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

Lecture 5

METHODS OF DATA COLLECTION AND DATA ANALYSIS IN RESEARCH

PhD, Acting
Associate Professor,
Doktyrbay Gulina

Aim of the Lecture

To introduce students to the fundamental methods of data collection and data analysis in scientific research, explain their principles and applications, and develop practical skills in selecting appropriate research methods, collecting reliable data, and interpreting research findings.

1. UNDERSTAND fundamental methods of scientific data collection and analysis.

2. IDENTIFY appropriate quantitative and qualitative research methods.

3. APPLY data collection techniques and analytical tools in biological research.

4. INTERPRET research data and evaluate the reliability of scientific findings.

Upon completion of the lecture, students will be able to:

1. Explain the main methods of data collection used in scientific research.
2. Differentiate between qualitative and quantitative data collection approaches.
3. Identify appropriate sampling strategies and measurement techniques.
4. Describe the basic methods of statistical data analysis and interpretation.
5. Evaluate data reliability, validity, and ethical considerations in research.
6. Apply appropriate data collection and analysis methods to biological and biotechnological research.

AGENDA & STRUCTURE

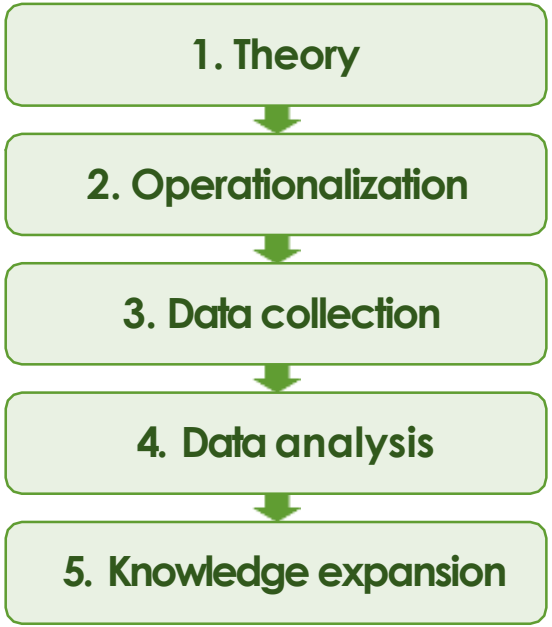
Thispresentation exploresthetwopillarsofempiricalresearch: collection andanalysis.

- Foundations of Research Data (Slides 3–5)
- Quantitative Data Collection Methods (Slides 6–10)
- Qualitative Data Collection Methods (Slides 11–14)
- Data Quality, Measurement & Pre-processing (Slides 15–18)
- Quantitative Data Analysis (Slides 19–23)
- Qualitative & Mixed-Methods Analysis (Slides 24–28)
- Workshop Task & Literature (Slides 29–30)



EMPIRICAL EVIDENCE AS THE FOUNDATION

Data bridges theory and empirical reality. Systematic, recorded observations and measurements validate hypotheses and drive the expansion of scientific knowledge.



CATEGORIZING BIOLOGICAL AND SOCIAL DATA

Data category	Meaning	Example
Primary	Collected firsthand	Laboratory sensors, interviews
Secondary	Pre-existing records	Census, archives
Quantitative	Numerical measurements	Counts and measurements
Qualitative	Descriptive, contextual observations	Descriptions and observations

ALIGNMENT IN RESEARCH DESIGN

Collection methods must match planned analytical techniques. Operationalization and measurement level directly influence data gathering choices —, misalignment undermines validity and reproducibility.



EXPERIMENTAL MEASUREMENT

Experimental Conditions: Independent variables a manipulated under carefully controlled laboratory conditions.Measurement Instruments: Spectrophotometers, biosensors, and analytical devices are used to obtain precise measurements.Data Recording: Data loggers record experimental measurements and monitor changes over time.Accuracy and Reliability: Precis measurement techniques minimize experimental errors and improve the reliability of research results.



SURVEYS & QUESTIONNAIRES

Surveys and questionnaires are widely used research methods for collecting structured data from individuals or populations.

