205. BUSINESS RESEARCH METHODS

UNIT – I

Nature and Scope of business research- Research process- Defining the research Problem and setting the objectives- Preparing research proposal-Review of literature.

Meaning of Research

In the modern complex world every society today is faced with serious social, economic & political problems. These problems need systematic, intelligent and Practical solutions. Problem solving is technical process. It requires the accumulation of new knowledge. ***Research provides the means for accumulating knowledge & wisdom***.

In other words, ***research is a systematic effort of gathering analysis & interpretation of problems confronted by humanity***. It is a thinking process and scientific method of studying a problem and finding solution. It is an in depth analysis based on reflective thinking.

Definitions:

Research in common parlance refers to a search for knowledge. One can also define research as a scientific and systematic search for pertinent information on a specific topic. Research is an academic activity and the term should be used in a technical sense.

A. William Emory defines Research as **"*any organized enquiry designed and carried out to provide information for solving a problem*"**

B. The new Oxford English Dictionary defines research is **"*the scientific investigation into and study of material, sources etc. in order to establish facts and the reach new conclusions*".**

C. Redman and Mory defines, research as **"*a systematised effort to gain new knowledge*''.**

D. **"*A careful investigation or inquiry especially through search for new facts in any branch of knowledge*"** Advanced Leaner"s Dictionary.

Characteristics of Research

The above definitions reveal the following characteristics of Research

1. Research is directed towards the solution of a problem.
2. Research is based upon observable experience or empirical evidence.

3. Research demands accurate observation and description.
4. Research involves gathering new data from primary sources or using existing data for a new purpose.
5. Research activities are characterized by carefully designed procedures.
6. Research requires expertise i.e., skill necessary to carryout investigation, search the related literature and to understand and analyze the data gathered.
7. Research is objective and logical – applying every possible test to validate the data collected and conclusions reached.
8. Research involves the quest for answers to unsolved problems.
9. Research requires courage.
10. Research is characterized by patient and unhurried activity.
11. Research is carefully recorded and reported.
12. A systematic approach must be followed for accurate data. Rules and procedures are an integral part of the process that set the objective. Researchers need to practice ethics and a code of conduct while making observations or drawing conclusions.
13. Research is based on logical reasoning and involves both inductive and deductive methods.
14. The data or knowledge that is derived is in real time from actual observations in natural settings.
15. There is an in-depth analysis of all data collected so that there are no anomalies associated with it.
16. Research creates a path for generating new questions. Existing data helps create more opportunities for research.
17. Research is analytical in nature. It makes use of all the available data so that there is no ambiguity in inference.
18. Accuracy is one of the most important aspects of research. The information that is obtained should be accurate and true to its nature.

    For example, laboratories provide a controlled environment to collect data. Accuracy is measured in the instruments used, the calibrations of instruments or tools, and the final result of the experiment.

Objectives of Research

The objectives of Research can be grouped under the following heads

1. To gain familiarity with a phenomenon or to achieve new insights to it.
2. To portray accurately the characteristics of a particular individual situation or a group.
3. To determine the frequency with which something occurs or with which it is associated with something else.
4. To test a hypothesis of a casual relationship between variables.

Motivations in Research

What makes people to undertake research?

The answer is as follows.

1. Desire to get a research degree along with its benefits.
2. Desire to face the challenge in the solving the unsolved Problem.
3. Desire to get intellectual joy of doing some creative work.
4. Desire to be of service to Society.
5. Desire to get respectability.

Importance of Research

"All progress is born of enquiry. Doubt is often better than overconfidence, for it leads to enquiry & enquiry leads to investigation". Research has an important role to guiding social plan. Knowledge of the society & the cultural behaviour of the people require proper planning for their well development. Because of knowledge & cultural behaviour of human being are interdependent. A reliable knowledge is needed for planning & this is possible only through research. Knowledge is a kind of power with which one can face the implication of a particular Phenomenon. Research provides the basis for all government policies in our economic system. Research helps us in making predictions.

1. A Tool for Building Knowledge and for Facilitating Learning
2. Means to Understand Various Issues and Increase Public Awareness
3. An Aid to Business Success
4. A Way to Prove Lies and to Support Truths
5. Means to Find, Gauge, and Seize Opportunities
6. A Seed to Love Reading, Writing, Analyzing, and Sharing Valuable Information
7. Nourishment and Exercise for the Mind

Research is equally important in seeking answer to various social problems

In addition to this, the significance of research can be understood with the following points.

1. To Professionals in research methodology, research means a source of live hood.
2. To Philosophers & thinkers research may mean the outlet for new ideas and insights.
3. To literary man research means the development of new styles & creative work.
4. To the intellectuals research mean the generalisation of new theories.

Nature and Scope of Business Research

The Nature of research methodology is the definition of the process of collecting the knowledge required to organise or solve the problem. The Scope is to examine the range of management education in India, its rating process, the viewpoint of faculty, HR executives and students.

Production Management: The research performs an important function in product development, diversification, introducing a new product, product improvement, process technologies, choosing a site, new investment etc.

Personnel Management/ HR: Research works well for job redesign, organization restructuring, development of motivational strategies and organizational development.

Marketing Management: Research performs an important part in choice and size of target market, the consumer behavior with regards to attitudes, life style, and influences of the target market. It is the primary tool in determining price policy, selection of channel of distribution and development of sales strategies, product mix, promotional strategies, etc.

Financial Management: Research can be useful for portfolio management, distribution of dividend, capital raising, hedging and looking after fluctuations in foreign currency and product cycles.

Materials Management: It is utilized in choosing the supplier, making the decisions relevant to make or buy as well as in selecting negotiation strategies.

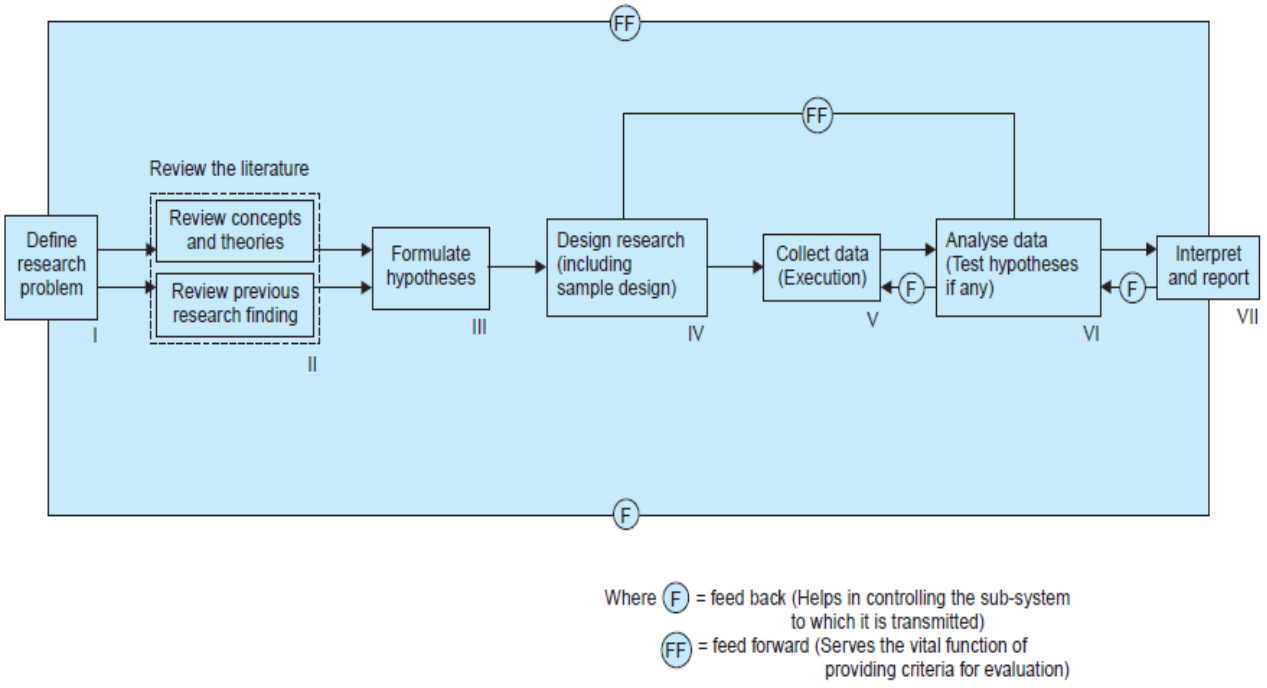
General Management: It contributes greatly in developing the standards, objectives, long-term goals, and growth strategies.

The Research Process / Steps

C.R. Kothari in his book, "Research Methodology: Methods & Techniques" presents a brief overview of a research process. He has given the following order concerning the Research Process.

1. Formulation the Research problem
2. Extensive Literature survey
3. Developing the hypothesis
4. Preparing the research design
5. Determining sample design
6. Collection of Data
7. Execution of the Project
8. Analysis of Data
9. Hypothesis testing
10. Generalisation & Interpretation
11. Preparation of the report.

Research process refers to various decision stages involved in a research project and the relationship between those stages. A researcher has to proceed systematically in the already planned directions with the help of number of steps in sequence.



Steps in the Research Design Process

The steps in the design process interact and often occur simultaneously. For example, the design of a measurement instrument is influenced by the type of analysis that will be conducted. However, the type of analysis is also influenced by the specific characteristics of the measurement instrument.

Step 1: Define the Research Problem

Problem definition is the most critical part of the research process. Research problem definition involves specifying the information needed by management. Unless the problem is properly defined, the information produced by the research process is unlikely to have any value. Coca-Cola Company researchers utilized a very sound research design to collect information on taste preferences. Unfortunately for Coca-Cola, taste preferences are only part of what drives the soft drink purchase decision.

Research problem definition involves four interrelated steps:

i. Management problem / opportunity clarification,
ii. Situation analysis,
iii. Model development,
iv. Specification of information requirements.

The basis goal of problem clarification is to ensure that the decision maker‟s initial description of the management decision is accurate and reflects the appropriate area of concern for research. If the wrong management problem is translated into a research problem, the probability of providing management with useful information is low.

Situation Analysis

The situation analysis focuses on the variables that have produced the stated management problem or opportunity. The factors that have led to the problem/opportunity manifestations and the factors that have led to management‟s concern should be isolated.

A situation analysis of the retail trade outflow problem revealed, among other things, that (1) the local

population had grown 25 percent over the previous five years, (2) buying power per capita appeared to be growing at the national rate of 3 percent a year, and (3) local retail sales of nongrocery items had increased approximately 20 percent over the past five years. Thus, the local retailer"s sales are clearly not keeping pace with the potential in the area.

Step 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.

Step 3: Development of working hypotheses: After extensive literature survey, researcher should state in clear terms the working hypothesis or hypotheses. Working hypothesis is tentative assumption made in order to draw out and test its logical or empirical consequences. As such the manner in which research hypotheses are developed is particularly important since they provide the focal point for research. They also affect the manner in which tests must be conducted in the analysis of data and indirectly the quality of data which is required for the analysis. In most types of research, the development of working hypothesis plays an important role. Hypothesis should be very specific and limited to the piece of research in hand because it has to be tested. The role of the hypothesis is to guide the researcher by delimiting the area of research and to keep him on the right track. It sharpens his thinking and focuses attention on the more important facets of the problem. It also indicates the type of data required and the type of methods of data analysis to be used

Step 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.

A. Exploration,
B. Description,
C. Diagnosis, and
D. 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.

Step 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.

Step 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.

Step 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.

Step 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.

Step 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.

Step 10. Generalisations and interpretation: If a hypothesis is tested and upheld several times, it may be possible for the researcher to arrive at generalisation, i.e., to build a theory. As a matter of fact, the real value of research lies in its ability to arrive at certain generalisations. If the researcher had no hypothesis to start with, he might seek to explain his findings on the basis of some theory. It is known as interpretation. The process of interpretation may quite often trigger off new questions which in turn may lead to further researches.

Step 11. Preparation of the report or the thesis: Finally, the researcher has to prepare the report of what has been done by him.

Selection and Formulation of Research Problems

Research Problem

Problem means a question or an issue to be examined. A research problem refers to some kind of problem which a researcher experiences or observes in the context of either a theoretical or practical situation. The researcher has to find out suitable course of action by which the objective can be attained optimally in the context of given environment. Thus, selection of research problem has high value to the society and the researcher must be able to identify those problems that need an urgent solution.

Requisites or Characteristics of a Good Research Problem

- Clear and unambiguous
- Logical and systematic
- Empirical
- Relation between variables
- Verifiable
- Interesting

Various Aspects of a Research Problem

For an effective formulation of the problem following aspects of the problem are to be considered by the researcher.

A. **Definition of the problem: -** Before one takes up a problem for the study one needs to define it properly. The issues for inquiry are to be identified clearly and specified in details. If any existing theoretical framework is tested, the particular theorem or theories must be identified.

Similarly if there are any assumptions made and terms used the meaning of them must be made clear. As far as possible the statement of the problem should not give any scope for ambiguity.

B. **Scope of the problem: -** The research scholar has to fix up the four walls of the study. The researcher must identify which of the aspects he is trying to prove. Taking the example of sickness he should specify. (1) Whether his study extends to all types of small scale industries, or limited to only few of them. (2) Whether the study is limited to find cause for sickness or also to prescribe certain prescriptions etc.

C. **Justification of the problem: -** Many a time research studies are put to the test of justification or relevance. In the scientific curiosity of the problems, th problem that needs urgent solution must be given preference.

D. **Feasibility of the problem: -** Although a problem needs urgent attention and is justifiable in several respects, one has to consider the feasibility of the same. Feasibility means the possibility of conducting the study successfully. The elements of time, data, Cost is to be taken into consideration before a topic is selected for study.

E. **Originality of the problem: -** In social sciences, particularly in commerce and management, there is no systematic compilation of the works already done or on hand. Two people may be doing a work more or less on similar topic. In such situations it is not advisable to continue work in the same manner. What is advisable is that, each of them should try to focus on different aspects, so that they could enrich the field of knowledge with their studies. Another problem faced by a researcher is that a problem which he intends to do is already worked out. Should he repeat the same or not? This depends upon the situation or circumstances which engage his attention.

Preparing Research Proposal

A research proposal is a document proposing a research project, generally in the sciences or academia, and generally constitutes a request for sponsorship of that research.

Proposals are evaluated on the cost and potential impact of the proposed research, and on the soundness of the proposed plan for carrying it out.

Research proposals generally address several key points:

- What research question(s) will be addressed, and how they will be addressed
- How much time and expense will be required for the research
- What prior research has been done on the topic
- How the results of the research will be evaluated
- How the research will benefit the sponsoring organization and other parties

Types

Research proposals may be solicited, meaning that they are submitted in response to a request with specified requirements, such as a request for proposal, or they may be unsolicited, meaning they are submitted without prior request. Other types of proposals include "pre-proposals", where a letter of intent or brief abstract is submitted for review prior to submission of a full proposal; continuation proposals, which re-literate an original proposal and its funding requirements in order to ensure continued funding; and renewal proposals, which seek continued sponsorship of a project which would otherwise be terminated.

Academic research proposals are generally written as part of the initial requirements of writing a thesis, research paper, or dissertation. They generally follow the same format as a research paper, with ***an introduction, a literature review, a discussion of research methodology and goals, and a conclusion***. This basic structure may vary between projects and between fields, each of which may have its own requirements.

