

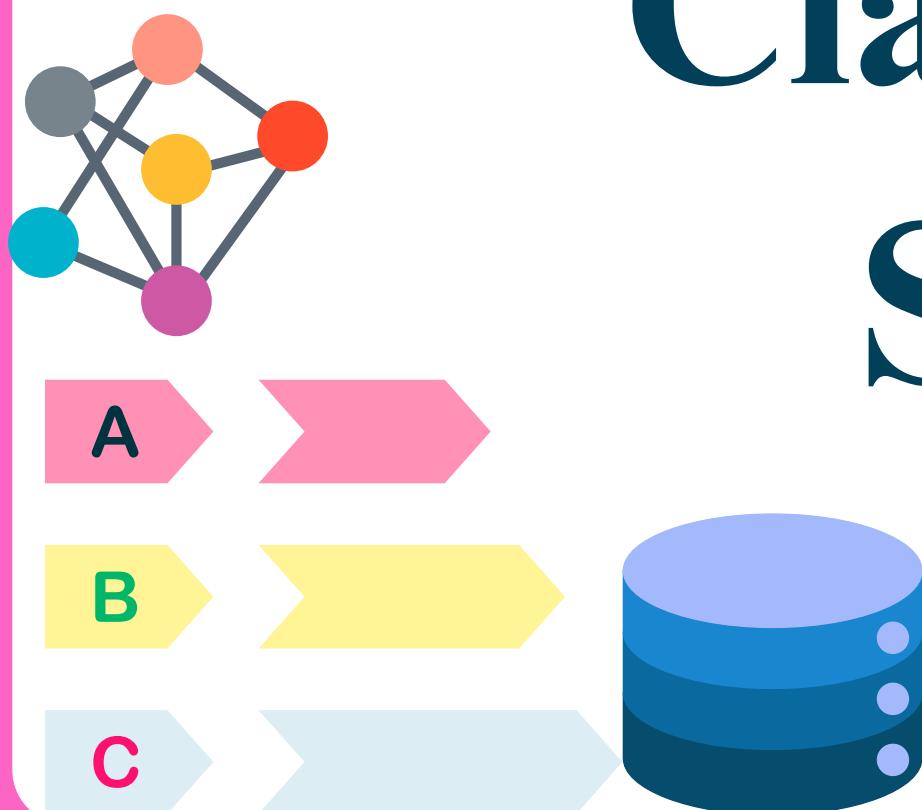
Day 1

Introduction to Analytical Reporting in Education



Session 3

Classification of Data and Sampling Techniques



Lesson Outcome

By the end of this session, participants will gain the knowledge and skills to:

- Understand the significance of data in decision-making.
- Classify data according to its attributes and measurement scales.
- Differentiate between variables, populations, and samples.
- Utilize different sampling techniques effectively for data collection.
- Calculate an appropriate sample size for various research scenarios.
- Develop tailored sampling strategies considering pertinent factors and limitations



Lesson Outline

- Definition and significance of data
- Classification of data based on attributes and scales of measurement.
- Understanding variables and their attributes.
- Differentiating between population and sample.
- Importance of sampling in data collection.
- Overview of various sampling techniques, their strengths, limitations, and real-world applications.
- Importance and methods of determining an appropriate sample size.

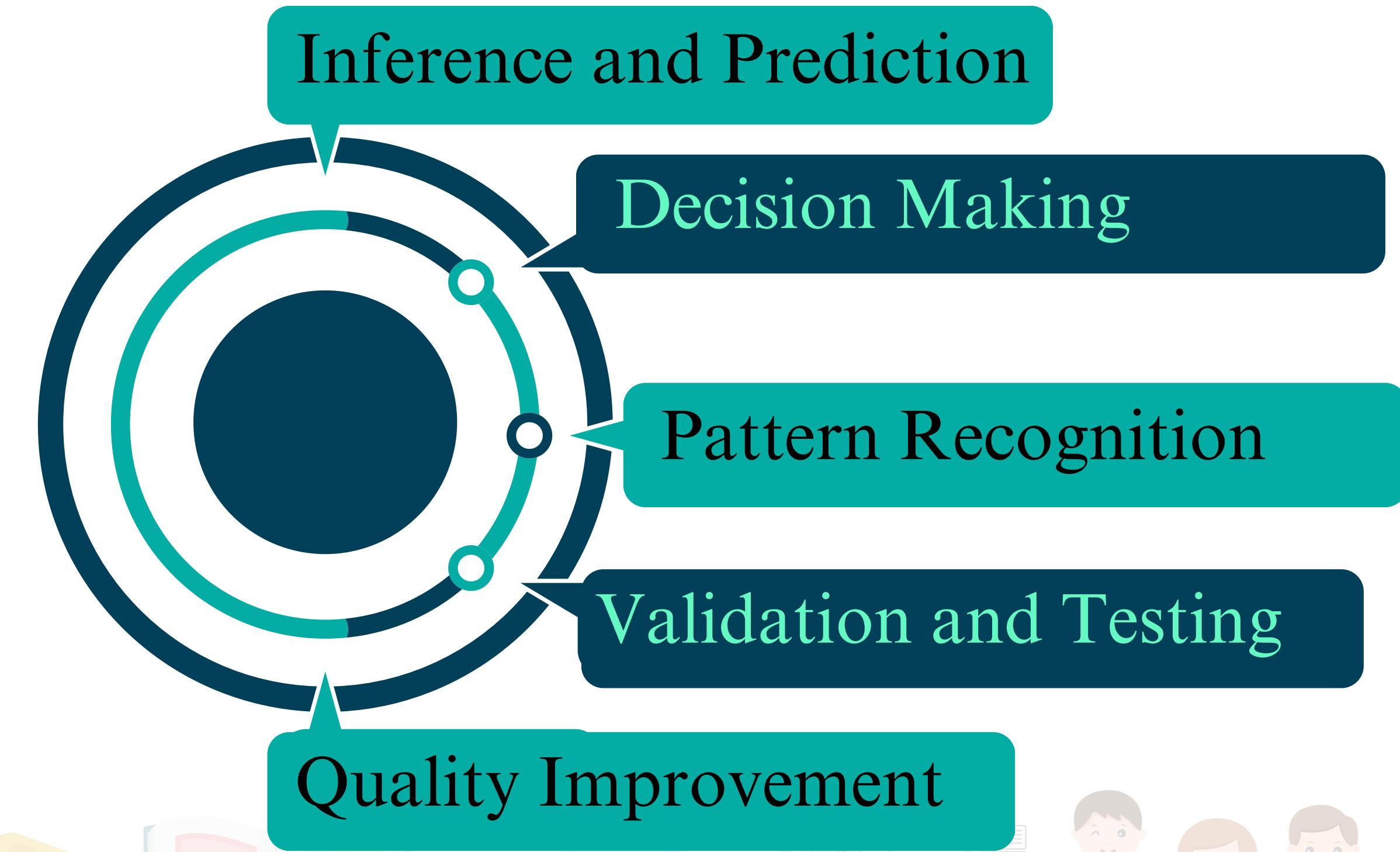


What is Data

Data is raw information that can be processed and analyzed to extract meaningful insights and facilitate decision-making.



Importance of Data



Importance of Data

Inference and Prediction: Data enables statistical inference and prediction by providing the basis for concluding populations from sample data.

Decision Making: Data-driven decision-making relies on statistical analysis to evaluate options, assess risks, and make informed choices based on empirical evidence.

Pattern Recognition: Statistical analysis of data facilitates pattern recognition, allowing for the identification of trends, relationships, and structures within datasets.