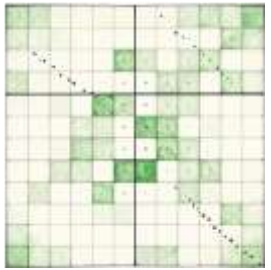
- Closed-ended questions provide quantitative data.
- Likert scales measure attitudes, opinions, and perceptions.
- Validated questionnaires improve data reliability and consistency.
- Standardized formats facilitate statistical analysis and comparison across large samples.



OBSERVATIONAL SAMPLING

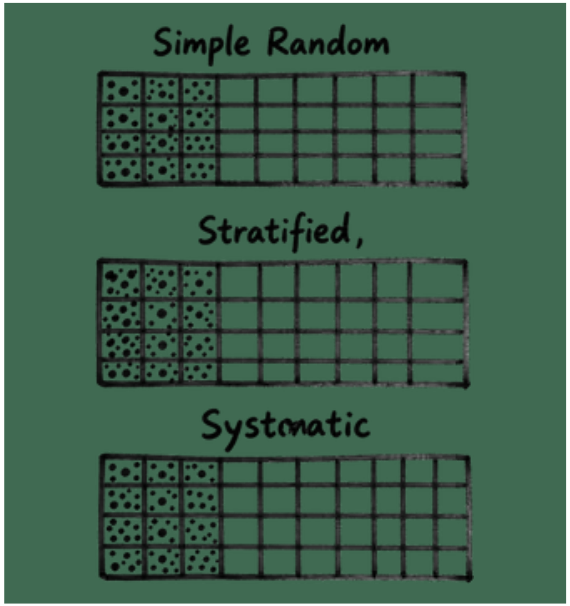
Point counts, line transects, and focal sampling quantify species occurrences and behaviors in natural settings without manipulation. Structured protocols reduce observer bias and ensure reproducible, comparable datasets across sites and time periods.





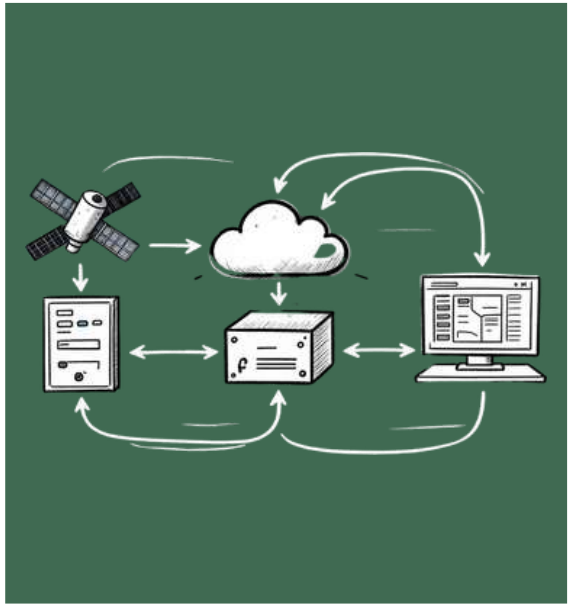
SAMPLING STRATEGIES IN QUANTITATIVE COLLECTION

Probability sampling (simple random, stratified, systematic) ensures representativeness and reduces bias. Non-probability methods (convenience, purposive) offer flexibility but limit generalizability and introduce selection bias.



DIGITAL AND AUTOMATED DATA COLLECTION

Digital technologies, including automated sensors, satellite systems, DNA sequencers, and web-scraping tools, enable continuous and large-scale data collection with minimal human error. These technologies transform raw measurements and digital signals into structured datasets, facilitating efficient data storage, processing, and analysis in modern scientific research.



QUALITATIVE COLLECTION: IN- DEPTH INTERVIEWS

Semi-structured or unstructured interviews with open-ended questions reveal meanings and contexts missed by quantitative tools. Researchers probe individual experiences, beliefs, and motivations in depth.



