1. Need of business research

The need to make intelligent, informed decisions ultimately motivates an organization to engage in business research. Not every decision requires research. Thus, when confronting a key decision, a manager must initially decide whether to conduct business research

The determination of the need for research centers on:
(1) Time constraints
(2) Availability of data
(3) Nature of the decision to be made.
(4) Value of the research information in relation to costs

Time Constraints

Systematic research takes time. In many instances, management believes that a decision must be made immediately, allowing no time for research. Decisions sometimes are made without adequate information or thorough understanding of the business situation. Although making decisions without researching a situation is not ideal, sometimes the urgency of a situation precludes the use of research. The urgency with which managers usually want to make decisions conflicts with researchers' desire for rigor in following the scientific method.

Availability of Data

Often managers already possess enough data, or information, to make sound decisions without additional research. When they lack adequate information, however, research must be considered.
This means that data need to be collected from an appropriate source. If a potential source of data exists, managers will want to know how much it will cost to get the data. If the data cannot be obtained, or it cannot be obtained in a timely fashion, this particular research project should not be conducted.

Nature of the Decision

The value of business research will depend on the nature of the managerial decision to be made. A routine tactical decision that does not require a substantial investment may not seem to warrant a substantial expenditure for research. For example, a computer company must update its operator's instruction manual when it makes minor product modifications. The research cost of determining the proper wording to use in the updated manual is likely to be too high for such a minor decision. The nature of the decision is not totally independent of the next issue to be considered: the benefits versus the costs of the research. In general, however, the more strategically or tactically important the decision, the more likely it is that research will be conducted.

Benefits versus Costs

Earlier we discussed some of the managerial benefits of business research. Of course, conducting research to obtain these benefits requires an expenditure of money. In any decision-making situation, managers must identify alternative courses of action and then weigh the value of each alternative against its cost. Business research can be thought of as an investment alternative. When deciding whether to make a decision without research or to postpone the decision in order to conduct research, managers should ask three questions:

1. Will the payoff or rate of return be worth the investment?

2. Will the information gained by business research improve the quality of the managerial decision enough to warrant the expenditure?

3. Is the proposed research expenditure the best use of the available funds?

2. Basic research vs Applied research

Basic Research	Applied Research
Research conducted without a specific decision in mind that usually does not address the needs of a specific organization.	Research conducted to address a specific business decision for a specific firm or organization
It focuses on theoretical concepts, principles, and the fundamental laws governing a particular phenomenon.	It focuses on the application of existing knowledge to address a specific question or challenge.
The primary outcome of basic research is the advancement of knowledge in a specific field. It lays the foundation for applied research.	The outcome of applied research is often a solution, product, or method that can be directly implemented to meet a practical need.
The outcome of applied research is often a solution, product, or method that can be directly implemented to meet a practical need.	Tends to have more immediate, tangible outcomes
Tends to have a higher tolerance for uncertainty and the possibility that the research may not have immediate practical applications.	Typically has a lower tolerance for uncertainty, as it is often driven by the need to solve specific problems or meet practical needs.
Allows for flexibility in exploration and may lead to unexpected discoveries with broad implications.	Requires a more focused and structured approach to address specific issues or challenges.
It attempts to expand the limits of knowledge in general and is not aimed at solving a particular pragmatic problem	

3. The scientific method of research

The scientific method is a systematic, logical approach used by scientists to investigate natural phenomena. It involves a series of steps designed to gather empirical evidence, formulate and test hypotheses, and draw objective conclusions. The scientific method provides a structured framework for conducting research and ensures that the process is objective, reproducible, and reliable.
Here are the typical steps involved in the scientific method:

Observation:
The process begins with the observation of a phenomenon or a set of facts in the natural world. Observations can be qualitative (descriptive) or quantitative (involving measurements).

Question:
Based on observations, a researcher formulates a specific question or a set of questions that address the observed phenomenon. The question should be clear, specific, and testable.

Hypothesis:
A hypothesis is a tentative and testable explanation for the observed phenomenon. It is a statement that predicts the relationship between variables and serves as a basis for further investigation. A good hypothesis is falsifiable, meaning it can be proven false through experimentation or observation.

Prediction:
A prediction is a statement about the expected outcome of an experiment or observation if the hypothesis is true. It helps guide the research and provides a way to test the validity of the hypothesis.

Experimentation:
Researchers design and conduct experiments to test the hypothesis and gather data. Experiments involve manipulating independent variables, measuring dependent variables, and controlling other factors (constants) to isolate the effects of the independent variable.

Data Collection:
During the experiment, researchers collect data by making observations and measurements. The data can be qualitative or quantitative and should be recorded accurately.

Analysis:
The collected data are analyzed using statistical methods or other analytical tools. The goal is to identify patterns, trends, or relationships in the data and determine whether these support or refute the hypothesis.

Conclusion:
Based on the analysis, researchers draw conclusions about the hypothesis. If the data support the hypothesis, it may be considered valid, but if the data contradict the hypothesis, it may need to be revised or rejected. Conclusions should be objective and supported by the evidence.

Communication:
Scientists communicate their findings through research papers, presentations, or other means. This allows other researchers to review, replicate, and build upon the study, contributing to the cumulative body of scientific knowledge.

Peer Review and Replication:
The scientific community relies on peer review to assess the validity and reliability of research. Other scientists review the methodology, results, and conclusions of a study before it is accepted for publication. Replication by independent researchers helps verify the results and ensure the reliability of the findings.