Review of literature

A literature review is an overview of the previously published works on a specific topic. The term can refer to a full scholarly paper or a section of a scholarly work such as a book, or an article. Either way, a literature review is supposed to provide the researcher/author and the audiences with a general image of the existing knowledge on the topic under question.

A good literature review can ensure that a proper research question has been asked and a proper theoretical framework and/or research methodology have been chosen. To be precise, a literature review serves to situate the current study within the body of the relevant literature and to provide context for the reader. In such case, the review usually precedes the methodology and results sections of the work.

Producing a literature review is often a part of graduate and post-graduate student work, including in the preparation of a thesis, dissertation, or a journal article. Literature reviews are also common in a research proposal or prospectus (the document that is approved before a student formally begins a dissertation or thesis).

A literature review can be a type of review article. In this sense, a literature review is a scholarly paper that presents the current knowledge including substantive findings as well as theoretical and methodological contributions to a particular topic.

Literature reviews are secondary sources and do not report new or original experimental work. Most often associated with academic-oriented literature, such reviews are found in academic journals and are not to be confused with book reviews, which may also appear in the same publication. Literature reviews are a basis for research in nearly every academic field.

Types:

The ***main types of literature reviews are: evaluative, exploratory, and instrumental.***

A fourth type, the systematic review, is often classified separately, but is ***essentially a literature review focused on a research question, trying to identify, appraise, select and synthesize all high-quality research evidence and arguments relevant to that question***. A meta-analysis is typically a systematic review using statistical methods to effectively combine the data used on all selected studies to produce a more reliable result.

Torraco (2016) describes an integrative literature review. The purpose of an integrative literature review is to generate new knowledge on a topic through the process of review, critique, and then synthesis of the literature under investigation.

UNIT – II

Research Designs: Exploratory Research Design and Qualitative Studies-Focus groups - Descriptive Research Design- Panel Research –Experimental Research Designs.

RESEARCH DESIGN

A research design is a "Blue Print" for collection, measurement and analysis of data. It outlines how the research will be carried out. It is like glue which sticks together the entire process of research. It provides answers to various questions like - What techniques will be used to gather data. What kind of sampling will be used? How time and cost constraints be dealt with? Etc.

Essentials of Research Design

1. The design should be an activity and time based plan
2. It is always based on research question
3. It guides the selection of sources and types of information
4. It indicates a framework for specifying the relationship among the study's variables
5. Outlines procedures for every research activity
6. It must be appropriate, efficient and economical
7. It should be flexible
8. It must be adequate

Types of Research Design

"*You cannot put the same shoe on every foot" - Syrus*

Although every problem and research objective may seem unique, there are usually enough similarities among problems and objectives to allow decisions to be made in advance about the best plan to resolve the problem. There are some basic research designs that can be successfully matched to given problems and research objectives.

Three traditional categories of research design:

- Exploratory
- Descriptive
- Causal

The choice of the most appropriate design depends largely on the objectives of the research and how much is known about the problem and these objectives. The overall research design for a project may include one or more of these three designs as part(s) of it.

Further, if more than one design is to be used, typically we progress from Exploratory toward Causal.

Basic Research Objectives and Research Design

Research Objective	Appropriate Design
To gain background information, to define terms, to clarify problems and develop hypotheses, to establish research priorities, to develop questions to be answered.	**Exploratory**
To describe and measure phenomena at a point in time.	**Descriptive**
To determine causality, test hypotheses, to make "if-then" statements, to answer questions.	**Causal**

Research Design: Exploratory Research

Exploratory research is most commonly unstructured, "informal" research that is undertaken to gain background information about the general nature of the research problem. Exploratory research is usually conducted when the researcher does not know much about the problem and needs additional information or desires new or more recent information. Exploratory research helps diagnose the dimensions of the problem so that successive research will be on target. It helps to set priorities for research. Exploratory research is used in a number of situations:

- To gain background information
- To define terms
- To clarify problems and hypotheses
- To establish research priorities

A variety of methods are available to conduct exploratory research:

- Secondary Data Analysis

- Experience Surveys
- Case Analysis
- Focus Groups
- Projective Techniques

Categories of Exploratory Research:

- **Experience Survey**s: Issues and ideas may be discussed with persons who have had personal experience in the field.
- **Secondary data analysis:** Another quick and economical source of background information is existing literature containing data that has been compiled for somepurpose other than the purpose in hand
- **Case Study method**: Obtains information from one or a few situations that are similar to the problem situation. Primary advantage is that an entire organisation or entity can be investigated in depth and with meticulous attention to detail.
- **Pilot Studies are used in different types of designs:** Within the context of exploratory research it covers some part of the research on a small scale. Major categories of pilot study include focus group interviews, projective techniques, and depth interviews.

Categories of Pilot Studies:

- **Focus Group interviews:** Unstructured, free flowing, group dynamic sessions that allow individuals the opportunity to initiate the topics of discussion. There is synergistic and spontaneous interaction among the respondents. Found to be highly advantageous.
- **Projective techniques:** An indirect means of questioning the respondents. Uses word association tests, sentence completion test, third person test, role playing technique and Thematic Apperception Test.

- **Depth interviews:** Unstructured, extensive interviews that encourage an individual to talk freely and in depth about a topic.

Qualitative Research

Qualitative research involves collecting and analyzing non-numerical data (e.g., text, video, or audio) to understand concepts, opinions, or experiences. It can be used to gather in-depth insights into a problem or generate new ideas for research.

Qualitative research is the opposite of quantitative research, which involves collecting andanalyzing numerical data for statistical analysis.

Qualitative research is commonly used in the humanities and social sciences, in subjects such as anthropology, sociology, education, health sciences, history, etc.

Approaches to qualitative research:

Qualitative research is used to understand how people experience the world. While there are many approaches to qualitative research, they tend to be flexible and focus on retaining rich meaning when interpreting data.

Common approaches include grounded theory, ethnography, action research, phenomenological research, and narrative research. They share some similarities, but emphasize different aims and perspectives.

Qualitative research approaches:

Approach	What does it involve?
Grounded theory	Researchers collect rich data on a topic of interest and develop theories inductively.
Ethnography	Researchers immerse themselves in groups or organizations to understand their cultures.
Action research	Researchers and participants collaboratively link theory to practice to drive social change.
Phenomenological research	Researchers investigate a phenomenon or event by describing and interpreting participants' lived experiences.
Narrative research	Researchers examine how stories are told to understand how participants perceive and make sense of their experiences.

Qualitative research methods:

Each of the research approaches involve using one or more data collection methods. These are some of the most common qualitative methods:

- **Observations:** Recording what you have seen, heard, or encountered in detailed field notes.
- **Interviews:** Personally asking people questions in one-on-one conversations.
- **Focus groups:** Asking questions and generating discussion among a group of people.

- **Surveys:** Distributing questionnaires with open-ended questions.
- **Secondary research:** Collecting existing data in the form of texts, images, audio or video recordings, etc.

Qualitative researchers often consider themselves "instruments" in research because all observations, interpretations and analyses are filtered through their own personal lens.

Qualitative data analysis:

Qualitative data can take the form of texts, photos, videos and audio. For example, you might be working with interview transcripts, survey responses, field notes, or recordings from natural settings.

Most types of qualitative data analysis share the same five steps:

- **Prepare and organize your data:** This may mean transcribing interviews or typing up fieldnotes.
- **Review and explore your data:** Examine the data for patterns or repeated ideas that emerge.
- **Develop a data coding system:** Based on your initial ideas, establish a set of codes that you can apply to categorize your data.
- **Assign codes to the data**: For example, in qualitative survey analysis, this may mean going through each participant's responses and tagging them with codes in a spreadsheet. As you go through your data, you can create new codes to add to your system ifnecessary.
- **Identify recurring theme:** Link codes together into cohesive, overarching themes.

There are several specific approaches to analyzing qualitative data. Although these methods share similar processes, they emphasize different concepts.

Qualitative data analysis:

Approach	When to use	Example
Content analysis	To describe and categorize common words, phrases, and ideas in qualitative data.	A market researcher could perform content analysis to find out what kind of language is used in descriptions of therapeutic apps.
Thematic analysis	To identify and interpret patterns and themes in qualitative data.	A psychologist could apply thematic analysis to travel blogs to explore how tourism shapes self-identity.
Textual analysis	To examine the content, structure, and design of texts.	A media researcher could use textual analysis to understand how news coverage of celebrities has changed in the past decade.

Approach	When to use	Example
Discourse analysis	To study communication and how language is used to achieve effects in specific contexts.	A political scientist could use discourse analysis to study how politicians generate trust in election campaigns.

Advantages of qualitative research:

Qualitative research often tries to preserve the voice and perspective of participants and can be adjusted as new research questions arise. Qualitative research is good for:

Flexibility: The data collection and analysis process can be adapted as new ideas or patterns emerge. They are not rigidly decided beforehand.

Natural settings: Data collection occurs in real-world contexts or in naturalistic ways.

Meaningful insights: Detailed descriptions of people's experiences, feelings and perceptions can be used in designing, testing or improving systems or products.

Generation of new ideas: Open-ended responses mean that researchers can uncover novel problems or opportunities that they wouldn't have thought of otherwise.

Disadvantages of qualitative research:

Researchers must consider practical and theoretical limitations in analyzing and interpreting their data. Qualitative research suffers from:

Unreliability: The real-world setting often makes qualitative research unreliable because of uncontrolled factors that affect the data.

Subjectivity: Due to the researcher's primary role in analyzing and interpreting data, qualitative research cannot be replicated. The researcher decides what is important and what is irrelevant in data analysis, so interpretations of the same data can vary greatly.

Limited generalizability: Small samples are often used to gather detailed data about specific contexts. Despite rigorous analysis procedures, it is difficult to draw generalizable conclusions because the data may be biased and unrepresentative of the wider population.

Labor-intensive: Although software can be used to manage and record large amounts of text, data analysis often has to be checked or performed manually.

Focus Groups

A focus group is a market research method that brings together 6-10 people in a room to provide feedback regarding a product, service, concept, or marketing campaign. A trained moderator leads a 30-90-minute discussion within the group that is designed to gather helpful information. The moderator arrives with a set list of 10-12 questions that will be shared with the group during their time together that are designed to elicit thoughtful responses from all the participants. The moderator's goal is to hear from everyone and to encourage many different opinions and ideas tobe shared.

Focus group participants are recruited based on their purchase history, demographics, psychographics, or behavior and typically do not know each other. To ensure that the maximum number of different ideas or reactions have been captured from participants, companies typically hold several focus groups, often in different cities; 3-4 is common.

While participants are responding to a moderator question, the moderator and/or other observers take notes.

History:

Formerly referred to as "focused interviews," focus groups were first used during World War II to assess reactions to radio programming. Today the technique has been expanded to evaluate consumer perceptions and reactions.

Focus Group Format:

During the focus group, the moderator takes participants through three different types of questions designed to gather as much information from them as possible. They include:

- ✓ Engagement questions
- ✓ Exploration questions
- ✓ Exit questions

Engagement questions: These are easy questions posed early on to introduce the participants to each other, to make them more at ease, and to familiarize them with the topic to be discussed, whether it's reacting to a new ad campaign for coffee or thinking about self-driving cars.

Exploration questions: Once participants have begun to relax and open up in the group, the moderator begins to ask deeper, probing questions about the topic and how the participants feel about it. These might include, "What makes you say that?" "and "What would be a better solution?"

Exit questions: After the moderator is confident the group has shared all that it can, wrap-up questions are posed to confirm that everything has been said. These might include, "Is there anything I haven't asked that I should have?"

Advantages of Focus Groups:

Focus groups are one type of market research method that are popular because they:

- Are generally lower cost than other methods
- Can generate results very quickly
- Are easy to conduct
- Can supplement verbal responses with body language and other non-verbal cues
- Information gathered is in respondents' own words, which is more accurate
- Technique is flexible and can be adjusted based on group behavior

Dis- advantages of Focus Groups:

Because a focus group involves multiple participants, the downsides of using this technique are generally related to the interactions between participants:

- Participants can be influenced by others in the group
- Domineering participants can skew the results
- Results from a small group can't always be generalized to a larger population

Types of Focus Groups:

Within the general category of focus groups are more specific types of groups that are designed for different scenarios. Some of these include:

- Mini focus groups
- Online focus groups
- Two-way focus group
- Dual moderator focus group
- Client participant focus group

Mini focus groups: Fewer participants are used, bringing the number down from 6-12 to four or five consumers.

Online focus groups: Consumers log into a website using video chat and participate remotely.

Two-way focus group: Focus groups are often conducted behind one-way glass, where researchers can take note of what's going on. In these types of groups, the whole group watches another and comments on what they observe and hear.

Dual moderator focus group: Instead of one moderator in the room, there are two – one to facilitate the discussion and the other to take notes.

Client participant focus group: When a representative of the company or product being studied watches or participates in the discussion.

In terms of collecting qualitative data from multiple respondents, focus groups are a popular market research tool.

Descriptive Research Design

Descriptive research aims to accurately and systematically describe a population, situation or phenomenon. It can answer what, where, when and how questions, but not why questions.

A descriptive research design can use a wide variety of research methods to investigate one or more variables. Unlike in experimental research, the researcher does not control or manipulate any of the variables, but only observes and measures them.

Types of Descriptive Research:

Descriptive research is classified into different types according to the kind of approach that is used in conducting descriptive research. The different types of descriptive research are highlighted below:

Descriptive-survey

Descriptive survey research uses surveys to gather data about varying subjects. This data aims to know the extent to which different conditions can be obtained among these subjects.

For example, a researcher wants to determine the qualification of employed professionals in Maryland. He uses a survey as his research instrument, and each item on the survey related to qualifications is subjected to a Yes/No answer.

This way, the researcher can describe the qualifications possessed by the employeddemographics of this community.

Descriptive-normative survey

This is an extension of the descriptive survey, with the addition being the normative element. In the descriptive-normative survey, the results of the study should be compared with the norm.

For example, an organization that wishes to test the skills of its employees by a team may have them take a skills test. The skills tests are the evaluation tool in this case, and the result of thistest is compared with the norm of each role.

If the score of the team is one standard deviation above the mean, it is very satisfactory, if within the mean, satisfactory, and one standard deviation below the mean is unsatisfactory.

Descriptive-status

This is a quantitative description technique that seeks to answer questions about real-life situations. For example, a researcher researching the income of the employees in a company, and the relationship with their performance.

A survey will be carried out to gather enough data about the income of the employees, then their performance will be evaluated and compared to their income. This will help determine whether a higher income means better performance and low income means lower performance or vice versa.

Descriptive-analysis

The descriptive-analysis method of research describes a subject by further analyzing it, which in this case involves dividing it into 2 parts. For example, the HR personnel of a company that wishes to analyze the job role of each employee of the company may divide the employees into the people that work at the Headquarters in the US and those that work from Oslo, Norway office.

A questionnaire is devised to analyze the job role of employees with similar salaries and who work in similar positions.

Descriptive classification

This method is employed in biological sciences for the classification of plants and animals. A researcher who wishes to classify the sea animals into different species will collect samples from various search stations, then classify them accordingly.

Descriptive-comparative

In descriptive-comparative research, the researcher considers 2 variables that are notmanipulated, and establish a formal procedure to conclude that one is better than the other. For example, an examination body wants to determine the better method of conducting tests betweenpaper-based and computer-based tests.

