MAA OMWATI DEGREE COLLEGE

HASSANPUR

NOTES

CLASS:- MBA 1ST SEM

SUBJECT: - BUSINESS RESEARCH METHOD (MC)

Unit-1

What is Research?

Research is a systematic inquiry that investigates hypothesis, suggests new interpretations of data or texts, and poses new questions for future research to explore.

Objectives of Research:

The purpose of research is to discover answers to questions through the application of scientific procedures.

The main aim of research is to find out the truth which is hidden and which has not been discovered as yet.

Though each research study has its own specific purpose, we may think of research objectives as falling into a number of following broad groupings:

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

SCOPE OF RESEARCH

1. To know the Business Competition – Environment
a. To know the Customers:- Know about customers needs,
b. To know the Product – Design, Price, Expectations
2. To know the Industry Competition:
3. Maturing of Management as a group of disciplines
4 Stakeholders demanding greater influence
5. To know the Global Competition
6. Government Intervention
7. Economical Data Collection

What is Business Research?

Business research is **a process of acquiring detailed information of all the areas of business and using suchAlso Read: Scope of MBA in International Business**

Business Research Example

Let say there's an automobile company that is planning to launch a car that runs on CNG. In order to promote cleaner fuel, the company will be involved in developing different plans and strategies to identify the demand for the car they are intending to launch. Other than this, the company will also look for competitors, the target audience, keeping in mind the distribution of CNG in India. Hence the researches are conducted on various ideas to formulate a sustainable and more efficient design.

Importance of Business Research

When it comes to the question why Business Research is important, it has an essential role to play in varied areas of business. Here are some of the reasons describing the importance of Business Research:

- It helps businesses gain better insights about their target customer's preferences, buying patterns, pain points, as well as demographics.

- Business Research also provides businesses with a detailed overview of their target markets, what's in trend, as well as market demand.
- By studying consumers' buying patterns and preferences as well as market trends and demands with the help of business research, businesses can effectively and efficiently curate the best possible plans and strategies accordingly.
- The importance of business research also lies in highlighting the areas where unnecessary costs can be minimized and those areas in a business which need more attention and can bring in more customers and hence boost profits.
- Businesses can constantly innovate as per their customers' preferences and interests and keep their attention towards the brand.
- Business Research also plays the role of a catalyst as it helps business thrive in their markets by capturing all the available opportunities and also meeting the needs and preferences of their customers.
- MANAGERIAL VALUE OF BUSINESS RESEARCH
- The prime managerial value of business research is that it reduces uncertainty by providing information that facilitates decisionmaking about strategies and the tactics used to achieve an organization¶s strategic goals. The decision-making process involvesthree interrelated stages.
- A.Identifying the existence of problems and opportunities. Before any
- strategy can be developed, an organization must determine where it wants to go and how it will get there. Business research canhelp managers plan strategies by determining the nature of situations by identifying the existence of problems or opportunities present in the organization.
- B.
- Diagnosis and Assessment . After an organization recognizes a problem or identifies a potential opportunity, an important aspect of business research is often the provision of diagnostic information that clarifies the situation. Managers need to gains insight about the underlying factors causing the situation.

Types of Research :

1. Descriptive Research
2. Exploratory Research
3. Applied Research
4. Fundamental Research

5. Quantitative Research
6. Qualitative Research
7. Conceptual Research
8. Empirical Research
9. Experimental Research
10. Historical Research

Specific to field of study

Graduate students learn about research methods used in their particular field of study. Whatever methods are used, there are many resources to support research, and any number of variations to the basic methods. Choose a method or variation that is manageable in your first years as a new faculty member to be certain to get your research agenda underway. Often the biggest challenge is often to get started, so establish research priorities early on, and create a plan to impliment them.

Quantitative Research

Quantitative research is inquiry into an identified problem, based on testing a theory, measured with numbers, and analyzed using statistical techniques. The goal of quantitative methods is to determine whether the predictive generalizations of a theory hold true. We will explore some of the issues and

challenges associated with quantitative research in this section. Seek the advice of faculty members who have conducted quantitative studies for advice, support and encouragement.

Qualitative Research

A study based upon a qualitative process of inquiry has the goal of understanding a social or human problem from multiple perspectives. Qualitative research is conducted in a natural setting and involves a process of building a complex and holistic picture of the phenomenon of interest. We will explore some of the issues and challenges associated with qualitative research in this section. Look for colleagues who engage in qualitative research to serve as a sounding board for procedures and processes you may use as a new faculty member.

Collaborative Research

There are many ways to collaborate and thereby enrich your work as a faculty member. Multiple perspectives offer a more complete view of an issue under study. Whether presenting or publishing, having several iterations of a study in different settings offers a means of validating findings. Collaboration may take many forms, including crossing disciplines, types of institutions, or engaging the community outside the campus environment. Explore the ways that collaboration may enrich your research. Some of the issues and challenges associated with collaborative research are explored in this section.

Practitioner Research

Simple definitions of practitioner research address the investigator, the setting and the purpose. The investigator is the practitioner, in workplace settings ranging from hospitals, to schools and communities. The general purpose is to better align the practitioner's purpose with their actions. There

are those who argue that practitioner research stems from a larger social justice movement within qualitative research. Even when social justice is not the sole motivating principle, an underlying commonality of purpose is the desire to improve upon and develop deeper insights into one's practice. Practitioner research by its nature offers practitioners a voice in the research conversation. Some consider it a bridge of sorts between theory and practice, although practitioners claim a rightful place in the research continuum. Some of the issues and challenges associated with practitioner research are explored in this section.

What are the various types of research?

1. **Pure research**
   a. This is called as the fundamental or the theoretical research.
   b. Is basic and original.
   c. Can lead to the discovery of a new theory.
   d. Can result in the development or refinement of a theory that already exists.
   e. Helps in getting knowledge without thinking formally of implementing it in practice based on the honesty, love and integrity of the researcher for discovering the truth.

2. Applied research

a. Based on the concept of the pure research.
b. Is problem oriented.
c. Helps in finding results or solutions for real life problems.
d. Provides evidence of usefulness to society.
e. Helps in testing empirical content of a theory.
f. Utilizes and helps in developing the techniques that can be used for basic research.
g. Helps in testing the validity of a theory but under some conditions.
h. Provides data that can lead to the acceleration of the process of generalization.

3. Exploratory research

a. Involves exploring a general aspect.
b. Includes studying of a problem, about which nothing or a very little is known.
c. Follows a very formal approach of research.
d. Helps in exploring new ideas.
e. Helps in gathering information to study a specific problem very minutely.
f. Helps in knowing the feasibility in attempting a study.

4. Descriptive research

a. Simplest form of research.
b. More specific in nature and working than exploratory research.
c. It involves a mutual effort.
d. Helps in identifying various features of a problem.
e. Restricted to the problems that are describable and not arguable and the problems in which validstandards can be developed for standards.
f. Existing theories can be easily put under test by empirical observations.
g. Underlines factors that may lead to experimental research.
h. It consumes a lot of time.
i. It is not directed by hypothesis.

5. Diagnostic study

a. Quite similar to the descriptive research.
b. Identifies the causes of the problems and then solutions for these problems.
c. Related to causal relations.
d. It is directed by hypothesis.
e. Can be done only where knowledge is advanced.

6. Evaluation study

a. Form of applied research.
b. Studies the development project.

c. Gives access to social or economical programmes.
d. Studies the quality and also the quantity of an activity.

7. Action research
a. Type of evaluation study.
b. Is a concurrent evaluation study.

Managerial Value of Business Research-

Research, may help managers in organizations to make decisions at the work place, sometimes they make good decisions and the problem solved, sometimes they make poor decision and the problem persists.-

The successful manager must know how to make the right decision by being knowledgeable about the various steps involved in finding solutions to problematic issue.

The decision-making process associated with the development and implementation of a business strategyinvolves four interrelated stages:

1.Identifying problems and opportunities

2.Diagnosing and assessing problems or opportunities

3.Selecting and implementing a course of action

4. Evaluating the course of action

5. Implementing the results to solve the problem

Theoretical Framework for Your Research in 4 Steps

If you're planning to conduct a study or write a research paper, consider using a theoretical framework to organize and document your work. A theoretical framework can help you comply with industry-standard rules for research, which can allow you to publish your study, as well as use your findings for a dissertation.

In this article, we discuss what a theoretical framework is, how it differs from a conceptual framework, why you'd use a theoretical framework and how to create one in four steps, plus we provide an example for you to use when creating your own framework.

What is a theoretical framework?

Before defining what a theoretical framework is, it's important to first understand what a theory is.

Theories propose well-substantiated explanations of phenomena, an observable fact or event, such as Newton's theory of gravitation. They allow researchers to make predictions and connect the issues they're investigating to existing knowledge, according to John Kuada, professor and author of "Research Methodology: A Project Guide for University Students."

Researchers use theoretical frameworks to explain the theories they're using within their research and provide their own research with context by identifying the assumptions that inform their work. A theoretical framework is often integrated into a literature review section near the beginning of a paper or experiment but can also be included as its own chapter or section.

Theoretical vs. conceptual framework

The terms "theoretical framework" and "conceptual framework" are often used interchangeably but there are key differences between the two:

- Theoretical and conceptual frameworks basically differ in terms of scope.
- While a theoretical framework explores the relationship between things in a given phenomenon in a broad and general way, a conceptual framework is more specific and represents the researcher's idea on how the research problem will be explored.
- A conceptual framework includes the ideas that are used to define research and evaluate data. They usually fall within a theoretical framework and are often laid out using visuals and graphics. Like theoretical frameworks, they're often included at the beginning of a paper or experiment.
- Conceptual frameworks consist of variables, concepts, theories and parts of other existing frameworks.

How to create a theoretical framework

The same research topic can be approached differently within different theoretical frameworks. For example, in psychology, different kinds of therapists might address trauma using various theories. Behavioral therapists may prefer to treat symptoms of trauma using exposure therapy while another therapist might recommend an eye movement desensitization and reprocessing (EMDR) session.

Here's how to create a theoretical framework for your study or research paper in four steps:

1. Define your objective

The first step in creating a theoretical framework is to define your research objective and then gather supporting documents. Consider what you hope to achieve with your study and what new information you aim to bring to your chosen field.

You can also think about the subject you're planning to study and how your research might affect the knowledge the scientific community already has about the subject matter. To prepare for a successful study, define each of your objectives clearly to help your audience understand them.

2. Write a problem statement

Once you define your objectives, you can write a problem statement to define the purpose of the study and what problems it's addressing. For example, if you're researching the moons of Mars you might create a problem statement like this:

Rather than providing in-depth information, try using concise and straightforward language in your problem statement to explain your study's purpose in an easy-to-read manner.

3. Present your research questions

A research question is an inquiry that you aim to address with your study results. These questions often provide a basis for your theories or hypotheses. They can also help your audience understand the study's purpose and the information you're seeking. This can clarify to your audience whether your study was successful.

A researcher studying Mars' moons might ask the following research questions:

- *How can I prove the astral body I'm observing is another moon?*
- *What tools can I use to verify my findings?*
- *What qualities am I looking for to verify the object shares an orbit with Mars?*
- *What factors affect the object's behavior in space?*
- *Why has no one else observed the object until now?*

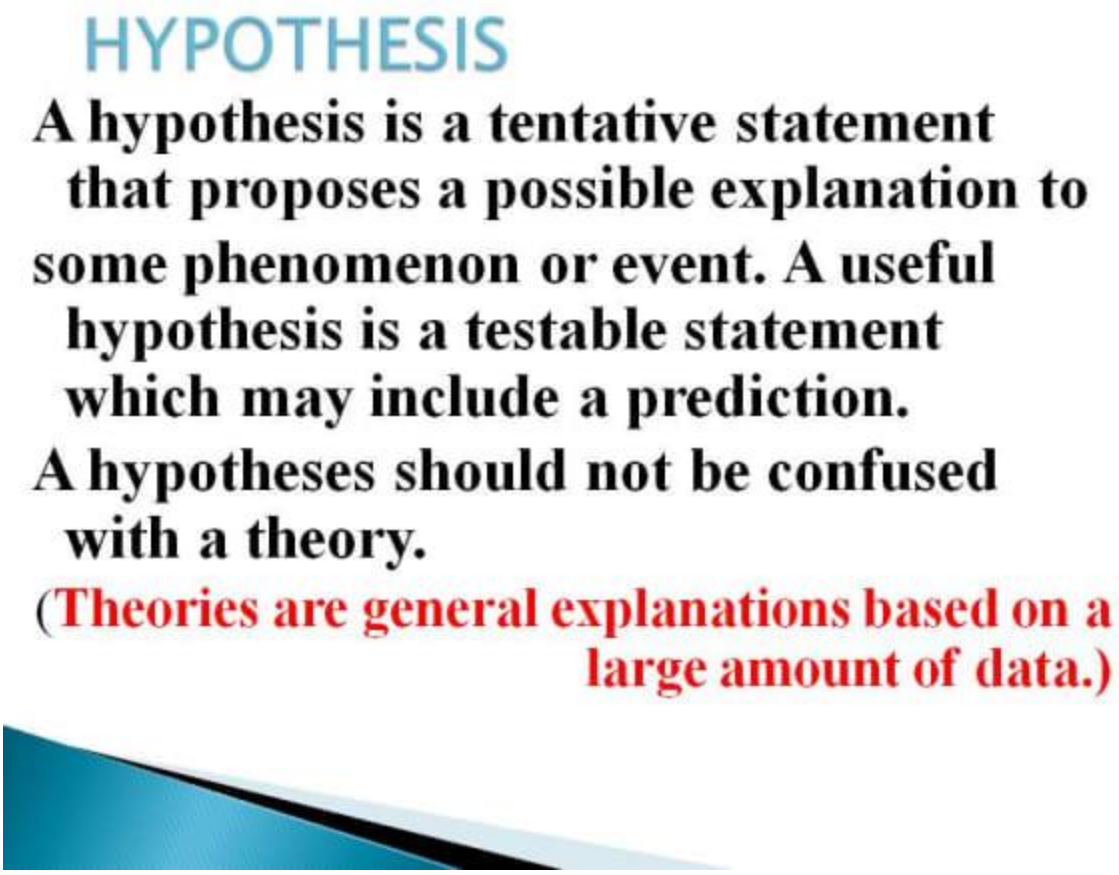
If you plan to answer more than one question with your study, consider adding a section of subquestions to support your main research question.

4. Create a literature review to highlight relevant theories

A literature review is a summary or analysis of all the research that you read before doing your own research. It collects key sources on a topic and outlines how those sources and other researchers' theories align with your study and how you're using their research to develop your theories. Since some research requires several theories from a variety of studies, it's helpful to define why each theory applies to your research in your theoretical framework.

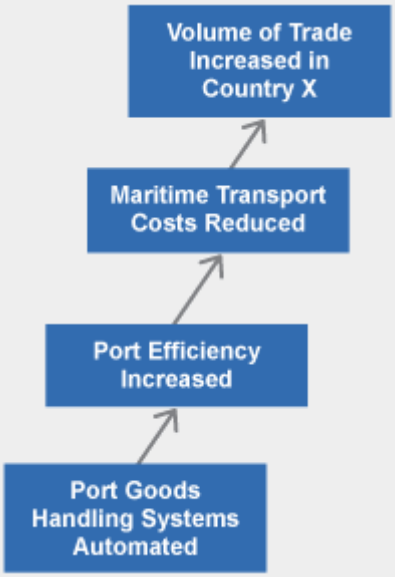
For example, if you're studying Mars' moons, you might evaluate research from astronomers, astrophysicists or other scientists to determine whether their work supports or denies the existence of another Mars moon.

Hypothesis Development in Business Research



DEVELOPMENT HYPOTHESES

A development hypothesis states what will occur if a particular intervention is undertaken, or a combination of several building blocks that are critical for bringing about a particular development outcome are put in place



While like those in the diagram about moving traded goods across borders are gender neutral, other development hypotheses included in a strategy may have different implications for men and women's involvement in export production, or their adoption of streamlined procedures introduced by a customs administration. For all development hypotheses that involve individual behavior, including responses to incentives,

A research hypothesis is a statement of expectation or prediction that will be tested by research. Before formulating your research hypothesis, read about the topic of interest

to you. From your reading, which may include articles, books and/or cases, you should gain sufficient information about your topic that will enable you to narrow or limit it and express it as a research question. The research question flows from the topic that you are considering. The research question, when stated as one sentence, is your Research Hypothesis. In some disciplines, the hypothesis is called a “thesis statement.”Other words for “hypothesized” are “posited,” “theorized” or “proposed”. Remember, your hypothesis must REQUIRE two or more disciplines, one of which is law. This is essential, since your paper is interdisciplinary and a demonstration of the interdisciplinary process. In your hypothesis, you are predicting the relationship between variables. Through the disciplinary insights gained in the research process throughout the year, you “prove” your hypothesis. This is a process of discovery to create greater understandings or conclusions. It is not a strict proof as in logic or mathematics. Following are some hints for the formulation of your hypothesis:

• 1. Be sure to read on the topic to familiarize yourself with it before making a final decision. You need to make certain that the topic is researchable in an interdisciplinary sense, meaning that there is sufficient published material on the topic in the legal literature and in the literature of a second or possibly a third discipline to write a 35-page paper.

• 2. As noted, a research hypothesis is more than just a topic. It has two elements (variables) that are in relation to each other. Remember that, within the word "hypothesis" is the word "thesis." Your hypothesis is what you propose to “prove” by your research. As a result of your research, you will arrive at a conclusion, a theory, or understanding that will be useful or applicable beyond the research itself.

• 3. Avoid judgmental words in your hypothesis. Value judgments are subjective and are not appropriate for a hypothesis. You should strive to be objective. Therefore the use of personal opinion is to be avoided.

• 4. Your hypothesis must involve an issue or question that cannot be answered exclusively by the discipline of law. You should try to limit your inquiry to the literatures of 2 or 3 disciplines. It is best to choose a hypothesis where you already have some level of familiarity with the disciplines that are most relevant to the topic.

• 5. Be sure that each term in your hypothesis is clearly understood and defined; do not deal in generalities or assume that the reader knows the meaning of a technical term.

• 6. Specify, if appropriate, whether you will be dealing with state or federal law or both on a comparative basis if appropriate.

• 7. Know that your hypothesis may change over time as your research progresses. You must obtain the professor's approval of your hypothesis, as well as any modifications to your hypothesis, before proceeding with any work on the topic. Your will be expressing your hypothesis in 3 ways:

• As a one-sentence hypothesis

• As a research question

• As a title for your paper Your hypothesis will become part of your research proposal

What is literature survey in research?
A **literature review** is a comprehensive summary of previous **research** on a topic. The **literature review surveys** scholarly articles, books, and other sources relevant to a particular area of **research**. The **review** should enumerate, describe, summarize, objectively evaluate and clarify this previous **research**
Different types of literature reviews

- Narrative or Traditional **literature** reviews. Narrative or Traditional **literature** reviews critique and summarise a body of **literature** about the thesis topic. ...
- Scoping Reviews. ...
- Systematic Quantitative **Literature Review**. ...
- Cochrane Reviews. ...
- Campbell Collaboration.

Some aspects of literature survey

Just like most academic papers, **literature reviews** also must contain at least **three** basic **elements**: an introduction or background information section; the body of the **review** containing the discussion of sources; and, finally, a conclusion and/or recommendations section to end the paper.

The Seven Steps of the Research Process

Research, as a tool for progress, satisfies mankind's curiosity to lots of questions. Whether you are a high school or college student, you have to take research subject for you to be able to receive your diploma. To ease your burden in doing research, here are the seven steps in the research process:

1. Identification of a research problem

A good research always starts with a good problem. You can observe people or things, visit places, read print materials, or consult experts to find the research problem that is right for you. The research problem guides you in formulating the hypothesis and interpretation of your findings so that you can formulate the right conclusion. A good research problem is important because it is the basis of all subsequent research activities you are going to undertake. Factors like area of interest, availability of fund, socio-economic significance of the study, and the safety measures to be undertaken should be considered in finding a good research problem.

2. Formulation of Hypothesis

After finding your research problem, the next step is to formulate your own hypothesis. A hypothesis is a theoretical statement in solving a logical relationship between variables. Do not be afraid if your hypothesis proves to be incorrect after the experimentation because it is only considered as an educated guess. Always remember that when you formulate a hypothesis, it should be based on the research problem being solved.

3. Review of Related Literature

A research problem is vague at first. To give you a vivid picture of the whole research, you shall read various publications or surf the internet to become aware of the previous works already done. In doing so, it could spur an idea that can be the subject of your investigation. The review of related literature can be taken from science books, magazines, journals, newspapers, or even in the internet.

4. Preparation of Research Design

A research design is the blueprint of the research you are going to undertake. It serves as the work plan of the whole study not only because it entails the resources needed in conducting the research but also the ways these resources are utilized.

