Semester V

L T P C

15PAU503 CORE -RESEARCH METHODOLOGY 5 - - 5

COURSE OBJECTIVES

- Research methodology represents the concept of the research process, research design, sampling techniques, and testing of hypothesis by applying parametric and non parametric test.
- This paper presents the various research tools and techniques in order to facilitate the research work.

LEARNING OUTCOME

- Make the students understand the concept of research methodology
- Enlighten the student's knowledge in sampling techniques
- Impart students knowledge in writing a good research report

Unit I Business Research : Meaning – Scope and Significance - Utility of business research - qualities of good researcher - Types of research – research Process – Identification, Selection and formulation of research problem – Hypothesis - Research design

Unit II Sampling : Methods and Techniques – sample size – sampling error – field work and data collection. Tools of data collection – interview schedule – questionnaire – observation – interview and mailed questionnaire – pilot study and final collection of data

Unit III Measurement and Scaling Techniques : Processing And Analysis of data – Editing – Coding – Transcription and tabulation - Statistical tools used in research – Interpretation and Report writing- types and contents and styles of report – steps in drafting reports.

Unit IV Measures of central tendency : Mean, Median and Mode – Standard Deviation – Correlation – Regression Models

Unit V Test of significance : 't'- test – large sample and 'f' test, test of significance for attributes – Analysis of Variance – Business Forecasting – Exponential Smoothing – Chi – Square test

Note : The question paper shall cover 60% theory and 40% problem

Text Book:

1. C.R Kothari, (2015), *Research Methodology*, Wishwa Prakashan Publications, New Delhi.

Reference Book:

1. Ravilochanan, (2013), *Research Methodology*, Maegham Publication, chennai.
2. S.P Gupta, (2014), *Statistical Methods and Techniques*, Sultan Chand & sons, New Delhi.
3. Rao K.V, (2012), *Research Methods for Management and Commerce*, sterling publishers pvt Ltd.,
4. Aamarchand, (2013), *Research Methods in commerce*, Emerld Publication, Chennai.

KARPAGAM ACADEMY OF HIGHER EDUCATION
(Deemed to be University Established Under Section 3 of UGC Act 1956)
Coimbatore – 641 021.

LECTURE PLAN
DEPARTMENT OF COMMERCE

Staff Name :Mr.T.Thirunavukkarasu
Subject Name : Research Methodology **Subject Code** : 15PAU503
Academic Year : 2017- 2018 **Class** : III B.Com - PA
Semester : V **Batch** : 2015-2018

LECTURE PLAN -UNIT-1

S.No	Lecture Duration (Hrs)	Topics to be covered	Support Material
1	1	Business Research – meaning, definition Scope of research Objectives of research	T.P:1-2
2	2	Utility of business research Significance of research, qualities of good research	R1.P:1.1 W1
3	1	Types of research	T.P:2-6
4	1	Types of research	T.P:2-6
5	2	Process of research Problems faced by the researcher in India	T.P:10-19, T.P:21-23
6	2	Research problem identification and selection of research problem Steps in research problem	T.P:24-28
7	2	Hypothesis- meaning, basic concepts relating to hypothesis	T.P:184 R1,P:8.1-8.8
8	1	Research design- meaning, need and importance	T.P:31-32
9	2	Types of research design	T.P:35-38
10	1	Recapitulation and Discussion of important questions	
Total No of Hours planned for Unit- I			**15**

Text Book:

T: C.R Kothari, (2015), *Research Methodology*, Wishwa Prakashan Publications, New Delhi.

Reference Book:

R1: Ravilochanan, (2013), *Research Methodology*, Maegham Publication, chennai.

Websites:
W1: https://www.slideshare.net /Meaning-Nature-and-Scope-utility-significnce

UNIT-II

S.No	Lecture Duration (Hrs)	Topics to be covered	Support Material
1	1	Sampling – meaning, important concept relating to sampling, sources of sampling	T.P:55-56 R1.P:21.2
2	2	Criteria for selecting a sample procedure Characteristics of good sampling, determination of sample size	T.P:57-58 R1.P:21.10
3	2	Sampling Types –Probability Sampling	T.P:58-60
4	2	Non probability sampling	T.P:60-63
5	1	Sampling error – types of error	R1.P:21.12
6	2	Filed work and data collection sources of data : primary data – observation	T.P:153-155
7	2	Interview method – types merits and demerits Questionnaire and schedule method of data collection – merits and demerits, mailed questionnaire	R1.P:11.4 T.P:100-102
8	2	Meaning of pilot study Final collection of data	W2 T.P:113-114
9	1	Recapitulation and Discussion of important questions	
Total No of Hours planned for Unit- II			**15**

Text Book:

T: C.R Kothari, (2015), *Research Methodology*, Wishwa Prakashan Publications, New Delhi.

Reference Book:

R1: Ravilochanan, (2013), *Research Methodology*, Maegham Publication, chennai.

Websites:

W2: www.wikipedia/pilotstudy.com

UNIT-III

S.No	Lecture Duration (Hrs)	Topics to be covered	Support Material
1	1	Measurement – meaning and sources Techniques for developing measurement tools	T.P:69-75 R1.P:20.2
2	2	Scaling – meaning, importance Scaling techniques	T.P:76-78 T.P:82-85
3	2	Data Processing And Analysing – Editing types	T.P:122-124
4	2	Concept of coding and classifications Tabulation meaning principles of tabulation	R1.P:14.3 T.127-129
5	2	Statistical tools used in research	W3
6	2	Interpretation – meaning, need, techniques Precautions	T.P:244-245 P.P:346-347
7	2	Report writing – meaning an significance	T.P.:347-348 R1.P.:16.1-16.13
8	2	Contents, styles of research report Steps in report writing, qualities	T.P:353-359
9	1	Recapitulation and Discussion of important questions	
Total No of Hours planned for Unit- III			**16**

Text Book:

T: C.R Kothari, (2015), *Research Methodology*, Wishwa Prakashan Publications, New Delhi.

Reference Book:

R1: Ravilochanan, (2013), *Research Methodology*, Maegham Publication, chennai.

Websites:

W3: www.arcjournals.org

UNIT-IV

S.No	Lecture Duration (Hrs)	Topics to be covered	Support Material
1	1	Measures of central tendency – introduction Meaning of mean, median, mode	R2.P:7.4-7.5
2	2	Mean – geometric mean problems Harmonic mean - problems	R2.P:7.5-7.10
3	2	Median – merits and demerits Calculation of median in individual series	R2.P:7.10-7.12
4	2	Calculation of median in Discrete series and continuous series	R2.P:7.12-7.19
5	2	Mode – merits and demerits, Calculation of mode in individual series, Discrete series and continuous series	R2.P:7.20-7.29
6	1	Standard deviation – calculation of standard deviation	R2.P:8.21-8.23
7	1	coefficient of variance	R2.P.8.24-8.29
8	2	Correlation – meaning, difference between correlation and regression Calculation of correlation	R2.P:10.5-10.10
9	1	Regression models – meaning, types	R2.P:11.3-11.4
10	2	Calculation of regression	R2.P:11.7-11.10
11	1	Recapitulation and Discussion of important questions	
Total No of Hours planned for Unit- IV			**16**

Text Book:

T: C.R Kothari, (2015), *Research Methodology*, Wishwa Prakashan Publications, New Delhi.

Reference Book:

R2: S.P Gupta, (2014), *Statistical Methods and Techniques*, Sultan Chand & sons, New Delhi.

UNIT-V

S.No	Lecture Duration (Hrs)	Topics to be covered	Support Material
1	1	Test of significance meaning and assumptions	T.P:195-197
2	1	t- test assumptions ad problems will be worked out	T.P:198-204
3	2	Paired t test – problem will be worked out	T.P:198-206
4	1	F test meaning and assumptions F test problem will be worked out	T.P:206-215
5	1	Test of significance of attributes Analysis of variance – meaning, problems	T.P:210-215 T.P:218-225
6	1	The concept of business forecasting Exponential smoothing meaning and methods	W4, W5
7	1	Chi square test- conditions, steps, problems will be worked out	R2.P:954-1002 T.P:237-250
8	1	Chi square test – problem will be worked out	T.P:237-250
9	1	Recapitulation and Discussion of important questions	
10	1	**Revision -Discussion of previous ESE question papers**	
11	1	**Discussion of previous ESE question papers**	
12	1	**Discussion of previous ESE question papers**	
Total No of Hours planned for Unit- V			**13**

Text Book:

T: C.R Kothari, (2015), *Research Methodology*, Wishwa Prakashan Publications, New Delhi.

Reference Book:

R1: Ravilochanan, (2013), *Research Methodology*, Maegham Publication, chennai.

R2: S.P Gupta, (2014), *Statistical Methods and Techniques*, Sultan Chand & sons, New Delhi.

Websites:
W4: www.wiki/exponential smoothing
W5: www.business/forecasting.com

UNIT I

SYLLABUS

Business Research : Meaning – Scope and Significance - Utility of business research - qualities of good researcher - Types of research – research Process – Identification, Selection and formulation of research problem – Hypothesis - Research design

MEANING OF RESEARCH

Research is an endeavour to discover, develop and verify knowledge. It is an intellectual act that begins with the asking of questions and progressiveness through the critical examination of evidence that is both relevant and reliable to the revelation of truth. Research can be defined as 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, usually using a scientific method. The primary purpose of research is discovering, interpreting, and the development of methods and systems for the advancement of human knowledge on a wide variety of scientific matters of our world and the universe.

DEFINITIONS OF RESEARCH

Webster's New International Dictionary: "Research is careful critical enquiry or examination in seeking facts or principles, diligent investigation in order to ascertain something.

John W. Best: Research may be defined as the systematic and objective analysis and recording to controlled observations that may lead to the development of generalization, principles of theories resulting in prediction and possible ultimate control of events

Robert Ross: Research is essentially an investigation, a recording and an analysis of evidence too the purpose of gaining knowledge

Clifford Woody: Research comprises of defining and redefining problems, formulating hypothesis or suggested solutions, collecting, organizing and evaluating data making deduction and reaching conclusion and at last carefully testing conclusions to determine whether they fit in formulating hypothesis.

John Dewey: Research is considered to be the formal, systematic, intensive process of carrying on the scientific method of analysis. It involves a more systematic structure of investigation, usually in some sort of formal record of procedures and a report of result or conclusions.

Fred Kerlinger: Research is an organized enquiry designed and carried out to provide information for solving a problem.

Redman and Mory: Systematized effort to gain new knowledge

PURPOSE OF RESEARCH

The purpose of research is to discover answers to questions through the application of scientific procedures. The main aim of research is to find out the truth which is hidden and which has not been discovered as yet.

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

TYPES OF RESEARCH

1) Descriptive Research

A descriptive study may be simple or complex. It determines who, what, where and how of a topic. It is concerned with describing the characteristics (e.g., the extent to which libraries are used) estimating the proportion of the people in a specified population who hold certain views or attitudes (e.g., how many favour the abolition of capital punishment?) predicting specifically (e.g., how may will cash their government bonds during a given period?) and discovering or testing whether certain variables are associated (e.g., people who spend a good deal of time for reading, go to movies often with each other)

Descriptive study may employ any of or all the methods of data collection such as interview, questionnaire, observation, tests and cumulative record cards. In the descriptive study the researcher must be careful to make a note of the bias and extravagance that may creep in at every stage of the study – formulating the objectives of the study; designing the methods of data collection; selecting the sample; collecting, processing and analyzing the data; and reporting the findings.

2) Analytical Research

Analytical study makes use of available information by analyzing and doing critical evaluation. Analytical study makes use of higher level statistical tools which are not commonly used.

3) Applied Research

Applied research aims at finding a solution for an immediate problem faced by any business organization. This research deals with real life situations. Example: "Why have sales decreased during the last quarter"? Market research is an example of applied research. Applied research has a practical problem-solving emphasis. It brings out many new facts.

Examples:

1. Use of fibre glass body for cars instead of metal.

2. To develop a new market for the product.

4) Fundamental Research

This is otherwise known as basic research or fundamental research. Gathering knowledge for knowledge's sake is known as fundamental research. It does not have any commercial potential. It is not connected to any practical problem. e.g. Theory of Relativity. It is only for the enrichment of the knowledge.

5) Quantitative Research

Quantitative researches are based on the measurements of quantity or amounts. It means that these type of researches deals with items which are expressed in numbers.

6) Qualitative Research

Qualitative researches deals with the qualitative phenomena. i.e. anything which cannot be expressed in numerical terms. Motivation research is an example of qualitative research.

7) Conceptual Research

Conceptual research is that related to some abstract idea(s) or theory. It is generally used by philosophers and thinkers to develop new concepts or to reinterpret existing ones.

8) Empirical Research

Empirical research relies on experience or observation alone, often without due regard for system and theory. It is data-based research, coming up with conclusions which are capable of being verified by observation or experiment. We can also call it as experimental type of research. In such a research it is necessary to get at facts firsthand, at their source, and actively to go about doing certain things to stimulate the production of desired information. In such a research, the researcher must first provide himself with a working hypothesis or guess as to the probable results. He then works to get enough facts (data) to prove or disprove his hypothesis. He then sets up experimental designs which he thinks will manipulate the persons or the materials concerned so as to bring forth the desired information. Such research is thus 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.

9) One-time research or Longitudinal Research

In the former case the research is confined to a single time-period, whereas in the latter case the research is carried on over several time-periods.

10) Field Method

Field study is a scientific enquiry aimed at discovering the relations and interactions among sociological, physiological and educational variables in real social structures and life situations like communities, schools, factories, organizations and institutions. Hence, it is called field study.

11) Exploratory Research

Explanatory research is carried, when the reason for a problem is not clear. In exploratory research, all possible reasons which are very obvious are eliminated, thereby directing the research to proceed further with limited options.

Example for Exploratory Research

Sales decline in a company may be due to:

- Inefficient service
- Improper price
- Inefficient sales force
- Ineffective promotion
- Improper quality

12) Formalized Research

Formalized research studies are those with substantial structure and with specific hypotheses to be tested.

13) Historical Research

This research is the induction of principles through research into the past and social forces which have shaped the present. Its aim is to apply reflective thinking to unsolved social problems by discovering past trends of events, facts and attitudes, and by tracing lines of development in human thought and action.

14) Decision Oriented Research

Decision-oriented research is always for the need of a decision maker and the researcher in this case is not free to embark upon research according to his own inclination.

15) Individual and Group Research

The research undertaken by an individual is called individual research. The bulk of research activities in universities, and colleges are made by the individual. The individual research is done on the basis of one's own judgement, interest and capacity.

Group research is undertaken by several researchers. Their activities are coordinated by a director, Research conducted by a firm, trade association and government agency is performed by a team of researchers under a project director. Research in colleges and universities financed by grants is done on a group basis.

16) Operations Research

This method of research has been done for solving problems by using scientific methods and quantitative techniques. While the researchers care to study the development of methods, the industrial operations researcher evinces interest in the applications of methods to solve the pressing or critical problems of their firm.

Research can also be classified as conclusion-oriented and decision-oriented. While doing conclusion-oriented research a researcher is free to pick up a problem, redesign the enquiry and is free to conceptualize as he wishes. Decision-oriented research always implies taking a rational decision. Operational research is an example of decision-oriented research.

SIGNIFICANCE OF RESEARCH

"All progress is born of inquiry. Doubt is often better than overconfidence, for it leads to inquiry, and inquiry leads to invention" is a famous Hudson Maxim in context of which the significance of research can well be understood. Increased amounts of research make progress possible. Research inculcates scientific and inductive thinking and it promotes the development of logical habits of thinking and organisation.

The role of research in several fields of applied economics, whether related to business or to the economy as a whole, has greatly increased in modern times. The increasingly complex nature of business and government has focused attention on the use of research in solving operational problems. Research, as an aid to economic policy, has gained added importance, both for government and business.

1) Research provides the basis for nearly all Government Policies in our Economic System

For instance, 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 these 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 as well examine the consequences of each of these alternatives. Decision-making may not be a part of research, but research certainly facilitates the decisions of the policy maker. Government has also 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 problems 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. Another area in government, where research is necessary, is collecting 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 is by no means a routine task, but it involves a variety of research problems. These days nearly all governments maintain large staff of research technicians or experts to carry on this work. Thus, in the context of government, research as a tool to economic policy has three distinct phases of operation, viz., (i) investigation of economic structure through continual compilation of facts; (ii) diagnosis of events that are taking place and the analysis of the forces underlying them; and (iii) the prognosis, i.e., the prediction of future developments.

2) Research has its Special Significance in Solving various Operational and Planning Problems of Business and Industry

Operations research and market research, along with motivational research, are considered crucial and their results assist, in more than one way, in taking business decisions. Market research is the investigation of the structure and development of a market for the purpose of formulating efficient policies for purchasing, production and sales. Operations research refers to the application of mathematical, logical and analytical techniques to the solution of business problems of cost minimization or of profit maximization or what can be termed as optimization problems. Motivational research of determining why people behave as they do is mainly concerned with market characteristics. In other words, it is concerned with the determination of motivations underlying the consumer (market) behaviour. All these are of great help to people in business and industry who are responsible for taking business decisions. Research with regard to demand and market factors has great utility in business. Given knowledge of future demand, it is generally not difficult for a firm, or for an industry to adjust its supply schedule within the limits of its projected capacity. Market analysis has become an integral tool of business policy these days. Business budgeting, which ultimately results in a projected profit and loss account, is based mainly on sales estimates which in turn depend on business research. Once sales forecasting is done, efficient production and investment programmes can be set up around which are grouped the purchasing and financing plans. Research, thus, replaces intuitive business decisions by more logical and scientific decisions.

3) Research is equally important for social scientists in studying social relationships and in seeking answers to various social problems.

It provides the intellectual satisfaction of knowing a few things just for the sake of knowledge and also has practical utility for the social scientist to know for the sake of being able to do something better or in a more efficient manner. Research in social sciences is concerned both with knowledge for its own sake and with knowledge for what it can contribute to practical concerns. “This double emphasis is perhaps especially appropriate in the case of social science. On the one hand, its responsibility as a science is to develop a body of principles that make possible the understanding and prediction of the whole range of human interactions. On the other hand, because of its social orientation, it is increasingly being looked to for practical guidance

In addition to what has been stated above, the significance of research can also be understood keeping in view the following points:

a) To those students who are to write a master's or Ph.D. thesis, research may mean careerism or a way to attain a high position in the social structure;
b) To professionals in research methodology, research may mean a source of livelihood;
c) To philosophers and thinkers, research may mean the outlet for new ideas and insights;
d) To literary men and women, research may mean the development of new styles and creative work;
e) To analysts and intellectuals, research may mean the generalizations of new theories.

Thus, research is the fountain of knowledge for the sake of knowledge and an important source of providing guidelines for solving different business, governmental and social problems. It is a sort of formal training which enables one to understand the new developments in one's field in a better way.

QUALITIES OF GOOD RESEARCH

1) Good research is Systematic

It means that research is structured with specified steps to be taken in a specified sequence in accordance with the well defined set of rules. Systematic characteristic of the research does not rule out creative thinking but it certainly does reject the use of guessing and intuition in arriving at conclusions.

2) Good research is Logical

This implies that research is guided by the rules of logical reasoning and the logical process of induction and deduction are of great value in carrying out research. Induction is the process of reasoning from a part to the whole whereas deduction is the process of reasoning from some premise to a conclusion which follows from that very premise.

3) Good research is Empirical

It implies that research is related basically to one or more aspects of a real situation and deals with concrete data that provides a basis for external validity to research results.

4) Good research is Replicable

This characteristic allows research results to be verified by replicating the study and thereby building a sound basis for decisions.

QUALITIES OF A GOOD RESEARCHER

A) General Qualities

1) Scientific Attitude

The first essential Quality of a successful research worker is that he must possess a scientific (systematic) frame (structure) of mind. He must have the determination (willpower / strength of mind) and ability to get the naked (hidden) facts and not to be influenced by one's own wishes.

As human beings he has certain praises (admiration) and prejudices (bias). He has also certain precarceived notions (ideas) about the problems being researched. He should keep all these things with him.

2) Imagination and Insight

Researcher must possess high degree of imagination. He should be able to go deeper and deeper into the realm (area) of abstract social phenomena (fact / event) and visualize the intangible aspects (features) of the society

3) Perseverance

Work of scientific research requires steady of mind. Researcher should not get easily discouraged. It is equally possible that he might subsequently feel that the choice of the problem was wrong. Inspite of all this he must have more courage to face the difficulties and work patiently and continuously over long periods

4) Quick Grasping Power

The researcher should possess the power to grasp the significance of things quickly

5) Clarity of Thinking

A good researcher should have clear idea about the terminology that he is going to use.

B) Specific Qualities

1) Knowledge of the Subject

The researcher should be enough knowledge in his area of research. Such knowledge helps him in preparing questionnaire and schedule to get proper information. He can enter into face to face discussion and remove any doubts arising the minds of the people regarding the study

2) Knowledge of the technique of Research

Researcher should have basic idea on tools used in his research

3) Personal Taste in the Study

A personal taste in the study will inspire him and keep his morale (confidence) in times of difficulties. A forced work is often monotonous and very tiresome

4) Familiarity about the Information

The researcher should be familiar with the people whom he is studying. Familiarity will help him to get intimate (close) information

5) Unbiased Attitude

The researcher should have no preconceptions (idea / bias) about the subject under study. He should go to his research with absolutely a clean state. He should maintain an open mind and look for data which would substantiate (validate / verify) and give his theory a new meaning

STEPS IN RESEARCH PROCESS

Research process consists of series of actions or steps necessary to effectively carry out research and the desired sequencing of these steps.

1. Formulating the Research Problem;
2. Extensive literature survey;
3. Developing the Hypothesis;
4. Preparing the Research Design;
5. Determining Sample Design;
6. Collecting the Data;
7. Execution of the Project;
8. Analysis of Data;
9. Hypothesis testing;
10. Generalizations and Interpretation, and
11. Preparation of the Report or Presentation of the Results

1) Formulating the Research Problem

There are two types of research problems, viz., those which relate to states of nature and those which relate to relationships between variables. At the very outset the researcher must single out the problem he wants to study, i.e., he must decide the general area of interest or aspect of a subject-matter that he would like to inquire into. Initially the problem may be stated in a broad general way and then the ambiguities, if any, relating to the problem be resolved. Then, the feasibility of a particular solution has to be considered before a working formulation of the problem can be set up. The formulation of a general topic into a specific research problem, thus, constitutes the first step in a scientific enquiry. Essentially two steps are involved in formulating the research problem, viz., understanding the problem thoroughly, and rephrasing the same into meaningful terms from an analytical point of view.

The best way of understanding the problem is to discuss it with one's own colleagues or with those having some expertise in the matter. In an academic institution the researcher can seek the help from a guide who is usually an experienced man and has several research problems in mind. Often, the guide puts forth the problem in general terms and it is up to the researcher to narrow it down and phrase the problem in operational terms. In private business units or in governmental organisations, the problem is usually earmarked by the administrative agencies with whom the researcher can discuss as to how the problem originally came about and what considerations are involved in its possible solutions.

The researcher must at the same time examine all available literature to get himself acquainted with the selected problem. He may review two types of literature the conceptual literature concerning the concepts and theories, and the empirical literature consisting of studies made earlier which are similar to the one proposed. The basic outcome of this review will be the knowledge as to what data and other materials are available for operational purposes which will enable the researcher to specify his own research problem in a meaningful context. After this the researcher rephrases the problem into analytical or operational terms i.e., to put the problem in as specific terms as possible. This task of formulating, or defining, a research problem is a step of greatest importance in the entire research process. The problem to be investigated must be defined unambiguously for that will help discriminating relevant data from irrelevant ones. Care must, however, be taken to verify the objectivity and validity of the background facts concerning the problem. Professor W.A. Neiswanger correctly states that the statement of the objective is of basic importance because it determines the data which are to be collected, the characteristics of the data which are relevant, relations which are to be explored, the choice of techniques to be used in these explorations and the form of the final report. If there are certain pertinent terms, the same should be clearly defined along with the task of formulating the problem. In fact, formulation of the problem often follows a sequential pattern where a number of formulations are set up, each formulation more specific than the preceding one, each one phrased in more analytical terms, and each more realistic in terms of the available data and resources.

2) Extensive Literature Survey

Once the problem is formulated, a brief summary of it should be written down. It is compulsory for a research worker writing a thesis for a Ph.D. degree to write a synopsis of the topic and submit it to the necessary Committee or the Research Board for approval. At this juncture the researcher should undertake extensive literature survey connected with the problem. For this purpose, the abstracting and indexing journals and published or unpublished bibliographies are the first place to go to. Academic journals, conference proceedings, government reports, books etc., must be tapped depending on the nature of the problem. In this process, it should be remembered that one source will lead to another. The earlier studies, if any, which are similar to the study in hand should be carefully studied. A good library will be a great help to the researcher at this stage.

3) Developing the Hypothesis

After extensive literature survey, researcher should state in clear terms the working hypothesis or hypotheses. Working hypothesis is tentative assumption made in order to draw out and test its logical or empirical consequences. As such the manner in which research hypotheses are developed is particularly important since they provide the focal point for research. They also affect the manner in which tests must be conducted in the analysis of data and indirectly the quality of data which is required for the analysis. In most types of research, the development of working hypothesis plays an important role. Hypothesis should be very specific and limited to the piece of research in hand because it has to be tested. The role of the hypothesis is to guide the researcher by delimiting the area of research and to keep him on the right track. It sharpens his thinking and focuses attention on the more important facets of the problem. It also indicates the type of data required and the type of methods of data analysis to be used.

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

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

Thus, working hypotheses arise as a result of a-priori thinking about the subject, examination of the available data and material including related studies and the counsel of experts and interested parties. Working hypotheses are more useful when stated in precise and clearly defined terms. It may as well be remembered that occasionally we may encounter a problem where we do not need working hypotheses, specially in the case of exploratory or formulative researches which do not aim at testing the hypothesis. But as a general rule, specification of working hypotheses in another basic step of the research process in most research problems.

4) Preparing the Research Design

The research problem having been formulated in clear cut terms, the researcher will be required to prepare a research design, i.e., he will have to state the conceptual structure within which research would be conducted. The preparation of such a design facilitates research to be as efficient as possible yielding maximal information. In other words, the function of research design is to provide for the collection of relevant evidence with minimal expenditure of effort, time and money. But how all these can be achieved depends mainly on the research purpose. Research purposes may be grouped into four categories, viz., (i) Exploration, (ii) Description, (iii) Diagnosis, and (iv) Experimentation. A flexible research design which provides opportunity for considering many different aspects of a problem is considered appropriate if the purpose of the research study is that of exploration. But when the purpose happens to be an accurate description of a situation or of an association between variables, the suitable design will be one that minimises bias and maximises the reliability of the data collected and analysed.

There are several research designs, such as, experimental and non-experimental hypothesis testing. Experimental designs can be either informal designs (such as before-and-after without control, after-only with control, before-and-after with control) or formal designs (such as completely randomized design, randomized block design, Latin square design, simple and complex factorial designs), out of which the researcher must select one for his own project.

The preparation of the research design, appropriate for a particular research problem, involves usually the consideration of the following:

a) The means of obtaining the information;
b) The availability and skills of the researcher and his staff (if any);
c) Explanation of the way in which selected means of obtaining information will be organized and the reasoning leading to the selection;
d) The time available for research; and
e) The cost factor relating to research, i.e., the finance available for the purpose.

5) Determining Sample Design

All the items under consideration in any field of inquiry constitute a ‘universe’ or ‘population’. A complete enumeration of all the items in the ‘population’ is known as a census inquiry. It can be presumed that in such an inquiry when all the items are covered no element of chance is left and highest accuracy is obtained. But in practice this may not be true. Even the slightest element of bias in such an inquiry will get larger and larger as the number of observations increases. Moreover, there is no way of checking the element of bias or its extent except through a resurvey or use of sample checks. Besides, this type of inquiry involves a great deal of time, money and energy. Not only this, census inquiry is not possible in practice under many circumstances. For instance, blood testing is done only on sample basis. Hence, quite often we select only a few items from the universe for our study purposes. The items so selected constitute what is technically called a sample.

The researcher must decide the way of selecting a sample or what is popularly known as the sample design. In other words, a sample design is a definite plan determined before any data are actually collected for obtaining a sample from a given population. Thus, the plan to select 12 of a city’s 200 drugstores in a certain way constitutes a sample design. Samples can be either probability samples or non-probability samples. With probability samples each element has a known probability of being included in the sample but the non-probability samples do not allow the researcher to determine this probability. Probability samples are those based on simple random sampling, systematic sampling, stratified sampling, cluster/area sampling whereas non-probability samples are those based on convenience sampling, judgement sampling and quota sampling techniques. A brief mention of the important sample designs is as follows:

a) Deliberate Sampling

Deliberate sampling is also known as purposive or non-probability sampling. This sampling method involves purposive or deliberate selection of particular units of the universe for constituting a sample which represents the universe. When population elements are selected for inclusion in the sample based on the ease of access, it can be called convenience sampling. If a researcher wishes to secure data from, say, gasoline buyers, he may select a fixed number of petrol stations and may conduct interviews at these stations. This would be an example of convenience sample of gasoline buyers. At times such a procedure may give very biased results particularly when the population is not homogeneous. On the other hand, in judgement sampling the researcher's judgement is used for selecting items which he considers as representative of the population. For example, a judgement sample of college students might be taken to secure reactions to a new method of teaching. Judgement sampling is used quite frequently in qualitative research where the desire happens to be to develop hypotheses rather than to generalise to larger populations.

b) Simple Random Sampling

This type of sampling is also known as chance sampling or probability sampling where each and every item in the population has an equal chance of inclusion in the sample and each one of the possible samples, in case of finite universe, has the same probability of being selected. For example, if we have to select a sample of 300 items from a universe of 15,000 items, then we can put the names or numbers of all the 15,000 items on slips of paper and conduct a lottery. Using the random number tables is another method of random sampling. To select the sample, each item is assigned a number from 1 to 15,000. Then, 300 five digit random numbers are selected from the table. To do this we select some random starting point and then a systematic pattern is used in proceeding through the table. We might start in the 4th row, second column and proceed down the column to the bottom of the table and then move to the top of the next column to the right. When a number exceeds the limit of the numbers in the frame, in our case over 15,000, it is simply passed over and the next number selected that does fall within the relevant range. Since the numbers were placed in the table in a completely random fashion, the resulting sample is random. This procedure gives each item an equal probability of being selected. In case of infinite population, the selection of each item in a random sample is controlled by the same probability and that successive selections are independent of one another.

c) Systematic Sampling

In some instances the most practical way of sampling is to select every 15th name on a list, every 10th house on one side of a street and so on. Sampling of this type is known as systematic sampling. An element of randomness is usually introduced into this kind of sampling by using random numbers to pick up the unit with which to start. This procedure is useful when sampling frame is available in the form of a list. In such a design the selection process starts by picking some random point in the list and then every nth element is selected until the desired number is secured.

d) Stratified Sampling

If the population from which a sample is to be drawn does not constitute a homogeneous group, then stratified sampling technique is applied so as to obtain a representative sample. In this technique, the population is stratified into a number of non-overlapping subpopulations or strata and sample items are selected from each stratum. If the items selected from each stratum is based on simple random sampling the entire procedure, first stratification and then simple random sampling, is known as stratified random sampling.

e) Quota Sampling

In stratified sampling the cost of taking random samples from individual strata is often so expensive that interviewers are simply given quota to be filled from different strata, the actual selection of items for sample being left to the interviewer's judgement. This is called quota sampling. The size of the quota for each stratum is generally proportionate to the size of that stratum in the population. Quota sampling is thus an important form of non-probability sampling. Quota samples generally happen to be judgement samples rather than random samples.

f) Cluster Sampling and Area Sampling

Cluster sampling involves grouping the population and then selecting the groups or the clusters rather than individual elements for inclusion in the sample. Suppose some departmental store wishes to sample its credit card holders. It has issued its cards to 15,000 customers. The sample size is to be kept say 450. For cluster sampling this list of 15,000 card holders could be formed into 100 clusters of 150 card holders each. Three clusters might then be selected for the sample randomly. The sample size must often be larger than the simple random sample to ensure the same level of accuracy because is cluster sampling procedural potential for order bias and other sources of error is usually accentuated. The clustering approach can, however, make the sampling procedure relatively easier and increase the efficiency of field work, specially in the case of personal interviews.

Area sampling is quite close to cluster sampling and is often talked about when the total geographical area of interest happens to be big one. Under area sampling we first divide the total area into a number of smaller non-overlapping areas, generally called geographical clusters, then a number of these smaller areas are randomly selected, and all units in these small areas are included in the sample. Area sampling is specially helpful where we do not have the list of the population concerned. It also makes the field interviewing more efficient since interviewer can do many interviews at each location.

g) Multi-stage Sampling

This is a further development of the idea of cluster sampling. This technique is meant for big inquiries extending to a considerably large geographical area like an entire country. Under multi-stage sampling the first stage may be to select large primary sampling units such as states, then districts, then towns and finally certain families within towns. If the technique of random-sampling is applied at all stages, the sampling procedure is described as multi-stage random sampling.

h) Sequential Sampling

This is somewhat a complex sample design where the ultimate size of the sample is not fixed in advance but is determined according to mathematical decisions on the basis of information yielded as survey progresses. This design is usually adopted under acceptance sampling plan in the context of statistical quality control.

In practice, several of the methods of sampling described above may well be used in the same study in which case it can be called mixed sampling. It may be pointed out here that normally one should resort to random sampling so that bias can be eliminated and sampling error can be estimated. But purposive sampling is considered desirable when the universe happens to be small and a known characteristic of it is to be studied intensively. Also, there are conditions under which sample designs other than random sampling may be considered better for reasons like convenience and low costs. The sample design to be used must be decided by the researcher taking into consideration the nature of the inquiry and other related factors.

6) Collecting the Data

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

Primary data can be collected either through experiment or through survey. If the researcher conducts an experiment, he observes some quantitative measurements, or the data, with the help of which he examines the truth contained in his hypothesis. But in the case of a survey, data can be collected by any one or more of the following ways:

a) By Observation

This method implies the collection of information by way of investigator's own observation, 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.

b) 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.

c) Through Telephone Interview

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.

d) By Mailing of Questionnaire

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.

e) 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, finanical 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.

7) Execution of the Project

Execution of the project is a very important step in the research process. If the execution of the project proceeds on correct lines, the data to be collected would be adequate and dependable. The researcher should see that the project is executed in a systematic manner and in time. If the survey is to be conducted by means of structured questionnaires, data can be readily machine-processed. In such a situation, questions as well as the possible answers may be coded. If the data are to be collected through interviewers, arrangements should be made for proper selection and training of the interviewers. The training may be given with the help of instruction manuals which explain clearly the job of the interviewers at each step. Occasional field checks should be made to ensure that the interviewers are doing their assigned job sincerely and efficiently. A careful watch should be kept for unanticipated factors in order to keep the survey as much realistic as possible. This, in other words, means that steps should be taken to ensure that the survey is under statistical control so that the collected information is in accordance with the pre-defined standard of accuracy. If some of the respondents do not cooperate, some suitable methods should be designed to tackle this problem. One method of dealing with the non-response problem is to make a list of the non-respondents and take a small sub-sample of them, and then with the help of experts vigorous efforts can be made for securing response.

8) Analysis of Data

After the data have been collected, the researcher turns to the task of analysing them. The analysis of data requires a number of closely related operations such as establishment of categories, the application of these categories to raw data through coding, tabulation and then drawing statistical inferences. The unwieldy data should necessarily be condensed into a few manageable groups and tables for further analysis. Thus, researcher should classify the raw data into some purposeful and usable categories. Coding operation is usually done at this stage through which the categories of data are transformed into symbols that may be tabulated and counted. Editing is the procedure that improves the quality of the data for coding. With coding the stage is ready for tabulation. Tabulation is a part of the technical procedure wherein the classified data are put in the form of tables. The mechanical devices can be made use of at this juncture. A great deal of data, 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). For instance, if there are two samples of weekly wages, each sample being drawn from factories in different parts of the same city, giving two different mean values, then our problem may be whether the two mean values are significantly different or the difference is just a matter of chance. Through the use of statistical tests we can establish whether such a difference is a real one or is the result of random fluctuations. If the difference happens to be real, the inference will be that the two samples come from different universes and if the difference is due to chance, the conclusion would be that the two samples belong to the same universe. Similarly, the technique of analysis of variance can help us in analysing whether three or more varieties of seeds grown on certain fields yield significantly different results or not. In brief, the researcher can analyse the collected data with the help of various statistical measures.

9) Hypothesis Testing

After analysing the data as stated above, the researcher is in a position to test the hypotheses, if any, he had formulated earlier. Do the facts support the hypotheses or they happen to be contrary? This is the usual question which should be answered while testing hypotheses. Various tests, such as Chi square test, t-test, F-test, have been developed by statisticians for the purpose. The hypotheses may be tested through the use of one or more of such tests, depending upon the nature and object of research inquiry. Hypothesis-testing will result in either accepting the hypothesis or in rejecting it. If the researcher had no hypotheses to start with, generalisations established on the basis of data may be stated as hypotheses to be tested by subsequent researches in times to come.