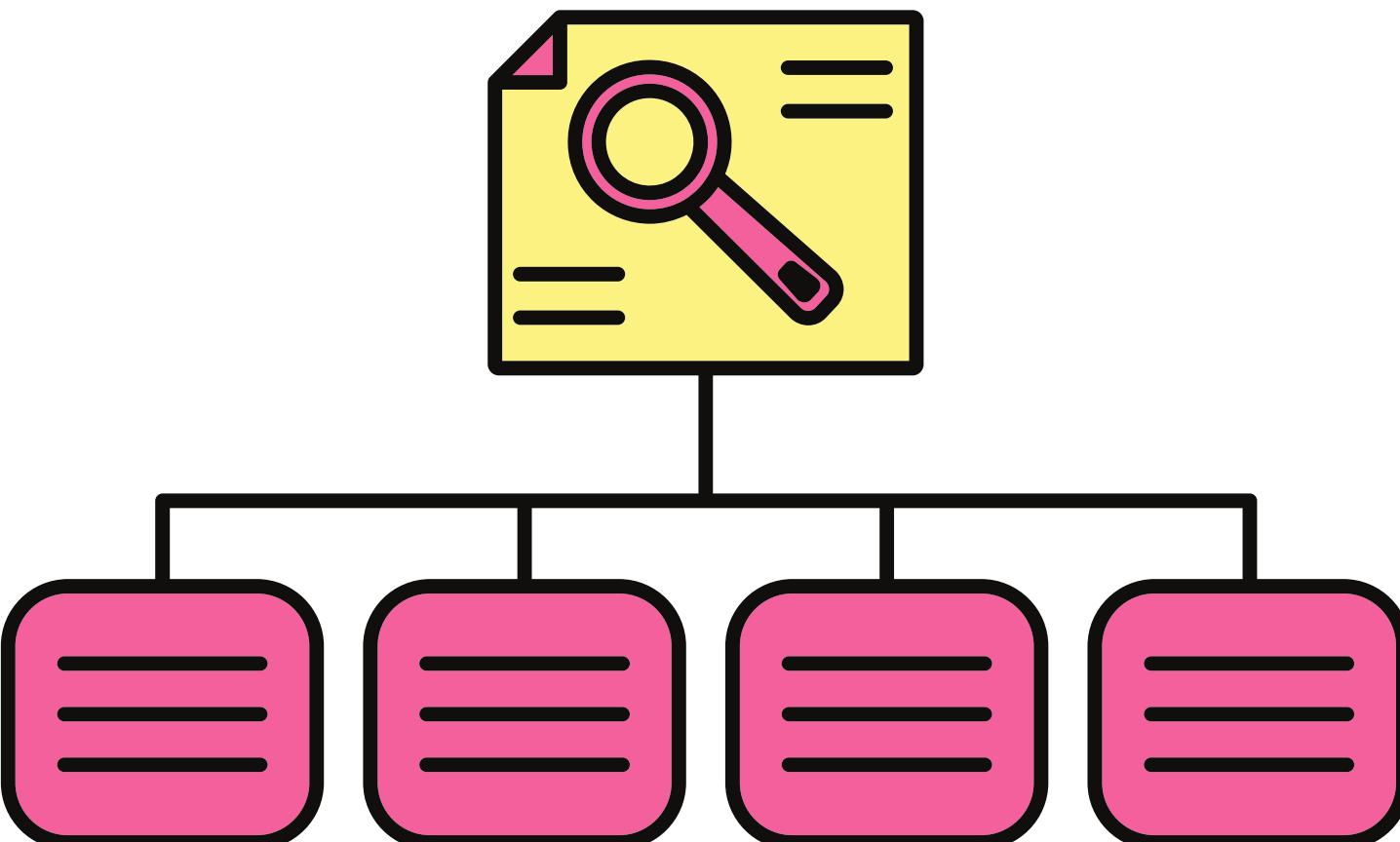
Validation and Testing: Data serves as the foundation for validating statistical models and hypotheses, supporting the assessment of data reliability and the testing of theoretical predictions.

Quality Improvement: Through data analysis, statisticians can identify areas for improvement and optimize processes, driving quality enhancement efforts and fostering innovation.

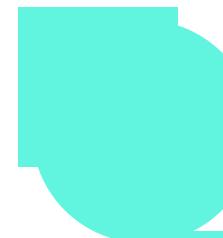
Classification of Data

Statistics is the science of data and data are obtained by measuring the value of one or more variables on the experimental units of the population or sample. Broadly speaking data may come from many ways such as:

- »»» **Based on Nature**
- »»» **Based on the Scale of Measurement**
- »»» **Based on Variability**
- »»» **Based on Source and Availability**
- »»» **Based on the Time Dimension**
- »»» **Based on Use and Purpose**



Based on Nature or variable



Qualitative Data **A B C**

- Also known as categorical data, qualitative data describes qualities or characteristics and cannot be measured numerically.
- **Examples** include gender, ethnicity, colour, and marital status.



Quantitative Data **1 2 3 4**

- Quantitative data consists of numerical values that can be measured and quantified.
- It can further be classified into discrete or continuous data types.



Based on the Scale of Measurement

	Nominal	Ordinal	Interval	Ratio
Categories	✓	✓	✓	✓
Rank order		✓	✓	✓
Equal spacing			✓	✓
True zero				✓

Based on the Scale of Measurement

Nominal Data: Nominal data represents categories or labels that do not have a specific order or ranking.

Examples: Gender (male, female), marital status (single, married, divorced), types of cars (sedan, SUV, truck).

Ordinal Data: Ordinal data represents categories with a natural order or ranking, but the differences between categories are not uniformly quantifiable.

Examples: Educational levels (elementary, middle, high school), Likert scale ratings (strongly agree, agree, neutral, disagree, strongly disagree), rank order (1st, 2nd, 3rd).

Based on the Scale of Measurement

Interval Data: Interval data represents measurements where the differences between values are consistent and quantifiable, but there is no true zero point.

Examples: Temperature measured in Celsius or Fahrenheit, dates (days of the week, months of the year), IQ scores.

Ratio Data: Ratio data represents measurements where there is a true zero point, and the differences between values are consistent and quantifiable.

Examples: Height, weight, time, income, number of children in a family.

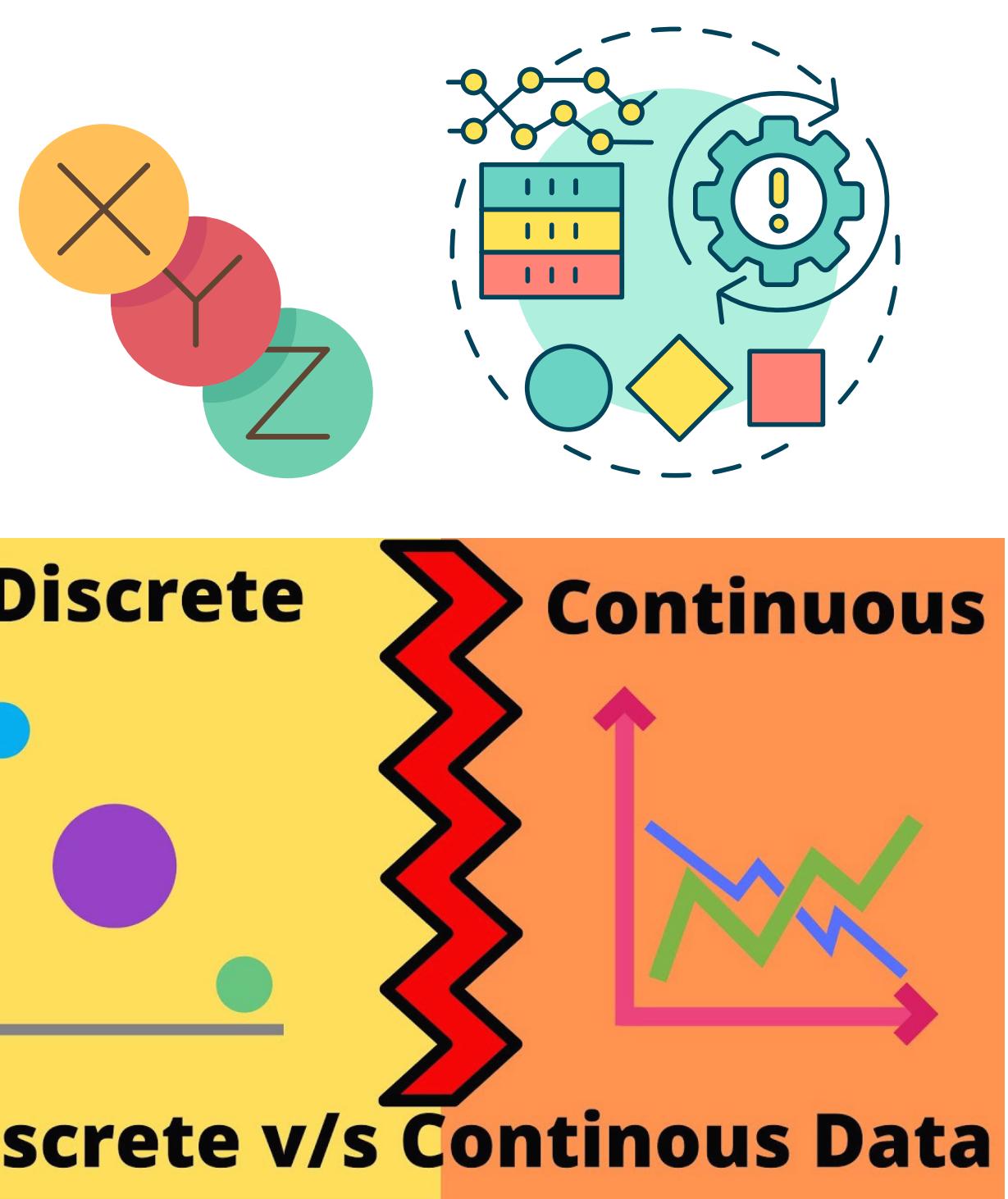
Based on Variability

Continuous Data: Continuous data represents measurements that can take any value within a given range.