4. Research proposition Vs Hypothesis

Research Proposition:

A research proposition is a broad statement or idea that suggests a topic or area of interest for research. It is often the starting point of the research process and serves as a preliminary expression of what the researcher intends to explore or investigate.
Research propositions are typically more general and may not be as specific or testable as hypotheses. They help guide the initial direction of the research and provide a basis for developing more focused research questions or hypotheses.
Example of a research proposition: "There is a relationship between social media usage and mental health outcomes in adolescents."

Hypothesis:
A hypothesis is a specific, testable, and falsifiable statement that predicts the relationship between variables. It is derived from the research proposition and is formulated after a thorough review of existing literature and preliminary observations.
Hypotheses propose a clear and precise expectation regarding the outcome of an experiment or study. They guide the research design and data collection, helping to determine if the observed results support or reject the hypothesis.
Example of a hypothesis based on the research proposition: "Increased social media usage is associated with higher levels of anxiety and depression in adolescents compared to those with lower social media usage."

5. Goals of theory

Suppose a researcher investigating business phenomena wants to know what caused the financial crisis. Another person wants to know if organizational structure influences leadership style. Both of these individuals want to gain a better understanding of the environment and be able to predict behavior; to be able to say that if we take a particular course of action we can expect a specific outcome to occur. These two issues—understanding and predicting—are the two purposes of theory.3 Accomplishing the first goal allows the theorist to gain an understanding of the relationship among various phenomena. For example, a financial advisor may believe, or theorize, that older investors tend to be more interested in investment income than younger investors. This theory, once verified, would then allow her to predict the importance of expected dividend yield based on the age of her customer. Thus a theory enables us to predict the behavior or characteristics of one phenomenon from the knowledge of another phenomenon. The value of understanding and anticipating future conditions in the environment or in an organization should be obvious. In most situations, of course, understanding and prediction go hand in hand. To predict phenomena, we must have an explanation of why variables behave as they do. Theories provide these explanations.

6. Characteristics of valuable information

In data science, the value of information is determined by its quality, relevance, and usefulness for solving specific problems or making informed decisions. Here are some key characteristics of valuable information in data science:

1. Accuracy:
-Definition: Accuracy refers to the correctness of the information.
-Importance: Accurate data ensures that the analysis and insights drawn from it are reliable and trustworthy.

2. Relevance:
- Definition: Relevant information is directly related to the problem or question at hand.
- Importance: Relevant data is essential for making informed decisions and drawing meaningful insights. Unnecessary or irrelevant information can lead to misleading conclusions.

3. Completeness:
- Definition: Complete information includes all the relevant data points needed for analysis.
- Importance: Incomplete data may result in biased or inaccurate conclusions. Complete datasets provide a more comprehensive view of the situation.

4. Consistency:
- Definition: Consistency ensures that data is uniform and doesn't contradict itself.
- Importance: Inconsistent data can lead to confusion and errors in analysis. Consistent data is necessary for producing reliable and valid results.

5. Timeliness:
- Definition: Timely information is up-to-date and relevant to the current context.
- Importance: Timely data is crucial for decision-making in dynamic environments. Outdated information may no longer reflect the current state of affairs.

6. Precision:
- Definition: Precision refers to the level of detail or granularity in the data.
- Importance: Precise data allows for more accurate and specific analyses. However, overly detailed data may not always be necessary and can lead to complexity without added value.

7. Validity:
- Definition: Valid data conforms to the rules and constraints of the data model or system.
- Importance: Validity ensures that the data is suitable for its intended purpose and fits within the defined parameters.

8. Reliability:

-Definition: Reliable data can be consistently reproduced and trusted for accuracy.

- Importance: Reliable data is essential for building trust in the findings and insights derived from the data. It allows for the replication of results by others.

9. Clarity:

- Definition: Clarity refers to the simplicity and transparency of the information.
- Importance: Clear and easily understandable data facilitates effective communication of insights. Complex or unclear data may lead to misunderstandings and misinterpretations.

10. Measurability:

- Definition: Measurable data can be quantified and expressed in numerical terms.
- Importance: Measurable data allows for quantitative analysis and statistical modeling, enabling more rigorous and objective decision-making.

11. Interpretability:

- Definition: Interpretability refers to the ease with which data can be understood and explained.
- Importance: Interpretable data is crucial for stakeholders to comprehend the findings and make informed decisions based on the analysis.

By ensuring these characteristics in the data used for analysis, data scientists can enhance the value of the information and contribute to more accurate, reliable, and actionable insights.

7. Input management

Input management is the process of receiving and storing data or information. It encompasses both digital and physical data, as well as the processes of converting that data into a format that can be used by computers.
Data input can be either manual or automatic. Manual input is entered by a person using a keyboard, scanner, or other input devices. Automatic input is generated by devices such as sensors or bar-code readers.
The process of inputting data also includes validation and error checking. This ensures that the data is accurate and complete before it is stored.

Five major sources of data input:

Internal records

Internal records, such as accounting reports of production costs and sales figures, provide considerable data that may become useful information for managers.

Proprietary Business research

The gathering of new data to investigate specific problems

Salesperson input

Salespeople are typically a business's boundary spanners, the link between the organization and the external environments. Since they are in touch with these outside entities, they commonly provide essential business data.

Behavioral tracking

Modern technology provides new ways of tracking human behavior. Global positioning satellite (GPS) systems allow management to track the whereabouts of delivery personnel at all times. This is the same system that provides directions through an automobile's navigation system.

