AL-FARABI KAZAKH NATIONAL UNIVERSITY
FACULTY OF BIOLOGY AND BIOTECHNOLOGY
DEPARTMENT OF BIOTECHNOLOGY

Lecture 2

FUNDAMENTALS OF SCIENTIFIC RESEARCH

PhD, Acting Associate Professor, Doktyrbay Gulina

Aim of the Lecture

To provide students with a comprehensive understanding of the fundamental principles, characteristics, and types of scientific research, and to explain the importance of research methodology in conducting reliable and evidence-based scientific investigations

By the end of this lecture, students will be able to:

1. Define the concept and purpose of scientific research.

2. Identify the main characteristics and principles of scientific research

3. Distinguish between basic, applied, and experimental research.

4. Explain the role of scientific methodology in research.

5. Describe the importance of objectivity, reliability, and validity in scientific investigations

WHAT IS SCIENTIFIC RESEARCH?

Definition: Scientificresearchisasystematic,controlled, empirical,and critical investigation of hypothetical propositions about presumed relationships among natural and social phenomena.

Core Purpose: To uncover universal truths, solve complex real-world problems, and expand the boundary of human knowledge through rigorous, reproducible methodologies.

Everyday Observation: Informal, unstructured, and subjective. Based on personal experience, intuition, and casual perception without controls or systematic protocols.

Scientific Investigation: Structured around rigorous methodologies designed to minimize subjective bias, ensure reproducibility, and produce verifiable, peer-reviewed empirical evidence.

CORE ATTRIBUTES OF SCIENCE

Empirical: Based ondirect,observable,and measurable evidencegatheredthrough systematic field or laboratory investigation. Claims must be grounded in real-world data, not speculation or belief.

Objective: Free from personal bias, subjective emotions, or unverified beliefs. Scientific findings must reflect reality as it is, not as the researcher wishes it to be.

Replicable: Independent researchers following the same protocol must obtain identical or compatible results. Replicability is the cornerstone of scientific credibility and trust.

Tentative: Open to revision or refutation when stronger empirical evidence emerges. Scientific knowledge evolves continuously as new data challenges or refines existing theor ies.

BASIC AND APPLIED RESEARCH HORIZONS

Basic Research: Expands foundational theoretical knowledge; driven by curiosity and scientific discovery with long-term horizons

Applied Research: Utilizes existing theory to address specific, urgent practical problems in technology, medicine, or social policy

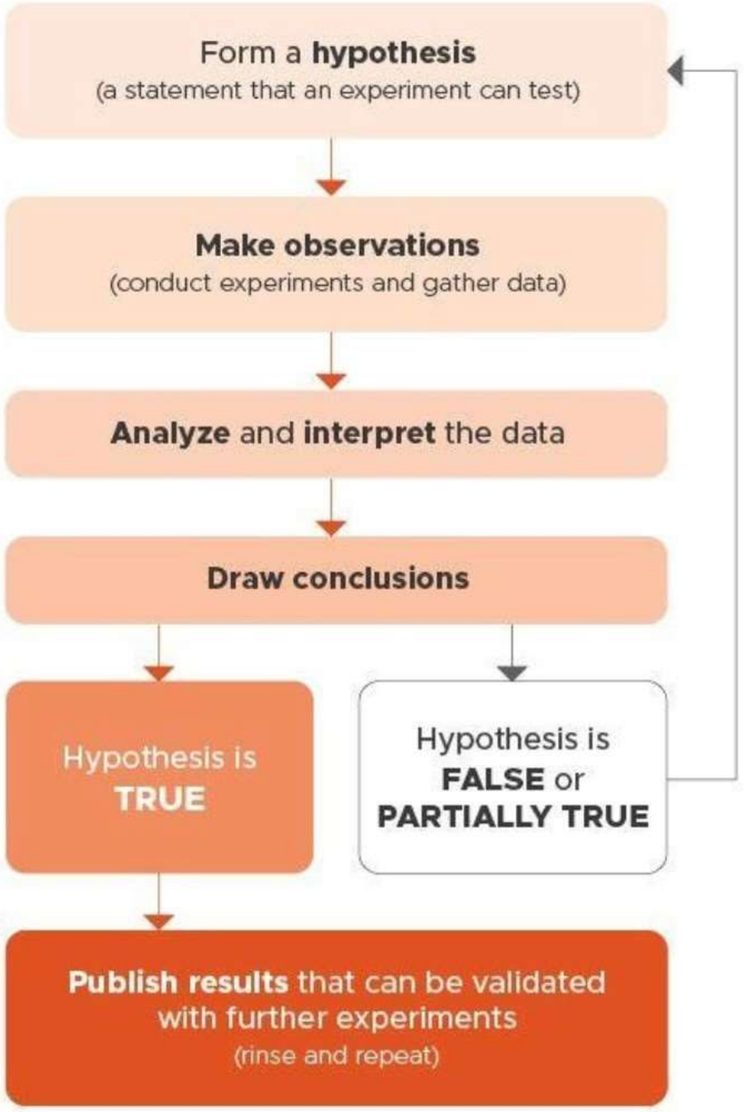
Mixed/Translational: Bridges theory and practice; translates basic findings into real-world solutions and innovations