Examples include height, weight, temperature, and time.

Discrete Data: Discrete data consists of distinct, separate values with no intermediate values possible.

Examples include the number of students in a class, the number of cars in a parking lot, or the number of goals scored in a soccer match.



Based on Source and Availability



Primary Data

Primary data is collected firsthand by the researcher for a specific research purpose. It includes surveys, interviews, observations, and experiments.

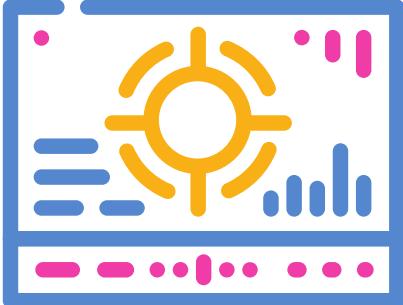


Secondary Data

Secondary data is obtained from existing sources, such as books, articles, databases, or official records. It's data that has been collected by someone else for a different purpose but can be repurposed for new analyses.

Based on Time Dimension

Cross-Sectional Data: Cross-sectional data is collected at a single point in time, providing a snapshot of a population or phenomenon at that moment.



Time-Series Data: Time-series data is collected over a sequence of equally spaced time intervals. It allows for the analysis of trends, patterns, and changes over time.



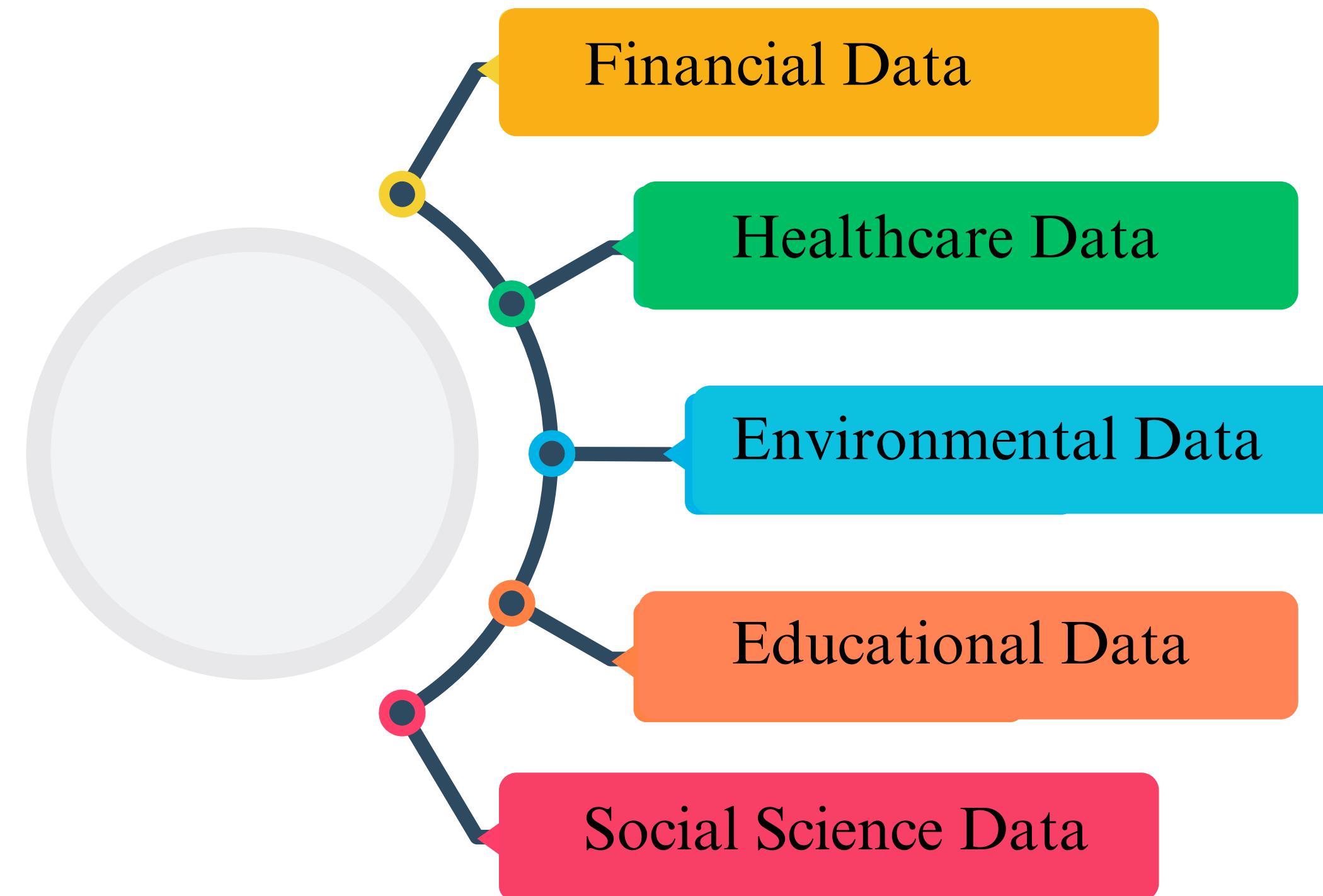
Based on Use and Purpose

Raw Data: Raw data is unprocessed and unanalyzed data collected directly from sources.



Processed Data: Processed data is organized, cleaned, and analyzed data ready for interpretation and decision-making