5. Actual experimentation

Actual experimentation is an implementation of the research design. In actual experimentation, you have to conduct an experiment to prove the validity of the hypothesis you have formulated. Actual experimentation includes the methodology that you have followed in doing your research. The methodology should be carefully planned prior to the actual experimentation to ensure the validity and accuracy of the result.

6. Results and Discussion

This is the heart of the research process because this is part where the findings of the research can be found. You can use table (not the table in your kitchen) and graph to interpret the results of your research.

7. Formulation of Conclusions and Recommendations

Conclusion is a statement where you will present the solution to the proposed problem based on the findings of the investigation. They are tied up to the questions investigated. Your conclusion will show whether or not your experiment worked. It should answer your hypothesis and research problem. In your concluding statement you can also infer on the possible benefits to society that your results might present. You can state any plans you might have to continue working on other aspects related to your area of study. We must remember that recommendations are based on conclusions and conclusions are based on findings

Conducting research is a tiresome task because it is a year-round activity. You have to be committed to become successful in making a good research which would benefit not only you but of everyone. The willingness in you in making future researches should always be there because doing research without your ‘heart’ and ‘mind’ on it is a burden on your part. Always enjoy doing it. Enjoy the ride. The more that you enjoy doing the task, the more that you ease the burden in conducting this difficult endeavor. Never hesitate to ask questions. Asking questions from other people who is aware of your research topic would help you arrive at the correct conclusions. Now that you have finally learned the steps in the research process, you can now start doing one.

Research Proposal

A research proposal **describes what you will investigate, why it's important, and how you will conduct your research**. The format of a research proposal varies between fields, but most proposals will contain at least these elements: Title page. Introduction. Literature review.

The purpose of the research proposal (it's job, so to speak) is to **convince** your research supervisor, committee or university that your research is **suitable** (for the requirements of the degree program) and **manageable** (given the time and resource constraints you will face).

STRUCTURE OF A RESEARCH PROPOSAL

- **Cover Page:** Contains your project title, your name, your supervisor's name, program/department, institution or affiliation, and date.
- **Table of Contents:** Outlines the contents of your entire proposal with respective page numbers.
- **Introduction:** Contains background and context, a problem statement, research questions, and the rationale behind the study.
- **Literature review:** Contains key concepts and theories that serve as the framework for your study as well as any gaps in research.
- **Research design and methods:** Contains research objectives, method, and potential limitations
- **Implications:** Explains how the study can be applied to the existing field of knowledge on the topic.
- **Reference list:** A list of references used to write the proposal.
- **Research schedule:** A timeline of research phases and how they will achieve the objective and meet deadlines.

In compliance with APA style, you can use these sections as headings for your document as well. Using section headings makes information more organized for the reader and allows them to follow the author's thoughts more clearly.

WRITING STYLE

Besides the contents of your proposal, you also need to pay attention to your writing style. It is going to be different from other papers or documents you may have had to write in the past. According to Academic Writer (n.d.), the following are some of the main elements of writing style. These are important to making your proposal sound respectful and professional.

Language

Instead of using common language, which is the type of language we use in normal conversations, you want to use the "language of research" or the "language of science." This means that if a term has two meanings, you should only use the term for the meaning that is the most relevant to your research. For example, if a chemist

uses the word "element" in a proposal, they use it only in the context of its scientific definition. This prevents the reader from getting confused throughout the document. Avoid creating new terms in your proposal and be sure to clearly define unfamiliar words at the beginning of the proposal (Locke et al., 2007). Lastly, you also want to avoid using first person in your proposal ("I will...") as it does not demonstrate professionalism in writing.

Tone

The tone of your writing should be professional and serious. In other words, use "academic voice" in your proposal writing. Academic voice is meant to convey your thoughts and distinguish them from other authors (Robbins, 2016). It is comprised of three elements ("What are the three elements," n.d.):

- Making declarative statements
- Avoiding casual language
- Demonstrating authority

These elements make your academic writing unique from other writers and present your thoughts in a professional manner.

Clarity

You want to ensure that your writing is precise so that readers have a clear understanding of your project. Proposals should exclude excessive jargon (technical terms), slang, and abbreviations. They should also make logical comparisons between ideas to prevent readers from getting confused or lost ("Academic Writer," n.d.). Here are some general tips for ensuring clarity in your writing:

- Using a term consistently throughout your paper (it refers to the same meaning throughout the document).
- Do not use excessive jargon or technical terms, and make sure you define any new terms.
- Draw comparisons between concepts to avoid ambiguity. This requires using proper word choice and sentence structure.

Conciseness

- Do not overuse passive voice

- Describe things precisely and "to the point."
- Assign one argument or idea per paragraph.
- Locate areas in your document to break up text into different paragraphs.
- Use a variety of sentence lengths.

Continuity

- Be mindful of how you use punctuation marks. This includes commas, dashes, and hyphens.
- Use transitional words (and, or, therefore, etc.) to maintain flow.

Tone

- Avoid using creative writing techniques, such as similes, metaphors, figurative language, and poetic devices.
- Do not use contractions in your proposal (can't, don't, etc.).
- Use words that reflect your involvement in research in your field.
- When writing about people, use respectful language.

Verb Tense

- Use appropriate verb tenses to reflect series of events and timelines in your proposal.

Unit-2

Research Design

Research design is the framework of research methods and techniques chosen by a researcher to conduct a study. The design allows researchers to sharpen the research methods suitable for the subject matter and set up their studies for success.

Creating a research topic explains the type of research (experimental, survey research, correlational, semi-experimental, review) and its sub-type (experimental design, research problem, descriptive case-study).

There are three main types of designs for research:

- Data collection
- Measurement
- Data Analysis

The research problem an organization faces will determine the design, not vice-versa. The design phase of a study determines which tools to use and how they are used.

Research Design Elements

Impactful research usually creates a minimum bias in data and increases trust in the accuracy of collected data. A design that produces the slightest margin of error in experimental research is generally considered the desired outcome. The essential elements are:

1. Accurate purpose statement
2. Techniques to be implemented for collecting and analyzing research
3. The method applied for analyzing collected details
4. Type of research methodology
5. Probable objections to research
6. Settings for the research study
7. Timeline
8. Measurement of analysis

Characteristics of Research Design

A proper design sets your study up for success. Successful research studies provide insights that are accurate and unbiased. You'll need to create a survey that meets all of the main characteristics of a design. There are four key characteristics:

- **Neutrality:** When you set up your study, you may have to make assumptions about the data you expect to collect. The results projected in the research should

be free from bias and neutral. Understand opinions about the final evaluated scores and conclusions from multiple individuals and consider those who agree with the results.

- **Reliability:** With regularly conducted research, the researcher expects similar results every time. You'll only be able to reach the desired results if your design is reliable. Your plan should indicate how to form research questions to ensure the standard of results.
- **Validity:** There are multiple measuring tools available. However, the only correct measuring tools are those which help a researcher in gauging results according to the objective of the research. The questionnaire developed from this design will then be valid.
- **Generalization:** The outcome of your design should apply to a population and not just a restricted sample. A generalized method implies that your survey can be conducted on any part of a population with similar accuracy.

The above factors affect how respondents answer the research questions, so they should balance all the above characteristics in a good design.

Exploratory Research

Exploratory Research: Types & Characteristics

Consider a scenario where a juice bar owner feels that increasing the variety of juices will enable an increase in customers. However, he is not sure and needs more information. The owner intends to conduct exploratory research to find out; hence, he decides to do exploratory research to find out if expanding their juices selection will enable him to get more customers or if there is a better idea.

Another example of exploratory research is a podcast survey template that can be used to collect feedback about the podcast consumption metrics both from existing listeners as well as other podcast listeners that are currently not subscribed to this channel. This helps the author of the podcast create curated content that will gain a larger audience. Let's explore this topic.

Content Index

Exploratory research: Definition

Exploratory research is defined as a research used to investigate a problem which is not clearly defined. It is conducted to have a better understanding of the existing problem, but will not provide conclusive results. For such a research, a researcher starts with a general idea and uses this research as a medium to identify issues, that can be the focus for future research. An important aspect here is that the researcher should be willing to change his/her direction subject to the revelation of new data or insight. Such a research is usually carried out when the problem is at a preliminary stage. It is often referred to as grounded theory approach or interpretive research as it used to answer questions like what, why and how.

Types and methodologies of Exploratory research

While it may sound difficult to research something that has very little information about it, there are several methods which can help a researcher figure out the best research design, data collection methods and choice of subjects. There are two ways in which research can be conducted namely primary and secondary.. Under these two types, there are multiple methods which can used by a researcher. The data gathered from these research can be qualitative or quantitative. Some of the most widely used research designs include the following:

Primary research methods

Primary research is information gathered directly from the subject. It can be through a group of people or even an individual. Such a research can be carried out directly by the researcher himself or can employ a third party to conduct it on their behalf. Primary research is specifically carried out to explore a certain problem which requires an in-depth study.

- **Surveys/polls**: Surveys/polls are used to gather information from a predefined group of respondents. It is one of the most important quantitative method.

Various types of surveys or polls can be used to explore opinions, trends, etc. With the advancement in technology, surveys can now be sent online and can be very easy to access. For instance, use of a survey app through tablets, laptops or even mobile phones. This information is also available to the researcher in real time as well. Nowadays, most organizations offer short length surveys and rewards to respondents, in order to achieve higher response rates.

For example: A survey is sent to a given set of audience to understand their opinions about the size of mobile phones when they purchase one. Based on such information organization can dig deeper into the topic and make business related decision.

- **Interviews:** While you may get a lot of information from public sources, but sometimes an in person interview can give in-depth information on the subject being studied. Such a research is a qualitative research method. An interview with a subject matter expert can give you meaningful insights that a generalized public source won't be able to provide. Interviews are carried out in person or on telephone which have open-ended questions to get meaningful information about the topic.

For example: An interview with an employee can give you more insights to find out the degree of job satisfaction, or an interview with a subject matter expert of quantum theory can give you in-depth information on that topic.

- **Focus groups:** Focus group is yet another widely used method in exploratory research. In such a method a group of people is chosen and are allowed to express their insights on the topic that is being studied. Although, it is important to make sure that while choosing the individuals in a focus group they should have a common background and have comparable experiences.

For example: A focus group helps a research identify the opinions of consumers if they were to buy a phone. Such a research can help the researcher understand what the consumer value while buying a phone. It may be screen size, brand value or even the dimensions. Based on which the organization can understand what are consumer buying attitudes, consumer opinions, etc.

- **Observations:** Observation research can be qualitative observation or quantitative observation. Such a research is done to observe a person and draw the finding from their reaction to certain parameters. In such a research, there is no direct interaction with the subject.

For example: An FMCG company wants to know how it's consumer react to the new shape of their product. The researcher observes the customers first reaction and collects the data, which is then used to draw inferences from the collective information.

Secondary research methods

Secondary research is gathering information from previously published primary research. In such a research you gather information from sources likes case studies, magazines, newspapers, books, etc.

- **Online research:** In today's world, this is one of the fastest way to gather information on any topic. A lot of data is readily available on the internet and the researcher can download it whenever he needs it. An important aspect to be noted for such a research is the genuineness and authenticity of the source websites that the researcher is gathering the information from.

For example: A researcher needs to find out what is the percentage of people that prefer a specific brand phone. The researcher just enters the information he needs in a search engine and gets multiple links with related information and statistics.

- **Literature research**: Literature research is one of the most inexpensive method used for discovering a hypothesis. There is tremendous amount of information available in libraries, online sources, or even commercial databases. Sources can include newspapers, magazines, books from library, documents from government agencies, specific topic related articles, literature, Annual reports, published statistics from research organizations and so on.

However, a few things have to be kept in mind while researching from these sources. Government agencies have authentic information but sometimes may come with a nominal cost. Also, research from educational institutions is generally overlooked, but in fact educational institutions carry out more number of research than any other entities.

Furthermore, commercial sources provide information on major topics like political agendas, demographics, financial information, market trends and information, etc.

For example: A company has low sales. It can be easily explored from available statistics and market literature if the problem is market related or organization related

or if the topic being studied is regarding financial situation of the country, then research data can be accessed through government documents or commercial sources.

- **Case study research:** Case study research can help a researcher with finding more information through carefully analyzing existing cases which have gone through a similar problem. Such analysis are very important and critical especially in today's business world. The researcher just needs to make sure he analyses the case carefully in regards to all the variables present in the previous case against his own case. It is very commonly used by business organizations or social sciences sector or even in the health sector.

For example: A particular orthopedic surgeon has the highest success rate for performing knee surgeries. A lot of other hospitals or doctors have taken up this case to understand and benchmark the method in which this surgeon does the procedure to increase their success rate.

Exploratory research: Steps to conduct a research

- **Identify the problem**: A researcher identifies the subject of research and the problem is addressed by carrying out multiple methods to answer the questions.
- **Create the hypothesis**: When the researcher has found out that there are no prior studies and the problem is not precisely resolved, the researcher will create a hypothesis based on the questions obtained while identifying the problem.
- **Further research**: Once the data has been obtained, the researcher will continue his study through descriptive investigation. Qualitative methods are used to further study the subject in detail and find out if the information is true or not.

Characteristics of Exploratory research

- They are not structured studies
- It is usually low cost, interactive and open ended.
- It will enable a researcher answer questions like what is the problem? What is the purpose of the study? And what topics could be studied?
- To carry out exploratory research, generally there is no prior research done or the existing ones do not answer the problem precisely enough.
- It is a time consuming research and it needs patience and has risks associated with it.
- The researcher will have to go through all the information available for the particular study he is doing.

- There are no set of rules to carry out the research per se, as they are flexible, broad and scattered.
- The research needs to have importance or value. If the problem is not important in the industry the research carried out is ineffective.
- The research should also have a few theories which can support its findings as that will make it easier for the researcher to assess it and move ahead in his study
- Such a research usually produces qualitative data, however in certain cases quantitative data can be generalized for a larger sample through use of surveys and experiments.

Advantages of Exploratory research

- The researcher has a lot of flexibility and can adapt to changes as the research progresses.
- It is usually low cost.
- It helps lay the foundation of a research, which can lead to further research.
- It enables the researcher understand at an early stage, if the topic is worth investing the time and resources and if it is worth pursuing.
- It can assist other researchers to find out possible causes for the problem, which can be further studied in detail to find out, which of them is the most likely cause for the problem.

Disadvantages of Exploratory research

- Even though it can point you in the right direction towards what is the answer, it is usually inconclusive.
- The main disadvantage of exploratory research is that they provide qualitative data. Interpretation of such information can be judgmental and biased.
- Most of the times, exploratory research involves a smaller sample, hence the results cannot be accurately interpreted for a generalized population.
- Many a times, if the data is being collected through secondary research, then there is a chance of that data being old and is not updated.

Importance of Exploratory research

Exploratory research is carried out when a topic needs to be understood in depth, especially if it hasn't been done before. The goal of such a research is to explore the problem and around it and not actually derive a conclusion from it. Such kind of research will enable a researcher to set a strong foundation for exploring his ideas, choosing the right research design and finding variables that actually are important

for the analysis. Most importantly, such a research can help organizations or researchers save up a lot of time and resources, as it will enable the researcher to know if it worth pursuing.

Descriptive Research

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.

When to use a descriptive research design

Descriptive research is an appropriate choice when the research aim is to identify characteristics, frequencies, trends, and categories.

It is useful when not much is known yet about the topic or problem. Before you can research why something happens, you need to understand how, when and where it happens.

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 necessary to 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 or interesting cases that challenge assumptions, add complexity, or reveal something new about a research problem.

Diagnostic Design:

Diagnostic refers to scientific differentiation among various conditions

• phenomenon for the purpose of accurately classifying these conditions. In its broadest sense diagnosis corresponds to the fact-finding aspects of

• clinical practice.

Its objective includes screening and classification personality description,

• prediction of outcome and attainment of insight by the client. The diagnostic research paradigm represents the most typical and simple

• problem solving strategy of the helper faced with problems and crises on the job.

It consists of:

• (a) the emergence of a problem

(b) a diagnosis of its causes

(c) formulation of all possible avenues of remediation, and

(d) recommendations for a possible solution.

Data for diagnosis can be obtained in four major ways:

- (i) A Case History or Interview

(ii) Clinical Observation

(iii) Informal Testing, and

(iv) Formal Standardised Testing.

Case Study Method

Case study method is concerned with everything that is significant in the history or development of the case. The purpose is to understand the life cycle, or an important part of the life cycle, or an important part of life of an individual unit. This unit may be a person, a family, a group, a social institution or an entire community. This method probes deeply and intensively, analyses interactions between the factors that produce change or growth. It emphasis longitudinal orgenetic approach showing development over a period of time. In each case, the element of typicalness is the focus of attention, with emphasis On the many factors that characterize the type.

- The characteristics of a good case study include an adequate data which is

valid, continues, carefully synthesized, confidential and which should be useful for follow up.

Interview Method

The nature of the personal relationship between interviewer and subject

- requires an expertness and sensibility that might well be called an art. It requires a skilful interviewer to obtain a maximum amount of useful data.

- The interviewers must try to establish a feeling of mental trust with the

- persons being interviewed being careful not to ask questions that might alarm them.

His attitude should convey a spirit of cooperation, acceptance and empathy

- while maintaining a degree of professional objectivity to guard against excessive emotional involvement and consequent ineffectiveness.

Clinical Observation:

Many attributes are in adequately identified through other standardiscal test

instruments or through interviews. The skilful diagnostician should be able to detect through observation or the

behaviour and through the proficient use of informal test.

The informal tests and observations of behaviour provide an opportunity to

• corroborate findings of the other two areas of assessment. Diagnostic tests provide a microscopic view of the component elements of

• some area of performance. Such tests enable the diagnostician to analyze the individual's functioning within specific subskill areas and supply direction for remediation.

Formal standardised tests are useful in testing particular aspects of behaviour

• however, it is important to know the limitations of the test and to use the information in proper perspectives. Collaborative Diagnosis Many of the problems which have been attached through field experiment

• have infact been problems related to highly developed professional skills in such areas as human relations training, therapy, community organizations etc. For such problems there are already available highly skilled professional

• people who may be called upon to conduct the manipulation. Once the basic role relationships are worked out, there should be a

• collaborative diagnosis of the situation by the researcher and at least some part of the client organization. 8 The purpose of this diagnosis is to assess various factors that will be involved

• in executing the research design the resistance that may be encountered, the dynamics of the situation, in regard to the problem of bring about a change. The ways of going about such a collaborative diagnosis will varies

• tremendously depending upon the problem to be studied, the setting, etc., all relevant experiences and information should be used, but whatever the source, a useful diagnosis will have to be formulated in theoretical terms. The next step should be joint planning based on an adequate diagnosis of the

• action that must be taken in order to manipulate the independent variable and to control other possibly confounding factors. Among the descriptive and diagnostic designs, descriptive research is

• particularly appropriate in the behavioural sciences; because many types of behaviour that interest the researcher can not be arranged in a realistic setting. Descriptive research, some times known as non experimental research, deals

• with the relationships between variables, the testing of hypothesis, and the development of generalization, principles, or theories that have universal validity. It concerns with functional relationships.

Experimental research

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.

If you don't have enough data to support your decisions, you must first determine the facts. Experimental research gathers the data necessary to help you make better decisions.

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 notable cause and effect.

You can conduct experimental research in the following situations:

- Time is a vital factor in establishing a relationship between cause and effect.
- Invariable behavior between cause and effect.
- You wish to understand the importance of the cause 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.

Methods of sampling from a population

We are currently in the process of updating this chapter and we appreciate your patience whilst this is being completed.

It would normally be impractical to study a whole population, for example when doing a questionnaire survey. Sampling is a method that allows researchers to infer information about a population based on results from a subset of the population, without having to investigate every individual. Reducing the number of individuals in a study reduces the cost and workload, and may make it easier to obtain high quality information, but this has to be balanced against having a large enough sample size with enough power to detect a true association. (Calculation of sample size is addressed in section 1B (statistics) of the Part A syllabus.)
If a sample is to be used, by whatever method it is chosen, it is important that the individuals selected are representative of the whole population. This may involve specifically targeting hard to reach groups. For example, if the electoral roll for a town was used to identify participants, some people, such as the homeless, would not be registered and therefore excluded from the study by default.

There are several different sampling techniques available, and they can be subdivided into two groups: probability sampling and non-probability sampling. In probability (random) sampling, you start with a complete sampling frame of all eligible individuals from which you select your sample. In this way, all eligible individuals have a chance of being chosen for the sample, and you will be more able to generalise the results from your study. Probability sampling methods tend to be more time-consuming and expensive than non-probability sampling. In non-probability (non-random) sampling, you do not start with a complete sampling frame, so some individuals have no chance of being selected. Consequently, you cannot estimate the effect of sampling error and there is a significant risk of ending up with a non-representative sample which produces non-generalisable results. However, non-probability sampling methods tend to be cheaper and more convenient, and they are useful for exploratory research and hypothesis generation.