THE ITERATIVE SCIENTIFIC CYCLE

Problem Formulation: Identifying a research gap or unexplained phenomenon to investigate.

Hypothesis Development: Formulating a testable prediction about variable relationships. **Experimentation & Data Collection:** Designing and conducting controlled experiments to gather evidence.

Data Analysis & Conclusion: Applying statistical methods to interpret empirical findings. Theory Revision & Refinement: Updating theoretical models based on new evidence and insights.



THE STARTING POINT OF INQUIRY

Identifying a BroadSubjectArea: Selecting ageneraldomainofscientificinterest **Narrowing to a Focused Topic:** Refining the subject into a specific, manageable scope Detecting the Research Gap: Identifying unexplained anomalies or missing knowledge Formulating the Research Problem: Defining a concise, scientifically relevant problem statement

Stating the Specific Research Question: Producing a focused, testable, and feasible inquiry question

LITERATURE REVIEW

Primary Sources: Originalpeer-reviewed journalarticles,dissertations, and first-hand empirical reports. Primary sources provide direct evidence and original findings that form the core of any rigorous literature review.

Secondary Literature: Textbooks, review articles, and meta-analyses that synthesize existing primary research. Secondary sources provide broader context, identify trends, and highlight consensus or contradictions across multiple studies.

Critical Evaluation: Systematically assessing the methodological rigor, sample validity, and analytical soundness of each source. Critical evaluation identifies limitations in prior studies and justifies the need for new research.

Theoretical Framing: Situating new research within established theoretical frameworks derived from the literature. Theoretical framing defines key concepts, clarifies relationships between variables, and anchors the study in scholarly tradition.