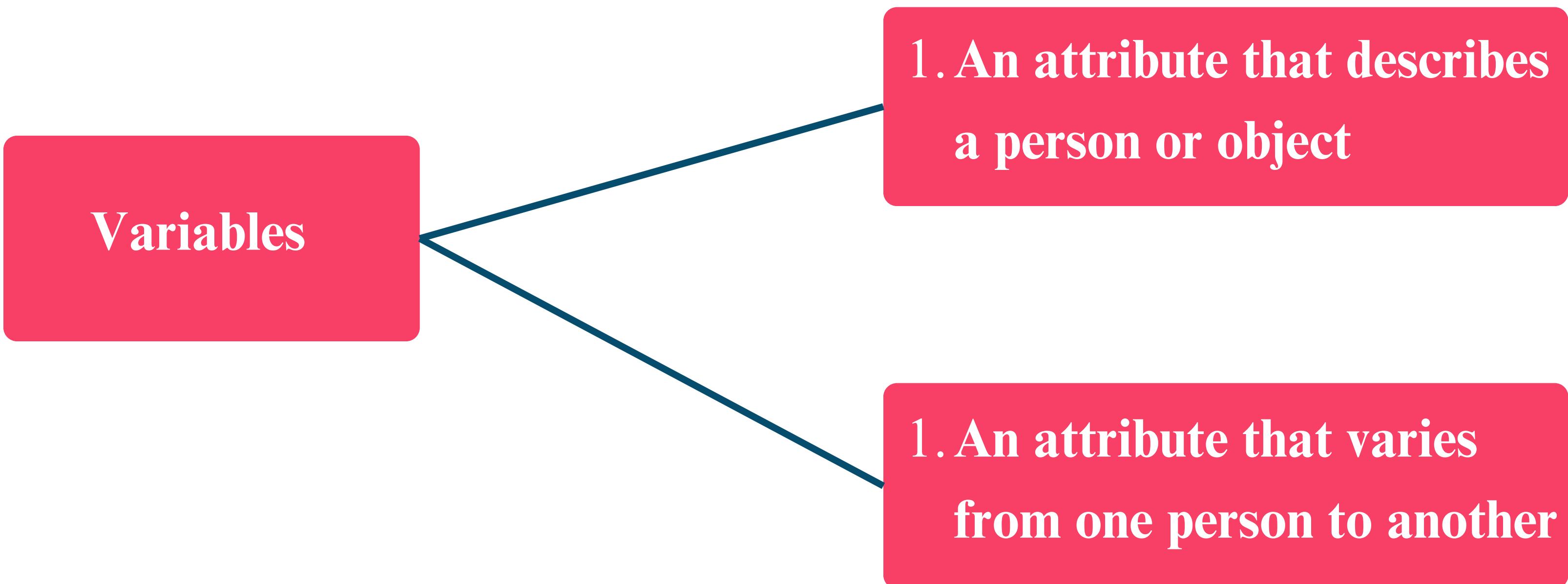
Data according to the Subject



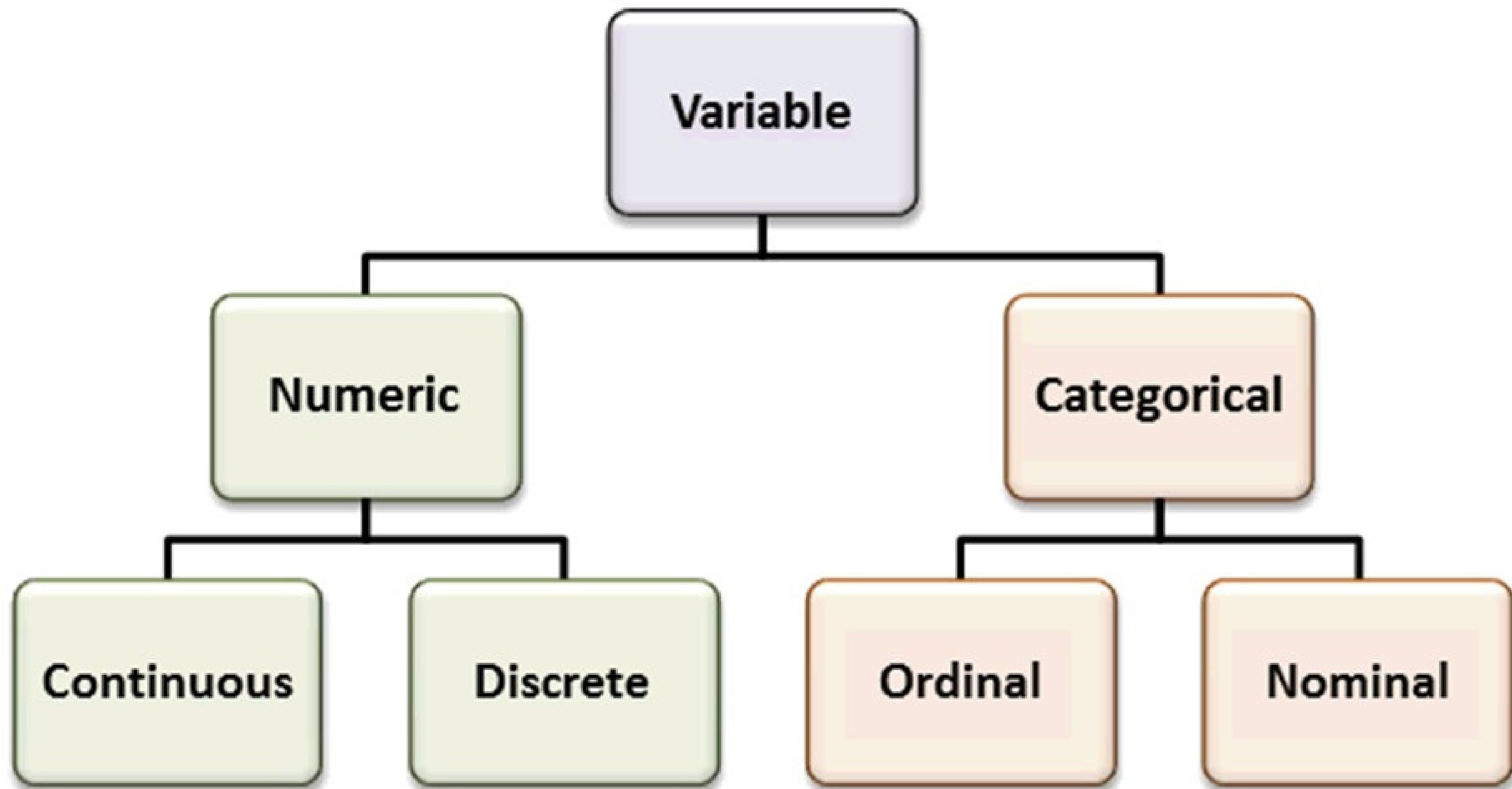
Understanding variables and their attributes

Variables: Variables are characteristics or attributes that can be measured in a research study. Variable mainly varies from person to person or things to things.

Example: Height, Weight, Gender

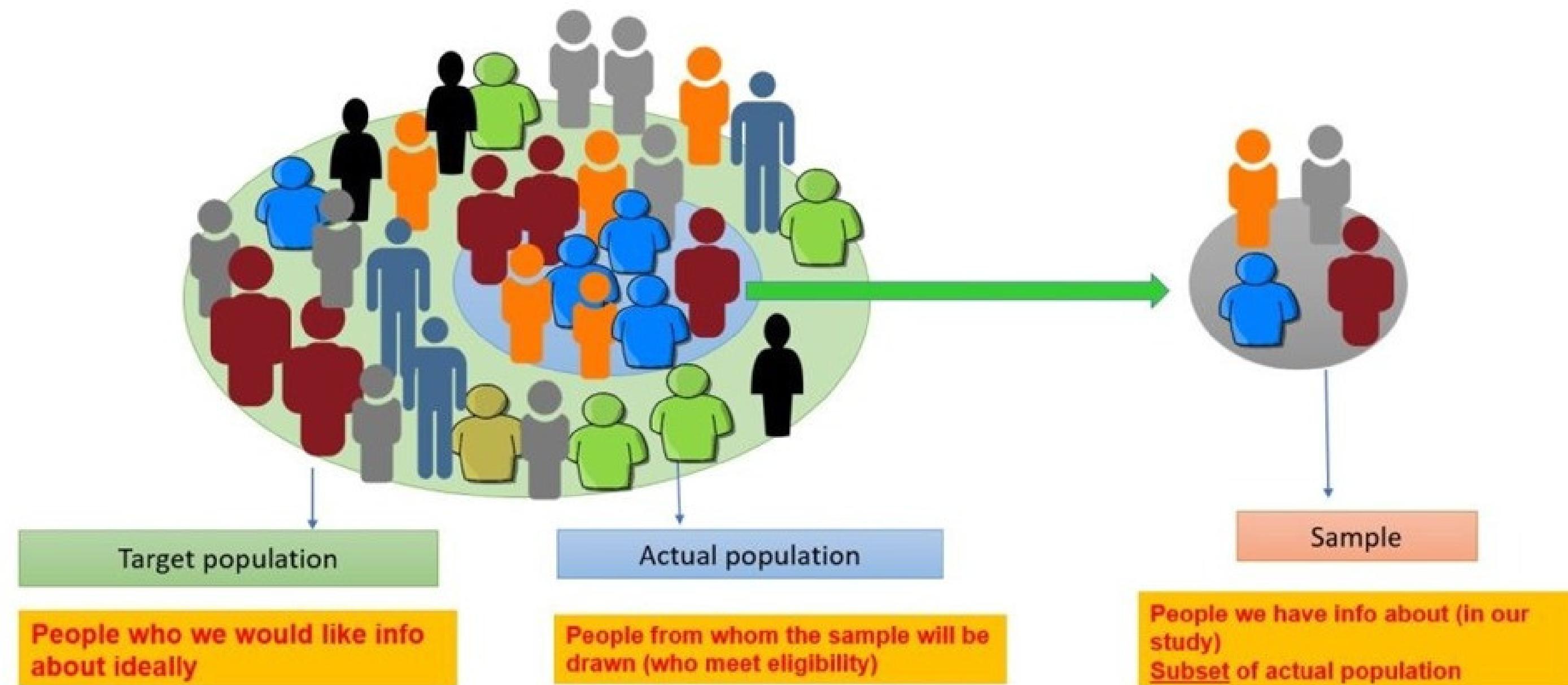


Types of Variables



Population, Sample, Sampling Techniques, Sample Size Determination

What do you mean by population and sample?