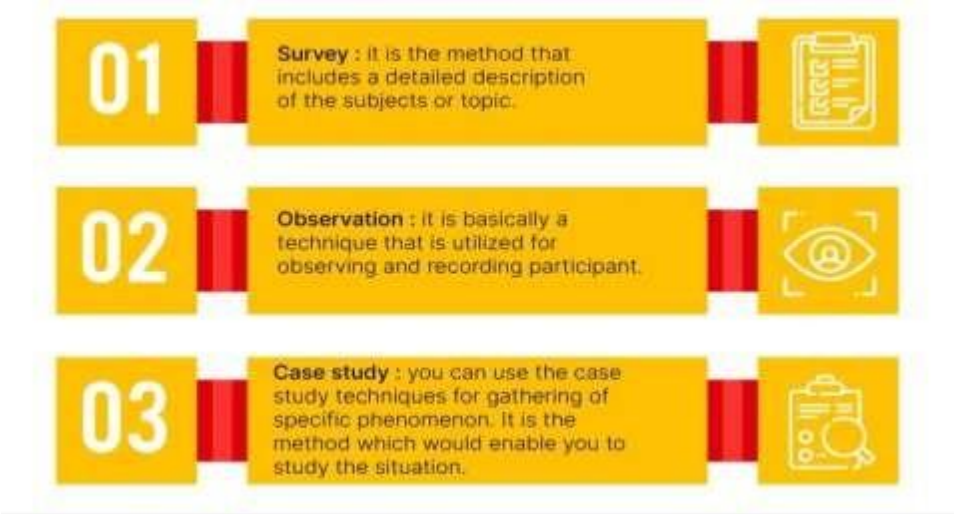
A random sample of potential participants of the test may be asked to use the 2 different methods, and factors like failure rates, time factors, and others will be evaluated to arrive at the best method.

Correlative Survey

Correlative surveys are used to determine whether the relationship between 2 variables is positive, negative, or neutral. That is, if 2 variables, say X and Y are directly proportional, inversely proportional or are not related to each other.

Descriptive research methods:

Descriptive research is usually defined as a type of quantitative research, though qualitative research can also be used for descriptive purposes. The research design should be carefully developed to ensure that the results are valid and reliable.



Surveys

Survey research allows you to gather large volumes of data that can be analyzed for frequencies, averages and patterns. Common uses of surveys include:

- Describing the demographics of a country or region
- Gauging public opinion on political and social topics
- Evaluating satisfaction with a company's products or an organization's services

Observations

Observations allow you to gather data on behaviours and phenomena without having to rely on the honesty and accuracy of respondents. This method is often used by psychological, social and market researchers to understand how people act in real-life situations.

Observation of physical entities and phenomena is also an important part of research in the natural sciences. Before you can develop testable hypotheses, models or theories, it's necessaryto observe and systematically describe the subject under investigation.

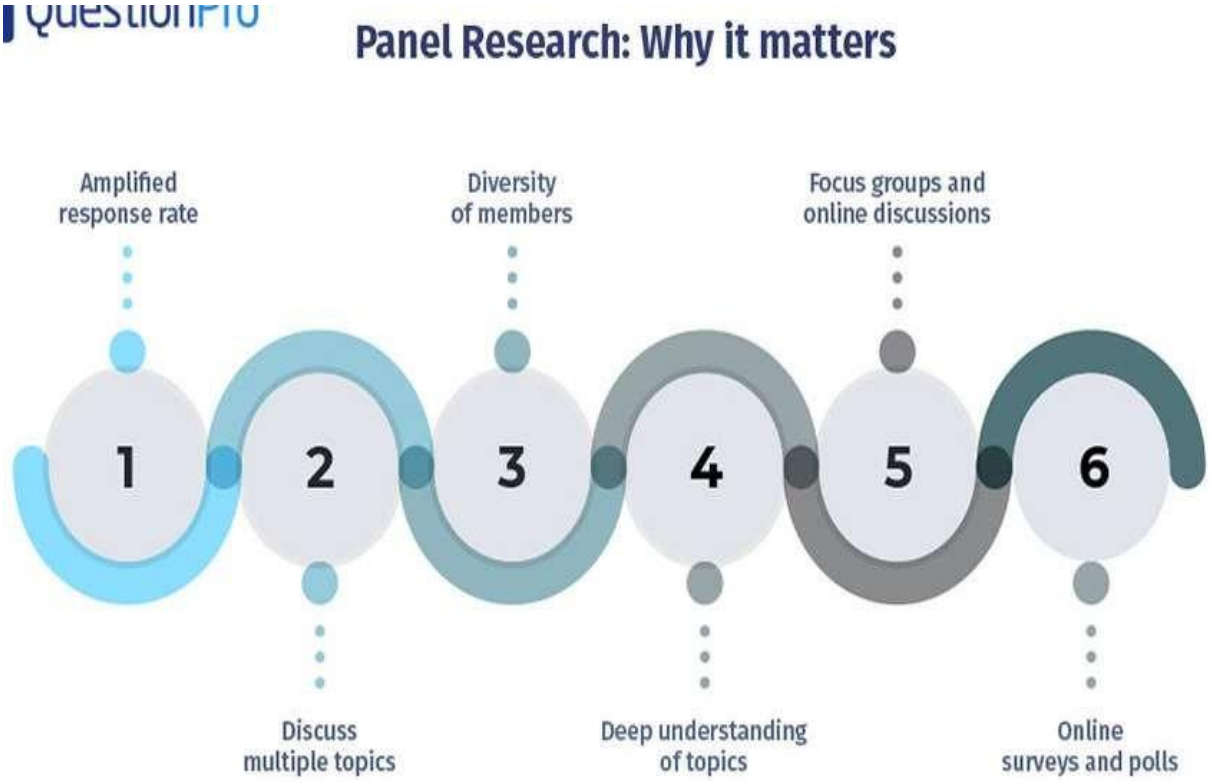
Case studies

A case study can be used to describe the characteristics of a specific subject (such as a person, group, event or organization). Instead of gathering a large volume of data to identify patterns across time or location, case studies gather detailed data to identify the characteristics of a narrowly defined subject.

Rather than aiming to describe generalizable facts, case studies often focus on unusual orinteresting cases that challenge assumptions, add complexity, or reveal something new about a research problem.

Panel Research

Panel research is a method for collecting data repeatedly, from a pre-recruited set of people. These individuals generally provide demographic, household and behavioral data, which can make conducting future studies easier. Technology, primarily the internet, has transformed panel research methodology by the ease of which we can access larger numbers of respondents. Panel research provides many advantages for companies including faster turnaround, higherparticipation rates, and cost savings. The quantitative data can provide companies with insights into pricing, effectiveness and sales projection of their products or brand.



Importance of Panel Research:

Building a quality research panel is very important because your data depends on it. With a research panel, you are able to build rich profiles of your members, which will help to ensure

that your reporting provides quality responses. If you are looking to launch a new mobile phone, you would want to target panelists who are interested in mobile phones and technology to yield more informed responses. Additionally, a well-managed panel of pre-recruited respondents allows for a faster response rate, as the participants have shown interest in participating in surveys by joining the panel.

Advantages of conducting panel research:

- The rate of research response is amplified as panel members have willingly signed up to participate in the research process.
- Different aspects of a particular subject can be discussed with panel members, unlike other research methods where a single topic needs to be discussed at a time. This makes panel research effective and less expensive.
- A panel consisting of a sizeable number of participants makes it easy for marketers to record behavioral changes across demographics due to the diversity of panel members.
- Capturing better details in panel research insights as panel members have a more sophisticated understanding of the research subject since they are profiled, screened and validated during recruitment.
- Qualitative market research methods such as focus groups, discussions, online interviews can be far more effective if they're conducted with well-recruited panel.
- Quantitative market research can be conducted to muster data and metrics-based inputs for survey research by sending out online surveys and online polls to a panel.

Experimental Research Designs

Definition:

Experimental research is research conducted with a scientific approach using two sets of variables. The first set acts as a constant, which you use to measure the differences of the second set. Quantitative research methods, for example, are experimental.

Any research conducted under scientifically acceptable conditions uses experimental methods. The success of experimental studies hinges on researchers confirming the change of a variable is based solely on the manipulation of the constant variable. The research should establish a notablecause and effect.

Types of experimental research design:

The classic experimental design definition is, "The methods used to collect data in experimental studies."

There are ***three primary types of experimental design***:

- Pre-experimental research design
- True experimental research design
- Quasi-experimental research design

The way you classify research subjects, based on conditions or groups, determines the type of design.

1. **Pre-experimental research design:** A group or various groups are kept under observation after implementing factors of cause and effect. You'll conduct this research to understandwhether further investigation is necessary for these particular groups.

You can break down pre-experimental research further in three types:

- One-shot Case Study Research Design
- One-group Pretest-posttest Research Design
- Static-group Comparison

2. True experimental research design: True experimental research relies on statistical analysis to prove or disprove a hypothesis, making it the most accurate form of research. Of the types of experimental design, only true design can establish a cause-effect relationship within a group. Ina true experiment, three factors need to be satisfied:

- There is a Control Group, which won't be subject to changes, and an ExperimentalGroup, which will experience the changed variables.
- A variable which can be manipulated by the researcher
- Random distribution

This experimental research method commonly occurs in the physical sciences.

3. Quasi-experimental research design: The word "Quasi" indicates similarity. A quasi-experimental design is similar to experimental, but it is not the same. The difference between the two is the assignment of a control group. In this research, an independent variable is manipulated, but the participants of a group are not randomly assigned. Quasi-research is used in field settings where random assignment is either irrelevant or not required.

Advantages of experimental research:

Experimental research allows you to test your idea in a controlled environment before taking it tomarket. It also provides the best method to test your theory, thanks to the following advantages:

- Researchers have a stronger hold over variables to obtain desired results.
- The subject or industry does not impact the effectiveness of experimental research. Anyindustry can implement it for research purposes.
- The results are specific.
- After analyzing the results, you can apply your findings to similar ideas or situations.
- You can identify the cause and effect of a hypothesis. Researchers can further analyzethis relationship to determine more in-depth ideas.
- Experimental research makes an ideal starting point. The data you collect is a foundationon which to build more ideas and conduct more research.

Whether you want to know how the public will react to a new product or if a certain foodincreases the chance of disease, experimental research is the best place to start.

205. BUSINESS RESEARCH METHODS

UNIT – III

Data sources: Secondary Vs Primary – Secondary data collection procedures- Primary Data Collection methods: Survey methods-Observational Studies.

Research data

Research data is any information that has been collected, observed, generated or created to validate original research findings. Although usually digital, research data also includes non-digital formats such as laboratory notebooks and diaries.

There are two types of data:

1. **Primary Data**
2. **Secondary Data**

1. Primary Data: It is a term for data collected at source. This type of information is obtained directly from first hand sources by means of surveys, observations and experimentation and not subjected to any processing or manipulation and also called primary data. Primary data means original data that has been collected specially for the purpose in mind. It means someone collected the data from the original source first hand.

Primary data has not been published yet and is more reliable, authentic and objective. For example population census conducted by the government of India after every 10 years.

2. Secondary data: It refers to the data collected by someone other than the user i.e. the data is already available and analysed by someone else. Common sources of secondary data include various published or unpublished data, books, magazines, newspaper, trade journals etc.

COLLECTION OF PRIMARY DATA:

Primary data is collected in the course of doing experimental or descriptive research by doing experiments, performing surveys or by observation or direct communication with respondents. Several methods for collecting primary data are given below-

1. Observation Method

It is commonly used in studies relating to behavioural science. Under this method observation becomes scientific tool and the method of data collection for the researcher, when it serves a formulated research purpose and is systematically planned and subjected to checks and controls.

(a) Structured (descriptive) and unstructured (exploratory) observation: When an observation is characterized by careful definition of units to be observed, style of observer, conditions of or observation and selection of pertinent data of observation it is a structured observation. When there characteristics are not thought of in advance or not present. it is a unstructured observation.

(b) Participant, Non-participant and disguised observation: When the observer observes by making himself more or less, the member of the group he is observing, it is participant observation but when the observer observes by detaching himself from the group under observation it is non-participant observation. If the observer observes in such manner that his presence is unknown to the people he is observing it is disguised observation.

(c) Controlled (laboratory) and uncontrolled(exploratory) observation: If the observation takes place in the natural setting it is a uncontrolled observation but when observation takes place according to some pre-arranged plans ,involving experimental procedure it is a controlled observation.

Advantages:

- Subjective bias is eliminated.
- Data is not affected by past behaviour or future intentions.
- Natural behaviour of the group can be recorded.

Limitations:

- Expensive methodology.
- Information provided is limited.
- Unforeseen factors may interfere with the observational task

2. INTERVIEW METHOD

This method of collecting data involves presentation of oral verbal stimuli and deeply in terms of oral- verbal responses. It can be achieved by two ways:-

A. **Personal interview:** It requires a person known as interviewer to ask questions generally in a face to face contact to the other person. It can be –

a. ***Direct personal investigation***- The interviewer has to collect the information personally from the services concerned.

b. ***Indirect oral examination***- The interviewer has to cross examine other persons who are supposed to have a knowledge about the problem.

c. **Structured interviews:** Interviews involving the use of pre-determined questions and of highly standard techniques of recording

d. **Unstructured interviews:** It does not follow a system of pre-determined questions and is characterized by flexibility of approach to questioning.

e. **Focused interview:** It is meant to focus attention on the given experience of the respondent and its effect. The interviewer may ask questions in any manner or sequence with the aim to explore reasons and motives of the respondent.

f. **Clinical interviews:** It is concerned with broad underlying feeling and motives or individuals life experience which are used as method to collect information under this method at the interviewer direction.

g. **Non directive interview:** The interviewer`s function is to encourage the respendent to talk about the given topic with a bare minimum of direct questioning.

Advantages:

- More information and in depth can be obtained.
- Samples can be controlled.
- There is greater flexibility under this method
- Personal information can as well be obtained.
- Mis-interpretation can be avoided by unstructured interview.

Limitations:

- It is an expensive method.
- More time consuming.
- Possibility of imaginary info and less frank responses.
- High skilled interviewer is required

B Telephonic interviews: It requires the interviewer to collect information by contacting respondents on telephone and asking questions or opinions orally.

2. QUESTIONNAIRE:

In this method a questionnaire is sent (mailed) to the concerned respondents who are expected to read, understand and reply on their own and return the questionnaire. It consists of a number of questions printed or typed in a definite order on a form or set of forms.

It is advisable to conduct a 'pilot study' which is the rehearsal of the main survey by experts for testing the questionnaire for weaknesses of the questions and techniques used.

Essential of a good questionnaire:

- ✓ It should be short and simple.
- ✓ Questions should process in a logical sequence.
- ✓ Technical terms and vague expressions must be avoided.
- ✓ Control questions to check the reliability of the respondent must be present.
- ✓ Adequate space for answers must be provided.
- ✓ Brief directions with regard to filling up of questionnaire must be provided.
- ✓ The physical appearances-quality of paper, colour etc. must be good to attract the attention of the respondent

Advantages:

- ❖ Free from bias of interviewer.
- ❖ Respondents have adequate time to give answers
- ❖ Respondents are easily and conveniently approachable
- ❖ Large samples can be used to be more reliable.