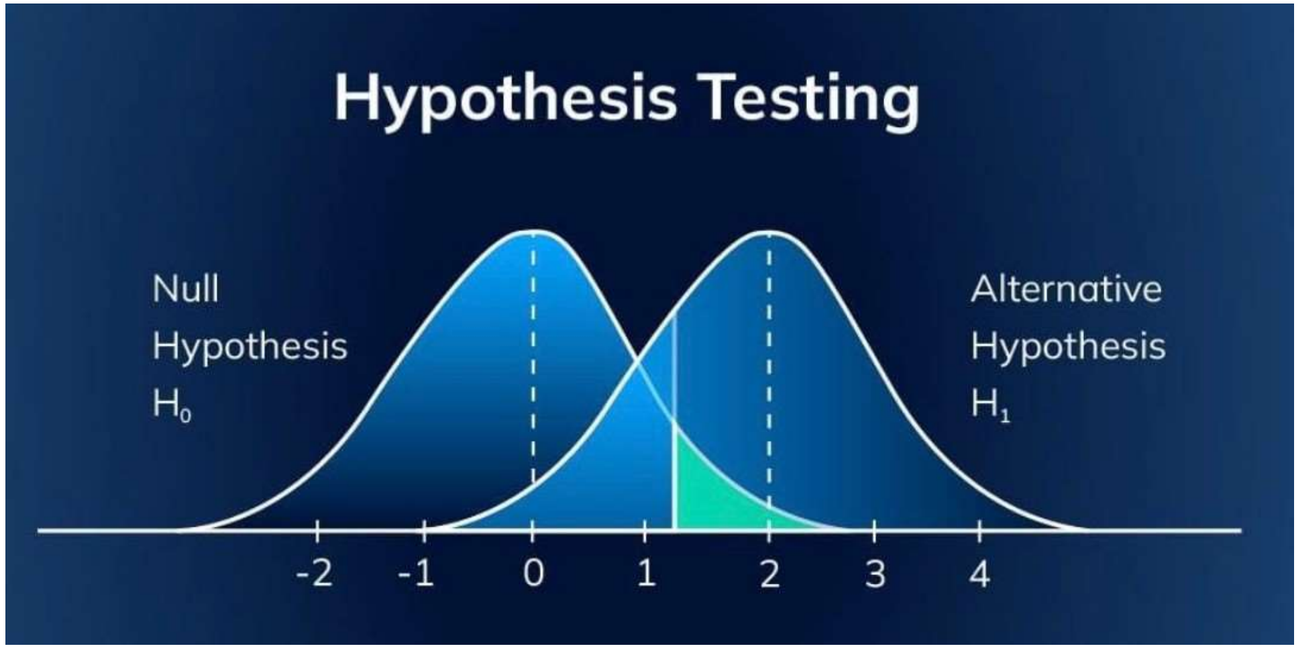
CRAFTING TESTABLE STATEMENTS

Null Hypothesis (H_0): Assumes no significant difference, effect, or relationship exists between the studied variables. It serves as the default position that the researcher attempts to disprove through empirical evidence.

Alternative Hypothesis (H_1): Posits that a real effect, difference, or relationship exists within the target population. It represents the researcher's prediction and is accepted when H_0 is statistically rejected.

Statistical Significance: A hypothesis is rejected or retained based on the p-value threshold ($\alpha = 0.05$). If $p < \alpha$, the null hypothesis is rejected in favor of the alternative hypothesis.

Operationalization: Translates abstract research questions into precise, measurable variables. A well-formed hypothesis must be specific, falsifiable, and directly testable through empirical data collection.

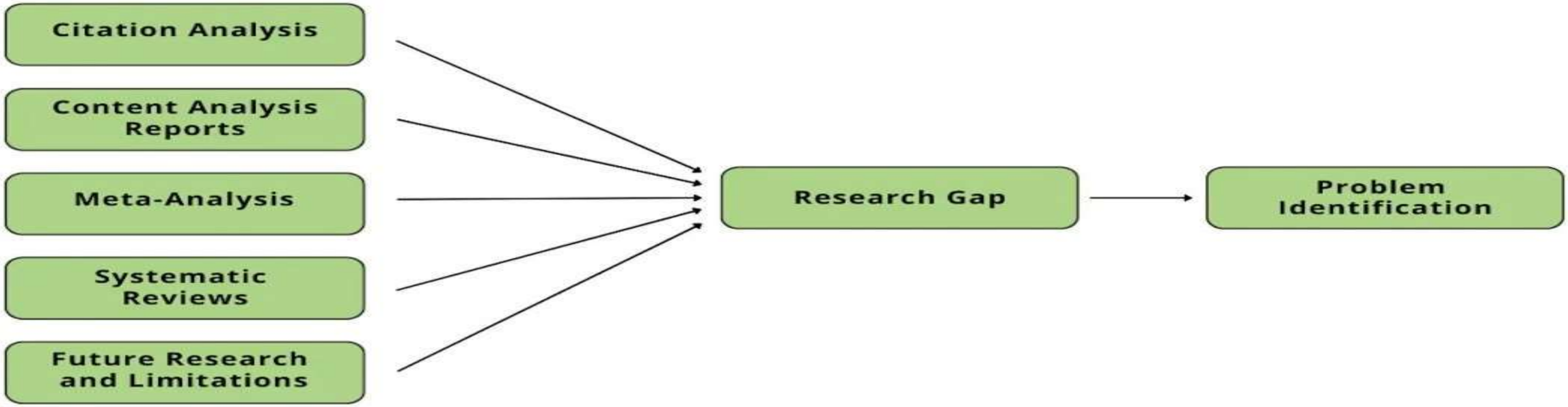


BLUEPRINT OF THE STUDY

Research Design: The architecturalblueprintdictatinghowdataiscollected, measured, and analyzed to answer research questions effectively.