Outside vendors and external distributors of data.

Outside vendors and external distributors market information as their products. Many organizations specialize in the collection and publication of high-quality information.

8. DSS and CRM

Decision Support System is a computer-based system that helps decision makers confront problems through direct interaction with databases and analytical software programs.
The purpose of a decision support system is to store data and transform them into organized information that is easily accessible to managers. Doing so saves managers countless hours so that decisions that might take days or even weeks otherwise can be made in minutes using a DSS.

A Customer Relationship Management system is the part of the DSS that addresses exchanges between the firm and its customers.
Modern decision support systems greatly facilitate customer relationship management.
It brings together information about customers including sales data, market trends, marketing promotions and the way consumers respond to them, customer preferences, and more. A CRM system describes customer relationships in sufficient detail so that financial directors, marketing managers, salespeople, customer service representatives, and perhaps the customers themselves can access information directly, match customer needs with satisfying product offerings, remind customers of service requirements, and know what other products a customer has purchased

9. Business ethics, Moral Standards and Ethical Dilemma

Business ethics is the application of morals to behavior related to the business environment or context.
Generally, good ethics conforms to the notion of "right," and a lack of ethics conforms to the notion of "wrong." Highly ethical behavior can be characterized as being fair, just, and acceptable.

Moral standards are principles that reflect beliefs about what is ethical and what is unethical.
More simply, they can be thought of as rules distinguishing right from wrong.

Ethical dilemma simply refers to a situation in which one chooses from alternative courses of actions, each with different ethical implications.
Each individual develops a philosophy or way of thinking that is applied to resolve the dilemmas they face. Many people use moral standards to guide their actions when confronted with an ethical dilemma. Others adapt an ethical orientation that rejects absolute principles. Their ethics are based more on the social or cultural acceptability of behavior.

10. Sources of conflict between senior management and research team

Money
Research budgets are a source of conflict between management and researchers. Financial managers often see research as a cost rather than as an investment or a way of lowering risk. Successful decisions that are supported by research are seldom attributed to the researcher. Thus, as is often true in many areas of business, managers often want to spend as little as possible on research

Time

The more quickly the research project is done, the less likely it is to be successful.
This doesn't mean it can't provide valuable information. It simply is not as certain that a quickly put together study will provide valuable answers as would a more deliberately planned project

Unit 2

1. Qualitative Vs Quantitative business research

Qualitative Research	Quantitative Research
A method for developing a better understanding of human and social sciences, in understanding human behaviour and personalities better	It is the method used to generate numerical data by using a lot of techniques such as logical, statistical and mathematical techniques
It employs a subjective approach	It employs an objective approach
It is generally expressed using words	It is expressed using graphs and numbers
It has open-ended questions	It has multiple choice questions
Qualitative research needs only a few respondents	Quantitative research requires many respondents
Qualitative research is holistic in nature	Quantitative Research is particularistic in nature
The reasoning used to synthesize data in this research is inductive	The reasoning used to synthesise data in this research is deductive
This method involves a process-oriented inquiry	This method does not involve a process-oriented inquiry
It develops the initial understanding of data	It recommends a final course of action
The data taken in the Qualitative research method is pretty verbal	The data taken in this method is pretty measurable
The objective of this research method is to engage and discover various ideas	The main objective of Quantitative research is to examine the cause and effect between the variables
It is one of the exploratory research methods	It is a conclusive research method

2. Exploratory research VS confirmatory research

Exploratory Research	Confirmatory Research
The primary goal is to explore and gain insights into a new or poorly understood phenomenon. It is used to generate ideas, hypotheses, and a better understanding of the research problem. Exploratory research is more flexible and open-ended.	The main purpose is to test specific hypotheses or theories that have been formulated prior to the study. It aims to confirm or refute existing theories and is more structured and formal than exploratory research.
It is characterized by flexibility. Researchers use qualitative methods such as literature reviews, interviews, focus groups, or observations to gather data. The approach is open, allowing for the discovery of new insights	It is more rigid in its design. Researchers follow a predefined research plan, often involving quantitative methods such as experiments, surveys, or statistical analyses. The focus is on minimizing bias and ensuring the reliability of results.
Focuses on hypothesis generation. It helps in formulating research questions and hypotheses that can be tested in subsequent confirmatory research.	Involves hypothesis testing. Researchers use statistical analyses to assess whether the collected data supports or contradicts the pre-existing hypotheses.
Involves qualitative methods to gather rich, descriptive data. Common methods include open-ended interviews, focus groups, case studies, or content analysis.	Primarily uses quantitative methods to collect numerical data that can be analyzed statistically. Surveys, experiments, and structured observations are common in confirmatory research.
Statistical analyses are generally less emphasized, and the focus is on understanding patterns, themes, or relationships in the data.	Rigorous statistical analyses are a key component to determine the significance of findings and draw valid conclusions.
Typically conducted at the beginning of the research process when the researcher is exploring a new area or problem.	Conducted after exploratory research, once hypotheses have been formulated. It follows a more structured and systematic approach.

3. Steps of problem definition

1. Understand the business situation—identify key symptoms

A situation analysis involves the gathering of background information to familiarize researchers and managers with the decision-making environment. The situation analysis can be written up as a way of documenting the problem-definition process.
The situation analysis begins with an interview between the researcher and management.

A] INTERVIEW PROCESS
Interrogative techniques simply involve asking multiple what, where, who, when, why, and how questions. They can also be used to provoke introspection, which can assist with problem definition.