Probability Sampling Methods

1. Simple random sampling

In this case each individual is chosen entirely by chance and each member of the population has an equal chance, or probability, of being selected. One way of obtaining a random sample is to give each individual in a population a number, and then use a table of random numbers to decide which individuals to include.[1] For example, if you have a sampling frame of 1000 individuals, labelled 0 to 999, use groups of three digits from the random number table to pick your sample. So, if the first three numbers from the random number table were 094, select the individual labelled "94", and so on.

As with all probability sampling methods, simple random sampling allows the sampling error to be calculated and reduces selection bias. A specific advantage is that it is the most straightforward method of probability sampling. A disadvantage of simple random sampling is that you may not select enough individuals with your characteristic of interest, especially if that characteristic is uncommon. It may also be difficult to define a complete sampling frame and inconvenient to contact them, especially if different forms of contact are required (email, phone, post) and your sample units are scattered over a wide geographical area.

2. Systematic sampling

Individuals are selected at regular intervals from the sampling frame. The intervals are chosen to ensure an adequate sample size. If you need a sample size *n* from a population of size *x*, you should select every x/n^{th} individual for the sample. For example, if you wanted a sample size of 100 from a population of 1000, select every $1000/100 = 10^{th}$ member of the sampling frame.
Systematic sampling is often more convenient than simple random sampling, and it is easy to administer. However, it may also lead to bias, for example if there are underlying patterns in the order of the individuals in the sampling frame, such that the sampling technique coincides with the periodicity of the underlying pattern. As a hypothetical example, if a group of students were being sampled to gain their opinions on college facilities, but the Student Record Department's central list of all students was arranged such that the sex of students alternated between male and female, choosing an even interval (e.g. every 20^{th} student) would result in a sample of all males or all females. Whilst in this example the bias is obvious and should be easily corrected, this may not always be the case.

3. Stratified sampling

In this method, the population is first divided into subgroups (or strata) who all share a similar characteristic. It is used when we might reasonably expect the measurement of interest to vary between the different subgroups, and we want to ensure

representation from all the subgroups. For example, in a study of stroke outcomes, we may stratify the population by sex, to ensure equal representation of men and women. The study sample is then obtained by taking equal sample sizes from each stratum. In stratified sampling, it may also be appropriate to choose non-equal sample sizes from each stratum. For example, in a study of the health outcomes of nursing staff in a county, if there are three hospitals each with different numbers of nursing staff (hospital A has 500 nurses, hospital B has 1000 and hospital C has 2000), then it would be appropriate to choose the sample numbers from each hospital *proportionally* (e.g. 10 from hospital A, 20 from hospital B and 40 from hospital C). This ensures a more realistic and accurate estimation of the health outcomes of nurses across the county, whereas simple random sampling would over-represent nurses from hospitals A and B. The fact that the sample was stratified should be taken into account at the analysis stage.

Stratified sampling improves the accuracy and representativeness of the results by reducing sampling bias. However, it requires knowledge of the appropriate characteristics of the sampling frame (the details of which are not always available), and it can be difficult to decide which characteristic(s) to stratify by.

4. Clustered sampling

In a clustered sample, subgroups of the population are used as the sampling unit, rather than individuals. The population is divided into subgroups, known as clusters, which are randomly selected to be included in the study. Clusters are usually already defined, for example individual GP practices or towns could be identified as clusters. In single-stage cluster sampling, all members of the chosen clusters are then included in the study. In two-stage cluster sampling, a selection of individuals from each cluster is then randomly selected for inclusion. Clustering should be taken into account in the analysis. The General Household survey, which is undertaken annually in England, is a good example of a (one-stage) cluster sample. All members of the selected households (clusters) are included in the survey.[1]

Cluster sampling can be more efficient that simple random sampling, especially where a study takes place over a wide geographical region. For instance, it is easier to contact lots of individuals in a few GP practices than a few individuals in many different GP practices. Disadvantages include an increased risk of bias, if the chosen clusters are not representative of the population, resulting in an increased sampling error.

Non-Probability Sampling Methods

1. Convenience sampling

Convenience sampling is perhaps the easiest method of sampling, because participants are selected based on availability and willingness to take part. Useful results can be obtained, but the results are prone to significant bias, because those who volunteer to take part may be different from those who choose not to (volunteer bias), and the sample may not be representative of other characteristics, such as age or sex. Note: volunteer bias is a risk of all non-probability sampling methods.

2. Quota sampling

This method of sampling is often used by market researchers. Interviewers are given a quota of subjects of a specified type to attempt to recruit. For example, an interviewer might be told to go out and select 20 adult men, 20 adult women, 10 teenage girls and 10 teenage boys so that they could interview them about their television viewing. Ideally the quotas chosen would proportionally represent the characteristics of the underlying population.

Whilst this has the advantage of being relatively straightforward and potentially representative, the chosen sample may not be representative of other characteristics that weren't considered (a consequence of the non-random nature of sampling). [2]

3. Judgement (or Purposive) Sampling

Also known as selective, or subjective, sampling, this technique relies on the judgement of the researcher when choosing who to ask to participate. Researchers may implicitly thus choose a "representative" sample to suit their needs, or specifically approach individuals with certain characteristics. This approach is often used by the media when canvassing the public for opinions and in qualitative research.

Judgement sampling has the advantage of being time-and cost-effective to perform whilst resulting in a range of responses (particularly useful in qualitative research). However, in addition to volunteer bias, it is also prone to errors of judgement by the researcher and the findings, whilst being potentially broad, will not necessarily be representative.

4. Snowball sampling

This method is commonly used in social sciences when investigating hard-to-reach groups. Existing subjects are asked to nominate further subjects known to them, so the sample increases in size like a rolling snowball. For example, when carrying out a survey of risk behaviours amongst intravenous drug users, participants may be asked to nominate other users to be interviewed.

Snowball sampling can be effective when a sampling frame is difficult to identify. However, by selecting friends and acquaintances of subjects already investigated, there is a significant risk of selection bias (choosing a large number of people with similar characteristics or views to the initial individual identified).

Bias in sampling

There are five important potential sources of bias that should be considered when selecting a sample, irrespective of the method used. Sampling bias may be introduced when:[1]

1. Any pre-agreed sampling rules are deviated from
2. People in hard-to-reach groups are omitted
3. Selected individuals are replaced with others, for example if they are difficult to contact
4. There are low response rates
5. An out-of-date list is used as the sample frame (for example, if it excludes people who have recently moved to an area)

Further potential problems with sampling strategies are covered in chapter 8 of this section (*"Sources of variation, its measurement and control"*).

What is Scale ?

A series of categories or items arranged in a progressive manner (in a continuous spectrum) based on magnitude or value is called 'scale'. Different responses of individuals are quantitatively placed on such scales. Every scale includes a lowest point, a highest point and some intermediate points (representing a particular activity or attitude). If the first point or item on the scale is of highest degree, the second point is higher in nature, than the third point; consequently, the third pint is higher in nature than the fourth point and so on.

The process of developing a continuum for placing the measured objects is called scaling. These measured objects may involve interests, opinions. attitude and other responses of the individuals. Thus, the next operation involved in measurement is the scaling process. In this, different qualitative aspects are associated with quantitative components.

Scale of Measurement in Research Methodology :

Measurement is followed by scaling procedure in a research activity. Researchers always face the problem of how to measure or quantify abstract concepts and how to relate one individual's response to that with another response. Hence, the problem is two-fold. First is, how to quantify a response in absolute terms, and the second is, how to relate it with other responses. This problem is resolved by scaling process, or assigning the numbers or values to responses and positioning them on a common scale. Scaling can then be defined as, "creating a continuum with two extreme limits and few immediate values between them". Hence, scaling is the process or procedure of quantifying and measuring abstract concepts like attitudes, opinions, emotions, etc. Scaling also quantitatively determines the position of an individual in a range of values.

Definition of Scaling

According to Edwards, scaling can be defined as,

"Procedure for the assignment of numbers (or other symbols) to a property of objects in order to impart some of the characteristics of numbers to the properties in question"

In a scale, the values are progressively placed according their value or magnitude in such a manner that moving forward in a scale will depict an item to be always higher than the immediate lower one.

Process of Scaling

Among all the commonly used direct techniques to measure attitudes, like word-association tests, sentence-completion tests, story-telling. performance of 'objective' tasks. physiological reactions, etc., the 'self-report method' is the most popular. In this method, the respondent is asked to give his opinion about a particular object freely. Described below is a step-by-step guide on how to construct a self-report :

1) Gathering Related Variables :

Once the subject of inquiry has been decided, it is time to collect all possible variables and related statements, which are then examined to remove those that are imprecise and ambiguous, so that the questionnaire is easily comprehensible and accurately answerable. Utmost care should be taken to ensure the use of simple, easily understandable language,

2) Shortlisting the Variables :

Once the pool of variables or questions has been finalized, it is time for scrutinizing the variables to ensure that the question or variable can be included in the scale. Some of the criteria for selecting the variables are as follows :

- Relevant with the enquiry,
- Indispensible,
- Easily comprehensible,
- Accurately answerable,
- Covers all the possible dimensions and angles of the subject of enquiry.

Before finalizing, its efficacy is needed to be tested either through a sample tested on a dummy or by assigning the task to a group of experts.

3) Formation of Scale :

Once the variables have been finalized, the scale is then tested for its validity and reliability, before being put to use in a survey.

Scale Construction Techniques

Techniques used in developing set of items or categories in a progressive manner (in a continuous spectrum) so as to measure an individual's attitude towards given event or object, refers to scale construction techniques. Different scale construction techniques in research methodology used in scaling are as follows :

1) Arbitrary Scales :

In this technique, large number of items or categories, which are able to represent the concept under study are collected and (then) measured in terms of same characteristic. Subjective selection of the researcher plays an important role here. Firstly, few related statements or items are selected by the researcher, and then these statements are filtered for being included in the measuring instrument. A list of filtered statements is presented to the respondents and they are required to tick the most suitable one.

2) Cumulative Scales :

A series of statements is used in developing scale for the respondents in cumulative scales. These are also called 'Louis Guttman's scales. These scales require

respondents to show their agreement or disagreement over the statements. These scales are called cumulative scales because a cumulative series is formed by different statements here. For example, the confirmation of an extreme position item should also lead to the confirmation of less extreme position items.

3) Consensus Scaling :

This method of scaling was given by L.L. Thurstone. In this, a panel of judges uses different attributes like attitude. relevance and ambiguity factor so as to select items or categories. Generally, differential scales are created by this method, which focus on measuring attitudes of respondents concerning specific issues like religion, war, etc. These are the most suitable scales for recording only a single attitude of the given concept. A high level of cost and effort is needed to create such scales. Judges' own attitudes may affect their functioning of assigning values to different statements.

4) Item Analysis :

In this type of scale construction technique, a group of respondents is given a test created by selecting different individual items. Following steps are involved in this technique :

- Firstly, large number of items expressing their positive or negative aspects is selected.
- Then, a group of respondents assigns scores to them.
    - Then for each respondent, total scores for all items are calculated. Then using these scores, they are divided into four constructs. Two middle constructs are then eliminated from the calculation.
    - Then average score is calculated separately for each item.
    - T-test is then used to compare average of items of different groups to select Items with noteworthy t-values.

In this, different statements expressing the positive or negative attitude towards the selected object are used to create the scale and the respondents are asked to check the most suitable one. Thus, the agreement or disagreement of respondents over different

statements is recorded. The most common example of item analysis is Likert scale. In this, several degrees of agreement or disagreement (generally five, but sometimes three or seven) are used by the respondents to respond to 'different statements. This type of scale is very easy to develop.

5) Factor Scales :

Several different techniques are involved in constructing such scales. The purpose of using different scales is to study multi-dimensional items, to find the relations between different dimensions and to identify any other dimensions of the items. In the end, limited set of factors are available each having interrelated dimensions. Scales developed through this technique are as follows :

i) Q-Sort Technique :

In this, the respondent is asked to sort the different given statements into prefixed categories.

ii) Semantic Differential (SD) :

Factor analysis of assumed interval scales results in semantic differential scales.

iii) Stapel Scale :

It is a non-verbal rating scale having even number of scales. Single adjectives are used in items or categories. It is necessary to measure both the dimensions consequently. There is no need of assuming the ratings or the interval equality.

iv) Mull-Dimensional Scaling :

In multi dimensional scaling, a multi-dimensional space is assumed to be present. A set of techniques are developed to deal with such multi-dimensional space.

v) Standardized Instruments :

In this, an available measuring instrument is selected for data collection. No new measuring instruments are developed. The opinion of expert is taken for selecting such standardized instruments.

Criteria for Good Scale

Any scale must satisfy the following criteria to prove its efficacy :

1) Validity :

Validity is the most critical aspect. It establishes or indicates the extent to which the scale does what it is supposed to do. Does it measure what it was intended to measure? In other words, whether or not the test has been useful or worthwhile.

2) Reliability :

This indicates whether or not the findings of a measuring instrument can be relied upon. To know this, the researcher needs to answer some questions, such as, Are the outcomes accurate? Can they be replicated? Are the findings consistent?, etc. While reliability contributes to validity, it is not true the other way round. A reliable instrument may or may not be necessarily valid.

3) Practicality :

Practicality refers to ease and economy, of constructing the measuring instrument as well as administration and interpretation of the test outcome. It measures the achievability, and practicality of an instrument. The benefits or results should justify the costs.

4) Sensitivity :

This refers to how well the test is standardized. It tests how much an instrument is able to measure accurately. For example, a test which requires respondents to merely say 'yes' or 'no', may not be very sensitive. On the other hand, if the respondent is asked to rule on a '5' or "7"-point scale, the scale may be considered to be highly sensitive. It may however be noted that all scales need not be highly sensitive. It largely depends upon the requirement of the test.

5) Generalisability :

This refers to whether or not one can generalize the findings, i.e., whether or not the findings can be applied to both similar and different situations. Whether or not the sample or respondents selected can be said to represent larger population. For example, exit polls can be said to have a high degree of generalisability.

6) Economy :

This factor emphasizes on the aspect that whether the instrument, is economical to be constructed and conducted. Tests are expensive to develop and administer. Therefore, the results or benefits of the tests must justify or warrant the costs.

7) Convenience :

A measuring instrument is convenient, if it is easy to conduct. The instruments that provide guidelines to use are much more convenient than those instruments that do not include this feature. Generally, it is considered that the requirement for convenience increases with the level of complexity of a measuring instrument.

Importance of Scaling

In business or management research, scaling is very crucial for the research process. It helps in measuring. and analyzing attitudes of different individuals. The exact behavior of an individual is reflected by such attitude analysis. Number of attitude measuring scales has been developed by researchers. For example, in order to measure the attitude of an individual about a particular tourist place, product or election candidate, i.e., if he visits, buys or votes, respectively, a suitable scale is developed. The different facts describing the importance of scaling are as follows :

1) Attitude Scoring :
Scaling is particularly used for altitude scoring of an individual. In scaling, with the help of an individual's responses a particular number or point is selected on given scale, which represents the attitude of that individual.

2) Broad Application :
Different management research processes as well as scientific inventions use scaling as their crucial element. Data collection methods like interviews, observations, surveys, etc,, also use scaling for attitude measurement of respondents.

3) Hypothesis Testing :
Scaling is also functional in hypothesis testing. Without effective measure of attitudes of different respondents, it is not easy to test hypothesis about the population.

4) Dimension Checking :

Scaling is useful in determining the dimensional aspects of different quantitative concepts or items. It helps in checking whether a particular item is single dimensional or multi-dimensional.

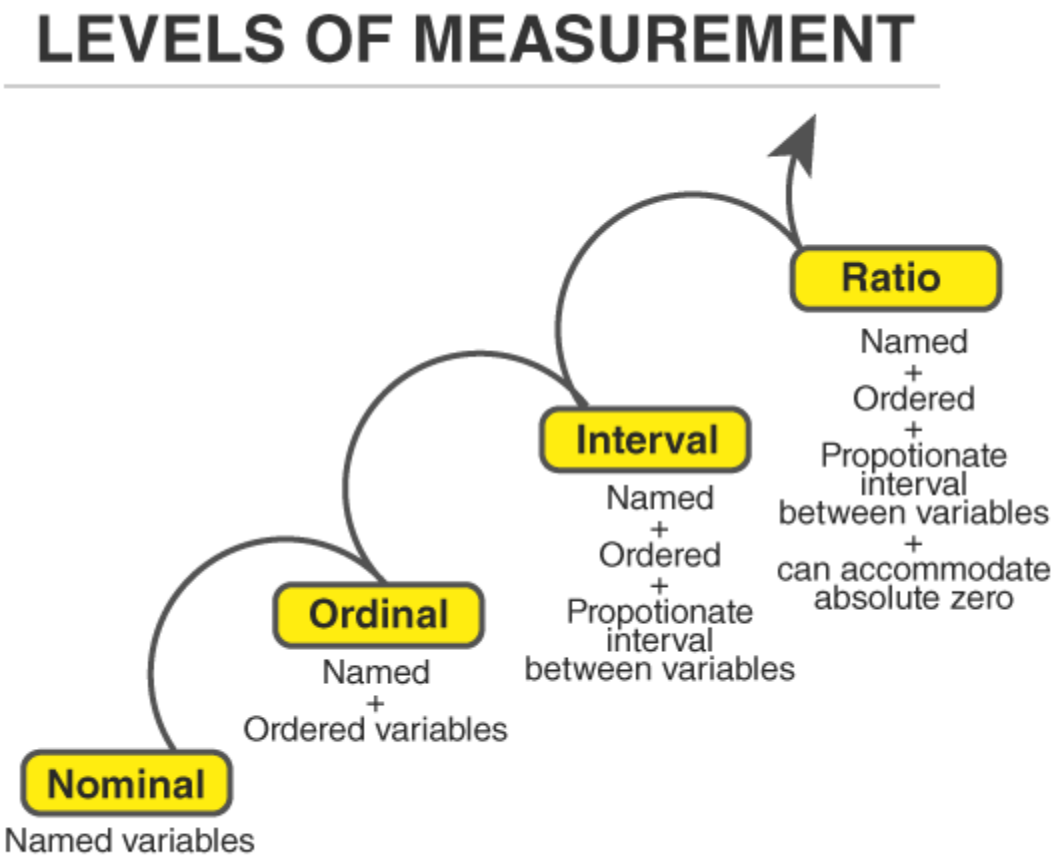
5) Others :

- It is an essential element of the exploratory research.
- It is used to check whether a set of questions is measuring single aspect or multiple aspects.

Levels of Measurements

There are four different scales of measurement. The data can be defined as being one of the four scales. The four types of scales are:

- Nominal Scale
- Ordinal Scale
- Interval Scale
- Ratio Scale



Nominal Scale

A nominal scale is the 1st level of measurement scale in which the numbers serve as "tags" or "labels" to classify or identify the objects. A nominal scale usually deals with the non-numeric variables or the numbers that do not have any value.

Characteristics of Nominal Scale

- A nominal scale variable is classified into two or more categories. In this measurement mechanism, the answer should fall into either of the classes.
- It is qualitative. The numbers are used here to identify the objects.
- The numbers don't define the object characteristics. The only permissible aspect of numbers in the nominal scale is "counting."

Example:

An example of a nominal scale measurement is given below:

What is your gender?

M- Male

F- Female

Here, the variables are used as tags, and the answer to this question should be either M or F.

Ordinal Scale

The ordinal scale is the 2nd level of measurement that reports the ordering and ranking of data without establishing the degree of variation between them. Ordinal represents the "order." Ordinal data is known as qualitative data or categorical data. It can be grouped, named and also ranked.

Characteristics of the Ordinal Scale

- The ordinal scale shows the relative ranking of the variables
- It identifies and describes the magnitude of a variable
- Along with the information provided by the nominal scale, ordinal scales give the rankings of those variables
- The interval properties are not known
- The surveyors can quickly analyse the degree of agreement concerning the identified order of variables

Example:

- Ranking of school students – 1st, 2nd, 3rd, etc.
- Ratings in restaurants

- Evaluating the frequency of occurrences
    - Very often
    - Often
    - Not often
    - Not at all

Assessing the degree of agreement

- Totally agree
- Agree
- Neutral
- Disagree
- Totally disagree

Interval Scale

The interval scale is the 3rd level of measurement scale. It is defined as a quantitative measurement scale in which the difference between the two variables is meaningful. In other words, the variables are measured in an exact manner, not as in a relative way in which the presence of zero is arbitrary.

Characteristics of Interval Scale:

- The interval scale is quantitative as it can quantify the difference between the values
- It allows calculating the mean and median of the variables
- To understand the difference between the variables, you can subtract the values between the variables
- The interval scale is the preferred scale in Statistics as it helps to assign any numerical values to arbitrary assessment such as feelings, calendar types, etc.

Example:

- Likert Scale
- Net Promoter Score (NPS)
- Bipolar Matrix Table

Ratio Scale

The ratio scale is the 4th level of measurement scale, which is quantitative. It is a type of variable measurement scale. It allows researchers to compare the differences or intervals. The ratio scale has a unique feature. It possesses the character of the origin or zero points.