FOCUS GROUP DISCUSSIONS

Small groups of 6–10 participants discuss research topics under structured moderation, capturing consensus, debate, and social interaction insights that individual interviews may miss.





PARTICIPANT OBSERVATION & ETHNOGRAPHY

Researchers embed themselves in subjects' environments, recording detailed field notes on interactions and contexts over time. This immersive naturalistic inquiry reveals meanings invisible to external observation.



DOCUMENT & ARCHIVAL DATA COLLECTION

Leveraging existing records — medical files, historical archives, policy reports, and social media logs — as longitudinal context sources without direct participant contact.



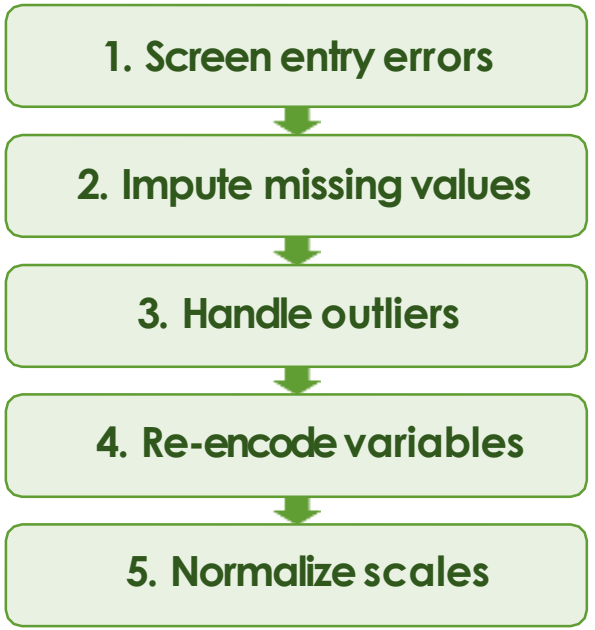
MEASUREMENT VALIDITY AND RELIABILITY

Reliability = consistency of results across repeated trials. Validity = accuracy in measuring the intended concept. Both are essential criteria for trustworthy research instruments.



DATA PRE- PROCESSING AND CLEANING

Raw datasets require systematic cleaning before analysis: screening for entry errors, imputing missing values, handling outliers, re-encoding variables, and normalizing scales to ensure analytical readiness.



LEVELS OF MEASUREMENT AND DATA SCALES

Scale	Properties	Example
Nominal	Categories without an order	Species, blood type
Ordinal	Ranked order; unequal intervals	Likert scale
Interval	Equal intervals; no true zero	Temperature, °C
Ratio	Equal intervals and a true zero	Mass, concentration

ETHICAL STANDARDS IN DATA COLLECTION

Informed consent, anonymization, secure data storage, and Institutional Review Board (IRB) compliance are mandatory safeguards protecting participants and ensuring data integrity throughout the research process.



DESCRIPTIVE STATISTICS

Measures of central tendency (mean, median, mode) summarize the center of a distribution. Measures of dispersion —standard deviation, variance, range, andIQR — quantify spread and variability within the dataset.



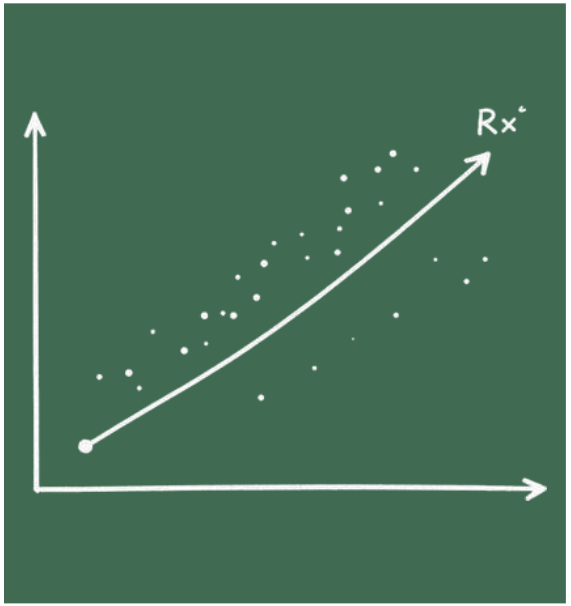
PARAMETRIC VS. NON-PARAMETRIC TESTS

Test family	Data / assumptions	Examples
Parametric	Normal distribution; interval / ratio data	t-test, ANOVA, Pearson r
Non-parametric	Non-normal or ordinal data	Mann–Whitney U, Kruskal–Wallis, Spearman r

Test selection depends on data level, sample size, and distribution shape.

