis and interpretation of data and experimentation
viii) Major findings
ix) Conclusion
x) Recommendation for further research.

4. Technical Reports

In the Technical report the major emphasis is on:

a) Research methodology utilised
b) Assumptions which were used for the frame of the study
c) The detailed presentation of the research outcome including the various limitations along with the supporting research data.

A general outline of a technical research report can be as follows:

i) Executive Summary of results: A brief review of the main findings just in two or three pages.

ii) Nature of the study: This section of the research report contains the brief description of the research objectives, research problem formulation and the research hypothesis. It also emphasizes on the data that needs to be acquired and type of analysis that would be performed on such data.

iii) Methods/techniques employed: Specific research methods and statistical techniques used in the research study are enumerated in this section. *For example,* if samples are obtained details on sampling design together with any limitations thereof should also be presented in this section.

iv) Data: In this section, details of type of data collected, sources of data and limitations of data collection are listed. In case any secondary data is utilized, the suitability of such data to the research at hand is also described. If the research study utilizes a survey methodology how this survey was conducted along with the data collected should also be detailed in this section.

v) Analysis of data and presentation of findings: In this section, data analysis together with various pictorial depiction of this analysis is presented. It also contains the research findings along with the supporting data in the form of tables, charts and exhibits.

vi) Conclusion: A detailed summary of the findings and the policy implications drawn from the results should be explained.

vii) Bibliography: Bibliography of various sources consulted should be prepared and attached.

Notes

viii) ***Technical Appendices:*** Appendices in a research report typically contain items like design of questionnaires mathematical formulation or description of a specific research technique utilized in the said research.

ix) ***Use of charts, graphs and diagrams:*** In a technical report simple presentation and ready availability of the findings remains an important consideration and as such the liberal use of charts and diagrams is considered desirable.

5.17 IMPORTANT PARTS OF A REPORT

1. The Preliminaries

The following aspects should be highlighted in the first part of the research report:

- Title of the report.
- Acknowledgement
- Preface
- Foreword
- Contents
- List of tables and illustrations

2. The Abstract

This is probably the most important part of the report because it may be the only part that some will read. It is a short summary of the complete project report. This enables those who are not sure whether they wish to read the complete report to make an informed decision. For those who intend to read the whole report, the abstract prepares them for what is to come. An abstract should contain four short paragraphs with the answers to the following questions:

- What are my research questions and why are they important?
- How did I go about answering the research question(s)?
- What did I find out?
- What conclusions do I draw regarding my research question(s)?

Smith (1991) lists five characteristics of a good abstract:

- It should be short. Try to keep it to a maximum of two sides of an A4-size paper sheet.
- It must be self-contained. Since it may be the only part of your report that some people see, it follows that it must summaries the complete content of your report.
- It must satisfy your reader's needs. Your reader must be told about the problem or central issue that the research addresses and the method adopted to solve it. It must also contain a brief statement of the main results and conclusions.
- It must have the same emphasis as the report, with the consequence that the reader should gain an accurate impression of the report's content from the abstract.
- It should be objective, precise and easy to read. The project report contents page should give you the outline structure for the abstract. Summarizing each section should give you an accurate resume of the content of the report. Do ensure that you stick to what you have written in the report. The abstract is not the place for elaborating any of your main themes. Be objective. You will need to write several drafts before you eliminate every word that is not absolutely necessary. The purpose is to convey the content of your report in as clear and brief a way as possible.

Notes

- Writing a good abstract is difficult. The obvious thing to do is to write it after you have finished the report. We suggest that you draft it when you start writing the report so that your story line is abundantly clear in your mind. You can then amend the draft when you have finished the report so that it conforms to the five principles above.

3. Research Design

The researcher should highlight the research design of the project. The researcher should answer the following questions:

- What is its basic design?
- What are the methods adopted to collect data?
- How is the study carried out?
- Is it an experimental/survey/historical data research method?
- If the study is an experimental one, what are the experimental manipulations?
- What type of questionnaire/interview/observations is used?
- If measurements were based on observation, what instructions are given to the observers?
- Who are the subjects?
- How many of them have been selected?
- How have they been selected?
- How have they been selected?
- Are the research instruments reliable?
- Do the research instruments have validity?

All these questions, when properly answered, can be used to estimate the probable limits of the findings' generalisability. The researcher has to take proper care to develop a well-planned research design, which is free from errors and limitations. To ensure the reliability and validity of the tools and instruments, a pilot study can be conducted to verify its strengths and utility.

4. Analysis of Data

Here, the researcher has to highlight the type of statistical analysis adopted to analyse the data. The analysis can be listed from simple descriptive analysis to complex multivariate analysis.