Characteristics of Ratio Scale:

- Ratio scale has a feature of absolute zero
- It doesn't have negative numbers, because of its zero-point feature
- It affords unique opportunities for statistical analysis. The variables can be orderly added, subtracted, multiplied, divided. Mean, median, and mode can be calculated using the ratio scale.
- Ratio scale has unique and useful properties. One such feature is that it allows unit conversions like kilogram – calories, gram – calories, etc.

Example:

An example of a ratio scale is:

What is your weight in Kgs?

- Less than 55 kgs
- 55 – 75 kgs
- 76 – 85 kgs
- 86 – 95 kgs
- More than 95 kgs

What is Research Ethics?

Research ethics provides guidelines for the responsible conduct of research. In addition, it educates and monitors scientists conducting research to ensure a high ethical standard. The following is a general summary of some ethical principles:

Honesty:

Honestly report data, results, methods and procedures, and publication status. Do not fabricate, falsify, or misrepresent data.

Objectivity:

Strive to avoid bias in experimental design, data analysis, data interpretation, peer review, personnel decisions, grant writing, expert testimony, and other aspects of research.

Integrity:

Keep your promises and agreements; act with sincerity; strive for consistency of thought and action.

Carefulness:

Avoid careless errors and negligence; carefully and critically examine your own work and the work of your peers. Keep good records of research activities.

Openness:

Share data, results, ideas, tools, resources. Be open to criticism and new ideas.

Respect for Intellectual Property:

Honor patents, copyrights, and other forms of intellectual property. Do not use unpublished data, methods, or results without permission. Give credit where credit is due. Never plagiarize.

Confidentiality:

Protect confidential communications, such as papers or grants submitted for publication, personnel records, trade or military secrets, and patient records.

Responsible Publication:

Publish in order to advance research and scholarship, not to advance just your own career. Avoid wasteful and duplicative publication.

Responsible Mentoring:

Help to educate, mentor, and advise students. Promote their welfare and allow them to make their own decisions.

Respect for Colleagues:

Respect your colleagues and treat them fairly.

Social Responsibility:

Strive to promote social good and prevent or mitigate social harms through research, public education, and advocacy.

Non-Discrimination:

Avoid discrimination against colleagues or students on the basis of sex, race, ethnicity, or other factors that are not related to their scientific competence and integrity.

Competence:

Maintain and improve your own professional competence and expertise through lifelong education and learning; take steps to promote competence in science as a whole.

Legality:

Know and obey relevant laws and institutional and governmental policies.

Animal Care:

Show proper respect and care for animals when using them in research. Do not conduct unnecessary or poorly designed animal experiments.

Human Subjects Protection:

When conducting research on human subjects, minimize harms and risks and maximize benefits; respect human dignity, privacy, and autonomy.

Unit-3

What is Data?

Data is a collection of facts, figures, objects, symbols, and events gathered from different sources. **Organizations collect data with various data collection methods to make better decisions**. Without data, it would be difficult for organizations to make appropriate decisions, so data is collected from different audiences at various points in time.

For instance, an organization must collect data on product demand, customer preferences, and competitors before launching a new product. If data is not collected beforehand, the organization's newly launched product may fail for many reasons, such as less demand and inability to meet customer needs.

Although data is a valuable asset for every organization, it does not serve any purpose until analyzed or processed to get the desired results.

What are Data Collection Methods?

Data collection methods are techniques and procedures used to gather information for research purposes. These methods can range from simple self-reported surveys to more complex experiments and can involve either quantitative or qualitative approaches to data gathering.

Some common data collection methods include surveys, interviews, observations, focus groups, experiments, and secondary data analysis. The data collected through these methods can then be analyzed and used to support or refute research hypotheses and draw conclusions about the study's subject matter.

Importance of Data Collection Methods

Data collection methods play a crucial role in the research process as they determine the quality and accuracy of the data collected. Here are some mejor importance of data collection methods.

- Determines the quality and accuracy of collected data.
- Ensures that the data is relevant, valid, and reliable.
- Helps reduce bias and increase the representativeness of the sample.
- Essential for making informed decisions and accurate conclusions.
- Facilitates achievement of research objectives by providing accurate data.
- Supports the validity and reliability of research findings.

The importance of data collection methods cannot be overstated, as it plays a key role in the overall success and validity of the research study.

Types of Data Collection Methods

The choice of data collection method depends on the research question being addressed, the type of data needed, and the resources and time available. You can categorize data collection methods into primary methods of data collection and secondary methods of data collection.

Primary Data Collection Methods

Primary data is collected from first-hand experience and is not used in the past. The data gathered by primary data collection methods are specific to the research's motive and highly accurate.

Primary data collection methods can be divided into two categories: quantitative methods and qualitative methods.

Quantitative Methods:

Quantitative techniques for market research and demand forecasting usually use statistical tools. In these techniques, demand is forecasted based on historical data. These methods of primary data collection are generally used to make long-term forecasts. Statistical methods are highly reliable as subjectivity is minimal in these methods.

Time Series Analysis: The term time series refers to a sequential order of values of a variable, known as a trend, at equal time intervals. Using patterns, an organization can predict the demand for its products and services for the projected time.

Smoothing Techniques: In cases where the time series lacks significant trends, smoothing techniques can be used. They eliminate a random variation from the historical demand. It helps in identifying patterns and demand levels to estimate future demand. The most common methods used in smoothing demand forecasting techniques are the simple moving average method and the weighted moving average method.

Barometric Method: Also known as the leading indicators approach, researchers use this method to speculate future trends based on current developments. When the past events are considered to predict future events, they act as leading indicators.

Qualitative Methods:

Qualitative methods are especially useful in situations when historical data is not available. Or there is no need of numbers or mathematical calculations. Qualitative research is closely associated with words, sounds, feeling, emotions, colors, and other elements that are non-quantifiable. These techniques are based on experience, judgment, intuition, conjecture, emotion, etc.

Quantitative methods do not provide the motive behind participants' responses, often don't reach underrepresented populations, and span long periods to collect the data. Hence, it is best to combine quantitative methods with qualitative methods.

Surveys: Surveys are used to collect data from the target audience and gather insights into their preferences, opinions, choices, and feedback related to their products and services. Most survey software often has a wide range of question types to select.

You can also use a ready-made survey template to save time and effort. Online surveys can be customized as per the business's brand by changing the theme, logo, etc. They can be distributed through several distribution channels such as email, website, offline app, QR code, social media, etc. Depending on the type and source of your audience, you can select the channel.

Once the data is collected, survey software can generate various reports and run analytics algorithms to discover hidden insights. A survey dashboard can give you statistics related to response rate, completion rate, filters based on demographics, export and sharing options, etc. Integrating survey builder with third-party apps can maximize the effort spent on online real-time data collection.

Polls: Polls comprise one single or multiple-choice question. You can go for polls when it is required to have a quick pulse of the audience's sentiments. Because they are short in length, it is easier to get responses from people.

Like surveys, online polls can also be embedded into various platforms. Once the respondents answer the question, they can also be shown how they stand compared to others' responses.

Interviews: In this method, the interviewer asks the respondents face-to-face or by telephone. In face-to-face interviews, the interviewer asks a series of questions to the interviewee in person and notes down responses. If it is not feasible to meet the person, the interviewer can go for a telephone interview. This form of data collection is suitable for only a few respondents. It is too time-consuming and tedious to repeat the same process if there are many participants.

Delphi Technique: In delphi method, market experts are provided with the estimates and assumptions of forecasts made by other experts in the industry. Experts may reconsider and revise their estimates and assumptions based on the information provided by other experts. The consensus of all experts on demand forecasts constitutes the final demand forecast.

Focus Groups: In a focus group, a small group of people, around 8-10 members, discuss the common areas of the problem. Each individual provides his insights on the issue concerned. A moderator regulates the discussion among the group members. At the end of the discussion, the group reaches a consensus.

Questionnaire: A questionnaire is a printed set of questions, either open-ended or closed-ended. The respondents must answer based on their knowledge and experience with the issue. The questionnaire is a part of the survey, whereas the questionnaire's end goal may or may not be a survey.

Secondary Data Collection Methods

Secondary data is the data that has been used in the past. The researcher can obtain data from the data sources, both internal and external, to the organization.

Internal sources of secondary data:

- Organization's health and safety records
- Mission and vision statements
- Financial Statements
- Magazines
- Sales Report
- CRM Software
- Executive summaries

External sources of secondary data:

- Government reports
- Press releases
- Business journals
- Libraries
- Internet

The secondary data collection methods can also involve quantitative and qualitative techniques. Secondary data is easily available and hence, less time-consuming and expensive than primary data. However, with the secondary data collection methods, the authenticity of the data gathered cannot be verified.

The secondary data collection methods can also involve quantitative and qualitative techniques. Secondary data is easily available, less time-consuming, and more expensive than primary data. However, with the secondary data collection methods, the authenticity of the data gathered cannot be verified.

Regardless of the data collection method of your choice, there must be direct communication with decision-makers so that they understand and commit to acting according to the results.

For this reason, we must pay special attention to the analysis and presentation of the information obtained. Remember that these data must be useful and functional to us, so the data collection method used has much to do with it.

Conclusion

The conclusion you obtain from your investigation will set the course of the company's decision-making, so present your report clearly, and list the steps you followed to obtain those results.

Make sure that whoever will take the corresponding actions understands the importance of the information collected and that it gives them the solutions they expect.

Remember that at QuestionPro, we can help you collect data easily and efficiently. Request a demo and learn about all the tools we have for you.

Data Analysis and Interpretation

Data Analysis & Interpretation

Data analysis is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decisionmaking. The purpose of Data Analysis is to extract useful information from data and taking the decision based upon the data analysis.

Data Analysis Process

The Data Analysis Process is nothing but gathering information by using a proper application or tool which allows you to explore the data and find a pattern in it. Based on that information and data, you can make decisions, or you can get ultimate conclusions.

Data Analysis consists of the following phases:

- Data Requirement Gathering
- Data Collection

Data Cleaning •

Data Analysis •

Data Interpretation•

Data Visualization•

Data Interpretation

Data interpretation is the process of reviewing data through some predefined processes which will help assign some meaning to the data and arrive at a relevant conclusion. It involves taking the result of data analysis. Data analysis is the process of ordering, categorizing, manipulating, and summarizing data to obtain answers to research questions. It is usually the first step taken towards data interpretation.

It is evident that the interpretation of data is very important, and as such needs to be done properly. Therefore, researchers have identified some data interpretation methods to aid this process.

Data Interpretation Methods

Data interpretation methods are how analysts help people make sense of numerical data that has been collected, analyzed and presented. Data, when collected in raw form, may be difficult for the layman to understand, which is why analysts need to break down the information gathered so that others can make sense of it. For example, when founders are pitching to potential investors, they must interpret data (e.g. market size, growth rate, etc.) for better understanding. There are 2 main methods in which this can be done, namely; quantitative methods and qualitative methods.

Qualitative Data Interpretation Method

The qualitative data interpretation method is used to analyze qualitative data, which is also known as categorical data. This method uses texts, rather than numbers or patterns to describe data. Qualitative data is usually gathered using a wide variety of person-toperson techniques, which may be difficult to analyze compared to the quantitative research method. Unlike the quantitative data which can be analyzed directly after it has been collected and sorted, qualitative data needs to first be coded into numbers before it can be analyzed. This is because texts are usually cumbersome, and will take more time and result in a lot of errors if analyzed in its original state. Coding done by the analyst should also be documented so that it can be reused by others and also analyzed. There are 2 main types of qualitative data, namely; nominal and ordinal data. These 2 data types are both interpreted using the same method, but ordinal data interpretation is quite easier than that of nominal data. In most cases, ordinal data is usually labelled with numbers during the process of data collection, and coding may not be required. This is different from nominal data that still needs to be coded for proper interpretation.

Quantitative Data Interpretation Method

The quantitative data interpretation method is used to analyze quantitative data, which is also known as numerical data. This data type contains numbers and is therefore analyzed with the use of numbers and not texts. Quantitative data are of 2 main types, namely; discrete and continuous data. Continuous data is further divided into interval data and ratio data, with all the data types being numeric. Due to its natural existence as a number, analysts do not need to employ the coding technique on quantitative data before it is analyzed. The process of analyzing quantitative data involves statistical modelling techniques such as standard deviation, mean and median.Some of the statistical methods used in analyzing quantitative data are highlighted below:

Mean•

The mean is a numerical average for a set of data and is calculated by dividing the sum of the values by the number of values in a dataset. It is used to get an estimate of a large population from the dataset obtained from a sample of the population.

Standard deviation•

This technique is used to measure how well the responses align with or deviates from the mean. It describes the degree of consistency within the responses; together with the mean, it provides insight into data sets.

Frequency distribution•

This technique is used to assess the demography of the respondents or the number of times a particular response appears in research. It is extremely keen on determining the degree of intersection between data points.

Advantages of Data Interpretation

It helps to make informed decisions and not just through guessing or predictions. •

It is cost-efficient •

The insights obtained can be used to set and identify trends in data.•

Conclusion

Data interpretation and analysis is an important aspect of working with data sets in any field or research and statistics. They both go hand in hand, as the process of data interpretation involves the analysis of data. Data interpretation is very important, as it helps to acquire useful information from a pool of irrelevant ones while making informed decisions. It is found useful for individuals, businesses, and researchers.

Data Processing: Editing, Coding, Tabulating

After collecting data, the method of converting raw data into meaningful statement; inclu processing, data analysis, and data interpretation and presentation.

Data reduction or processing mainly involves various manipulations necessary for preparing the analysis. The process (of manipulation) could be manual or electronic. It involves editing, catego open-ended questions, coding, computerization and preparation of tables and diagrams.

Editing data:

Information gathered during data collection may lack uniformity. Example: Data collected questionnaire and schedules may have answers which may not be ticked at proper places, or some may be left unanswered. Sometimes information may be given in a form which needs reconstru category designed for analysis, e.g., converting daily/monthly income in annual income and s researcher has to take a decision as to how to edit it.

Editing also needs that data are relevant and appropriate and errors are modified. Occasion investigator makes a mistake and records and impossible answer. "How much red chilies do you month" The answer is written as "4 kilos". Can a family of three members use four kilo chilies in a m correct answer could be "0.4 kilo".

Care should be taken in editing (re-arranging) answers to open-ended questions. Example: So "don't know" answer is edited as "no response". This is wrong. "Don't know" means that the respo not sure and is in a double mind about his reaction or considers the questions personal and does no answer it. "No response" means that the respondent is not familiar wi situation/object/event/individual about which he is asked.

Coding of data:

Coding is translating answers into numerical values or assigning numbers to the various catego variable to be used in data analysis. Coding is done by using a code book, code sheet, and a compu Coding is done on the basis of the instructions given in the codebook. The code book gives a numer for each variable.

Now-a-days, codes are assigned before going to the field while constructing the questionnaire/s Pose data collection; pre-coded items are fed to the computer for processing and analysis. For ope questions, however, post-coding is necessary. In such cases, all answers to open-ended questions a in categories and each category is assigned a code.

Manual processing is employed when qualitative methods are used or when in quantitative studies sample is used, or when the questionnaire/schedule has a large number of open-ended questions, accessibility to computers is difficult or inappropriate. However, coding is done in manual processing

Data classification/distribution:

Sarantakos (1998: 343) defines distribution of data as a form of classification of scores obtaine various categories or a particular variable. There are four types of distributions:

1. Frequency distribution
2. Percentage distribution
3. Cumulative distribution
4. Statistical distributions

1. Frequency distribution:

In social science research, frequency distribution is very common. It presents the frequency of occurr certain categories. This distribution appears in two forms:

Ungrouped: Here, the scores are not collapsed into categories, e.g., distribution of ages of the student (MC) class, each age value (e.g., 18, 19, 20, and so on) will be presented separately in the distribution.

Grouped: Here, the scores are collapsed into categories, so that 2 or 3 scores are presented toget group. For example, in the above age distribution groups like 18-20, 21-22 etc., can be formed)

2. Percentage distribution:

It is also possible to give frequencies not in absolute numbers but in percentages. For instance saying 200 respondents of total 2000 had a monthly income of less than Rs. 500, we can say respondents have a monthly income of less than Rs. 500.

3. Cumulative distribution:

It tells how often the value of the random variable is less than or equal to a particular reference valu

4. Statistical data distribution:

In this type of data distribution, some measure of average is found out of a sample of responde kind of averages are available (mean, median, mode) and the researcher must decide which is mo to his purpose. Once the average has been calculated, the question arises: how representative a i.e., how closely the answers are bunched around it. Are most of them very close to it or is there a of variation?

Tabulation of data:

After editing, which ensures that the information on the schedule is accurate and categorized in form, the data are put together in some kinds of tables and may also undergo some other forms of analysis.

Table can be prepared manually and/or by computers. For a small study of 100 to 200 persons, the little point in tabulating by computer since this necessitates putting the data on punched cards survey analysis involving a large number of respondents and requiring cross tabulation involving two variables, hand tabulation will be inappropriate and time consuming.

Usefulness of tables:

Tables are useful to the researchers and the readers in three ways:

1. The present an overall view of findings in a simpler way.
2. They identify trends.
3. They display relationships in a comparable way between parts of the findings.

By convention, the dependent variable is presented in the rows and the independent variable in the

Hypothesis Testing

Hypothesis testing is one of the most important concepts in Statistics which is heavily used by **Statisticians**, **Machine Learning Engineers**, and **Data Scientists**.

In hypothesis testing, Statistical tests are used to check whether the **null hypothesis** is rejected or not rejected. These Statistical tests assume a null hypothesis of no relationship or no difference between groups.

So, In this article, we will be discussing the statistical test for hypothesis testing including both parametric and non-parametric tests.

Parametric Tests

The basic principle behind the parametric tests is that we have a fixed set of parameters that are used to determine a probabilistic model that may be used in Machine Learning as well.

Parametric tests are those tests for which we have prior knowledge of the population distribution (i.e, normal), or if not then we can easily approximate it to a normal distribution which is possible with the help of the Central Limit Theorem.

Parameters for using the normal distribution is –

- Mean
- Standard Deviation

Eventually, the classification of a test to be parametric is completely dependent on the population assumptions. There are many parametric tests available from which some of them are as follows:

- To find the confidence interval for the population means with the help of known standard deviation.
- To determine the confidence interval for population means along with the unknown standard deviation.
- To find the confidence interval for the population variance.
- To find the confidence interval for the difference of two means, with an unknown value of standard deviation.

Non-parametric Tests

In Non-Parametric tests, we don't make any assumption about the parameters for the given population or the population we are studying. In fact, these tests don't depend on the population.

Hence, there is no fixed set of parameters is available, and also there is no distribution (normal distribution, etc.) of any kind is available for use.

This is also the reason that nonparametric tests are also referred to as **distribution-free tests**.

In modern days, Non-parametric tests are gaining popularity and an impact of influence some reasons behind this fame is –

- The main reason is that there is no need to be mannered while using parametric tests.
- The second reason is that we do not require to make assumptions about the population given (or taken) on which we are doing the analysis.
- Most of the nonparametric tests available are very easy to apply and to understand also i.e. the complexity is very low.