Internal Validity: Ensures that observed outcomes are truly caused by the independent variable, free from confounding influences.

External Validity: The degree to which findings can be generalized beyond the study to real-world contexts and broader populations.



QUANTITATIVE RESEARCH

Numerical Measurement: Relieson precise numerical dataandmathematicalmodeling to quantify phenomena. Variables are measured using standardized instruments to produce objective, comparable results across studies.

Statistical Analysis: Applies inferential and descriptive statistics to test hypotheses and identify significant patterns. Results are evaluated using p-values, confidence intervals, and effect sizes to ensure rigor.

Standardized Instruments: Uses validated, pre-tested tools such as surveys, sensors, and lab equipment to ensure consistency and minimize measurement error across all data collection sessions.

Generalizability: Deductive approach moves from general theory to specific empirical tests. Representative sample sizes and controlled settings allow findings to be generalized to broader target populations.

QUALITATIVE RESEARCH

In-Depth Interviews: One-on-oneconversations exploringparticipants' perspectives, experiences, and meanings. Can be structured, semi-structured, or unstructured to elicit rich, detailed narrative responses.

Focus Groups: Facilitated group discussions among selected participants to gather diverse viewpoints. Interaction between participants generates insights not obtainable through individual interviews alone.

Thematic Analysis: Systematic process of coding transcripts, identifying patterns, and grouping codes into overarching themes. Transforms unstructured text into structured, theory- grounded interpretations.

Contextual Insights: Qualitative inquiry captures deep social, cultural, and behavioral context that quantitative data cannot reveal. Findings are interpreted within the natural setting of the participants studied.

MIXED METHODS RESEARCH

Convergent Design:Collects quantitativeandqualitativedatasimultaneously, then merges results to compare and validate findings from both strands.

Explanatory Sequential: Begins with quantitative data collection and analysis, then uses qualitative data to explain or elaborate on the initial numerical results.