B] IDENTIFYING SYMPTOMS
Probing is an interview technique that tries to draw deeper and more elaborate explanations from the discussion.

2. Identify key problem(s) from symptoms

Identifying key problems from symptoms involves, figuring out what's really causing the issues we see enabling a more focused approach to addressing the root causes rather than just treating surface-level manifestations.

3. Write managerial decision statement and corresponding research objectives

Writing managerial decision statements involves expressing the key decisions that managers need to make, while corresponding research objectives are specific, measurable goals that guide the research process in providing relevant information to support those managerial decisions.

4. Determine the unit of analysis

Determining the unit of analysis involves specifying the specific entities or subjects under investigation, defining the distinct units upon which data will be collected and analyzed in a research study.

5. Determine the relevant variables

Determining the relevant variables involves identifying and selecting the key factors or characteristics that are essential to the research study, influencing the outcomes or behaviors being investigated.

6. Write research questions and/or research hypotheses

Writing research questions and/or research hypotheses involves formulating clear and concise inquiries or statements that guide the research process by articulating the specific aspects to be investigated or tested in a study.`

4. Proposal as a contract/planning tool

Proposal as a contract

When an external consultant or research supplier is hired to conduct the research, their written proposal acts as a bid, detailing the specific services they offer. Clients usually seek multiple proposals to compare and evaluate the quality of different research suppliers before making a decision. A wise researcher will not agree to do a research job for which no written proposal exists.

The proposal is like a contract that outlines what the research user is purchasing, essentially serving as a preliminary version of the final research report without the actual results. To prevent misunderstandings and ensure clarity, it's important for both the researcher and the client to sign the proposal, creating a documented agreement on the scope of the research and reducing the likelihood of conflicts after the study is completed.

Proposal as a Planning tool

Creating a research proposal makes researchers think carefully about every step of the research process, turning unclear plans and big ideas into clear, detailed statements about specific events. The proposal, submitted to management for approval, undergoes evaluation based on its potential to provide useful information within a reasonable budget. Revisions are common after the initial review, and the proposal helps managers ensure that the research will obtain the right information and achieve the desired outcomes, aligning with the client's needs.

5. Ethnography, Phenomenology, Hermeneutics

Ethnography

Ethnography represents ways of studying cultures through methods that involve becoming highly active within that culture. Participant-observation typifies an ethnographic research approach. Participant observation means the researcher becomes immersed within the culture that he or she is studying and draws data from his or her observations. Sometimes, researchers become long-term employees of an organization, blending into its culture, and over time, colleagues start behaving naturally around them. This allows the researcher to observe behaviors that employees might not openly share otherwise. For example, a researcher studying the ethical behavior of salespeople might struggle to get a car salesperson to admit to potentially deceptive sales tactics in a standard interview. However, by using ethnographic techniques and becoming part of the workplace, the salesperson may be more open, leading to more accurate insights into the culture of car selling

Phenomenology

Phenomenology represents a philosophical approach to studying human experiences based on the idea that human experience itself is inherently subjective and determined by the context in which people live.The phenomenological researcher focuses on how a person's behavior is shaped by the relationship he or she has with the physical environment, objects, people, and

Hermeneutics

Hermeneutics is an approach to understanding phenomenology that relies on analysis of texts in which a person tells a story about him or herself.

Hermeneutic unit refers to a text passage from a respondent's story that is linked with a key theme from within this story or provided by the researcher

6. Personal interview vs Focus group, Focus group moderator

Personal Interviews	Focus Group Interviews
A personal interview is a form of direct communication in which an interviewer asks respondents questions face-to-face. This versatile and flexible method is a two-way conversation between interviewer and respondent	An unstructured and free flowing interview with a small group of around six to ten people. Focus groups are led by a trained moderator who follows a flexible format encouraging dialogue among respondents.
Personal interviews are well-suited for gathering detailed and individualized information.	Focus group interviews are employed to understand group dynamics, explore shared opinions, and uncover collective perspectives on a given topic.
Data collected in personal interviews is often more detailed and personalized.	Data from focus group interviews include interactions between participants
Analysis of personal interviews involves a detailed examination of each individual's responses.	Analysis of focus group interviews focuses on the interactions within the group.

.

Focus Group Moderator

A person who leads a focus group interview and ensures that everyone gets a chance to speak and contribute to the discussion.

The moderator must be able to develop rapport with the group to promote interaction among all participants. The moderator should be someone who is really interested in people, who listens carefully to what others have to say, and who can readily establish rapport, gain people's confidence, and make them feel relaxed and eager to talk.

7. Variable and types of variables (categorical, discrete, continuous etc), Dependent vs independent variable

Variable

A variable is anything that varies or changes from one instance to another. Variables can exhibit differences in value, usually in magnitude or strength, or in direction.

In research, a variable is either observed or manipulated, in which case it is an experimental variable.

TYPES OF VARIABLES

Continuous variable

A continuous variable is one that can take on a range of values that correspond to some quantitative amount.

Categorical variable

A categorical variable is one that indicates membership in some group.

Categorical variables sometimes represent quantities that take on only a small number of values

Dependent variable

A process outcome or a variable that is predicted and/or explained by other variables

Independent variable

A variable that is expected to influence the dependent variable in some way.

Dependent variable	Independent variable
A process outcome or a variable that is predicted and/or explained by other variables	A variable that is expected to influence the dependent variable in some way.
It is the outcome or response that researchers are interested in understanding or explaining.	It is the presumed cause or factor that is believed to have an effect on the dependent variable.