Population, Sample, Sampling Techniques, Sample Size Determination

Population: A collection of all elements interested id called population. The population refers to the entire group of individuals or entities that the researcher is interested in studying.

For example, if the study focuses on primary school students in Bangladesh, the population would be all primary school students in the country.

Sample: A representative part of the population is known as a sample. The sample should be representative of the population to ensure that study findings can be generalized.

Example: To assess reading proficiency of third-grade students in a school district by selecting a sample of 100 randomly selected students from different schools, allowing for generalization of results to all third-grade students in the district.



Population, Sample, Sampling Techniques, Sample Size Determination

Census: A survey that includes every number of populations is called a Census. A Census is defined as a complete enumeration of the elements of a population or study objects.

Example: The sixth population and Housing Census-2022.

Sample Survey: A survey that includes every part of the sample is known as a sample survey.

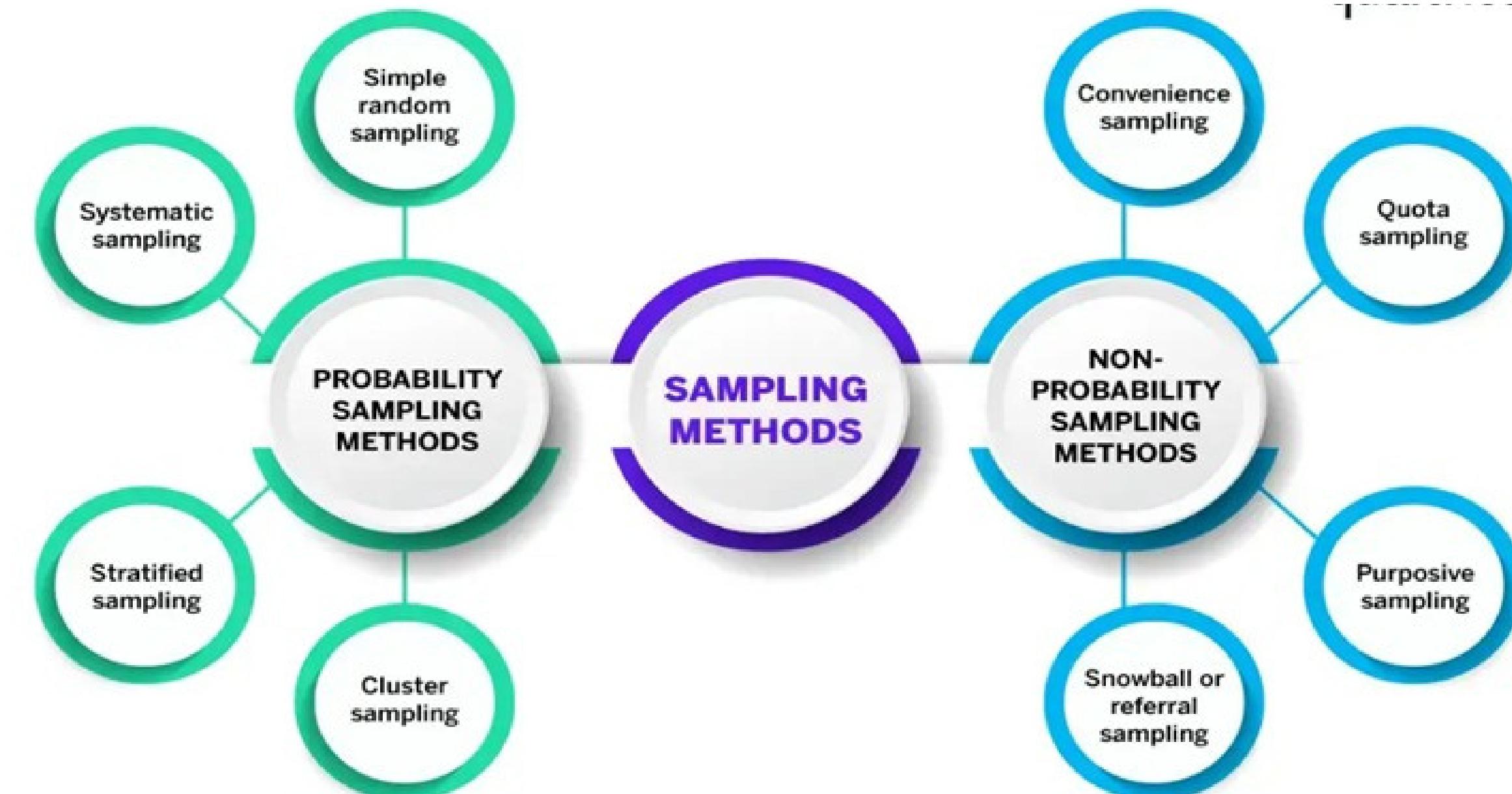
Example: Bangladesh Demographic Health Survey (BDHS), Household Income and Expenditure Survey (HIES).



Sampling Techniques

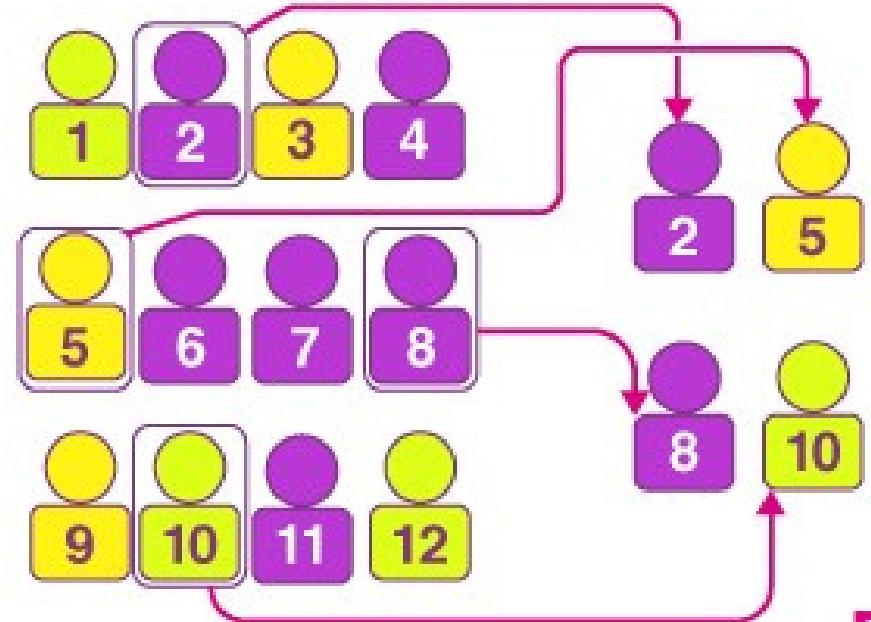
Researchers will select appropriate sampling techniques to ensure the sample is representative and unbiased.