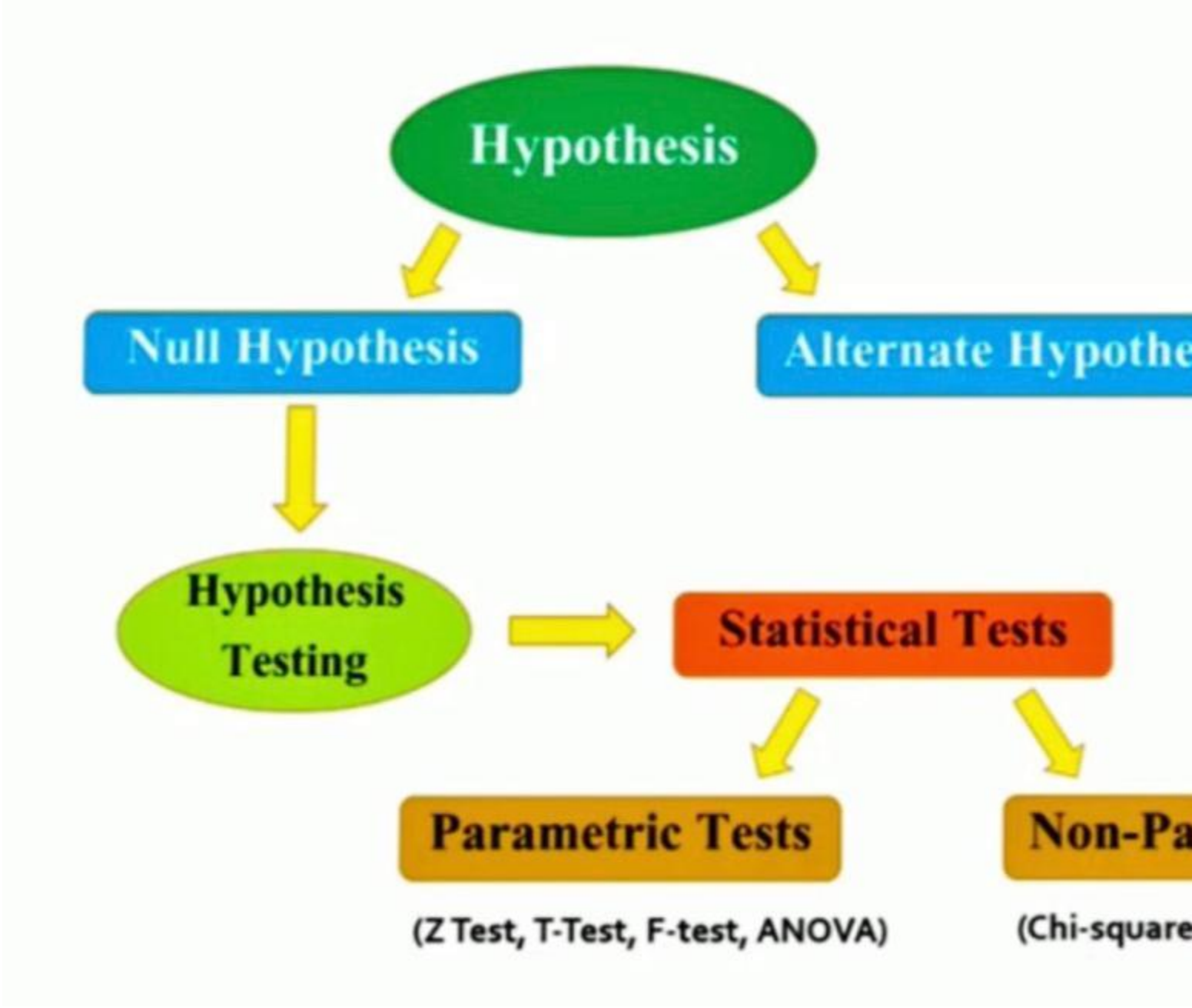


Image Source: Google Images

T-Test

1. It is a parametric test of hypothesis testing based on **Student's T distribution**.

2. It is essentially, testing the significance of the difference of the mean values when the sample size is small (i.e, less than 30) and when the population standard deviation is not available.

3. Assumptions of this test:

- Population distribution is normal, and
- Samples are random and independent
- The sample size is small.
- Population standard deviation is not known.

4. Mann-Whitney 'U' test is a non-parametric counterpart of the T-test.

A T-test can be a:

One Sample T-test: To compare a sample mean with that of the population mean.

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

where,

$\bar{x}$ is the sample mean

s is the sample standard deviation

n is the sample size

μ is the population mean

Two-Sample T-test: To compare the means of two different samples.

where,

$\bar{x}_1$ is the sample mean of the first group

$\bar{x}_2$ is the sample mean of the second group

S_1 is the sample-1 standard deviation

S_2 is the sample-2 standard deviation

n is the sample size

Conclusion:

- If the value of the test statistic is greater than the table value -> **Rejects the null hypothesis**.
- If the value of the test statistic is less than the table value -> **Do not reject the null hypothesis**.

Z-Test

1. It is a parametric test of hypothesis testing.

2. It is used to determine whether the means are different when the population variance is known and the sample size is large (i.e, greater than 30).

3. Assumptions of this test:

- Population distribution is normal
- Samples are random and independent.
- The sample size is large.
- Population standard deviation is known.

A Z-test can be:

One Sample Z-test: To compare a sample mean with that of the population mean.

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

$\bar{x}$ = sample mean

μ = population mean

σ = population standard deviation

n = sample size

Image Source: Google Images

Two Sample Z-test: To compare the means of two different samples.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

where,

$\bar{x}_1$ is the sample mean of 1st group

$\bar{x}_2$ is the sample mean of 2nd group

σ_1 is the population-1 standard deviation

σ_2 is the population-2 standard deviation

n is the sample size

F-Test

1. It is a parametric test of hypothesis testing based on **Snedecor F-distribution**.

2. It is a test for the null hypothesis that two normal populations have the same variance.

3. An F-test is regarded as a comparison of equality of sample variances.

4. F-statistic is simply a ratio of two variances.

5. It is calculated as:

$F = s_1^2/s_2^2$

$$s^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{X}\right)^2}{n-1}$$

6. By changing the variance in the ratio, F-test has become a very flexible test. It can then be used to:

- Test the overall significance for a regression model.
- To compare the fits of different models and
- To test the equality of means.

7. Assumptions of this test:

- Population distribution is normal, and
- Samples are drawn randomly and independently.

ANOVA

1. Also called as **Analysis of variance**, it is a parametric test of hypothesis testing.

2. It is an extension of the T-Test and Z-test.

3. It is used to test the significance of the differences in the mean values among more than two sample groups.

4. It uses F-test to statistically test the equality of means and the relative variance between them.

5. Assumptions of this test:

- Population distribution is normal, and
- Samples are random and independent.
- Homogeneity of sample variance.

6. One-way ANOVA and Two-way ANOVA are is types.

7. F-statistic = variance between the sample means/variance within the sample

Chi-Square Test

1. It is a non-parametric test of hypothesis testing.

2. As a non-parametric test, chi-square can be used:

- test of goodness of fit.
- as a test of independence of two variables.

3. It helps in assessing the goodness of fit between a set of observed and those expected theoretically.

4. It makes a comparison between the expected frequencies and the observed frequencies.

5. Greater the difference, the greater is the value of chi-square.

6. If there is no difference between the expected and observed frequencies, then the value of chi-square is equal to zero.

7. It is also known as the **"Goodness of fit test"** which determines whether a particular distribution fits the observed data or not.

8. It is calculated as:

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

O = the frequencies observed

E = the frequencies expected

$\sum$ ***= the 'sum of'***

9. Chi-square is also used to test the independence of two variables.

10. Conditions for chi-square test:

- Randomly collect and record the Observations.
- In the sample, all the entities must be independent.
- No one of the groups should contain very few items, say less than 10.

- The reasonably large overall number of items. Normally, it should be at least 50, however small the number of groups may be.

11. Chi-square as a parametric test is used as a test for population variance based on sample variance.

12. If we take each one of a collection of sample variances, divide them by the known population variance and multiply these quotients by (n-1), where n means the number of items in the sample, we get the values of chi-square.

13. It is calculated as:

Chi-square test

$$\chi^2 = \frac{\sigma s^2}{\sigma p^2}(\text{n-1})$$

$$\chi^2 = \sum \frac{(Oij - Eij)^2}{Eij}$$

Degree of Freedom	Probability of Exceeding the Critical Value								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.0
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.6
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.2
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.
	Not Significant							*Significant*	

Mann-Whitney U-Test

1. It is a non-parametric test of hypothesis testing.

2. This test is used to investigate whether two independent samples were selected from a population having the same distribution.

3. It is a true non-parametric counterpart of the T-test and gives the most accurate estimates of significance especially when sample sizes are small and the population is not normally distributed.

4. It is based on the comparison of every observation in the first sample with every observation in the other sample.

5. The test statistic used here is "U".

6. Maximum value of "U" is '$n_1 * n_2$' and the minimum value is zero.

7. It is also known as:

- Mann-Whitney Wilcoxon Test.
- Mann-Whitney Wilcoxon Rank Test.

8. Mathematically, U is given by:

$$\mathbf{U_1 = R_1 - n_1(n_1+1)/2}$$

where n_1 is the sample size for sample 1, and R_1 is the sum of ranks in Sample 1.

$$\mathbf{U_2 = R_2 - n_2(n_2+1)/2}$$

When consulting the significance tables, the smaller values of U_1 and U_2 are used. The sum of two values is given by,

$$\mathbf{U_1 + U_2 = \{ R_1 - n_1(n_1+1)/2 \} + \{ R_2 - n_2(n_2+1)/2 \}}$$

Knowing that $R_1+R_2 = N(N+1)/2$ and $N=n_1+n_2$, and doing some algebra, we find that the sum is:

$$\mathbf{U_1 + U_2 = n_1 * n_2}$$

Kruskal-Wallis H-test

1. It is a non-parametric test of hypothesis testing.

2. This test is used for comparing two or more independent samples of equal or different sample sizes.

3. It extends the Mann-Whitney-U-Test which is used to comparing only two groups.

4. One-Way ANOVA is the parametric equivalent of this test. And that's why it is also known as '**One-Way ANOVA on ranks**.

5. It uses ranks instead of actual data.

6. It does not assume the population to be normally distributed.

7. The test statistic used here is "H".

Wilcoxon Matched-Pairs Signed-Ranks Test

Another popular nonparametric test for matched or paired data is called the Wilcoxon Signed Rank Test. Like the Sign Test, it is based on difference scores, but in addition to analyzing the signs of the differences, it also takes into account the magnitude of the observed differences.

Let's use the Wilcoxon Signed Rank Test to re-analyze the data in Example 4 on page 5 of this module. Recall that this study assessed the effectiveness of a new drug designed to reduce repetitive behaviors in children affected with autism. A total of 8 children with autism enroll in the study and the amount of time that each child is engaged in repetitive behavior during three hour observation periods are measured both before treatment and then again after taking the new medication for a period of 1 week. The data are shown below.

Child	Before Treatment	After 1 Week of Treat
1	85	75
2	70	50
3	40	50
4	65	40
5	80	20
6	75	65
7	55	40
8	20	25

First, we compute difference scores for each child.

Child	Before Treatment	After 1 Week of Treatment	
1	85	75	
2	70	50	
3	40	50	

4	65	40	
5	80	20	
6	75	65	
7	55	40	
8	20	25	

The next step is to rank the difference scores. We first order the *absolute values of the difference scores* and assign rank from 1 through n to the smallest through largest absolute values of the difference scores, and assign the mean rank when there are ties in the absolute values of the difference scores.

Observed Differences		**Ordered Absolute Values of Differences**
10		-5
20		10
-10		-10
25		10
60		15
10		20
15		25
-5		60

The final step is to attach the signs ("+" or "-") of the observed differences to each rank as shown below.

Observed Differences		**Ordered Absolute Values of Difference Scores**
10		-5
20		10
-10		-10

25		10
60		15
10		20
15		25
-5		60

Similar to the Sign Test, hypotheses for the Wilcoxon Signed Rank Test concern the population median of the difference scores. The research hypothesis can be one- or two-sided. Here we consider a one-sided test.

H_0: The median difference is zero versus

H_1: The median difference is positive $\alpha=0.05$

Test Statistic for the Wilcoxon Signed Rank Test

The test statistic for the Wilcoxon Signed Rank Test is W, defined as the smaller of W+ (sum of the positive ranks) and W- (sum of the negative ranks). If the *null hypothesis* is true, we expect to see similar numbers of lower and higher ranks that are both positive and negative (i.e., W+ and W- would be similar). If the *research hypothesis* is true we expect to see more higher and positive ranks (in this example, more children with substantial improvement in repetitive behavior after treatment as compared to before, i.e., W+ much larger than W-).

In this example, W+ = 32 and W- = 4. Recall that the sum of the ranks (ignoring the signs) will always equal n(n+1)/2. As a check on our assignment of ranks, we have n(n+1)/2 = 8(9)/2 = 36 which is equal to 32+4. The test statistic is W = 4.

Next we must determine whether the observed test statistic W supports the null or research hypothesis. This is done following the same approach used in parametric testing. Specifically, we determine a critical value of W such that if the observed value of W is less than or equal to the critical value, we reject H_0 in favor of H_1, and if the observed value of W exceeds the critical value, we do not reject H_0.

Table of Critical Values of W

The critical value of W can be found in the table below:

To determine the appropriate one-sided critical value we need sample size (n=8) and our one-sided level of significance (α=0.05). For this example, the critical value of W is 6 and the decision rule is to reject H_0 if $W \leq 6$. Thus, we reject H_0, because $4 \leq 6$. We have statistically significant evidence at α =0.05, to show that the median difference is positive (i.e., that repetitive behavior improves.)

Note that when we analyzed the data previously using the Sign Test, we failed to find statistical significance. However, when we use the Wilcoxon Signed Rank Test, we conclude that the treatment result in a statistically significant improvement at α=0.05. The discrepant results are due to the fact that the Sign Test uses very little information in the data and is a less powerful test.

Example:

A study is run to evaluate the effectiveness of an exercise program in reducing systolic blood pressure in patients with pre-hypertension (defined as a systolic blood pressure between 120-139 mmHg or a diastolic blood pressure between 80-89 mmHg). A total of 15 patients with pre-hypertension enroll in the study, and their systolic blood pressures are measured. Each patient then participates in an exercise training program where they learn proper techniques and execution of a series of exercises. Patients are instructed to do the exercise program 3 times per week for 6 weeks. After 6 weeks, systolic blood pressures are again measured. The data are shown below.

Patient	Systolic Blood Pressure Before Exercise Program	Systolic Blood After Exercise
1	125	118
2	132	134
3	138	130
4	120	124
5	125	105
6	127	130

7	136	130
8	139	132
9	131	123
10	132	128
11	135	126
12	136	140
13	128	135
14	127	126
15	130	132

Is there is a difference in systolic blood pressures after participating in the exercise program as compared to before?

- **Step1.** Set up hypotheses and determine level of significance.

  H_0: The median difference is zero versus

  H_1: The median difference is not zero $\alpha=0.05$

- **Step 2.** Select the appropriate test statistic.

The test statistic for the Wilcoxon Signed Rank Test is W, defined as the smaller of W+ and W- which are the sums of the positive and negative ranks, respectively.

- **Step 3.** Set up the decision rule.

The critical value of W can be found in the table of critical values. To determine the appropriate critical value from Table 7 we need sample size (n=15) and our two-sided level of significance ($\alpha=0.05$). The critical value for this two-sided test with n=15 and $\alpha=0.05$ is 25 and the decision rule is as follows: Reject H_0 if $W \leq 25$.

- **Step 4.** Compute the test statistic.

Because the before and after systolic blood pressures measures are paired, we compute difference scores for each patient.

Patient	Systolic Blood Pressure Before Exercise Program	Systolic Blood Pressure After Exercise Program	(
1	125	118	
2	132	134	
3	138	130	
4	120	124	
5	125	105	
6	127	130	
7	136	130	
8	139	132	
9	131	123	
10	132	128	
11	135	126	
12	136	140	
13	128	135	
14	127	126	
15	130	132	

The next step is to rank the ordered absolute values of the difference scores using the approach outlined in Section 10.1. Specifically, we assign ranks from 1 through n to the smallest through largest absolute values of the difference scores, respectively, and assign the mean rank when there are ties in the absolute values of the difference scores.

Observed Differences		Ordered Absolute Values of Differences	
7		1	

-2		-2	
8		-2	
-4		-3	
20		-4	
-3		-4	
6		4	
7		6	
8		-7	
4		7	
9		7	
-4		8	
-7		8	
1		9	
-2		20	

The final step is to attach the signs ("+" or "-") of the observed differences to each rank as shown below.

Observed Differences		**Ordered Absolute Values of Differences**	**Ranks**	**Signed Ranks**
7		1	1	1
-2		-2	2.5	-2.5
8		-2	2.5	-2.5
-4		-3	4	-4
20		-4	6	-6

-3		-4	6	-6
6		4	6	6
7		6	8	8
8		-7	10	-10
4		7	10	10

In this example, W+ = 89 and W- = 31. Recall that the sum of the ranks (ignoring the signs) will always equal n(n+1)/2. As a check on our assignment of ranks, we have n(n+1)/2 = 15(16)/2 = 120 which is equal to 89 + 31. The test statistic is W = 31.

- **Step 5.** Conclusion.

We do not reject H_0 because 31 > 25. Therefore, we do not have statistically significant evidence at α=0.05, to show that the median difference in systolic blood pressures is not zero (i.e., that there is a significant difference in systolic blood pressures after the exercise program as compared to before).

Mann Whitney U Test (Wilcoxon Rank Sum Test)

The modules on hypothesis testing presented techniques for testing the equality of means in two independent samples. An underlying assumption for appropriate use of the tests described was that the continuous outcome was approximately normally distributed or that the samples were sufficiently large (usually $n_1 \geq 30$ and $n_2 \geq 30$) to justify their use based on the Central Limit Theorem. When comparing two independent samples when the outcome is not normally distributed and the samples are small, a nonparametric test is appropriate.

A popular nonparametric test to compare outcomes between two independent groups is the Mann Whitney U test. The Mann Whitney U test, sometimes called the Mann Whitney Wilcoxon Test or the Wilcoxon Rank Sum Test, is used to test whether two samples are likely to derive from the same population (i.e., that the two populations have the same shape). Some investigators interpret this test as comparing the medians between the two populations. Recall that the parametric test compares the means (H_0: $\mu_1=\mu_2$) between independent groups.

In contrast, the null and two-sided research hypotheses for the *nonparametric test* are stated as follows:

H_0: The two populations are equal versus

H_1: The two populations are not equal.

This test is often performed as a two-sided test and, thus, the research hypothesis indicates that the populations are not equal as opposed to specifying directionality. A one-sided research hypothesis is used if interest lies in detecting a positive or negative shift in one population as compared to the other. The procedure for the test involves pooling the observations from the two samples into one combined sample, keeping track of which sample each observation comes from, and then ranking lowest to highest from 1 to n_1+n_2, respectively.

Example:

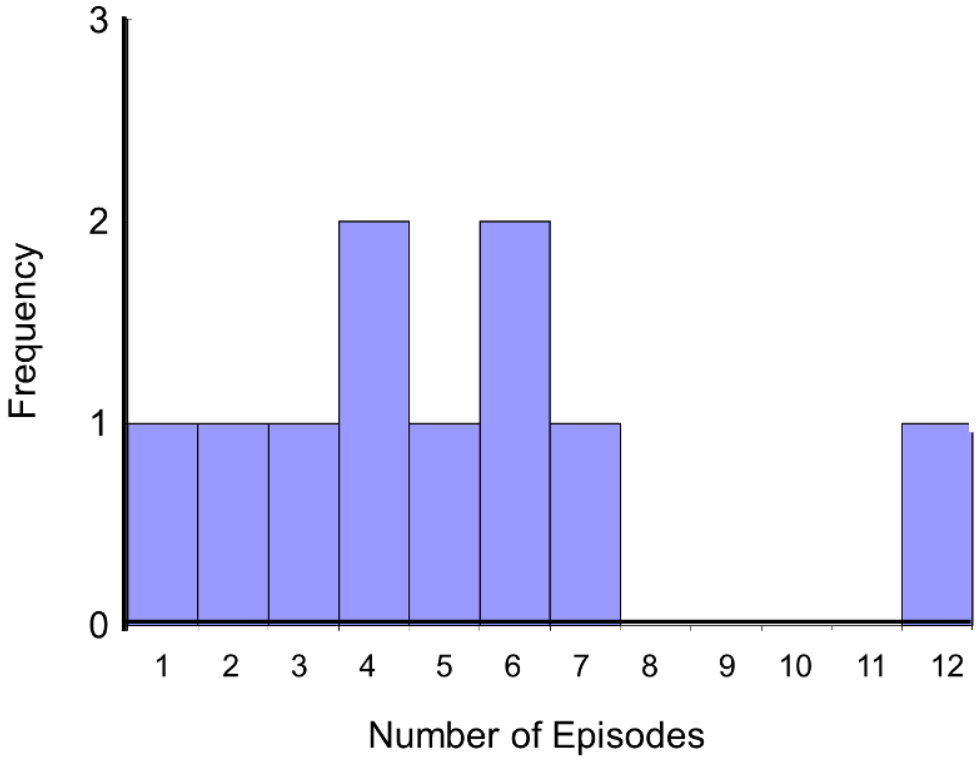
Consider a Phase II clinical trial designed to investigate the effectiveness of a new drug to reduce symptoms of asthma in children. A total of n=10 participants are randomized to receive either the new drug or a placebo. Participants are asked to record the number of episodes of shortness of breath over a 1 week period following receipt of the assigned treatment. The data are shown below.

Placebo	7	5	6	4	12
New Drug	3	6	4	2	1

Is there a difference in the number of episodes of shortness of breath over a 1 week period in participants receiving the new drug as compared to those receiving the placebo? By inspection, it appears that participants receiving the placebo have more episodes of shortness of breath, but is this statistically significant?

In this example, the outcome is a count and in this sample the data do not follow a normal distribution.

Frequency Histogram of Number of Episodes of Shortness of Breath



In addition, the sample size is small ($n_1=n_2=5$), so a nonparametric test is appropriate. The hypothesis is given below, and we run the test at the 5% level of significance (i.e., $\alpha=0.05$).

H_0: The two populations are equal versus

H_1: The two populations are not equal.

Note that if the null hypothesis is true (i.e., the two populations are equal), we expect to see similar numbers of episodes of shortness of breath in each of the two treatment groups, and we would expect to see some participants reporting few episodes and some reporting more episodes in each group. This does not appear to be the case with the observed data. A test of hypothesis is needed to determine whether the observed data is evidence of a statistically significant difference in populations.

The first step is to assign ranks and to do so we order the data from smallest to largest. This is done on the combined or total sample (i.e., pooling the data from the two treatment groups (n=10)), and assigning ranks from 1 to 10, as follows. We also need to keep track of the group assignments in the total sample.