8. Primary VS Secondary data

Primary Data	Secondary Data
Primary data is original data collected directly from firsthand sources for a specific research purpose.	Secondary data is pre-existing data that has been collected for a purpose other than the current research project.
Researchers gather primary data through methods such as surveys, interviews, observations, experiments, or focus groups.	This data is not collected by the researcher but is obtained from sources such as books, articles, government reports, databases, or previously conducted research studies.
Primary data is unique to the study and is collected specifically to address the research questions or objectives	Secondary data is not original to the current study; it has been collected for a different purpose or by someone else.

9. Sources of secondary data

Secondary data can be classified as either internal to the organization or external.

Internal data
Internal data should be defined as data that originated in the organization, or data created, recorded, or generated by the organization

External Data
External data are generated or recorded by an entity other than the researcher's organization

10. Evaluation of secondary data through questions

Source Credibility
Where did the data come from?
Is the source reputable and trustworthy?
Were the data collected by a reliable and unbiased organization?

Data Collection Methods:
How was the data collected?
Are the methods used to collect the data appropriate for the research question?
Were standardized procedures followed during data collection?

Data Accuracy:
Can the accuracy of the data be verified?
Are there any known errors or discrepancies in the data?
Were the data collected using reliable instruments or procedures?

Data Consistency:
Are there inconsistencies or discrepancies in the data?
Do the data align with other reliable sources or previously collected data?
Were the data collected consistently across different variables?

Sampling Representativeness:
How was the sample selected?
Does the sample represent the population of interest?
Are there any biases in the sampling method?

Unit 3

1. What can be observed

Observational studies gather a wide variety of information about behavior.

Physical actions,
Such as shopping patterns (in-store or via a Web interface) or television viewing;

Verbal behavior
Such as sales conversations or the exchange between a worker and supervisor;

Expressive behavior
Such as tone of voice, facial expressions, or a coach stomping his foot;

Spatial relations and locations, such as traffic patterns;
Temporal patterns, such as amount of time spent shopping, driving, or making a business decision;

Physical objects,
Such as the amount of newspapers recycled or number of beer cans in the trash;

Verbal and pictorial records,
Such as the content of advertisements or the number of minorities pictured in a company brochure

2. Direct observation

A straightforward attempt to observe and record what naturally occurs; the investigator does not create an artificial situation.

Direct observation can produce detailed records of what people actually do during an event. The observer plays a passive role, making no attempt to control or manipulate a situation, instead merely recording what occurs.

Many types of data can be obtained more accurately through direct observation than by questioning. For example, recording traffic counts or observing the direction of customer movement within a supermarket can help managers design store layouts that maximize the exposure of departments that sell impulse goods. A manufacturer can determine the number of facings, shelf locations, display maintenance, and other characteristics that improve store conditions. If directly questioned in a survey, most shoppers would be unable to accurately portray the time they spent in each department. The observation method, in contrast, could determine this without difficulty.

3. Demand characteristics

The term demand characteristic refers to an experimental design element that unintentionally provides subjects with hints about the research hypothesis.

Experimenter Bias and Demand Effects

Experiment bias happens when a researcher unintentionally influences participants or outcomes due to their own expectations, beliefs, or personal involvement in the study, which can compromise the study's fairness and accuracy.

When participants adjust their responses to align with the experimenter's expectations, they may not accurately represent their actual behavior in real-life situations, creating a "demand effect." For instance, if participants in an advertising study know the researcher is assessing attitude changes based on an advertisement, they might provide responses they believe the experimenter wants to hear, rather than reflecting a genuine effect of the experimental treatment.

Hawthorne Effect

The Hawthorne Effect is when people change their behavior because they know they are being observed or studied, often leading to alterations in their actions that don't necessarily represent their typical conduct

4. Errors in survey research and it's types

Random Sampling Error
Random sampling error is a function of sample size. As sample size increases, random sampling error decreases.

Random sampling error is a type of error that occurs when the individuals or items selected for a sample from a larger population are not perfectly representative of that population due to chance. In simpler terms, it's the discrepancy between the characteristics of a sample and the characteristics of the entire population that arises purely by random chance.

Systematic Error
Systematic sampling error occurs when there is a consistent and recurring pattern of deviation between the sample and the population due to a systematic flaw in the sampling method. Unlike random sampling error, which is caused by chance, systematic sampling error is introduced by a specific feature or flaw in the sampling process.

Respondent Error
Respondent error, also known as respondent bias, occurs when survey participants provide inaccurate or misleading information due to various factors. This type of error can impact the reliability and validity of survey data

5. Structured vs Unstructured, Disguised vs Undisguised questionnaire

Structured	Unstructured
In a structured questionnaire, the questions are formulated in a standardized and predetermined manner with fixed response options.	In an unstructured questionnaire, questions are open-ended, allowing respondents to provide detailed and unrestricted responses.
Closed-ended questions dominate, providing respondents with specific response choices.	Questions are open-ended, encouraging respondents to express their thoughts in their own words.
The format is formal and rigid, ensuring uniformity in data collection.	The format is flexible, allowing for a range of responses that may not be predefined.
Responses are easily quantifiable and lend themselves well to statistical analysis.	Responses are often qualitative and require more extensive analysis.
Facilitates efficient data processing and analysis.	Allows for the discovery of unexpected themes or issues.