Limitations:

- ❖ Low rate of return of duly filled questionnaire.
- ❖ Control over questions is lost once it is sent.
- ❖ It is inflexible once it is sent.
- ❖ Possiblitty of ambiguous omission of replies.
- ❖ Time taking and slow process.

3. Schedules

This method of data collection is similar to questionnaire method with difference that schedule are being filled by the enumerations specially appointed for the purpose. Enumerations explain the aims and objects of the investigation and may remove any misunderstanding and help the respondents to record answer. Enumerations should be well trained to perform their job,he/she should be honest hardworking and patient. This type of data is helpful in extensive enquiries however it is very expensive.

COLLECTION OF SECONDARY DATA:

A researcher can obtain secondary data from various sources. Secondary data may either be published data or unpublished data.

Published data are available in:

A. Publications of government.
B. Technical and trade journals.
C. Reports of various businesses, banks etc.
D. Public records.
E. Stastistical or historical documents.

Unpublished data may be found in letters, diaries, unpublished biographies or work. Beforeusing secondary data it must be checked for the following characteristics-

1. Reliability of data:

- ✓ Who collected the data?
- ✓ From what source?
- ✓ Which method?
- ✓ Time?
- ✓ Possibility of bias?
- ✓ Accuracy?

2. Suitability of data: The object scope and nature of the original enquiry must be studies and then carefully scrutinize the data for suitability.

3. Adequency: The data is considered inadequate if the level of accuracy achieved in data is found inadequate or if they are related to an area which may be either narrower or wider than the area of the present enquiry.

205. BUSINESS RESEARCH METHODS

UNIT – IV

Sampling Designs: Probability and Non-probability sampling methods- Sample Size determination- Attitude Measurement and Measurement Scales.

PROBABILITY AND NON-PROBABILITY SAMPLING

A probability sampling is one in which every unit in the population has a chance (greater than zero) of being selected in the sample, and this probability can be accurately determined. The combination of these traits makes it possible to produce unbiased estimates of population totals, by weighting sampled units according to their probability of selection.

Example: We want to estimate the total income of adults living in a given street. We visit each household in that street, identify all adults living there, and randomly select one adult fromeach household. (For example, we can allocate each person a random number, generated froma uniform distribution between 0 and 1, and select the person with the highest number in each household). We then interview the selected person and find their income. People living on their own are certain to be selected, so we simply add their income to our estimate of the total. But a person living in a household of two adults has only a one-in-two chance of selection. To reflect this, when we come to such a household, we would count the selected person's income twice towards the total. (The person who is selected from that household can be loosely viewed as also representingthe person who isn't selected.)

In the above example, not everybody has the same probability of selection; what makes it a probability sample is the fact that each person's probability is known. When every element in the population *does* have the same probability of selection, this is known as an 'equal probability of selection' (EPS) design. Such designs are also referred to as 'self-weighting' because all sampled units are given the same weight.

Probability sampling includes: Simple Random Sampling, Systematic Sampling, Stratified Sampling, Probability Proportional to Size Sampling, and Cluster or Multistage Sampling. These various ways of probability sampling have two things in common:

1. Every element has a known nonzero probability of being sampled and
2. Involves random selection at some point.

Non Probability Sampling ; - Non Probability Sampling is any sampling method where some elements of the population have *no* chance of selection (these are sometimes referred to as 'out of coverage'/'under covered'), or where the probability of selection can't be accurately determined. It involves the selection of elements based on assumptions regarding the population of interest, which forms the criteria for selection. Hence, because the selection of elements is nonrandom, non probability sampling does not allow the estimation of sampling errors. These conditions give rise to exclusion bias, placing limits on how much information a sample can provide about the population. Information about the relationship between sample and population is limited, making it difficult to extrapolate from the sample to the population.

Example: We visit every household in a given street, and interview the first person to answer the door. In any household with more than one occupant, this is a nonprobability sample, because some people are more likely to answer the door (e.g. an unemployed person who spends most of their time at home is more likely to answer than an employed housemate who might be at work when the interviewer calls) and it's not practical to calculate these probabilities.

Non probability sampling methods include accidental sampling, quota sampling and purposive sampling. In addition, non-response effects may turn *any* probability designinto a non-probability design if the characteristics of non- response are not well understood, since non response effectively modifies each element's probability of being sampled.

Sampling Methods

Within any of the types of frame identified above, a variety of sampling methods can be employed, individually or in combination. Factors commonly influencing the choice between these designs include:

- Nature and quality of the frame
- Availability of auxiliary information about units on the frame
- Accuracy requirements, and the need to measure accuracy
- Whether detailed analysis of the sample is expected
- Cost/operational concerns

Simple Random Sampling

In a simple random sample (SRS) of a given size, all such subsets of the frame are given an equal probability. Each element of the frame thus has an equal probability of selection: the frame is not subdivided or partitioned. Furthermore, any given *pair* of elements has the same chance of selection as any other such pair (and similarly for triples, and so on). This minimizes bias and simplifies analysis of results. In particular, the variance between individual results within thesample is a good indicator of variance in the overall population, which makes it relatively easy to estimate the accuracy of results.

However, SRS can be vulnerable to sampling error because the randomness of the selection may result in a sample that doesn't reflect the makeup of the population. For instance, a simple random sample of ten people from a given country will *on average* produce five men and five women, but any given trial is likely to over represent one sex and under represent the other. Systematic and stratified techniques, discussed below, attempt to overcome this problem by using information about the population to choose a more representative sample.

SRS may also be cumbersome and tedious when sampling from an unusually large target population. In some cases, investigators are interested in research questions specific to subgroups of the population. For example, researchers might be interested in examining whether cognitive ability as a predictor of job performance is equally applicable across racial groups. SRS cannot accommodate the needs of researchers in this situation because it does not provide subsamples ofthe population. Stratified sampling, which is discussed below, addresses this weakness of SRS.

Simple random sampling is always an EPS design (equal probability of selection), but not all EPS designs are simple random sampling.

Systematic Sampling

Systematic sampling relies on arranging the study population according to some ordering scheme and then selecting elements at regular intervals through that ordered list. Systematic sampling involves a random start and then proceeds with the selection of every *k*th element from then onwards. In this case, k = (population size/sample size). It is important that the starting point is not automatically the first in the list, but is instead randomly chosen from within the first to the *k*th

element in the list. A simple example would be to select every 10th name from the telephone directory (an 'every 10th' sample, also referred to as 'sampling with a skip of 10').

As long as the starting point is randomized, systematic sampling is a type of probability sampling. It is easy to implement and the stratification induced can make it efficient, *if* the variable by which the list is ordered is correlated with the variable of interest. 'Every 10th' sampling is especially useful for efficient sampling from databases.

For example, suppose we wish to sample people from a long street that starts in a poor area (house No. 1) and ends in an expensive district (house No. 1000). A simple random selection of addresses from this street could easily end up with too many from the high end and too few from the low end (or vice versa), leading to an unrepresentative sample. Selecting (e.g.) every 10th street number along the street ensures that the sample is spread evenly along the length of the street, representing all of these districts. (Note that if we always start at house #1 and end at #991, the sample is slightly biased towards the low end; by randomly selecting the start between #1 and #10, this bias is eliminated.

However, systematic sampling is especially vulnerable to periodicities in the list. If periodicity is present and the period is a multiple or factor of the interval used, the sample is especially likely to be *un*representative of the overall population, making the scheme less accurate than simple random sampling.

For example, consider a street where the odd-numbered houses are all on the north (expensive) side of the road, and the even-numbered houses are all on the south (cheap) side. Under the sampling scheme given above, it is impossible to get a representative sample; either the houses sampled will *all* be from the odd-numbered, expensive side, or they will *all* be from the even- numbered, cheap side.

Another drawback of systematic sampling is that even in scenarios where it is more accurate than SRS, its theoretical properties make it difficult to *quantify* that accuracy. (In the two examples of systematic sampling that are given above, much of the potential sampling error is due to variation between neighbouring houses - but because this method never selects two neighbouring houses, the sample will not give us any information on that variation.)

As described above, systematic sampling is an EPS method, because all elements have the same probability of selection (in the example given, one in ten). It is *not* 'simple random sampling' because different subsets of the same size have different selection probabilities - e.g. the set {4, 14, 24,..., 994} has a one-in-ten probability of selection, but the set {4,13,24,34,...} has zero probability of selection.

Systematic sampling can also be adapted to a non-EPS approach; for an example, see discussion of PPS samples below.

Stratified Sampling

Where the population embraces a number of distinct categories, the frame can be organized by these categories into separate "strata." Each stratum is then sampled as an independent sub- population, out of which individual elements can be randomly selected. There are several potential benefits to stratified sampling.

First, dividing the population into distinct, independent strata can enable researchers to draw inferences about specific subgroups that may be lost in a more generalized random sample.

Second, utilizing a stratified sampling method can lead to more efficient statistical estimates (provided that strata are selected based upon relevance to the criterion in question, instead of availability of the samples). Even if a stratified sampling approach does not lead to increased statistical efficiency, such a tactic will not result in less efficiency than would simple random sampling, provided that each stratum is proportional to the group's size in the population.

Third, it is sometimes the case that data are more readily available for individual, pre-existing strata within a population than for the overall population; in such cases, using a stratified sampling approach may be more convenient than aggregating data across groups (though this may potentially be at odds with the previously noted importance of utilizing criterion-relevant strata).

Finally, since each stratum is treated as an independent population, different sampling approaches can be applied to different strata, potentially enabling researchers to use the approach best suited (or most cost-effective) for each identified subgroup within the population.

There are, however, some potential drawbacks to using stratified sampling. First, identifying strata and implementing such an approach can increase the cost and complexity of sample selection, as well as leading to increased complexity of population estimates. Second, when examining multiple criteria, stratifying variables may be related to some, but not to others, further complicating the design, and potentially reducing the utility of the strata. Finally, in some cases (such as designs with a large number of strata, or those with a specified minimum sample size per group), stratified sampling can potentially require a largersample than would other methods (although in most cases, the required sample size would be no larger than would be required for simple random sampling.

A stratified sampling approach is most effective when three conditions are met

1. Variability within strata are minimized
2. Variability between strata are maximized
3. The variables upon which the population is stratified are strongly correlated with the desired dependent variable.

Advantages over other sampling methods

1. Focuses on important subpopulations and ignores irrelevant ones.
2. Allows use of different sampling techniques for different subpopulations.
3. Improves the accuracy/efficiency of estimation.
4. Permits greater balancing of statistical power of tests of differences between strata by sampling equal numbers from strata varying widely in size.

Disadvantages

1. Requires selection of relevant stratification variables which can be difficult.
2. Is not useful when there are no homogeneous subgroups.
3. Can be expensive to implement.

Post stratification

Stratification is sometimes introduced after the sampling phase in a process called "post stratification". This approach is typically implemented due to a lack of prior knowledge of an appropriate stratifying variable or when the experimenter lacks the necessary information to create a stratifying variable during the sampling phase. Although the method is susceptible to the pitfalls of post hoc approaches, it can provide several benefits in the right situation. Implementation usually follows a simple random sample. In addition to allowing for stratification on an ancillary variable, post stratification can be used to implement weighting, which can improve the precision of a sample's estimates.

Oversampling

Choice-based sampling is one of the stratified sampling strategies. In choice-based sampling, the data are stratified on the target and a sample is taken from each stratum so that the rare target class will be more represented in the sample. The model is then built on thisbiased sample. The effects of the input variables on the target are often estimated with more precision with the choice-based sample even when a smaller overall sample size is taken compared to a random sample. The results usually must be adjusted to correct for the oversampling.

Probability-Proportional-To-Size Sampling

In some cases the sample designer has access to an "auxiliary variable" or "size measure", believed to be correlated to the variable of interest, for each element in the population. These data can be used to improve accuracy in sample design. One option is to use the auxiliary variable as a basis for stratification, as discussed above.

Another option is probability-proportional-to-size ('PPS') sampling, in which the selection probability for each element is set to be proportional to its size measure, up to a maximum of 1. In a simple PPS design, these selection probabilities can then be used as the basis for Poisson sampling. However, this has the drawback of variable sample size, and

different portions of the population may still be over- or under-represented due to chance variation in selections. To address this problem, PPS may be combined with a systematic approach.

Example: Suppose we have six schools with populations of 150, 180, 200, 220,260, and 490 students respectively (total 1500 students), and we want to use student population as the basis for a PPS sample of size three. To do this, we could allocate the first school numbers 1 to 150, the second school 151 to 330 (= 150 + 180), the third school 331 to 530, and so on to the last school (1011 to 1500). We then generate a random start between 1 and 500 (equal to 1500/3) and count through the school populations by multiples of 500. Ifour random start was 137, we would select the schools which have been allocated numbers 137, 637, and 1137, i.e. the first, fourth, and sixth schools.

The PPS approach can improve accuracy for a given sample sizes by concentrating sample on large elements that have the greatest impact on population estimates. PPS sampling is commonly used for surveys of businesses, where element size varies greatly and auxiliary information is often available - for instance, a survey attempting to measure the number of guest-nights spent in hotels might use each hotel's number of rooms as an auxiliary variable. In some cases, an older measurement of the variable of interest can be used as an auxiliary variable when attempting to produce more current estimates.

Cluster Sampling

Sometimes it is more cost-effective to select respondents in groups ('clusters'). Sampling is often clustered by geography, or by time periods. (Nearly all samples are in some sense 'clustered' in time - although this is rarely taken into account in the analysis.) For instance, if surveying households within a city, we might choose to select 100 city blocksand then interview every household within the selected blocks.

Clustering can reduce travel and administrative costs. In the example above, an interviewer can make a single trip to visit several households in one block, rather than having to drive to a different block for each household.

It also means that one does not need a sampling frame listing all elements in thetarget population. Instead, clusters can be chosen from a cluster-level frame, with an element- level frame created only for the selected clusters. In the example above, the sample only requires a block-level city map for initial selections, and then a household-level map of the 100 selected blocks, rather than a household-level map of the whole city.

Cluster sampling generally increases the variability of sample estimates above that of simple random sampling, depending on how the clusters differ between themselves, as compared with the within-cluster variation. For this reason, cluster sampling requires a larger sample than SRS to achieve the same level of accuracy - but cost savings from clustering might still make this a cheaper option.

Cluster sampling is commonly implemented as multistage sampling. This is a complex form of cluster sampling in which two or more levels of units are embedded one in the other. The first stage consists of constructing the clusters that will be used to sample from.In the second stage, a sample of primary units is randomly selected from each cluster (rather than using all units contained in all selected clusters). In following stages, in each ofthose selected clusters, additional samples of units are selected, and so on. All ultimate units (individuals, for instance) selected at the last step of this procedure are then surveyed. This technique, thus, is essentially the process of taking random subsamples of preceding random samples.

Multistage sampling can substantially reduce sampling costs, where the complete population list would need to be constructed (before other sampling methods could be applied). By eliminating the work involved in describing clusters that are not selected, multistage sampling can reduce the large costs associated with traditional cluster sampling.

Quota Sampling

In quota sampling, the population is first segmented into mutually exclusive sub-groups, just as in stratified sampling. Then judgement is used to select the subjects or units from each segment based on a specified proportion. For example, an interviewer may be told to sample 200 females and 300 males between the age of 45 and 60.