Classification of Sampling Techniques

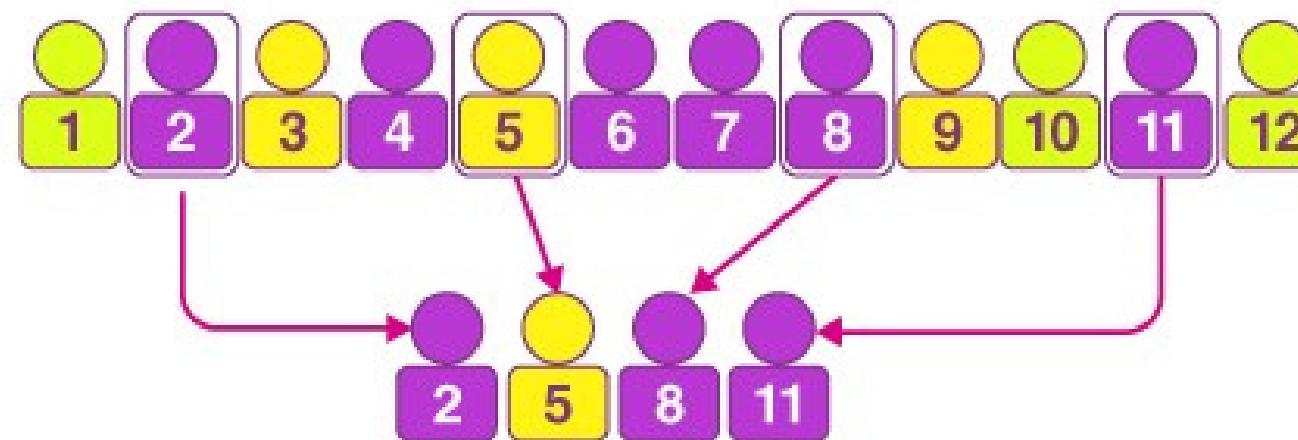


Probability Sampling Methods

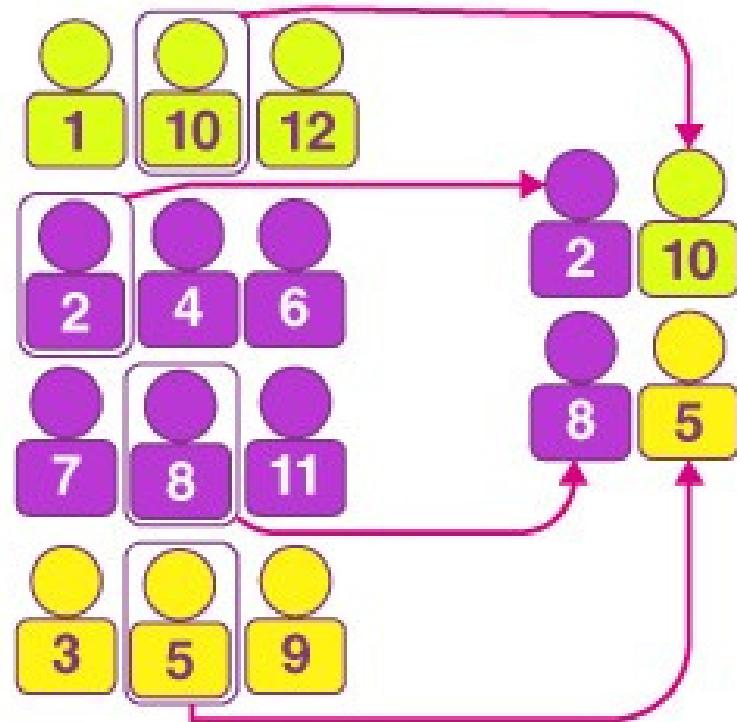
Simple Random Sampling



Systematic Sampling

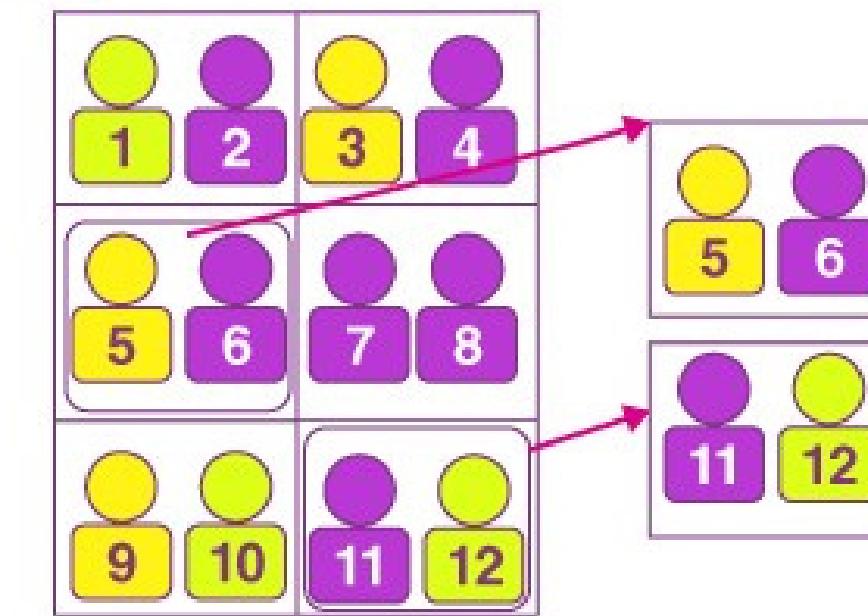


Stratified Sampling



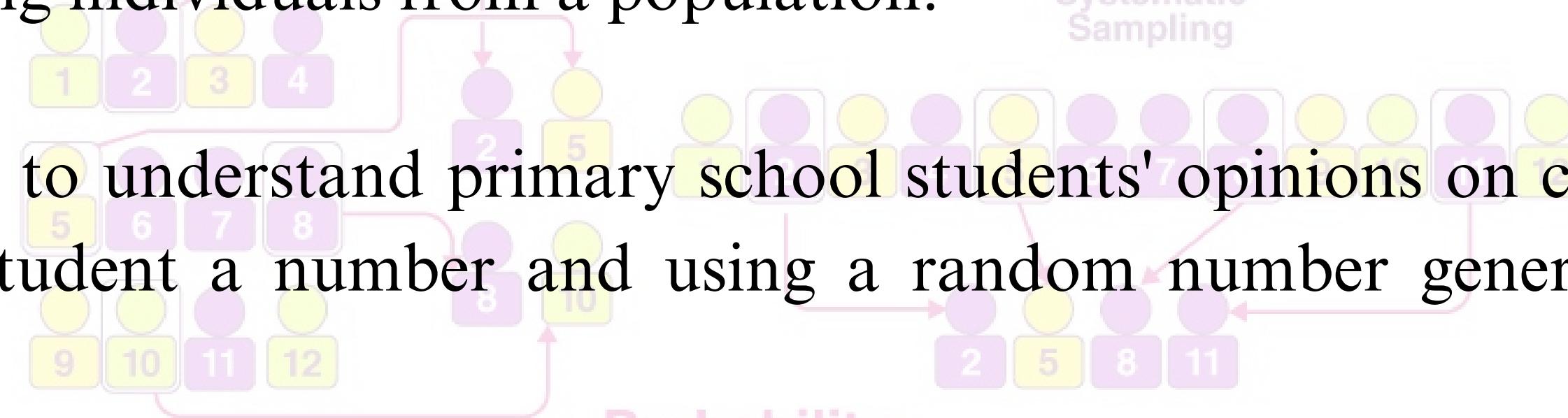
Probability sampling Methods

Clustered Sampling



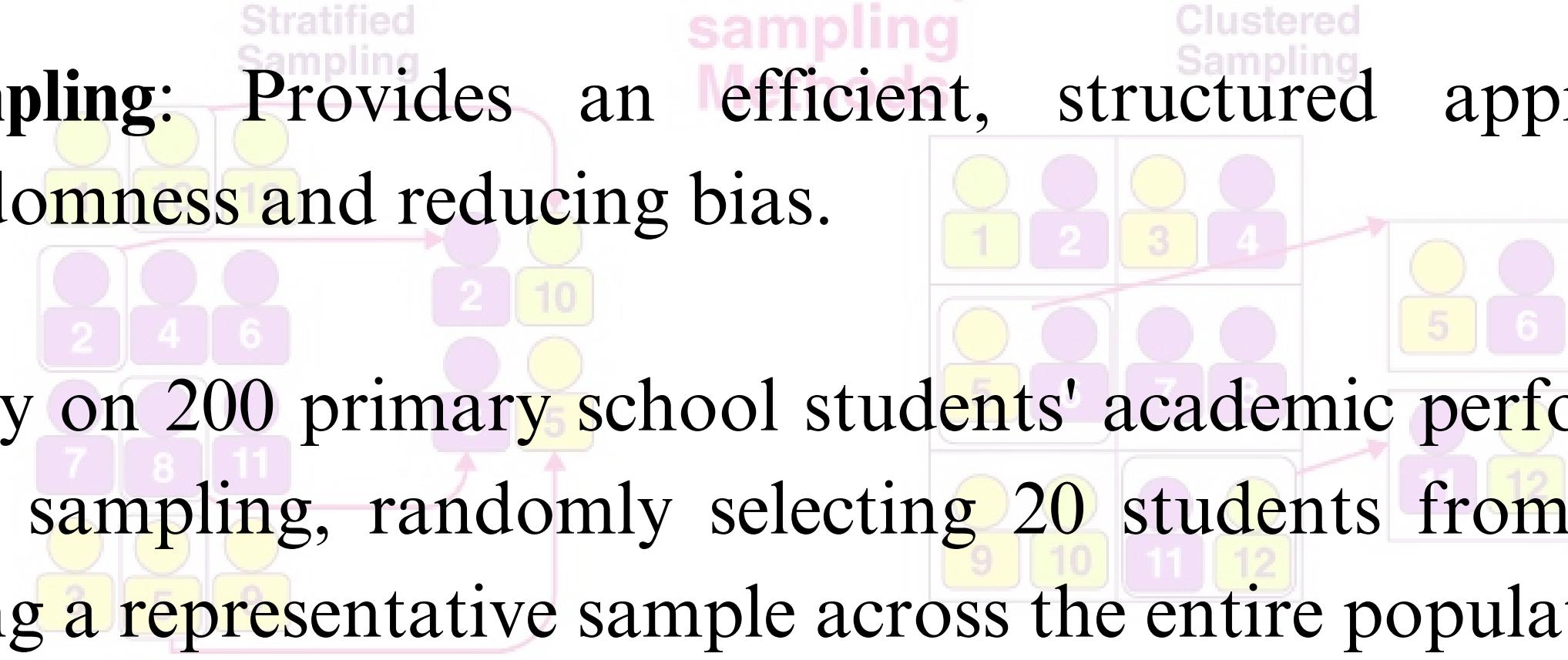
Probability Sampling Methods

Simple Random Sampling: Enables unbiased representation and statistical validity by randomly selecting individuals from a population.



Example: Survey to understand primary school students' opinions on cafeteria food by assigning each student a number and using a random number generator to select a sample.

Systematic Sampling: Provides an efficient, structured approach to sampling, maintaining randomness and reducing bias.



Example: A study on 200 primary school students' academic performance is conducted using systematic sampling, randomly selecting 20 students from a list sorted by ID numbers, ensuring a representative sample across the entire population.

Probability Sampling Methods

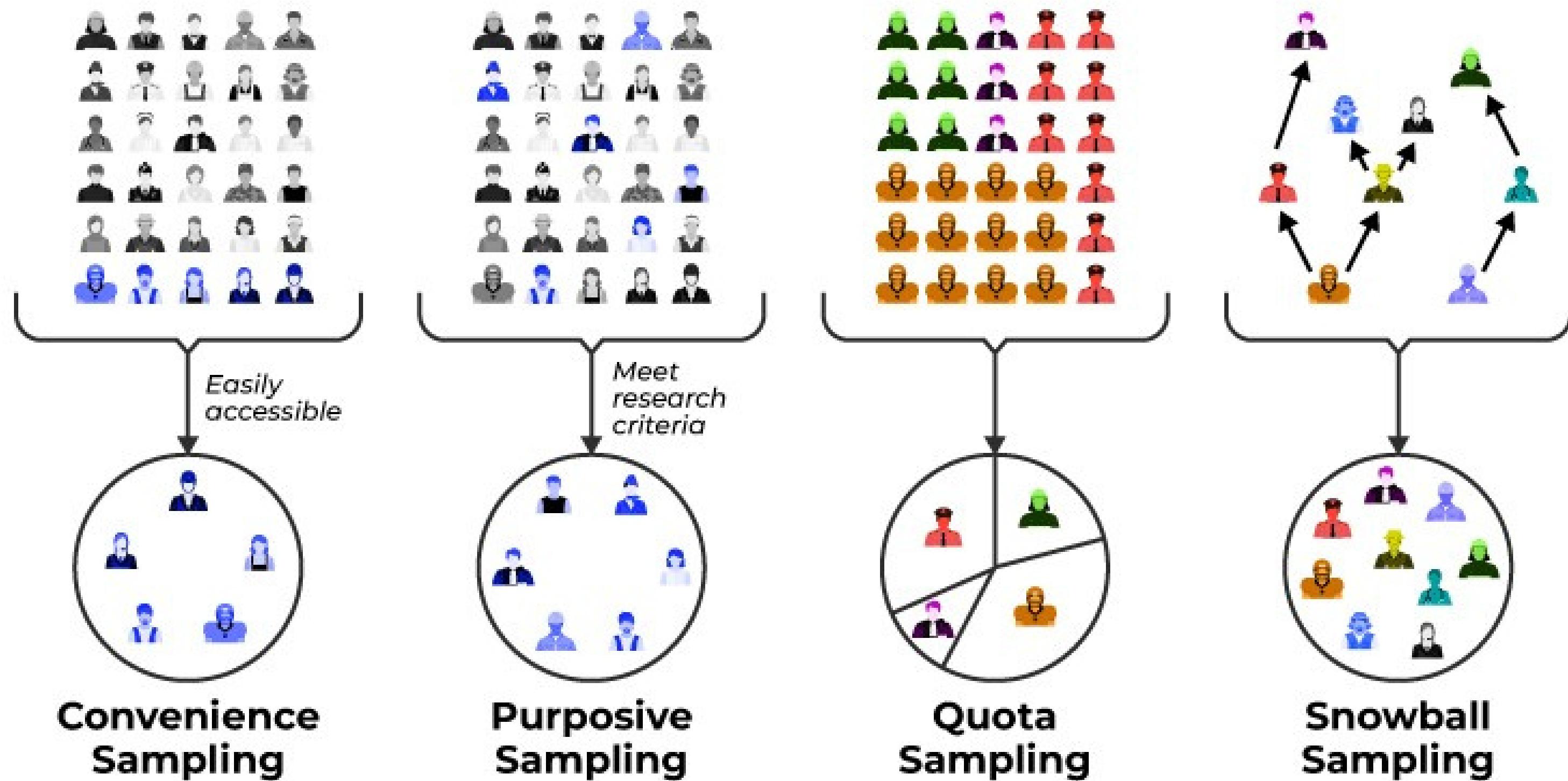
Stratified Sampling: Divides the population into subgroups based on relevant characteristics, then randomly selects samples from each subgroup.

Example: Studying primary school students' reading habits in a diverse city, divide students into different grade levels and randomly select a sample from each to ensure a representative sample.

Cluster Sampling: Divides the population into clusters or groups, then randomly selects entire clusters as sampling units.