Disguised	Undisguised
the true purpose of the research is intentionally hidden or not fully disclosed to the respondents.	the true purpose and nature of the research are openly communicated to the respondents.
The research objectives, sponsor, or nature of the study may be concealed from the respondents.	The research objectives and sponsor are clearly stated, and respondents are aware of the study's purpose.
Disguised questionnaires often use indirect or veiled language to gather unbiased responses.	Questions are straightforward, and the language is direct in conveying the study's intentions.
Commonly employed in studies where revealing the true purpose might lead to response bias or influence participant behavior.	Often used in studies where transparency is not expected to significantly impact participant responses or behavior.
Encourages more natural and unbiased responses	May be more ethical in situations where informed consent is a priority.

6. Types of questions in questionnaires

A] Simple-Dichotomy (dichotomous) Question

The simple-dichotomy (dichotomous) question requires the respondent to choose one of two alternatives. The answer can be a simple “yes” or “no” or a choice between “this” and “that.”

For example:
Did you have any overnight travel for work-related activities last month? **Yes or No**

B] Determinant-Choice Question

The determinant-choice question requires the respondent to choose one—and only one—response from among several possible alternatives.

For example:
Please give us some information about your flight. In which section of the aircraft did you sit?

Option 1: First Class
Option 2: Business Class
Option 3: Coach Class

C] Frequency-Determination Question

The frequency-determination question is a determinant-choice question that asks for an answer about the general frequency of occurrence. For example:

How frequently do you watch MTV?

1. Every day
2. 5–6 times a week
3. 2–4 times a week
4. Once a week
5. Less than once a week
6. Never

D] Checklist Question

The checklist question allows the respondent to provide multiple answers to a single question.
The respondent indicates past experience, preference, and the like merely by checking off items.
In many cases the choices are adjectives that describe a particular object. A typical checklist question might ask the following:

Please check which, if any, of the following sources of information about investments you regularly use.

Personal advice of your broker(s)
Brokerage newsletters
Brokerage research reports
Investment advisory service(s)
Conversations with other investors
Web page(s)
None of these
Other (please specify) _____

7. Pretesting

It is a screening procedure that involves a trial run with a group of respondents to iron out fundamental problems in the survey design
Pretesting refers to the process of testing a survey instrument, such as a questionnaire or survey questionnaire, before it is administered to the actual study participants. The purpose of pretesting is to identify and address potential issues, errors, or ambiguities in the survey instrument, ensuring its clarity, reliability, and validity

Unit 4

1. Levels of scale measurement (nominal ordinal interval ratio)

Nominal scales
Nominal scales represent the most elementary level of measurement
Nominal scales are a type of measurement scale in which data is categorized into distinct, non-ordered groups or categories without any inherent ranking or numerical value assigned. These categories represent different qualitative attributes or distinct identities, but they do not imply any quantitative relationship between them. Examples include categories like colors, gender, or types of fruit.

Ordinal Scales
Ordinal scales are a type of measurement scale that orders data into distinct categories or ranks based on the relative magnitude or order of the attributes, but the intervals between categories are not uniformly defined. While ordinal scales convey the order or ranking of items, they do not provide information about the precise differences in magnitude between them. Examples include rankings in a competition, customer satisfaction levels (e.g., low, medium, high), or education levels (e.g., high school, college, postgraduate).

Interval Scales
Interval scales are a type of measurement scale where the intervals between consecutive points are equal and meaningful. They provide a ranking of values and allow for comparisons of the magnitude of differences between values. However, interval scales do not have a true zero point, making ratios between values meaningless. Examples include the Celsius temperature scale and IQ scores

Ratio Scales
Ratio scales are a type of measurement scale that not only maintains equal and meaningful intervals between consecutive points but also features a true zero point, signifying the absence of the measured quantity. This allows for meaningful ratios between values, making it possible to say one value is, for instance, twice or three times as much as another. Examples of ratio scales include height, weight, income, and age.

2. Ranking rating sorting

Ranking
A ranking task requires the respondent to rank order a small number of stores, brands, feelings, or objects on the basis of overall preference or some characteristic of the stimulus.

Rating
A measurement task that requires respondents to estimate the magnitude of a characteristic or quality that a brand, store, or object possesses

Sorting

A measurement task that presents a respondent with several objects or product concepts and requires the respondent to arrange the objects into piles or classify the product concepts.

3. Coding and reverse coding

Coding

Coding, in the context of data analysis or programming, involves assigning labels, symbols, or numerical values to represent different categories, variables, or information in a systematic way. It is a process that simplifies and organizes data, making it more manageable for analysis or computational tasks.

Reverse coding

Reverse coding, also known as reverse scoring, is a technique in survey design or psychological assessments where the scoring of certain items or questions is reversed to counteract response bias. It involves changing the scoring direction of specific items so that higher scores on those items now reflect the opposite of the original construct being measured. This helps ensure more accurate and unbiased interpretation of the overall scores

4. Criteria for good measurement

Reliability

Reliability refers to the consistency and stability of a measurement tool. A reliable measure produces consistent results when used under the same conditions.
High reliability is indicated by consistent outcomes across repeated measurements.

INTERNAL CONSISTENCY

Internal consistency is a measure of the degree to which different items or questions within a measurement instrument, such as a survey or a test, consistently measure the same underlying construct or concept. It assesses the reliability of the instrument by examining the extent to which items that are supposed to measure the same thing produce similar results.

TEST RETEST RELIABILITY

Test-retest reliability is a measure of the consistency or stability of a test or measurement over time. It assesses whether the same individuals, when tested with the same instrument on two separate occasions, produce similar or consistent scores.