It is this second step which makes the technique one of non-probability sampling. In quota sampling the selection of the sample is non-random. For example interviewers might be

tempted to interview those who look most helpful. The problem is that these samples may be biased because not everyone gets a chance of selection. This random element is its greatest weakness and quota versus probability has been a matter of controversy for several years.

Accidental Sampling

Accidental sampling (sometimes known as grab, convenience or opportunity sampling) is a type of non probability sampling which involves the sample being drawn from that part of the population which is close to hand. That is, a population is selected because it isreadily available and convenient. It may be through meeting the person or including a person in the sample when one meets them or chosen by finding them through technological means such as the internet or through phone. The researcher using such a sample cannot scientifically make generalizations about the total population from this sample because it would not be representative enough. For example, if the interviewer were to conduct such a survey at a shopping center early in the morning on a given day, the people that he/she could interview would be limited to those given there at that given time, which would not represent the views of other members of society in such an area, if the survey

were to be conducted at different times of day and several times per week. This type of sampling is most useful for pilot testing. Several important considerations for researchers using convenience samples include:

1. Are there controls within the research design or experiment which can serve to lessen the impact of a non-random convenience sample, thereby ensuring the results will be more representative of the population?
2. Is there good reason to believe that a particular convenience sample would or should respond or behave differently than a random sample from the same population?
3. Is the question being asked by the research one that can adequately be answeredusing a convenience sample?

In social science research, snowball sampling is a similar technique, where existing study subjects are used to recruit more subjects into the sample. Some variants of snowball sampling, such as respondent driven sampling, allow calculation of selection probabilities and are probability sampling methods under certain conditions.

4. Duplication.

SAMPLE SIZE

Sample size is the number of items to be selected from the universe. It should be optimum. Formulas, tables, and power function charts are well known approaches to determine sample size.

Steps for Using Sample Size Tables

1. Postulate the effect size of interest, α, and β.
2. Check sample size table
    1. Select the table corresponding to the selected α
    2. Locate the row corresponding to the desired power
    3. Locate the column corresponding to the estimated effect size.
    4. The intersection of the column and row is the minimum sample size required.

The Factors Considering While Deciding the Size of the Sample

a. Nature of the population.
b. Complexity of tabulation.
c. Problem relate with collection of data.
d. Type of sampling.
e. Basic information.
f. Degree of accuracy required for the study.

Characteristics of Good Sample Design

a. Representative.
b. Viable.
c. The selected sample design should not cause more errors.
d. A good sample design able to control systematic bias efficiently.
e. If the sample is well design and selected, decision makers can use this info with confidence.

Criteria of Selecting a Sampling Procedure

1. Nature of the problem.
2. Goal of researchers.

3. Geographical area covered by the survey.
4. Size of the population under study.
5. Extent of fact available about population.
6. Availability of funds
7. Available time for study.
8. Desired reliability of the result.

Criteria Used For Selecting Sampling Techniques

- The purpose of the survey.
- Measurability.
- Degree of precision.
- Information about population.
- The nature of the population.
- The geographical area covered by the survey.
- Fund availability.
- Time.
- Economy.

Errors in Sample Surveys

Survey results are typically subject to some error. Total errors can be classified into sampling errors and non-sampling errors. The term "error" here includes systematic biases as well as random errors.

Sampling Errors and Biases

Sampling errors and biases are induced by the sample design. They include:

1. **Selection Bias:** When the true selection probabilities differ from those assumed in calculating the results.
2. **Random Sampling Error:** Random variation in the results due to the elements in the sample being selected at random.

Sampling Bias

Sampling analysis involve to type of cost namely cost of collecting data and cost of an incorrect inference resulting from the data. They are to causes for incorrect inference resulting from data. They are

I. Systematic bias
II. Sampling errors

Causes of systematic bias

- Unsuitable sample frame or source list.
- Faulty measuring device.
- Non respondent
- Indeterminacy principle.
- Usual bias in reporting data.

Sampling errors

The errors which arise due to the use of sampling survey are known as sampling errors. These are random variation in the sample estimate around the true population parameters.

Type of sampling errors

Biased errors: These errors are occurring due to the faulty selection of sampling method due to the prejudice of the researchers.

Unbiased errors: This type of bias is occurring due to chance difference between the items included in the sample.

Causes of bias

Bias may arise due to,

1. Faulty process selection.
2. Faulty work during the collection of information.
3. Faulty method of analysis.

Non-Sampling Error

Non-sampling errors are other errors which can impact the final survey estimates, caused by problems in data collection, processing, or sample design. They include:

1. **Over coverage**: Inclusion of data from outside of the population.
2. **Under coverage**: Sampling frame does not include elements in the population.
3. **Measurement error**: e.g. when respondents misunderstand a question, or find it difficult to answer.
4. **Processing error**: Mistakes in data coding.
5. **Non-response**: Failure to obtain complete data from all selected individuals.

After sampling, a review should be held of the exact process followed in sampling, rather than that intended, in order to study any effects that any divergences might have on subsequent analysis. A particular problem is that of non-response.

Two major types of non-response exist: unit non-response (referring to lack of completion of any part of the survey) and item non-response (submission or participation in survey but failing to complete one or more components/questions of the survey). In survey sampling, many of the individuals identified as part of the sample may be unwilling to participate, not have the time to participate (opportunity cost), or survey administrators may not have been able to contact them. In this case, there is a risk of differences, between respondents and non-respondents, leading to biased estimates of population parameters. Thisis often addressed by improving survey design, offering incentives, and conducting follow-up studies which make a repeated attempt to contact the unresponsive and to characterize their similarities and differences with the rest of the frame. The effects can also be mitigated by weighting the data when population benchmarks are available or by imputing data based on answers to other questions.

Non-response is particularly a problem in internet sampling. Reasons for this problem include improperly designed surveys,] over-surveying (or survey fatigue), and the fact that potential participants hold multiple e-mail addresses, which they don't use anymore or don't check regularly.

MEASUREMENT AND SCALING

The word scale or scaling is generally used for indicating measurements or measuring something. Many aspects of social phenomena like emotion, attitude, faiths etc. are not measurable directly. They are not quantitative in nature. In social phenomena there are two types of variables- quantitative and qualitative. Measurement or scaling implies conversion of qualitative data into quantitative data and then measuring them. Various kinds of statistical measurements are used for this purpose. Scaling is an attempt to bring about greater accuracy which is desired in both physical and social sciences.

Essentials of Scaling (Criteria for Good Scaling)

- Continuum means judging the scalability of the phenomenon under study.
- Reliability means that it should consistently produce the same result when applied to the same design.
- Validity implies correct measurement. A scale is valid if it measures correctly what is expected to measure.
- Weighting items means proper weights are to be provided to the attributes involved in the study because they are not of equal importance.
- Equality of units is a desirable characteristic but not essential for sound scientific procedure.
- In order to make the units equal, sometimes subtraction or addition can be made.

Measurement

Measurement is a systematic way of assigning numbers or names to objects and their attributes. It is easy to assign numbers in respect of properties of some objects, properties like weight, height.

Why Measurement?

"When you can measure what you are talking about and express it in numbers, you know something about it" Lord Kelvin

Rules of Measurement

- A rule is a guide which instructs us to do.
- Operational definitions help us to specify the rules for assigning the numbers.

Concept	Conceptual definition	Operational definition	Measurement
Job Challenge	Reflects a worker's desire for stimulation and challenge in his job and ability to exercise skills	I have an opportunity to develop my own special abilities I am given a chance to do things I do best I love my work Interval between entering the door and receiving receipt.	Very true/ somewhat true/ not very true/ not at all true
Shopping time	Time spent in a shop	Interval between entering the door and receiving receipt	Time in minutes

Scales

Scales are devised for measuring variable in social science research. Scaling is the procedure for determining the quantitative measure of abstract concepts like leadership style, brand image of product etc. It is therefore -

- Any series of items which is progressively arranged according to value or magnitude into which an item can be placed according to its quantification.
- A continuous spectrum or series of categories
- Used to represent, usually quantitatively, an item's or person's place in that spectrum

Scale Classification

(A) Level of Measurement

i. Nominal scales
ii. Ordinal scales
iii. Interval scales

iv. Ratio scale

Nominal Scale

Numbers or letters assigned to objects which serve as labels for identification or classification.

They are Scales "in name only"

Ordinal Scale

Arranges objects or alternatives according to their magnitude in an ordered relationship. eg rating career opportunities as excellent, good, average poor or very poor.

Interval Scale

Interval scales not only indicate order, they also measure order or distance in units of equal intervals. When an interval scale is used to measure psychological attribute, the researcher can comment on the magnitude of differences or compare the average differences on attributes that are measured, but cannot determine the actual strength of the attitude towards an object

Ratio Scale

A ratio scale has equal interval properties of an interval scale but has two additional features: Has a true zero. Possible to indicate the complete absence of a property.

The numerals of the ratio scale have the qualities of real numbers and can be added, subtracted, multiplied, Divided and expressed in ratio relationships.

Analysis of Scales

Type of scale	Numerical Operation	Descriptive statistics
Nominal	Counting	Frequency in each category Percentage in each category Mode
Ordinal	Rank ordering	Median Range Percentile ranking

Interval	Arithmetic operations on intervals between numbers	Mean Standard Deviation Variance
Ratio	Arithmetic operations on actual quantities	Geometric mean Coefficient of variation

Lickert Scale

Respondents indicate their attitude by checking how strongly they agree or disagree with carefully constructed statements that range from the very positive to the very negative towards the attitudinal object. Individuals generally choose from five alternatives: strongly agree, agree, uncertain, disagree and strongly disagree

Example of Lickert scale:

1) If the price of raw materials fell firms would reduce the price of their food product.	SA	A	N	DA	SD
2) Without Government regulation the firms would exploit the customers.					
3) Most food companies are so concerned about making profit they do not care about quality.					
4) The food industry spends a great deal of money making sure that its manufacturing is hygienic.					

Semantic Differential

An attitude measure consisting of a series of seven –point bipolar rating scales allowing response to a concept. Bi Polar adjectives such as “good and bad”, “clean or dirty” anchor the beginning and end poles of the scale.

Numerical Scales

An attitude rating scale similar to a semantic differential except that it uses numbers as response options to identify response positions instead of verbal descriptions. Usually five point scale or seven point scale

Constant Sum Scale

A measure of attitudes in which the respondents are asked to divide a constant sum to indicate the relative importance of the attributes

Stapel Scale

Measures the direction and intensity of an attitude simultaneously. E.g. To measure attitudes towards a supervisor

Supervisor's name

+3

+2

+1

Supportive

-1

-2

-3

Scaling Techniques

Some of the important scaling techniques are as follows:

- **Rating Scale:** It means measuring an attribute by judgement in a continuum. For eg. Students are rated by their teachers. In rating three elements are taken into consideration namely judges, continuum and subjects. Judges must be impartial. Continuum must accommodate the attributes realistically. Subjects should be amenable to rating. Rating scales are generally used for measuring the attitudes andthe intensity of attitudes.

- **Ranking Scales:** Ranking scales are identical to rating scales. In ranking scales, we make relative the score will place judgements against other similar objects. There are two generally used approaches of ranking scales namely.

    1. **Method of Paired Comparison** where the respondent can express his attitude by making a choice between two objects.

    2. **Method of Ranking Order** where the respondents are asked to rank their choices.

- **Attitude Scales:** In this type of scale, the attitude of an individual towards a matter can be known from the score of his responses given on a questionnaire. The score will place him in a scale. He simply expresses his likes or dislikes, agreement or disagreement with the issue involved as given in the forms of questions. On the basis of reply, he is assigned a score which indicates his position. In the attitude scale some relevant statements are to be considered by the respondents. The statements are found in such a way as to be intimately related to the attribute which is sought to be measured.

- **Factor Scaling:** This is a type of scaling in which multi dimensions of a complex attitude is identified.

205. BUSINESS RESEARCH METHODS

UNIT – V

Preliminary steps of data analysis: editing, coding and tabulation- Hypothesis testing – Tests of Significance for small and large samples- t-test, z-test, chi-square test, ANOVA: one-way,two-way - Presenting Results: Written and Oral Reports

EDITING, CODING AND TABULATION

Data continues to be in raw form, unless and until they are processed and analyzed. Processing is a statistical method by which the collected data is so organized the further analysis and interpretation of data become easy. It is an intermediary stage between the collection of data and their analysis and interpretation.

Processing stages

There are four important stages in the processing of data. They are;

1. Editing
2. Coding
3. Tabulation

Editing

As soon as the researcher receives the data, he should screen it for accuracy. Editing is the process of examining the data collected through various methods to detect errors and omissions and correct them for further analysis. Though editing, it is ensured that the collected data are accurate, consistent with other facts gathered, uniformly entered and well- arranged so that further analysis is made easier

Practical guidelines for editing

While editing care has to be taken to see that the data are as accurate and complete as possible. The following points are to be noted;

1. The editor should familiarize with the copy of instructions given to the interviewers.

2. The original entry, if found incorrect, should not be destroyed or erased. On the other hand, it should be crossed out in such a manner that it is still eligible.
3. Any, modification to the original entry by the editor must be specifically indicated.
4. All completed schedules must bear signature of the editor an d the date.
5. Incorrect answer to the questions can be corrected only if the editor is absolutely sure of the answer, otherwise leave it as such.
6. Inconsistent, incomplete or missing answers should not be used.
7. Sere that all numerical answers are converted to same units.

Coding

Coding is the process by which r response categories are summarized by numerals or other symbols to carry out subsequent operations of data analysis. This process of assigning numerals or symbols to the responses is called coding. It facilitates efficient analysis of the collected data and helps in reducing several replies to a small number of classes which contain the critical information required for analysis. In general it reduces the huge amount of information collected in to a form that is amenable to analysis.

Steps in coding

1. Study the answers carefully.
2. Develop a coding frame by listing the answers and by aligning codes to each of them.
3. Prepare a coding manual with the detail of variable names, codes and instructions.
4. If the coding manual has already been prepared before the collection of the data, make the required additions for the open ended and partially coded questions.

Coding rules

1. Give each respondent a code number for identification.
2. Provide code number for each question.
3. All responses including „don‟t know‟, „no opinion‟. Etc is to be coded.
4. Assign additional codes to partially coded questions.

Tabulation

Tabulation is the next step to classification. It is an orderly arrangement of data in rows and columns. It is defined as the "measurement of data in columns and rows". Data presented in tabular form is much easier to read and understand than the data presented in the text the main purpose of tabulation is to prepare the data for final analysis. It is a stage between classification of data and final analysis.

Objectives of Tabulation

1. To clarify the purpose of enquiry
2. To make the significance of data clear.
3. To express the data in least possible space.
4. To enable comparative study.
5. To eliminate unnecessary data
6. To help in further analysis of the data.

Types of Tables

Simple Table

- Here the data are presented only for one variable or characteristic. Any frequency distribution of a single variable is simple table

Complex table

- In complex table, two or more characteristics are shown. If the study is related to more than two variables, it is called multivariate analysis. They may be of the following tables.

a. One- way table

In this type of table, data of only one characteristic will be shown. It means that when one type of information is secured about different groups or individuals, it can be displayed with the help of one- way table

b. Two- way table

When mutually related attributes of a phenomenon are to be displayed, two way tables are used. In other words, this table shows two types of characteristics.

c. Three-way table

It displays three types of attributes. It is used when three inter- related or mutually

related attributes or characteristics of a phenomenon are to be displayed,.

d. Manifold tables

When information about different mutually attributes or characteristics of a phenomenon are to be displayed, manifold table is used. Such tables display information about various characteristics or attributes.