10) Generalizations and Interpretation

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

11) Preparation of the Report or Presentation of the Results

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

1) The layout of the report should be as follows: (i) the preliminary pages; (ii) the main text, and (iii) the end matter.

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

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

a) **Introduction:** It should contain a clear statement of the objective of the research and an explanation of the methodology adopted in accomplishing the research. The scope of the study along with various limitations should as well be stated in this part.
b) **Summary of Findings**: After introduction there would appear a statement of findings and recommendations in non-technical language. If the findings are extensive, they should be summarised.
c) **Main Report**: The main body of the report should be presented in logical sequence and broken-down into readily identifiable sections.

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

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

2) Report should be written in a concise and objective style in simple language avoiding vague expressions such as ‘it seems,’ ‘there may be’, and the like.

3) Charts and illustrations in the main report should be used only if they present the information more clearly and forcibly.

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

RESEARCH DESIGN

The formidable problem that follows the task of defining the research problem is the preparation of the design of the research project, popularly known as the “research design”. Decisions regarding what, where, when, how much, by what means concerning an inquiry or a research study constitute a research design. “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.”1 In fact, the research design is the conceptual structure within which research is conducted; it constitutes the blueprint for the collection, measurement and analysis of data. As such the design includes an outline of what the researcher will do from writing the hypothesis and its operational implications to the final analysis of data. More explicitly, the desing decisions happen to be in respect of:

1. What is the study about?
2. Why is the study being made?
3. Where will the study be carried out?

4. What type of data is required?
5. Where can the required data be found?
6. What periods of time will the study include?
7. What will be the sample design?
8. What techniques of data collection will be used?
9. How will the data be analyzed?
10. In what style will the report be prepared?

Keeping in view the above stated design decisions, one may split the overall research design into the following parts:

a) the sampling design which deals with the method of selecting items to be observed for the given study;
b) the observational design which relates to the conditions under which the observations are to be made;
c) the statistical design which concerns with the question of how many items are to be observed and how the information and data gathered are to be analysed; and
d) the operational design which deals with the techniques by which the procedures specified in the sampling, statistical and observational designs can be carried out.

From what has been stated above, we can state the important features of a research design as under:

i) It is a plan that specifies the sources and types of information relevant to the research problem.
ii) It is a strategy specifying which approach will be used for gathering and analysing the data.
iii) It also includes the time and cost budgets since most studies are done under these two constraints.

In brief, research design must, at least, contain—(a) a clear statement of the research problem; (b) procedures and techniques to be used for gathering information; (c) the population to be studied; and (d) methods to be used in processing and analyzing data.

COMPONENTS OF RESEARCH DESIGN

1) Title of the Study

Enough information should be given in the title, to identify the study. The researcher should consider the following while selecting a title

- The title should be specific to the area of the study
- The title should indicate the topic of the study
- The language of the title should be professional in nature but not pedantic (dull)
- The title should be as brief as possible

2) Introduction

Under this heading a brief explanation of the genesis of the problem should be given.

3) Statement of the Problem

After a brief introduction explaining the genesis of the problem, the researcher should state the problem. While stating the problem use of clear, simple and concise statement is preferable

4) Review of Previous Studies

Under this head the researcher presents what is so far known about the problem under consideration. All related studies need not be discussed. The researcher may describe the most important ones. A review of the previous studies enables the researcher to know the different areas covered by various studies, to concentrate on the areas where little research has been carried out, to look into various merits and shortcomings of certain studies already completed and to verify the present findings with that of the previous ones

5) Scope of the Study

This heading gives an idea about the extend of the study. The scope of the study is dependent on several factors such as the time and money available with the investigator, availability of the sample, co-operation of the respondents and the like

6) Objectives of the Study

The task of the researcher is to lay down the objectives precisely. The objectives enlighten the researcher's own mind and lead to more efficient enquiry. Once the objectives are selected, the study can be undertaken with required accuracy and within the given resources. The objectives mentioned should be well within the scope of the study

7) Hypothesis to be Tested

Hypothesis is a proposition, condition or principle which is assumed, perhaps without belief in order to draw logical conclusions. Hypotheses are formulated to explain observed facts, conditions, or behaviours and to serve as a guide in the research process. Each hypothesis is individually tested to determine whether it is tenable (reasonable) or not. Hypothesis should be stated in clear, concise and understandable language

8) Operational Definition of Concepts

All terms that might be ambiguous should be clarified. A clear understanding of the terms used in the study is important. It is necessary to identify and label the variables. The variables can be labeled as independent variable and dependent variable. An independent variable is the factor which is measured, manipulated or selected by the experiments. A dependent variable is that factor which is measured to determine the effect of independent variable

9) Geographical Area to be covered

Under this head the area to be covered by the study is mentioned

10) Reference Period

The period of study can be mentioned under this heading

11) Methodology

The researcher should first determine the kind of information needed to answer the research questions. Secondly he must know the sources of data and finally he must know the means by which he will gather information which is known as methodology

12) Sampling

Sampling involves taking a portion of population, making observation on this smaller group and then generalizing the findings to be applied to a large population. The small group that is observed is called the sample and the large group is called population. The sample is the portion of the population and it must be representative of the population. If the sample is biased, the findings of the study cannot be generalized

13) Tools for Collection of Data

The choice of method for collecting the data is governed by the subject matter, the unit of enquiry and the scale of the study. A study of the behavior of a group would call for observational techniques, for a simple enquiry among the cross section of population, a questionnaire is adequate. A survey of general population entailing many complicated questions would call for personal interviewing

14) Plan of Analysis

Once the data have been collected, they must be reduced to meaningful results by statistical analysis so that the conclusions for generalization can be drawn from them. The researcher should describe how he plans to organize the data. He should decide the statistical treatment. The appropriateness of the technique should be discussed. He must discuss the procedure for treating the data

15) Research Report

The results should be communicated. The format consists of three parts

- Part I Preliminary Pages, which contain title page, approval sheet, preface (if any), table of contents, list of tables (if any) and list of figures (if any)
- Part II Body of the report, which covers content chapters
- Part III Supplementary pages which included bibliography appendix (if any) and index (if any)

16) Time Schedule

The researcher has to work out a time schedule for his research work. The time required includes the following:

- Time to be used for preparing the theoretical background
- Time to be used for preparing the data gathering devices such as questionnaire, interview schedule, record sheet, interviewer's manual and time and expenses sheets
- Time to be used for data collection
- Time to be used for processing the data
- Time to be used for writing the report
- Time to be used for submitting the thesis

17) Financial Budget

It is desirable to work out the budget which gives an idea about the money needed to complete the project. The cost estimates of the project will include stationery, printing, sample selection, field work, mailing, processing, tabulating, preparation of report and overheads.

The research design will differ depending on the research purpose. The research purposes may purport to gain familiarity with a phenomenon, portray accurately the characteristics of a particular individual, situation or group, determine the frequency with which something occurs and test a hypothesis of casual relationship between variables. In practice, these different types of studies are not always sharply distinguishable. The research design will be more or less the same for any type of study

METHODS OF RESEARCH DESIGN

Different research designs can be conveniently described if we categorize them as: (1) research design in case of exploratory research studies; (2) research design in case of descriptive and diagnostic research studies, and (3) research design in case of hypothesis-testing research studies.

1) Research design in case of exploratory research studies

Exploratory research studies are also termed as formulative research studies. The main purpose of such studies is that of formulating a problem for more precise (accurate) investigation or of developing the working hypotheses from an operational point of view. The major emphasis in such studies is on the discovery of ideas and insights. As such the research design appropriate for such studies must be flexible enough to provide opportunity for considering different aspects of a problem under study. Inbuilt flexibility in research design is needed because the research problem, broadly defined initially, is transformed into one with more precise meaning in exploratory studies, which fact may necessitate changes in the research procedure for gathering relevant data. Generally, the following three methods in the context of research design for such studies are talked about: (a) the survey of concerning literature; (b) the experience survey and (c) the analysis of 'insight-stimulating' examples.

The survey of concerning literature happens to be the most simple and fruitful method of formulating precisely the research problem or developing hypothesis. Hypotheses stated by earlier workers may be reviewed and their usefulness be evaluated as a basis for further research.

It may also be considered whether the already stated hypotheses suggest new hypothesis. In this way the researcher should review and build upon the work already done by others, but in cases where hypotheses have not yet been formulated, his task is to review the available material for deriving the relevant hypotheses from it.

Besides, the bibliographical survey of studies, already made in one's area of interest may as well as made by the researcher for precisely formulating the problem. He should also make an attempt to apply concepts and theories developed in different research contexts to the area in which he is himself working. Sometimes the works of creative writers also provide a fertile ground for hypothesis-formulation and as such may be looked into by the researcher.

Experience survey means the survey of people who have had practical experience with the problem to be studied. The object of such a survey is to obtain insight into the relationships between variables and new ideas relating to the research problem. For such a survey people who are competent and can contribute new ideas may be carefully selected as respondents to ensure a representation of different types of experience. The respondents so selected may then be interviewed by the investigator. The researcher must prepare an interview schedule for the systematic questioning of informants. But the interview must ensure flexibility in the sense that the respondents should be allowed to raise issues and questions which the investigator has not previously considered. Generally, the experience-collecting interview is likely to be long and may last for few hours. Hence, it is often considered desirable to send a copy of the questions to be discussed to the respondents well in advance. This will also give an opportunity to the respondents for doing some advance thinking over the various issues involved so that, at the time of interview, they may be able to contribute effectively. Thus, an experience survey may enable the researcher to define the problem more concisely (briefly) and help in the formulation of the research hypothesis. This survey may as well provide information about the practical possibilities for doing different types of research.

Analysis of ‘insight-stimulating’ examples is also a fruitful method for suggesting hypotheses for research. It is particularly suitable in areas where there is little experience to serve as a guide. This method consists of the intensive study of selected instances of the phenomenon in which one is interested. For this purpose the existing records, if any, may be examined, the unstructured interviewing may take place, or some other approach may be adopted. Attitude of the investigator, the intensity of the study and the ability of the researcher to draw together diverse information into a unified interpretation are the main features which make this method an appropriate procedure for evoking insights.

Now, what sort of examples is to be selected and studied? There is no clear cut answer to it. Experience indicates that for particular problems certain types of instances are more appropriate than others. One can mention few examples of ‘insight-stimulating’ cases such as the reactions of strangers, the reactions of marginal individuals, the study of individuals who are in transition from one stage to another, the reactions of individuals from different social strata and the like. In general, cases that provide sharp contrasts or have striking features are considered relatively more useful while adopting this method of hypotheses formulation.

Thus, in an exploratory of formulative research study which merely leads to insights or hypotheses, whatever method or research design outlined above is adopted, the only thing essential is that it must continue to remain flexible so that many different facets of a problem may be considered as and when they arise and come to the notice of the researcher.

2) Research design in case of descriptive and diagnostic research studies

Descriptive research studies are those studies, which are concerned with describing the characteristics of a particular individual, or of a group, whereas diagnostic research studies determine the frequency with which something occurs or its association with something else. The studies concerning whether certain variables are associated are examples of diagnostic research studies. As against this, studies concerned with specific predictions, with narration of facts and characteristics concerning individual, group or situation are all examples of descriptive research studies. Most of the social research comes under this category. From the point of view of the research design, the descriptive as well as diagnostic studies share common requirements and as such we may group together these two types of research studies. In descriptive as well as in diagnostic studies, the researcher must be able to define clearly, what he wants to measure and must find adequate methods for measuring it along with a clear cut definition of 'population' he wants to study. Since the aim is to obtain complete and accurate information in the said studies, the procedure to be used must be carefully planned. The research design must make enough provision for protection against bias and must maximize reliability, with due concern for the economical completion of the research study. The design in such studies must be rigid and not flexible and must focus attention on the following:

a. Formulating the objective of the study (what the study is about and why is it being made?)
b. Designing the methods of data collection (what techniques of gathering data will be adopted?)
c. Selecting the sample (how much material will be needed?)
d. Collecting the data (where can the required data be found and with what time period should the data be related?)
e. Processing and analyzing the data.
f. Reporting the findings.

In a descriptive/diagnostic study the first step is to specify the objectives with sufficient precision (accuracy) to ensure that the data collected are relevant. If this is not done carefully, the study may not provide the desired information.

Then comes the question of selecting the methods by which the data are to be obtained. In other words, techniques for collecting the information must be devised. Several methods (viz., observation, questionnaires, interviewing, examination of records, etc.), with their merits and limitations, are available for the purpose and the researcher may user one or more of these methods which have been discussed in detail in later chapters. While designing data-collection procedure, adequate safeguards against bias and unreliability must be ensured. Whichever method is selected, questions must be well examined and be made unambiguous; interviewers must be instructed not to express their own opinion; observers must be trained so that they uniformly record a given item of behaviour. It is always desirable to pretest the data collection instruments before they are finally used for the study purposes. In other words, we can say that "structured instruments" are used in such studies.

In most of the descriptive/diagnostic studies the researcher takes out sample(s) and then wishes to make statements about the population on the basis of the sample analysis or analyses. More often than not, sample has to be designed. Here we may only mention that the problem of designing samples should be tackled in such a fashion that the samples may yield accurate information with a minimum amount of research effort. Usually one or more forms of probability sampling, or what is often described as random sampling, are used.

To obtain data free from errors introduced by those responsible for collecting them, it is necessary to supervise closely the staff of field workers as they collect and record information. Checks may be set up to ensure that the data collecting staff perform their duty honestly and without prejudice. "As data are collected, they should be examined for completeness, comprehensibility, consistency and reliability."

The data collected must be processed and analysed. This includes steps like coding the interview replies, observations, etc.; tabulating the data; and performing several statistical computations. To the extent possible, the processing and analysing procedure should be planned in detail before actual work is started. This will prove economical in the sense that the researcher may avoid unnecessary labour such as preparing tables for which he later finds he has no use or on the other hand, re-doing some tables because he failed to include relevant data. Coding should be done carefully to avoid error in coding and for this purpose the reliability of coders needs to be checked. Similarly, the accuracy of tabulation may be checked by having a sample of the tables re-done. In case of mechanical tabulation the material (i.e., the collected data or information) must be entered on appropriate cards, which is usually done by punching holes corresponding to a given code. The accuracy of punching is to be checked and ensured. Finally, statistical computations are needed and as such averages, percentages and various coefficients must be worked out. Probability and sampling analysis may as well be used. The appropriate statistical operations, along with the use of appropriate tests of significance should be carried out to safeguard the drawing of conclusions concerning the study.

Last of all comes the question of reporting the findings. This is the task of communicating the findings to others and the researcher must do it in an efficient manner. The layout of the report needs to be well planned so that all things relating to the research study may be well presented in simple and effective style.

Thus, the research design in case of descriptive/diagnostic studies is a comparative design throwing light on all points narrated above and must be prepared keeping in view the objective(s) of the study and the resources available. However, it must ensure the minimisation of bias and maximisation of reliability of the evidence collected. The said design can be appropriately referred to as a *survey design* since it takes into account all the steps involved in a survey concerning a phenomenon to be studied.

3) Research design in case of hypothesis-testing research studies

Hypothesis-testing research studies (generally known as experimental studies) are those where the researcher tests the hypotheses of causal relationships between variables. Such studies require procedures that will not only reduce bias and increase reliability, but will permit drawing inferences about causality. Usually experiments meet this requirement.

Hence, when we talk of research design in such studies, we often mean the design of experiments.

Professor R.A. Fisher's name is associated with experimental designs. Beginning of such designs was made by him when he was working at Rothamsted Experimental Station (Centre for Agricultural Research in England). As such the study of experimental designs has its origin in agricultural research. Professor Fisher found that by dividing agricultural fields or plots into different blocks and then by conducting experiments in each of these blocks, whatever information is collected and inferences drawn from them, happens to be more reliable. This fact inspired him to develop certain experimental designs for testing hypotheses concerning scientific investigations. Today, the experimental designs are being used in researches relating to phenomena of several disciplines. Since experimental designs originated in the context of agricultural operations, we still use, though in a technical sense, several terms of agriculture (such as treatment, yield, plot, block etc.) in experimental designs.

*** * * * ***

POSSIBLE QUESTIONS

PART A (ONE MARKS)

(ONLINE EXAMINATION)

PART – B

1. Describe the types of research with examples?
2. Explain the scope of research in modern times?
3. Explain the significance of research in detail.
4. Describe the process of research with diagrams.
5. Explain the characteristics and qualities of good research?
6. Discuss the problems faced by the researchers in India?
7. Discuss the importance of research in modern times?
8. Explain the types of research with suitable examples?
9. Define research and explain the significance of research in detail?
10. Discuss the steps followed in research process with suitable chart?
11. Explain the criteria's to be followed to develop the good hypothesis?

UNIT II

SYLLABUS

Sampling : Methods and Techniques – sample size – sampling error – field work and data collection. Tools of data collection – interview schedule – questionnaire – observation – interview and mailed questionnaire – pilot study and final collection of data

SAMPLING

All items in any field of inquiry constitute a 'Universe' or 'Population. A complete enumeration of all items in the 'population' is known as a census inquiry. It can be presumed that in such an inquiry, when all items are covered, no element of chance is left and highest accuracy is obtained. But in practice this may not be true. Even the slightest element of bias in such an inquiry will get larger and larger as the number of observation increases. Moreover, there is no way of checking the element of bias or its extent except through a resurvey or use of sample checks. Besides, this type of inquiry involves a great deal of time, money and energy. Therefore, when the field of inquiry is large, this method becomes difficult to adopt because of the resources involved. At times, this method is practically beyond the reach of ordinary researchers. Perhaps, government is the only institution which can get the complete enumeration carried out. Even the government adopts this in very rare cases such as population census conducted once in a decade.

Further, many a time it is not possible to examine every item in the population, and sometimes it is possible to obtain sufficiently accurate results by studying only a part of total population. In such cases there is no utility of census surveys. Then, the researcher may make use of Sampling.

- Sampling is nothing but a proportion of Population
- However, it needs to be emphasised that when the universe is a small one, it is no use resorting to a sample survey. When field studies are undertaken in practical life,

considerations of time and cost almost invariably lead to a selection of respondents i.e., selection of only a few items. The respondents selected should be as representative of the total population as possible in order to produce a miniature cross-section. The selected respondents constitute what is technically called a ‘sample’ and the selection process is called ‘sampling technique.’ The survey so conducted is known as ‘sample survey’.

- Algebraically, let the population size be N and if a part of size n (which is < N) of this population is selected according to some rule for studying some characteristic of the population, the group consisting of these n units is known as ‘sample’.
- Researcher must prepare a sample design for his study i.e., he must plan how a sample should be selected and of what size such a sample would be.

STEPS IN SAMPLE DESIGN

1) Type of Universe

The first step in developing any sample design is to clearly define the set of objects, technically called the universe, to be studied. The universe can be finite or infinite. In finite universe the number of items is certain, but in case of an infinite universe the number of items is infinite, i.e. we cannot have any idea about the total number of items. The population of a city, the number of workers in a factory and the like are examples of finite universes, whereas the number of stars in the sky, listeners of a specific radio programme, throwing a dice etc. are examples of infinite universes.

2) Sampling Unit

A decision has to be taken concerning a sampling unit before selecting sample. Sampling unit may be a geographical one such as state, district, village, etc., or a construction unit such as a house, flat, etc., or it may be a social unit such as family, club, school, etc., or it may be an individual. The researcher will have to decide one or more of such units that he has to select for his study.

3) Source List

It is also known as 'sampling frame' from which sample is to be drawn. It contains the names of all items of universe. If source list is not available, researcher has to prepare it. Such a list should be comprehensive, correct, reliable and appropriate. It is extremely important for the source list to be as representative of the population as possible.

4) Size of Sample

This refers to the number of items to be selected from the universe to constitute a sample. This is a major problem before a researcher. The size of sample should neither be excessively large, nor too small. It should be optimum. An optimum sample is one which fulfills the requirements of efficiency, representativeness, reliability and flexibility. While deciding the size of sample, researcher must determine the desired precision as also an acceptable confidence level for the estimate. The size of population variance needs to be considered as in case of larger variance usually a bigger sample is needed. The size of population must be kept in view for this also limits the sample size. The parameters of interest in a research study must be kept in view, while deciding the size of sample. Costs too dictate the size of sample that we can draw. As such, budgetary constraint must invariably be taken into consideration when we decide the sample size.

5) Parameters of Interest

In determining the sample design, one must consider the question of the specific population parameters which are of interest. For instance, we may be interested in estimating the proportion of persons with some characteristics in the population or we may be interested in knowing some average or the other measure concerning the population. There may be important sub-groups in the population about whom we would like to make estimates. All this has a strong impact upon the sample design we would accept.

6) Budgetary Constraint Cost considerations, from practical point of view, have a major impact upon decisions relating to not only the size of the sample but also to the type of sample. This fact can even lead to the use of a non-probability sample.

7) Sampling Procedure

Finally, the researcher must decide the type of sample he will use i.e. he must decide about the technique to be used in selecting the items for the sample. In fact, this technique or procedure stands for the sample design itself. There are several sample designs out of which the researcher must choose one for his study. Obviously, he must select that design which, for a given sample size and for a given cost, has a smaller sampling error.

CHARACTERISTICS OF A GOOD SAMPLE DESIGN

- Sample design must result in a truly representative sample.
- Sample design must be such which results in a small sampling error.
- Sample design must be viable in the context of funds available for the research study.
- Sample design must be such so that systematic bias can be controlled in a better way.
- Sample should be such that the results of the sample study can be applied, in general, for the universe with a reasonable level of confidence.

DETERMINATION OF SAMPLE SIZE

1) Nature of Universe

Universe may be either homogenous or heterogeneous in nature. If the items of the universe are homogenous, a small sample can serve the purpose. But if the items are heterogeneous, a large sample would be required. Technically, this can be termed as the dispersion factor.

2) Number of Classes Proposed

If many class-groups are to be formed, a large sample would be required because a small sample might not be able to give a reasonable number of items in each class-group.

3) Nature of Study

If items are to be intensively and continuously studied, the sample should be small. For a general survey the size of the sample should be large, but a small sample is considered appropriate in technical surveys.

4) Type of Sampling

Sampling technique plays an important part in determining the size of the sample. A small random sample is apt to be much superior to a larger but badly selected sample.

5) Standard of Accuracy and Acceptable Confidence Level

If the standard of accuracy or the level of precision is to be kept high, we shall require relatively larger sample. For doubling the accuracy for a fixed significance level, the sample size has to be increased fourfold.

6) Availability of Finance

In practice, size of the sample depends upon the amount of money available for the study purposes. This factor should be kept in view while determining the size of sample for large samples result in increasing the cost of sampling estimates.

7) Other Considerations

Nature of units, size of the population, size of questionnaire, availability of trained investigators, the conditions under which the sample is being conducted, the time available for completion of the study are a few other considerations to which a researcher must pay attention while selecting the size of sample.

SAMPLING TECHNIQUES

1) Probability Sampling

In probability sample, every unit in the population has equal chances for being selected as a sample unit.

2) Non–probability Sampling

In non probability sampling, units in the population has unequal or zero chances for being selected as a sample unit.

RANDOM SAMPLING PROCEDURES

The importance of randomness in sampling needs no emphasis, It is a means for securing a representative sample. How can a random sample be drawn? The layman tends to think that random sampling means picking out units "at random", i.e., in a haphazard or hit-and-miss way. Experience shows that the human being is an extremely poor instrument for the conduct of a random selection. To ensure true randomness the method of selection must be independent of human judgement. There are basic procedures.

1) Lottery Method

This is the simplest and most familiar procedure of random sampling. If a sample of 10 students is to be drawn out of a list of 50 students in a section, take 50 equal size chips or slips of paper; number them from 1 to 50 each bearing only one number. Roll each slip. Put the rolled slips in a global container and thoroughly shuffle or mix them. Take 10 chips from the container one after another. Each time before drawing a chip, mix the chips in the container thoroughly. The units bearing the numbers of chips drawn constitute the random sample.

(i) Sampling with Replacement

After a number is selected by draw, it may be replaced and consequently it has a chance of being selected again. Such method is known as sampling with replacement. This is usually referred to as unrestricted random sampling.

(ii) Sampling without Replacement

Selected numbers is set aside, and so in subsequent draws, it does not get a chance of being selected again. This type of sampling is known as sampling without replacement. This is a form of restricted sampling.

2) Use of Table o Random Numbers

This is a less cumbersome, but equally valid procedure of sample selection. Tables of random numbers have been developed by Kendall and Smith (1939), Fisher and Yates (1963) and Tippett (1927). To select a random sample out of a given frame, one should simply start to read numbers from a Table of Random Numbers at any randomly selected point and pick out numbers within the range of the frame. Let us suppose that random sample of 50 is to be selected from a College populations of 500 Commerce Students. We can use any table of random numbers.

10	09	73	25	33	76	52	01
37	54	20	48	05	64	89	47
08	42	26	89	53	19	64	50
09	01	90	25	29	09	37	67
12	80	79	99	70	80	15	73
66	06	57	47	17	34	07	27
31	06	01	08	05	45	57	18
85	26	97	76	02	02	05	16
63	57	33	21	35	05	32	54
73	79	64	57	53	03	52	96

Let us suppose, we start at the top of the left hand second column. As the population consists of a three-digit figure, read three-digit columns, i.e., read 097, 542, 422, 019 and so on. All the numbers within the range of 1 to 500 may be picked out. Then the sample will consist of:

097, 422, 019, 065. 060, 269 and so on. In the above reading, 542, 807, 573, etc., are rejected because they are over 500.

When the researcher reaches the bottom of a column, he can simply move one digit to the right and start at the top of the column again, and read numbers in three-digits: 973, 420, 226, 190, 079 and so on.

The main advantage of the use of a Table of Random Numbers are:

Easy to use and ready accessibility

The Table of Random numbers is ideal for obtaining a random sample from relatively small populations. When populations are quite large say lakhs or crores, drawing numbers from the table becomes tedious.

3) Use of Computer

If the population is very large and if computer facilities are available, a computer may be used for drawing a random sample. The computer can be programmed to print out a series of random numbers as the researcher desires.

TYPES OF SAMPLING

A) Probability Sampling

1) Simple Random Sampling

This sampling technique gives each element an equal and independent chance of being selected. An equal chance means equal probability of selection, e.g., in a population of 300, each element theoretically has $1/300^{th}$ chance of being selected. In a population of 1000, each element has $1/1000^{th}$ chance of being selected. Equal probability selection method is described as Epsem sampling. An independent choice means that the draw of one element will not affect the chances of other elements being selected.

Where some elements are purposely excluded from the sample, the resulting sample is not a random one, Hence, all elements should be included in the sample frame to draw a random sample.

Merits

1. All elements in the population have an equal chance of being selected
2. Of all the probability sampling techniques, simple random sampling is the easiest to apply
3. It is the most simple type of probability sampling to understand
4. It does not required a prior knowledge of the true composition of the population
5. The amount of sampling error associated with any sample drawn can easily be computed

Demerits

1. It is often impractical, because of non-availability of population list, or of difficulty in enumerating the population. For example, it is difficult to get a current accurate list of households in a city of a list of landless rural agricultural labourers who migrate from area to area in search of employment or a list of households of a nomadic tribe
2. The use of simple random sampling may be wasteful because we fail to use all of the known information about the population
3. This technique does not ensure proportionate representation to various groups constituting the population
4. The sampling error in this sampling is greater than that in other probability samples of the same size, because it is less precise than other methods
5. The size of the sample required to ensure its representativeness is equally larger under this type of sampling than under other random sampling techniques
6. A simple random design may be expensive in time and money

2) Stratified Random Sampling

This is an improved type of random sampling. In this method, the population is sub-divided into homogenous groups or strata, and from each stratum, random sample is drawn. For example university students may be divided on the basis of discipline, and each discipline group may again be divided into juniors and seniors; and the employees of a business undertaking may be divided into managers and non-managers and each of those two groups may be sub-divided into salary-grade wise strata.

a) Proportionate Stratified Sampling

This sampling involves drawing a sample from each stratum in proportion to the latter's share in the total population.

Specialization	No. of Students	Proportion of Each Stream
Production	40	0.4
Finance	20	0.2
Marketing	30	0.3
Rural Development	10	0.1
Total	100	1.0

Merits

1. It enhances the representativeness of the sample by giving proper representation to all sub-groups in the population
2. It gives higher statistical efficiency that the given by simple random sampling for a given sample size
3. It is easy to carry out this sample method
4. This method gives a self-weighing sample, the population mean can be estimated simply by calculating the sample mean

Demerits

1. A prior knowledge of the composition of the population and the distribution of the population characteristics are required to adopt this method
2. This method is very expensive in time and money. Of course its greater efficiency may offset the additional cost
3. The identification of the strata might lead to classification errors. Some elements maybe included into the wrong strata. This may vitiate the interpretation of survey results.

b) Disproportionate Stratified Sampling

This method does not give proportionate representation to strata (group). It necessarily involves giving over representation to some strata and under representation to others. There may be several disproportionate schemes. All strata may be given equal weight, even though their shares in the total population vary. Alternatively some substrata may be given greater weight and others lesser weight. When is such disproportionate weighing preferable? **Example :** Drawing one per cent as sample irrespective on the numbers of members in the sample

Merits

1. It is less time consuming compared with proportionate sampling, because the researcher is not necessarily concerned about the proportionate representativeness of his resulting sample as in the latter method
2. It facilitates giving appropriate weighting to particular groups, which are small but more important

Demerits

1. This method does not give each stratum proportionate representation. Hence, the resulting sample may be less representative
2. This method requires a prior knowledge of the composition of the population, which is not always possible

3. This method is also subject to classification errors. It is possible that the researcher may misclassify certain elements
4. Though disproportionate sampling is a means for developing an optimal stratification scheme, its practical feasibility is doubtful because one generally does not know the relative variability in the strata nor the relative costs

3) Systematic Sampling

This method of sampling is an alternative to random sampling. It consists of taking every K^{th} item in the population after a random start with an item form 1 to K. **For example**, suppose it is desired to select a sample of 20 students, from a list of 300 students, divide the population total of 300 by 20, the quotient is 15. Select a number at random between 1 and 15, using lottery method or a table of random numbers. Suppose the selected number is 9. Then the students numbered 9, 24, 39 are selected as the sample.

As the interval between sample units is fixed, this method is also known as fixed interval method.

Merits

1. It is much simpler than random sampling. It is easy to use
2. It is easy to instruct to field investigators to use this method
3. This method may require less time. A researcher operating on a limited time schedule will prefer this method
4. This method is cheaper than simple random sampling
5. It is easier to check whether every ‘k’th has been included in the sample
6. Sample is spread evenly over the population
7. It is statistically more efficient than a simple random sample when population elements are ordered chronologically, by size, class, etc., Then systematic sampling gives a better representative sample

Demerits

1. This method ignores all elements between two 'k'th elements selected. Further, except the first element, other selected elements are not chosen at random. Hence, this sampling cannot be considered to be a probably sampling in the strict sense of the term
2. As each element does not have an equal chance of being selected, the resulting sample is not a random one. For studies aiming at estimation or generalizations, this disadvantage would be serious one
3. This method may sometimes give a biased sample. If by chance, several 'k' th elements chosen represent a particular group, that group would be over-represented in the sample

4) Cluster Sampling

Where the population elements are scattered over a wider area and a list of population elements is not readily available, the use of simple or stratified random sampling method would be too expensive and time consuming. In such cases cluster sampling is usually adopted.

Cluster sampling means random selection of sampling units consisting of population elements. Each such sampling unit is a cluster of population elements. Then from each selected sampling unit, a sample of population elements is drawn by either simple random selection or stratified random selection.

Example: Suppose a researcher wants to select a random sample of 1000 households out of 40000 estimated households in a city for a survey. A direct sample of individual households would be difficult to select, because a list of households does not exist and would be too costly to prepare. Instead, he can select a random sample of a few blocks / wards. The number of blocks to be selected depends upon the average number of estimated households per block. Suppose the average number of households per block is 200, then 5 blocks comprise the sample. Since the number of households per block varies, the actual sample size depends on the block which happen to be selected. Alternatively, he can draw a sample of more blocks and from each blocks a certain number of households may be selected by systematic sampling.

Merits

1. This method is much easier and more convenient to apply when large populations are studied or large geographical areas are covered. Even a ready list of population elements is not necessary. A researcher can simply draw a random sample of geographical sections and adopt single or multistage sampling depending on the vastness of the area covered by the study
2. The cost of this method is much less when compared with other sampling methods
3. This method promotes the convenience of field work at it could be done in compact places
4. Sampling under this method does not require more time
5. Units of study can be readily substituted for other units within the same random section
6. This method is flexible. Where it involves multistage sampling, it is possible to employ different types of sampling in successive stages

Demerits

1. The cluster size may vary and this variation could increase the bias of the resulting sample. For example, if the researcher were to interview all adults in households in each selected street the number of adults would vary from house to house. There would be certain bias resulting from the large coverage of big families
2. The sampling error in this method of sampling is greater. Thus, this method is statistically less efficient than other probability sampling methods
3. Adjacent units of study (e.g. households) tend to have more similar characteristics than do units distantly apart. This affects the 'representativeness' of the sample and this effect is reflected in a greater sampling error.

5) Area Sampling

This is an important form of cluster sampling. In larger field surveys, clusters consisting of specific geographical areas like districts, taluks, villages or blocks in a city are randomly drawn. As the geographical areas are selected as sampling units in such cases, their sampling is called area sampling. It is not a separate method of sampling, but forms a part of cluster sampling.

In a country like India where a state (previously known as province) is divided into districts, districts into talukas and talukas into towns and villages, area sampling is done on the basis of these administrative units in multi-stages.

6) Multi-stage Sampling

In this method, sampling is carried out in two or more stages. The population is regarded as being composed of a number of first stage sampling units. Each of them is made up of a number of second stage units and so forth. That is, at each stage, a sampling unit is a cluster of the sampling units of the subsequent stage. First, a sample of the first stage sampling units is drawn, then from each of the selected first stage sampling unit, a sample of the second stage sampling units is drawn. The procedure continues down to the final sampling units or population elements. Appropriate random sampling method is adopted at each stage.

Merits

It results in concentration of fieldwork in compact small areas and consequently in a saving of time, labour and money

1. It is more convenient, efficient and flexible than single-stage sampling
2. It obviates the necessity of having a sampling frame covered the entire population

Demerits

The major disadvantage of the multi-stage sampling is that the procedure of estimating sampling error and cost advantage is complicated. It is difficult for a non-statistician to follow estimation procedure.

B. Non-Probability Sampling

1) Convenience Sampling

This is non-probability sampling. It means selecting sample units in a just 'hit an miss' fashion. E.g. Interviewing people whom we happen to meet. This sampling also means selecting whatever sampling units are conveniently available e.g. a teacher may select students in his class. This method is also known as accidental sampling because the respondents whom the researcher meets accidentally are included in the sample.

Merits

1. Cheapest and simplest
2. It does not require a list of population
3. It does not require any statistical expertise

Demerits

1. Convenience sampling is highly biased, because of the researcher's subjectivity, and so it does not yield a representative sample
2. It is the least reliable sampling method. There is no way of estimating the representativeness of the sample
3. The findings cannot be generalized

2) Purposed or Judgement Sampling

This method means deliberate selection of sample units that conform to some pre-determined criteria. This is known as judgement sampling. This involves selection of cases which we judge as the most appropriate ones for the given study. It is based on the judgement of the researcher or some expert. It does not aim at securing a cross section of a population.

The chance that a particular case be selected for the sample depends on the subjective judgement of the researcher. For example, A researcher may deliberately choose industrial undertakings in which quality circles are believed to be functioning successfully and undertakings in which quality circles are believed to be a total failure

Merits

1. It is less costly and more convenient
2. It guarantees inclusion of relevant elements in the sample. Probability sampling plans cannot give such guarantee

Demerits

1. This does not ensure the representativeness of the sample
2. This is less efficient for generalizing when compared with random sampling
3. This method requires more prior extensive information about the population one studies. Without such information, it is not possible to adjudge the suitability of the sample items to be selected
4. The method does not lend itself for using inferential statistics, because, this sampling does not satisfy the underlying assumption of randomness.

3) Quota Sampling

This is a form of convenient sampling involving selection of quota groups of accessible sampling units by traits such as sex, age, social class etc., when the population is known to consist of various categories by sex, age, religion, social class, etc., in specific proportions, each investigator may be given an assignment of quota groups specified by the pre-determined traits in specific proportions. He can then select accessible persons, belonging to those quota groups in the area assigned to him.

Example

Sex	Numbers	Age		Social Class	
Male	11	20-40	5	Higher	3
Female	9	41-50	8	Middle	10
	20	51-60	4	Lower	7
		Above 60	3		**20**
			20		

Merits

1. It is considerably less costly than probability sampling
2. It takes less time
3. There is no need for a list of population. Thus, quota sampling is a suitable method of sampling a population for which no suitable frame is available
4. Field work can easily be organized. Strict supervision need not be required

Demerits

1. It may not yield a precise representative sample, and it is impossible to estimate sampling error. The findings, therefore, are not generalizable to any significant extent
2. Interviewers may tent to choose the most accessible persons; they may ignore slums or areas difficult to reach. Thus, they may fail to secure a representative sample within their quota groups
3. Strict control of field work is difficult
4. The quota of sampling is subject to higher degree of classification error, because the investigators are likely to base their classification of respondents' social status and economic status mostly on their impressions about them

5. It is difficult for sampling on more than three variable dimensions. This is because the number of categories to be selected is a multiplication of the number of values in each variable. For instances, if we want to sample proportionate number of persons by sex, social status and age and these variables consist of two, three and three categories respectively.