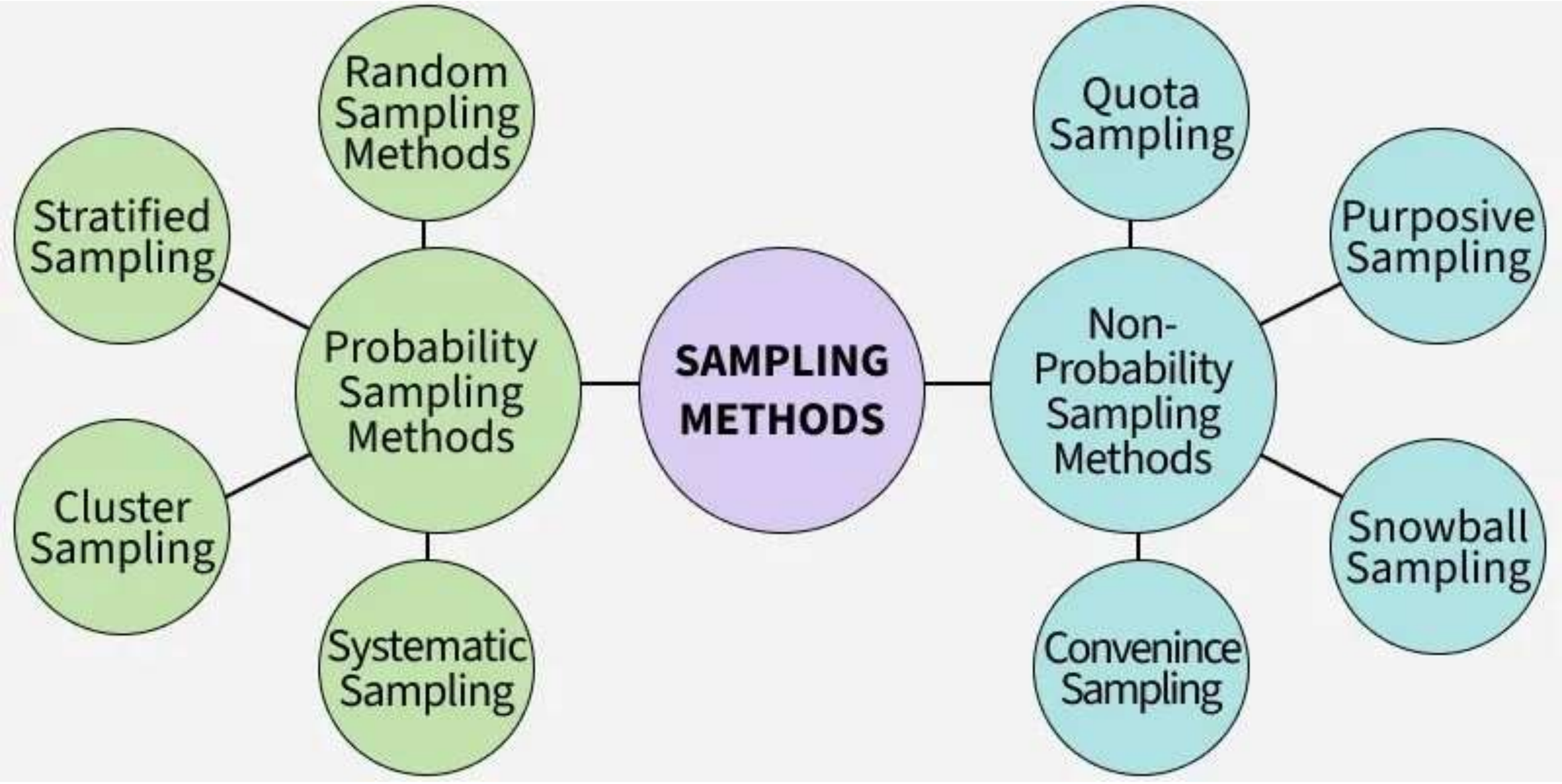
Exploratory Sequential: Starts with qualitative exploration to identify themes, then builds a quantitative instrument to test and generalize the findings.

CLASSIFYING EMPIRICAL APPROACHES

Experimental Studies: Researcheractively manipulates an independentvariableto measure its causal impact on a dependent variable under controlled conditions

Observational Studies: Phenomena are recorded as they occur naturally — no manipulation of variables; ideal when intervention is ethically or practically impossible

Quasi-Experimental: Combines elements of both: some control over variables without full randomization; used when true experiments are unfeasible Experimental: High internal validity Observational: High external validity


SAMPLING METHODS
Probability Sampling Methods
Random Sampling Methods
Stratified Sampling
Cluster Sampling
Systematic Sampling
Non-Probability Sampling Methods
Quota Sampling
Purposive Sampling
Snowball Sampling
Convenince Sampling

SAMPLING TECHNIQUES

Sampling Method	Description	Example
Random Sampling	• Gathering a representative sample from a population where each member in the population has an **equal chance** of being selected.	• Using a **random number generator** to select students in a class to complete a task.
Stratified Sampling	• Smaller groups or **strata** within the sample are represented **proportionally** to the population.	• Finding out a favourite soap opera from different age **categories** of people in a year group.
Systematic Sampling	• Every member in the population is given a **number**. After the first member is chosen at random, the remaining members are chosen from a given **interval**.	• A **list of people** with their first names in alphabetical order are numbered. The 5th person is **chosen randomly**, followed by every subsequent 8th person.
Non Random Sampling	• **Convenience** sampling is used for **ease** of data collection. **Volunteers** usually collect data.	• Asking people at a given location about how long their commute to work is.
Capture Recapture	• **Collecting** a sample of data from **one location** at **different points in time, marking** the individuals to estimate a population size.	• A sample of woodlice were **captured, marked and released**. Another sample of woodlice was captured 5 days later and the number of marked woodlice was counted.