Parts of a statistical table

Following are the important parts of a statistical table.

a. Title of the table

The title of the table is placed above the table. If there are more than one table in a research, each should bear a number for easy reference.

b. Caption or title of the column

It is also termed as "box head". There may be sub- captions under the main caption.

c. Stub (row heading)

Stub refers to the title given to rows

d. Body (main data)

This is the main body of information needed for the research work.

e. End note (foot note)

This is placed below the table to convey the expansions of abbreviations to caption, stub or main body.

f. Source note

If the table is based on outside information, it should be mentioned in the source note below.

HYPOTHESIS

Hypothesis is a tentative statement showing the relationship between two or more variables, the reliability and validity of which is to be tested and verified. It expresses the nature and degree of relationship between variables. Hypotheses are -

- Assumptions
- Tentative statements
- Propositions
- Answering the questions
- Proposed solution to a problem
- Statements which are to be tested
- To be accepted of rejected
- To be verified empirically on the basis of sample

Why Hypothesis

- Gives the direction of research
- Specifies the sources of data
- Determines the data needs
- Type of research
- Appropriate techniques of research
- Contributes to the development of theory

Role of Hypothesis

- It guides the direction of the study
- It identifies facts that are relevant and those that are not
- It suggests which form of research design is likely to be most appropriate
- It provides a frame work for organising the conclusions that result

Sources of Hypothesis

- Observation –based on the behavior pattern
- Relation between price and demand is hypothesized,
- The sales and ad may be hypothesized
- Analogies casual observations in nature
- Poor people buy more lottery
- Intuitions and personal experiences –
- The story of Newton and falling of apple,
- The wisdom of Budha under the banyan tree
- A sparking in our mind at particular occasions
- Findings of studies
- State of Knowledge – the theorems may be modified
- Culture –castes, beliefs, habits, behaviour
- Contribution of research – the rejection of certain hypothesis may lead to further research
- Theory –large concerns earn more profit, return on capital is an index of business success

Different Types of Hypothesis

a. **Descriptive Hypothesis** – Describing the characteristics of a variable (may be an object, person, organisation, event, and situation)

   Eg. Employment opportunity of commerce graduates is more than the arts students.

b. **Relational Hypothesis** – Establishes relationship between two variables. It may be positive, negative or nil relationship.

Eg. High income leads to high savings

c. **Causal Hypothesis** – The change in one variable leads to change in another variable i.e. Dependent and independent variables, one variable is a cause and the other one is the effect

d. **Statistical Hypothesis** – association or difference between two variables are hypothesized

e. **Null Hypothesis** – it points out there is no difference between two populations in respect of same property.

f. **Alternative Hypothesis-** when we reject the null hypothesis, we accept another hypothesis known as alternate hypothesis.

g. Working Hypothesis

h. Complex Hypothesis

i. How to test

- State the two hypotheses - null and alternative
- Decide the test statistic t, Z, F, Chi-square
- Fix the level of significance
- Make the computations
- Take the decision
- Type 1 error and Type 11 error
- Degree of freedom (based on probability, distribution)

RESEARCH REPORT

A report is a detailed description of what has been done and how it has been done with respect to a particular area or topic. The purpose of the written report is to present the results of your research, but more importantly to provide a persuasive argument to readers of what you have found. It is the end product of a research activity. It is highly skilled work it is the final stage of the research work.

Need For Research Report

- The aim of research is the search for knowledge.
- A research report is needed to evolve a theory or a principle.
- Reporting is a process through which a basic ground is prepared for exchange of ideas or thoughts.
- Reporting helps the researcher to make specific recommendation for course of action.
- The research ability of a candidate is revealed through the final report he presents.
- A research report is highly useful for policy formulators, practitioners, general public and others.

Functions of Research Report

- Research report serves as a means for presenting the problem studied, methods and techniques used, findings, conclusions and recommendation in an organised manner.
- It serves as a reference material for future use in the same or related area.
- It serves as a means for judging the quality of the research project.
- It is a means for evaluating research ability.
- It provides systematic knowledge on problems and issues analysed.

Qualities of a Good Report

- Clarity
- Continuity
- Consistency
- Brevity

Tuesday 12/12/17

1) A random Sample of 400 flower stems has an average length of 10cm. can this be regarded as a Sample drawn from a large population with a mean of 10.2cm and standard deviation of 2.25cm

Sol Given that $n = 400$;

mean of the Sample $\bar{x} = 10cm$

Mean of the population $\mu = 10.2cm$

Standard Deviation $\sigma = 2.25cm$

H_0: The Sample has been drawn from the Normal population with mean $\mu = 10.2cm$.

H_1: The Sample has not been drawn from the Normal population. $\mu \neq 10.2cm$.

Hence it is decided to conduct two tail test

$H_0: \mu = 10.2$

$H_1: \mu \neq 10.2$

Two tailed test ని ఎందుకు ఉపయోగిస్తారు

Computation of Test statistic

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{10 - 10.2}{2.25/\sqrt{400}} = \frac{-0.2}{\frac{2.25}{20}} = \frac{-0.2}{0.1125} = -1.7778$$

$|Z| = 1.77$

Assuming Level of Significance as 5%, we have critical value of Z for Two-Tail Test $\Rightarrow Z_{cri} = 1.96$ ✓

$Z_{cal}(1.77) < Z_{cri}(1.96)$

So, Accept null Hypothesis.

Hence the Sample can be assumed to be drawn from the Normal population with mean = 10.2cm and standard Deviation 2.25cm

2) The mean life-time of a Sample of 100 fluorescent light tubes produced by a company is computed to be 1570 hours with a standard Deviation of 120 hrs. The company claims that the Average life of the tubes produced by the company is 1600 hrs. using the level of significance of

0.05 , is the claim Acceptable

Sol Given that $n = 100$

Mean of the Sample $\bar{x} = 1570$ hrs

Mean of the population $\mu = 1600$ hrs

Standard Deviation $\sigma = 120$ hrs.

H_0: The Sample has been drawn from the company with $\mu = 1600$ hrs.

H_1: The Sample has not been drawn from the company $\mu \neq 1600$ hrs.

$H_0: \mu = 1600$

$H_1: \mu \neq 1600$.

Hence to conduct large sample Test $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$= \dfrac{1570 - 1600}{120/\sqrt{100}}$

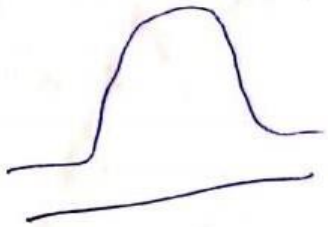
$= \dfrac{-30}{\frac{120}{10}} = \dfrac{-30}{12} = -2.5$

$|z| = 2.5$

At 0.05 Level of Significance $Z_{cri} = 1.96$ ✓

We observe $Z_{cal}(2.5) > Z_{cri}(1.96)$.

So H_0 Rejected.



3) Suppose that 100 Tyres of a certain brand lasted on the average 21,431 miles with a Standard Deviation of 1295 miles using $\alpha = 0.05$, Test the Null hypothesis $\mu = 22000$ miles against alternative Hypotheses $\mu < 22000$ miles

Sol Given that $H_0: \mu = 22000$

$H_1: \mu < 22000$

$\bar{x} = 21431$

$\sigma = 1295$

Hence Test statistic $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$= \dfrac{21431 - 22000}{1295/\sqrt{100}} = \dfrac{-569}{129.5} = -4.39$

$|z| > 4.39$ $Z_{cal} = -4.39$

$Z_{cri} = -1.645$

We observe $Z_{cal}(-4.39) < Z_{cri}(-1.645)$, So H_0 Accepted.

This test left tailed test left tailed test & modulus of 1.96 also

4) The means of two large samples of sizes 1000 and 2000 are 67.5 and 68.0 respectively. Test the quality of means of the two populations each with a standard deviation of 2.5

Sol $n_1 = 1000$; $n_2 = 2000$

$\bar{x}_1 = 67.5$; $\bar{x}_2 = 68.0$; $\sigma_1 = \sigma_2 = 2.5 \Rightarrow \sigma$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

Hence decided Two Tailed Test

Test Statistic $\Rightarrow$ $$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{67.5 - 68}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$$= \frac{-0.5}{2.5\sqrt{0.0010 + 0.0005}}$$

$$= \frac{-0.5}{2.5(0.0387)} \Rightarrow \frac{-0.5}{0.0968}$$

$|Z| = 5.16$ $\qquad Z = -5.16$

At 5% level of significance $Z_{cri} = 1.96$

$Z_{cal}(5.16) > Z_{cri}(1.96)$

Hence Rejected

left tailed test

t- Test Small Sample Test $n < 12$

1) In a 12 Test runs over a marked course, a newly designed motor boat averages $\bar{x} = 33.6$ secs with a standard deviation of 2.3 secs. Assuming that it is reasonable to treat the data as a random sample from a Normal population, Test the Null hypothesis $\mu = 35$ against Alternative hypothesis $\mu < 35$ at the level of significance $\alpha = 0.05$

Sol Given that $H_0: \mu = 35$ secs ; $\bar{x} = 33.6$; $n = 12$

$H_1: \mu < 35$ secs ; $\sigma = 2.3$; $\mu = 35$

Test statistics

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{33.6 - 35}{2.3/\sqrt{12}} \Rightarrow \frac{-1.4}{2.3/3.4641} = \frac{-1.4}{0.6640} = -2.12$$

$t_{cal} \; -2.12$

The critical Region is in the left Tail

[ఈ test లో < ఉన్నది కాబట్టి left tailed test తీసుకోవాలి so |t|]

from the tables

$$t_{0.05,\,11} = -1.796$$

$$1 - \alpha = 1 - 0.05 = 0.95$$

$df = n - 1 \Rightarrow 12 - 1 = 11$

~~We observe~~ We observe

$$t_{cal}\,(-2.12) < t_{cri} \text{ or } t_{0.05,11}\,(-1.796)$$

Hence Accepted.

2) A machine which produces Mica insulating washers for use in Electric devices is set to turn out washers having a thickness of 10mm. A sample of 10 washers has an Average thickness of 9.52 mm with a standard Deviation of 0.60 m.m. Test whether the sample is drawn from the Given population. [Two tailed test]

Sol Given that $n = 10$; $\sigma = 0.60$

$\bar{x} = 9.52$; $\mu = 10$

[| | లేవి two tailed test గా తీసుకోవాలి]

$H_0: \mu = 10$

$H_1: \mu \neq 10$

Test statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \Rightarrow \frac{9.52 - 10}{0.60/\sqrt{10}} \Rightarrow \frac{-0.48}{0.60/3.1623}$$

$$= \frac{-0.48}{0.1897}$$

$$t_{cal} = -2.53$$

Because it is two tailed Test

$|t| = 2.53$

Given $n = 10$

$df = n-1 = 10-1$

$= 9$

At 5% (0.05) Level of Significance

$t_{0.05, 9} = 1.833$

We observe $t_{cal}(2.53) > t_{0.05,9}(1.833)$

Hence Rejected.

So, the Sample is not drawn from the population.

Table value కన్నా ఎక్కువ అయితే Rejected
తక్కువ అయితే Accepted

3) A random sample of 10 boys had the following IQs: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100 does this data support the assumption of a population mean IQ of 100.

Sol

$H_0: \mu = 100$

$H_1: \mu \neq 100$

Test statistic

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

$$\bar{x} = \frac{1}{n}\Sigma x_i$$

$$= \frac{972}{10}$$

$$= 97.2$$

$$S = \sqrt{\frac{1}{n-1}\Sigma(x_i - \bar{x})^2}$$

$$S = \sqrt{\frac{1}{10-1} \times 1833.60}$$

$$S = \sqrt{\frac{1}{9} \times 1833.60}$$

$$S = \sqrt{203.73}$$

$$S = 14.27$$

x	$x - \bar{x}$	$(x-\bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
$\Sigma x = 972$		$\Sigma(x-\bar{x})^2 = 1833.60$

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

$$= \frac{97.2 - 100}{14.27/\sqrt{10}} = \frac{-2.8}{14.27/3.1623} = \frac{-2.8}{4.5125} = -0.62$$

Because it is two-tailed Test $|t| = 0.62$ $\quad |t| = |-0.62|$

$t_{0.05, 9} = 1.833$ $\quad t = 0.62$

We observe

$t_{cal}(0.62) < t_{cri}(1.833)$

Hence Accepted.

3) Test of Significance of difference b/w two Means ("small" Sample) :-

$$t = \frac{\bar{x_1} - \bar{x_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim^{\text{Tending}} t_{n_1+n_2-2}$$

↓ degrees of freedom.

Where n_1 = 1st sample size ; $\bar{x_1}$ = mean of the first sample

n_2 = 2nd sample size ; $\bar{x_2}$ = Mean of 2nd sample

$$S^2 = \frac{1}{n_1+n_2-2}\left[(n_1-1)S_1^2 + (n_2-1)S_2^2\right]$$

$$S_1^2 = \frac{1}{(n_1-1)} \Sigma (x_i - \bar{x_1})^2$$

$$S_2^2 = \frac{1}{(n_2-1)} \Sigma (x_i - \bar{x_2})^2$$

4) To Examine the Hypothesis that Husbands are more intelligent than Wives, an Investigator took a Sample of 10 couples and conducted a test which measures the IQ as follows

Husbands :	117	105	97	105	123	109	86	78	107	107
Wives :	106	98	87	104	116	95	90	69	108	85

Test the Hypothesis with a reasonable Test at 0.05 level of Significance.