Example: To evaluate the effectiveness of a new teaching methodology in primary schools across a state, randomly select a few schools from different regions and collect data from all students.

Non-Probability Sampling Methods



Non-Probability Sampling Methods

Non-Probability Sampling Non-probability sampling relies on the personal judgment of the researcher rather than the chance to select sample elements.

Convenience Sampling: Samples are selected based on the ease of access or availability.

Examples: Automobile registrations, telephone directories etc.

Judgment sampling: Judgmental sampling involves researcher-selected elements based on subjective judgment, offering targeted insights without generalizability.

Example: If a sample of 12 students is selected from the class of 100 students for analyzing their study hours, then the investigator may select 12 students who in his judgment representative of the class.

Convenience
Sampling

Purposive
Sampling

Quota
Sampling

Snowball
Sampling

Non-Probability Sampling Methods

Quota sampling: Quota sampling utilizes predetermined categories to select sample elements conveniently, aiming for representative samples at lower costs, often yielding results comparable to probability sampling.

Example: In a specific TV channel ESPN watching a survey in a town, the interviewers may interview 70 housewives, 80 teenagers, 30 business men and 20 jobless people.

Snowball sampling: Snowball sampling efficiently targets rare population characteristics through initial random selection and referrals, with low sampling variance or costs. Snowball sampling has been particularly used to Study drug cultures;

Heroin addiction;