5. The Results

Once the analysis is over, the results can be depicted in a tabulated form, with appropriate illustrations. A detailed presentation of the findings of the study is a major part of the researchreport. These can be supported in the form of tables and charts together with a validation of results. Since it comprises the main body of the report, it generally extends over several chapters. It is advisable to project summarized results rather than raw data. All the results should be presented in logical sequence and split into readily identifiable sections. All relevant results must find a place in the report. All the results of the report should address the research problems stated earlier in the report, illustrating whether the results support or reject the hypothesis. But ultimately the researcher must rely on his own judgement in deciding the outline of his report.

Interpretation of results

- To find the relationships among the variables that are studied and observing the commonality, uniqueness, diversity etc. among them.

Notes

- To observe the role of extraneous variables. How they affect the various phenomena studied.
- To ensure validity; the results can be cross-checked with others through consultation.
- To consider all the relevant factors affecting the problem before generalising it to the whole population.

The prime tasks of interpretation are to bring to the surface the gist of the findings. A researcher should explain why the findings are so, in objective terms. He should try to bring out the principles involved in the observations. He can also make reasonable prediction. On the basis of interpretation of an exploratory study, a new hypothesis can be formulated for experimental research. During interpretation, unconnected, isolated facts should not be discarded, but should be explained properly. Interpretation leads to the establishment of some explanatory concepts arising out of the connection between the underlying processes and principles, and the observed facts from a working model. A researcher's task is to identify and disengage such principles and processes. Interpretation can also provide a theoretical conception, which can be the basis of further researcher and new knowledge. Thus, continuity in research can be established and the quest for knowing the unknown can be sustained.

Prerequisites for good Interpretation

i) While drawing inferences from the analysis of data, the researcher has to ensure that the inferences are free from any biases and mistakes that may arise due to both subjective and objective factors. This can be minimised by: checking whether (a) the data are appropriate, trustworthy and adequate for drawing inferences b) the data reflect good homogeneity and (c) proper analysis has been done through statistical methods.

ii) The researcher should also check for personal bias (subjective element) while interpreting the results. There are so many pitfalls that have to be avoided while observing and interpreting the results. Some of them are: stereotyping (conforming with existing results), preoccupation with set results, projecting his own views on the subject, snap judgements, lack of appreciation for others' feelings, prejudicial treatment and so on. The researcher must remain vigilant about all such things so that false generalisations may not take place. He should be well-equipped with statistical measures and must know their correct use for drawing inferences concerning his study.

iii) The researcher must always keep in view that the task of interpretation is very much intertwined with analysis and cannot be separated. He should take precautions about there liability of data, computational checks, validation and comparison of results.

iv) The researcher should also pay attention to the hidden factors underlying the results. Broad generalizations should be avoided because the coverage may be restricted to a particular time, area and conditions.

v) Originality and creativity are critical in interpreting the results. While linking the relationship between theoretical orientation and empirical observation, the researcher has to make use of his originality and creativity in developing concepts and models. He must pay special attention to this aspect while engaged in the task of interpretation.

6. Summary

It is a generally practice to conclude the report with a very brief summary. In business reports, it is called an executive summary. Here, all the aspects of the research report are given in capsule form.

7. Reference Material

The listing of reference material comes at the end of any research report. Appendices with all technical data such as questionnaires, sample information, mathematical derivations etc. should be included

Notes at the end. The bibliography, listed in alphabetical order, should be added in the last section. Similarly, the researcher has to prepare an index (an alphabetical listing of names, places and topics along with the page numbers in the book or report in which they are mentioned). That should invariably be given at the end of the report.

8. Other Considerations

i) ***Use of quotations:*** The appropriate use of quotations will enrich the effective presentation of research reports. Quotations should be placed within quotation marks and double-spaced. In case the quotation is lengthy, it can be typed in single space and indented at least half an inch to the right of the normal text margin.

ii) ***Punctuation and abbreviations:*** The researcher has to take care to check punctuation marks such as commas, full stops, colons, semicolons etc. these punctuation marks can be checked and verified in listing the bibliography, references, citations, documentations etc. For example, in listing the reference, the author's name is followed by a comma. After the comma, the title of the book is given; the article (such as 'a', 'an', 'the' etc.) is omitted and only the first word, proper nouns and adjectives are capitalized. A comma follows the title. Information concerning the edition is given next. This entry is followed by a comma. The place of publication is then stated; it may be mentioned in an abbreviated form.