4) Snowball Sampling

This is the colourful name for a technique of building up a list or a sample of a special population by using an initial set of its members as informants. For example, if a researcher wants to study the problem faced by Indians through some source like Indian Embassy. Then he can ask each one of them to supply names of other Indians known to them, and continue this procedure until he gets an exhaustive list from which he can draw a sample or make a census survey.

This sampling technique may also be used in socio-metric studies. For example, the members of a social group may be asked to name the persons with whom they have social contacts, each one of the persons so named may also be asked to do so, and so on. The researcher may thus get a constellation of associates and analyse it.

Merits

1. It is very useful in studying social groups, informal group in a formal organization, and diffusion of information among professionals of various kinds
2. It is useful for smaller populations for which no frames are readily available

Demerits

1. The major disadvantages of snowball sampling is that it does not allow the use of probability statistical methods. Elements included are dependent on the subjective choice of the original selected respondents
2. It is difficult to apply this method when the population is large
3. It does not ensure the inclusion of all elements in the lists

SAMPLING ERROR

A survey aims at estimating or inferring selected population characteristics or parameters by studying either population or a sample of the population. The research results may either from the 'true values' of the parameters under study. Such differences are known as Errors and Biases. The errors of a survey may be classified into (a) Sampling Errors (b) Sampling Biases (c) Non-sampling errors and (d) Non-sampling biases.

1) Sampling Errors

The errors which arise because of studying only a part of the total population are called sampling errors. These may arise due to non-representativeness of the samples and the inadequacy of sample size. When several samples are drawn from a population, their results would not be identical. The degree of variation of sample results is measured by standard deviation and it is known as the standard error of the concerned statistic. As sample size increases the magnitude of the error decreases. Sample size and sampling error are thus negatively correlated.

2) Sampling Biases

The average of the estimates of a population parameter derived from an infinite number of samples is called the expected value of the estimator. The difference between this value and 'true value' of the parameter is the bias. Bias may arise (1) if the sampling is done by a non-random method. (2) if the sampling frame is incomplete or inaccurate and (3) if some sections of the population are not available / refuse to cooperate. Any of these factors will cause non-compensating errors which cannot be reduced by an increase in sample size. The only sure way of avoiding bias arising through the sampling method is to use a random method. Randomness is an essential part of the protection against selection bias.

3) Non-Sampling Errors

These are errors which arise from sources other than sampling. They include errors of observation, errors of measurement and errors of responses. Data are collected through the methods of observation or interviewing. The physical procedures of observation or interviewing are subject to imperfection which cause errors. Measurement errors consist of errors in processing and analysis. Errors of response include incorrect responses of the respondents, mistakes in noting their response etc.,

4) Non-Sampling Biases

These biases pose problems for scientific measurement. They affect both the population sample value and account for the difference between the population value and the true value. They consist of biases of observation and non-observation, response biases and process biases. Biases of observation are caused by obtaining and recording observations incorrectly. Non-observation biases arise from failure to obtain observations on some segments of the population due to either non-coverage or non-response. The latter may be due to refusals, non-at-homes, lost forms, etc., response biases consist of biases arising from imperfections in field observation or interviewing. Processing biases are produced during coding, tabulating and computing.

5) Total Error

In sampling theory, a popular model combines sampling and non-sampling errors and biases into the Total error. This total error is the square root of the sums of squares of variable errors and squares of bias. It is often called the root means square error. The variable error are caused only by sampling errors, and VE equals the standard error of sampling. Bias is mostly caused by measurement biases. The total error depends on the length of both the legs. The sampling error / standard error leg can be shortened by improving the sample design and by increasing the sample size. The length of biases leg may be reduced by improving the tool of data collection, the precision of method of data collection, filed work, coding, processing and analysis. The measurement of sampling error does not pose much problem but the measurement of non-sampling errors require special procedures and it is a costly effect. Hence, the reduction of non-sampling error is a challenge to the researchers.

POSSIBLE QUESTIONS

PART A (ONE MARKS)

(ONLINE EXAMINATION)

PART – B (8 MARKS)

1. Discuss the primary methods of data collection with examples?
2. Describe the steps in sampling design?
3. Explain the criteria for selection of good sampling design.
4. Define questionnaire method of data collection and explain its types. How it is differ from
5. Schedule method.
6. Analyze the interview method of data collection with pros and cons?
7. Explain the various types of probability sample designs?
8. Discuss the various tools used to collect the data?
9. Explain the techniques involved in data processing?
10. Describe the advantages and disadvantages of case study method?

UNIT III

SYLLABUS

Measurement and Scaling Techniques : Processing And Analysis of data – Editing – Coding – Transcription and tabulation - Statistical tools used in research – Interpretation and Report writing- types and contents and styles of report – steps in drafting reports

.SCALING TECHNIQUES

In research we quite often face measurement problem (since we want a valid measurement but may not obtain it), especially when the concepts to be measured are complex and abstract and we do not possess the standardised measurement tools. Alternatively, we can say that while measuring attitudes and opinions, we face the problem of their valid measurement. Similar problem may be faced by a researcher, of course in a lesser degree, while measuring physical or institutional concepts

As such we should study some procedures which may enable us to measure abstract concepts more accurately. This brings us to the study of scaling techniques.

MEANING OF SCALE

Scaling describes the procedures of assigning numbers to various degrees of opinion, attitude and other concepts. This can be done in two ways viz., (i) making a judgement about some characteristic of an individual and then placing him directly on a scale that has been defined in terms of that characteristic and (ii) constructing questionnaires in such a way that the score of individual's responses assigns him a place on a scale. It may be stated here that a scale is a continuum, consisting of the highest point (in terms of some characteristic e.g., preference, favourableness, etc.) and the lowest point along with several intermediate points between these two extreme points. These scale-point positions are so related to each other that when the first point happens to be the highest point, the second point indicates a higher degree in terms of a given characteristic as compared to the third point and the third point indicates a higher degree as compared to the fourth and so on

Numbers for measuring the distinctions of degree in the attitudes/opinions are, thus, assigned to individuals corresponding to their scale-positions. All this is better understood when we talk about scaling technique(s). Hence the term 'scaling' is applied to the procedures for attempting to determine quantitative measures of subjective abstract concepts. Scaling has been defined as a "procedure for the assignment of numbers (or other symbols) to a property of objects in order to impart some of the characteristics of numbers to the properties in question."

MEASUREMENT OF SCALES

1) Nominal scale

Nominal scale is simply a system of assigning number symbols to events in order to label them. The usual example of this is the assignment of numbers of basketball players in order to identify them. Such numbers cannot be considered to be associated with an ordered scale for their order is of no consequence; the numbers are just convenient labels for the particular class of events and as such have no quantitative value. Nominal scales provide convenient ways of keeping track of people, objects and events. One cannot do much with the numbers involved. For example, one cannot usefully average the numbers on the back of a group of football players and come up with a meaningful value. Neither can one usefully compare the numbers assigned to one group with the numbers assigned to another. The counting of members in each group is the only possible arithmetic operation when a nominal scale is employed. Accordingly, we are restricted to use mode as the measure of central tendency. There is no generally used measure of dispersion for nominal scales. Chi-square test is the most common test of statistical significance that can be utilized, and for the measures of correlation, the contingency coefficient can be worked out.

Nominal scale is the least powerful level of measurement. It indicates no order or distance relationship and has no arithmetic origin. A nominal scale simply describes differences between things by assigning them to categories. Nominal data are, thus, counted data. The scale wastes any information that we may have about varying degrees of attitude, skills, understandings, etc. In spite of all this, nominal scales are still very useful and are widely used in surveys and other ex-post-facto research when data are being classified by major sub-groups of the population.

2) Ordinal Scale

The lowest level of the ordered scale that is commonly used is the ordinal scale. The ordinal scale places events in order, but there is no attempt to make the intervals of the scale equal in terms of some rule. Rank orders represent ordinal scales and are frequently used in research relating to qualitative phenomena. A student's rank in his graduation class involves the use of an ordinal scale. One has to be very careful in making statement about scores based on ordinal scales. For instance, if Ram's position in his class is 10 and Mohan's position is 40, it cannot be said that Ram's position is four times as good as that of Mohan. The statement would make no sense at all. Ordinal scales only permit the ranking of items from highest to lowest. Ordinal measures have no absolute values, and the real differences between adjacent ranks may not be equal. All that can be said is that one person is higher or lower on the scale than another, but more precise comparisons cannot be made.

Thus, the use of an ordinal scale implies a statement of 'greater than' or 'less than' (an equality statement is also acceptable) without our being able to state how much greater or less. The real difference between ranks 1 and 2 may be more or less than the difference between ranks 5 and 6. Since the numbers of this scale have only a rank meaning, the appropriate measure of central tendency is the median. A percentile or quartile measure is used for measuring dispersion. Correlations are restricted to various rank order methods. Measures of statistical significance are restricted to the non-parametric methods.

3) Interval Scale

In the case of interval scale, the intervals are adjusted in terms of some rule that has been established as a basis for making the units equal. The units are equal only in so far as one accepts the assumptions on which the rule is based. Interval scales can have an arbitrary zero, but it is not possible to determine for them what may be called an absolute zero or the unique origin. The primary limitation of the interval scale is the lack of a true zero; it does not have the capacity to measure the complete absence of a trait or characteristic. The Fahrenheit scale is an example of an interval scale and shows similarities in what one can and cannot do with it. One can say that an increase in temperature from 30° to 40° involves the same increase in temperature as an increase from 60° to 70°, but one cannot say that the temperature of 60° is twice as warm as the temperature of 30° because both numbers are dependent on the fact that the zero on the scale is set arbitrarily at the temperature of the freezing point of water. The ratio of the two temperatures, 30° and 60°, means nothing because zero is an arbitrary point.

Interval scales provide more powerful measurement than ordinal scales for interval scale also incorporates the concept of equality of interval. As such more powerful statistical measures can be used with interval scales. Mean is the appropriate measure of central tendency, while standard deviation is the most widely used measure of dispersion. Product moment correlation techniques are appropriate and the generally used tests for statistical significance are the 't' test and 'F' test.

4) Ratio Scale

Ratio scales have an absolute or true zero of measurement. The term 'absolute zero' is not as precise as it was once believed to be. We can conceive of an absolute zero of length and similarly we can conceive of an absolute zero of time. For example, the zero point on a centimeter scale indicates the complete absence of length or height. But an absolute zero of temperature is theoretically unobtainable and it remains a concept existing only in the scientist's mind. The number of minor traffic-rule violations and the number of incorrect letters in a page of type script represent scores on ratio scales. Both these scales have absolute zeros and as such all minor traffic violations and all typing errors can be assumed to be equal in significance. With ratio scales involved one can make statements like "Jyoti's" typing performance was twice as good as that of "Reetu." The ratio involved does have significance and facilitates a kind of comparison which is not possible in case of an interval scale.

Ratio scale represents the actual amounts of variables. Measures of physical dimensions such as weight, height, distance, etc. are examples. Generally, all statistical techniques are usable with ratio scales and all manipulations that one can carry out with real numbers can also be carried out with ratio scale values. Multiplication and division can be used with this scale but not with other scales mentioned above. Geometric and harmonic means can be used as measures of central tendency and coefficients of variation may also be calculated.

Thus, proceeding from the nominal scale (the least precise type of scale) to ratio scale (the most precise), relevant information is obtained increasingly. If the nature of the variables permits, the researcher should use the scale that provides the most precise description. Researchers in physical sciences have the advantage to describe variables in ratio scale form but the behavioural sciences are generally limited to describe variables in interval scale form, a less precise type of measurement.

IMPORTANT SCALING TECHNIQUES

A) Rating Scale

In rating scale, the rater makes judgment about some characteristic of a subject and places him directly on some point on the scale. Rating scale may be either a graphic rating scale or an itemized random scale.

1) Graphic Rating Scale

It is quite simple and various points are usually put along the line to form a continuum and the rater indicates his rating by making a tick mark at the appropriate point on a line that runs from one extreme to other.

2) The Itemized Rating Scale

It is also known as numerical scale, presents a series of statements from which a respondent select one as best reflecting his evaluation. These statements are ordered progressively in terms of more or less of some property. It provide more information and meaning to the rater and increases reliability. This form is difficult to develop and the statement may not say exactly what the respondent would like to express.

The advantage of rating scale is that the results attained from these scales are comparable favorably with alternative methods. They require less time, interesting to us and have wide range of applications. One of the disadvantages of rating scales is that it suffers from lack of reliability and validity.

B) Attitude Scales

The attitude scales are constructed with sets of rating scales designed to measures one or more aspects of an individual's or group's attitude towards some objects. The individual responses to various scales may be aggregated to provide a single attitude for the individual. There are different attitude scales. They are

1) Likert's Summated Scale

It was developed by Likert, which is frequently used in the measurement of social attitude. It uses only the definitely favorable and unfavorable statement and does not take into account the intermediate position, and the respondent is asked to react. The respondent indicates his agreements or disagreement with each statement. Each response is given a numerical score and the score are totaled to measure the respondent's attitude. The overall scale represents the respondent's position on the continuum of favorable – unfavorable towards an issue.

Half of the statements (usually 15) included in the questionnaire are favorable and the rest ones are unfavorable.

Strongly Agree	-5	or	+2
Agree	-4	or	+1
Undecided	-3	or	0
Disagree	-2	or	-1
Strongly Disagree	-1	or	-2

Procedure

i. At the out set, subjects are dividing in to a couple of arbitrarily defined groups. For instance, those subjects with top 25% of all total scores and those with the lowest 25% of all total scores are constructed to be in possession of the most favorable and least favorable attitudes

ii. Thereafter the researcher calculates the mean score for each statement separately

iii. The difference between the two mean scores, in respect of each statement is calculated

iv. Finally, all statements are ranked according to their difference in mean scores. Those with mean differences near zero are considered poor and therefore eliminated.

Advantages

i. Item analysis increases the degree of homogeneity or internal consistency in the set of statements

ii. This method is less difficult

iii. Since a wide range of answers are given to the subject, they don't find it difficult to respond and express the intensity of their feeling

iv. Since there is no involvement of the outside group of judges in selection of the statements, it does not suffer from the problem of subjectivity

Disadvantages

i. Ties in rank are likely to occur quite frequently due to equality in total score values

ii. The response pattern of two persons having exactly identical scores may be significantly different

iii. It suffers from the problem of interpretation which does not arise in Thurstone's scale

iv. The subject is required to respond to all statement, whereas in Thurstone scale, he is required to check only those statement with which he agrees

2) Thurstone's Equal Appearing Interval Scale

L.L. Thurstone is the inventor of this scale. This scale consists of 15 to 20 statements which form a continuum of attitudes towards a subject ranging from the most favorable to the least favorable.

Procedure

i. The researcher gather a large number of statements, usually 20 or more, that express various points of view towards a group, idea or practice

ii. These statements are then submitted to a panel of judges, each of whom arranges them in eleven group or piles ranging from one extreme to another in position. Each of the judges is requested to place generally in the first pile the statements which he thinks are unfavorable issue, in the second pile to place those statement which he thinks all next most unfavorable and he goes on doing so in this manner till in the 11^{th} pile he put the statement which he considers to be most favorable

iii. This sorting by each judge a composite position for each of the items

iv. For items that are retained, each is given its median scale value between one and 11^{th} as established by the panel. In other words the scale value of any one statement is computed as the median position to which it is assigned by the group of judges

v. A final selection of statement is then made. For this purpose a sample of statement, whose median scores is spread evenly from one extreme to the other is taken. The statement so selected constitutes the final scale to be administered to respondents

vi. After developing the scale, the respondent is asked during the administration of scale to check the statements with which they agree. Respondents score is equal to the average of scale values attached to the items he endorses. This average is either median or mean.

Disadvantages

i. The procedure involved in the construction is very costly and time consuming

ii. The scale does not allow subjects to express the intensity of their feelings

iii. The scale values assigned to statements are influenced by the attitudes, background and intelligence of judges who see things differently from actual respondents

iv. This method is not completely objective, it involves subjective decision

3) Bogardus's Scale of Social Distance

Bogardus used cumulative scale containing a number of questions regarding a particular issue. The respondent is required to express his agreement or disagreement over that issue. The respondents who answer favorably have higher aggregate score than those who answer unfavorably. The score is computed by counting the number of items which is responded favorably by the respondent. The respondent is placed on a particular position on the scale on the basis of his scores. Bogardus used this scale to know social distance by measuring the attitude of individual towards a particular social group. In this scale a number of suggested relationships are listed to which member of an ethic group are admitted. The respondent is asked to indicate to which social group is admissible to him for each specified relationship in terms of his willingness to accept social distance.

The respondent is asked to tick off against each of 7 categories of relationship, he is willing to accept a average number of a particular ethic, social or nationality group, because the respondents first feeling reactions are known by this. In this scale the respondent has to express his reaction to each race as a group without having any regard to any individual members of a group, whether he likes of dislikes.

In order to calculate social distance mathematically weights are assigned to different categories of relationship.

Procedure

i. The weight and percentage response in respect of each category are to be placed in rows
ii. The percentage responses are to be multiplied by its weight
iii. The product is to be summed up so as to indicate social distance

Limitations

i. Although it is expected that the respondent has to express his reaction to each race as a group without having any regard to any individual members of that group, in reality the influence of any individual members may not be wiped out from mind of respondent while giving preferences
ii. The score is this scale does not indicates the actual extent or exact degree of preference of a group
iii. It is not always possible on the part of the respondent to be acquainted completely with a group, difficulty to express his attitude towards that group

4) Guttman's Cumulative Scale

Cumulative scale or Louis Guttmann's scalograms analysis, consist of a series of statements to which a respondent expresses his agreement or disagreement. The special feature is that they are cumulative in nature. A respondent replies favorably to item no.3 and also to item no. 2 and 1. The score of the individual is worked out by counting the number of point concerning the number of statement he answers favorably. If the investigator knows this total score, he can estimate as to how a respondent has answered individual statements consulting cumulative scales. The major scale of this type of cumulative scales is the Guttman's scalogram.

Scalogram analysis refers to the procedure for determining whether a set of items forms an unidimensional scale. A scale is said to be unidimensional if the responses fall into a pattern in which endorsement of item reflecting extreme position results. The perfect scale in terms of Guttman's technique implies that an informant who responds to given question will have a high total score than informants who responded to it negatively.

Procedure

i. The area of concept in first defined with reference to problem in hand
ii. Ten or twelve statement are selected which are assumed to be representative of selected area
iii. The statements are arranged in form of 3 or 5 point scale so that subjects can indicates the intensity of their attitude for each item
iv. 10 or 12 items are submitted to a sample of 100 or more respondents who will check the items. This step is designed to determine the scalability of item
v. After the total score is attained for each person by adding up the weights of categories checked, the questionnaires are arranged in rank order from high to low according to total score
vi. A table of scalogram is presented from data as the questionnaire by recording separately the response of each person to each category by providing a column for each person and a row for each category
vii. The reproducibility of each item is determined on the basis that none with a tower score ranks higher in any item than any person with a higher total score. By this method the error of reproducibility is minimized.

Limitation

i. This method is not frequently used for simple reason that its development procedure is tedious and complex
ii. As regards the measurement of attitude towards some objects or prediction of behaviour relating to such objects, the unidimensional scale may not always prove to be most effective basis.

Measurement is a systematic way of assigning number or names to object and their attitudes. And scaling is the procedure for determining the quantitative measure of abstract concepts. A scale consists of a set of statements logically related referring to same attitude. Thus a scale may be used to measure the characteristic of respondent or to evaluate object presented to him.

PROCESSING OF DATA

The data, after collection, has to be processed and analyzed in accordance with the outline laid down for the purpose at the time of developing the research plan. This is essential for a scientific study and for ensuring that we have all relevant data for making contemplated comparisons and analysis. Technically speaking, processing implies editing, coding, classification and tabulation of collected data so that they are amenable to analysis. The term analysis refers to the computation of certain measures along with searching for patterns of relationship that exist among data-groups. Thus, "in the process of analysis, relationships or differences supporting or conflicting with original or new hypotheses should be subjected to statistical tests of significance to determine with what validity data can be said to indicate any conclusions". But there are persons (Selltiz, Jahoda and others) who do not like to make difference between processing and analysis. They opine that analysis of data in a general way involves a number of closely related operations which are performed with the purpose of summarising the collected data and organising these in such a manner that they answer the research question(s). We, however, shall prefer to observe the difference between the two terms as stated here in order to understand their implications more clearly.

EDITING

Editing is a process of examining the collected data to detect errors and omissions and to correct these when possible. As a matter of fact, editing involves a careful scrutiny of the completed questionnaire / schedules. Editing is done to assure that the data are accurate, consistent (constant), uniformly entered (Male=1, Female=2), as completed as possible and have been well arranged to facilitate coding and tabulation

Types of Editing

A) Field Editing

1. Field Editing – is the review of the reporting forms by the investigator for completing (translating or rewriting) what the latter has written in abbreviated or illegible (unreadable) form at the time of recording the respondent's responses
2. This type of editing is necessary in view of the fact that individual writing styles often can be difficult for others to decipher (read)
3. This sort of editing should be done as soon as possible after the interview, preferably on the very day or on the next day
4. While doing field editing, the investigator must restrain (control) himself and must not correct error of omission by simply guessing what the informant would have said if the question had been asked

B) Central Editing

1. Central Editing **–** should take place when all forms of schedules have been completed and returned to the office
2. This type of editing implies that all forms should get a through editing by a single editor in a small study and by a team of editors in case of large enquiry
3. Editor may correct the obvious (noticeable) errors (ex.) entry in the wrong place, entry recorded in the months when it should have been recorded in weeks,

4. In case of inappropriate (Wrong) on missing replies, the editor can sometimes determine the proper answer by reviewing the other information in the schedule
5. At times, the respondent can be contacted for clarification
6. The editor must strike out the answer if the same is inappropriate and he has no basis for determining the correct answer or the response. In such a case an editing entry of 'No answer' is called for
7. All wrong replies, which are quite obvious (unclear) must be dropped from the final results

Points to be Considered by Editor While Editing of Data

1. They should be familiar with instructions given to the interviewers and coders as well as with the editing instructions supplied to them for the purpose
2. They must make entries on the form in some distinctive (unique) and that too in a standardized form
3. They should initial all answers which they change or supply
4. Editor's initials and the date of editing should be placed on each completed form or schedule
5. While crossing-out (removing or deleting) an original entry for one reason or another, they should just draw a single line on it so that the same may remain legible

CODING

Coding refers to the process of assigning numerals or other symbols to answers so that responses can be put into a limited number of categories or classes. Such classes should be appropriate to the research problem under consideration. They must also possess the characteristics of exhaustiveness and also that of mutual exclusively which means that a specific answers can be placed in one and only one cell in a given category set. Another rule to be observed is that of unidimensionality by which is meant that every class is defined in terms of only one concept

Coding is necessary for efficient analysis and through it the several replies may be reduced to a small number of classes which contain the critical information required for analysis. Coding decisions should usually be taken at the designing stage or the questionnaire. This makes it possible to precode the questionnaire choices and which in turn is helpful for computer tabulation as one can straight forward key punch from the original questionnaire. But in case of hand coding some standard method may be used. One such standard method is to code in the margin with a colored pencil. The other method can be to transcribe the data from the questionnaire to a coding sheet. Whatever method is adopted, one should see that coding errors are altogether eliminated or reduced to the minimum level.

CLASSIFICATION

Most research studies result in a large volume of raw data which must be reduced into homogeneous groups if we are to get meaningful relationships. This fact necessitates classification of data which happens to be the process of arranging data in groups or classes on the basis of common characteristics. Data having a common characteristic are placed in one class and in this way the entire data get divided into a number of groups or classes. Classification can be one of the following two types, depending upon the nature of the phenomenon involved:

1) Classification according to Attributes

As stated above, data are classified on the basis of common characteristics which can either be descriptive (such as literacy, sex, honesty, etc.) or numerical (such as weight, height, income, etc.). Descriptive characteristics refer to qualitative phenomenon which cannot be measured quantitatively; only their presence or absence in an individual item can be noticed. Data obtained this way on the basis of certain attributes are known as statistics of attributes and their classification is said to be classification according to attributes. Such classification can be simple classification or manifold classification. In simple classification we consider only one attribute and divide the universe into two classes—one class consisting of items possessing the given attribute and the other class consisting of items which do not possess the given attribute. But in manifold classification we consider two or more attributes simultaneously, and divide that data into a number of classes (total number of classes of final order is given by 2n, where n = number of attributes considered).* Whenever data are classified according to attributes, the researcher must see that the attributes are defined in such a manner that there is least possibility of any doubt/ambiguity concerning the said attributes.

2) Classification according to Class-Intervals

Unlike descriptive characteristics, the numerical characteristics refer to quantitative phenomenon which can be measured through some statistical units. Data relating to income, production, age, weight, etc. come under this category. Such data are known as statistics of variables and are classified on the basis of class intervals. For instance, persons whose incomes, say, are within Rs 201 to Rs 400 can form one group, those whose incomes are within Rs 401 to Rs 600 can form another group and so on. In this way the entire data may be divided into a number of groups or classes or what are usually called, 'class-intervals.' Each group of class-interval, thus, has an upper limit as well as a lower limit which are known as class limits. The difference between the two class limits is known as class magnitude. We may have classes with equal class magnitudes or with unequal class magnitudes. The number of items which fall in a given class is known as the frequency of the given class. All the classes or groups, with their respective frequencies taken together and put in the form of a table, are described as group frequency distribution or simply frequency distribution. Classification according to class intervals usually involves the following three main problems:

i) How may classes should be there? What should be their magnitudes?

There can be no specific answer with regard to the number of classes. The decision about this calls for skill and experience of the researcher. However, the objective should be to display the data in such a way as to make it meaningful for the analyst. Typically, we may have 5 to 15 classes. With regard to the second part of the question, we can say that, to the extent possible, class-intervals should be of equal magnitudes, but in some cases unequal magnitudes may result in better classification. Hence the researcher's objective judgement plays an important part in this connection. Multiples of 2, 5 and 10 are generally preferred while determining class magnitudes. Some statisticians adopt the following formula, suggested by H.A. Sturges, determining the size of class interval:

$$i = R/(1 + 3.3 \log N)$$

where

i = size of class interval;

R = Range (i.e., difference between the values of the largest item and smallest item among the given items);

N = Number of items to be grouped.

It should also be kept in mind that in case one or two or very few items have very high or very low values, one may use what are known as open-ended intervals in the overall frequency distribution. Such intervals may be expressed like under Rs 500 or Rs 10001 and over. Such intervals are generally not desirable, but often cannot be avoided. The researcher must always remain conscious of this fact while deciding the issue of the total number of class intervals in which the data are to be classified.

ii) How to Choose Class Limits?

While choosing class limits, the researcher must take into consideration the criterion that the mid-point (generally worked out first by taking the sum of the upper limit and lower limit of a class and then divide this sum by 2) of a class-interval and the actual average of items of that class interval should remain as close to each other as possible. Consistent with this, the class limits should be located at multiples of 2, 5, 10, 20, 100 and such other figures. Class limits may generally be stated in any of the following forms:

Exclusive type class intervals: They are usually stated as follows:

10–20

20–30

30–40

40–50

The above intervals should be read as under:

10 and under 20

20 and under 30

30 and under 40

40 and under 50

Thus, under the exclusive type class intervals, the items whose values are equal to the upper limit of a class are grouped in the next higher class. For example, an item whose value is exactly 30 would be put in 30–40 class interval and not in 20–30 class interval. In simple words, we can say that under exclusive type class intervals, the upper limit of a class interval is excluded and items with values less than the upper limit (but not less than the lower limit) are put in the given class interval.

Inclusive type class intervals: They are usually stated as follows:

11–20

21–30

31–40

41–50

In inclusive type class intervals the upper limit of a class interval is also included in the concerning class interval. Thus, an item whose value is 20 will be put in 11–20 class interval. The stated upper limit of the class interval 11–20 is 20 but the real limit is 20.99999 and as such 11–20 class interval really means 11 and under 21.

When the phenomenon under consideration happens to be a discrete one (i.e., can be measured and stated only in integers), then we should adopt inclusive type classification. But when the phenomenon happens to be a continuous one capable of being measured in fractions as well, we can use exclusive type class intervals.

iii) How to determine the Frequency of Each Class?

This can be done either by tally sheets or by mechanical aids. Under the technique of tally sheet, the class-groups are written on a sheet of paper (commonly known as the tally sheet) and for each item a stroke (usually a small vertical line) is marked against the class group in which it falls. The general practice is that after every four small vertical lines in a class group, the fifth line for the item falling in the same group, is indicated as horizontal line through the said four lines and the resulting flower (IIII) represents five items. All this facilitates the counting of items in each one of the class groups.

Alternatively, class frequencies can be determined, specially in case of large inquires and surveys, by mechanical aids i.e., with the help of machines viz., sorting machines that are available for the purpose. Some machines are hand operated, whereas other work with electricity. There are machines which can sort out cards at a speed of something like 25000 cards per hour. This method is fast but expensive.

TABULATION

When a mass of data has been assembled, it becomes necessary for the researcher to arrange the same in some kind of concise (brief / short) and logical order. This is known as tabulation. Thus, tabulation is the process of summarizing raw data and displaying the same in compact form. In a broader sense, tabulation is an orderly arrangement of data in column and rows

Need for Tabulation

1. It conserves (save) space and reduce explanatory and descriptive statement to a minimum
2. It facilitates the process of comparison
3. It facilitates the summation (abstract) of items and the detection of errors and omissions
4. It provides a basis for various statistical computations (calculation)

Principles of Tabulation

1. Every table should have a clear, concise (short) and adequate title so as to make the table intelligible without reference to the text and this title should always be placed just above the body of the table
2. Every table should be given a distinct number to facilitate easy reference
3. The column and the row headings of the table should be clear and brief
4. The units of measurements under each heading or sub-heading must always be indicated (i.e.) in Rs.
5. Explanatory footnotes, if any, concerning the table should be placed directly beneath (under) the table, along with the reference symbols used in the table
6. Source or sources from where the data in the table have been obtained must be indicated juts below the table
7. Usually the columns are separated from one another by lines which make the table more readable and attractive. Lines are always drawn at the top and bottom of the table and below the captions

8. There should be thick lines to separate the data under one class from the data under another class and the lines separating the sub-divisions of the classes should be comparatively thin lines
9. The columns may be numbered to facilitate reference
10. Those columns whose data are to be compared should be kept side by side. Similarly, percentages and / or averages must also be kept close to the data
11. It is generally considered better to approximate (rough / estimated) figures before tabulation as the same would reduce unnecessary details in the table itself
12. In order to emphasize the relative significance of certain categories, different kinds of type, spacing and indentations (marks) may be used
13. It is important that all column figures be properly aligned. Decimal points and (+) or (-) signs should be in perfect alignment
14. Abbreviations should be avoided to the extent possible and ditto marks should not be used in the table
15. Miscellaneous and exceptional items, if any, should be usually placed in the last row of the table
16. Table should be made as logical, clear, accurate and simple as possible. If the data happen to be very large, they should not be crowded in a single table for that would make the table unwieldy and inconvenient
17. Total of rows should normally be placed in the extreme right column and that of column should be placed at the bottom
18. The arrangement of the categories in a table may be chronological, geographical, alphabetical or according to magnitude to facilitate comparison. Above all, the table must suit the needs and requirements of an investigation

INTERPRETATION

After collecting and analyzing the data, the researcher has to accomplish the task of drawing inferences followed by report writing. This has to be done very carefully, otherwise misleading conclusions may be drawn and the whole purpose of doing research may get vitiated. It is only through interpretation that the researcher can expose relations and processes that underlie his findings. In case of hypotheses testing studies, if hypotheses are tested and upheld several times, the researcher may arrive at generalizations. But in case the researcher had no hypothesis to start with, he would try to explain his findings on the basis of some theory. This may at times result in new questions, leading to further researches. All this analytical information and consequential inference(s) may well be communicated, preferably through research report, to the consumers of research results who may be either an individual or a group of individuals or some public/private organisation.

Meaning of Interpretation

Interpretation refers to the task of drawing inferences from the collected facts after an analytical and/or experimental study. In fact, it is a search for broader meaning of research findings. The task of interpretation has two major aspects viz., (i) the effort to establish continuity in research through linking the results of a given study with those of another and (ii) the establishment of some explanatory concepts. "In one sense, 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."1 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.

Need for Interpretation

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.
4. The interpretation of the findings of exploratory research study often results into hypotheses for experimental research and as such interpretation is involved in the transition from exploratory to experimental research. Since an exploratory study does not have a hypothesis to start with, the findings of such a study have to be interpreted on a post-factum basis in which case the interpretation is technically described as 'post factum' interpretation.

TECHNIQUES OF INTERPRETATION

Interpretation is not an easy job and it requires a great skill on the part of the investigator. The investigator gets the required expertise to apply the techniques. The techniques of interpretation are given below:

1) Relationship between Variables

The basic object of every analytical research is to find out the relationship between any two variables. There may be three types of relationship

- Symmetrical Relationship
- Reciprocal Relationship
- Asymmetrical Relationship

The interpretation of data can be made with the help of these relationships

2) Percentages

Percentages are used in making comparison between two or more series of data. They are also used to describe the relationships.

3) Averages

There are three forms of averages such as arithmetic mean, median, mode. Though there are other measures of central tendency, the above three measures are commonly used. Instead of using long statistical tables, the use of average makes the interpretation very simple.

4) Dispersion

Dispersion refers to the amount or the magnitude of the spread. Measures of dispersion include range, inter quartile range, average deviation and standard deviation. These measures help to interpret the data more scientifically

PRECAUTIONS IN INTERPRETATION

1. At the outset, researcher must invariably satisfy himself that (a) the data are appropriate, trustworthy and adequate for drawing inferences; (b) the data reflect good homogeneity; and that (c) proper analysis has been done through statistical methods
2. The researcher must remain cautious about the errors that can possibly arise in the process of interpreting results. Errors can arise due to false generalization and/or due to wrong interpretation of statistical measures, such as the application of findings beyond the range of observations, identification of correlation with causation and the like. Another major pitfall is the tendency to affirm that definite relationships exist on the basis of confirmation of particular hypotheses. In fact, the positive test results accepting the hypothesis must be interpreted as “being in accord” with the hypothesis, rather than as “confirming the validity of the hypothesis”. The researcher must remain vigilant about

all such things so that false generalization may not take place. He should be well equipped with and must know the correct use of statistical measures for drawing inferences concerning his study.

3. He must always keep in view that the task of interpretation is very much intertwined with analysis and cannot be distinctly separated. As such he must take the task of interpretation as a special aspect of analysis and accordingly must take all those precautions that one usually observes while going through the process of analysis viz., precautions concerning the reliability of data, computational checks, validation and comparison of results.
4. He must never lose sight of the fact that his task is not only to make sensitive observations of relevant occurrences, but also to identify and disengage the factors that are initially hidden to the eye. This will enable him to do his job of interpretation on proper lines. Broad generalisation should be avoided as most research is not amenable to it because the coverage may be restricted to a particular time, a particular area and particular conditions. Such restrictions, if any, must invariably be specified and the results must be framed within their limits.
5. The researcher must remember that "ideally in the course of a research study, there should be constant interaction between initial hypothesis, empirical observation and theoretical conceptions. It is exactly in this area of interaction between theoretical orientation and empirical observation that opportunities for originality and creativity lie." He must pay special attention to this aspect while engaged in the task of interpretation.

REPORT WRITING

Research report is considered a major component of the research study for the research task remains incomplete till the report has been presented and/or written. As a matter of fact even the most brilliant hypothesis, highly well designed and conducted research study, and the most striking generalizations and findings are of little value unless they are effectively communicated to others. The purpose of research is not well served unless the findings are made known to others. Research results must invariably enter the general store of knowledge. All this explains the significance of writing research report. There are people who do not consider writing of report as an integral part of the research process. But the general opinion is in favour of treating the presentation of research results or the writing of report as part and parcel of the research project. Writing of report is the last step in a research study and requires a set of skills somewhat different from those called for in respect of the earlier stages of research. This task should be accomplished by the researcher with utmost care; he may seek the assistance and guidance of experts for the purpose.