Sol Given $n_1 = 10$; $n_2 = 10$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

x_i	x_j	$(x_i - \bar{x}_1)$	$(x_i - \bar{x}_1)^2$	$(x_j - \bar{x}_2)$	$(x_j - \bar{x}_2)^2$
117	106	13.6	184.96	10.2	104.04
105	98	1.6	2.56	2.2	4.84
97	87	−6.4	40.96	−8.8	77.44
105	104	1.6	2.56	8.2	67.24
123	116	19.6	384.16	20.2	408.04
109	95	5.6	31.36	−0.8	0.64
86	90	−17.4	302.76	−5.8	33.64
78	69	−25.4	645.16	−26.8	718.24
107	108	3.6	12.96	12.2	148.84
107	85	3.6	12.96	−10.8	116.64
1034	958		1620.4		1679.6

$$\bar{x}_1 = \frac{\Sigma x_i}{n} = \frac{1034}{10} = 103.4$$

$$\bar{x}_2 = \frac{\Sigma x_j}{n} = \frac{958}{10} = 95.8$$

$$S_1^2 = \frac{1}{n_1 - 1} \Sigma (x_i - \bar{x}_1)^2$$

$$= \frac{1}{10-1} 1620.4 \Rightarrow \frac{1}{9} \times 1620.4$$

$$= 180.04$$

$$\Rightarrow S_1 = \sqrt{180.04}$$

$$= 13.42$$

$$S_2^2 = \frac{1}{n_2 - 1} \Sigma (x_j - \bar{x}_2)^2$$

$$= \frac{1}{10-1} \times 1679.6 \Rightarrow \frac{1}{9} \times 1679.6 = 186.62$$

$$\Rightarrow S_2 = \sqrt{186.62} = 13.66$$

$$S^2 = \frac{1}{n_1+n_2-2}\left[(n_1-1)S_1^2 + (n_2-1)S_2^2\right]$$

$$S^2 = \frac{1}{(10+10-2)}\left[(10-1)(180.04) + (10-1)186.62\right]$$

$$S^2 = \frac{1}{18}\left[(9\times 180.04) + (9\times 186.62)\right]$$

$$S^2 = \frac{1}{18}\left[1620.36 + 1679.58\right]$$

$$S^2 = \frac{3299.94}{18}$$

$$S^2 = 183.33$$

$$S = \sqrt{183.33}$$

$$S = 13.54$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \Rightarrow \frac{103.4 - 95.8}{13.54\sqrt{\frac{1}{10}+\frac{1}{10}}}$$

$$\Rightarrow \frac{7.6}{13.54\sqrt{0.1+0.1}} \Rightarrow \frac{7.6}{13.54\times 0.4472}$$

$$\Rightarrow \frac{7.6}{6.0553} \Rightarrow t = 1.25$$

$t_{cal} = 1.25$

critical value

$t_{n_1+n_2-2}$

$\Rightarrow t_{10+10-2} = t_{18}$

$t_{0.05,18} = 1.734$

We observe $t_{cal}\ (1.25) < t_{0.05,18}\ (1.734)$

So, Null Hypothesis Accepted

5) The Time taken by workers in performing a job by Method I and Method 2 are given below

Method I	20	16	26	27	23	22	
Method II	27	33	42	35	32	34	38

Do the data show that the variances of Time Distribution from means population from which these samples are drawn donot differ Significantly.

Jan 2016

6) Two kinds of Manures were applied to 16 Sixteen one hectare (1000 acres) plots. other conditions remaining the same, the yields in quintals are given below

Manure I	18	20	36	50	49	36	34	49	41
Manure II	29	28	26	35	30	44	46		

Is there any Significant difference between the mean yield.

Sol

**Chi-Square Test (χ^2) [774 page Table]

* chi-square criteria are based on observed and expected cell frequencies. This Test play an important role in many problems dealing with the analysis of count data in Business and Industry.

→ The rxc tables in chi-square test is called contingency Table. No(÷, x, −, +)

→ The rxc tables, Association is found between two attributes A&B. under consideration. If the hypothesis is (rejected), then the two attributes are considered to be associated, otherwise they are considered Independent.

For Example :- Whether a given medicine 'A' is effective in curing a desease 'B', Whether A given fertiliser 'A' is effective in increasing the crop yield 'B' (or) Whether the advertisement A has impact on the Sales 'B'.

→ In chi-square test, the rows and columns table which is called contigency table is as follows.

	A_1	A_2		A_c	
B_1	A_1B_1	A_2B_1		A_cB_1	B_1
B_2	A_1B_2	A_2B_2		A_cB_2	B_2
B_r	A_1B_r	A_2B_r		A_cB_r	B_r
	A_1	A_2		A_c	(N)

* In chi-square test, the frequencies given in the cells is called Observed frequencies ('o') and the frequencies to be calculated are called Expected frequencies.

* The formula used to calculate the Expected frequencies

$$e(A_1,B_1) = \frac{(A_1 \cdot B_1)}{N}$$

$$e(A_2B_1) = \frac{(A_2 \cdot B_1)}{N}$$

. . . .

$$e(A_c,B_r) = \frac{(A_c \cdot B_r)}{N}$$

So, finally

$$\chi^2_{cal} = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{[o_{ij} - e_{ij}]^2}{e_{ij}}$$

$$\chi^2_{cri} \sim (r-1)(c-1)df$$

Model - I

1) In an experiment on Immunization of cattle from tuberculosis the following particulars are available.

	Affected	Not-affected
Innoculated	12	26
Non-Innoculated	16	6

Analyse and Discuss the effect of vaccine in controlling the tuberculoses.

Sol Let H_0: The variables are independent (vaccine & Tuberculosis)

H_1: The variables are dependent.

12	26	38
16	6	22
28	32	60

Observed cell frequency

$O_{11} = 12$

$O_{12} = 26$

$O_{21} = 16$

$O_{22} = 6$

60

Expected frequencies

$e_{11} = \frac{38 \times 28}{60} = 17.73$

$e_{12} = \frac{38 \times 32}{60} = 20.27$

$e_{21} = \frac{22 \times 28}{60} = 10.27$

$e_{22} = \frac{22 \times 32}{60} = 11.73$

Now $\chi^2_{cal} = \frac{(12-17.73)^2}{17.73} + \frac{(26-20.27)^2}{20.27} + \frac{(16-10.27)^2}{10.27} + \frac{(6-11.73)^2}{11.73}$

$= \frac{(-5.73)^2}{17.73} + \frac{(5.73)^2}{20.27} + \frac{(5.73)^2}{10.27} + \frac{(-5.73)^2}{11.73}$

$= \frac{32.83}{17.73} + \frac{32.83}{20.27} + \frac{32.83}{10.27} + \frac{32.83}{11.73}$

$= 1.85 + 1.62 + 3.19 + 2.79$

$\chi^2_{cal} = 9.43$

$\chi^2_{\alpha, df} = \chi^2_{0.05, (r-1)(c-1)}$

$= \chi^2_{(0.05)(2-1)(2-1)}$

$= \chi^2_{0.05,1} = 3.841$

$\chi^2_{cri} = 3.841$

condition

$\chi^2_{cal} < \chi^2_{cri}$ (Accept)

$\chi^2_{cal} > \chi^2_{cri}$ (Reject)

We observe

$\chi^2_{cal}(9.43) > \chi^2_{cri}(3.841)$

So, H_0 Rejected

i.e H_1 is accepted

H_0 is now depended

There exist an Association between vaccine & Tuberculosis.

2) In an anti-malarial campaign in certain area, Quinine was given for a population in that area and the no. of fever cases recorded are given below

	Fever	No-fever
Quinine :	20	792
No quinine :	220	2216

Discuss the usefulness of Quinine.

3) The following data represents a person ability in mathematics and his interest in statistics.

		Ability in mathematics		
		Low	Avg	High
Interest in Statistics :	Low	63	42	15
	Avg	58	61	31
	High	14	47	29

Use 0.01 level of significance and test for Independence b/w persons ability in maths and in statistics.

Sol Let H_0 = They are independent

H_1 = They are dependent

63	42	15	120
58	61	31	150
14	47	29	90
135	150	75	360

Cell frequency

$O_{11} = 63$

$O_{12} = 42$

$O_{13} = 15$

$O_{21} = 58$

$O_{22} = 61$

$O_{23} = 31$

$O_{31} = 14$

$O_{32} = 47$

$O_{33} = 29$

Expected frequency

$e_{11} = \frac{120 \times 135}{360} = 45$

$e_{12} = \frac{120 \times 150}{360} = 50$

$e_{13} = \frac{120 \times 75}{360} = 25$

$e_{21} = \frac{150 \times 135}{360} = 56.25$

$e_{22} = \frac{150 \times 150}{360} = 62.5$

$$e_{23} = \frac{150 \times 75}{360} = 31.25$$

$$e_{31} = \frac{90 \times 135}{360} = 33.75$$

$$e_{32} = \frac{90 \times 150}{360} = 37.5$$

$$e_{33} = \frac{90 \times 75}{360} = 18.75$$

$$\chi^2_{cal} = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{[O_{ij} - e_{ij}]^2}{e_{ij}}$$

$$= \frac{(63-45)^2}{45} + \frac{(42-50)^2}{50} + \frac{(15-25)^2}{25} + \frac{(58-56.25)^2}{56.25} + \frac{(61-62.5)^2}{62.5} + \frac{(31-31.25)}{31.25}$$

$$+ \frac{(14-33.75)^2}{33.75} + \frac{(47-37.5)^2}{37.5} + \frac{(29-18.75)^2}{18.75}$$

$$= \frac{324}{45} + \frac{64}{50} + \frac{100}{25} + \frac{4}{56.25} + \frac{2.25}{62.5} + \frac{0.0625}{31.25} + \frac{390.1}{33.75} + \frac{90.25}{37.5} + \frac{105.1}{18.75}$$

$$= 7.2 + 1.28 + 4 + 0.036 + 0.002 + 11.56 + 2.41 + 5.61$$

$$\boxed{\chi^2_{cal} = 32.10}$$

$$\chi^2_{cri;(0.01)(r-1)(c-1)} = \chi^2_{cri;(0.01)(3-1)(3-1)} \Rightarrow \chi^2_{cri(0.01,4)} = 13.277.$$

(2x2)

* χ^2_{cal} (32.10) > χ^2_{cri} (13.277) ⟶ Rejected H_0 & accept H_1.

So, the ability of Maths and Interest in statistic of person is dependent.

Model - II chi-square Test of Goodness of fit

* chi-square test is used to check the significance of the discrepency (deviation) b/w Theory and Experiment. It is given by Karl pearson (P) in 1900 and is popularly known as "chi-square Test of goodness of fit".

If O_i $(i = 1,2,3,4, \ldots n)$ is a set of observed (or) Experimental frequencies and e_i $(i = 1,2,3, \ldots n)$ is a set of expected frequencies (or) Theoritical frequencies

chi-square test is given by

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - e_i)^2}{e_i} \sim \chi^2_{(n-1)} \cdot df$$

1) find the value of chi-square for the following data and test the goodness of fit

observed frequency : 14, 15, 18, 20, 15, 10

Expected frequency : 17, 10, 15, 25, 10, 15

Sol H_0: The given distribution fit is good

H_1: The given distribution fitness of good doesnot Exist

$$\chi^2_{cal} = \sum_{i=1}^{n} \frac{(O_i - e_i)^2}{e_i} \sim \chi^2_{(n-1)} df$$

$$= \frac{(14-17)^2}{17} + \frac{(15-10)^2}{10} + \frac{(18-15)^2}{15} + \frac{(20-25)^2}{25} + \frac{(15-10)^2}{10} + \frac{(10-15)^2}{15}$$

$$= \frac{9}{17} + \frac{25}{10} + \frac{9}{15} + \frac{25}{25} + \frac{25}{10} + \frac{25}{15}$$

$$= 0.53 + 2.5 + 0.6 + 1 + 2.5 + 1.67$$

$$\chi^2_{cal} = 8.8$$

$df = n - 1 = 6 - 1 = 5$

$\chi^2_{cri;(0.05,5)} = 11.070$

$\chi^2_{cal}(8.8) < \chi^2_{cri}(11.070)$. It is accepted H_0 So given distribution fit is good.

2) A personnel manager is interested in trying to determine whether absenteeism is greater on one day of the week than on the otherday of the Week. He has the following record for the past years.

Days of Week :	mon	Tues	Wed	Thurs	Fri
No. of Absentees :	66	57	54	48	75

Test Whether Absenteism is uni-formly destributed over the Week. By using test of goodness of fit

Sol Ho: They are uni-formly Distributed (fit is good)

H1: They are not uni-formly Distributed (fit is not good).

* Expected frequency = $\frac{66+57+54+48+75}{5} = \frac{300}{5} = 60$

uni-formly Absentees

$$\chi^2_{cal} = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} \sim \chi^2_{df}$$

$$= \frac{(66-60)^2}{60} + \frac{(57-60)^2}{60} + \frac{(54-60)^2}{60} + \frac{(48-60)^2}{60} + \frac{(75-60)^2}{60}$$

$$= \frac{36}{60} + \frac{9}{60} + \frac{36}{60} + \frac{144}{60} + \frac{225}{60}$$

$$= \frac{450}{60} = 7.5$$

$\boxed{\chi^2_{cal} = 7.5}$

o_i	e_i	$(o_i - e_i)$	$(o_i - e_i)^2$	$\frac{(o_i - e_i)^2}{60}$
				7.5

$$\chi^2_{cri} = \chi^2_{0.05\,(5-1)} = \chi^2_{0.05\,(4)} = 9.488$$

df = 4

$\boxed{\chi^2_{cal}(7.5) < \chi^2_{cri}(9.488)}$ Ho is Accepted

Hence there is uni-formly Distribution of absence / Week fit of good exist.

3) The following figures show the Distribution of digits in numbers chosen at random from a telephone directory

Digit : 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

frequency : 1026, 1107, 997, 966, 1075, 933, 1107, 972, 964, 853

Test Whether the digits may be taken to occur equally frequently in the Directory.

Sol H_0: They are uni-formly (equally) -frequented

H_1: They are not uni-formly -frequented.

Excepted -frequency : $\frac{\Sigma f}{n} = \frac{10000}{10} = 1000$

$$\chi^2_{cal} = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} \sim \chi^2_{(0.05)\,df}$$

o_i	e_i	$(o_i - e_i)$	$(o_i - e_i)^2$	$\frac{(o_i - e_i)^2}{e_i}$
1026	1000	26	676	0.676
1107	1000	107	11449	11.449
997	1000	−3	9	0.009
966	1000	−34	1156	1.156
1075	1000	75	5625	5.625
933	1000	−67	4489	4.489
1107	1000	107	11449	11.449
972	1000	−28	784	0.784
964	1000	−36	1296	1.296
853	1000	−147	21609	21.604
			58542	58542 = χ^2

$$\chi^2_{cal} = \frac{58542}{1000} = 58.542$$

$$\chi^2_{cri} = \chi^2_{0.05(10-1)} \Rightarrow \chi^2_{0.05,9} = 16.919$$

$$\chi^2_{cal}\ (58.542) > \chi^2_{cri}\ (16.919)$$

Wednesday 13/12/17

4) The following is the distribution of the hourly number of Trucks arriving at a companies awarehouse

Trucks arrivals :	0	1	2	3	4	5	6	7	8
frequency :	52	151	130	102	45	12	5	1	2

fit a poisson distribution and Test the Goodness of fit at 0.05 level of Significance.

Sol

H_0 : The fit is good.

H_1 : The fit is not good.

x_i	f_i	$f_i x_i$
0	52	0
1	151	151
2	130	260
3	102	306
4	45	180
5	12	60
6	5	30
7	1	7
8	2	16
	500	1010

$$\text{Mean} = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

$$= \frac{1010}{500}$$

$$= 2.02$$

In poisson Distribution

λ = mean

$\lambda = 2.02$

	$P(x) = N \cdot \frac{\lambda^x e^{-\lambda}}{x!}$	Expected frequency
0	$\frac{500 \cdot (2.02)^0 \cdot e^{-2.02}}{0!} = 500 \times 0.1327$	66.
1	$\frac{500 \cdot (2.02)^1 \cdot e^{-2.02}}{1!} = 500 \times 2.02 \times 0.1327$	134
2	$\frac{500 \cdot (2.02)^2 \cdot e^{-2.02}}{2!} = \frac{500 \times 4.0804 \times 0.1327}{2!}$	135
3	$\frac{500 \cdot (2.02)^3 \cdot e^{-2.02}}{3!} = \frac{500 \times 8.2424 \times 0.1327}{6}$	91
4	$\frac{500 (2.02)^4 e^{-2.02}}{4!} = \frac{500 \times 16.6497 \times 0.1327}{24}$	46
5	$\frac{500 \cdot (2.02)^5 e^{-2.02}}{5!} = \frac{500 \times 33.6323 \times 0.1327}{120}$	19
6	$\frac{500 \cdot (2.02)^6 e^{-2.02}}{6!} = \frac{500 \times 67.9373 \times 0.1327}{720}$	6
7	$\frac{500 \cdot (2.02)^7 e^{-2.02}}{7!} = \frac{500 \times 137.23 \times 0.1327}{5040}$	2
8	$\frac{500 \cdot (2.02)^8 e^{-2.02}}{8!} = \frac{500 \times 277.211 \times 0.1327}{40320}$	1

O_i	e_i	$O_i - e_i$	$(O_i - e_i)^2$	$(O_i - e_i)^2/e_i$
52	66	-14	196	2.97
151	134	17	289	2.16
130	135	-5	25	0.19
102	91	11	121	1.33
45	46	-1	1	0.02
12	19	-7	49	2.57
5	6	-1	1	0.17
1	2	-1	1	0.50
2	1	1	1	1
				10.90

$\chi^2_{cal} = 10.90$

$\chi^2_{cri} \Rightarrow \chi^2_{0.05,\ 8\ (9-1)} = 15.507$

$\chi^2_{cal} < \chi^2_{cri}$

Accept Ho.