5.19 PRECAUTIONS IN PREPARING REPORT

A report is an important way of communicating research findings to others. A good research report is one that does this task efficiently and effectively. Hence, the following precautions must be taken while preparing it:

1. While determining the length of the report, one should keep in mind the fact that it should be long enough to cover the subject but short enough to maintain interest. In fact, report writing should not be a means to learning more and more about less and less.
2. Abstract terminology and technical jargon should be avoided. The report should be able to convey the matter as simply as possible. In other words, this means that reports should be written in an objective style in simple language, avoiding expressions such as 'it seems', 'there may be' and the like.
3. Readers are often interested in acquiring quick knowledge of the main findings and as such the report must make the findings readily accessible. For this purpose, charts, graphs and statistical tables may be used for the various results in the main report in addition to summaries of important findings.
4. The layout of the research should be well thought out. It must be appropriate and in accordance with the objective of the research problem.
5. The report should be free from grammatical mistakes and must be prepared strictly according to the rules of composition of research reports such as the use of quotation marks, footnotes, documentation, punctuation and use of abbreviations in footnotes and the like.
6. The report must present a logical analysis of the subject matter. It must reflect a structure wherein the different pieces of analysis relating to the research problem fit well.
7. A research report should show originality and should necessarily be an attempt to solve some intellectual problem. It must contribute to the solution of a problem and must add to the store of knowledge.

Notes

8. Towards the end, the report must also state the policy implications of the problem under consideration. It is usually considered desirable for a report to make a forecast of the probable future of the subject concerned and indicate the kind of research that still needs to be done in that particular field.
9. Appendices should be enlisted for all the technical data in the report.
10. Bibliography of sources consulted is a must for a good report.
11. An index is also considered an essential part of a good report and as such must be prepared and appended at the end.
12. The report must have an attractive appearance. It should be neat and clean, whether typed or printed.
13. Calculated confidence limits must be mentioned and the various constraints experienced in conducting the research study stated.
14. The objective of the study, the nature of the problem, the methods employed and the technique of analysis adopted must all be stated at the beginning of the report in the form of an introduction.

5.20 SUMMARY

Data analysis is the process of bringing order, structure and meaning to the mass of collected data. It is a messy, ambiguous, time consuming, creative and fascinating process. It does not proceed in a linear fashion; it is not neat. Data analysis is a search for answers about relationships among categories of data.

Interpretation is the process of making sense of numerical data that has been collected, analyzed and presented.

Interpretation is the act of explaining, reframing, or otherwise showing your own understanding of something. Interpretation provides a theoretical conception which can serve as a guide for the further research work.

Data interpretation refers to the implementation of processes through which data is reviewed for the purpose of arriving at an informed conclusion. The interpretation of data assigns a meaning to the information analyzed and determines its signification and implications.

A report may be defined as the presentation of tangible output of the efforts of the research. A research report starts with the statement of the issue on which the study was focused. It contains the statement of the procedure adopted the stages covered during the research survey and the findings and conclusions arrived at. In fact, it is the statement and description of the significant facts that are necessary for an understanding of the conclusions drawn.

Koontz and O'Donnell define report as, "a documentation in which by the purpose of providing information a specified problem is researched and analyzed and conclusions, thoughts and sometimes references are presented". In a nut shell, a business report is any factual, objective document that serves a business purpose.

Report writing is an integral part of a research process. Research reports are written to communicate to the world at large the results of the research, field work and other activities.

Research report writing is the oral or written presentation of evidence and the findings in such a way that it is readily understood and assessed by the reader and enables him to verify the validity of the conclusions. Research report writing is the culmination of the research investigation.

Notes

Source refers to the person/persons who initiated the report. Voluntary reports are prepared on own initiative and they require to be more comprehensive. The background of the subject should be more carefully planned. Authorized reports are those which are prepared as a response to a request.

Routine or periodic reports are submitted on a recurring basis which may be weekly, monthly, daily etc. Some routine reports may be prepared in preprinted computerized form. Due to the routine nature of report it requires only less introduction than the special reports. Special reports are non recurring in nature and they present the results of specific one time studies or investigations.

Structure of the report deals with the way in which the ideas will be subdivided and developed. The structure of the report depends on its type viz., informational, analytical, investigative etc.

The research objectives provide the purpose of the research. The objectives may be research questions and associated investigative questions. In correlational study the hypothesis statements are included. Hypothesis are declarative statements describing the relationship between two or more variables. They state clearly the variables of concern the relationships among them and the target group being studied.

Producing the report involves adding elements such as graphics and designing the page layout to give the report an attractive and contemporary appearance. The appearance of the report meets the eyes of the reader first and plays an important role in creating impression.

Information reports are meant to understand the existing situation in terms of parameters of business, economy, technology, market or research scenario. The information reports may also provide the background for subsequent decision reports and research reports.

5.21 SELF ASSESSMENT QUESTIONS

1. Give the meaning of Interpretation. Explain the Importance of Interpretation.
2. Discuss the techniques of Interpretation.
3. What is reports? Explain the purpose of report.
4. Explain the essentials of a report.
5. What are the characteristics of research report.
6. Discuss the types of reports.
7. Explain the principles of drafting a research report.
8. Discuss the basis of reports.
9. Explain the components/format/layout of the research report.
10. Briefly explain the methods of research report writing.
11. What is report writing? Explain the steps in report writing.

*****