TYPES OF REPORTS

1. Research reports vary greatly in length and type. In each individual case, both the length and the form are largely dictated (determined / ordered) by the problems at hand
2. For instance, business firms prefer reports in the letter form, just one or two pages in length.
3. Banks, Insurance organizations and financial institutions are generally fond of (having a liking for) the short balance-sheet type of tabulation for their annual reports to their customers and shareholders
4. Mathematicians prefer to write the results of their investigations in the form of algebraic (numerical) notations (Symbols + / -)
5. Chemists (Scientist trained in Chemistry) report their results in symbols and formulae (method)
6. Students of literature usually write long reports presenting the critical analysis of some writer or period or the like with a liberal use of quotations from the works of the author under discussion
7. In the field of education and psychology, the favorite form in the report on the results of experimentation accompanied by the detailed statistical tabulations

8. Clinical psychologists and social pathologists (diagnosis of disease) frequently find it necessary to make use of the case-history (all the relevant information previous gatherer) form
9. News items in the daily papers are also forms of report writing. They represent firsthand on-the-scene accounts of the events described or compilation (collection) of interviews with persons who were on the scene (area). In such report the first paragraph usually contains the important information in detail and the succeeding paragraphs contain material which is progressively less and less important
10. Book – reviews which analyze the content of the book and report on the author's intentions, his success or failure in achieving his aims, his language, his style, scholarship (learning / research / study), bias or his point of view, such reviews also happen to be a kind of short report
11. The reports prepared by governmental bureaus (agency) special commissions, and similar other organizations are generally very comprehensive (full / complete) reports on the issues involved. Such reports are usually considered as important research products
12. Similarly, Ph.D.theses and dissertation are also a form of report-writing, usually completed by students in academic institutions
13. The above narration throws light on the fact that the results of a research investigation can be presented in a number of ways viz., a technical report, popular report,
14. Which method of presentation to be used in a particular study depends on the circumstances under which the study arose (take place) and the nature of the results
15. A technical report is used whenever a full written report of the study is required whether for record-keeping or for public dissimilation
16. A popular report is used if the research results have policy implications.

A) TECHNICAL REPORT

In technical report the main emphasis is on (i) the methods employed (ii) assumptions made in the course of the study (iii) the detailed presentation of the findings including their limitations and supporting data

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

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 of the problem at hand be fully assessed.

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 (explained). 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. Even in 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

B) POPULAR REPORT

The popular report is one which emphasis on simplicity and attractiveness. The simplification should be sought (required) 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 popular report

1) Findings and their Implications

Emphasis (importance) in the report is given on the findings of most practical interest and on the implication of these findings

2) Recommendation for Action

Recommendations for action on the basis of the findings of the study is made in this section of the report

3) Objectives 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 the 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

LAYOUT OF THE RESEARCH REPORT

Anybody, who is reading the research report, must necessarily be conveyed enough about the study so that he can place it in its general scientific context (background), judge the adequacy of its methods and thus form an opinion of how seriously the findings are to be taken. For this purpose there is the need for proper layout of the report. The layout of the report means as to what the research report should contain. A comprehensive (complete) layout of the research report should comprise (a) Preliminary Pages (b) The Main Text and (c) The end matter

1) Preliminary Pages

In its preliminary pages the report should carry a title and date, followed by acknowledgements in the form of 'preface' or 'foreword'. Then there should be a table of contents followed by list of tables and illustrations so that the decision-maker or anybody interested in reading the report can easily locate the required information in the report

2) Main Text

The main text provides the complete outline of the research report along with all details. Title of the research study is repeated at the top of the first page of the main text and then follows the other details on pages numbered consecutively, beginning with the second page. Each main section of the report should begin on a new page. The main text of the report should have the following sections:

i) Introduction

The purpose of introduction is to introduce the research project to the readers. It should contain a clear statement of the objectives of research i.e. enough background should be give to make clear to the reader why the problem was considered worth investigating. A brief summary of other relevant research may also be stated so that the present study can be seen in that context. The hypotheses of the study, if any, and the definitions of the major concepts employed in the study should be explicitly stated in the introduction of the report

The methodology adopted in conducting the study must be fully explained. The scientific reader would like to know in detail about thing: How was the study carried out? What was its basic design? If the study was an experimental one, then what were the experimental manipulations? If the data were collected by means of questionnaires or interviews, then exactly what questions were asked? If measurements were based on observation, then what instructions were given to the observers? Regarding the sample used in the study the reader should told: Who were the subject (Respondents)? How many were there?

How were they selected? All these questions are crucial for estimating the probable limits of generalizability of the findings. The statistical analysis adopted must also be clearly stated. In addition to all this, the scope of the study should be stated the boundary lines be demarcated. The various limitations, under which the research project was completed, must also be narrated

ii) Statement of Findings and Recommendations

After introduction, the research report must contain a statement of findings and recommendations in non-technical language so that it can be easily understood by all concerned. If the findings happen to be extensive, at this point they should be put in the summarized form

iii) Results

A detailed presentation of the findings of the study, with supporting data in the form of tables and charts together with a validation of results, is the next step in writing the main text of the report. This generally comprises the main body of the report, extending over several chapters. The result section of the report should contain statistical summaries and reductions of the data rather than the raw data. All the results should be presented in logical sequence and splitted into readily identifiable sections. All relevant results must find a place in the report. But how one is to decide about what is relevant is the basic question. Quite often guidance comes primarily from the research problem and from the hypotheses, if any, with which the study was concerned. But ultimately the researcher must rely on his own judgement in deciding the outline of his report. “Nevertheless”, it is still necessary that he states clearly the problem with which he was concerned, the procedure by which he worked on the problem, the conclusions at which he arrived, and the bases for his conclusions

iv) Implications of the Results

Toward the end of the main text, the researcher should again put down the results of his research clearly and precisely. He should, state the implications that flow from the results of the study, for the general reader is interested in the implications for understanding the human behaviour.

a. A statement of the inferences drawn from the present study which may be expected to apply in similar circumstances
b. The conditions of the present study which may limit the extent of legitimate (lawful / rightful) generalizations of the inferences drawn from the study
c. The relevant questions that still remain unanswered or new questions raised by the study along with suggestions for the kind of research that would provide answers for them

It is considered a good practice to finish the report with a short conclusion which summarizes and recapitulates (sum up) the main points of the study. The conclusion drawn from the study should be clearly related to the hypotheses that were stated in the introductory section. At the same time, a forecast of the probable future of the subject and an indication of the kind of research which needs to be done in that particular filed is useful and desirable

v) Summary

It has become customary (Usual) to conclude the research report with a very brief summary (abstract / synopsis), resting in brief the research problem, the methodology, the major findings and the major conclusions drawn from the research results.

3) End Matter

At the end of the report, appendices should be enlisted (join up) in respect of all technical data such as questionnaires, sample information, mathematical derivation and the like ones. Bibliography of sources consulted should also be given. Index should invariably be given at the end of the report. The value of index lies in the fact that it works as a guide to the reader for the contents in the report

STEPS IN WRITING REPORTS

Research reports are the product of slow, painstaking (careful / meticulous), accurate inductive (logical / reasonable) work

1) Logical Analysis of the Subject Matter

It is the first step which is primarily concerned with the development of a subject. There are two ways in which to develop a subject – (a) logically and (b) Chronologically (arranging data as per time of occurrence). The logical development is made on the basis of mental connections and associations between the one thing and another by means of analysis. Logical treatment often consists in developing the material from the simple possible to the most complex structures. Chronological development is based on a connection or sequence in time or occurrence. The directions for doing or making something usually follow the chronological order

2) Preparation of the Final Outline

It is the next step in writing the research report. Outlines (hints) are the framework upon which long written works are constructed. They are in aid (help) to the logical organization of the material and a reminder of the points to be stressed in the report.

3) Preparation of the Rough Draft

This follows the logical analysis of the subject and the preparation of the final outline. Such a step is of utmost importance for the researcher now sits to write down what he has done in the context of his research study. He will write down the procedure adopted by him in collecting the material for his study along with various limitations faced by him, the technique of analysis for adopted by him, the broad findings and generalizations and the various suggestions he wants to offer regarding the problem concerned

4) Rewriting and Polishing of the Rough Draft

This step happens to be most difficult part of all formal writing. Usually this step requires more time than the writing of the rough report. The careful revision makes the difference between a mediocre (ordinary) and a good piece of writing. While rewriting and polishing, one should check the report for weaknesses in logical development or presentation. The researcher should also 'see whether or not the material, as it is presented, has unity and cohesion (organization); does the report stand upright and firm and exhibit a definite pattern, like a marble arch? Or does it resemble an old wall of moldering cement and loose bricks. In addition, the researcher should give due attention to the fact that in his rough draft he has been consistent (reliably / steady) or not. He should check the mechanics of writing – grammar, spelling and usage

5) Preparation of the Final Bibliography

Next in order comes the task of the preparation of the final bibliography. The bibliography, which is generally appended (add on) to the research report, is a list of books in some way pertinent to the research which has been done. It should contain all those works which the researcher has consulted. The bibliography should be arranged alphabetically and may be divided into two parts; the first part may contain the names of books and pamphlets, and the second part may contain the name of magazines and newspaper articles. Generally, this pattern of bibliography is considered convenient and satisfactory from the point of view of reader, thought it is not the only way of presenting bibliography

6) Writing the Final Draft

This consists the last step. The final draft should be written in a concise (brief) and objective style and in simple language, avoiding vague expressions such as "it seems". While writing the final draft, the researcher must avoid abstract (theoretical) terminology and technical jargon. Illustrations and examples based on common experiences must be incorporated in the final draft as they happen to most effective in communicating the research findings to others. A research report should not be dull (boring), but must enthuse (motivate) people and maintain interest and must show originality. It must be remembered that every report should be an attempt to solve some intellectual problem and must contribute to the solution of a problem and must add to the knowledge of both the researcher and the reader

CONTENTS OF REPORTS

The format is divided into three broad categories namely (1) Preliminary Section (2) Main body of the text (3) Reference Section

1. The preliminary section will vary according to the type of research one has undertaken. The headings below this section are meant as a guide and one may decide to omit some of them or to amalgamate when this seems appropriate
2. The main body of the report is divided into five sub-sections, namely introduction, review of the literature, design of the study, presentation and analysis of data, summary and conclusions. Introduction is the starting point. In introduction the researcher should say why he / she undertook the research – what the problem is and why it is important. Depending on the particular piece of research, the review of the literature is introduced. This division will show what is known already and how our research will fill a gap in knowledge or replicate earlier work.
3. The description of the design of our research will depend on the particular project the researcher has undertaken. The section presentation and analysis of data will interpret the results of the research. Tables, figures will substantiate the analysis. The summary division will organize the results and point out the implications of the findings for policy or for other researchers and show what further research needs to be done.
4. The last category namely the references section should include bibliography, appendices and index in a proper form.

*****

POSSIBLE QUESTIONS

PART A (ONE MARKS)

(ONLINE EXAMINATION)

PART – B (8 MARKS)

PART B

1. Discuss the types of measurement scales with suitable example.
2. Elucidate the layout of research report covering all relevant points.
3. Explain the scaling techniques and its significance in business research in detail
4. Discuss the precautions should be followed in writing a good research report.
5. Define tabulation and explain the principles of tabulation.
6. What do you understand by research report? What are the steps followed in it?
7. Discuss the various tools used to collect the data.
8. Explain the techniques involved in data processing.
9. Explain the term tabulation and discuss the principles of tabulation.
10. Explain the significance of research report in modern times.

UNIT IV

SYLLABUS

Measures of central tendency : Mean, Median and Mode – Standard Deviation – Correlation – Regression Models

MEASURES OF CENTRAL TENDENCY

Raw data are difficult to comprehend. Classification facilitates, many a time, quick and easy understanding of diversified nature of data. A single representative value serves the purpose in a better manner.

Quantitative data in a mass exhibit certain general characteristics. They show a tendency to concentrate at certain values, usually somewhere in the centre of the distribution. Measures of this tendency are called measures of central tendency or averages. This tendency toward centralization, though not universal, has established the expression “measure of central tendency” to describe an average. The terms is imbedded in statistical language, but it is not always pertinent.

An average is a value which is typical or representative of a set of data.

A measure of central tendency gives a single representative value for a set of usually unequal values. The single value is the point of location around which the individual values of the set cluster. The measures of central tendency are hence known as ‘measures of location’. They are popularly called averages.

ARITHMETIC MEAN

Arithmetic mean is the total of the values of the items divided by their number.

Direct Method

Sum 1: The expenditure of 10 families in Rupees are given below.

Family	A	B	C	D	E	F	G	H	I	J
Exp.	30	70	10	75	500	8	42	250	40	36

Calculate the Arithmetic Mean.

Family	**Expenditure (Rs.)**
A	30

Family	Expenditure (Rs.)
B	70
C	10
D	75
E	500
F	8
G	42
H	250
I	40
J	36
Total	**$\sum X$= 1061**

AM = A + ($\sum$fd/n)

= 150 + (192 / 12)

= 150 + 16

= Rs. 166

Short Cut Method

Sum 2: The expenditure of 10 families in Rupees are given below

Family	A	B	C	D	E	F	G	H	I	J
Exp.	30	70	10	75	500	8	42	250	40	36

Solution

Family	Expenditure (Rs.) X	d=X-A; A = 100
A	30	-70
B	70	-30
C	10	-90
D	75	-25
E	500	400
F	8	-92
G	42	-58
H	250	150
I	40	-60
J	36	-64
Total		**$\sum d=61$**

N = 12 and $\sum d$ = 192

AM = A + ($\sum$d/N)

= 100 + (61/10)

= 100 + 6.10

= Rs. 106.10

Sum 3: The monthly income of 12 families in a town is given below:

S.No.	1	2	3	4	5	6	7	8	9	10	11	12
Income	280	180	96	98	104	85	80	94	100	75	600	200

Calculate the arithmetic mean by taking 150 as the assumed mean.

Solution

S.No.	Income (Rs.) X	d=X-A; A=150
1	280	130
2	180	30
3	96	-54
4	98	-52
5	104	-46
6	85	-65
7	80	-70
8	94	-56
9	100	-50
10	75	-75
11	600	450
12	200	50
Total	**-**	**192**

N = 12 and $\sum d = 192$

AM = A + ($\sum$d/N)

= 150 + (192 / 12)

= 150 + 16

= Rs. 166

Step Deviation Method

Sum 4: Find the arithmetic mean by step deviation method.

Marks	20	30	40	50	50	60	70	80	90	90

Solution

Marks X	**d=X-A/C A= 50; C=10**
20	-3
30	-2
40	-1
50	0
50	0
60	1
70	2
80	3
90	4
90	4
Total	**∑d=8**

N = 10 ; ∑d = 8

AM = A + (C∑d/N)

= 50 + (10 X 8) / 10

= 50 + 8

= 58

Arithmetic Mean for Discrete Series

Sum 5: Calculate the mean number of persons per house. Given

No. of Persons per house	2	3	4	5	6	Total

No. of Houses	10	25	30	25	10	100

Solution:

No. of Persons per house X	No. of Houses f	fx
2	10	20
3	25	75
4	30	120
5	25	125
6	10	60
Total	**N=100**	**∑fx=400**

AM = ∑fx / N

= 400 / 100

= 4

Arithmetic Mean for Discrete Series – Short Cut Method

Sum 6: Calculate the arithmetic mean.

Marks	40	50	54	60	68	80	Total
No, of Students	10	18	20	39	15	8	110

Solution

Let A = 60

Marks (X)	No. of Students (f)	d=X-A; A=60	fd

40	10	-20	-200
50	18	-10	-180
54	20	-6	-120
60	39	0	0
68	15	8	120
80	8	20	160
Total	**N=110**	**--**	**∑fd=-220**

AM = A + (∑fd/N)

= 60 + (-220 / 110)

= 60 – 2

= 58

Arithmetic Mean for Discrete Series – Step Deviation Method

Sum 7: Calculate the arithmetic mean from the following discrete series.

Daily Wages (Rs.)	75	100	120	150	200	Total
No. of Labourers	5	12	20	14	9	60

Solution

Daily Wage (Rs.) X	**No. of Labourers (f)**	**d=X-A/C; A=120 C=5**	**fd**

75	5	-9	-45
100	12	-4	-48
120	20	0	0
150	14	6	84
200	9	16	144
Total	**N=60**	**--**	**∑fd=135**

AM = A + (C∑fd/N)

= 120 + (5 X 135) / 60

= 120 + 11.25

= Rs. 131.25

Continuous Series – Exclusive Class Intervals

Sum 8: Calculate Arithmetic mean for the following:

Marks	20-30	30-40	40-50	50-60	60-70	70-80
No. of Students	5	8	12	15	6	4

Solution

Marks	**No. of Students (f)**	**Mid value (m)**	**fm**
20-30	5	25	125
30-40	8	35	280
40-50	12	45	540
50-60	15	55	825

60-70	6	65	390
70-80	4	75	300
Total	**N=50**	**--**	**∑fm=2460**

AM = A + ∑fm/N

= 2460 / 50

= 49.20

Sum 9: From the following data, compute arithmetic mean by short cut method

Marks Obtained	0-10	10-20	20-30	30-40	40-50	50-60
No. of Students	5	10	25	30	20	10

Solution

Marks Obtained	**No.of Students (f)**	**Mid value (m)**	**d=m-A A=25**	**fd**
00-10	5	5	-20	-100
10-20	10	15	-10	-100
20-30	25	25	0	0
30-40	30	35	10	300
40-50	20	45	20	400
50-60	10	55	30	300
Total	**N=100**	**--**	**--**	**∑fd=800**

AM = A + ∑fd/N

= 25 + (800/100)

= 25 + 8

= 33

Continuous Series –Inclusive Class Intervals

Sum 10: The annual profits of 90 companies are given below. Find the arithmetic mean.

Annual Profit (Rs.)	0-19	20-39	40-59	60-79	80-99
No. of Companies	5	17	32	24	12

Solution

Annual Profit (Rs.)	**No. of Companies (f)**	**Mid value (m)**	**fm**
00-19	5	9.5	47.50
20-39	17	29.5	501.50
40-59	32	49.5	1584.00
60-79	24	69.5	1668.00
80-99	12	89.5	1074.00
Total	**N=90**	**--**	**∑fm=4875.00**

AM = ∑fm/N

= 4875 / 90

= Rs. 54.17

Continuous Series –Inclusive Class Intervals – Step Deviation Method

Sum 11: The annual profits of 90 companies are given below. Find the arithmetic mean.

Annual Profit (Rs.)	0-19	20-39	40-59	60-79	80-99
No. of Companies	5	17	32	24	12

Solution

Annual Profit	**No. of Companies (f)**	**Mid value (m)**	**d=m-A/C A=49.5; C=20**	**fd**
00-19	5	9.5	-2	-10
20-39	17	29.5	-1	-17
40-59	32	49.5	0	0
60-79	24	69.5	1	24
80-99	12	89.5	2	24
Total	**N=90**	**--**	**--**	**∑fd=21**

AM = A + (C∑fd/N)

= 49.5 + (20 X 21)/90

= 49.5 + 4.67

= Rs. 54.17

Continuous Series Less than Cumulative Frequencies

Sum 12: Calculate the mean height

Height Below (Cms)	150	155	160	165	170	175	180	185
No. of Soldiers	0	23	77	152	266	419	472	500

Solution

Height Below (Cms)	**No. of Soldiers**	**Height (cms)**	**No. of Soldiers (f)**	**Mid value (m)**	**d=m-A/C A=167.5 C=5**	**fd**
150	0	150-155	23	152.50	-3	-69
155	23	155-160	54	157.50	-2	-108
160	77	160-165	75	162.50	-1	-75
165	152	165-170	114	167.50	0	0
170	266	170-175	153	172.50	1	153
175	419	175-180	53	177.50	2	106
180	472	180-185	28	182.50	3	84
185	500					
Total	**-**	**-**	**N=500**	**-**	**-**	**∑fd=91**

Mean Height = A + (C∑fd/N)

= 167.50 + (5 X 91)/500

= 167.50 + 0.91

= 168.41 Cms.

Continuous Series – More than Cumulative Frequencies

Sum 13: Calculate the arithmetic mean from the following data:

Weight (Above)	20	25	30	35	40
No. of Boys	160	145	100	50	9

Solution

Weight Above (Kgs.)	**No. of Boys**	**Weight (Kgs.)**	**No. of Boys (f)**	**Mid value (m)**	**fm**
20	160	20-25	15	22.50	337.50
25	145	25-30	45	27.50	1237.50
30	100	30-35	50	32.50	1625.00
35	50	35-40	41	37.50	1537.50
40	9	40-	9	42.50	382.50
Total	**-**	**-**	**N=60**	**-**	**∑fm=5120**

AM = ∑fm/N

= 5120/160

= 32.00 Kgs.

MEDIAN

Median is the value of the middle most items when all the items are in the order of magnitude.

Arithmetic mean is calculated on the basis of magnitudes or values of all the items. But median is concerned with the position or place of the items in a series. 'Which is the middle most item' is the question.

Median divides the series into two equal parts. Half of the items will be equal to or less than the median; half of the items will be equal or more than the median.

Sum1: Find median for the following

6	9	21	5	7	-2	0	32	9

Solution

Values in Ascending Order

-2, 0, 5, 6, 7, 9, 9, 21, 32

Position of Median is = N+1/2 = 9+1/2 = 5

Median = 7 (It is the value at 5th Position)

Sum2: Find Median for the following data

57	58	61	42	38	65	72	66

Solution

Values in ascending order: 38, 42, 57, 58, 61, 65, 66, 72

Position of Median is = N+1/2

= 8+1/2

= 4.5, a fraction

Value at (N/2 = 8/2) = 4th Position = 58

Value at (N/2+1 = 4+1) = 5th Position = 61

Median = Value at 4th Position + Value at 5th Position / 2

= 58 + 61 / 2

= 59.5

Discrete Series

Sum 3: Consider the following data and compute Median.

Value (X)	0	1	2	3	Total
Frequency (f)	1	2	5	3	11

Solution

Value (x)	Frequency (f)	Cumulative Frequency (cf)
0	1	1
1	2	3
2	5	8
3	3	11
Total	**11**	

The position of median is N+1/2 = 11+1/2 = 6. 6 lies between the cumulative frequencies 3 and 8. Hence, the value at 6th position is the value corresponding to the cumulative frequency 8 in the table. It is 2. Hence, Median = 2.

When N+1/2 is a fraction, the two middle most items are to be identified in a similar manner and the mean of those two items is to be found.

Sum 4: The marks (out of a maximum of 10) scored by the students of a class are given below. Find the Median mark.

Mark	3	4	5	6	7	8	9	10	Total
No. of Students	1	5	6	7	10	15	10	5	59

Solution

Mark (x)	No. of Students (f)	Cumulative Frequency (cf)
3	1	1
4	5	6
5	6	12
6	7	19
7	10	29
8	15	44
9	10	54
10	5	59
Total	**59**	

N+1/2 = 59+1/2=30. When all the 59 items are in ascending order, which is in 30th position? It is included in cf=44. Median =8

Sum 5: Find the median from the following data.

Wages (Rs.)	50	75	100	150	250	Total
No. of Labourers	8	14	10	5	3	40

Solution

Wage Rs. (X)	No. of Labourers (f)	Cumulative Frequency (cf)
50	8	8
75	14	22
100	10	32
150	5	37
250	3	40

Total	**40**	

Median = N+1/2

= 40+1/2

= 20.5

Looking corresponding to cf=22, Wage (x) at 20th position= 75.

Wage (x) at 21st position 75.

Median = Wage at 20th Position + Wage at 21st Position / 2

= 75+75/2

= Rs. 75

Sum 6: Find the Median

No. of Cars sold in a Day	10	15	17	18	21	Total
No. of Days	4	16	12	5	3	40

Solution:

No. of Cars sold in a Day (x)	No. of Days (f)	Cumulative Frequency (cf)
10	4	4
15	16	20
17	12	32
18	5	37
21	3	40
Total	**40**	

N+1/2 = 40+1/2

= 20.5

X at 20^{th} position = 15

X at 21^{st} position = 17

Median = X at 20^{th} Position + X at 21^{st} Position / 2

= 15+17/2

= 16

Median – Continuous Series

Sum 7: Calculate the median height

Height (cms.)	145-150	150-155	155-160	160-165	165-170	170-175
No. of Students	2	5	10	8	4	1

Solution

Height (cms.)	**No. of Students (f)**	**Cumulative Frequency (cf)**
145-150	2	2
150-155	5	7
155-160	10	17
160-165	8	25
165-170	4	29
170-175	1	30

Total	**N=30**	

L = 155; f = 10; i = 160-155=5; cf = 7

M = L + (i (N_2-cf)/f)

= 155 + (5 (15-7))/10

= 155 + (5X8)/10

= 155 + 4

= 159 Cms.

Sum 8: Calculate the median from the following data:

Marks	10-25	25-40	40-55	55-70	70-85	85-100
Frequency	6	20	44	26	3	1

Solution

Marks	Frequency (f)	Cumulative Frequency (cf)
10-25	6	6
25-40	20	26
40-55	44	70
55-70	26	96
70-85	3	99
85-100	1	100

N/2 = 100/2 = 50

Median class interval: 40-55

L = 40; f = 44; cf = 26; i = 55 – 40 = 15

M = L + (i (N_2-cf)/f)

= 40 + (15 (50-26))/44

= 40 + (15 X 24) / 44

= 40 + 8.18

= 48.18 (or) 48

Sum 9: Calculate the median for the following:

Value	0-9	10-19	20-29	30-39	40-49	50-59	60-69
Frequency	328	720	640	598	524	378	244

Solution

Value	**Frequency (f)**	**True Class Intervals**	**Cumulative Frequency (cf)**
00-09	328	-0.5-9.5	328
10-19	720	9.5-19.5	1048
20-29	664	19.5-29.5	1712
30-39	598	29.5-39.5	2310
40-49	524	39.5-49.5	2834
50-59	378	49.5-59.5	3212
60-69	244	59.5-69.5	3456
Total	**N=3456**		

Second order limit – First upper limit = 10 – 9

= 1

Half of the difference = ½ = 0.5

0.5 has been added to each upper limit to and 0.5 has been subtracted from each lower limit to get the boundaries of the true class intervals. It is the required form for the calculation of median.

N/2 = 3456/2 = 1728. Hence, the median class interval is 29.5 – 39.5

L = 29.5; f = 598; i = 39.5-29.5=10 ; cf = 1712

M = L + (i (N_2-cf)/f)

= 29.5 + (10 (1728-1712))/598

= 29.5 + (10 X 16) / 598

= 29.5 + 0.27

= 29.77

Sum 10: Convert the following "Less than cumulative frequency" distribution into an ordinary "Frequency distribution" and then calculate the median age.

Age (Less than)	10	20	30	40	50	60	70	80
No. of Persons	4	16	40	76	96	112	120	125

Solution

Age (Less than)	**Number of Persons**	**Age**	**No. of Persons (f)**	**Cumulative Frequency (cf)**
10	4	00-10	4	
20	16	10-20	12	4

30	40	20-30	24	16
40	76	30-40	36	40
50	96	40-50	20	76
60	112	50-60	16	96
70	120	60-70	8	112
80	125	70-80	8	120
Total	.	-	**N=125**	125

N/2 = 125/2 = 62.5; Median class interval: 30-40

L = 30; f = 36; i = 40-30=10; cf=40

M = L + (i (N_2-cf)/f)

= 30 + (10 (62.5-40))/36

= 30 + 6.25

= 36.25

Sum 11: Calculate the Median

Annual Expenditure on Ad.	0	4	8	12	16
No. of Years	50	35	25	15	6

Solution

Annual Expenditure on Ad. (more than Rs. Lakhs)	**No. of Years**	**Annual Expenditure on Ad. (Rs. Lakhs)**	**No. of Years (f)**	**Cf**

0	50	0-4	15	15
4	35	4-8	10	25
8	25	8-12	10	35
12	15	12-16	9	44
16	6	16-	6	50
Total			**N=50**	**-**

N/2 = 50/2 = 25; Median class: 4-8

L = 4; f = 10; i = 8-4=4; cf = 15

M = L + (i (N_2-cf)/f)

= 4 + (4 (25-15))/10

= 4 + (4X10)/10

= 4+4

= Rs. 8 Lakhs

Sum 12: Compute median from the following data:

Mid values	115	125	135	145	155	165	175	185	195
Frequency	6	25	48	72	116	60	38	22	3

Solution

Mid values (m)	**Frequency (f)**	**Class Interval**	**Cumulative Frequency (cf)**
115	6	110-120	6
125	25	120-130	31
135	48	130-140	79
145	72	140-150	151

155	116	150-160	267
165	60	160-170	327
175	38	170-180	365
185	22	180-190	387
195	3	190-200	390
Total	**N=390**		

Difference between successive mid values = 10

Half of the difference = 5

N/2 = 390/2 = 195

Median class: 150-160

L = 150; f = 116 ; i = 160-150=10; cf = 151

M = L + (i (N_2-cf)/f)

= 150 + (10 (195-151))/116

= 150 + (10X44)/116

= 150 + 3.79

= 153.79

MODE

Mode is the value which has the greatest frequency density.

Z or M_o denotes Mode.

According to Croxton and Cowden, "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values". Mode is called the most typical or fashionable value of a distribution.

In individual observations and discrete series, the mode is most often available by inspection. The value which has the greatest frequency is mode. Such values have the greatest frequency density also. The difference between the greatest frequency and the next lower frequency may be nominal in a few cases. Mode is then determined on the basis of the greatest frequency density. That is, on the basis of its frequency and the neighbouring frequencies of each value. It is found out by forming a grouping table and analysis table of frequencies.

In the words of Murray R Spiegel, "The mode of a set of numbers is that value which occurs with the greatest frequency, i.e. it is the most common value. The mode may not exist, and even if it does exist it may not be unique". There is no mode when all the observations occur equal number of times. If one value occurs distinctly more times than any other value, that value is the mode. The set which has only one mode is said to be unimodal. Sets with two modes are said to be bimodal. Sets which have more than two modes are said to be multimodal. In a few situations due to fluctuations of sampling it becomes a difficult task to identify a single value with greatest frequency in a sample even though the population is undoubtedly unimodal.

Sum 1: Determine the mode on the following:

1. 320, 395, 342, 444, 551, 395, 425, 417, 395, 401, 390, 400
2. 3, 6, 7, 5, 8, 4, 9
3. 25, 32, 24, 27, 32, 27, 25, 32, 24, 27, 25, 24
4. 0, 2, 5, 6, 9, 5, 6, 14, 6, 15, 5, 6, 5

Solution

Mode = 395	Because its frequency, 3, is higher than others. The frequency of others is 1 each. (This is an example for unimodal distribution)
No Mode	Because all the values have equal frequency (1). (This is an example for a distribution which has no mode)
No Mode	Because all the values occur equal number of times (3 times each). (This is an example for a distribution which has no mode)
Modes 5& 6	Because they occur equal number of times and they occur greater number of times than other values. (This is an example for bimodal distribution. Grouping table of frequencies also shows that 5 and 6 have greatest and equal frequency densities. Mode is said to be ill defined in this case. Hence, the answers are not given)

Discrete Series

Sum 2: Determine the Mode:

Size of Dress	18	20	22	24
No. of Sets Produced	55	120	108	45

Solution:

Mode = 20

Sum 3: Find the Modal Size

Size of Shoes	3	4	5	6	7	8	9
No. of Pairs Sold	10	25	32	38	61	47	34

Solution

Mode = 7

Sum 4: Calculate the mode from the following:

Size	10	11	12	13	14	15	16	17	18
Frequency	10	12	15	19	20	8	4	3	2

Solution

Greatest frequency is 20. Mode need not be 14 because the difference between the greatest frequency 20 and the next lower frequency 19 is very small. Further, 19 has the support of the neighbouring frequency 15 while 20 has the support of 8 only.

Grouping table and analysis table are formed as explained earlier.

Grouping Table

Size (x)	**Frequency (f) (1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
10	10					
		22				
11	12			37		
			27			
12	15				**46**	
		34				
13	19					**54**
			39			
14	**20**			**47**		
		28				
15	8				32	
			12			
16	4					15
		7				
17	3			9		
18	2					

Analysis Table

Size X	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
10							-

11					1		1
12		1			1	1	3
13		1	1	1	1	1	5
14	1		1	1		1	4
15				1			1
16							-
17							-
18							-

Mode = 13

Continuous Series

Sum 5: Calculate the Mode.

Daily Wage in Rs.	50-60	60-70	70-80	80-90	90-100
No. of Labourers	40	62	75	100	65

Solution

L=80	The lower boundary of the modal class interval
f_1=100	The frequency of the modal class
f_0=75	The frequency of the class preceding the modal class and
f_2=65	The frequency of the class succeeding the modal class and
i=90-80=10	The size of the modal class interval

$$Z = L + (iD_1/ (D_1+D_2))$$

= 80 + (10 X 25/ (25 + 35))

= 80 + (250/60)

= 80 + 4.17

= 84.17

Sum 6: Find out the mode for the following data using grouping and analysis table.

CI	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency	9	12	15	16	17	15	10	13

Solution

Grouping Table

Size (x)	**Frequency (f) (1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
0-5	9					
		21				
5-10	12			36		
			27			
10-15	15				**43**	
		31				
15-20	16					**48**
			33			
20-25	**17**			**48**		

Size (x)	**Frequency (f) (1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
		32				
25-30	15				42	
			25			
30-35	10					38
		23				
35-40	13					

Analysis Table

Size X	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
00-05							-
05-10					1		1
10-15					1	1	2
15-20			1	1	1	1	4
20-25	1	1	1	1		1	5
25-30		1		1			2
30-35							-
35-40							-

Modal Class Interval = 20 – 25

L = 20; i = 25-20=5; D_1 = 17-16=1; D_2 = 17-15=12

Z = $L + (iD_1/ (D_1+D_2))$

= 20 + (5X1/(1+2))

= 20 + (5/3)

= 20 + 1.67

= 21.67

Sum 7: Calculate the mode

Interval	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16
Frequency	45	50	65	70	30	25	20	18

Solution

Grouping Table

Size (x)	**Frequency (f) (1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
0-2	45					
		95				
2-4	50			**160**		
			115			
4-6	65				**185**	
		135				
6-8	**70**					**165**
			100			
8-10	30			125		
		55				
10-12	25				75	
			45			

Size (x)	Frequency (f) (1)	(2)	(3)	(4)	(5)	(6)
12-14	20					63
		38				
14-16	18					

Analysis Table

Size X	1	2	3	4	5	6	Total
0-2				1			1
2-4			1	1	1		3
4-6		1	1	1	1	1	5
6-8	1	1			1	1	4
8-10						1	1
10-12							-
12-14							-
14-16							-

The interval 4-6 does not have greatest frequency, But it has greatest frequency density consider.

L=4	The lower boundary of the modal class interval
f_1=50	The frequency of the class preceding the modal class
f_2=70	The frequency of the class succeeding the modal class
i=6-4=2	The size of the modal class interval

Z = $L + (if_2/ (f_1+f_2))$

= 4 + (2X70/(50+70))

= 4 + (140/120)

= 4 + 1.17

= 5.17

Sum 8: Calculate the mode

Marks	0-19	20-39	40-59	60-79	80-99
No. of Students	5	20	35	20	12

Solution

Marks	**No. of Students**	**Marks**
0-19	5	-0.5-19.5
20-39	20	19.5-39.5
40-59	35	39.5-59.5
60-79	20	59.5-79.5
80-99	12	79.5-99.5

Greatest Frequency = 35 ; Modal Class = 39.5 – 59.5

L = 39.5; i=59.5-39.5=20; D_1= 35-20=15; D_2= 35-20=15

Z = L + (iD_1/ (D_1+D_2))

= 39.5 + (20X15/(15+15))

= 39.5 + (300/30)

= 39.5 + 10

= 49.5

Sum 9: Calculate the mode.

Annual Profit (Rs.)	10-19	20-29	30-39	40-49	50-59	60-69
No. of Companies	12	15	16	17	10	9

Solution

Grouping Table

Size (x)	**Frequency (f) (1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
9.5-19.5	12					
		27				
19.5-29.5	15			**43**		
			31			
29.5-39.5	16				**48**	
		33				
39.5-49.5	**17**					**43**
			27			
49.5-59.5	10			36		
		19				
59.5-69.5	9					

Analysis Table

Size X	1	2	3	4	5	6	Total
9.5-19.5				1			1
19.5-29.5			1	1	1		3
29.5-39.5		1	1	1	1	1	5
39.5-49.5	1	1			1	1	4
49.5-59.5						1	1
59.5-69.5							-

29.5-39.5 has the greatest frequency density. Its frequency is not the greatest. Hence, consider

L=29.5	The lower boundary of the modal class interval
f_1=15	The frequency of the class preceding the modal class
f_2=17	The frequency of the class succeeding the modal class
i=39.5-29.5-=10	The size of the modal class

Z = L + (if_2/ (f_1+f_2))

= 29.5 + (10X17/(15+17))

= 29.5 + (170/32)

= 29.5 + 5.31

= Rs. 34.81

Sum 10: Calculate the Mode

No. of Days Absent	No. of Students
Less than 3	10

Less than 6	25
Less than 9	38
Less than 12	48
Less than 15	51
Less than 18	52

Solution

Data are written in continuous class intervals first. Corresponding frequencies are then found.

No. of Days Absent	**No. of Students**	**Number of Days Absent**	**No. of Students**
Less than 3	10	0-3	10
Less than 6	25	3-6	15
Less than 9	38	6-9	13
Less than 12	48	9-12	10
Less than 15	51	12-15	3
Less than 18	52	15-18	1

Greatest frequency = 15; Modal class interval = 3-6

L=3; i=6-3=3; D_1=15-10=5; D_2=15-13=2

Z = L + (iD_1/ (D_1+D_2))

= 3 + (3X5/(5+2))

= 3 + (15/7)

= 3 + 2.14

= 5.14

Sum 11: Calculate the mode of the following frequency distribution

Wages Above (Rs.)	0	20	40	60	80	100
No. of Workers	50	45	34	16	6	0

Solution

Data are written in continuous class intervals first. Corresponding frequencies are then found;

Wages Above (Rs.)	**No. of Workers**	**Wages (Rs.)**	**No. of Workers**
0	50	00-20	5
20	45	20-40	11
40	34	40-60	18
60	16	60-80	10
80	6	80-100	6
100	0		

Greatest Frequency = 18; Modal Class Interval = 40-60

L=40; i=60-40=20; D_1=18-11=7; D_2=18-10=8

Z = L + (iD_1/ (D_1+D_2))

= 40 + (20 X 7/(7+8))

= 40 + (140/15)

= 40 + 9.33

= 49.33

Sum 12: Calculate the mode

Central Value	70	90	110	130	150
Frequency	43	78	83	125	87

Solution

Data are written in continuous class interval form.