Validity

Validity assesses the extent to which a measurement tool accurately measures what it is intended to measure. It ensures that the instrument is capturing the concept or construct of interest.

5.Purpose of sampling

The purpose of sampling in research is to select a subset of individuals or elements from a larger population for the purpose of making inferences or generalizations about the entire population. Sampling is a practical and efficient way to study a population without having to examine every single member.

The key purposes of sampling include

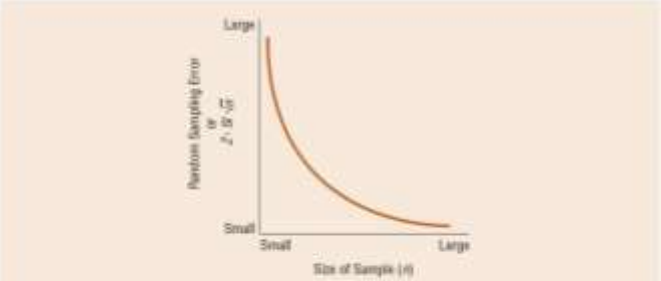
Representativeness
Efficiency
Feasibility
Precision
Inference
Minimizing Bias
Logistics
Statistical Analysis

6. Probability vs non probability sampling

Probability	Non Probability
Probability sampling is a sampling technique where each member of the population has a known and non-zero chance of being selected for the sample.	Non-probability sampling is a sampling technique where the likelihood of any particular member of the population being chosen for the sample is unknown or cannot be determined.
It involves random selection, and every individual or element in the population has an equal opportunity to be chosen.	The selection is based on the judgment or convenience of the researcher.
Probability sampling methods aim to create samples that are representative of the entire population.	Non-probability samples are often less representative of the entire population compared to probability samples.
Probability sampling allows for the application of statistical techniques to make inferences about the entire population based on the characteristics of the sample	The goal may be more focused on exploring specific characteristics or phenomena rather than making population-level inferences.
Common types of probability sampling include simple random sampling, stratified random sampling, systematic sampling, and cluster sampling	Non-probability sampling is often more practical and cost-effective in certain situations, especially when time and resources are limited.

7. Relationship between sample size and random error

The relationship between sample size and random error is characterized by an inverse relationship—larger sample sizes contribute to reduced random error, leading to more precise and reliable statistical estimates. Researchers must carefully consider sample size in study design to balance the need for precision with practical constraints



Inverse Relationship:
There is an inverse relationship between sample size and random error. As sample size increases, random error tends to decrease, and as sample size decreases, random error tends to increase
The population variability also affects the sampling error. More variable populations give rise to larger errors as the samples or the estimates calculated from different samples are more likely to have greater variation. The effect of the variability within the population can be reduced by increasing the sample size to make it more representative of the survey population. Various sample design options also affect the size of the sampling error. For example, stratification reduces sampling error whereas cluster sampling tends to increase it

8. Normal Distribution

A normal distribution, also known as a Gaussian distribution or bell curve, is a symmetrical probability distribution characterized by a bell-shaped curve. The normal distribution is a fundamental concept in statistics and probability theory
It has several key properties:

Symmetry
The normal distribution is symmetric around its mean, which is located at the center of the distribution. The left and right tails of the distribution are mirror images of each other.

Bell Shaped Curve
The curve of the normal distribution is bell-shaped, with the highest point at the mean. As values move away from the mean in either direction, the frequency of occurrence decreases.

Mean, Median & Mode
In a normal distribution, the mean, median, and mode are all equal and located at the center of the distribution. This is a unique property of the normal distribution.

Standard Deviation
The spread or dispersion of values in a normal distribution is determined by the standard deviation. The larger the standard deviation, the wider the distribution.

Z-score
The z-score is a measure of how many standard deviations a particular data point is from the mean in a normal distribution. It is calculated as the difference between a data point and the mean divided by the standard deviation.

Unit 5

1. Cross Tabulation

Cross-tabulation, also known as contingency table analysis, is a statistical technique used to analyze the relationship between two categorical variables. It involves creating a table that displays the joint distribution of the variables, showing how the frequencies or percentages are distributed across the different categories of each variable.

Cross-tabulation involves two categorical variables. These variables can be nominal or ordinal, representing different categories or groups.

A contingency table is created to organize the joint distribution of the two variables. The rows represent categories of one variable, and the columns represent categories of the other variable.

The intersection of each row and column in the contingency table contains the frequency or count of observations that fall into that specific combination of categories.

Often, percentages or proportions are calculated within each cell, representing the proportion of observations in that cell relative to the total sample.

Cross-tabulation allows for the visual and quantitative analysis of relationships between the two variables. It helps identify patterns, trends, or associations between categories.

2. Data Transformation

Data transformation is a process in which the original data is converted or manipulated into a new format to make it more suitable or meaningful for analysis. It is a common step in the data preprocessing phase and is often employed to address issues such as skewed distributions, non-linearity, or to meet the assumptions of certain statistical methods

Collapsing or combining adjacent categories of a variable is a common form of data transformation used to reduce the number of categories.

For example, many researchers believe that less response bias will result if interviewers ask respondents for their year of birth rather than their age. This presents no problem for the research analyst, because a simple data transformation is possible. The raw data coded as birth year can easily be transformed to age by subtracting the birth year from the current year.