VARIABLES IN RESEARCH

Independent Variable: The factor directly manipulated or categorized by the researcher to observe its effect

Dependent Variable: The outcome measured to observe the effect of the independent variable on subjects

Confounding Variable: Uncontrolled factors that can mask or distort the true relationship between variables

ENSURING DATA INTEGRITY

Reliability: Consistencyofresultsacrossrepeatedmeasurementsunderidentical conditions and over time

Validity: Accuracy of an instrument in measuring the specific theoretical concept it is designed to assess

Measurement Error: Systematic or random deviations that reduce the precision and accuracy of collected data

DATA COLLECTION METHODS

LabInstruments:Highprecisionandcontrolinphysicalandbiologicalsciences. Requires significant resources and expertise.

Surveys & Questionnaires: Enables large sample coverage in social sciences. Subject to self-reporting bias and response errors.

Observational Protocols: Captures natural contextual data in ethology and ecology. Vulnerable to observer bias and subjectivity.

Pre-Testing & Calibration: All instruments must be pre-tested and calibrated to minimize measurement errors before field or lab execution.

DATA PRE-PROCESSING PIPELINE

RawData: Collectedempirical datafromfield surveys,lab instruments, andsensorlogs.

Screening: Inspect and validate entries. Detect entry errors and missing values. Cleaning:

Remove errors, outliers, and duplicates. Impute missing data.

Transform: Normalize and encode variables. Rescale, standardize, and recode variables.

Analysis Ready: A structured dataset for analysis that meets statistical assumptions for testing.

Data Preprocessing Pipeline



Data Collection

Gathering raw data from various sources

Data Understanding & Exploration

Analyzing data to gain insights

Handling Missing Values

Addressing incomplete data entries

Handling Outliers

Identifying and managing extreme data points

Data Type Conversion

Changing data formats as needed

Encoding Categorical Variables

Converting categorical data into numerical form

Feature Scaling

Normalizing data to a standard range

Feature Engineering

Creating new features from existing ones

Feature Selection/ Dimensionality Reduction

Choosing relevant features and reducing complexity

Train-Test Split & Final Checks

Dividing data for training and testing, ensuring quality

DESCRIPTIVE STATISTICS

Mean: The arithmetic averageofallvaluesinadataset.Calculatedbysummingallobservations and dividing by the total number of datapoints (n).

Median: The middle value when data is ordered from lowest to highest. Resistant to outliers and skewed distributions.
Mode: The most frequently occurring value in a dataset. A dataset may be unimodal, bimodal, or multimodal.

Range: The difference between the maximum and minimum values. Provides a simple measure of total spread across the dataset.

Variance: The average of squared deviations from the mean. Quantifies how far data points are spread from the central value.

Standard Deviation: The square root of variance, expressed in original data units. Indicates typical distance of observations from the mean.

INFERENTIAL STATISTICS & HYPOTHESIS TESTING

Null Hypothesis (H_0): Assumes no significant effect or relationship exists between variables. Rejected when sample evidence is sufficiently strong.

Population Inference: Inferential statistics use probability theory to generalize findings from a sample to the broader target population.

p-value Threshold: If $p \leq \alpha$ (0.05), reject H_0. Observed results are unlikely due to chance alone, supporting the alternative hypothesis.

Confidence Intervals: A 95% CI provides a plausible range for the true population parameter, reflecting estimation precision and sampling variability.

QUALITATIVE DATA ANALYSIS

RawTextTranscription: Verbatim transcription ofinterviews, audio,and field notes

Generating Initial Codes: Labeling meaningful segments with descriptive short codes
Grouping into Categories: Clustering related codes into broader conceptual categories
Developing Overarching Themes: Identifying recurring patterns that form key thematic insights

Theoretical Interpretation: Grounding themes in theory through auditing and peer debr iefing

SCIENTIFIC PAPER STRUCTURE

Introduction: Defines theresearchgap, theoretical background, and coreresearch questions. Establishes why the study is necessary and relevant to the field.

Methods: Describes procedures, sample selection, instruments, and analysis tools in sufficient detail to allow full replication by independent researchers.

Results: Presents statistical and empirical findings objectively and without interpretation. Data is reported using tables, figures, and descriptive statistics.

Discussion: Interprets findings in relation to existing literature, addresses study limitations, and outlines implications and directions for future research.

CITATION STANDARDS & ATTRIBUTION

CitationStyles:Standardized formats(APA,IEEE,Vancouver)ensureconsistent attribution of sources across all academic disciplines.