Central Value	**Frequency**	**Class Intervals**
70	43	60-80
90	78	80-100
110	83	100-120
130	125	120-140
150	87	140-160

Greatest Frequency = 125

Modal Class Interval : 120-140

L=120; i=140-120=20; D_1=125-83=42; D_2=125-87=38

Z = L + (iD_1/ (D_1+D_2))

= 120 + (20X42/(42+38))

= 120 + (840/80)

= 120 + 10.50

= 130.50

MEASURES OF DISPERSION

In a series, all the items are not equal. There is difference or variation among the values. The degree of variation is evaluated by various measures of dispersion.

Averages are central values. They enable comparison of two or more sets of data. They are not sufficient to depict the true nature of the sets. For example, consider the following marks of two students

Student I	Student II
68	85
75	90
65	80
67	25
70	65

Both have got a total of 345 and an average of 69 each. The fact is that the second student has failed in one paper. When the averages alone are considered, the two student are equal.

Less variation is a desirable characteristic. First student has less variation. That is, he is almost equally good in all the subjects. To quote Simpson and Kafka, “An average does not tell the full story. It is hardly fully representative of a mass, unless we know the manner in which the individual items scatter around it. A further description of the series is necessary if we are to guage how representative the average is”.

STANDARD DEVIATION

Standard deviation is the root mean square deviation of the values from their arithmetic mean.

SD is the abbreviation and σ (read, sigma) is the symbol. Mean square deviation of the values from their AM is Variance and is denoted by σ^2. SD is the positive square root of variance. Karl Pearson introduced the concept of standard deviation in 1893. SD is also called root mean square deviation. It is a mathematical deficiency of mean deviation to ignore negative sign. Standard deviation possesses most of the desirable properties of a good measure of dispersion. The corresponding relative measure is Coefficient of Variation. It is very popular and so extremely used as raise a doubt whether there is any other relative measure of dispersion.

Coefficient of Variation = Standard Deviation / Arithmetic Mean X 100

Sum 1: Find SD for the following: 77, 73, 75, 70, 72, 76, 75, 72, 74, 76

Solution

X	**$x = X - \overline{X}$**	**x^2**
77	3	9
73	-1	1
75	1	1
70	-4	16
72	-2	4
76	2	4
75	1	1
72	-2	4
74	0	0
76	2	4
$\sum X=740$	$\sum (X-\overline{X})=0$	$\sum x^2=44$

Arithmetic Mean, $\overline{X}$ $= \sum X/N$

$= 740/10$

$= 74$

Standard Deviation, σ $= \sqrt{\sum x^2/N}$

$= \sqrt{44/10}$

$= \sqrt{4.4}$

$= 2.10$

Sum 2: 10 Students of B.Com. class of a College have obtained the following marks in Statistics out of 100 marks. Calculate the standard deviation.

S.No.	1	2	3	4	5	6	7	8	9	10
Marks	5	10	20	25	40	42	45	48	70	80

Solution

S.No.	**Marks X**	**X^2**
1	5	25
2	10	100
3	20	400
4	25	625
5	40	1600
6	42	1764
7	45	2025
8	48	2304
9	70	4900
10	80	6400
Total	**$\sum X=385$**	**$\sum X^2=20143$**

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$$

$$= \sqrt{20143/10 - (385/10)^2}$$

$= \sqrt{20143 - (38.5)^2}$

$= \sqrt{20143 - 148225}$

$= \sqrt{532.05}$

$= 23.07$

Deviations taken from Assumed Mean

Sum 3: For the data below, calculate standard deviation 40, 50, 60, 70, 80, 90, 100

Solution

X	X=X-A; A=70	d^2
40	-30	900
50	-20	400
60	-10	100
70	0	0
80	10	100
90	20	400
100	30	900
Total	**∑d=0**	**$\sum d^2$=2800**

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$$

$= \sqrt{2800/7 - (0/7)^2}$

$= \sqrt{400 - 0^2}$

$= \sqrt{400}$

$= 20$

Step Deviation Method

Sum 4: Given below are the marks obtained by 5 B.Com. Students

Roll No.	101	102	103	104	105
Marks	10	30	20	25	15

Calculate Standard Deviation

Solution

Roll No.	**Marks (X)**	**d=X-A/C: A=20; C=5**	**d^2**
101	10	-2	4
102	30	2	4
103	20	0	0
104	25	1	1
105	15	-1	1
Total		**$\sum d=0$**	**$\sum d^2=10$**

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$$

$$= \sqrt{10/5 - (0/5)^2} \times 5$$

$$= \sqrt{2 - 0^2} \times 5$$

$$= \sqrt{2} \times 5$$

$$= 1.4142 \times 5$$

$$= 7.07$$

Discrete Series – Deviation taken from Actual Mean

Sum 5: Calculate the Standard Deviation of the following series.

X	6	9	12	15	18
f	7	12	13	10	8

Solution

X	**f**	**fX**	**$X=X-\bar{X}$**	**X^2**	**fX^2**
6	7	42	-6	36	252
9	12	108	-3	9	108
12	13	156	0	0	0
15	10	150	3	9	90
18	8	144	6	36	288
Total	**N=50**	**∑fx=600**	**-**	**-**	**$\sum fx^2$=738**

Arithmetic Mean = ∑fx/N

= 600 / 50

= 12.00

Standard Deviation, $\sigma = \sqrt{\sum fX^2/N}$

$= \sqrt{738/50}$

$= \sqrt{14.76}$

$= 3.84$

Discrete Series – Direct Method

Sum 6: Calculate the Standard Deviation

No. of Goals Scored in a Match	0	1	2	3	4	5
No. of Matches	1	2	4	3	0	2

Solution:

X	**f**	**fX**	**fx^2**
0	1	0	0
1	2	2	2
2	4	8	16
3	3	9	27
4	0	0	0
5	2	10	50
Total	**N=12**	**$\sum fx=29$**	**$\sum fx^2=95$**

$$\sigma = \sqrt{\frac{\Sigma fx'^2}{N} - \left(\frac{\Sigma fx'}{N}\right)^2} \times C$$

$$= \sqrt{95/12 - (29/12)^2}$$

$$= \sqrt{7.9167 - (2.4167)^2}$$

$$= \sqrt{7.9167 - 5.8404}$$

$$= \sqrt{2.0763}$$

$$= 1.44$$

Discrete Series – Deviations taken from Assumed Mean

Sum 7: Calculate Standard Deviation from the following data

X	6	9	12	15	18
f	7	12	19	10	2

X	f	d=X-A; A=12	fd	fd^2
6	7	-6	-42	252
9	12	-3	-36	108

12	19	0	0	0
15	10	3	30	90
18	2	6	12	72
Total	**N=50**	**-**	**$\sum fd=-36$**	**$\sum fd^2=522$**

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$$

$$= \sqrt{522/50 - (-36/50)^2}$$

$$= \sqrt{10.44 - (0.72)^2}$$

$$= \sqrt{10.4400 - 0.5184}$$

$$= \sqrt{9.9216}$$

$$= 3.15$$

Discrete Series – Deviations taken from Assumed Mean

Sum 8: The weekly salaries of a group of employees are given in the following table. Find the mean and standard deviation of the salaries.

Salary (Rs.)	75	80	85	90	95	100
No. of Persons	3	7	18	12	6	4

Solution:

Salary (Rs.)	**No. of Persons (f)**	**d=X-A/C A=85; C=5**	**fd**	**fd^2**
75	3	-2	-6	12
80	7	-1	-7	7

85	18	0	0	0
90	12	1	12	12
95	6	2	12	24
100	4	3	12	36
Total	**N=50**	**-**	**∑fd=23**	**$\sum fd^2$=91**

Arithmetic Mean = A + (C∑fd/N)

= 85 + (5X23/50)

= 85 + 2.3

= Rs.87.30

Continuous Series – Deviation taken from Actual Mean

Sum 9: Find the Standard Deviation.

Class Interval	0-10	10-20	20-30	30-40	40-50	Total
Frequency	2	5	9	3	1	20

Solution

Class Interval	**Frequency (f)**	**Mid Value (m)**	**fm**	**m-X X=23**	**$(m-X)^2$**	**$f(m-X)^2$**
0-10	2	5	10	-18	324	648
10-20	5	15	75	-8	64	320
20-30	9	25	225	2	4	36
30-40	3	35	10	12	144	432
40-50	1	45	45	22	484	484

Total	N=20		$\sum fm=460$	-	-	**$\sum f(m-\bar{X})^2=1920$**

Arithmetic Mean $= \sum fm/N$

$= 460/20$

$= 23$

SD σ $= \sqrt{\sum f(m-\bar{X})^2/N}$

$= \sqrt{1920/20}$

$= \sqrt{96}$

$= 9.80$

Continuous Series – Direct Method

Sum 10: The following data were obtained while observing the life span of a few neon lights of a company. Calculate SD.

Life Span (Yrs.)	4-6	6-8	8-10	10-12	12-14	Total
No. of Neon Lights	10	17	32	21	20	100

Solution

Life Span (Yrs.)	**No. of Neon Lights (f)**	**Mid Value (m)**	**fm**	**fm^2**
4-6	10	5	50	250

6-8	17	7	119	833
8-10	32	9	288	2592
10-12	21	11	231	2541
12-14	20	13	260	3380
Total	N=100		$\sum fm=948$	$\sum fm^2=9596$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$$

$$= \sqrt{9596/100 - (948/100)^2}$$

$$= \sqrt{95.96 - (9.48)^2}$$

$$= \sqrt{95.9600 - 89.8704}$$

$$= \sqrt{6.0896}$$

$$= 2.47$$

Continuous Series – Deviation taken from Assumed Mean

Sum 11: Calculate the standard deviation of the following series.

No. of Students in 00 (Below)	2	6	10	14	18	22	26
No. of Colleges	0	7	19	42	61	72	80

Solution

No. of Students in 00	No. of Colleges	No. of Students	No. of Colleges	Mid value	d=m-A	fd	fd^2

(Below)		**in 00**	**f**	**(m)**	**A=12**		
2	0	2-6	7	4	-8	-56	448
6	7	6-10	12	8	-4	-48	192
10	19	10-14	23	12	0	0	0
14	42	14-18	19	16	4	76	304
18	61	18-22	11	20	8	88	704
22	72	22-26	8	24	12	96	1152
26	80						
Total	**-**	**-**	**N=80**	**-**	**-**	**$\sum fd=156$**	**$\sum fd^2=2800$**

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$$

$$= \sqrt{2800/80 - (156/80)^2}$$

$$= \sqrt{35 - (1.95)^2}$$

$$= \sqrt{35.0000-3.8025}$$

$$= \sqrt{31.1975}$$

$$= 5.59$$

Continuous Series – Step Deviation Method

Sum 12: Calcuate the standard deviation of the following frequency distribution.

Annual Profit (Rs.)	20-40	40-60	60-80	80-100	100-120	120-140	140-160
No. of Banks	10	14	25	48	33	24	16

Solution

Annual Profit (Rs.)	No. of Banks	Mid value (m)	d=m-A A=90; C=20	fd	fd^2
20-40	10	30	-3	-30	90
40-60	14	50	-2	-28	56
60-80	25	70	-1	-25	25
80-100	48	90	0	0	0
100-120	33	110	1	33	33
120-140	24	130	2	48	96
140-160	16	150	3	48	144
Total	**N=170**	**-**	**-**	**$\sum$fd=46**	**$\sum fd^2$=444**

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

$$= \sqrt{444/170 - (46/170)^2} \times 20$$

$$= \sqrt{2.6118 - (0.2706)^2} \times 20$$

$$= \sqrt{2.6118 - 0.0732} \times 20$$

$$= \sqrt{2.5386} \times 20$$

$$= 1.5933 \times 20$$

$$= \text{Rs. } 31.87$$

CORRELATION

The term correlation refers to the relationship between the variables. Simple correlation refers to the relationship between two variables. There may be fluctuation or co-variation between the values of the variables. The direction of change and the closeness of the relationship are found.

DEFINITION

Correlation Analysis attempts to determine the degree of relationship between variables- Ya-Kun-Chou.

. Correlation is an analysis of the covariation between two or more variables.- A.M.Tuttle.

Types of Correlation

1) Positive and Negative Correlation

It depends upon the direction of change of the variables. If the two variables tend to move together in the same direction (ie) an increase in the value of one variable is accompanied by an increase in the value of the other, (or) a decrease in the value of one variable is accompanied by a decrease in the value of other, then the correlation is called positive or direct correlation. Price and supply, height and weight, yield and rainfall, are some examples of positive correlation.

If the two variables tend to move together in opposite directions so that increase (or) decrease in the value of one variable is accompanied by a decrease or increase in the value of the other variable, then the correlation is called negative (or) inverse correlation. Price and demand, yield of crop and price, are examples of negative correlation.

2) Linear and Non-linear Correlation

If the ratio of change between the two variables is a constant then there will be linear correlation between them.

Consider the following

X	2	4	6	8	10	12
Y	3	6	9	12	15	18

Here the ratio of change between the two variables is the same. If we plot these points on a graph we get a straight line.

If the amount of change in one variable does not bear a constant ratio of the amount of change in the other. Then the relation is called Curvi-linear (or) non-linear correlation. The graph will be a curve.

3) Simple and Multiple Correlation

When we study only two variables, the relationship is simple correlation. For example, quantity of money and price level, demand and price. But in a multiple correlation we study more than two variables simultaneously. The relationship of price, demand and supply of a commodity are an example for multiple correlation.

4) Partial and Total Correlation

The study of two variables excluding some other variable is called Partial correlation. For example, we study price and demand eliminating supply side. In total correlation all facts are taken into account.

KARL PEARSON'S COEFFICIENT OF CORRELATION

This is also called product moment correlation coefficient. This is denoted by r. This is covariance between the two variables divided by the product of their standard deviations. This can be calculated by using any one of the formulae. Choice of formula depends on the nature of the data.

Sum 1: The following table gives aptitude test scores and productivity indices of 8 randomly selected workers:

Aptitude Scores	57	58	59	59	60	61	62	64
Productivity Index	67	68	65	68	72	72	69	71

Calculate the correlation coefficient between Aptitude score and productivity index.

Solution

X	**Y**	**X=X-X X=60**	**Y=Y-Y Y=69**	**XY**	**X^2**	**Y^2**
57	67	-3	-2	6	9	4
58	68	-2	-1	2	4	1
59	65	-1	-4	4	1	16
59	68	-1	-1	1	1	1
60	72	0	3	0	0	9
61	72	1	3	3	1	9
62	69	2	0	0	4	0
64	71	4	2	8	16	4
$\sum x=480$	$\sum y=552$	$\sum x=0$	$\sum y=0$	$\sum xy=24$	$\sum x^2=36$	$\sum y^2=44$

Mean of X = ∑X/N = 480/8 = 60

Mean of Y = ∑Y /N = 552/8 = 69

$$r = \frac{\Sigma XY}{\sqrt{\sum X^2 . \sum Y^2}}$$

$= 24 / \sqrt{36 X 44}$

= 0.6030

Sum 2: Compute the coefficient of Correlation between X –Advertisement Expenditure and Y-Sales

X	10	12	18	8	13	20	22	15	5	17
Y	88	90	94	86	87	92	96	94	88	85

Solution

X	Y	XY	X^2	Y^2
10	88	880	100	7744
12	90	1080	144	8100
18	94	1692	324	8836
8	86	688	64	7396
13	87	1131	169	7569
20	92	1840	400	8464
22	96	2112	484	9216
15	94	1410	225	8836
5	88	440	25	7744
17	85	1445	289	7225
∑X=140	**∑Y=900**	**∑XY=12718**	**$\sum x^2$=2224**	**$\sum y^2$=81130**

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

$= \quad 10 \text{ X } 12718 - 140 \text{ X } 900 \sqrt{10 \text{ X } 2224 - (140)^2} \sqrt{10 \text{ X } 81130 - (900)^2}$

$= \quad 1180 \sqrt{2640 \text{ X } 1300}$

$= \quad 0.6370$

Sum 3: Calculate the coefficient of correlation between Expenditure on Advertisement in Rs. ‘000 (X) and Sales in Rs. Lakhs (Y) after allowing a time lag of two months.

	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**
X	40	45	47	50	53	60	57	51	48	45
Y	75	69	65	64	70	71	75	83	90	92

Solution: As a time lag of two months is to be allowed, the following pairs of values are available

X	**Y**	**XY**	**X^2**	**Y^2**
40	65	2600	1600	4225
45	64	2880	2025	4096
47	70	3290	2209	4900
50	71	3550	2500	5041
53	75	3975	2809	5625
60	83	4980	3600	6889
57	90	5130	3249	8100
51	92	4692	2601	8464

$\sum X$=403	**$\sum Y$=610**	**$\sum XY$=31097**	**$\sum x^2$=20593**	**$\sum y^2$=47340**

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$= \quad 8 \text{ X } 31097 - 403 \text{ X } 610 / \sqrt{8 \text{ X } 20593 - (403)^2} \sqrt{8 \text{ X } 47340 - (610)^2}$

$= \quad 2946 / \sqrt{2335 \text{ X } 6620}$

$= \quad 0.7493$

Sum 4: From the following data, compute the coefficient of correlation between X and Y.

	X	**Y**
Sum of squares of deviations from the arithmetic mean	8250	724
Sum of products of deviations of X an Y from respective means	2350	
No. of pairs of observations	10	

Solution

$$r = \frac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}}$$

$= 2350 / \sqrt{8250} \ \sqrt{724}$

= 0.9615

Sum 5: Compute the coefficient of Correlation between X –Advertisement Expenditure and Y-Sales

X	10	12	18	8	13	20	22	15	5	17
Y	88	90	94	86	87	92	96	94	88	85

Solution

X	**Y**	**u=X-a/c A=15;C=1**	**v=Y-b b=90;d=1**	**uv**	**U^2**	**V^2**
10	88	-5	-2	10	25	4
12	90	-3	0	0	9	0
18	94	3	4	12	9	16
8	86	-7	-4	28	49	16
13	87	-2	-3	6	4	9
20	92	5	2	10	25	4
22	96	7	6	42	49	36
15	94	0	4	0	0	16
5	88	-10	-2	20	100	4
17	85	2	-5	-10	4	25
Total	**-**	**$\sum u=-10$**	**$\sum v=0$**	**$\sum uv=118$**	**$\sum u^2=274$**	**$\sum v^2=130$**

$$r = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{\left[n\Sigma u^2 - (\Sigma u)^2\right].\left[(n\Sigma v^2) - (\Sigma v)^2\right]}}$$

$\sqrt{\quad}$

$= 10 \text{ X } 118 – (-10)\ (0) \ / \ 10 \text{ X } 274 – (-10)^2 \sqrt{10 \text{ X } 130 – (0)^2}$

$= 1180 \ / \sqrt{2640} \ \sqrt{1300}$

$= 0.6370$

Sum 6: From the following data find the percentage of variation in Y that is explained by the variation in X; N=11; $\sum X=117$; $\sum Y=260$; $\sum X^2=1313$; $\sum Y^2=6580$; $\sum XY=2827$

Solution

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2]\,[n\sum y^2 - (\sum y)^2]}}$$

$= \quad 11 \text{ x } 2827 – (117)\ (260) \ / \sqrt{11 \text{ X } 1313 – (117)^2} \sqrt{11 \text{ X } 6580 – (260)^2}$

$= \quad 677 \ / \ \sqrt{754} \ \sqrt{4780}$

$= \quad 0.3566$

Sum 7: Find Karl Pearson's Coefficient of Correlation from the marks secured by 10 students in Accountancy and Statistics

Accountancy	45	70	65	30	90	40	50	75	85	60
Statistics	35	90	70	40	95	40	60	80	80	50

Solution

X	**Y**	**X=X-X X=61**	**Y=Y-Y Y=64**	**XY**	**X^2**	**Y^2**
45	35	-16	-29	464	256	841
70	90	9	26	234	81	676
65	70	4	6	24	16	36
30	40	-31	-24	744	961	576
90	95	29	31	899	841	961
40	40	-21	-24	504	441	576
50	60	-11	-4	44	121	16
75	80	14	16	224	196	256
85	80	24	16	384	576	256
60	50	-1	-14	14	1	196
∑x=610	**∑y=640**	**∑x=0**	**∑y=0**	**∑xy=3535**	**$\sum x^2$=3490**	**$\sum y^2$=4390**

Mean of X = ∑X/N = 610/10 =61

Mean of Y = ∑Y/N = 640/10=64

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 . \Sigma Y^2}}$$

$= 3535 / \sqrt{3490}\sqrt{4390}$

$= 0.9031$

Sum 8: Calculate Karl Pearson's Coefficient of Correlation from the following data

Roll Nos.	**1**	**2**	**3**	**4**	**5**
Marks in Accountancy	48	35	17	23	47
Marks in Statistics	45	20	40	25	45

Solution

X	**Y**	**X=X-X X=34**	**Y=Y-Y Y=35**	**XY**	**X^2**	**Y^2**
48	45	14	10	140	196	100
35	20	1	-15	-15	1	225
17	40	-17	5	-85	289	25
23	25	-11	-10	110	121	100
47	45	13	10	130	169	100
∑x=170	**∑y=175**	**∑x=0**	**∑y=0**	**∑xy=280**	**$\sum x^2$=776**	**$\sum y^2$=550**

Mean of X = ∑X/N = 170/5 = 34

Mean of Y = ∑Y/N = 175/5 = 35

$$r = \frac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}}$$

$= 280 \sqrt{776 \text{ X } 550}$

= 280 / 653.299

= 0.429

Sum 9: The following table gives indices of industrial production of registered unemployed (in hundred thousand). Calculate the value of the coefficient so obtained.

Year	1991	1992	1993	1994	1995	1996	1997	1998
Index of Production	100	102	104	107	105	112	103	99
Number Unemployed	15	12	13	11	12	12	19	26

Solution

X	**Y**	**X=X-X X=34**	**Y=Y-Y Y=35**	**XY**	**X^2**	**Y^2**
100	15	-4	0	0	16	0
102	12	-2	-3	6	4	9
104	13	0	-2	0	0	4
107	11	3	-4	-12	9	16
105	12	1	-3	-3	1	9
112	12	8	-3	-24	64	9
103	19	-1	4	-4	1	16
99	26	-5	11	-55	25	121
$\sum x=832$	**$\sum y=120$**	**$\sum x=0$**	**$\sum y=0$**	**$\sum xy=92$**	**$\sum x^2=120$**	**$\sum y^2=184$**

Mean of X = $\sum X/N$ = 832/8 = 104

Mean of Y = $\sum Y/N$ = 120/8 = 15

$$r = \frac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}}$$

$= -91 \sqrt{120 \text{ X } 184}$

= 0.619

SPEARMAN'S RANK CORRELATION

The Karl Pearson's method is based on the assumption that the population being studied is normally distributed. When it is known that the population is not normal or when the shape of the distribution is not known, there is need for a measure of correlation that involves no assumption about the parameter of the population.

It is possible to avoid making any assumption about the populations being studied by ranking the observations according to size and basing the calculations on the ranks rather than upon the original observations. It does not matter which way the items are ranked, item number one may be the largest or it may be the smallest. Using ranks rather than actual observations gives the coefficient of the rank correlation.

This method of finding out co-variability or the lack of it between two variables was developed by the British Psychologist Charles Edward Spearman in 1904. This measure is especially useful when quantitative measures for certain factors (such as in the evaluation of leadership ability or the judgement of female beauty) cannot be fixed, but the individual in the group can be arranged in order thereby obtaining for each individual a number indicating his (her) rank in the group.

$$\rho = 1 - \left(\frac{6\sum d^2}{N(N^2-1)} \right)$$

Where there is no tie. d-difference between X and Y ranks.

$$\rho = 1 - \left(\frac{6\left(\sum d^2 + (m(m^2-1)/12)\right)}{N(N^2-1)} \right)$$

When one value occurs m times

$$\rho = 1 - \left[\frac{6\,(\sum d^2 + (m(m^2-1)/12) + ((m(m^2-1)/12)}{N\,(N^2-1)} \right]$$

When more than one value is repeated

It is calculated when ranks are given or when rank correlation coefficient is required. Rank correlation coefficient also lies between -1 and +1.

Sum 1: Rankings of 10 trainees at the beginning (x) and at the end (y) of a certain course are given below:

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
X	1	6	3	9	5	2	7	10	8	4
Y	6	8	3	7	2	1	5	9	4	10

Calculate Spearman's rank correlation coefficient.

Solution

X	**Y**	**d**	**D^2**
1	6	-5	25
6	8	-2	4
3	3	0	0
9	7	2	4
5	2	3	9
2	1	1	1
7	5	2	4
10	9	1	1

X	Y	d	D^2
8	4	4	16
4	10	-6	36
Total	**-**	**$\sum d=0$**	**$\sum d^2=100$**

$$\rho = 1-\left[\frac{6\sum d^2}{N(N^2-1)}\right]$$

$$\rho = 1-\left[\frac{6X100}{10X99}\right]$$

$= 1 – 0.6061$

$= 0.3939$

Sum 2: From the data given below, calculate the rank correlation coefficient.

X	21	36	42	37	25
Y	47	40	37	42	43

Solution

X	Y	X	Y	d	D^2
21	47	5	1	4	16
36	40	3	4	-1	1
42	37	1	5	-4	16
37	42	2	3	-1	1

25	43	4	2	2	4
Total	**-**	**-**		**$\sum d=0$**	**$\sum d^2=38$**

$$\rho = 1 - \left[\frac{6\sum d^2}{N(N^2-1)}\right]$$

$$\rho = 1 - \left[\frac{6 X 38}{5(5^2-1)}\right]$$

$$\rho = 1 - \left[\frac{6 X 38}{5 X 24}\right]$$

$= 1 – 1.9$

$= -0.9$

Sum 3: Find the rank correlation coefficient for the percentage of marks secured by a group of 8 students in Economics and Statistics.

Marks in Economics	50	60	65	70	75	40	70	80
Marks in Statistics	80	71	60	75	90	82	70	50

Solution

X	Y	X	Y	d	D^2
50	80	7	3	4	16

60	71	6	5	1	1
65	60	5	7	-2	4
70	75	3.5	4	-0.5	0.25
75	90	2	1	1	1
40	82	8	2	6	36
70	70	3.5	6	-2.5	6.25
80	50	1	8	-7	49
Total	**-**	**-**	**-**	**$\sum d=0$**	**$\sum d^2=113.50$**

$$\rho = 1-\left[\frac{6\ (\sum d^2 + (m(m^2-1)/12)}{N\ (N^2-1)}\right]$$

$$\rho = 1-\left[\frac{6\ (113.5 + 0.5)}{8\ (8^2-1)}\right]$$

$$\rho = 1-\left[\frac{6 \text{ X } 114}{8 \text{ X } 63}\right]$$

= 1-1.3571

= -0.3571

Sum 4: Marks obtained by 8 students in Accountancy and Statistics are given below. Compute rank correlation.

X	15	20	28	12	40	60	20	80
Y	40	30	50	30	20	10	30	60

Solution:

$$\rho = 1 - \left[\frac{6\,(\sum d^2 + (m(m^2-1)/12) + ((m(m^2-1)/12)}{N\,(N^2-1)}\right]$$

$$\rho = 1 - \left[\frac{6\,(81.5+0.5+2)}{8\,(8^2-1)}\right]$$

$$\rho = 1 - \left[\frac{6 \text{ X } 84}{8 \text{ X } 63}\right]$$

$$= 0$$

REGRESSION

After having established the fact that two variables are closely related we may be interested in estimating (predicting) the value of one variable given the value of another. For example, if we know that advertising and sales are correlated we find out expected amount of sales for a given advertising expenditure or the required amount of expenditure for attaining a given amount of sales. Similarly, if we know that the yield of rice and rainfall are closely related we may find out the amount of rain required to achieve a certain production figure.

Regression analysis reveals average relationship between two variables and this makes possible estimation or prediction.

The dictionary meaning of the term 'regression' is the act of returning or going back. The term 'regression' was first used by Sir Francis Galton (1822-1911) in 1877 while studying the relationship between the height of fathers and sons. This term was introduced by him in the paper 'Regression towards fathers and sons revealed a very interesting relationship, i.e., tall fathers tend to have tall sons and short fathers short sons, but the average height of the sons of a group of tall fathers is less than that of the fathers and the average fathers. The line describing the tendency to regress or going back was called by Galton s 'Regression Line'. The term is still used to describe that line drawn for a group of points to represent the trend present, but it no longer necessarily carries the original implication of "stepping back" that Galton intended. These days there is a growing tendency to the modern writers to use the term estimating line instead of regression line because the expression estimating line is more clarificatory in character.

DEFINITIONS

Regression is the measure of the average relationship between two or more variables in terms of the original units of the data.

The term 'regression analysis' refers to the methods by which estimates are made of the values of a variable from a knowledge of the values of one or more other variables and to the measurement of the errors involved in this estimation process – Morris Hamburg.

One of the most frequently used techniques in economics and business research to find a relation between two or more variables that are related causally, is regression analysis – Taro Yamane

Regression analysis attempts to establish the 'nature of the relationship' between variables – that is, to study the functional relationship between the variables and thereby provide a mechanism for prediction, or forecasting" – Ya Lun Chou

It is clear from the above definitions that regression analysis is a statistical device with the help of which we are in a position to estimate (or predict) the unknown values of one variable from known values of another variable. The variable which is used to predict the variable of interest is called the independent variable or exploratory variable and the variable we are trying to predict is called the dependent variable or "explained variable. The independent variable is denoted by X and the dependent variable by Y. The analysis used is called the simple regression analysis – simple because there is only one predictor or independent variable, and linear because of the assumed linear relationship between the dependent and the independent variables. The term "linear" means that an equation of a straight line of the form Y=a+bx, where a and b are constants, is used to describe the average relationship that exists between the two variables.

It should be noted that the term 'dependent' and 'independent' refer to the mathematical or functional meaning of dependence – they do not imply that there is necessarily any cause and effect relationship between the variables. What is meant is simply that estimates of values of the dependent variable Y may be obtained for given values of the independent variable X from a mathematical function involving X and Y. In that sense, the values of Y are dependent upon the values of X. The X variable may or may not be causing change in the Y variable. For example, while estimating sales of a product from figures on advertising expenditures, sale is generally taken as the dependent variable. However, there may or may not be causal connection between these two factors in the sense that changes in advertising expenditures cause change in sales. In fact, in certain cases, the cause-effect relation may be just opposite of what appears to be the obvious one.

Uses of Regression Analysis

Regression analysis is a branch of statistical theory that is widely used in almost all the scientific disciplines. In economics it is the basic technique for measuring or estimating the relationship among economic variables that constitute the essence of economic theory and economic life. For example, if we know that two variables, Price (X) and Demand (Y), are closely related we can find out the most probable value of X for a given value of Y or the most probable value of Y for a given value of X. Similarly, if we closely related, we can find out the expected price for a certain price for a certain amount of tax levy. Thus, we find that the study of regression is of considerable help to the economists and businessmen. The uses of regression are not confined to economics and business field only. Its applications are extended to almost all the natural, physical and social sciences. The regression analysis attempts to accomplish the following:

1. Regression analysis provides estimates of values of the dependent variable from values of the independent variable. The device used to accomplish this estimation procedure is the regression line. The regression line describes the average relationship existing between X and Y variables. i.e., it displays mean values of X for given values of Y. The equation of

this line, known as the regression equation, provides estimates of the dependent variable when values of the independent variable are inserted into the equation.

2. A second goal of regression analysis is to obtain a measure of the error involved in using the regression line as a basis for estimation. For this purpose the standard error of estimate is calculated. This is a measure of the scatter or spread of the observed values of Y around the corresponding values estimated from the regression line. If the line fits the data closely, that is, if there is little scatter of the observations around the regression line, good estimates can be made of the Y variable. On the other hand, if there is a great deal of scatter or the observations around the fitted regression line, the line will not produce accurate estimates of the dependent variable

3. With the help of regression coefficients we can calculate the correlation coefficient. The square of correlation coefficient (r) called coefficient of determination, measures the degree of association of correlation that exists between the two variables. It assesses the proportion of variance in the dependent variable that has been accounted for by the regression equation. In general, the greater the value of r^2 the better is the fit and the more useful the regression equations as a predictive device.

DIFFERENCE BETWEEN CORRELATION AND REGRESSION

Correlation	Regression
Correlation is the relationship between variables. It is expressed numerically	Regression means going back. The average relation between the variables is given as an equation
Between two variables, none is identified as independent variable	One of the variables is independent variable and the other is dependent variable in any particular context
Correlation does not mean causation. One variable need not be the cause and the other, effect	Independent variable may be the 'the cause' and dependent variable, 'the effect'
There is spurious or nonsense correlation	There is no such possibility. Regression is considered only when the variables are related
Correlation coefficient is independent of change of origin and scale	Regression coefficients are independent of change or origin but are affected by change of scale

Correlation	Regression
Correlation coefficient is a number -1 and +1	The two regression coefficients have the same sign, + or -. One of them can be greater than 1 numerically. But they can not be greater than 1 numerically simultaneously.
Correlation coefficient is not in any unit of measurement	Each regression coefficient is in the unit of measurement of the dependent variable
Correlation coefficient indicates the direction of co-variation and the closeness of the linear relation between two variables	Regression equations give the value of the dependent variable corresponding to any value of the independent variable
The significance of the sample correlation coefficient can be tested. The limits between which the population correlation coefficient is expected to lie can be found	Target can be reached. The value of the independent variable can be chosen so as to get the target value of the dependent variable. For example, a specific amount can be spent on advertisement to get the targeted revenue.

Sum 1: From the following data, obtain the two regression equations:

X	6	2	10	4	8
Y	9	11	5	8	7

Solution

X	Y	XY	X^2	Y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49

$\sum X=30$	$\sum Y=40$	$\sum XY=214$	$\sum X^2=220$	$\sum Y^2=340$

Let the regression equation of Y on X be Y=A+BX

The normal equations are $\sum Y = NA + B\sum X$

$\sum XY = A\sum X + B\sum X^2$

By substituting the value from the table,

5A + 30B = 40 Say (1)

30A +220B = 214 Say (2)

(1) X 6 30A + 180B = 240 Say (3)

(2) – (3) 40 B = -26

B = -26 / 40

= -0.6500

From (1), 5A-30X0.6500 = 40

A = 40 + 19.5 / 5

= 11.90

The regression equation of Y on X is

Y = 11.90 – 0.6500X

Let the regression equation of X on Y be X = A + BY

The normal equations are $\sum X = NA + B\sum Y$

$\sum XY = A\sum Y + B\sum Y^2$

By substituting the values from the table,

5A + 40B = 30 Say (4)

40A + 340 B = 214 Say (5)

(4) X 8 40A = 320 B = 240 Say (6)

(5) – (6) 20B = - 26

B = -26/20

= -1.300

From (4), 5A+40 X (-1.30) = 30

A = 30 + 52 /5

= 16.40

The regression equation of X on Y is X = 16.40 – 1.300

Sum 2: You are given the following data:

	X	Y
Arithmetic Mean	36	85
Standard Deviation	11	8
Correlation coefficient between X and Y	0.66	

(a) Find the two regression equations
(b) Estimate the value of X when Y=75

Solution

bxy = rσx/σy = 0.66 x 11 / 8 = 0.9075

byx = rσy/σx = 0.66 X 8 / 11 = 0.4800

a) Regression equation of Y on X

$Y - \overline{Y} = byx\,(X - \overline{X})$

Y – 85 = 0.4800 (X-36)

= 0.4800 X – 17.28

Y = 67.72 + 0.4800 X

Regression equation of X on Y

$X - \overline{X} = bxy\,(Y - \overline{Y})$

X – 36 = 0.9075 (Y-85)

= 0.9075 – 77.14

X = 0.9075 Y – 41.14

b) When Y=75, X = 0.9075 X 75 – 41.14

= 26.92

Sum 3: From the following information on values of two variables X and Y find the two regression lines and the correlation coefficient.