		Total Sample (Ordered Smallest to Largest)	**R**

Placebo	New Drug	Placebo	New Drug	Placebo
7	3		1	
5	6		2	
6	4		3	
4	2	4	4	4.5
12	1	5		6
		6	6	7.5
		7		9
		12		10

Note that the lower ranks (e.g., 1, 2 and 3) are assigned to responses in the new drug group while the higher ranks (e.g., 9, 10) are assigned to responses in the placebo group. Again, the goal of the test is to determine whether the observed data support a difference in the populations of responses. Recall that in parametric tests (discussed in the modules on hypothesis testing), when comparing means between two groups, we analyzed the difference in the sample means relative to their variability and summarized the sample information in a test statistic. A similar approach is employed here. Specifically, we produce a test statistic based on the ranks.

First, we sum the ranks in each group. In the placebo group, the sum of the ranks is 37; in the new drug group, the sum of the ranks is 18. Recall that the sum of the ranks will always equal n(n+1)/2. As a check on our assignment of ranks, we have n(n+1)/2 = 10(11)/2=55 which is equal to 37+18 = 55.

For the test, we call the placebo group 1 and the new drug group 2 (assignment of groups 1 and 2 is arbitrary). We let R_1 denote the sum of the ranks in group 1 (i.e., R_1=37), and R_2 denote the sum of the ranks in group 2 (i.e., R_2=18). If the null hypothesis is true (i.e., if the two populations are equal), we expect R_1 and R_2 to be similar. In this example, the lower values (lower ranks) are clustered in the new drug group (group 2), while the higher values (higher ranks) are clustered in the placebo group (group 1). This is suggestive, but is the observed difference in the sums of the ranks simply due to chance? To answer this we will compute a test statistic to summarize the sample information and look up the corresponding value in a probability distribution.

Test Statistic for the Mann Whitney U Test

The test statistic for the Mann Whitney U Test is denoted **U** and is the ***smaller*** of U_1 and U_2, defined below.

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

where R_1 = sum of the ranks for group 1 and R_2 = sum of the ranks for group 2.

For this example,

$$U_1 = 5(5) + \frac{5(6)}{2} - 37 = 3$$

$$U_2 = 5(5) + \frac{5(6)}{2} - 18 = 22$$

In our example, U=3. Is this evidence in support of the null or research hypothesis? Before we address this question, we consider the range of the test statistic U in two different situations.

Situation #1

Consider the situation where there is ***complete separation*** of the groups, supporting the ***research hypothesis*** that the two populations are not equal. If all of the higher numbers of episodes of shortness of breath (and thus all of the higher ranks) are in the placebo group, and all of the lower numbers of episodes (and ranks) are in the new drug group and that there are no ties, then:

$$R_1 = 6+7+8+9+10 = 40 \text{ and } R_2 = 1+2+3+4+5 = 15$$

and

$$U_1 = 5\left(5\right) + \frac{5(6)}{2} - 40 = 0 \text{ and } U_2 = 5\left(5\right) + \frac{5(6)}{2} - 15 = 25$$

Therefore, when there is clearly a difference in the populations, U=0.

Situation #2

Consider a second situation where l***ow and high scores are approximately evenly distributed in the two groups***, supporting the ***null hypothesis*** that the groups are equal. If ranks of 2, 4, 6, 8 and 10 are assigned to the numbers of episodes of

shortness of breath reported in the placebo group and ranks of 1, 3, 5, 7 and 9 are assigned to the numbers of episodes of shortness of breath reported in the new drug group, then:

$R_1 = 2+4+6+8+10 = 30$ and $R_2 = 1+3+5+7+9 = 25$ R_1= 2+4+6+8+10 = 30 and R_2= 1+3+5+7+9 = 25,

and

$$U_1 = 5(5) + \frac{5(6)}{2} - 30 = 10 \text{ and } U_2 = 5\left(5\right) + \frac{5(6)}{2} - 25 = 15$$

When there is clearly no difference between populations, then U=10.

Thus, smaller values of U support the research hypothesis, and larger values of U support the null hypothesis.

Key Concept:

For any Mann-Whitney U test, the theoretical range of U is from 0 (complete separation between groups, H_0 most likely false and H_1 most likely true) to n_1*n_2 (little evidence in support of H_1).

In every test, **U_1+U_2 is always equal to n_1*n_2.** In the example above, U can range from 0 to 25 and smaller values of U support the research hypothesis (i.e., we reject H_0 if U is small). The procedure for determining exactly when to reject H_0 is described below.

In every test, we must determine whether the observed U supports the null or research hypothesis. This is done following the same approach used in parametric testing. Specifically, we determine a critical value of U such that if the observed value of U is less than or equal to the critical value, we reject H_0 in favor of H_1 and if the observed value of U exceeds the critical value we do not reject H_0.

The critical value of U can be found in the table below. To determine the appropriate critical value we need sample sizes (for Example: $n_1=n_2=5$) and our two-sided level of significance ($\alpha=0.05$). For Example 1 the critical value is 2, and the decision rule is to reject H_0 if $U \leq 2$. We do not reject H_0 because $3 > 2$. We do not have statistically significant evidence at $\alpha = 0.05$, to show that the two populations of numbers of episodes of shortness of breath are not equal. However, in this example, the failure to

reach statistical significance may be due to low power. The sample data suggest a difference, but the sample sizes are too small to conclude that there is a statistically significant difference.

Table of Critical Values for U

Example:

A new approach to prenatal care is proposed for pregnant women living in a rural community. The new program involves in-home visits during the course of pregnancy in addition to the usual or regularly scheduled visits. A pilot randomized trial with 15 pregnant women is designed to evaluate whether women who participate in the program deliver healthier babies than women receiving usual care. The outcome is the **APGAR score** measured 5 minutes after birth. Recall that APGAR scores range from 0 to 10 with scores of 7 or higher considered normal (healthy), 4-6 low and 0-3 critically low. The data are shown below.

Usual Care	8	7	6	2	5	8	
New Program	9	9	7	8	10	9	

Is there statistical evidence of a difference in APGAR scores in women receiving the new and enhanced versus usual prenatal care? We run the test using the five-step approach.

- **Step 1.** Set up hypotheses and determine level of significance.

H_0: The two populations are equal versus

H_1: The two populations are not equal. $\alpha = 0.05$

- **Step 2.** Select the appropriate test statistic.

Because APGAR scores are not normally distributed and the samples are small ($n_1=8$ and $n_2=7$), we use the Mann Whitney U test. The test statistic is U, the smaller of

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \text{ and } U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

where R_1 and R_2 are the sums of the ranks in groups 1 and 2, respectively.

- **Step 3.** Set up decision rule.

The appropriate critical value can be found in the table above. To determine the appropriate critical value we need sample sizes (n_1=8 and n_2=7) and our two-sided level of significance (α=0.05). The critical value for this test with n_1=8, n_2=7 and α =0.05 is 10 and the decision rule is as follows: Reject H_0 if $U \leq 10$.

- **Step 4.** Compute the test statistic.

The first step is to assign ranks of 1 through 15 to the smallest through largest values in the total sample, as follows:

		Total Sample (Ordered Smallest to Largest)		R
Usual Care	**New Program**	**Usual Care**	**New Program**	**Usual Care**
8	9	2		1
7	8	3		2
6	7	5		3
2	8	6	6	4.5
5	10	7	7	7
8	9	7		7
7	6	8	8	10.5
3		8	8	10.5
			9	
			9	
			10	
				R_1=45.5

Next, we sum the ranks in each group. In the usual care group, the sum of the ranks is R_1=45.5 and in the new program group, the sum of the ranks is R_2=74.5. Recall that the sum of the ranks will always equal n(n+1)/2. As a check on our assignment of ranks, we have n(n+1)/2 = 15(16)/2=120 which is equal to 45.5+74.5 = 120.

We now compute U_1 and U_2, as follows:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 8(7) + \frac{8(9)}{2} - 45.5 = 46.5$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 8(7) + \frac{7(8)}{2} - 74.5 = 9.5$$

Thus, the test statistic is U=9.5.

- **Step 5.** Conclusion:

We reject H_0 because $9.5 \leq 10$. We have statistically significant evidence at $\alpha = 0.05$ to show that the populations of APGAR scores are not equal in women receiving usual prenatal care as compared to the new program of prenatal care.

Example:

A clinical trial is run to assess the effectiveness of a new anti-retroviral therapy for patients with HIV. Patients are randomized to receive a standard anti-retroviral therapy (usual care) or the new anti-retroviral therapy and are monitored for 3 months. The primary outcome is viral load which represents the number of HIV copies per milliliter of blood. A total of 30 participants are randomized and the data are shown below.

Standard Therapy	7500	8000	2000	550	1250	1000	2250	6800	3400	6
New Therapy	400	250	800	1400	8000	7400	1020	6000	920	1

Is there statistical evidence of a difference in viral load in patients receiving the standard versus the new anti-retroviral therapy?

- **Step 1.** Set up hypotheses and determine level of significance.

  H_0: The two populations are equal versus

  H_1: The two populations are not equal. $\alpha=0.05$

- **Step 2.** Select the appropriate test statistic.

Because viral load measures are not normally distributed (with outliers as well as limits of detection (e.g., "undetectable")), we use the Mann-Whitney U test. The test statistic is U, the smaller of

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \text{ and } U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

where R_1 and R_2 are the sums of the ranks in groups 1 and 2, respectively.

- **Step 3.** Set up the decision rule.

The critical value can be found in the table of critical values based on sample sizes (n_1=n_2=15) and a two-sided level of significance (α=0.05). The critical value 64 and the decision rule is as follows: Reject H_0 if $U \leq 64$.

- **Step 4**. Compute the test statistic.

The first step is to assign ranks of 1 through 30 to the smallest through largest values in the total sample. Note in the table below, that the "undetectable" measurement is listed first in the ordered values (smallest) and assigned a rank of 1.

		Total Sample (Ordered Smallest to Largest)	
Standard Anti-retroviral	**New Anti-retroviral**	**Standard Anti-retroviral**	**New Anti-retroviral**
7500	400		undetectable
8000	250		250
2000	800	400	400
550	1400	550	
1250	8000	670	
1000	7400		800
2250	1020		920
6800	6000	970	
3400	920	1000	
6300	1420		1020
9100	2700	1040	
970	4200	1250	

1040	5200		1400
670	4100		1420
400	undetectable	2000	
		2250	
			2700
		3400	
			4100
			4200
			5200
			6000
		6300	
		6800	
			7400
		7500	
		8000	8000
		9100	

Next, we sum the ranks in each group. In the standard anti-retroviral therapy group, the sum of the ranks is R_1=245; in the new anti-retroviral therapy group, the sum of the ranks is R_2=220. Recall that the sum of the ranks will always equal n(n+1)/2. As a check on our assignment of ranks, we have n(n+1)/2 = 30(31)/2=465 which is equal to 245+220 = 465. We now compute U_1 and U_2, as follows,

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = \mathbf{15(15)} + \frac{15(16)}{2} - \mathbf{245} = \mathbf{100}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = \mathbf{15(15)} + \frac{15(16)}{2} - \mathbf{220} = \mathbf{125}$$

Thus, the test statistic is U=100.

- **Step 5.** Conclusion.

We do not reject H_0 because 100 > 64. We do not have sufficient evidence to conclude that the treatment groups differ in viral load.

Unit-4

Concept of Independent and Dependent Variable

A variable is something you're trying to measure. It can be practically anything, such as objects, an time, feelings, events, or ideas. If you're studying how people feel about different television sh variables in that experiment are television shows and feelings. If you're studying how different fertilizer affect how tall plants grow, the variables are type of fertilizer and plant height.

There are two key variables in every experiment

1. Independent Variables

The independent variable and the dependent variable. The independent variable is the variable whos isn't affected by any other variable in the experiment. Either the scientist has to change the ind variable herself or it changes on its own; nothing else in the experiment affects or changes it. Two of common independent variables are age and time. There's nothing you or anything else can do to or slow down time or increase or decrease age. They're independent of everything else.

2. Dependent Variable

The dependent variable is what is being studied and measured in the experiment. It's what changes a of the changes to the independent variable. An example of a dependent variable is how tall y different ages. The dependent variable (height) depends on the independent variable (age).

An easy way to think of independent and dependent variables is, when you're conducting an experin independent variable is what you change, and the dependent variable is what changes because of t can also think of the independent variable as the cause and the dependent variable as the effect.

It can be a lot easier to understand the differences between these two variables with examples, so some sample experiments below.

Examples of Independent and Dependent Variables in Experiment

Below are overviews of three experiments, each with their independent and dependent variables ide

Experiment 1: You want to figure out which brand of microwave popcorn pops the most kernels so y the most value for your money. You test different brands of popcorn to see which bag pops the mos kernels.

- Independent Variable: Brand of popcorn bag (It's the independent variable becaus actually deciding the popcorn bag brands)
- Dependent Variable: Number of kernels popped (This is the dependent variable be what you measure for each popcorn brand)

Experiment 2: You want to see which type of fertilizer helps plants grow fastest, so you add a diffe of fertilizer to each plant and see how tall they grow.

- Independent Variable: Type of fertilizer given to the plant
- Dependent Variable: Plant height

Experiment 3: You're interested in how rising sea temperatures impact algae life, so you experiment that measures the number of algae in a sample of water taken from a specific ocean varying temperatures.

- Independent Variable: Ocean temperature
- Dependent Variable: The number of algae in the sample

For each of the independent variables above, it's clear that they can't be changed by other variab experiment. You have to be the one to change the popcorn and fertilizer brands in Experiments 1 a the ocean temperature in Experiment 3 cannot be significantly changed by other factors. Changes these independent variables cause the dependent variables to change in the experiments.

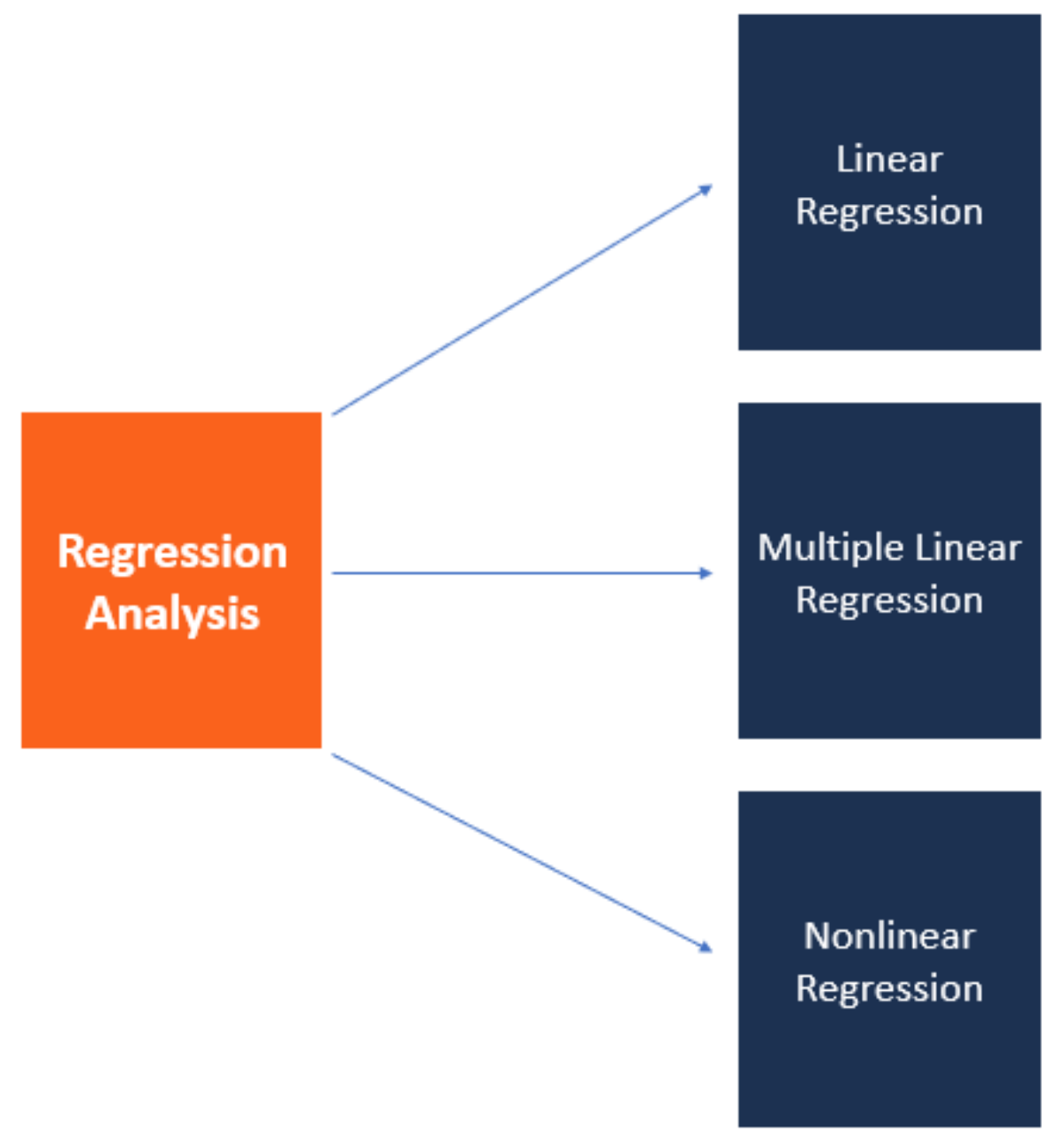
Multiple Regression

Multiple regression generally explains the relationship between multiple independent or predictor variables and one dependent or criterion variable. A

dependent variable is modeled as a function of several independent variables with corresponding coefficients, along with the constant term. Multiple regression requires two or more predictor variables, and this is why it is called multiple regression.

Multiple Regression

Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables. It can be utilized to assess the strength of the relationship between variables and for modeling the future relationship between them.



Regression analysis includes several variations, such as linear, multiple linear, and nonlinear. The most common models are simple linear and multiple linear. Nonlinear regression analysis is commonly used for more complicated data sets in which the dependent and independent variables show a nonlinear relationship.

Regression analysis offers numerous applications in various disciplines, including finance.

Regression Analysis – Linear Model Assumptions

Linear regression analysis is based on six fundamental assumptions:

1. The dependent and independent variables show a linear relationship between the slope and the intercept.
2. The independent variable is not random.
3. The value of the residual (error) is zero.
4. The value of the residual (error) is constant across all observations.
5. The value of the residual (error) is not correlated across all observations.
6. The residual (error) values follow the normal distribution.

Regression Analysis – Simple Linear Regression

Simple linear regression is a model that assesses the relationship between a dependent variable and an independent variable. The simple linear model is expressed using the following equation:

$$Y = a + bX + \epsilon$$

Where:

- **Y** – Dependent variable
- **X** – Independent (explanatory) variable
- **a** – Intercept
- **b** – Slope
- **ϵ** – Residual (error)

Regression Analysis – Multiple Linear Regression

Multiple linear regression analysis is essentially similar to the simple linear model, with the exception that multiple independent variables are used in the model. The mathematical representation of multiple linear regression is:

$$Y = a + bX_1 + cX_2 + dX_3 + \epsilon$$

Where:

- **Y** – Dependent variable
- **X_1, X_2, X_3** – Independent (explanatory) variables
- **a** – Intercept
- **b, c, d** – Slopes
- **ϵ** – Residual (error)

Multiple linear regression follows the same conditions as the simple linear model. However, since there are several independent variables in multiple linear analysis, there is another mandatory condition for the model:

- **Non-collinearity:** Independent variables should show a minimum correlation with each other. If the independent variables are highly correlated with each other, it will be difficult to assess the true relationships between the dependent and independent variables.

Factor Analysis

Factor analysis is a technique that is used to reduce a large number of variables into fewer numbers of factors. This technique extracts maximum common variance from all variables and puts them into a common score. As an index of all variables, we can use this score for further analysis. Factor analysis is part of general linear model (GLM) and this method also assumes several assumptions: there is linear relationship, there is no multicollinearity, it includes relevant variables into analysis, and there is true correlation between variables and factors. Several methods are available, but principal component analysis is used most commonly.

Types of factoring:

There are different types of methods used to extract the factor from the data set:

1. **Principal component analysis:** This is the most common method used by researchers. PCA starts extracting the maximum variance and puts them into the first factor. After that, it removes that variance explained by the first factors and then starts extracting maximum variance for the second factor. This process goes to the last factor.

2. **Common factor analysis:** The second most preferred method by researchers, it extracts the common variance and puts them into factors. This method does not include the unique variance of all variables. This method is used in SEM.

3. **Image factoring:** This method is based on correlation matrix. OLS Regression method is used to predict the factor in image factoring.

4. **Maximum likelihood method:** This method also works on correlation metric but it uses maximum likelihood method to factor.

5. **Other methods of factor analysis:** Alfa factoring outweighs least squares. Weight square is another regression based method which is used for factoring.