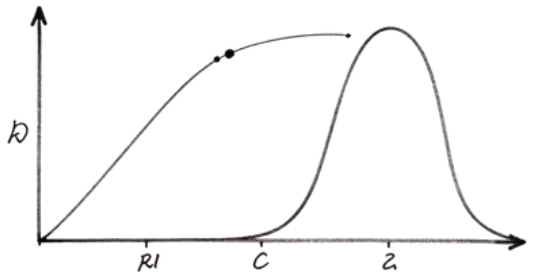
CORRELATION & REGRESSION ANALYSIS

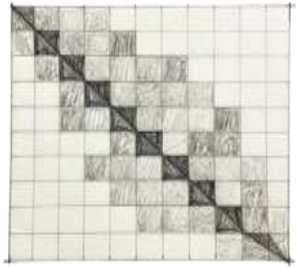
PCA and other machine learning methods reduce dimensionality in genomic and gene expression datasets. Clustering algorithms group biological samples, revealing hidden patterns and relationships across thousands of variables.



EXPLORATORY DATA ANALYSIS

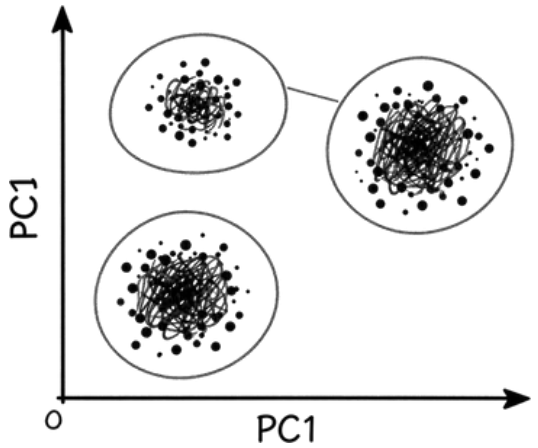
Box plots, heatmaps, density curves, and pair plots reveal data structure, outliers, and correlations before formal modeling — guiding selection of appropriate statistical tests.





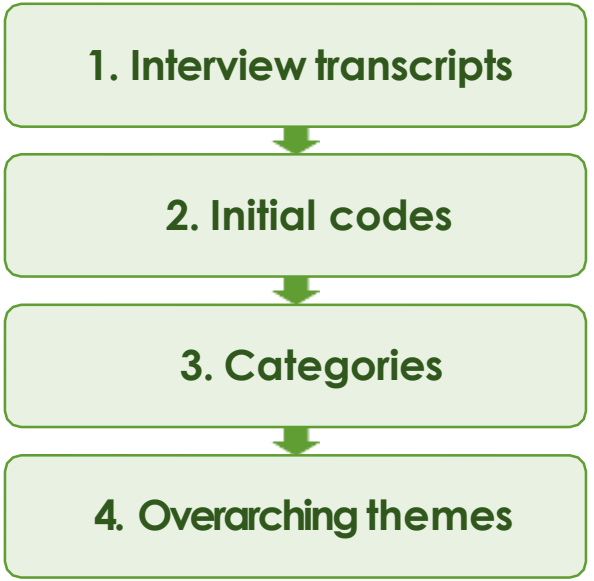
BIOINFORMATICS & HIGH-DIMENSIONAL DATA

PCA and machine learning reduce dimensionality in genomic and expression datasets. Clustering algorithms classify biological sample groups, revealing hidden patterns across thousands of variables.