N=10; $\sum X=20$; $\sum Y=40$; $\sum X^2=240$; $\sum Y^2=410$; $\sum XY=200$

Solution

$\bar{X} = \sum X/N = 20/10 = 2.00$

$\bar{Y} = \sum Y/N = 40/10 = 4.00$

$$bxy = \frac{N\sum XY-(\sum X)(\sum Y)}{N\sum Y^2 - (\sum Y)^2} \text{ as } \sum X\neq 0 \text{ and } \sum Y\neq 0$$

$$= \frac{10 \times 200 - 20 \times 40}{10 \times 410 - (40)^2}$$

$$= \frac{2000 - 800}{4100-1600}$$

$$= \frac{1200}{2500} = 0.4800$$

$$byx = \frac{N\sum XY-(\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} \text{ as } \sum X \neq 0 \text{ and } \sum Y \neq 0$$

$$= \frac{1200}{10X240-(20)^2}$$

$$= \frac{1200}{2000} = 0.6000$$

Regression equation of Y on X

$Y - \overline{Y} = byx\,(X-\overline{X})$

$Y-4 = 0.6000\,(X-2)$

$= 0.6000\,X - 1.20$

$Y = 2.80 + 0.6000X$

Regression equation of X on Y

$X - \overline{X} = bxy\,(Y-\overline{Y})$

$X-2 = 0.4800\,(Y-4)$

$= 0.4800\,Y - 1.92$

$X = 0.08 + 0.4800Y$

Sum 4: Calculate the two regression equations from the following data:

X	10	12	13	12	16	15
Y	40	38	43	45	37	43

Also estimate Y when X=20.

Solution

X	Y	XY	X^2	Y^2
10	40	400	100	1600
12	38	458	144	1444
13	43	559	169	1849
12	45	540	144	2025
16	37	592	256	1369
15	43	645	225	1849
∑X=78	**∑Y=246**	**∑XY=3192**	**$\sum X^2$=1038**	**$\sum Y^2$=10136**

$\bar{X}$ = ∑X/N = 78/6 = 13.00

Y = ∑Y/N = 246/6 = 41.00

$$bxy = \frac{N\sum XY-(\sum X)(\sum Y)}{\text{--------------------------}}$$

$N\sum Y^2 - (\sum Y)^2$

$$= \frac{6 \times 3192 - 78 \times 246}{6 \times 10316 - (246)^2}$$

$$= \frac{19152 - 19188}{60816 - 60516}$$

$$= \frac{-36}{300} = 0.1200$$

$$byx = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2}$$

$$= \frac{-36}{6 \times 1038 - (78)^2}$$

$$= \frac{-36}{6228 - 6084} = 0.6000$$

$$= \frac{-36}{144} = 0.2500$$

Regression equation of Y on X

$Y - \bar{Y}$ = byx $(X-\bar{X})$

Y-41 = -0.2500 (X-13)

= -0.2500 X + 3.25

Y = 44.25 + 0.25X

When X = 20, Y=44.25-0.25 X 20 = 39.25

Regression equation of X on Y

$X - \bar{X}$ = bxy $(Y-\bar{Y})$

X-13 = -0.1200 (Y-41)

= -0.1200 Y + 4.92

X = 17.92 - 0.12Y

Sum 5: From the data given below, find two regression equations

Marks in Mathematics	25	28	35	32	31	36	29	38	34	32
Marks in Statistics	43	46	49	41	36	32	31	30	33	39

Solution

X	**Y**	**$x = X-\bar{X}$** **$\bar{X}=32$**	**$y = Y-\bar{Y}$** **$\bar{Y}=38$**	**XY**	**X^2**	**Y^2**

25	43	-7	5	-35	49	25
28	46	-4	8	-32	16	64
35	49	3	11	33	9	121
32	41	0	3	0	0	9
31	36	-1	-2	2	1	4
36	32	4	-6	-24	16	36
29	31	-3	-7	21	9	49
38	30	6	-8	-48	36	64
34	33	2	-5	-10	4	25
32	39	0	1	0	0	1
∑X=320	**∑Y=380**	**∑X=0**	**∑Y=0**	**∑XY=-93**	**$\sum X^2$=140**	**$\sum Y^2$=398**

Solution

$\bar{X} = \sum X/N$ = 320/10 = 32.00

$\bar{Y} = \sum Y/N$ = 380/10 = 38.00

bxy = $\sum xy/\sum y^2$ = -93/398 = -0.2337

byx = $\sum xy/\sum x^2$ = -93/140 = -0.6643

a) Regression equation of Y on X

$Y - \bar{Y}$ = byx $(X-\bar{X})$

Y-38 = -0.6643 X (X-32)

= -0.6643 X + 21.26

= 59.26 – 0.6643X

Regression equation of X on Y

$X - \overline{X} = bxy\,(Y - \overline{Y})$

$X - 32 = -0.2337\,(Y - 38)$

$= -0.2337Y + 8.88$

$= 40.88 - 0.2337Y$

POSSIBLE QUESTIONS

PART A (ONE MARKS)

(ONLINE EXAMINATION)

PART – B (8 MARKS)

1. Calculate the mean, median and mode of the data given below.

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No of students:	8	15	22	20	10	5

2. Explain the types of correlation in detail.

3. The weekly salaries of a group of employees are given in the following table. Find the mean and standard deviation of the salaries.

Salary (in Rs)	75	80	85	90	95	100
No. of Persons	3	7	18	12	6	4

4. Write short note on : i. Regression ii. Correlation

5. Find the mean median and mode for the following data and verify the empirical relations.

Class	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
Frequency	3	7	13	17	12	10	8	8	6	6

6. Explain the uses of regression and how it is differ from correlation.
7. Find the coefficient correlation between x and y from the following data:

X	5	10	5	11	12	4	3	2	7	1
Y	1	6	2	8	5	1	4	6	5	2

8.Write short note on:

i. Mean ii. Median iii. Mode iv. Standard deviation

9. Find the median for the following frequency distribution.

Number of days	**Absent**	**Number of students**
Less than	5	29
Less than	10	224
Less than	15	465
Less than	20	582
Less than	25	634
Less than	30	644
Less than	35	650
Less than	40	653
Less than	45	655

10. Define regression and distinguish between Regression and Correlation

UNIT V

SYLLABUS

Test of Significance 't'- test – large sample and 'f' test, test of significance for attributes – Analysis of Variance – Business Forecasting – Exponential Smoothing – Chi – Square test

HYPOTHESIS

Hypothesis is usually considered as the principal instrument in research. Its main function is to suggest new experiments and observations. In fact, many experiments are carried out with the deliberate object of testing hypotheses. Decision-makers often face situations wherein they are interested in testing hypotheses on the basis of available information and then take decisions on the basis of such testing. In social science, where direct knowledge of population parameter(s) is rare, hypothesis testing is the often used strategy for deciding whether a sample data offer such support for a hypothesis that generalisation can be made. Thus hypothesis testing enables us to make probability statements about population parameter(s). The hypothesis may not be proved absolutely, but in practice it is accepted if it has withstood a critical testing. Before we explain how hypotheses are tested through different tests meant for the purpose, it will be appropriate to explain clearly the meaning of a hypothesis and the related concepts for better understanding of the hypothesis testing techniques.

What is Hypothesis?

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

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

CHARACTERISTICS OF HYPOTHESIS

1. Hypothesis should be clear and precise. If the hypothesis is not clear and precise, the inferences drawn on its basis cannot be taken as reliable.
2. Hypothesis should be capable of being tested. In a swamp of untestable hypotheses, many a time the research programmes have bogged down. Some prior study may be done by researcher in order to make hypothesis a testable one. A hypothesis "is testable if other deductions can be made from it which, in turn, can be confirmed or disproved by observation."
3. Hypothesis should state relationship between variables, if it happens to be a relational hypothesis.
4. Hypothesis should be limited in scope and must be specific. A researcher must remember that narrower hypotheses are generally more testable and he should develop such hypotheses.
5. Hypothesis should be stated as far as possible in most simple terms so that the same is easily understandable by all concerned. But one must remember that simplicity of hypothesis has nothing to do with its significance.

6. Hypothesis should be consistent with most known facts i.e., it must be consistent with a substantial body of established facts. In other words, it should be one which judges accept as being the most likely.
7. Hypothesis should be amenable to testing within a reasonable time. One should not use even an excellent hypothesis, if the same cannot be tested in reasonable time for one cannot spend a life-time collecting data to test it.
8. Hypothesis must explain the facts that gave rise to the need for explanation. This means that by using the hypothesis plus other known and accepted generalizations, one should be able to deduce the original problem condition. Thus hypothesis must actually explain what it claims to explain; it should have empirical reference.

FORMULATION OF HYPOTHESIS

1. hypothesis may originate in different ways. A cultural environment may give rise to it. In India, for example, religion and custom dominate the way of life. This has had it reaction on economic values and individual initiative in various walks of life. Such a situation could give rise to any number of hypothesis; sociological, cultural, political and economic
2. A second source of hypothesis is folk wisdom or current popular beliefs and practices suggesting both the problems and the hypothesis
3. Analogies are often a spring of valuable hypothesis. Students of sociology in the course of their studies would have come across analogies wherein a society is compared to a biological organism, the natural law to the social law, thermo-dynamics to social dynamics etc.,
4. The history of science provides an eloquent testimony to the fact that personal and idiosyncratic experiences of the scientist contributes a great deal to the type and form of questions he may ask as also to the kinds of tentative answers to these questions that he can provide
5. Hypotheses may also rest on the findings of other studies. The researcher on the basis of the findings of other studies may hypothesise that similar relationship between specified

variables will hold good in the present study, too. This is a very common way of researchers who design their study with a view to replicating another study conducted in a different concrete context

6. Another source of hypothesis formulation in cases which are expectations to accepted theory
7. Personal experience and individual reaction may give rise to hypotheses
8. A hypotheses may turn from a body of theory which, by way of logical deduction, may lead to the production that if certain conditions are present, certain results will follow. Theory is indeed an extremely fertile seed-bed of hypotheses.

PROCEDURE FOR HYPOTHESIS TESTING

To test a hypothesis means to tell (on the basis of the data the researcher has collected) whether or not the hypothesis seems to be valid. In hypothesis testing the main question is: whether to accept the null hypothesis or not to accept the null hypothesis? Procedure for hypothesis testing refers to all those steps that we undertake for making a choice between the two actions i.e., rejection and acceptance of a null hypothesis. The various steps involved in hypothesis testing are stated below:

1) Making a Formal Statement

The step consists in making a formal statement of the null hypothesis (H0) and also of the alternative Hypothesis (Ha). This means that hypothesis should be clearly stated, considering the nature of the research problem. For instance, Mr. Mohan of the Civil Engineering Department wants to test the load bearing capacity of an old bridge which must be more than 10 tons, in that case he can state his hypotheses as under:

Null Hypothesis H_0 : μ =10 tons

Alternative Hypothesis H_a: μ > 10 tons

Take another example. The average score in an aptitude test administered at the national level is 80. To evaluate a state's education system, the average score of 100 of the state's students selected on random basis was 75. The state wants to know if there is a significant difference between the local scores and the national scores. In such a situation the hypotheses may be stated as under:

Null Hypothesis $H_0 : \mu = 80$

Alternative Hypothesis $H_a : \mu \neq 80$

The formulation of hypotheses is an important step which must be accomplished with due care in accordance with the object and nature of the problem under consideration. It also indicates whether we should use a one-tailed test or a two-tailed test. If Ha is of the type greater than (or of the type lesser than), we use a one-tailed test, but when Ha is of the type "whether greater or smaller" then we use a two-tailed test.

2) Selecting a Significance Level

The hypotheses are tested on a pre-determined level of significance and as such the same should be specified. Generally, in practice, either 5% level or 1% level is adopted for the purpose. The factors that affect the level of significance are: (a) the magnitude of the difference between sample means; (b) the size of the samples; (c) the variability of measurements within samples; and (d) whether the hypothesis is directional or non-directional (A directional hypothesis is one which predicts the direction of the difference between, say, means). In brief, the level of significance must be adequate in the context of the purpose and nature of enquiry.

3) Deciding the Distribution to Use

After deciding the level of significance, the next step in hypothesis testing is to determine the appropriate sampling distribution. The choice generally remains between normal distribution and the t-distribution. The rules for selecting the correct distribution are similar to those which we have stated earlier in the context of estimation.

4) Selecting a Random Sample and Computing an Appropriate Value

Another step is to select a random sample(s) and compute an appropriate value from the sample data concerning the test statistic utilizing the relevant distribution. In other words, draw a sample to furnish empirical data.

5) Calculation of the Probability

One has then to calculate the probability that the sample result would diverge as widely as it has from expectations, if the null hypothesis were in fact true.

6) Comparing the Probability

Yet another step consists in comparing the probability thus calculated with the specified value for α, the significance level. If the calculated probability is equal to or smaller than the α value in case of one-tailed test (and α /2 in case of two-tailed test), then reject the null hypothesis (i.e., accept the alternative hypothesis), but if the calculated probability is greater, then accept the null hypothesis. In case we reject H_0, we run a risk of (at most the level of significance) committing an error of Type I, but if we accept H_0, then we run some risk (the size of which cannot be specified as long as the H_0 happens to be vague rather than specific) of committing an error of Type II.

STATISTICS

‘t’ Test

When the size of sample is small (less than 30). In particular, it will no longer be possible for us to assume (a) that the random sampling distribution of a statistic is approximately normal and (b) that values given by the sample data are sufficiently close to the population values and can be used in their place for the calculation of the standard error of the estimate.

The removal of these assumptions makes it necessary to use entirely new techniques to deal with the problems of small samples. The division between the theories of large and small samples is, therefore, a very real one, though it is not always easy to draw a precise line of demarcation. It should be noted that as a rule, the methods and the theory of small samples are applicable to large samples, though the reverse is not true.

While dealing with small samples our main interest is not to estimate the population values as in true in large samples; rather our interest lies in testing a given hypothesis, i.e., in ascertaining whether observed values could have arisen by sampling fluctuations from some value given in advance. For example, if a sample of 15 gives a correlation coefficient of +0.4, we shall be interested not so much in the value of the correlation in the parent population, but more generally whether this value could have been arisen from an uncorrelated population. i.e. whether it is significant of correlation in the parent population.

It should be noted that the investigator who works with very small samples must know that his estimates will vary widely from sample to sample. Moreover, he must be satisfied with relatively wide confidence intervals. Precision of statement is less, of course, the wider the intervals employed. Each inference drawn from large sample results in far more precise in the limits it sets up than is an inference based on a much smaller sample.

Student's 't' Distribution

Theoretical work on t-distribution was done by W.S. Gosset (1876-1937) in the early 1900. Gosset was employed by the Guinness and Son, a Dulbin bravery, Ireland, which did not permit employees to publish research findings under their own names. So Gosset adopted the pen name "Student" and published his findings under their name. Thereafter, the t-distribution is commonly called Student's t-distribution or simply student's distribution.

The t-distribution is used when sample size is 30 or less and the population standard deviation is unknown.

The 't' statistic is defined as

$t= X-\mu/S \times \sqrt{n}$

Where $S=\sqrt{\sum(X-X)^2/ n-1}$

Sum 1: The manufacturer of a certain make of electric bulbs claims that his bulbs have a mean life of 25 months with a standard deviation of 5 months. A random sample of 6 such bulbs gave the following values

Life of Months	24	26	30	20	20	18

Can you regard the producer's claim to be valid at 1% level of significance? (Given that the table values of the appropriate test statistics at the said level are 4.032, 3.707 and 3.499 for 5, 6 and 7 degrees of freedom respectively)

Solution

Let us taken the hypothesis that there is no significant difference in the mean life of bulbs in the sample and that of the population, Applying t-test:

X	**$(X-\bar{X})$ X**	**X^2**
24	+1	1
26	+3	9
30	+7	49
20	-3	9
20	-3	9
18	-5	25
$\sum X=138$		$\sum X^2=102$

$t = \bar{X}-\mu / SD \times \sqrt{n}$

Average of X= $\sum X/n$ = 138/6 = 23

SD $= \sqrt{\sum X^2/n-1} = \sqrt{102/5} = \sqrt{20.4} = 4.517$

$$= \frac{(23-25)}{4.517}\sqrt{6} = \frac{2 \times 2.449}{4.517} = 1.084$$

ν=n-1 = 6-1 = 5

For ν=5 $t_{0.01}$=4.032

The calculated value of t is less than the table value. The hypothesis is accepted. Hence, the producer's claim is not valid at 1% level of significance.

Sum 2: A random sample of size 16 has 53 as mean. The sum of the squares of the deviation taken from mean is 135. Can this sample be regarded as taken from the population having 56 as mean? Obtain 95% and 99% confidence limits of the mean of the population. (For v=15, $t_{0.05}$=2.13 for v=15, $t_{0.01}$=2.95)

Solution

Let us take the hypothesis that there is no significant difference between the sample mean and hypothetical mean, Applying t-test.

$t = X-\mu / SD \times \sqrt{n}$

SD $= \sqrt{\sum X^2/n-1} = \sqrt{135/15} = 3$

$= (53-56) / 3 \quad X \sqrt{16}$

$= 3 \text{ X } 4/ 3$

$= 4$

v=16-1=15. For v=16, $t_{0.05}$=2.13.

The calculated value is more than the table value. The hypothesis is rejected. Hence, the sample has not come from a population having 56 as mean.

Sum: 3 The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data:

Item	1	2	3	4	5	6	7	8	9	10
Life in ‘000 hours	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accept the hypothesis that the average life time of bulbs is 4000 hours.

Solution

Let us take the hypothesis that there is no significant difference in the sample mean and the hypothetical population mean

X	**$(X-\bar{X})$**	**$(X-\bar{X})^2$**
4.2	-0.2	0.04
4.6	+0.2	0.04
3.9	-0.5	0.25
4.1	-0.3	0.09
5.2	+0.8	0.64
3.8	-0.6	0.36
3.9	-0.5	0.25
4.3	-0.1	0.01
4.4	0.0	0.00
5.6	+1.2	1.44
$\sum X=44$		$\sum(X-X)^2=3.12$

$t= \bar{X}-\mu/S \times \sqrt{n}$

Where $S= \sqrt{\sum(X-X)^2/ n-1}$

$$t = \frac{4.4-4}{0.589}\sqrt{10} = \frac{0.4 \times 3.162}{0.589} = 0.2148$$

$\nu=n-1 = 10-1 = 9$

For ν=9 $t_{0.05}$=2.262

The calculated value of t is less than the table value. The hypothesis is accepted. The average life time of the bulbs could be 4000 hours.

Sum 4: Two types of drugs were used on 5 and 7 patients for reducing their weight.

Drug A was imported and drug B indigenous. The decrease in the weight after using the drugs for six months was as follows:

Drug A	10	12	13	11	14		
Drug B	8	9	12	14	15	10	9

Is there a significant difference in the efficacy of the two drugs? If not, which drug should you buy (For ν=10, $t_{0.05}$=2.228)

Solution

Let us take the hypothesis that there is no significant difference in the efficacy of the two drugs.

	—				

X_1	**$(X_1 - X_1)$**	**$(X_1 - \bar{X}_1)^2$**	**X_2**	**$(X_2 - \bar{X}_2)$**	**$(X_2 - \bar{X}_2)^2$**
10	-2	4	8	-3	9
12	0	0	9	-2	4
13	+1	1	12	+1	1
11	-1	1	14	+3	9
14	+2	4	15	+4	16
			10	-1	1
			9	-2	4
$\sum X_1=60$		**$\sum (X_1 - X_1)^2=10$**	$\sum X_2=77$		**$\sum (X_2 - X_2)^2=44$**

$$t = \frac{\bar{X}_1 + \bar{X}_2}{S} \times \sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

$$S = \sqrt{\frac{\sum(X_1-\bar{X}_1)^2 + \sum(X_2-\bar{X}_2)^2}{n_1+n_2-2}}$$

$\sqrt{10 + 44}$ $\sqrt{54}$

S= ---------- = ---- = 2.324

5+7-2 10

12-11 5 X 7 1.708

= -------- ------ = ------- = 0.735

2.324 5+7 2.324

$\nu = n_1 + n_2 - 2 = 5 + 7 - 2 = 10$

$\nu = 10$, $t_{0.05} = 2.228$

The calculated value of t is less than the table value, the hypothesis is accepted. Hence, there is no significance in the efficacy of two drugs. Since drug B is indigenous and there is no difference in the efficacy of imported and indigenous drug, we should buy indigenous drug, i.e., B.

Sum 5 : For a random sample of 10 persons, fed on diet A, the increased weight in pounds in a certain period were:

10	6	16	17	13	12	8	14	15	9

For another random sample of 12 persons, fed on diet B, the increase in the same period were:

7	13	22	15	12	14	18	8	21	23	10	17

Test whether the diets A and diet B differ significantly as regards their effect on increase in weight, Given the following.

Degrees of Freedom	19	20	21	22	23
Value at t at 5% level	2.09	2.09	2.08	2.07	2.07

Solution: Let us take the null hypothesis that A and B do not differ significantly weight regard to their effect on increase in weight.

X_1	$(X_1 - \bar{X}_1)$	$(X_1 - \bar{X}_1)^2$	X_2	$(X_2 - \bar{X}_2)$	$(X_2 - \bar{X}_2)^2$
10	-2	4	7	-8	64
6	-6	36	13	-2	4
16	+4	16	22	+7	49
17	+5	25	15	0	0
13	+1	1	12	-3	9
12	0	0	14	-1	1
8	-4	16	18	+3	9
14	+2	4	8	-7	49
15	+3	9	21	+6	36
9	-3	9	23	+8	64

			10	-5	25
			17	+2	4
$\sum X_1=120$		$\sum \mathbf{(X_1 - X_1)^2=120}$	$\sum X_2=180$		$\sum \mathbf{(X_2 - X_2)^2=314}$

Mean increase in weight of 10 persons fed on diet A

Mean increase in weight of 10 persons fed on diet B

$$t = \frac{\bar{X}_1+\bar{X}_2}{\bar{S}} \;\bar{X}\; \sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

$$S = \frac{\sqrt{\sum(X_1-X_1)^2 + \sum(X_2-X_2)^2}}{n_1+n_2-2}$$

$$SD = \frac{120 + 314}{10+12-2} = \frac{434}{20} = 4.66$$

X_1=12, X_2=15, n_1=10, n_2 = 12, S=4.66. Substituting the values in the above formula:

$$t = \frac{12 - 15}{4.66} \times \frac{10 \times 12}{10 + 12} = \frac{3}{4.66} \times 2.34 = 1.51$$

For v=20, the table value of t at 5 per cent level is 2.09. The calculated value is less than the table value and hence the experiment provides no evidence against the hypothesis. We, therefore, conclude that diets A and B do not differ significantly as regards their effect on increase in weight is concerned.

Sum 6: In a test given to two groups of students, the marks obtained are as follows:

I Group	18	20	36	50	49	36	34	49	41
II Group	29	28	26	35	30	44	46		

Examine the significance of difference between the arithmetic mean of the marks secured by the students of the above two groups. (The value of t at 5% level of significance for v=14 is 2.14)

Solution

Let us take the hypothesis that there is no significant difference in the arithmetic mean of the marks secured by the students of the two groups.

Group I X_1	**$(X_1 - \bar{X}_1)$**	**$(X_1 - \bar{X}_1)^2$**	**Group II X_2**	**$(X_2 - \bar{X}_2)$**	**$(X_2 - \bar{X}_2)^2$**
18	-19	361	29	-5	25
20	-17	289	28	-6	36
36	-1	1	26	-8	64
50	+13	169	35	+1	1
49	+12	144	30	-4	16
36	-1	1	44	+10	100
34	-3	9	46	+12	144
49	+12	144			
41	+4	16			
$\sum X_1=333$		**$\sum (X_1 - X_1)^2=1134$**	$\sum X_2=238$		**$\sum (X_2 - X_2)^2=386$**

$$t = \frac{\bar{X}_1 + \bar{X}_2}{\bar{S}} \times \sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

$$S = \sqrt{\frac{\sum(X_1-X_1)^2 + \sum(X_2-X_2)^2}{n_1+n_2-2}}$$

1134 + 386

$$SD = \frac{\quad}{9 + 7 - 2} = 10.42$$

$$t = \frac{37-34}{10.42} \quad \frac{9 \times 7}{9 + 7} = \frac{3}{10.42} \times 1.984 = 0.571$$

$\nu=n_1+n_2-2 = 9 + 7 - 2 = 14$; For $\nu = 14$, $t_{0.05}= 2.14$

The calculated value of t is less than the table value and hence the hypothesis hold true. We, therefore, conclude that the mean marks of the students of the two groups do not differ significantly.

F-Test

The F-test is named in honor of the great Statistician R.A. Fisher. The object of F – test is to find out whether the two independent estimates of population variance differ significantly, or whether the two samples may be regarded as drawn from the normal populations having the same variance. For carrying out the test of significance, we calculate the ratio F. F is defined as

$F=S_1^2/S_2^2$, Where $S_1^2 = \sum(X_1-\bar{X}_1)^2/n_1-1$ and

$S_2^2 = \sum(X_2-\bar{X}_2)^2/n_2-1$

It should be noted that S_1^2 is always the larger estimate of variance, i.e., $S_1^2 > S_2^2$

$$F = \frac{\text{Larger estimate of Variance}}{\text{Smaller estimate of Variance}}$$

The calculated value of F is compared with the table value for v1 and v2 at 5% or 1% level of significance. If calculated value of F is greater than the table value then the F ratio is considered significant and the null hypothesis is rejected. On the other hand, if the calculated value of F is less than the table value the null hypothesis is accepted and it is inferred that both the samples have come from the population having same variance.

Since F Test is based on the ratio of two variances, it is also known as the Variance Ratio Test. The ratio of two variances follows a distribution called the F distribution named after the famous statistician R.A. Fisher.

Sum 1: Two random samples were drawn from two normal populations and their values are :

A	66	67	75	76	82	84	88	90	92		
B	64	66	74	78	82	85	87	92	93	95	97

Test whether the two populations have the same variance at the 5% level of significance. (F=3.36) at 5% level of significance level v1=10 and v2=8.

Solution

Let us take the hypothesis that the two populations have the same variance

A X_1	**$(X_1-\bar{X}_1)$; X1**	**X_1^2**	**B X_2**	**$(X_2-\bar{X}_2)$; X_2**	**c X_2^2**
66	-14	196	64	-19	361
67	-13	169	66	-17	289
75	-5	25	74	-9	81
76	-4	16	78	-5	25
82	2	4	82	-1	1
84	4	16	85	2	4
88	8	64	87	4	16
90	10	100	92	9	81
92	12	144	93	10	100
			95	12	144
			97	14	196
$\sum X_1$=720	**$\sum X_1$=0**	**$\sum X_1^2$=734**	**$\sum X_2$=913**	**$\sum X_2$=0**	**$\sum X_2^2$=1298**

Average of $X_1 = \sum X_1/n1$ = 720 / 9 =80;

Average of $X_2 = \sum X_2/n2$ = 913/ 11 =83

$S_1^2 \quad = \sum(X_1-X_1)^2/n_1-1$

$= 734 / 9-1 = 91.75$

$S_2^2 \quad = \sum(X_2-X_2)^2/n_2-1$

$= 1298 / 11-1 = 129.80$

$F \quad = S_1^2/S_2^2$

$= 129.8 / 191.75$

$= 1.415$

The calculated value of F is less than the table value. The hypothesis is accepted. Hence, it may be calculated that the two populations have the same variance.

Sum 2: In a sample of 8 observations, the sum of squared deviations of items from the mean was 84.4. In another sample of 10 observations, the value was found to be 102.60. Test whether the difference is significant at 5% level.

You are given that at 5% level, critical value of F for v1=7 and v2=9 degrees of freedom is 3.29 and for v1=8 and v2=10 degrees of freedom, its value is 3.07.

Solution

Let us take hypothesis that the difference in the variance of the two samples is not significant. We are given

$S_1^2 = \sum(X_1-X_1)^2/n_1-1 = 84.4/7 = 12.06$

$S_2^2 = \sum(X_2-X_2)^2/n_2-1 = 102.3/9 = 11.40$

$F = S_1^2/S_2^2 = 12.06/11.40 = 1.06$

The calculated value of F is less than the table value. Hence, we accept the hypothesis and conclude that the difference in the variance of two samples is not significant at 5% level.

Sum 3: Two samples are drawn from two normal populations. From the following data test whether the two samples have the same variance at 5% level.

Sample 1	60	65	71	74	76	82	85	87		
Sample 2	61	66	67	85	78	63	85	86	88	91

Solution: Let us take the hypothesis that the two populations have the same variance.

$F = S_1^2/S_2^2$

Sample 1 X_1	**$(X_1-\bar{X}_1)$; X1**	**X_1^2**	**Sample 2 X_2**	**$(X_2-\bar{X}_2)$; X_2**	**X_2^2**
60	-15	225	61	-16	256
65	-10	100	66	-11	121
71	-4	16	67	-10	100
74	-1	1	85	8	64

76	1	1	78	1	1
82	7	49	63	-14	196
85	10	100	85	8	64
87	12	144	86	9	81
			88	11	121
			91	14	196
$\sum X_1$=600	**$\sum X_1$=0**	**$\sum X_1^2$=636**	**$\sum X_2$=770**	**$\sum X_2$=0**	**$\sum X_2^2$=1200**

Average of X_1 = 600/8 = 75

Average of X_2 = 770/10 = 77

$S_1^2 = \sum(X_1-X_1)^2/n_1-1 = 636 / 8-1 = 90.857$

$S_2^2 = \sum(X_2-X_2)^2/n_2-1 = 1200/10-1 = 133.33$

$F = S_1^2/S_2^2$

$= 133.33/90.857$

$= 1.467$

For v1=9 and v2=7, $F_{0.05}$ = 3.68. The calculated value of F is less than the table value. The hypothesis holds good and hence we conclude that the two populations have the same variance.

Sum 4: The following data present the yields in Quintals of common ten subdivisions of equal area of two agricultural plots.

Plot 1	6.2	5.7	6.5	6.0	6.3	5.8	5.7	6.0	6.0	5.8
Plot 2	5.6	5.9	5.6	5.7	5.8	5.7	6.0	5.5	5.7	5.5

Test whether two samples taken from two random populations have the same variance. (5% point of F for v1=9 and v2=9 is 3.18)

Solution: Let us take the null hypothesis that the samples come from populations having the same variance

Plot 1 X_1	**$(X_1-\overline{X}_1)$; X1**	**X_1^2**	**Plot 2 X_2**	**$(X_2-\overline{X}_2)$; X_2**	**c X_2^2**
6.2	0.2	0.04	5.6	-0.1	0.01
5.7	-0.3	0.09	5.9	0.2	0.04
6.5	0.5	0.25	5.6	-0.1	0.01
6.0	0	0	5.7	0	0
6.3	0.3	0.09	5.8	0.1	0.01
5.8	-0.2	0.04	5.7	0	0
5.7	-0.3	0.09	6.0	0.3	0.09
6.0	0	0	5.5	-0.2	0.04
6.0	0	0	5.7	0	0
5.8	-0.2	0.04	5.5	-0.2	0.04
$\sum X_1=60$	**$\sum X_1=0$**	**$\sum X_1^2=0.64$**	**$\sum X_2=57$**	**$\sum X_2=0$**	**$\sum X_2^2=0.24$**

F $= S_1^2/S_2^2$

$S_1^2 = \sum(X_1-X_1)^2/n_1-1 = 0.64 / 9 = 0.071$

$S_2^2 = \sum(X_2-X_2)^2/n_2-1 = 0.24 / 9 = 0.027$

$F = S_1^2/S_2^2 = 0.071/0.027 = 2.63$

The value of F for 9 and 6 at 5% level of significance is 3.18. The calculated value is less than the table value. The hypothesis holds true. Hence, the samples come from population having the same variance.

CHI-SQUARE

The chi-square test is an important test amongst the several tests of significance developed by statisticians. Chi-square, symbolically written as χ^2 (Pronounced as Ki-square), is a statistical measure used in the context of sampling analysis for comparing a variance to a theoretical variance. As a non-parametric test, it "can be used to determine if categorical data shows dependency or the two classifications are independent. It can also be used to make comparisons between theoretical populations and actual data when categories are used." Thus, the chi-square test is applicable in large number of problems. The test is, in fact, a technique through the use of which it is possible for all researchers to (i) test the goodness of fit; (ii) test the significance of association between two attributes, and (iii) test the homogeneity or the significance of population variance.

Chi-square is an important non-parametric test and as such no rigid assumptions are necessary in respect of the type of population. We require only the degrees of freedom (implicitly of course the size of the sample) for using this test. As a non-parametric test, chi-square can be used (i) as a test of goodness of fit and (ii) as a test of independence.

As a test of goodness of fit, χ^2 test enables us to see how well does the assumed theoretical distribution (such as Binomial distribution, Poisson distribution or Normal distribution) fit to the observed data. When some theoretical distribution is fitted to the given data, we are always interested in knowing as to how well this distribution fits with the observed data. The chi-square test can give answer to this. If the calculated value of χ^2 is less than the table value at a certain level of significance, the fit is considered to be a good one which means that the divergence between the observed and expected frequencies is attributable to fluctuations of sampling. But if the calculated value of χ^2 is greater than its table value, the fit is not considered to be a good one.

As a test of independence, χ^2 test enables us to explain whether or not two attributes are associated. For instance, we may be interested in knowing whether a new medicine is effective in controlling fever or not, χ^2 test will helps us in deciding this issue. In such a situation, we proceed with the null hypothesis that the two attributes (viz., new medicine and control of fever) are independent which means that new medicine is not effective in controlling fever. On this basis we first calculate the expected frequencies and then work out the value of χ^2 . If the calculated value of χ^2 is less than the table value at a certain level of significance for given degrees of freedom, we conclude that null hypothesis stands which means that the two attributes are independent or not associated (i.e., the new medicine is not effective in controlling the fever). But if the calculated value of χ^2 is greater than its table value, our inference then would be that null hypothesis does not hold good which means the two attributes are associated and the association is not because of some chance factor but it exists in reality (i.e., the new medicine is effective in controlling the fever and as such may be prescribed). It may, however, be stated here that χ^2 is not a measure of the degree of relationship or the form of relationship between two attributes, but is simply a technique of judging the significance of such association or relationship between two attributes.

In order that we may apply the chi-square test either as a test of goodness of fit or as a test to judge the significance of association between attributes, it is necessary that the observed as well as theoretical or expected frequencies must be grouped in the same way and the theoretical distribution must be adjusted to give the same total frequency as we find in case of observed distribution. χ^2 is then calculated as follows:

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

If two distributions (observed and theoretical) are exactly alike, χ2 = 0; but generally due to 2 sampling errors, χ2 is not equal to zero and as such we must know the sampling distribution of χ2 so that we may find the probability of an observed χ2 being given by a random sample from the hypothetical universe. Instead of working out the probabilities, we can use ready table which gives probabilities for given values of χ2. Whether or not a calculated value of χ2 is significant can be ascertained by looking at the tabulated values of χ2 for given degrees of freedom at a certain level of significance. If the calculated value of χ2 is equal to or exceeds the table value, the difference between the observed and expected frequencies is taken as significant, but if the table value is more than the calculated value of χ2, then the difference is considered as insignificant i.e., considered to have arisen as a result of chance and as such can be ignored.

As already stated, degrees of freedom play an important part in using the chi-square distribution and the test based on it, one must correctly determine the degrees of freedom. If there are 10 frequency classes and there is one independent constraint, then there are (10 – 1) = 9 degrees of freedom. Thus, if 'n' is the number of groups and one constraint is placed by making the totals of observed and expected frequencies equal, the d.f. would be equal to (n – 1). In the case of a contingency table (i.e., a table with 2 columns and 2 rows or a table with two columns and more than two rows or a table with two rows but more than two columns or a table with more than two rows and more than two columns), the d.f. is worked out as follows: d.f. = (c – 1) (r – 1), where 'c' means the number of columns and 'r' means the number of rows.

CONDITIONS FOR THE APPLICATION OF χ2 TEST

1. Observations recorded and used are collected on a random basis.
2. All the items in the sample must be independent.
3. No group should contain very few items, say less than 10. In case where the frequencies are less than 10, regrouping is done by combining the frequencies of adjoining groups so that the new frequencies become greater than 10. Some statisticians take this number as 5, but 10 is regarded as better by most of the statisticians.
4. The overall number of items must also be reasonably large. It should normally be at least 50, howsoever small the number of groups may be.
5. The constraints must be linear. Constraints which involve linear equations in the cell frequencies of a contingency table (i.e., equations containing no squares or higher powers of the frequencies) are known are know as linear constraints.

Sum 1: A die is thrown 132 times with following results:

Number turned up	1	2	3	4	5	6
Frequency	16	20	25	14	29	28

Is the die unbiased?

Solution

Let us take the hypothesis that the die is unbiased. If that is so, the probability of obtaining any one of the six numbers is 1/6 and as such the expected frequency of any one number coming upward is 132 ×1/6 = 22. Now we can write the observed frequencies along with expected frequencies and work out the value of $\chi 2$ as follows:

No. Turned UP	Observed Frequency	Expected Frequency	(Oi – Ei)	(Oi – Ei)	(Oi – Ei) /Ei
1	16	22	–6	36	36/22
2	20	22	–2	4	4/22
3	25	22	3	9	9/22
4	14	22	–8	64	64/22
5	29	22	7	49	49/22
6	28	22	6	36	36/22

$\sum (O-E)^2/E = 9$

Hence, the calculated χ^2=9

DF = (n-1) (6-1)=5

The table value of $\chi 2$ for 5 degrees of freedom at 5 per cent level of significance is 11.071. Comparing calculated and table values of $\chi 2$, we find that calculated value is less than the table value and as such could have arisen due to fluctuations of sampling. The result, thus, supports the hypothesis and it can be concluded that the die is unbiased.