Factor loading:
Factor loading is basically the correlation coefficient for the variable and factor. Factor loading shows the variance explained by the variable on that particular factor. In the SEM approach, as a rule of thumb, 0.7 or higher factor loading represents that the factor extracts sufficient variance from that variable.
Eigenvalues: Eigenvalues is also called characteristic roots. Eigenvalues shows variance explained by that particular factor out of the total variance. From the commonality column, we can know how much variance is explained by the first factor out of the total variance. For example, if our first factor explains 68% variance out of the total, this means that 32% variance will be explained by the other factor.
Factor score: The factor score is also called the component score. This score is of all row and columns, which can be used as an index of all variables and can be used for further analysis. We can standardize this score by multiplying a common term. With this factor score, whatever analysis we will do, we will assume that all variables will behave as factor scores and will move.

Criteria for determining the number of factors: According to the Kaiser Criterion, Eigenvalues is a good criteria for determining a factor. If Eigenvalues is greater than one, we should consider that a factor and if Eigenvalues is less than one, then we

should not consider that a factor. According to the variance extraction rule, it should be more than 0.7. If variance is less than 0.7, then we should not consider that a factor.

Rotation method: Rotation method makes it more reliable to understand the output. Eigenvalues do not affect the rotation method, but the rotation method affects the Eigenvalues or percentage of variance extracted. There are a number of rotation methods available: (1) No rotation method, (2) Varimax rotation method, (3) Quartimax rotation method, (4) Direct oblimin rotation method, and (5) Promax rotation method. Each of these can be easily selected in SPSS, and we can compare our variance explained by those particular methods.

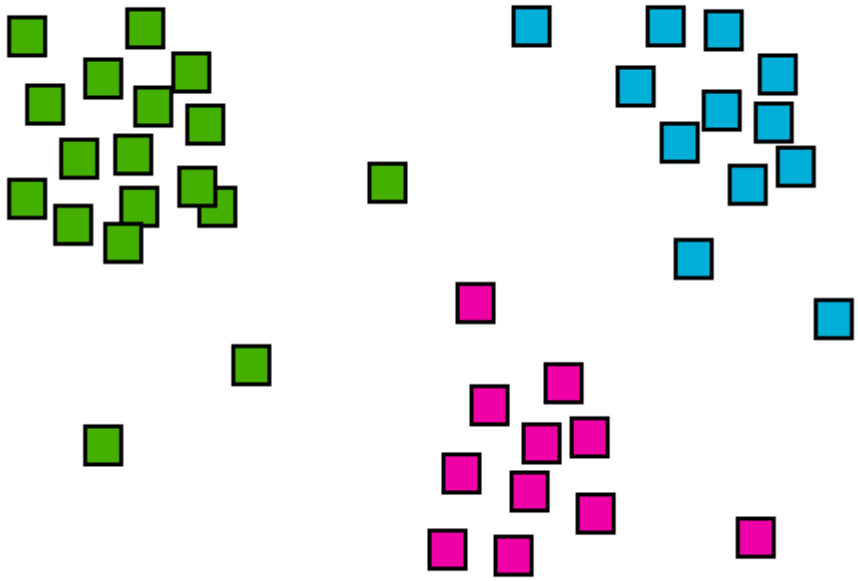
Assumptions:

1. **No outlier:** Assume that there are no outliers in data.
2. **Adequate sample size:** The case must be greater than the factor.
3. **No perfect multicollinearity:** Factor analysis is an interdependency technique. There should not be perfect multicollinearity between the variables.
4. **Homoscedasticity:** Since factor analysis is a linear function of measured variables, it does not require homoscedasticity between the variables.
5. **Linearity:** Factor analysis is also based on linearity assumption. Non-linear variables can also be used. After transfer, however, it changes into linear variable.
6. **Interval Data:** Interval data are assumed.

Cluster Analysis

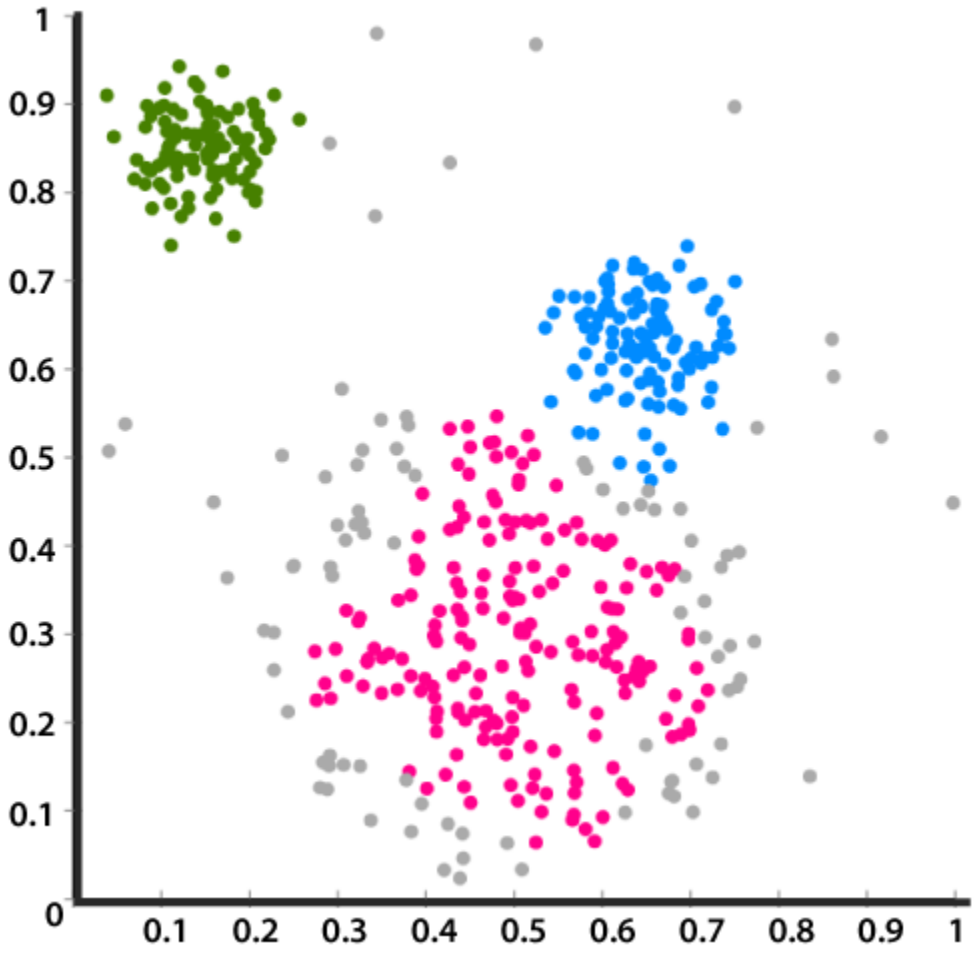
Cluster analysis foundations rely on one of the most fundamental, simple and very often unnoticed ways (or methods) of understanding and learning, which is grouping "objects" into "similar" groups. This process includes a number of different algorithms and methods to make clusters of a similar kind. It is also a part of data management in statistical analysis.

When we try to group a set of objects that have similar kind of characteristics, attributes these groups are called **clusters**. The process is called **clustering**. It is a very difficult task to get to know the properties of every individual object instead, it would be easy to group those similar objects and have a common structure of properties that the group follows.



What is Cluster Analysis?

Cluster analysis is a multivariate data mining technique whose goal is to groups objects (eg., products, respondents, or other entities) based on a set of user selected characteristics or attributes. It is the basic and most important step of data mining and a common technique for statistical data analysis, and it is used in many fields such as data compression, machine learning, pattern recognition, information retrieval etc.



Clusters should exhibit high internal homogeneity and high external heterogeneity.

What does this mean?

When plotted geometrically, objects within clusters should be very close together and clusters will be far apart.

Related Articles:

- Data Collection And Organization
- Data Sets
- Statistics

Types of Cluster Analysis

The clustering algorithm needs to be chosen experimentally unless there is a mathematical reason to choose one cluster method over another.It should be noted that an algorithm that works on a particular set of data will not work on another set of data. There are a number of different methods to perform cluster analysis. Some of them are,

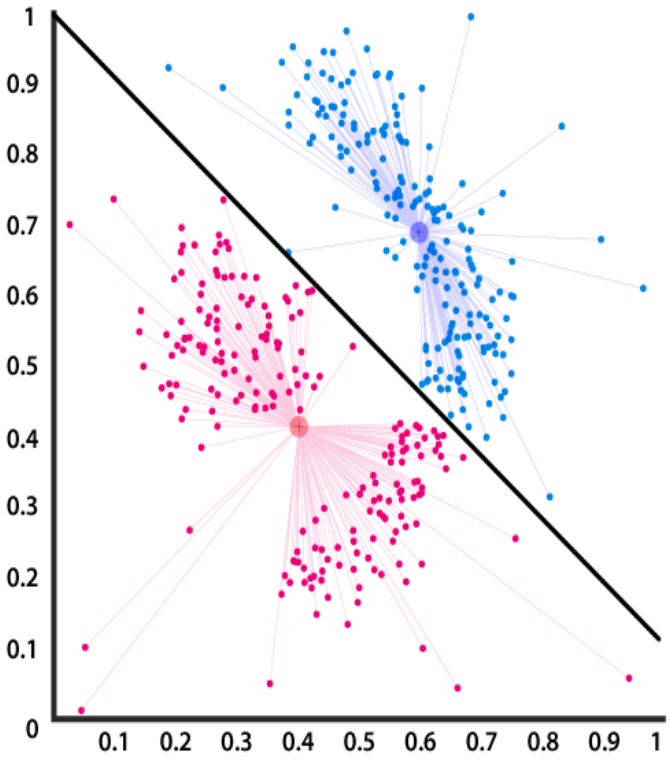
Hierarchical Cluster Analysis

In this method, first, a cluster is made and then added to another cluster (the most similar and closest one) to form one single cluster. This process is repeated until all subjects are in one cluster. This particular method is known as **Agglomerative method**. Agglomerative clustering starts with single objects and starts grouping them into clusters.

The divisive method is another kind of Hierarchical method in which clustering starts with the complete data set and then starts dividing into partitions.

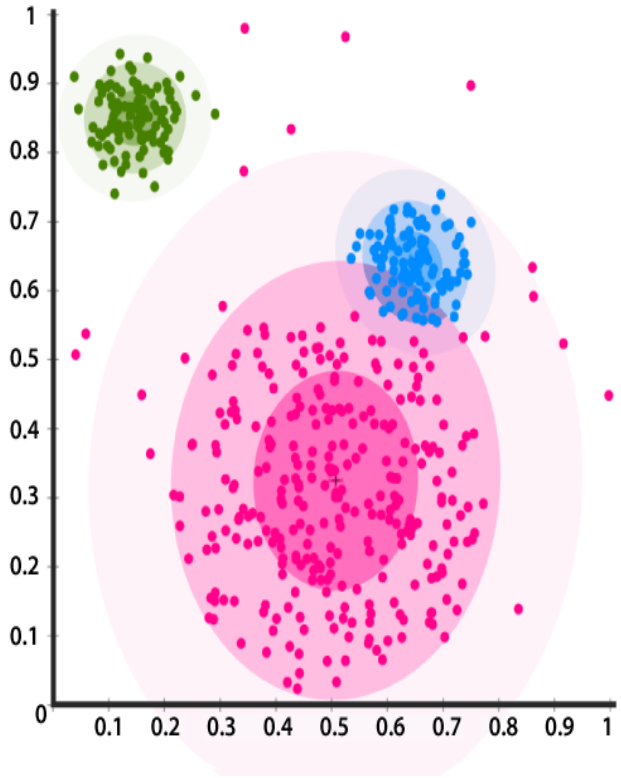
Centroid-based Clustering

In this type of clustering, clusters are represented by a central entity, which may or may not be a part of the given data set. K-Means method of clustering is used in this method, where k are the cluster centers and objects are assigned to the nearest cluster centres.



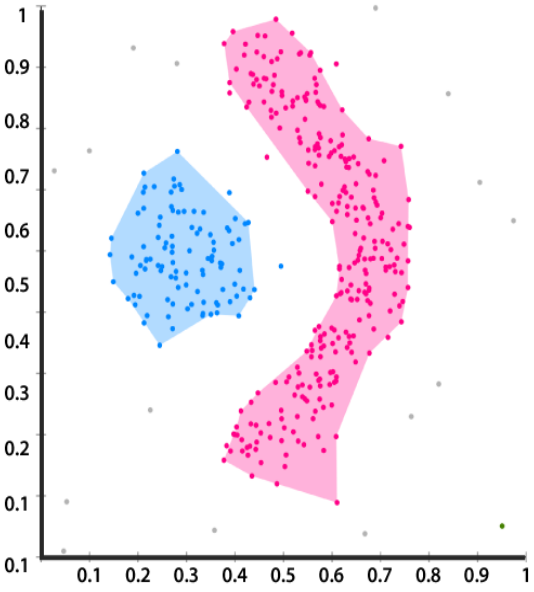
Distribution-based Clustering

It is a type of clustering model closely related to statistics based on the modals of distribution. Objects that belong to the same distribution are put into a single cluster.This type of clustering can capture some complex properties of objects like correlation and dependence between attributes.



Density-based Clustering

In this type of clustering, clusters are defined by the areas of density that are higher than the remaining of the data set. Objects in sparse areas are usually required to separate clusters.The objects in these sparse points are usually noise and border points in the graph.The most popular method in this type of clustering is DBSCAN.



To learn more on the cluster and other statistics-related topics, visit BYJU’S.

Applications and Examples

It is the principal job of exploratory data mining, and a common method for statistical data analysis. It is used in many fields, such as machine learning, image analysis, pattern recognition, information retrieval, data compression, bioinformatics and computer graphics.

It can be used to examine patterns of antibiotic resistance, to incorporate antimicrobial compounds according to their mechanism of activity, to analyse antibiotics according to their antibacterial action.

Cluster analysis can be a compelling data-mining means for any organization that wants to recognise discrete groups of customers, sales transactions, or other kinds of behaviours and things. For example, insurance providing companies use cluster analysis to identify fraudulent claims and banks apply it for credit scoring.

What Is a Research Report?

A research report is a well-crafted document that outlines the processes, data, and findings of a systematic investigation. It is an important document that serves as a first-hand account of the research process, and it is typically considered an objective and accurate source of information.

In many ways, a research report can be considered as a summary of the research process that clearly highlights findings, recommendations, and other important details. Reading a well-written research report should provide you with all the information you need about the core areas of the research process.

Features of a Research Report

So how do you recognize a research report when you see one? Here are some of the basic features that define a research report.

- It is a detailed presentation of research processes and findings, and it usually includes tables and graphs.
- It is written in a formal language.
- A research report is usually written in the third person.
- It is informative and based on first-hand verifiable information.
- It is formally structured with headings, sections, and bullet points.
- It always includes recommendations for future actions.

Types of Research Report

The research report is classified based on two things; nature of research and target audience.

Nature of Research

- **Qualitative Research Report**

This is the type of report written for qualitative research. It outlines the methods, processes, and findings of a qualitative method of systematic investigation. In educational research, a qualitative research report provides an opportunity for one to apply his or her knowledge and develop skills in planning and executing qualitative research projects.

A qualitative research report is usually descriptive in nature. Hence, in addition to presenting details of the research process, you must also create a descriptive narrative of the information.

- **Quantitative Research Report**

A quantitative research report is a type of research report that is written for quantitative research. Quantitative research is a type of systematic investigation that pays attention to numerical or statistical values in a bid to find answers to research questions.

In this type of research report, the researcher presents quantitative data to support the research process and findings. Unlike a qualitative research report that is mainly descriptive, a quantitative research report works with numbers; that is, it is numerical in nature.

Target Audience

Also, a research report can be said to be technical or popular based on the target audience. If you're dealing with a general audience, you would need to present a popular research report, and if you're dealing with a specialized audience, you would submit a technical report.

- **Technical Research Report**

A technical research report is a detailed document that you present after carrying out industry-based research. This report is highly specialized because it provides

information for a technical audience; that is, individuals with above-average knowledge in the field of study.

In a technical research report, the researcher is expected to provide specific information about the research process, including statistical analyses and sampling methods. Also, the use of language is highly specialized and filled with jargon.

Examples of technical research reports include legal and medical research reports.

- **Popular Research Report**

A popular research report is one for a general audience; that is, for individuals who do not necessarily have any knowledge in the field of study. A popular research report aims to make information accessible to everyone.

It is written in very simple language, which makes it easy to understand the findings and recommendations. Examples of popular research reports are the information contained in newspapers and magazines.

Importance of a Research Report

- **Knowledge Transfer:** As already stated above, one of the reasons for carrying out research is to contribute to the existing body of knowledge, and this is made possible with a research report. A research report serves as a means to effectively communicate the findings of a systematic investigation to all and sundry.
- **Identification of Knowledge Gaps:** With a research report, you'd be able to identify knowledge gaps for further inquiry. A research report shows what has been done while hinting at other areas needing systematic investigation.
- In market research, a research report would help you understand the market needs and peculiarities at a glance.
- A research report allows you to present information in a precise and concise manner.
- It is time-efficient and practical because, in a research report, you do not have to spend time detailing the findings of your research work in person. You can easily send out the report via email and have stakeholders look at it.

Guide to Writing a Research Report

A lot of detail goes into writing a research report, and getting familiar with the different requirements would help you create the ideal research report. A research

report is usually broken down into multiple sections, which allows for a concise presentation of information.

Structure and Example of a Research Report

- **Title**

This is the title of your systematic investigation. Your title should be concise and point to the aims, objectives, and findings of a research report.

- **Table of Contents**

This is like a compass that makes it easier for readers to navigate the research report.

- **Abstract**

An abstract is an overview that highlights all important aspects of the research including the research method, data collection process, and research findings. Think of an abstract as a summary of your research report that presents pertinent information in a concise manner.

An abstract is always brief; typically 100-150 words and goes straight to the point. The focus of your research abstract should be the 5Ws and 1H format – What, Where, Why, When, Who and How.

- **Introduction**

Here, the researcher highlights the aims and objectives of the systematic investigation as well as the problem which the systematic investigation sets out to solve. When writing the report introduction, it is also essential to indicate whether the purposes of the research were achieved or would require more work.

In the introduction section, the researcher specifies the research problem and also outlines the significance of the systematic investigation. Also, the researcher is expected to outline any jargons and terminologies that are contained in the research.

- **Literature Review**

A literature review is a written survey of existing knowledge in the field of study. In other words, it is the section where you provide an overview and analysis of different research works that are relevant to your systematic investigation.

It highlights existing research knowledge and areas needing further investigation, which your research has sought to fill. At this stage, you can also hint at your

research hypothesis and its possible implications for the existing body of knowledge in your field of study.

- **An Account of Investigation**

This is a detailed account of the research process, including the methodology, sample, and research subjects. Here, you are expected to provide in-depth information on the research process including the data collection and analysis procedures.

In a quantitative research report, you'd need to provide information surveys, questionnaires and other quantitative data collection methods used in your research. In a qualitative research report, you are expected to describe the qualitative data collection methods used in your research including interviews and focus groups.

- **Findings**

In this section, you are expected to present the results of the systematic investigation.

- **Discussion**

This section further explains the findings of the research, earlier outlined. Here, you are expected to present a justification for each outcome and show whether the results are in line with your hypotheses or if other research studies have come up with similar results.

- **Conclusions**

This is a summary of all the information in the report. It also outlines the significance of the entire study.

- **References and Appendices**

This section contains a list of all the primary and secondary research sources.

Tips for Writing a Research Report

- **Define the Context for the Report**

As is obtainable when writing an essay, defining the context for your research report would help you create a detailed yet concise document. This is why you need to create an outline before writing so that you do not miss out on anything.

- **Define your Audience**

Writing with your audience in mind is essential as it determines the tone of the report. If you're writing for a general audience, you would want to present the information in a simple and relatable manner. For a specialized audience, you would need to make use of technical and field-specific terms.

- **Include Significant Findings**

The idea of a research report is to present some sort of abridged version of your systematic investigation. In your report, you should exclude irrelevant information while highlighting only important data and findings.

- **Include Illustrations**

Your research report should include illustrations and other visual representations of your data. Graphs, pie charts, and relevant images lend additional credibility to your systematic investigation.

- **Choose the Right Title**

A good research report title is brief, precise, and contains keywords from your research. It should provide a clear idea of your systematic investigation so that readers can grasp the entire focus of your research from the title.

- **Proofread the Report**

Before publishing the document, ensure that you give it a second look to authenticate the information. If you can, get someone else to go through the report, too, and you can also run it through proofreading and editing software.

How to Gather Research Data for Your Report

- **Understand the Problem**

Every research aims at solving a specific problem or set of problems, and this should be at the back of your mind when writing your research report. Understanding the problem would help you to filter the information you have and include only important data in your report.

- **Know what your report seeks to achieve**

This is somewhat similar to the point above because, in some way, the aim of your research report is intertwined with the objectives of your systematic investigation. Identifying the primary purpose of writing a research report would help you to identify and present the required information accordingly.