The given distribution fit is good.

ANOVA (Analysis of variance) :-

* The Analysis of Variance was developed by fisher. ANOVA helps us to Test the Significance of the differences of variances among morethan two Sample means. Using ANOVA, We can make inferences about whether the samples are drawn from the population having same mean. ANOVA is used in such situations as comparing the mileage given by (5) five different brands of fuel, comparing the first year earnings of Graduates of half-a-dozen business Schools.

In ANOVA comparison will be made between means of more than two samples. ANOVA is the classification and cross classification of statistical data with a view of testing whether the means of specific classification differ significantly or they are Homogeneous.

ANOVA TABLE

The Technique of ANOVA includes a Table showing the Source of Variance, the Sum of squares, the degrees of freedom, mean Square (variance) and the formula for F-ratio

$$F_{statistic} = \frac{\text{Variance b/w the Samples}}{\text{Variance within the Samples}}$$

ANOVA is classified into

1. one way classification – one way ANOVA.
2. Two way classification – Two way ANOVA. (Manifold ANOVA)

One Way ANOVA :-

The Simplest form of ANOVA is a one way Model which is used with Simple random Samples in order to compare the effect of Single Independent variable on the dependent Variable. The following is the Specimen (model) of ANOVA Table.

Source of Variation	Sum of Squares	df	Mean Square	F-ratio
b/w Samples	SSC	K-1	MSC = SSC / (columns ← K-1)	MSC / MSE
Within Sample	SSE (SST - SSC)	N-K	MSE = SSE / N-K	
Total	SST	N-1 (↙ rows x colums)		

1) Below are given the yield (kgs per acre) for five Trial plots of varities of Treatment

Plot NO	Treatment			
	1	2	3	4
1	42	48	68	80
2	50	66	52	94
3	62	68	76	78
4	34	78	64	82
5	52	70	70	66

Check the Hypothesis whether the treatments have the same effect.

Sol

plot	Treatment			
	X_1	X_2	X_3	X_4
1	42	48	68	80
2	50	66	52	94
3	62	68	76	78
4	34	78	64	82
5	52	70	70	66
Total	240	330	330	400

Let T = Sum of all observations

$\Rightarrow 42+50+62+34+52+48+66+68+\cdots\cdots+66 = 1300$

Now $\frac{T^2}{N} = \frac{(1300)^2}{5\times4}$ $\quad N = r\times c$

$= 84500$ (correction factor)

SST = Sum of squares of all observations − correction factor

$= [42^2+50^2+62^2+34^2+52^2+\cdots\cdots+66^2] - 84500$

$= 88736 - 84500$

$SST = 4236$

$$SSC = \frac{(\Sigma X_1)^2}{N_1} + \frac{(\Sigma X_2)^2}{N_2} + \frac{(\Sigma X_3)^2}{N} + \dots + \frac{(\Sigma X_n)^n}{N_n} - \frac{T^2}{N}$$

$$= \frac{(240)^2}{5} + \frac{(330)^2}{5} + \frac{(330)^2}{5} + \frac{(400)^2}{5} - 84500$$

$$= \frac{57600}{5} + \frac{108900}{5} + \frac{108900}{5} + \frac{160000}{5}$$

$$= 11520 + 21780 + 21780 + 32000 - 84500$$

$$= 87080 - 84500$$

$$SSC = 2580$$

$$SSE = SST - SSC$$

$$= 4236 - 2580 = 1656$$

$$MSC = \frac{SSC}{K-1} = \frac{2580}{4-1} = 860$$

(K-1 → columns)

$$MSE = \frac{SSE}{N-K} = \frac{1656}{20-4} = 103.5$$

Source of Variation	Sum of Squares	df	Mean Square	Fcal
b/w Samples	SSC = 2580	K-1 = 3	MSC = 860	$\frac{860}{103.5} = 8.3$ ($\frac{MSC}{MSE}$)
Within Sample	SSE = 1656	N-K = 16 (20-4)	MSE = 103.5	
Total	SST	N-1		

$F_{cal} = 8.3$

df $\Rightarrow (K-1), (N-K)$

$\Rightarrow (4-1)(20-4) \Rightarrow (3, 16)$

$\Rightarrow F_{table} = 2.24$

$F_{cal} > F_{table}$

H_0 Rejected.

Thursday
14/12/17

Two-way ANOVA

In Two-way ANOVA, the data is classified according to two criteria.

Procedure for Two-way ANOVA:-

step ①: Assuming that the means of all columns is equal that is the effect of all factors in one kind of Treatment are equal i.e.,

$\alpha_1 = \alpha_2 = \alpha_3 = \cdots\cdots = \alpha_c$ (c – columns)

step ②:- Assuming the means of all rows are equal that is the effect of all factors in the second Treatment are equal i.e.,

$\beta_1 = \beta_2 = \beta_3 = \cdots\cdots = \beta_r$ (r – rows)

step ③: compute T = Sum of all values

step ④: compute SST = Sum of all squares of observations $- \frac{T^2}{N}$

step ⑤: compute

$$SSC \quad \frac{(\Sigma x_1)^2}{N_1} + \frac{(\Sigma x_2)^2}{N_2} + \cdots + \frac{(\Sigma x_n)^2}{N_n} - \frac{T^2}{N}$$

where $\Sigma x_1, \Sigma x_2, \ldots, \Sigma x_n$ are column Totals.

step ⑥: compute

$$SSR \quad \frac{(\Sigma x_{j1})^2}{N_1} + \frac{(\Sigma x_{j2})^2}{N_2} + \cdots + \frac{(\Sigma x_{jn})^2}{N_n} - \frac{T^2}{N}$$

where $\Sigma x_{j1}, \Sigma x_{j2}, \ldots, \Sigma x_{jn}$ are Row totals

step ⑦:- $SSE = SST - SSC - SSR$

step ⑧:- $MSC = \frac{SSC}{C-1}$, $MSR = \frac{SSR}{R-1}$, $MSE = \frac{SSE}{(r-1)(c-1)}$

step ⑨:- $F_c = \frac{MSC}{MSE}$ $F_r = \frac{MSR}{MSE}$

step 10:- Degrees of freedom for $F_c = (c-1), (c-1)(r-1)$

$F_r = (r-1), (c-1)(r-1)$

F_c = column wise comparison

F_r = Row wise comparison

Source of Variation	Sum of Squares	df	Mean Square	F ratio
columns	SSC	$c-1$	MSC	$F_c = \frac{MSC}{MSE}$
Rows	SSR	$r-1$	MSR	
Error	SSE	$(c-1)(r-1)$	MSE	$F_r = \frac{MSR}{MSE}$
Total	SST	$N-1$		

1) Three varities of crops A, B, C are tested with four replications as given below. The yields are shown in the following Table

Replications

Variety	1	2	3	4
A	6	4	8	6
B	7	6	6	9
C	8	5	10	9

Test whether there is a difference between Replications and also the crop varieties differ significantly.

Sol

Replications

Variety	1	2	3	4	
A	6	4	8	6	24
B	7	6	6	9	28
C	8	5	10	9	32
Total	21	15	24	24	84

C - Ho: There is a significant difference in replications

R - Ho: There is a significant difference in variety of crops.

Let T = Sum of all observations

= 84

N = Number of observations = 12

$$\frac{T^2}{N} = \frac{84^2}{12} = \frac{7056}{12} = 588$$

SST = Sum of squares of all observations − Correction factor $\frac{T^2}{N}$

$$= (6^2 + 7^2 + 8^2 + 4^2 + 6^2 + 5^2 + 8^2 + 6^2 + 10^2 + 9^2 + 6^2 + 9^2) - 588$$

$$= 624 - 588$$

$$\boxed{SST = 36}$$

$$SSC = \frac{(\Sigma X_1)^2}{N_1} + \frac{(\Sigma X_2)^2}{N_2} + \frac{(\Sigma X_3)^2}{N_3} - \frac{T^2}{N}$$

$$= \frac{21^2}{3} + \frac{15^2}{3} + \frac{24^2}{3} + \frac{24^2}{3} - 588$$

$$= 606 - 588$$

$$\boxed{SSC = 18}$$

$$SSR = \frac{24^2}{4} + \frac{28^2}{4} + \frac{32^2}{4} - \frac{T^2}{N}$$

$$= 596 - 588$$

$$= 8 \Rightarrow SSR = 8$$

$$SSE = SST - SSC - SSR$$

$$= 36 - 18 - 8$$

$$= 10$$

$$MSC = \frac{SSC}{c-1} = \frac{18}{4-1} = \frac{18}{3} = 6$$

$$MSR = \frac{SSR}{r-1} = \frac{8}{3-1} = \frac{8}{2} = 4$$

$$MSE = \frac{SSE}{(c-1)(r-1)}$$

$$= \frac{10}{(4-1)(3-1)}$$

$$= \frac{10}{6}$$

$$= 1.67$$

Source of Variation	Sum of squares	df	Mean squares	Fratio
Columns	18	(4-1) 3	6	$F_c = \frac{MSC}{MSE} = \frac{6}{1.67} = 3.59$
Rows	8	(3-1) 2	4	$F_r = \frac{MSR}{MSE} = \frac{4}{1.67} = 2.39$
Error	10	(4-1)(3-1) 6	1.67	
Total	36	11		

Columns

$df = (C-1), (C-1)(r-1)$

$= (4-1), (3-1)(4-1)$

$= 3, 6$

$F_{cr} = 4.76$

$F_{cal} < F_{cr}$;

Hypothesis Accepted

Rows

$df = (r-1), (C-1)(r-1)$

$= (3-1), (4-1)(3-1)$

$= 2, 6$

$F_{cr} = 5.14$

$F_{cal} < F_{cr}$;

Hypothesis Accepted

2016 case study

2) The following data represents the sales (Rs '000) per month of four brands of soaps allocated among four cities

Brand	City A	B	C	D
I	12	48	30	29
II	42	57	54	39
III	19	42	21	36
IV	28	39	44	50

Test whether

i) the mean sales of three brands are equal

ii) the mean sales of the soaps in each city are equal

Sol

Band	City A	B	C	D	
I	12	48	30	29	— 119
II	42	57	54	39	— 192
III	19	42	21	36	— 118
IV	28	39	44	50	— 161
	1	1	1	1	590
	101	186	149	154	

C - H0: Sales of four brands are equal

R - H0: Sales of four soaps are equal in each city

T = Sum of all observations

$\boxed{T = 590}$

N = No. of observation

$\boxed{N = 16}$

Correction factor $= \frac{T^2}{N}$

$= \frac{590^2}{16} \Rightarrow \frac{348100}{16}$

$= 21756.25$

SST = Sum of square of all observations − correction factor $\left(\frac{T^2}{N}\right)$

$= (12^2 + 42^2 + 19^2 + 28^2 + \ldots\ldots + 50^2) - 21756.25$

$= 24242 - 21756.25$

$\boxed{SST = 2485.75}$

$SSC = \frac{(\Sigma X_1)^2}{N_1} + \frac{(\Sigma X_2)^2}{N_2} + \frac{(\Sigma X_3)^2}{N_3} + \frac{(\Sigma X_4)^2}{N_4} - \frac{T^2}{N}$

$= \frac{101^2}{4} + \frac{186^2}{4} + \frac{149^2}{4} + \frac{154^2}{4} - 21756.25$

$= 22678.5 - 21756.25$

$\boxed{SSc = 922.25}$

$SSR = \frac{(\Sigma x_{j1})^2}{N_1} + \frac{(\Sigma x_{j2})^2}{N_2} + \frac{(\Sigma x_{j3})^2}{N_3} + \frac{(\Sigma x_{j4})^2}{N_4} - \frac{T^2}{N}$

$= \frac{119^2}{4} + \frac{118^2}{4} + \frac{161^2}{4} + \frac{192^2}{4} - 21756.25$

$= 22717.5 - 21756.25$

$\boxed{SSR = 961.25}$

$$SSE = SSC - SST - SSR$$

$$= 2485.75 - 922.25 - 961.25$$

$$\boxed{SSE = 602.25}$$

$$MSC = \frac{SSC}{C-1} = \frac{922.25}{4-1} = 307.4167$$

$$\boxed{MSC = 307.42}$$

$$MSR = \frac{SSR}{r-1} = \frac{961.25}{4-1}$$

$$\boxed{MSR = 320.42}$$

$$MSE = \frac{SSE}{(r-1)(C-1)}$$

$$= \frac{602.25}{(4-1)(4-1)} = \frac{602.25}{9}$$

$$\boxed{MSE = 66.92}$$

Source of variation	Sum of squares	df	Mean Square	F ratio
columns	922.25	3	307.42	$Fc = \frac{MSC}{MSE} = \frac{307.42}{66.92}$
Rows	961.25	3	320.42	$Fc = 4.5938$
Error	602.25	9	66.92	$Fr = \frac{MSR}{MSE} = \frac{320.42}{66.92}$
Total	2485.75	15		$Fr = 4.788$

column

$df = (C-1), (C-1)(r-1)$

$= (4-1), (4-1)(4-1)$

$= 3, 9$

$F_{cri} = 3.86$

$F_{cal} > F_{cri}$

Hypothesis rejected

Rows

$df = (r-1), (C-1)(r-1)$

$= (4-1), (4-1)(4-1)$

$= 3, 9$

$F_{cri} = 3.86$

$F_{cal} > F_{cri}$

Hypothesis Rejected.

2015 case study

3) four experiments determine the moisture content of samples of a powder, Taking a Sample from each of the Six consignments of load. Their Assessments are given below

observers	consignments 1	2	3	4	5	6
1	9	10	9	10	11	11
2	12	11	9	11	10	10
3	11	10	10	12	11	10
4	12	13	11	14	12	10

Analyse the data and Discuss whether there is any significant difference between the consignments or between the observers.

2014 case study

4) You are given the following data relating to the number of table fans sold by four sales man in 3 months April, may, June

Months	Salesman A	B	C	D
April	50	40	48	39
may	46	48	50	45
June	39	44	40	39

i) Is there Significand difference in the Sales made by the four Salesman

ii) Is there a Significant difference in the Sales made during different months.