Sampling

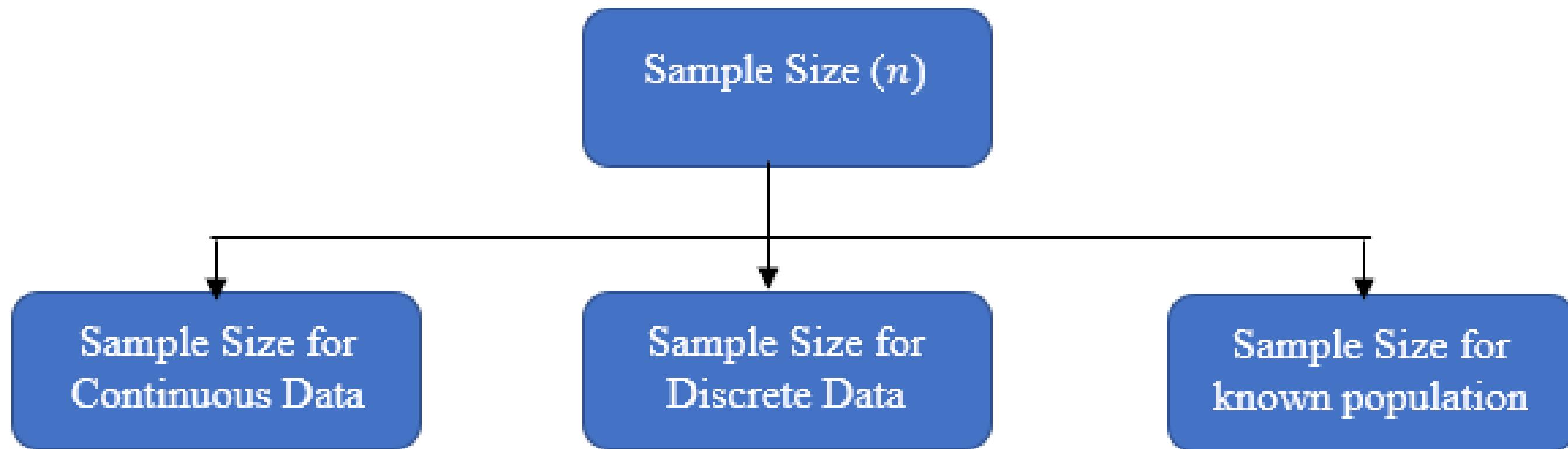
Purposive
Sampling

Quota
Sampling

Snowball
Sampling

Sample Size Determination

Sample size determination is the process of selecting the appropriate number of participants or units for a research study.



$$n = \left(\frac{Z\sigma}{E}\right)^2$$

$$n = \frac{Z^2(\hat{p}\hat{q})}{E^2}$$

$$n = \frac{N}{1 + NE^2}$$

Sample Size Determination

Here,

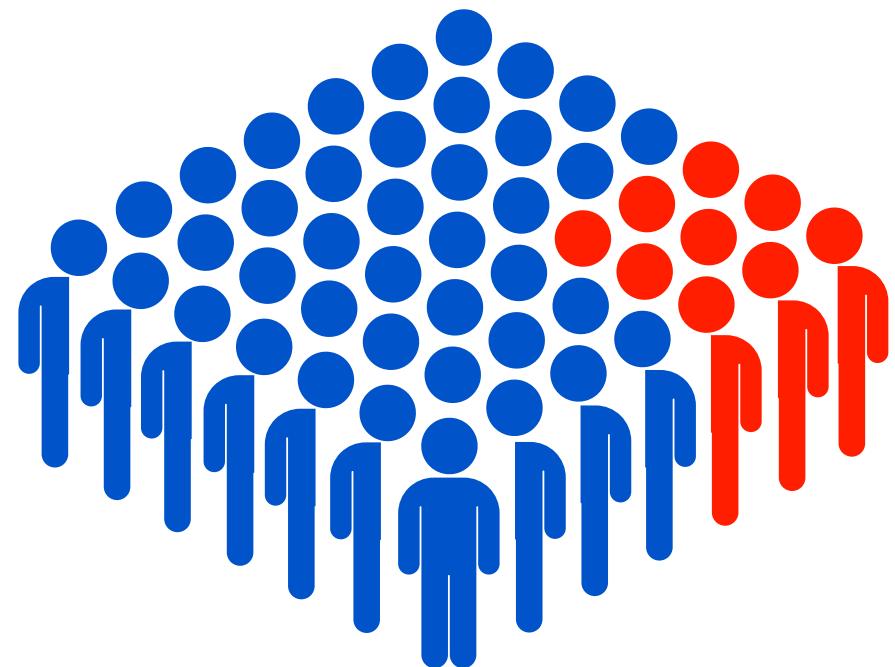
n is the sample size,

Z is the statistic corresponding to the level of confidence

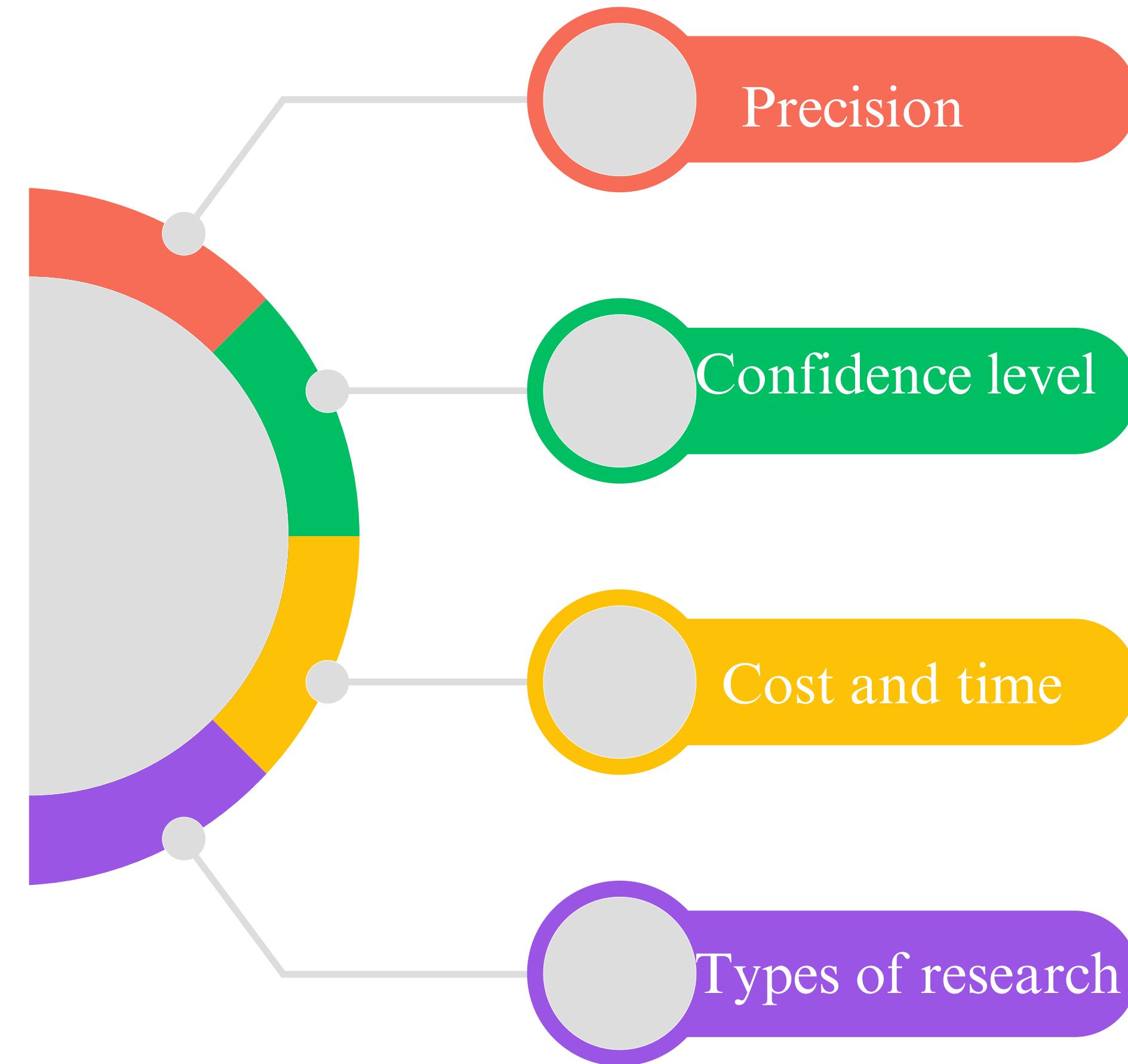
\hat{p} is the expected prevalence (that can be obtained from the same studies or a pilot study conducted by the researchers),

$\hat{q} = (1 - \hat{p})$ and

E is precision (corresponding to effect size)



Importance of Sample Size in Research



Importance of Sample Size in Research

Precision: Larger sample sizes provide more accurate population parameter estimates, enhancing the precision of study findings. Smaller samples may lead to biased or unreliable results.

Confidence level: Higher sample sizes increase the level of confidence in the study findings. Smaller samples may not provide enough information for valid conclusions.

Cost and time: Larger sample sizes may require more resources for data collection, processing, and analysis. Researchers must balance these trade-offs.

Types of research: Sample size is crucial for different research types, with qualitative studies typically using smaller sizes, while experimental studies may require larger ones to minimize extraneous variables.

Thank You