3. Type I Type II errors

Type I and Type II errors are concepts in statistical hypothesis testing, representing the errors that can occur when making decisions about a null hypothesis. These errors are inherent in the testing process and are related to the acceptance or rejection of a null hypothesis based on sample data

As the significance level (α) decreases the probability of Type II errors (β) may increase, and vice versa.

Type I Error (False Positive or Alpha Error):
Occurs when the null hypothesis is incorrectly rejected when it is actually true.
Denoted as α (alpha)
The probability of committing a Type I error is equal to the chosen significance level (α), typically set at 0.05 or 0.01.

Type II Error (False Negative or Beta Error):
Occurs when the null hypothesis is incorrectly not rejected when it is actually false
Denoted as β (beta).
The probability of committing a Type II error is influenced by factors such as the sample size, effect size, and the chosen significance level.

4. Correlation

Correlation is a statistical technique used to measure the degree of association or relationship between two variables. It quantifies the strength and direction of a linear relationship between two continuous variables. The result of a correlation analysis is expressed as a correlation coefficient, which ranges from -1 to 1.
The correlation coefficient is a numerical value that represents the strength and direction of the linear relationship between two variables.
It ranges from -1 to 1, where:

r=1: Perfect positive correlation
r=−1: Perfect negative correlation
r=0: No linear correlation

If r is positive, it indicates a positive or direct correlation, meaning that as one variable increases, the other tends to increase as well.
If r is negative, it indicates a negative or inverse correlation, meaning that as one variable increases, the other tends to decrease

5. ANOVA vs MANOVA

ANOVA	MANOVA
ANOVA is used to compare means among three or more independent groups or levels of a single categorical independent variable.	MANOVA extends ANOVA to situations where there are multiple dependent variables.
ANOVA is typically applied to analyze a single dependent variable, which should be continuous and normally distributed	MANOVA allows the simultaneous analysis of multiple dependent variables, assessing whether there are significant differences in the vector of means across groups.
The assumption of homogeneity of variances is important in ANOVA, meaning that the variance within each group should be approximately equal.	Similar to ANOVA, MANOVA assumes homogeneity of variances across groups for each dependent variable. Additionally, it assumes multivariate normality.
If ANOVA indicates a significant difference, post-hoc tests (e.g., Tukey's HSD or Bonferroni) can be conducted to identify which specific groups differ from each other.	If the multivariate test is significant, post-hoc tests (e.g., Bonferroni correction) can be conducted for each dependent variable separately.

Comparing mean exam scores of students across different teaching methods (e.g., Method A, Method B, Method C).	Assessing whether there are significant differences in the mean scores of both Math and English exams across different teaching methods.

6. When to choose one tailed vs two tailed tests

The decision to use a one-tailed or two-tailed test in statistical hypothesis testing depends on the nature of the research question and the specific hypotheses being tested

Use One Tailed if	Use Two Tailed if
Use when there is a specific directional hypothesis, and you are interested in detecting an effect in one direction only (either an increase or a decrease).	Use when the hypothesis is non-directional, and you want to detect whether there is a significant effect, regardless of the direction.
Appropriate when there is a clear, a priori expectation or theoretical basis for the direction of the effect.	Appropriate when there is no specific expectation regarding the direction of the effect
Useful when the cost or consequences of only considering one direction are justified.	Useful when you want to be sensitive to the possibility of an effect in either direction.
Generally has higher statistical power than a two-tailed test because it concentrates the testing in one direction	Lower statistical power compared to a one-tailed test because the critical region is split between two directions.
More likely to detect a significant effect if it exists in the specified direction.	Requires a larger sample size to achieve the same power as a one-tailed test.

7. T test, Z test, Chi squared goodness of fit test (what they are, when do you need to use them? How to choose which test to use? etc)

T-Test

A t-test is a statistical test used to determine if there is a significant difference between the means of two groups.
It's commonly used when working with small sample sizes.

Z-Test

A Z-test is a statistical method used to assess whether a sample mean is significantly different from a known or hypothesized population mean
It is applicable when the population standard deviation is known and is often used in situations with larger sample size

Chi Squared Goodness Of Fit Test

The Chi-squared goodness-of-fit test is a statistical test used to determine whether there is a significant difference between the observed and expected frequencies of categorical data
It is commonly used when dealing with categorical variables and is used to assess whether the distribution of observed categorical data matches an expected distribution.

How to choose which test to use?

Use a t-test when comparing means of two groups or means of paired observations with continuous data and small to moderate sample sizes.

Use a z-test when comparing a sample mean to a known or hypothesized population mean, and the population standard deviation is known, especially with large sample sizes.

Use a Chi-squared goodness-of-fit test when dealing with categorical data to assess whether observed frequencies match expected frequencies.

8. Univariate vs Bivariate vs Multivariate Techniques

Univariate	Bivariate	Multivariate
Univariate analysis involves the examination and interpretation of a single variable in isolation.	Bivariate analysis involves the study of the relationship between two variables.	Multivariate analysis involves the simultaneous study and analysis of multiple variables.
Visualizing the distribution through histograms, box plots, or summary statistics	Visualizing relationships through scatter plots or correlation coefficients.	Common techniques include multivariate regression, factor analysis, and cluster analysis
Simple analysis focused on a single variable.	Examining the interrelation between two variables.	Handling the complexity of multiple variables interacting with each other
Eg. Analyzing the scores of students in a class without considering other factors.	Eg. Investigating the relationship between study hours and exam scores.	Eg. Assessing the impact of study hours, sleep, and diet on academic performance.