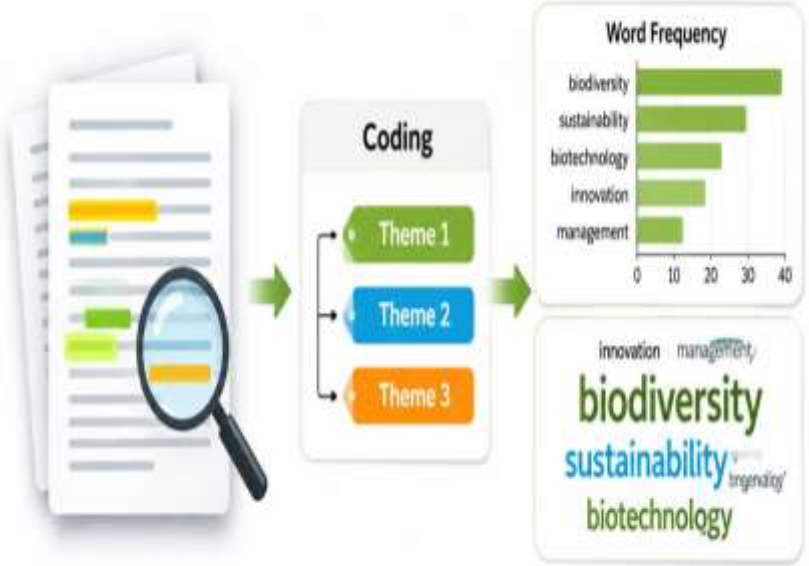
QUALITATIVE ANALYSIS: THEMATIC ANALYSIS

Coding transcripts, grouping codes into categories, and developing overarching themes that capture conceptual patterns across the data. Thematic analysis transforms raw interview text into structured, meaningful insights.



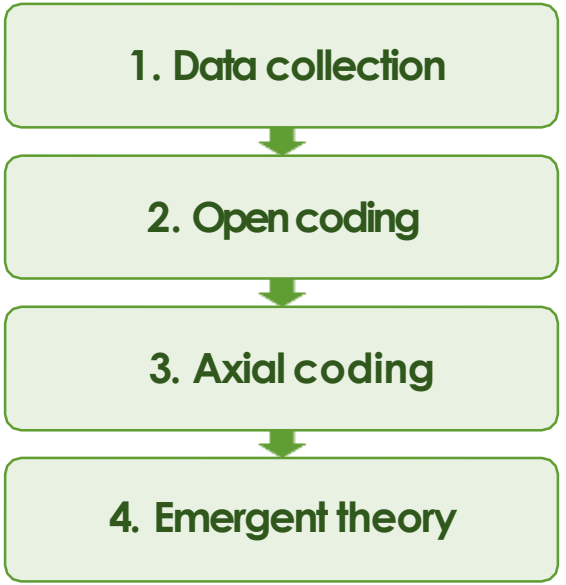
QUALITATIVE ANALYSIS: CONTENT ANALYSIS

Content analysis categorizes words and concepts either qualitatively (thematic coding) or quantitatively (word frequencies, occurrence counts). It transforms unstructured text into structured, interpretable patterns—bridging qualitative depth and quantitative rigor.