- **Identify your audience**

Knowing your target audience plays a crucial role in data collection for a research report. If your research report is specifically for an organization, you would want to present industry-specific information or show how the research findings are relevant to the work that the company does.

- **Create Surveys/Questionnaires**

A survey is a research method that is used to gather data from a specific group of people through a set of questions. It can be either quantitative or qualitative.

A survey is usually made up of structured questions, and it can be administered online or offline. However, an online survey is a more effective method of research data collection because it helps you save time and gather data with ease.

You can seamlessly create an online questionnaire for your research on Formplus. With the multiple sharing options available in the builder, you would be able to administer your survey to respondents in little or no time.

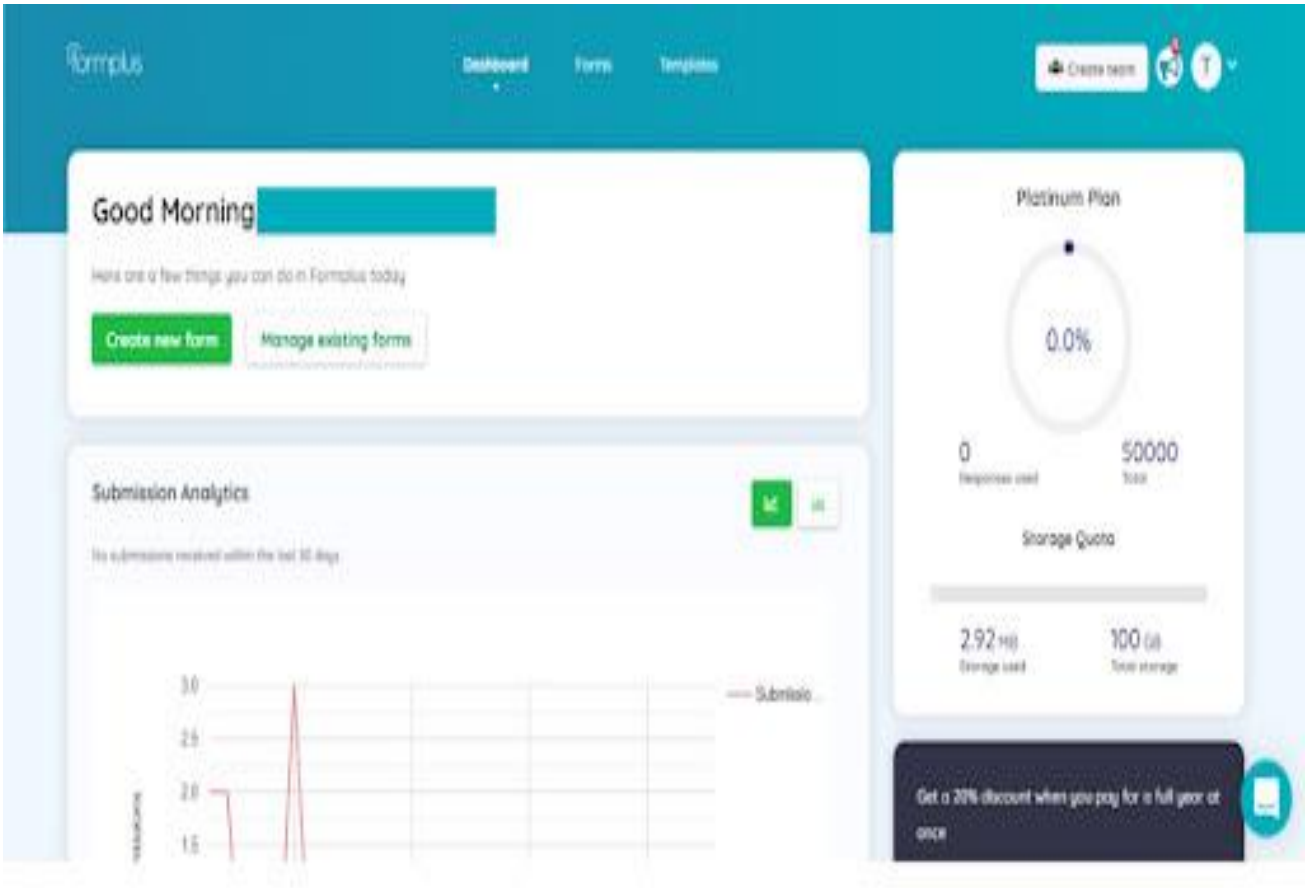
Formplus also has a report summary tool that you can use to create custom visual reports for your research.

Step-by-step guide on how to create an online questionnaire using Formplus

- **Sign into Formplus**

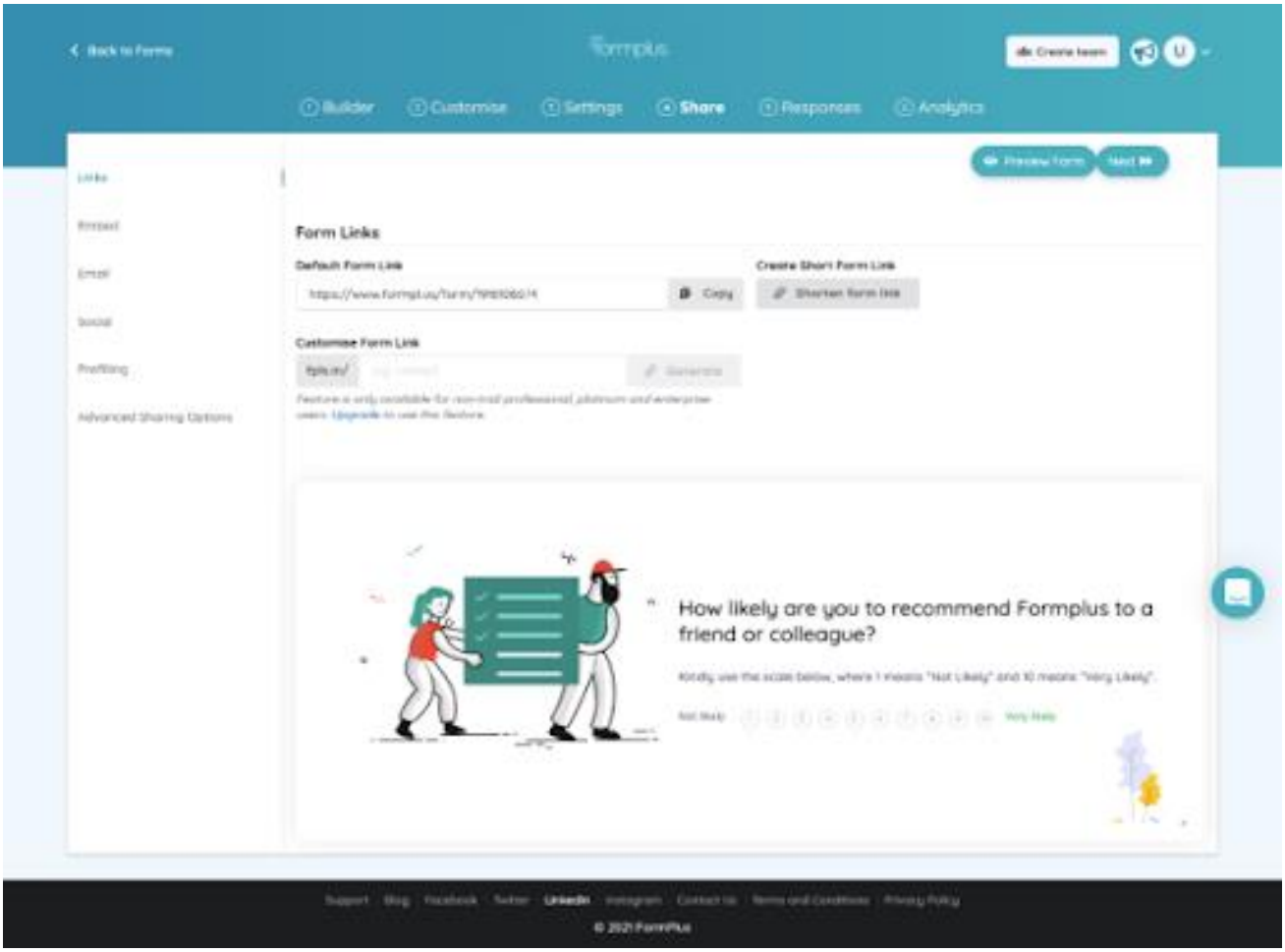
In the Formplus builder, you can easily create different online questionnaires for your research by dragging and dropping preferred fields into your form. To access the Formplus builder, you will need to create an account on Formplus.

Once you do this, sign in to your account and click on **Create new form** to begin.



1. **Edit Form Title**: Click on the field provided to input your form title, for example, “Research Questionnaire.”
2. **Edit Form**: Click on the edit icon to edit the form.
3. **Add Fields**: Drag and drop preferred form fields into your form in the Formplus builder inputs column. There are several field input options for questionnaires in the Formplus builder.
4. Edit fields
5. Click on “Save”
6. **Form Customization:** With the form customization options in the form builder, you can easily change the outlook of your form and make it more unique and personalized. Formplus allows you to change your form theme, add background images, and even change the font according to your needs.

7. **Multiple Sharing Options:** Formplus offers various form-sharing options, which enables you to share your questionnaire with respondents easily. You can use the direct social media sharing buttons to share your form link to your organization's social media pages. You can also send out your survey form as email invitations to your research subjects too. If you wish, you can share your form's QR code or embed it on your organization's website for easy access.



Conclusion

Always remember that a research report is just as important as the actual systematic investigation because it plays a vital role in communicating research findings to everyone else. This is why you must take care to create a concise document summarizing the process of conducting any research.

In this article, we've outlined essential tips to help you create a research report. When writing your report, you should always have the audience at the back of your mind, as this would set the tone for the document.

Ingredients of Research Report

Key Considerations/Factors:

While preparing research report, following issues must be considered:

(i) Objectives

(ii) Type of problem/subject

(iii) Nature and type of research

(iv) Audience or users of research work

(v) Size of report

(vi) Form of writing – handwritten, typed, or computerized.

(vii) Time and cost

(viii) Language

(ix) Contents of report

(x) Order of contents

(xi) Number of copies

(xii) Format – type and size of paper; lengths width, and depth of report; and pattern of writing including paragraph, indent, numbering, font size and type, colouring, etc.

(xiii) Binding (for soft, and, particularly, for hard copy) – type, quality of material, colour, etc., related issues.

Report writing differs from person to person depending on personality, imaginative and creative abilities, experience, and training. However, most researchers agree that following general principles must be kept in mind to produce a better research report. These principles are often called as qualities or requirements of a good report.

1. Selectiveness:

It is important to exclude the matter, which is known to all. Only necessary contents should be included to save time, costs, and energy. However, care should be taken that the vital points should not be missed.

2. Comprehensiveness:

Report must be complete. It must include all the necessary contents. In short, it must contain enough detail to covey meaning.

3. Cost Consideration:

It must be prepared within the budgeted amount. It should not result into excessive costs.

4. Accuracy:

As far as possible, research report must be prepared carefully. It must be free from spelling mistakes and grammatical errors.

5. Objectivity:

Report must be free from personal bias, i.e., it must be free from one's personal liking and disliking. The report must be prepared for impersonal needs. The facts must be stated boldly. It must reveal the bitter truth. It must suit the objectives and must meet expectations of the relevant audience/readers.

6. Clarity:

Report must reveal the facts clearly. Contents and conclusions drawn must be free from ambiguities. In short, outcomes must convey clear-cut implications.

7. Preciseness:

Research report must not be unnecessarily lengthy. It must contain only necessary parts with adequate description.

8. Simplicity:

Report must be simple to understand. Unnecessary technical words or terminologies (jargons) should be avoided.

9. Proper Language:

Researcher must use a suitable language. Language should be selected as per its target users.

10. Reliability:

Research report must be reliable. Manager can trust on it. He can be convinced to decide on the basis of research reposts.

11. Proper Format:

An ideal repost is one, which must be prepared as per commonly used format. One must comply with the contemporary practices; completely a new format should not be used.

12. Attractive:

Report must be attractive in all the important regards like size, colour, paper quality, etc. Similarly, it should use liberally the charts, diagrams, figures, illustrations, pictures, and multiple colours.

Difference Between Reference and Bibliography

BASIS FOR COMPARISON	REFERENCE	BIBLIOGRAPHY
Meaning	Reference implies the list of sources, that has been referred in the research work.	Bibliography is about listing out materials which has been consult during the research work.
Based on	Primary Sources	Both Primary and Secondary Sou
Arrangement	Alphabetically and numerically	Numerically
Includes	Only in-text citations, that have been used in the assignment or project.	Both in-text citations and other sources, that are used to generate idea.
Supporting argument	A reference can be used to support an argument.	A bibliography cannot be used to support an argument.
Used for	Thesis and Dissertation	Journal Papers and Research wor

Difference Between Reference and Bibliography

Comparison Chart

BASIS FOR COMPARISON	REFERENCE	BIBLIOGRAPHY
Meaning	Reference implies the list of sources, that has been referred in the research work.	Bibliography is about listing out materials which has been consult during the research work.

BASIS FOR COMPARISON	REFERENCE	BIBLIOGRAPHY
Based on	Primary Sources	Both Primary and Secondary Sou
Arrangement	Alphabetically and numerically	Numerically
Includes	Only in-text citations, that have been used in the assignment or project.	Both in-text citations and other sources, that are used to generate idea.
Supporting argument	A reference can be used to support an argument.	A bibliography cannot be used to support an argument.
Used for	Thesis and Dissertation	Journal Papers and Research wor

Definition of Reference

Reference can be understood as the act of giving credit to or mentioning the name of, someone or something. In research methodology, it denotes the items which you have reviewed and referred to, in the text, in your research work. It is nothing but a way to acknowledge or indirectly showing gratitude, towards the sources from where the information is gathered.

While using references, one thing is to be noted that you go for reliable sources only, because it increases credence and also supports your arguments. It may include, books, research papers, or articles from magazines, journals, newspapers, etc., interview transcripts, internet sources such as websites, blogs, videos watched, and so forth.

These are used to inform the reader about the sources of direct quotations, tables, statistics, photos etc. that are included in the research work.

Definition of Bibliography

At the end of the research report, bibliography is added, which contains a list of books, magazines, journals, websites or other publications which are in some way relevant to the topic under study, that has been consulted by the researcher during the research. In finer terms, it comprises of all the references cited in the form of footnotes and other important works that the author has studied.

The bibliography is helpful to the reader in gaining information regarding the literature available on the topic and what influenced the author. For better presentation and convenient reading, the bibliography can be grouped into two parts, wherein the first part lists out the names of books and pamphlets consulted, and the other contains the names of magazines and newspapers considered.

Types of Bibliography

- **Bibliography of works cited**: It contains the name of those books whose content has been cited in the text of the research report.
- **Selected Bibliography**: As it is evident from the name itself, selected bibliography covers only those works which the author assumes that are of major interest to the reader.
- **Annotated Bibliography**: In this type of bibliography, a small description of the items covered is given by the author to ensure readability and also improve the usefulness of the book.

Guidelines for the Preparation of a Bibliography

Prepared by the Bibliography Committee, Collection Development and Evaluation Section, Reference and User Services Division, American Library Association, 1992. Revised by the RUSA Standards Committee and approved by the RUSA Board of Directors, June, 2001. Reviewed and updated in June, 2008, by Collection Development Policies and Assessment Committee and approved by RUSA's Standards and Guidelines Committee, July, 2009. Approved by the RUSA Board of Directors, March 2010.

Introduction

These guidelines originated as the "Criteria for Evaluating a Bibliography," adopted by the Reference Services Division Board in 1971. A revision of those guidelines under the current name, "Guidelines for the Preparation of a Bibliography" was approved by the RASD board of directors in 1982 and can be found in RQ22 (Fall 1982): 31-32. The RASD Bibliography Committee based many parts of the 1982 revision on the "Criteria for the Evaluation of Enumerative Bibliographies" prepared by the Committee on Bibliographical Services for Canada in 1979."

"The 1992 committee has proposed less extensive revisions than were adopted in 1982 but wants to strengthen wording regarding annotations and multiple points of access and to make clearer what is meant by standard bibliographic form."

The 2007-09 RUSA/CODES Collection Development Policies and Assessment Committee has proposed minor changes to the document to include electronic bibliographies, author information, timeliness and history.

The original Guidelines were adopted in 1971 by the Reference Services Division Board and revised and updated in 1982 and in 1992 by the Bibliography Committee of the Collection Development and Evaluation Section of the Reference and User Services Association. The 2001 revision was prepared by RUSA's Standards and Guidelines Committee to reflect technological developments since the wide dissemination of print or electronic bibliographies and to make more explicit sound principles involved in the preparation of a bibliography, regardless of its format. They are intended for use by all in the library community.

For purposes of these Guidelines, a 'bibliography' is a systematic list of bibliographic units within a subject (see 3.2). Bibliographies may exist as stand alone works or may appear at the end of research documents. As appropriate to the audience, the author(s) may elect to use a more common term to describe the final bibliography, such as "pathfinder," "finding aid," or "research guide."

Purpose

Ensure that the bibliography fills a significant need in order to justify its compilation.

Fit the subject into the general scheme of available bibliographical sources without unnecessary duplication. If similar bibliographies exist, review them and then explicitly state the unique contribution of this new one.

Clearly state the subject in the title and define the subject in a preliminary statement.

Scope

Clearly define the scope.

Strive for completeness within the stated limitations (period, geographical area, medium, language, library holdings, quality, intended audience, etc.).

Identify and describe each different format appropriately.

Methodology

Provide sources consulted and information on the method of compilation.

Include all available bibliographic units within the subject. A bibliographic unit is an entity in a bibliography: books, journal articles, reports, manuscripts, sound and video recordings, individual web pages and/or entire web sites, computer programs or printouts, films, charts, etc. Identify all items not personally examined by the author(s).

Organization

Principles of organization

Organize the material suitably for both the subject and the targeted users.

Arrange the material so it is possible to use the bibliography from at least one organizational approach without consulting supporting documentation such as an index.

Provide multiple means of access as appropriate. Means of access include both the useful arrangement of materials and the available

methods to search those materials.

Develop the scheme for a classified bibliography so that it is logical and easy for users to understand.

Utilize recognized navigation features and other sound principles relating to layout and file size for bibliographies published on the World Wide Web.

4.2 Necessary components

Provide a statement of scope and purpose for every bibliography.

Provide an explanation of how to use the bibliography.

Provide a key to all abbreviations used in the bibliography.

Provide both a table of contents and an index or indexes. For an electronic version, include search engine or keyword searching capability.

Describe indexes with sufficient detail to provide acceptable levels of recall and precision. Utilize terminology of the indexes appropriate to both subject and intended users.

Provide cross-references adequate for normal reference purposes.

Provide multiple indexes if required for complete access to the materials.

4.3 Desirable features

Consider utilizing entry numbers for bibliographic units.

Consider including location of copies of bibliographical units, if not readily available.

Include links to available full text in electronic bibliographies if there are no copyright issues.

Annotations/Notes

Provide annotations or notes at one of the following three levels:

Informative- Use informative notes chiefly when the nature or reason for inclusion of a title is not clear. Limit use of this minimal level of description to those bibliographies that approach comprehensiveness for the area they are covering.

Annotated- For descriptive annotations include enough of the contents to enable users to decide whether or not they want to view the original. Provide annotations at least at this level for any bibliography designated "annotated."

Critical or Evaluative- Have a knowledgeable person in the field write discriminating critical evaluations and ensure that the value of each item is assessed in relationship to other works in the area. Provide annotations at this level for any bibliography designated "critical" or "evaluative."

5.2 In each case, provide succinct and informative annotations or notes written on a level suitable for the intended users. If another source is drawn upon for the annotation, acknowledge the source appropriately.

Bibliographic Form

Provide sufficient information to identify the bibliographic unit easily for the purpose of the bibliography and needs of the intended user.

Consistently follow a recognized standard bibliographic form. Examples of these standards include, but are not limited to, those described in *The Chicago Manual of Style*, *The MLA Style Manual*, and the *Publication Manual of the American Psychological Association.*

Timeliness

Minimize the time lag between completion of bibliographies and its publication. In the introductory material clearly state when the bibliography was completed.

Issue bibliographies intended to be current as closely as possible after the publication of the bibliographical units listed.

Periodically review the bibliography's life cycle and evaluate its current relevance.

Accuracy

Ensure that citations are correct and free from typographical errors.

Ensure that information provided in annotations and elsewhere is factually accurate and grammatically correct.

Consider provision for corrections after publication.

Format of the Work

Produce the bibliography with clear and appropriate format and typeface.

Produce printed volumes sturdy enough to withstand anticipated use.

Design the bibliography to keep its price within the

means of potential users without sacrificing important features that facilitate its use.

Cumulation

Cumulate ongoing bibliographies whenever possible.

Distribution

Properly advertise and distribute published bibliographies in either print or electronic format as appropriate to the format, including notice to whatever standard national bibliograp