Sum 2: Find the value of □ 2 for the following information:

Class	***A***	***B***	***C***	***D***	***E***
Observed frequency	8	29	44	15	4
Theoretical (or expected) frequency	7	24	38	24	7

Solution : Since some of the frequencies less than 10, we shall first re-group the given data as follows and then will work out the value of $\chi 2$

Class	**Observed Frequency**	**Expected Frequency**	**O-E**	**$(O-E)^2/E$**
A and *B*	(8 + 29) = 37	(7 + 24) = 31	6	36/31
C	44	38	6	36/38
D and *E*	(15 + 4) = 19	(24 + 7) = 31	–12	144/31

$\sum (O-E)^2/E = 6.76$ (Approximate)

Sum 3: Genetic theory states that children having one parent of blood type A and the other of blood type B will always be of one of three types, A, AB, B and that the proportion of three types will on an average be as 1 : 2 : 1. A report states that out of 300 children having one A parent and B parent, 30 per cent were found to be types A, 45 per cent per cent type AB and remainder type B. Test the hypothesis by $\chi 2$ test.

Solution: The observed frequencies of type A, AB and B is given in the question are 90, 135 and 75 respectively.

The expected frequencies of type A, AB and B (as per the genetic theory) should have been 75, 150 and 75 respectively.

We now calculate the value of $\chi 2$ as follows:

Type	**Observed Frequency**	**Expected Frequency**	**O-E**	**(O-E)^2**	**(O-E)^2/E**
A	90	75	15	225	225/75 = 3
AB	135	150	–15	225	225/150 = 1.5
B	75	75	0	0	0/75 = 0

$\sum (O-E)^2/E = 3 + 1.5 = 4.5$

Table value of $\chi 2$ for 2 d.f. at 5 per cent level of significance is 5.991.

The calculated value of $\chi 2$ is 4.5 which is less than the table value and hence can be ascribed to have taken place because of chance. This supports the theoretical hypothesis of the genetic theory that on an average type A, AB and B stand in the proportion of 1 : 2 : 1.

Sum 4: The table given below shows the data obtained during outbreak of smallpox:

Particulars	Attacked	Not attacked	Total
Vaccinated	31	469	500
Not vaccinated	185	1315	1500
Total	**216**	**1784**	**2000**

Test the effectiveness of vaccination in preventing the attack from smallpox. Test your result with the help of $\chi 2$ at 5 per cent level of significance.

Solution: Let us take the hypothesis that vaccination is not effective in preventing the attack from smallpox i.e., vaccination and attack are independent.

Group	Observed Frequency	Expected Frequency	O-E	$(O-E)^2$	$(O-E)^2/E$
AB	31	54	–23	529	529/54 = 9.796
Ab	469	446	+23	529	529/44 = 1.186
aB	158	162	+23	529	529/162 = 3.265
ab	1315	1338	–23	529	529/1338 = 0.395

$\sum (O-E)^2/E = 14.462$

Degrees of freedom in this case = $(r – 1) (c – 1) = (2 – 1) (2 – 1) = 1$.

The table value of χ^2 for 1 degree of freedom at 5 per cent level of significance is 3.841. The calculated value of χ^2 is much higher than this table value and hence the result of the experiment does not support the hypothesis. We can, thus, conclude that vaccination is effective in preventing the attack from smallpox.

Sum 5: Two research workers classified some people in income groups on the basis of sampling studies. Their results are as follows:

Investigators		**Income groups**		**Total**
	Poor	**Middle**	**Rich**	
A	160	30	10	200
B	140	120	40	300
Total	**300**	**150**	**50**	**500**

Show that the sampling technique of at least one research worker is defective.

Solution

Let us take the hypothesis that the sampling techniques adopted by research workers are similar (i.e., there is no difference between the techniques adopted by research workers). This being so, the expectation of A investigator classifying the people in

OF	EF	(O-E)	$(O-E)^2/E$
160	120	40	1600/120 = 13.33
30	60	–30	900/60 = 15.00

10	20	–10	100/20 = 5.00
140	180	–40	1600/180 = 8.88
120	90	30	900/90 = 10.00
40	30	10	100/30 = 3.33

$\sum (O-E)^2/E = 55.54$

Degrees of freedom = (c – 1) (r – 1) = = (3 – 1) (2 – 1) = 2.

The table value of χ2 for two degrees of freedom at 5 per cent level of significance is 5.991. The calculated value of χ2 is much higher than this table value which means that the calculated value cannot be said to have arisen just because of chance. It is significant. Hence, the hypothesis does not hold good. This means that the sampling techniques adopted by two investigators differ and are not similar. Naturally, then the technique of one must be superior than that of the other.

Sum 6: In an anti a malarial campaign in a certain area, quinine was administered to 812 persons out of a total population of 3248. The number of fever cases is shown below:

Treatment	**Fever**	**No Fever**	**Total**
Quinine	20	792	812
No Quinine	220	2216	2436
Total	**240**	**3008**	**3248**

Discuss the usefulness of Quinine in checking malaria.

Solution

Let us take the hypothesis that quinine is not effective in checking malaria.

Observed Frequency	**Expected Frequency**	**$(O-E)^2$**	**$(O-E)^2/E$**
20	60	1600	26.667
220	180	1600	8.889
792	752	1600	2.128
2216	2256	1600	0.709
			$(\sum((O-E)^2/E)=38.393$

$$\chi^2 = \sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

= 38.393

v= (r-1)(c-1) = (2-1) (2-1)=1

v=1, $\chi^2_{0.05}$ = 3.84

The calculated value of χ^2 is greater than the table value. The hypothesis is rejected. Hence, quinine is useful in checking malaria.

Sum 7: Based on information on 1000 randomly selected fields about the tenancy status of the cultivation of these fields and use of fertilizers, collected in an agro-economic survey, the following classification was noted:

Treatment	**Owned**	**Rented**	**Total**
Using Fertilizer	416	184	600
Not using Fertilizer	64	336	400

Total	**480**	**520**	**1000**

Would you conclude that owner cultivators are more inclined towards the use of fertilizer at 5% level? Carry out chi-square test as per testing procedure.

Solution

Let us take the hypothesis that ownership of fields and the use of fertilizers are independent attributes.

Observed Frequency	**Expected Frequency**	**$(O-E)^2$**	**$(O-E)^2/E$**
416	288	16384	56.889
64	192	16384	85.333
184	312	16384	52.513
336	208	16384	78.769
			$(\sum((O-E)^2/E) = 273.504$

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

= 273.504

v= (r-1)(c-1) = (2-1) (2-1)=1

v=1, $\chi^2_{0.05} = 3.84$

The calculated value of χ^2 is much more than the table value. The hypothesis is rejected. Hence, it can be concluded that owner's cultivators are more inclined towards the use of fertilizers.

Sum 8: In an experiment on immunization of cattle from tuberculosis, the following results were obtained.

Particulars	**Affected**	**Not Affected**
Inoculated	12	26
Not Inoculated	16	6

Calculate χ^2 and discuss the effect of vaccine in controlling suspectability to tuberculosis. (5% value of χ^2 for one degree of freedom =3.84)

Solution:

Let us take the hypothesis that the vaccine is not effective in controlling susceptibility to tuberculosis.

Observed Frequency	**Expected Frequency**	**$(O-E)^2$**	**$(O-E)^2/E$**
12.5	17.7	27.04	1.528
15.5	10.3	27.04	2.625
25.5	20.3	27.04	1.332
6.5	11.7	27.04	2.311

			$(\sum((O-E)^2/E)=7.796$

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

= 7.796

v= (r-1)(c-1) = (2-1) (2-1)=1

v=1, $\chi^2_{0.05}$ = 3.84

Since the calculated value of χ^2 is greater than the table value, the hypothesis is not true. We, therefore, conclude that vaccine is effective in controlling susceptibility to tuberculosis.

Sum 9: From the data given below about the treatment of 250 patients suffering from a disease, state whether the new treatment is superior to the conventional treatment.

Treatment	Favourable	Unfavourable	Total
New	140	30	170
Conventional	60	20	80
Total	**200**	**50**	**250**

Solution

Observed Frequency	Expected Frequency	$(O-E)^2$	$(O-E)^2/E$

140	136	16	0.118
60	64	16	0.250
30	34	16	0.471
20	16	16	1.000
			$(\sum((O-E)^2/E)=1.839$

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

= 1.839

v= (r-1)(c-1) = (2-1) (2-1)=1

v=1, $\chi^2_{0.05}$ = 3.84

The calculated value of χ^2 is less than the table value. The hypothesis is accepted. Hence, there is no significant difference between the new and conventional treatment.

Sum 10: 1000 students at College level are graded according to IQ and their economic conditions. Use Chi-square test to find out whether there is any association between economic conditions and the level of IQ.

Economic Condition	**Intelligent Quotient**			**Total**
	High	**Medium**	**Low**	
Rich	160	300	140	600

Poor	140	100	160	400
Total	**300**	**400**	**300**	**1000**

Solution

Let us take the hypothesis that there is no association between economic conditions and the level of IQ. On the basis of this hypothesis the expected frequencies corresponding to (a) and (b) are:

Observed Frequency	**Expected Frequency**	**$(O-E)^2$**	**$(O-E)^2/E$**
160	180	400	2.222
140	120	400	3.333
300	240	3600	15.000
100	160	3600	22.500
140	180	1600	8.889
160	120	1600	13.333
			$(\sum((O-E)^2/E)=65.277$

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

$= 65.277$

$v= (r-1)(c-1) \quad = (2-1)\ (3-1)=2$

$v=2,\ \chi^2_{0.05} = 5.99$

The calculated value of χ^2 is much greater than the table value. The hypothesis is rejected. Hence, there is association between economic condition and the level of IQ.

ANALYSIS OF VARIANCE (ANOVA)

Analysis of variance (abbreviated as ANOVA) is an extremely useful technique concerning researches in the fields of economics, biology, education, psychology, sociology, and business/industry and in researches of several other disciplines. This technique is used when multiple sample cases are involved. As stated earlier, the significance of the difference between the means of two samples can be judged through either z-test or the t-test, but the difficulty arises when we happen to examine the significance of the difference amongst more than two sample means at the same time. The ANOVA technique enables us to perform this simultaneous test and as such is considered to be an important tool of analysis in the hands of a researcher. Using this technique, one can draw inferences about whether the samples have been drawn from populations having the same mean.

The ANOVA technique is important in the context of all those situations where we want to compare more than two populations such as in comparing the yield of crop from several varieties of seeds, the gasoline mileage of four automobiles, the smoking habits of five groups of university students and so on. In such circumstances one generally does not want to consider all possible combinations of two populations at a time for that would require a great number of tests before we would be able to arrive at a decision. This would also consume lot of time and money, and even then certain relationships may be left unidentified (particularly the interaction effects). Therefore, one quite often utilizes the ANOVA technique and through it investigates the differences among the means of all the populations simultaneously.

Professor R.A. Fisher was the first man to use the term 'Variance'* and, in fact, it was he who developed a very elaborate theory concerning ANOVA, explaining its usefulness in practical field. Later on Professor Snedecor and many others contributed to the development of this technique. ANOVA is essentially a procedure for testing the difference among different groups of data for homogeneity. "The essence of ANOVA is that the total amount of variation in a set of data is broken down into two types, that amount which can be attributed to chance and that amount which can be attributed to specified causes."1 There may be variation between samples and also within sample items. ANOVA consists in splitting the variance for analytical purposes. Hence, it is a method of analyzing the variance to which a response is subject into its various components corresponding to various sources of variation. Through this technique one can explain whether various varieties of seeds or fertilizers or soils differ significantly so that a policy decision could be taken accordingly, concerning a particular variety in the context of agriculture researches. Similarly, the differences in various types of feed prepared for a particular class of animal or various types of drugs manufactured for curing a specific disease may be studied and judged to be significant or not through the application of ANOVA technique. Likewise, a manager of a big concern can analyze the performance of various salesmen of his concern in order to know whether their performances differ significantly.

Thus, through ANOVA technique one can, in general, investigate any number of factors which are hypothesized or said to influence the dependent variable. One may as well investigate the differences amongst various categories within each of these factors which may have a large number of possible values. If we take only one factor and investigate the differences amongst its various categories having numerous possible values, we are said to use one-way ANOVA and in case we investigate two factors at the same time, then we use two-way ANOVA. In a two or more way ANOVA, the interaction (i.e., inter-relation between two independent variables/factors), if any, between two independent variables affecting a dependent variable can as well be studied for better decisions.

BASIC PRINCIPLES OF ANOVA

The basic principle of ANOVA is to test for differences among the means of the populations by examining the amount of variation within each of these samples, relative to the amount of variation between the samples. In terms of variation within the given population, it is assumed that the values of (X_{ij}) differ from the mean of this population only because of random effects i.e., there are influences on (X_{ij}) which are unexplainable, whereas in examining differences between populations we assume that the difference between the mean of the jth population and the grand mean is attributable to what is called a 'specific factor' or what is technically described as treatment effect. Thus while using ANOVA, we assume that each of the samples is drawn from a normal population and that each of these populations has the same variance. We also assume that all factors other than the one or more being tested are effectively controlled. This, in other words, means that we assume the absence of many factors that might affect our conclusions concerning the factor(s) to be studied.

In short, we have to make two estimates of population variance viz., one based on between samples variance and the other based on within samples variance. Then the said two estimates of population variance are compared with F-test, wherein we work out.

$$F = \frac{\text{Estimate of population variance based on between sample variance}}{\text{Estimate of population variance based on within sample variance}}$$

Sum 1: Set up an analysis of variance table for the following per acre production data for three varieties of wheat, each grown on 4 plots and state if the variety differences are significant.

Plot of land		**Per acre production data** **Variety of wheat**	
	A	**B**	**C**
1	6	5	5
2	7	5	4
3	3	3	3
4	8	7	4

Solution: We can solve the problem by the direct method or by short-cut method, but in each case we shall get the same result. We try below both the methods.

Solution through direct method: First we calculate the mean of each of these samples:

Average of X_1 = 6+7+3+8 / 4 = 6

Average of X_2 = 5+5+3+7 / 4 = 5

Average of X_3 = 5+4+3+4 / 4 = 4

Mean of the sample means or $\overline{\overline{X}} = \dfrac{\overline{X}_1 + \overline{X}_2 + \overline{X}_3}{k}$

= 6 + 5 + 3 /3

= 5

Now we work out SS between and SS within Samples

$$SS \text{ between} = n_1\left(\overline{X}_1 - \overline{\overline{X}}\right)^2 + n_2\left(\overline{X}_2 - \overline{\overline{X}}\right)^2 + n_3\left(\overline{X}_3 - \overline{\overline{X}}\right)^2$$

$= 4(6-5)^2 + 4(5-5)^2 + 4(4-5)^2 5$

$= 4 + 0 + 4$

$= 8$

$$SS \text{ within} = \Sigma\left(X_{1i} - \overline{X}_1\right)^2 + \Sigma\left(X_{2i} - \overline{X}_2\right)^2 + \Sigma\left(X_{3i} - \overline{X}_3\right)^2, \qquad i = 1, 2, 3, 4$$

$= \{(6-6)^2 + (7-6)^2 + (3-6)^2 + (8-6)^2\} +$

$\{(5-5)^2 + (5-5)^2 + (3-5)^2 + (7-5)^2\} +$

$\{(5-4)^2 + (4-4)^2 + (3-4)^2 + (4-4)^2\}$

$\{0 + 1 + 9 + 4\} + \{0 + 0 + 4 + 4\} + \{1 + 0 + 1 + 0\}$

$14 + 8 + 2$

24

$$SS \text{ for total variance } = \Sigma\left(X_{ij} - \overline{\overline{X}}\right)^2 \qquad i = 1, 2, 3\ldots$$

$$j = 1, 2, 3\ldots$$

$= (6-5)^2 + (7-5)^2 + (3-5)^2 + (8-5)^2$

$+ (5-5)^2 + (5-5)^2 + (3-5)^2$

$+ (7-5)^2 + (5-5)^2 + (4-5)^2$

$+ (3-5)^2 + (4-5)^2$

$= 1 + 4 + 4 + 9 + 0 + 0 + 4 + 4 + 0 + 1 + 4 + 1$

$= 32$

Source of Variation	SS	d.f.	MS	F-ratio	5% F-limit(from the F-table)
Between sample	8	(3 – 1) = 2	8/2 = 4.00	4.00/2.67 = 1.5	*F*(2, 9) = 4.26
Within sample	24	(12 – 3) = 9	24/9 = 2.67		
Total	32	(12 – 1) = 11			

The above table shows that the calculated value of F is 1.5 which is less than the table value of 4.26 at 5% level with d.f. being v1 = 2 and v2= 9 and hence could have arisen due to chance. This analysis supports the null-hypothesis of no difference is sample means. We may, therefore, conclude that the difference in wheat output due to varieties is insignificant and is just a matter of chance.

Sum 2: To assess the significance of possible variation in performance in a certain test between the grammar schools of a city, a common test was given to a number of students taken at random from the senior fifth class of each of the four schools concerned. The results are given below. Make an analysis of variance of data.

Schools			
A	**B**	**C**	**D**
8	12	18	13
10	11	12	9
12	9	16	12
8	14	14	16
7	4	4	15

Solution

Sample 1 X_1		**Sample 2** X_2	**Sample 3** X_3	**Sample 4** X_4
8		12	18	13
10		11	12	9
12		9	16	12
8		14	6	16
7		4	8	15
Total	**45**	**50**	**60**	**65**
Avg. of X	**9**	**10**	**12**	**13**

Grand Mean of $\bar{\bar{X}}$ = $\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \bar{X}_4$ / N

= 9 + 10 + 12 + 13 /4

= 44/4

= 11

Grand Mean of all Samples of $\bar{\bar{X}}$ = 45 + 50 + 60 + 65 / 20

= 11

Variance between Samples

To obtain variance between samples, calculate square of the deviation of the various samples from the grand average. The mean of sample I is 9 but the grand mean is 11. So we will take the difference between 9 and 11 and square it. Similarly for sample II the mean is

10 but the grand average is 11 and so will take the difference between 10 and 11 and square it. Thus we will have the following table.

Sample 1 $(\bar{X}_1 - \bar{\bar{X}}_1)^2$	**Sample 2** $(\bar{X}_2 - \bar{\bar{X}}_2)^2$	**Sample 3** $(\bar{X}_3 - \bar{\bar{X}}_3)^2$	**Sample 4** $(\bar{X}_4 - \bar{\bar{X}}_4)^2$
4	1	1	4
4	1	1	4
4	1	1	4
4	1	1	4
4	1	1	4
20	**5**	**5**	**20**

Sum of the squares between the samples = 20 + 5 + 5 + 20 = 50

Mean sum of squares between the samples is 50 / 4-1 = 16.7 (because there are four samples and the degrees of freedom are 4-1=3)

Variance within the Samples

Here we find the sum of the squares is the deviations of various items in a sample from the mean values of respective samples. Thus for the first sample, then mean is 9 and so we will take deviations from 10, and so on. The squared deviations are given in the following table.

Sample 1		**Sample 2**		**Sample 3**		**Sample 4**	
					—		

X_1	$(X_1- X_1)^2$	X_2	$(X_2- \bar{X}_2)^2$	X_3	$(X_3- X_3)^2$	X_4	$(X_4- \bar{X}_4)^2$
8	1	12	4	18	36	13	0
10	1	11	1	12	0	9	16
12	9	9	1	16	16	12	1
8	1	14	16	6	36	16	9
7	4	4	36	8	16	15	4
$\sum(X_1- X_1)^2=16$		**$\sum(X_2- X_2)^2=58$**		**$\sum(X_3- X_3)^2=104$**		**$\sum(X_4- X_4)^2=30$**	

Total sum of squares within the samples = 16 + 58 + 104 + 30 = 208

Mean Sum of Squares within the Samples

= 208 / 20-4 = 208 / 16 = 13

It is advisable to check up the calculations by finding out total variation. Total variation is calculated by taking the squares of the deviation of each item from the grand average.

Sample 1		**Sample 2**		**Sample 3**		**Sample 4**	
X_1	$(X_1- X_1)^2$	X_2	$(X_2- X_2)^2$	X_3	$(X_3- X_3)^2$	X_4	$(X_4- X_4)^2$
8	9	12	1	18	49	13	4
10	1	11	0	12	1	9	4
12	1	9	4	16	25	12	1
8	9	14	9	6	25	16	25

7	16	4	49	8	9	15	16
$\sum(X_1 - X_1)^2 = 36$		**$\sum(X_2 - X_2)^2 = 63$**		**$\sum(X_3 - X_3)^2 = 109$**		**$\sum(X_4 - X_4)^2 = 50$**	

Total sum of squares = 36 + 63 + 109 + 50 = 258

Degrees of Freedom = 20 – 1 = 19

When we add the sum of squares between samples and sum of squares within the samples, we get the same total, i.e. 50 + 208 = 258. Hence, our calculations are correct.

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**
Between Samples	50	3	16.7
Within Samples	208	16	13.0
Total	**258**	**19**	

$$F = \frac{\text{Variance between Samples}}{\text{Variance within Samples}} = \frac{16.7}{13} = 1.285$$

The table value of F for v1=3 and v2=16 at 5% level of significance =3.24. The calculated value of F is less than the table value and hence the difference in the mean values of the sample is not significant. i.e., the samples could have come from the same variance.

Sum 3: The three samples below have been obtained from normal populations with equal variances. Test the hypothesis that the sample means are equal.

8	7	12
10	5	9
7	10	13
14	9	12
11	9	14

The table value of F at 5% level of significance for v1=1 and v2=12 is 3.88.

Solution

Let us taken the null hypothesis that there is no significant difference in the means of three samples.

X_1	X_2	X_3
8	7	12
10	5	9
7	10	13
14	9	12
11	9	14
Total **50**	**40**	**60**
$\bar{X}$ **10**	**8**	**12**

$X = 10 + 8 + 12 /3 \quad =10$

Variance between Samples

$(X_1\text{-}X)^2$	$(X_2\text{-}X)^2$	$(X_3\text{-}X)^2$
0	4	4
0	4	4
0	4	4
0	4	4
0	4	4
0	**20**	**20**

Sum of Squares between samples = 0 + 20 + 20 =40

Variance within Samples

$\mathbf{X_1}$	$\mathbf{(X_1- \bar{X}_1)^2}$	$\mathbf{X_2}$	$\mathbf{(X_2- \bar{X}_2)^2}$	$\mathbf{X_3}$	$\mathbf{(X_3- \bar{X}_3)^2}$
8	4	7	1	12	0
10	0	5	9	9	9
7	9	10	4	13	1
14	16	9	1	12	0
11	1	9	1	14	4

	30		**16**		**14**

Sum of Squares within Samples = 30 + 16 + 14 = 60

ANOVA TABLE

Source of Variation	Sum of Squares	V	Mean Square
Between	40	2	22
Within	60	12	5
Total	**100**	**14**	

F=20/5 = 4

For v1=2 and v2=12, $F_{0.05}$ = 3.88

The calculated value of F is greater than the table value. The hypothesis is rejected. Hence, there is significant difference in the sample means.

BUSINESS FORECASTING

The growing competition rapidity of change in circumstances and the trend towards automation demand that decisions in business are not based purely on guesses and hunches rather on a careful analysis of data concerning the future course of events. More time and attention is given to the future than to the past and the question 'what is likely to happen'?

takes precedence over ‘what has happened?’ though no attempt to answer the first can be taken without the facts and figures being available to answer the second.

When estimates of future conditions are made on a systematic basis, the process is referred to as “forecasting” and the figure or statement obtained is known as a “forecast”. In a word where the future is not known with certainty, virtually every business and economic decision rests upon a forecast of future conditions. In fact, when a man assumes the responsibility of running a business, he automatically takes the responsibility for attempting to forecast the future and to a very large extent, his success or failure would depend upon the ability to forecast successfully the future course of events. Forecasting aims at reducing the areas of uncertainty that surround management decision-making with respect to costs, profit, sales, production, pricing, capital investment and so forth. If the future were known with certainty, forecasting would be unnecessary. Decisions could be made and plans formulated on a once-and-for-all basis, without the need for subsequent revision. But uncertainty does not exist, future outcomes are rarely assured and, therefore, organized system of forecasting is necessary rather than the establishment of predictions that are based on hunches, intuition or guesses. In fact, a good manager is not so much one who can minimize the effect of past mistakes but rather one who can successfully manage the future. The forecasts may be wrong, but they must be made.

The question is not ‘forecast or no forecast’? Instead, it is: “What is kind of forecast?” It may be noted that the value of a forecast is not merely its accuracy but the fact that making it requires a balanced consideration of factors influencing future developments, right or wrong.

Role of Forecasting in Business

It should be realized at the outset that the object of business forecasting it not to determine a curve or series of figures that will tell exactly what will happen, say, a year in advance, but it is to make analysis based on definite statistical data, which will enable and execute to take advantage of future conditions to a greater extent than he could do without them. In many respects, the future tends to move like the past. This is a good thing, since without some element of continuity between past, present and future, there would be little possibility of successful prediction. But history is not likely to repeat itself and we would hardly expect economic conditions next year or over the next ten years to follow a clear-cut precedent. Yet, frequently past patterns prevail sufficiently to justify using the past as a basis for predicting the future.

While forecasting, one should note that it is impossible to forecast the future precisely – there always must be some range of error allowed for in the forecast. Statistical forecasts are those in which we can use the mathematical theory of probability of measure the risks of errors in predictions.

Methods of Forecasting

There is nothing new about business forecasting as, for centuries, the businessmen have tried to adjust themselves in such a manner as to make the best out of the future conditions. The rule-of-thumb method has been widely practised in business. It consists of deciding about the future in terms of past experience and familiarity with the problem at hand. Even today, this method is very widely used in business. However, it can lead to absurd conclusions if employed by the inexperienced.

In recent years, the techniques of forecasting have improved to a marked degree and are applicable to almost every sphere of business activity. Attempts are being made to make forecasting as scientific as possible. The base of scientific forecasting is statistics, i.e., numerical data on business trends which many businessmen fail to acquaint themselves with. However, forecasting business change involves more than an analysis of statistical data – it also embodies the prediction of economic change, such as secular trend, seasonal variation and a consideration of cause and effect. To handle the increasing variety of managerial forecasting problems, several forecasting techniques have been developed in recent years. Forecasting techniques vary from simple expert guesses to complex analyses of mass data. Each technique has its special use and care must be taken to select the correct technique for a particular situation. Also before applying a method of forecasting, the following questions should be answered:

1. What is the purpose of the forecast – how it is to be used?

2. What are the dynamics and components of the system for which the forecast will be made?
3. How important is the past in estimating the future?

The following are some of the important methods of forecasting:

1. Business Barometers
2. Extrapolation
3. Regression Analysis
4. Econometric Models
5. Forecasting by the use of Time Series Analysis
6. Opinion Polling
7. Causal Models
8. Exponential Smoothing
9. Survey Method

A forecast is usually a combination of several techniques.

1) Business Barometers

Of great assistance in practical forecasting is a series that can be used “index” or “indicator” is also widely, through loosely, used in business statistics; sometimes the term is used to mean simply an indicator of the present economic situations and sometimes it is used to designate an indicator of future conditions.

The following are some of the important series which aid businessmen in forecasting:

1. Gross National Product
2. Employment
3. Wholesale Prices
4. Consumer Prices
5. Industrial Production
6. Volume of bank deposits and currency outstanding

7. Consumer Credit
8. Disposable Personal Income
9. Departmental Store Sales
10. Stock Prices
11. Bond Yields

The list by no means exhaustive, nor is the arrangement necessarily in order of importance. Several of the above series are composite average or total – or indexes of these averages or totals. Analysis also should be made of the major components of these series.

Index numbers relating to different activities in the field of production, trade, finance, etc., may also be combined into a general index of Business Activity. This general index refers to the general conditions of trade and industry. But the behaviour of individual industries of trades might show a different trend from that of the Composite Business Activity Index. Also general boom or depression may be reflected in a majority of separate industries and trades, yet some industries and trade might show quite contrary tendencies. Hence the study of general business conditions, as revealed by the Composite Business Index, should be supplemented by special studies of individual business based on separate indices. The trends indicated by barometers will guide the businessmen as to whether the stocks of goods should be increased or released or whether to increase investment or not, etc.,

2) Extrapolation

Extrapolation is the simplest yet often a useful method of forecasting. In many forecasting situations the most reasonable expectation is that the variable will follow its already established path. Extrapolation relies on the relative constancy in the pattern of past movements in some time series. Strictly speaking, nothing needs to be known causation – why the series moves as it does. But in practice the justification does involve the nature of the growth process being described. Extrapolation is used frequently for sales forecasts and for other estimates when “better” forecasting methods may not be justified.

Since extrapolation assumes that the variable will follow its established pattern of growth, the problem is to determine accurately the appropriate trend curve and the values of its parameters. Numerous alternative trend curves can be used for the purpose of business forecasting. Some of the most useful ones are:

1. Arithmetic Trend: The straight-line arithmetic trend assumes that growth will be by a constant absolute amount each year
2. Semi-log trend: The semi-logarithmic trend assumes a constant percentage increase each year. Since the annual increment is constant in logarithms, this line translates into a straight line when drawn on paper within a logarithmic vertical scale
3. Modified Exponential Trend: This curve assumes that each increment of growth will be a constant per cent less than 100 of the previous one. The line tends generally to approach, but never quite reach a constant asymptote, which may be thought of as an upper limit.
4. Logistic Curve: The logistic curve has both an upper asymptote and a lower asymptote. It assumes a 'law of growth' involving increasing increments from an initial low value and then gradual slowing down of growth as 'maturity' is approached
5. The Gompertz Curve: The Gompertz is curve with similar properties as prescribed above and is often used to describe growth of industrial output

Selection of an appropriate growth curve can be by empirical and theoretical considerations. Empirically, it is a question of selecting the curve that best fits the past movement of the data. Theoretical matters which intervene in that logic may support a particular growth pattern. For example, population growth, when there are no restraints, implies a geometric pattern of growth, as has been known since Malthus. With limited resources, however, population is sometimes thought to grow along a logistic curve. Lest these theoretical consideration notions be taken too seriously, it should be emphasized that empirical consideration may lead us quickly to a more realistic and less restrictive notion of the relevant growth curve.

3) Regression Analysis

The regression approach offers many valuable contributions to the solution of the forecasting problem. It is the means by which we select from among the many possible or theoretically suggestive relationships between variable in a complex economy those which will be useful for forecasting. With it, one makes the jump from intuitive evaluation on the relationship between two variables to precise quantified knowledge. If two variables are functionally related then knowledge of one will make possible an estimate of the other. For example, if we know that advertising expenditure and sales are correlated then for a given advertising expenditure, we can find out the probable increase in sales or vice-versa.

Regression relationship may involve one predicted or dependent and one independent variable – simple regression, or it may involve relationships between the variable to be forecast and several independent variables – multiple regression. Statistical techniques to estimate the regression equations are often fairly complex and time-consuming but there are many computer programmes now available that estimate simple and multiple regressions quickly. Regression analysis is also the econometrician's standard tool in his research.

With the help of regression analysis, the level of industry sales is usually forecasted first and the company forecast is then made on the basis of predicted behaviour of the whole industry. The demand for most commodities often depends upon multiplicity of aggregate economic variables such as GNP, disposable income, prices, population, employment, industrial production. A careful analysis of these factors is made. It may be pointed out that the factors selected for analysis should be kept as low as possible because the danger of seriously imprecise results is multiplied as regressions are made more complex or if curvilinear relationships are used.

There are two dangers in using regression analysis for forecasting:

1. There is possibility of a mechanistic approach accepting with little question the relationship which the calculations reveal – perhaps that with the highest r^2 – and applying it to the forecast. There are many possibilities for spurious correlation among

time series as many series move together over time even where there is no conceivable connection between them

2. There is the risk that estimated regression is false. The forecaster must always use his judgement and knowledge of the facts and of the underlying theory

4) Econometric Models

Econometric techniques, which originated in the eighteenth century, have recently gained in popularity for forecasting, Much of the revival of econometrics is attributed to the growth of computer technology. The term econometrics refers to the application of mathematical economic theory and statistical procedures to economic data in order to verify economic theorems and to establish quantitative results in economics. An econometrician is, therefore, an economist, a statistician and a mathematician, all in one. Econometric models take the form of a set of simultaneous equations. The values of the constants in such equations are supplied by a study of statistical time series and a large number of equations may be necessary to produce an adequate model. The work of computations is greatly facilitated by electronic data processing equipment like computer etc.,

At the present time, most short-term forecasting uses only statistical methods with little qualitative information. However, in the years to come when most large companies develop and refine econometrics models of their major business, this tool of forecasting will become more popular. However, it should be remembered that the development of an econometric model requires sufficient data so that the correct relationships can be established. Hence, when data are scarce – for example, when a product is first introduced into the market – this method cannot profitably be employed.

The econometric model is, in principle, the most formal, since the forecast is based on an explicit mathematical model. The model states in detail and in quantitative terms, the way in which the various aspects of the economy are interrelated. Theoretically, the model makes possible a wholly mechanical forecast, because once values have been estimated for the exogenous variables, the solution of the model gives specific values for the predicted variables. But in actual practice, qualitative and quantitative forecasters have tended to come together. The 'artist' forecaster has become fully aware of the fact that he needs quantitative relationships, while the econometric forecaster has learnt that in some instances, quantitative relationships have to be modified by qualitative factors.

The econometric model provides the forecaster with a record of the prediction with a clear statement of the assumptions concerning exogenous variables and the solution of the model – it is often possible or at least it is made easier to trace and reproduce the causes for success as well as failures. One can learn just where errors were made and, hopefully, where improvements can be made. Thus, discredited hypothesis may be dropped and new ones can be substituted which ultimately will lead to better understanding of the economic system and business fluctuations.

The econometric models are not very popular in practice because it is probably neither necessary nor feasible for every business forecaster to construct his own model of the economy. The effort and cost involved in a fully developed econometric model are well beyond most forecasting operations. Thus, most forecaster will probably rely for some time on the basic aggregate model developed at research institutes or at universities. These models may be used to make predictions and to test out alternative assumptions about Government policy or the other exogenous aspects of the economy. With the help of the models and hopefully sector analysis of his own industry the business forecaster will be in a better position to augment other more familiar approaches. The greater the understanding of the various forecasting methods and of their inter-relationships, the better the forecasts will be.

5) Forecasting by the use of Time Series Analysis

Time series analysis helps to identify and explain:

1. Any regular or systematic variation in the series of data which is due to seasonality – the 'Seasonals"
2. Cyclical patterns
3. Trends in the data
4. Growth rates of these trends. Unfortunately, most existing methods identify only the seasonals, the combined effect of trends and cycles and the irregular or chance component. That is, they do not separate trend from cycles

This is not to say that those other effects are not to some degree manageable. The suggestion is rather that the analysis for trend and seasonal effects and the projection of these two sets of forces should be understood to be the first step in the forecast and that, taking into account such cyclical and residual forces as may be manageable, further refinements may be made.

Many statisticians consider time series analysis a somewhat useless tool. One critic, M.J. Moroney, said that 'Economic forecasting like weather forecasting in England is only valid for the next six hours or so. Beyond that it is sheer guesswork"

In any event, although the limitation of time series analysis must be understood, the importance and usefulness of the procedures should not be underestimated. The analysis serves two purposes:

1. It does provide an initial approximation forecast that takes into account those empirical regularities which may, with reasonable assurance, be expected to persist
2. After the trend and seasonal effects have been identified and measured, the original data may be adjusted for these influences, yielding a new historical time series consisting of the trend and seasonally adjusted data. This new time series may be helpful in the analysis and interpretation of cyclical and residual influences

It should be noted that this method of forecasting can be used only when several years' data for a product or product line are available and when relationships and trends are both clear and relatively stable.

6) Opinion Polling

Opinion poll is the survey of opinion of experts, i.e., knowledgeable persons in the field whose views carry lot of weight. For example, a survey of opinion of sales representatives, wholesalers, retailers, etc,, shall be of great help in formulating demand projections. The survey Research Centre of the University of Michigan conducts an annual pool regarding the future plans of consumers. The answer to many questions are translated into short-run demand for colour television sets, automobiles and other consumer products.

7) Causal Models

A causal model is the most sophisticated kind of forecasting tool, It expresses mathematically the relevant causal relationships, and may include pipeline considerations (i.e., inventories) and market survey information. It may also directly incorporate the results of a time series analysis.

The causal model takes into account everything known of the dynamics of the flow system and utilizes predictions of related events such as competitive action, Strikes and promotions. If the data are available, the model generally include factors for each location in the flow chart and connects these by questions to describe overall product flow. If certain kinds of data are lacking, initially it may be necessary to make assumptions about some of the relationships and then track what is happening to determine if the assumptions are true. Typically, a causal model is continually revised as more knowledge about the system becomes available. S

8) Exponential Smoothing

This method is an outgrowth of the recent attempts to maintain the smoothing function of moving averages without then corresponding drawbacks and limitations. Exponential smoothing is a special kind of weighted average and is found extremely useful in short-term forecasting of inventories and sales.

When this method is applied and we wish to forecast the value of a time series for the period t + 1 on the basis of information available just after the period t the forecast is best considered as a function of two components : the actual value of the series for period t and the forecasted value for the same period made in the previous period t-1. The use of both realized and estimated values available now for predicting future values is better than the use of either alone, since the actual value in period t might have been unduly influenced by random factors and, or the conditions that led to the forecast for period, t, may not hold any longer.