Bibliographic Ethics: Proper attribution credits prior researchers and upholds intellectual honesty in all publications.

Reference Managers: Tools like Zotero, Mendeley, and EndNote automate in-text citations and bibliography generation.

ETHICAL PRINCIPLES IN RESEARCH

Informed Consent:Participantsmustvoluntarilyagreetoparticipatewithfullknowledge of all risks and procedures involved.

Confidentiality: Individual participant data must be protected, anonymized, and kept strictly confidential throughout the study.

Minimize Harm: Research designs must prioritize the safety and well-being of all human and animal subjects at all times.

SCIENTIFIC MISCONDUCT

Plagiarism:Presentinganotherauthor'swork,ideas, or words as one's own without proper citation or attribution

Fabrication: Inventing or fabricating data, results, or findings and recording them as real empirical evidence

Falsification: Manipulating research materials, equipment, or data to distort or misrepresent true findings



Research Misconduct Process

RIO receives allegation

Assessment

- RIO reviews the concern to see if it meets fabrication, falsification, and/or plagiarism (FFP) and gathers additional information to see if a scientific review (Inquiry) is needed
- Non-FFP issues are referred to the appropriate unit/college



Data Sequestration

- RIO works with IT and college leadership to secure all available data associated with the concern both remotely (e.g., backups from SharePoint) and physically (e.g., making a copy of a hard drive) prior to or at the time the Respondent is notified to preserve the data and protect the Respondent from concerns that they had the opportunity to tamper with evidence



Inquiry

- Subject matter expert performs preliminary scientific review of the concern to identify if there are potential errors or issues and recommend whether there is a need for more in-depth investigation
- Report of review submitted to DO for their determination



Investigation

- Subject matter experts (at least 3) performs final scientific review to identify how the issues arose, who was responsible, the extent of the issues and whether the identified concerns were a result of research misconduct
- Report of review submitted to DO for their determination
- Based on DO's determination, corrective actions are recommended

THE PEER REVIEW PROCESS

1. Author Submission: The author submits the manuscript to an academic journal.

2. Editor Screening: The editor checks scope and quality fit.

3. Peer Review Evaluation: Expert reviewers assess rigor and validity.

4. Revision & Decision: Authors revise based on reviewer feedback.

5. Final Publication: The accepted manuscript is published in the journal.

BROADENING THE IMPACT OF SCIENCE

OpenAccess: Publishing researchfindingsin freelyaccessiblejournals and repositories,removing paywalls so that scientists, practitioners, and the public worldwide can read and build upon new knowledge without financial barriers.

Open Data: Making raw datasets, codebooks, and supporting materials publicly available in standardized repositories, enabling independent verification, replication, and secondary analysis by the broader scientific community.

Open Source Code: Sharing analysis scripts, algorithms, and computational tools openly so other researchers can inspect, reproduce, and extend computational methodologies, promoting transparency and accelerating scientific progress.

Public Engagement: Communicating scientific findings to non-specialist audiences through outreach, citizen science initiatives, and science communication platforms to foster public trust and societal benefit from research.

PRACTICAL ASSIGNMENT

Scenario:Doesadigitallearning platform improve biologyperformance?

Null & Alternative Hypothesis: Formulate H_0 (no effect) and H_1 (significant effect) for the digital learning platform study.

Variable Identification: Identify the independent variable, dependent variable, and two possible confounding variables.

Research Design: Select either experimental or observational design for the digital platform versus biology performance study.

Justify Your Design: Explain why experimental or observational design is most appropriate in 2–3 sentences.

REFERENCES & READINGS

Creswell,J.W., &Creswell, J. D.(2018).ResearchDesign:Qualitative, Quantitative, and Mixed Methods Approaches. SAGE Publications.
Bryman, A. (2016): Social Research Methods (5th ed.). Oxford University Press.

Field, A. (2018): Discovering Statistics Using IBM SPSS Statistics (5th ed.). SAGE Publications.

American Psychological Association (2020). Publication Manual of the APA (7th ed.). APA Publishing.

Kothari, C. R. (2004): Research Methodology: Methods and Techniques (2nd ed.). New Age International.