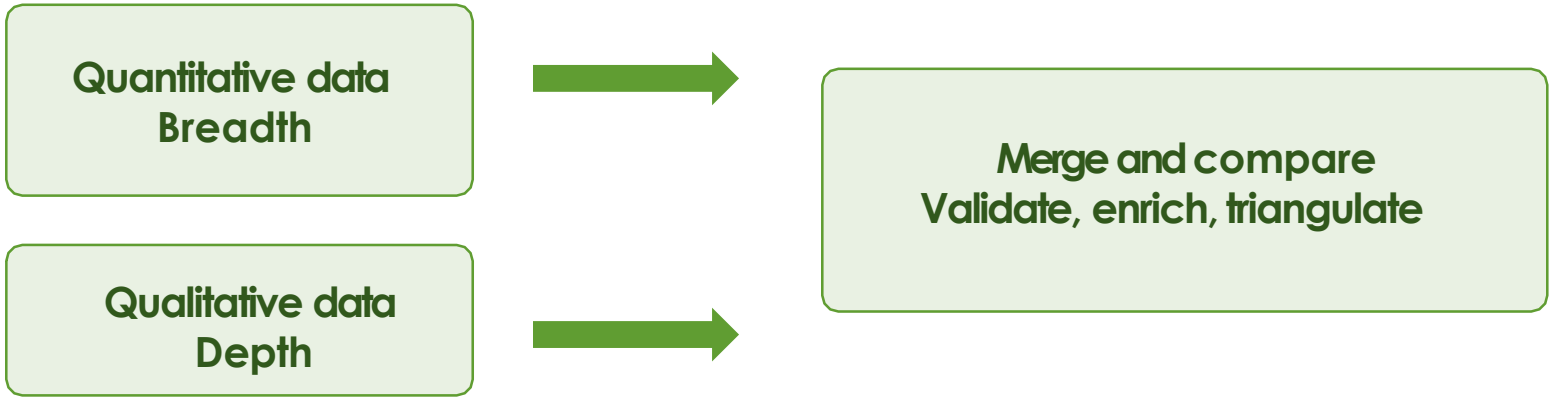
GROUNDED THEORY & NARRATIVE APPROACHES

Grounded theory uses iterative, inductive coding to build emergent theory directly from data. Narrative analysis examines story structure, meaning-making, and lived experience. Both approaches rely on constant comparison:



MIXED-METHODS DATA INTEGRATION

Combining breadth and depth throughconvergentparallel designs— quantitative and qualitative data streams are collected simultaneously, then merged to validate, enrich, and triangulate research insights.



SOFTWARE & DIGITAL TOOLS FOR DATA ANALYSIS

QUANTITATIVE ANALYSIS

R, Python, SPSS, and GraphPad Prism are widely used for statistical modeling, regression analysis, data processing, and visualization.

QUALITATIVE ANALYSIS (CAQDAS)

NVivo, MAXQDA, and ATLAS.ti support systematic data coding, thematic analysis, qualitative interpretation, and mixed-methods integration.



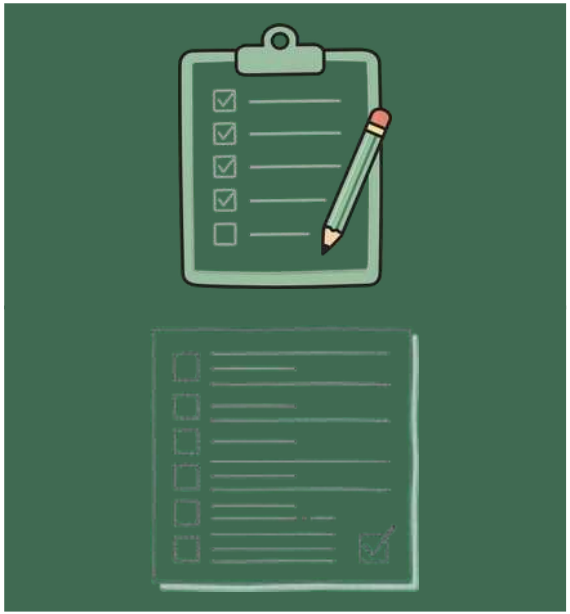
WORKSHOP ASSIGNMENT

Scenario: Evaluatehybrid work schedule effects on stress and productivity among medical lab technicians.

- Propose one quantitative and one qualitative data collection tool.

- Define the dependent variable and its level of measurement.

- Select analysis methods (statistical test + qualitative approach) with brief justification.



REFERENCES & SUGGESTED READING

Creswell & Creswell (2018) – Research Design: Qualitative, Quantitative & Mixed Methods

Field (2018) – Discovering Statistics Using IBM SPSS Statistics

Saunders, Lewis & Thornhill (2019) – Research Methods for Business Students

Braun & Clarke (2006) – Using Thematic Analysis in Psychology