9) Survey Method

The survey method is very widely used as a tool of forecasting for the existing and new products. Field surveys are conducted and the necessary information, both, quantitative and qualitative obtained. Forecasts are made out about likely demand expenditure on consumer durables etc., Attitude of consumers about consumption of different items provides very useful information,

An important limitation of the survey method is that the information supplied may be biased. However, bias could be reduced to minimal by proper designing of questionnaires, skilfully conducting the interviews and verifying the results by other methods.

* * * * *

POSSIBLE QUESTIONS

PART A (ONE MARKS)

(ONLINE EXAMINATION)

PART – B (8 MARKS)

1. Explain the assumptions of Chi-square test.

2. Define hypothesis. Explain the procedure of hypothesis testing.

3. 200 digits are chosen at random from a set of tables. The frequencies of the digits are as follows.

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	18	19	23	21	16	25	22	20	21	15

Use Chi square test to assess the correctness of the hypothesis that the digits were

Distributed in equal numbers in the tables from which they were chosen.

4. State the basic assumptions of the analysis of variance.

5. The sales manager of a large company conducted a sample survey in states A and B taking 400 samples in each case. The results were

	State A	**State B**
Average Sales	Rs. 2,500	Rs. 2,200
Standard Deviation	Rs. 400	Rs. 550

Test whether the average sales is the same in the two states at 1% significant Level (Table value is : 2.58)

6. Describe the concept of exponential smoothing in detail.

7.Describe the steps involved in chi square test.

8. The Height of 10 males of a given locality are found to be 175, 168, 155, 170, 152, 170, 175, 160, 160, 165 cms. Based on this sample of 10 items test the hypothesis that the height of males is 170cms. Calculate t test at 5% significant level (table value : 2.76)

KARPAGAM ACADEMY OF HIGHER EDUCATION

(DEEMED UNIVERSITY ESTABLISHED UNDER SECTION 3 OF UGC ACT 1956)

DEPARTMENT OF COMMERCE (III - B.COM - PA)

RESEARCH METHODOLOGY (15PAU0503)

UNIT I

1	A study to gain familiarity with a phenomenon to achieve new insights is :	Exploratory research	Descriptive Research	Diagnostic Research	Hypothesis testing Research	Exploratory research
2	Study to determine the frequency with which something occurs uis called	Diagnostic Research	Descriptive Research	Exploratory research	Hypothesis testing Research	Diagnostic Research
3	Study to portray accurately characteristics of a particular individual ,	Exploratory research	Descriptive Research	Diagnostic Research	Hypothesis testing Research	Descriptive Research
4	In social science and business research we often use:	Ex post facto research	Hypothesis testing Research	Diagnostic Research	Exploratory research	Ex post facto research
5	If the researcher has no control over the variables it is termed as :	Exploratory research	Ex post facto research	Diagnostic Research	Hypothesis testing Research	Ex post facto research
6	Critical evaluation made by the researcher with the facts and information already	Exploratory research	Analytical Research	Diagnostic Research	Hypothesis testing Research	Analytical Research
7	Gathering knowledge for knowledge sake is termed as :	Exploratory research	Pure Research	Diagnostic Research	Hypothesis testing Research	Pure Research

8	Research aiming to find an solutions for an immediate problem is called:	Applied Research	Exploratory research	Diagnostic Research	other research	Applied Research
9	Research to find reason , why people think or do certain things is an example	Qualitative research	Quantitative Research	Applied Research	Fundamental research	Qualitative research
10	Data based research coming up with conclusions which are capable of being verified	Empirical Research	Pure Research	Conceptual Research	Basic Research	Empirical Research
11	To develop new concepts or to reinterpret existing ones, philosophers and thinkers	Empirical Research	Conceptual Research	Pure Research	Basic Research	Conceptual Research
12	A researcher is free to pick up a problem in	Conclusion-oriented Research	Decision-oriented Research	Historical Research	Diagnostic Research	Conclusion-oriented Research
13	A researcher will not be free to embark upon research according to his own	Historical Research	Decision-oriented Research	Conclusion-oriented Research	Diagnostic Research	Decision-oriented Research
14	Decision-oriented Research is always for the need of the	Researcher	Society	Decision maker	Others	Decision maker
15	Research carried out on over several time periods are called	One-time Research	Longitudinal Research	Field setting Research	Clinical Research	Longitudinal Research
16	Research studies with substantial structure and specific hypothesis to be	Formalized Research	Empirical Research	Historical Research	One-time Research	Formalized Research
17	Characteristics of research that allows research results to be verified by replicating	Logical	Empirical	Historical	Replicable	Replicable

18	The methods or techniques used by researchers in performing research	Research Methods	Research techniques	Research Methodology	Research process	Research Methods
19	The way to systematically solve the research problem is called:	Research Methods	Research techniques	Research process	Research Methodology	Research Methodology
20	Conceptual Literature comprises of :	Concepts and Theories	Earlier studies	Scientific enquiry	Specific Details	Concepts and Theories
21	Empirical literature comprises of :	Concepts and Theories	Earlier studies	Scientific enquiry	Specific Details	Earlier studies
22	if we want to identify the reasons for human behavior we can best apply	Qualitative research	Quantitative research	Fundamental research	Analytical research.	Qualitative research
23	Attitude or opinion research is	qualitative research	quantitative research	conceptual research	fundamental research.	qualitative research
24	To determine the frequency with which something occurs or associated with	descriptive research	formulative research	diagnostic research	hypothesis resear	diagnostic research
25	Research means ________	find out new things	find out old things	find out methods	find out techniqu	find out new things
26	. Depth interview defined as	Medge Jhon	Dr. Fay. B. Karpf	P.V. Young	Morton.	P.V. Young
27	Primary data is _______ source of collection	Second	Indirect	non direct	Direct	Direct

28	Over Rapporting __________	to collect the data	for good interaction	should be avoided	should be used	should be avoided
29	Sample size should be possible with the help of	Society	Available resources	Person	hypothesis	Available resources
30	Good research is	Progressive	Systematic	Informative	non informative	Systematic
31	Failure to set out clearly the objectives of research are bound to lead to	Partly true	partly false	false	true	true
32	Which one if the following is not the essentials of sampling?	Representative	Adequacy	Homogeneity	Reliability	Homogeneity
33	The main objective of research design is	Budgeting on time	Budgeting on money	Number of respondents to be selected	prepare a structure	prepare a structure
34	The concept which can take on different quantitative values is called as _______	variable	information	sources	data	variable
35	The pre determined plots or the blocks and different treatment are used as	units	experimental units	tratments	control groups	experimental units
36	Which one of the following is not the source of research problem?	Specialization	Observation	Consultations	Resource development	Resource development
37	The formal, systematic and intensive process of carrying on a scientific method of	Research Design	Research	Interpretation	Research analysis	Research

38	The research, which has the purpose of improving a product or a process testing	Statistical research	Applied research	Domestic research	Biological researc	Applied research
39	The research that describes what was in the past is	Schedule	Research	Historical research	Ancient research	Historical research
40	According to the nature of the sample chosen for the study what varies	Data Collection	Sampling	Methodology	Research Problem	Data Collection
41	The device which would retain the actual wording of the respondents is	Recording	)Writing	Copying	Editing	Recording
42	It refers to the relationship between facts or to the of them in some meaningful	Science	Fact	Theory	Economics	Theory
43	What is regarded as an empirically verifiable observation?	Science	Research	Theory	Social Sciences	Research
44	It is a tentative statement about something validity of which is usually unknown	Null hypothesis	Hypothesis	Research hypothesis	Statistical hypothesis	Hypothesis
45	Formulative research studies emphasise on	Bibliographical survey	Discovery of insights	Sample representatives	Specific prediction.	Discovery of insights
46	Research depends upon	random method	systematic method	horizontal method	vertical method	systematic method
47	Which of the following method of sampling belongs to the category of	Judgement sampling	Quota sampling	Convenience sampling	Stratified sampling	Convenience sampling

48	Basic principle of an experimental design is	Duplication	Replication	Simplification	Multiplication	Replication
49	The content of a structured interview is	Statement based on experience	Controlled observation	Predetermined questions	Feelings and beliefs	Predetermined questions
50	Social microscope is	Statistical data	Interview	Observation	Case study method	Case study method
51	In an experimental research a group is exposed to usual conditions is termes	experimental group	cntrol group	confounded relationship	non experimental group	cntrol group
52	A Standard test must have the following quality	) Objectivity	Length	Interest	Nonverbal content	Objectivity
53	----------- constitutes the blue print collection measurement and analysis of	Applied thinking	Operational research	Subjective assessment	Research design	Research design
54	Phenomena, which can take on qualitatively different values even in decimal	Non-continuous	Continuous variable	Independent variables	dependent variables	Continuous variable
55	______________ research is based on the measurement of quantity or amount	Quantitiative	Qualititative	experimental	non experimental	Quantitiative
56	_____________ research is concerned with qualitative phenomena	Quantitative	Qualitative	non experimental	experimental	Qualitative
57	_____________ research is related to some abstract idea or theory	Conceptual	Empirical	Conceptual & Empirical	applied	Conceptual

58	______________ research relies on experience or observation alone	Conceptual	Empirical	Field setting	Historical	Empirical
59	______________ research, which utilizes historical sources like documents,	Field setting	Conceptual	Empirical	Historical	Historical
60	________________ scale is simply a system of assigning number	Ordinal	Interval	Nominal	ratio	Nominal

KARPAGAM ACADEMY OF HIGHER EDUCATION

(DEEMED UNIVERSITY ESTABLISHED UNDER SECTION 3 OF UGC ACT 1956)

DEPARTMENT OF COMMERCE (III - B.COM - PA)

RESEARCH METHODOLOGY (15PAU0503)

UNIT II

S.NO	Questions	Option 1	Option 2	Option 3	Option 4	Answer
1	Deliberate sampling is also known as :	Purposive sampling	Probability Sampling	Random Sampling	Judgment sampling	Purposive sampling
2	When population elements are selected for inclusion in the sample based on the ase of access:	Convenience sampling	Judgment sampling	Random Sampling	Probability Sampling	Convenience sampling
3	Researcher's Judgment used for selecting items, which he considers as representative of the population, is	Judgment Sampling	Convenience sampling	Random Sampling	Systematic Sampling	Judgment Sampling
4	Simple random sampling is also called:	Chance Sampling	Convenience sampling	Random Sampling	Systematic Sampling	Chance Sampling
5	To select a sample as every 10^{th} house on one side of a street is an example of:	Non Probability sampling	Systematic Sampling	Convenience sampling	Random Sampling	Systematic Sampling
6	To draw a sample from non-homogeneous group, the sampling used is:	Stratified Sampling	Deliberate Sampling	Convenience sampling	Random Sampling	Stratified Sampling

7	Quota Sampling is an important form of :	Probability Sampling	Non-Probability Sampling	Convenience Sampling	Systematic Sampling	Non-Probability Sampling
8	Grouping the population and selecting groups for inclusion in the sampling is called:	Cluster Sampling	Area Sampling	Stratified Sampling	Systematic Sampling	Cluster Sampling
9	Sampling that helps the researcher to carry out research where there is no specific list of population concerned	Cluster Sampling	Systematic Sampling	Stratified Sampling	Area Sampling	Area Sampling
10	Sampling determined according to mathematical decisions on the basis o information yielded as Survey	Sequential Sampling	Multi-Stage Sampling	Area sampling	Quota Sampling	Sequential Sampling
11	Data Collected by filling up the Schedules by the enumerators on the basis of replies given by respondents.	Questionnaire	Schedule	Interview	observation	Schedule
12	Method of selecting items to be observed for the given study is called:	Sampling design	Statistical Design	Operational Design	Observational Design	Sampling design
13	Research Design that relates to conditions under which observations are to be made is called:	Sampling design	Statistical Design	Operational Design	Observational Design	Observational Design
14	A concept which can take on different quantitative values is called a	Variable	Research	Research Design	data	Variable
15	______________ data are those which are collected afresh and for the first time	Primary	Secondary	case study method	warranty cards	Primary
16	________________ data are to be originally collected	Secondary	Primary	warranty cards	case study method	Primary

17	___________ data's are those which have already been collected by someone else.	Primary	Secondary	Primary & Secondary	case study method	Secondary
18	______________ Method is most commonly used method specially ;in studies relating to behavioural	Interview	Questionnaires	Schedules	Observation	Observation
19	___________ Most of collecting data involves presentation of oral verbal stimuli	Questionnaires	Interview	Observation	Schedule	Interview
20	The method of collecting inform through personal interview is usually carried out in a _____________ way	Structured	unstructured	formal	informal	Structured
21	______________ Interview is meant to focus attention on the given experience of the respondent and its	Focused	Clinical	Structured	Directive	Focused
22	____________________ Interview is concerned with broad underlying feelings orMotivations	Unstructured	Sttructured	Clinical	Non-directive	Sttructured
23	The main Number of sources of data is	2	3	4	1	2
24	Number of methods of collection of primary data is	2	3	4	5	4
25	Number of questions in a questionnaire should be	5	10	maximum	minimum	minimum
26	Sources of secondary data are	Published sources	Unpublished sources	Neither Published sources nor Unpublished	both Published sources and Unpublished sources	Neither Published sources nor Unpublished sources

27	compared with primary data , secondary data are	more reliable	less reliable	equally reliable	none of these	less reliable
28	In Quantitative classification data are classified on the basis of	attributes	time	location	magnitudes	magnitudes
29	A ------- source is one that itself collects the data.	Primary	Secondary	Published	un published	Secondary
30	The data which is compiled from the records of others is called-----data	Primary	Secondary	un published	Published	Published
31	What type of data will be original in character	unpublised	source data	primary data	secondary data	primary data
32	. What type of data are those which have already been by someone else	secondary data	primary data	source data	unpublised	secondary data
33	Primary data can be collected through ,	direct method	indirect method	other mathods	direct and indirec method	direct and indirec method
34	Which method of data collection is used in studies relating to behavioral sciences	mailed questionnaire	through post	observation method	indirect collection	observation method
35	Which type of data collection is most commonly used method	mailed questionnaire	through post	observation method	indirect collection	observation method
36	If observations takes place in the natural setting, it may be termed as	uncontrolled observation	controlled observation	personal observation	controlled observation	uncontrolled observation

37	If observation takes place according to definite prearranged plans, it is called	uncontrolled observation	controlled observation	personal observation	uncontrolled observation	controlled observation
38	__________ is asking questions face to face	indirect method	mailed questionnaire	through post	personal interview.	personal interview.
39	________ is meant to focus attention on the given experience of the respondent and its effects.	facing interview	sending post	facing indirect interview	mailing through net.	facing interview
40	In __________ interview, the interviewers function is to simply encourage the respondent to talk	Direct	in direct	through post	through mail	in direct
41	More information and depth can be obtained in ________ method	through mail	questionnaire	indirect interview	interview	interview
42	Which method of data collection is very popular	questionnaire method	pilot study	mailed questionnaire	through post	questionnaire method
43	Which method of data collection is used especially in case of big enquiries	questionnaire method	pilot study	mailed questionnaire	through post	questionnaire method
44	Which method of data collection is very much like the collection of data through questionnaires	schedules	inquires	finding	posting method	schedules
45	Journals, books, magazines etc.. are useful sources of collecting	primary data	secondary data	case study method	warranty cards	secondary data
46	Case study method is a very popular method for ,	quantitative method	qualitative method	non qualitative	non quantitative	qualitative method

47	Which method involves careful and complete observation of a unit	pilot study	questionnaire study	case study	schedule	case study
48	The collected raw data to detect errors and are called ,	editing	coding	classification	tabulation	editing
49	When editing is done to assure that the data are	accurate	informal	formal	additional	accurate
50	Pilot study should be taken for,	pre- testing the questionnaire	post testing the questionnaire	pre- testing the hypothesis	post testing the hypothesis	pre- testing the questionnaire
51	Questionnaire should be contain	simple and easy	complex	not understandable	maximum	simple and easy
52	t-test is an important	parametric test	. non-parametric test	normal distribution test	. statistical test	parametric test
53	Random sampling is	Chance sampling	Non probability sampling	Complex sampling	Deliberate sampling.	Chance sampling
54	Probability sampling is	accidental sampling	quata sampling	snow ball sampling	systematic sampling.	systematic sampling.
55	Non – probability sampling is	systematic sampling.	Random sampling.	Cluster sampling.	Quota sampling.	Quota sampling.
56	Which is not a method of non-probability sampling.	accidental sampling	quata sampling.	Purposive sampling	Random sampling.	Random sampling.
57	Which is not a method of probability sampling.	Systematic sampling.	cluster sampling.	Area sampling.	Purposive sampling.	Purposive sampling.

58	What is to be Quota sampling is a method of	Probability sampling	Unrestricted sampling	Non-probability sampling	cluster sample	Non-probability sampling
59	Random sampling conducted to test a questionnaire?	Experiment	Research	Pilot survey	Interview.	Experiment
60	The aim of schedule is	To collect the data in an objective manner	Reliability	Subjective	consistent	To collect the data in an objective manner

KARPAGAM ACADEMY OF HIGHER EDUCATION

(DEEMED UNIVERSITY ESTABLISHED UNDER SECTION 3 OF UGC ACT 1956)

DEPARTMENT OF COMMERCE (III - B.COM - PA)

RESEARCH METHODOLOGY (15PAU0503)

UNIT V

S.NO	Questions	Option 1	Option 2	Option 3	Option 4
1	in research through linking the results	Interpretation	Analysis	Editing	Coding
2	aking comparison between two or mo	etrical Relations	Percentage	Reciprocal Relationship	Asymetrical Relationship
3	n, Median and Mode is an example of	Average	Deviation	Dispersion	Relation
4	to the amount or the magnitude of th	Dispersion	Central Tendency	Time Series	Non-Parametric Test
5	idy for the research task remains inco	Data Collection	Research Report	Hypothesis	Objectives
6	ng of ___is the last step in a research	Summary	Report	Conclusion	Suggestion
7	connections and associations between	Illogical	Irrational	Unscientific	Logical
8	based on a connection or sequence in	Chronological	Logical	Irrational	Unscientific
9	ework upon which long written works	alysis of the Sub	ation of the Final C	reparation of the Rough Dra	ing and Polishing of the Rough
10	nals referred by the researcher are me	Bibliography	Webliography	Citation	Quotation
11	collected from website are mentione	Bibliography	Webliography	Citation	Quotation
12	rts in the letter form, just one or two	Business firms	Banks	Mathematicians	Chemists
13	t type of tabulation for their annual re	Business firms	Banks	Mathematicians	Chemists
14	t type of tabulation for their annual re	Business firms	Insurance	Mathematicians	Chemists
15	t type of tabulation for their annual re	Business firms	inancial Institution	Mathematicians	Chemists
16	sults of their investigations in the for	Mathematicians	Chemists	Students of Literature	Education
17	eport their results in symbols and forr	Mathematicians	Chemists	Students of Literature	Education
18	writer or period or the like with a lib	Mathematicians	Chemists	Students of Literature	Education
19	rt on the results of experimentation ac	Mathematicians	Chemists	Students of Literature	Education

20	rt on the results of experimentation ac	Mathematicians	Chemists	Students of Literature	Psychology
21	ind it necessary to make use of the cas	Psychology	linical Psychologis	News Items	Book Reviews
22	ind it necessary to make use of the cas	Psychology	Social Pathologists	News Items	Book Reviews
23	e content of the book and report on the	News Items	Book Reviews	Internet	Debates
24	reports are generally very comprehens	Business firms	Banks	Government	Financial Institutions
25	t all informations are presented in exh	Popular	Technical	General	Common
26	ns are presented by liberal use of grap	Popular	Technical	General	Common
27	e findings and the policy implications	lethods Employe	Data	Analysis of Data	Conclusion
28	review of the main findings is present	Nature of Study	Methods Employed	Summary	Analysis
29	various sources consulted be prepared	Bibliography	Webliography	Citation	Quotation
30	aire, mathematical derivations, elabor	Bibliography	Webliography	Citation	Appendices
31	pared and be given invariably in the re	Bibliography	Webliography	Index	Appendices
32	thor names in an alphabetical order is	Bibliography	Webliography	Indexing	Appendices
33	for data collection is affixed in the rep	Appendices	Bibliography	Webliography	Indexing
34	gs of the study through oral verbal sti	Oral Presentation	Vritten Presentatio	Printed	Typed
35	uotations in the report and the notice	Footnotes	Quotations	Proverbs	Maxim
36	iness firms report should not exceed _	One	Two	Three	Four
37	balance sheet or financial statements a	Banks	Hospitals	Government	University
38	d with algebraic notations is an examp	Mathematics	Chemistry	Physics	Language
39	king use of symbols and formulae is a	Mathematics	Chemistry	Physics	Language
40	proach where scale is developed in A	rbitrary approach	onsensus approach	interval scale	ratio scale
41	I t is used to measure concepts	onsensus approa	arbitrary approach	interval scale	ratio scale
42	sed to find out the relationship betwee	rbitrary approach	onsensus approach	ratio scale	factor scale
43	ethod are used to find out the individ	onsensus approa	arbitrary approach	item analysis approach	factor approach
44	__ is used to assigning number to sym	nominal scale	ordinal scale	interval scale	ratio scale
45	_____ places events in order	nominal scale	ordinal scale	interval scale	ratio scale
46	Normally rank orders represent in ___	ordinal scale	nominal scale	interval scale	ratio scale

47	cale intervals are adjusted in terms of	ordinal scale	nominal scale	interval scale	ratio scale
48	__ is an absolute or true zero measure	ratio scale	nominal scale	interval scale	ordinal scale
49	_ represents the actual amount of vari	ordinal scale	nominal scale	interval scale	ratio scale
50	: is used to deduct errors and omission	coding	editing	classification	tabulation
51	n the review of the reporting forms by	field editing	central editing	classification	coding
52	ocess of assigning numerals or other sy	coding	editing	classification	tabulation
53	ne by hand or by mechanical or electr	coding	editing	classification	tabulation
54	ication and division is possible in ___	ordinal scale	nominal scale	ratio scale	factor scale
55	ypothesized relationship is to be tested	Orders	Values	Prediction	Thoughts
56	nderlying dimension that account of f	factor	sign test	multivariate techniques	correlation
57	______ is not a method of factor anal	centroid method	pal components m	maximum liklihood method	minimum liklihood method
58	: term path analysis is first introduced	sewall wright	kothari	gilbreth	poission
59	alysis consists methods of classifying	Cluster	factor	structure	MDS
60	____ Test is based on plus or minus si	MDS	Cluster	sign test	MDS

Answer
Interpretation
Percentage
Average
Dispersion
Research Report
Report
Logical
Chronological
Preparation of the Final Outline
Bibliography
Webliography
Business firms
Banks
Insurance
Financial Institutions
Mathematicians
Chemists
Students of Literature
Education

Psychology
Clinical Psychologists
Social Pathologists
Book Reviews
Government
Technical
Popular
Conclusion
Summary
Bibliography
Appendices
Index
Indexing
Appendices
Oral Presentation
Footnotes
One
Banks
Mathematics
Chemistry
Arbitrary approach
arbitrary approach
factor scale
item analysis approach
nominal scale
ordinal scale
ordinal scale

interval scale
ratio scale
ratio scale
editing
field editing
coding
tabulation
ratio scale
Prediction
factor
minimum liklihood method
sewall wright
Cluster
sign test

KARPAGAM ACADEMY OF HIGHER

(DEEMED UNIVERSITY ESTABLISHED UNDER SE

DEPARTMENT OF COMMERCE (III

RESEARCH METHODOLOGY (15

UNIT V

S.NO	Questions	Option 1	Option 2	Option 3
1	The word statistics is used as-------	singular word	a plural word	both singular and plural words
2	In chronological classification data	time	attributes	class intervals
3	Bar diagrams are -------------- dimer	two	three	one
4	Diagrams and graphs are tools of --	collection of data	presentation	analysis
5	In a two dimensional diagram ------	only height is consider	only width is considere	height,width and thickness are considere
6	Data are generally obtained from---	Primary sources	Secondary sources	Both primary and secondary sources
7	In geographical classification data a	area	attributes	time

8	In qualitative classification data are	area	attributes	time
9	In quantitative classification data ar	area	attributes	time
10	Number of source of data is---------	2	3	4
11	Squares and rectangles are-----------	Two dimensional diag	One dimensional diagra	Three dimensional diagram
12	Data originally collected for an inve	Tabulation	Primary data	Secondary data
13	The heading of a row in a statistical	stub	caption	title
14	Statistics can ----------	prove anything	disprove anything	neither prove nor disprove anything but
15	Statistics is also a science of---------	estimates	both a and b	probabilities
16	Statistics is---------	quantitative science	a qualitative science	both quantitative and qualitative science
17	Statistics considers-------------	a single item	a set of item	either a single item or a set of item
18	Statistics can be considered as-------	an art	a science	both an art and science

19	Mid value= ------	lower boundary/2	upper boundary/2	lower boundary+ upper boundary)/2
20	The origin of the word statistics has	statista	status	statistik
21	Graphs of frequency distribution are	histogram	pie diagram	bar chart
22	cubes are-----------------------	Two dimensional diag	One dimensional diagra	Three dimensional diagram
23	------------------ is the difference be	class interval	frequency	number of items
24	Exclusive class intervals suit -------	discrete variables	continuous variables	both
25	To represent two or more sets of int	bar diagram	pie diagram	histogram
26	Histogram is a graph of -------------	Time series	frequency distribution	cumulative frequency distribution
27	Univariate data consists of ----------	one variable	two variables	three variable
28	Which one of the following is a me	Median	range	variation
29	The total of the values of the items	Median	Arithmetic mean	mode

30	The sum of the deviations of the va	– 1	one	two
31	Find the Mean of the following valu	5	18	41
32	Which of the followings represents	First quartile	Third quartile	Second quartile
33	Which of the measure of central ten	Mode	Median	sixth deciles
34	Sum of square of the deviations abo	Maximum	one	zero
35	Median is the value of --------------	First	second	Middle most
36	Find the Median of the following da	160	175	176
37	The position of the median for an in	(N + 1) / 2	(N + 2) / 2	N/2
38	Mode is the value, which has	Average frequency de	less frequency density	greatest frequency density
39	A frequency distribution having two	unimodal	bimodal	trimodal
40	Mode has --------------- stable than m	less	more	same

41	Which of the following is not a mea	Range	quartile deviation	standard deviation
42	Range of the given values is given l	L- S	L+S	S+L
43	Which one of the following is relati	Range	Q.D	S.D
44	Which one of the following refers t	Relationship between	Relationship between t	Average relationship between two varial
45	If r = +1, then the relationship betw	perfectly positive	perfectly negative	no correlation
46	If r = - 1, then the relationship betw	perfectly positive	perfectly negative	no correlation
47	If r = 0, then the relationship betwe	Perfectly positive	perfectly negative	no correlation
48	Coefficient of correlation value lies	1 and –1	0 and 1	0 and ∞
49	While drawing a scatter diagram if	Perfect positive correl	simple positive correl	Perfect negative correlation
50	The range of the rank correlation co	0 to 1	–1 to 1	0 to ∞
51	If r =1, then the angle between two	Zero degree	sixty degree	ninety degree

52	Regression coefficient is independe	Origin	scale	both origin and scale
53	If the correlation coefficient betwee	Positive	negative	not certain
54	If the correlation coefficient betwee	Positive	negative	not certain
55	If b_{xy} and b_{yx} represent regression coeffici	Less than one	greater than one	equal to one
56	Rank correlation was discovered by	R.A.Fisher	Sir Francis Galton	Karl Pearson
57	If two regression coefficients are po	Zero	negative	positive
58	If two-regression coefficients are n	Positive	negative	zero
59	The relationship between two varial	coefficient of correlati	Scatter diagram	Correlogram
60	Correlation coefficient is the --------	Mode	Geometric mean	Arithmetic mean

R EDUCATION

CTION 3 OF UGC ACT 1956)

- B.COM - PA)

PAU0503)

Option 4	Answer

neither singular nor plural word	both singular and plural words
location	time
multi	one
summarization	presentation
Both height and width are considered	only height is considered
neither primary nor secondary sources	Both primary and secondary sources
location	area

location	attributes
magnitude	magnitude
1	2
Multi dimensional diagram	Two dimensional diagram
Published data	Primary data
heading	stub
none of these	neither prove nor disprove anything but it is just a tool
neither a nor b	both a and b
neither quantitative nor qualitative	both quantitative and qualitative science
neither a single item or a set of item	a set of item
neither an art nor a science	both an art and science

lower boundary+ upper boundary	lower boundary+ upper boundary)/2
statistique	status
circle	histogram
Multi dimensional diagram	Three dimensional diagram
range	range
neither	continuous variables
multiple bar diagram	multiple bar diagram
normal distribution	frequency distribution
four	one variable
correlation	Median
range	Arithmetic mean

zero	zero
20	18
Q.D	Second quartile
Mean	Median
Minimum	Minimum
last	Middle most
180	175
N/4	(N + 1) / 2
graetest frequency	greatest frequency density
modal	bimodal
most	less

median	median
LS	L- S
coefficient of variation	coefficient of variation
Relationship between two things	Relationship between two variables
high positive	perfectly positive
low Positive	perfectly negative
both positive and negative	no correlation
0 and –1.	1 and –1
no correlation	Perfect negative correlation
$-\infty$ to ∞	–1 to 1
thirty degree	ninety degree

neither origin nor scale.	Origin
zero	negative
zero	Positive
equal to zero	Less than one
Spearman	Spearman
one	positive
one	Positive
rank correlation	Scatter diagram
median	Geometric mean

KARPAGAM ACADEMY OF HIGHER EDUCATIO

(DEEMED UNIVERSITY ESTABLISHED UNDER SECTION 3 OF

DEPARTMENT OF COMMERCE (III - B.COM - PA

RESEARCH METHODOLOGY (15PAU0503)

UNIT V

S.NO	Questions	Option 1	Option 2
1	__ test should be employed, when the sample size is less than 30	T	ANOVA
2	Student test is also known as ___	t' Test	Vertical
3	The variable ____ranges from minus infinity to plus infinity	ANOVA	Chisquare
4	Like the standard normal distribution, the t-distribution is ____ and has a mean zero	Symmetrical	Irregular
5	When the sample size is greater than __, it is known as large sample	30	40
6	To find mean difference among more than two groups____ test is employed	Analysis of Va	Chisquare
7	To find mean difference between two groups ___ is employed	't' Test	F Test
8	To find whether there exists any association between variables ___test is employed	't' Test	F Test
9	___ is a statistical technique specially designed to test whether the means of more than two quantitative populati	Analysis of Va	Chisquare
10	Chi-square is an example of _____ test	Parametric	Non-Parametri
11	‘t’ test is an example of _____ test	Parametric	Non-Parametri
12	To find difference between observed frequency and expected frequency is known as __	't' Test	F Test
13	What will be the degrees of freedom, if the Chi-square table has 3 rows and 3 columns ___	2	3

14	What will be the degrees of freedom, if the Chi-square table has 4 rows and 3 columns ___	3	4
15	What will be the degrees of freedom, if the Chi-square table has 3 rows and 2 columns ___	1	2
16	__has been defined as a “procedure for the assignment of numbers to a property of objects	Scaling	Measurement
17	__scale is simply a system of assigning number symbols to events in order to label them	Nominal	Ordinal
18	___is usually considered as the principal instrument in research	Hypothesis	Report
19	Hypothesis should state relationship between__, if it happens to be a relational hypothesis	Variables	Data
20	A ___hypothesis is one which predicts the direction of the difference between two items	Directional	Vertical
21	Level of significance measures __ percentage	Data	Error
22	__ denotes the chances of occurrence	Chance	Likelihood
23	If the calculated probability is equal to or smaller than the α value in case of one-tailed test then reject the ___hyp	Valid	Applicable
24	If we are to compare method *A* with method *B* about its superiority and if we proceed on the assumption that bot	Null Hypothesi	Alternative Hyp
25	If a researcher think that the method *A* is superior or the method *B* is inferior, we are then stating what is termed	Null Hypothesi	Alternative Hyp
26	___hypothesis is also known as statistical hypothesis	Alternative	Null
27	___level is the maximum value of the probability of rejecting H_0	Implication	Connotation
28	___ test can be applied, while using nominal scaling	Chisquare	Mean
29	Chisquare test can be employed only when the sample size is above ___	30	40
30	___ test is employed to find association between variables	Correlation	Regression
31	___ test is employed when the sample size is less than 30	Analysis of Va	t'
32	Questions with three or five point scale answer is known as	Graphic Rating	Itemized Rating

33	Assigning number for just identification purpose is known as ___	Nominal	Ordinal
34	When asnweres offered by respondents are ranked, it is known as ___ scaling	Nominal	Ordinal
35	Small sample test is also known as --------------------.	Exact test	t – test
36	The formula for χ^2 is -------------	å(O–E)^2/E	å(E+O)^2/E
37	If a statistic ‘t’ follows student’s t distribution with n degrees of freedom then t^2 follows ------	c^{2}distribution with	c^{2}distribution with
38	The distribution used to test goodness of fit is------------	F distribution	c^{2}distribution
39	Degree of freedom for statistic chi-square incase of contingency table of order 2x2 is----------	3	4
40	If the sample size is greater than 30, then the sample is called --------------.	Large sample	small sample
41	If the sample size is less than 30, then the sample is called --------------.	Large sample	small sample
42	Z – test is applicable only when the sample size is -----------.	zero	one
43	An assumption of t – test is population of the sample is -----------.	Binomial	Poisson
44	The degrees of freedom of chi – square test is ------------.	(r – 1)(c – 1)	(r + 1)(c + 1)
45	In F – test, the variance of population from which samples are drawn are ----------.	equal	not equal
46	The value of Z test at 5% level of significance is ------------.	3.96	2.96
47	In -------, the variance of population from which samples are drawn are equal	t-test	Chi-Square test
48	F – statistics is ---------------------.	Variance betw	Variance withir
49	Analysis of variance utilizes:	t-test	Chi-Square test
50	F – test whish is also known as ------	Chi-Square tes	Z-test
51	The technique of analysis of variance refered to as ------	ANOVA	F – test

52	The two variations, variation within the samples and variations between the samplesare tested for their significar	Chi- square tes	F – test
53	Under ------- classification , the influence of only one attribute or factor is considered.	two way	three way
54	Under --------- classification , the influence of two attribute or factors is considered	two way	three way
55	Completely randomized design is similar to ----------	three way	one way
56	Randomized block design is similar to ----------	two way	three way
57	ANOVA is the technique of analysis of ------	standard deviat	variance
58	Under one way classification , the influence of only ----- attribute or factor is considered	two	three
59	Under two way classification , the influence of only ----- attribute or factor is considered	four	two
60	The degrees of freedom for two samples in t – test is ---------.	$n_1 + n_2 + 1$	$n_1 + n_2 - 2$

N

UGC ACT 1956)

)

Option 3	Option 4		Answer
Chisquare	Correlation		T
Chisquare	Correlation		t' Test
t' distribution	Regression		t' distribution
Asymmetrical	Uneven		Symmetrical
50	60		30
Correlation	F Test		Analysis of Variance
Chisquare	Analysis of Variance		't' Test
Chisquare	Analysis of Variance		Chisquare
Correlation	F Test		Analysis of Variance
Psychometric	Frequency		Non-Parametric
Psychometric	Frequency		Parametric
Chisquare	Analysis of Variance		Chisquare
4	5		4

5	6		6
3	4		2
Numbering	Calculation		Scaling
Interval	Ratio		Nominal
Table of Conte	List of Charts		Hypothesis
Report	Objectives		Variables
Horizontal	alternative		Directional
Inaccuracy	Fault		Error
Odds	Probability		Probability
Null	Legtimate		Null
Substitute Hyp	Option Hypothesis		Null Hypothesis
Substitute Hyp	Option Hypothesis		Alternative Hypothesis
Convinving	Authoratative		Null
Significance	Substance		Significance
Median	Mode		Chisquare
50	60		30
Chisquare	t' Test		Chisquare
Chisquare	Correlation		t'
Attitude Scales	Likert’s Summated Scale		Likert’s Summated Scale

Interval	Ratio		Nominal
Interval	Ratio		Ordinal
normal test	F-test		t – test
å(O-E) /E	å(O-E) ^{2}/O		å(O–E) ^{2}/E
χ^{2}distribution wi	c^{2}distribution with (n+1) degree		c^{2}distribution with (n-1) degrees of freedom
t distribution	Z distribution		c^{2}distribution
2	1		1
population	Null hypothesis		Large sample
population	alternative hypothesis		small sample
small	large		large
normal	exponential		normal
(r + 1)(c – 1)	(r – 1)(c + 1)		(r – 1)(c – 1)
small	large		equal
1.96	0.96		1.96
Z-test	F-test		F-test
Variance betwe	Variance within the rows /		Variance between the samples / variance within the samples
Z-test	F-test		F-test
varience ratio t	t-test		varience ratio test
Z – test	Chi- square test		ANOVA

t-test	Z – test		F – test
one way	many		one way
one way	many		two way
two way	t test		one way
one way	many		two way
mean	range		variance
one	many		one
three	one		two
$n_1 + n_2 + 2$	$n_1 + n_2 - 1$		$n_1 + n_2 - 2$