INTRODUCTION TO STATISTICS FOR DATA SCIENCE

What is Statistics?

- Statistics is the science from learning from data

What is Statistics?

- Statistics is the science from learning from data

Steps to learn from data

- Defining the problem

What is Statistics?

- Statistics is the science from learning from data

Steps to learn from data

- Defining the problem
- Collecting the data

What is Statistics?

- Statistics is the science from learning from data

Steps to learn from data

- Defining the problem
- Collecting the data
- Summarizing the data

What is Statistics?

- Statistics is the science from learning from data

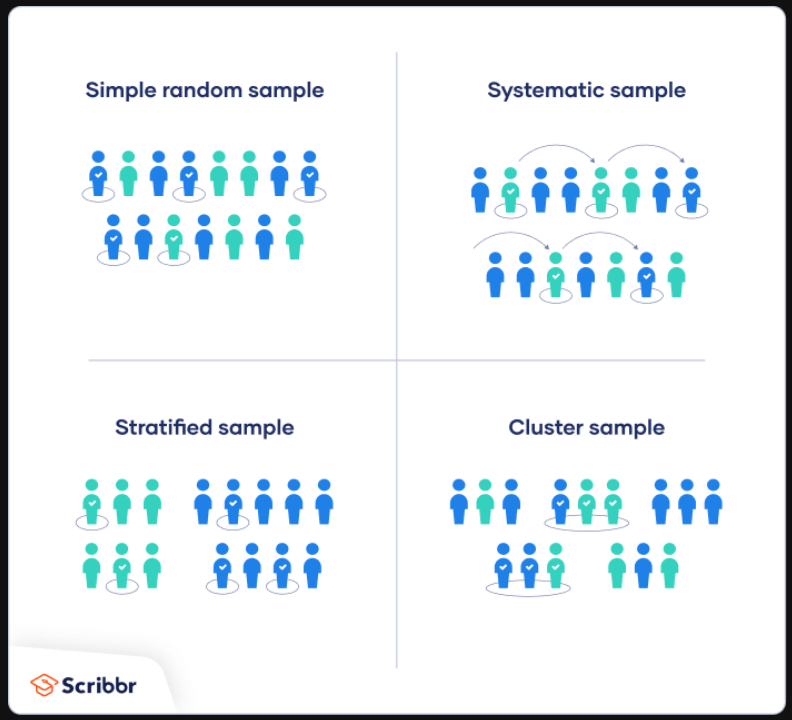
Steps to learn from data

- Defining the problem
- Collecting the data
- Summarizing the data
- Analyzing the data, interpreting the analyses, and communicating the results.

Population and Sample

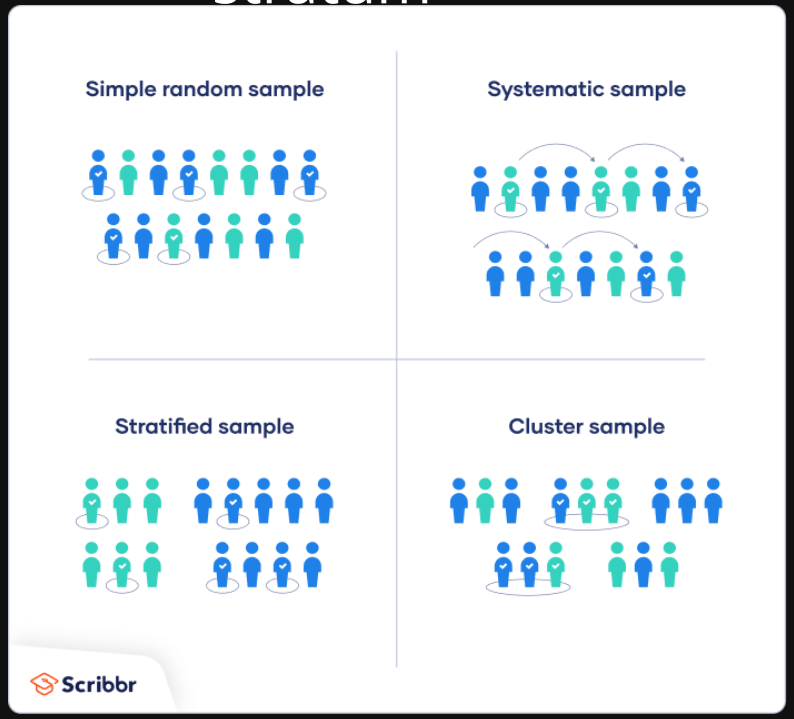
Types of Sampling

1. Simple random sampling: Each case is equally likely to be selected



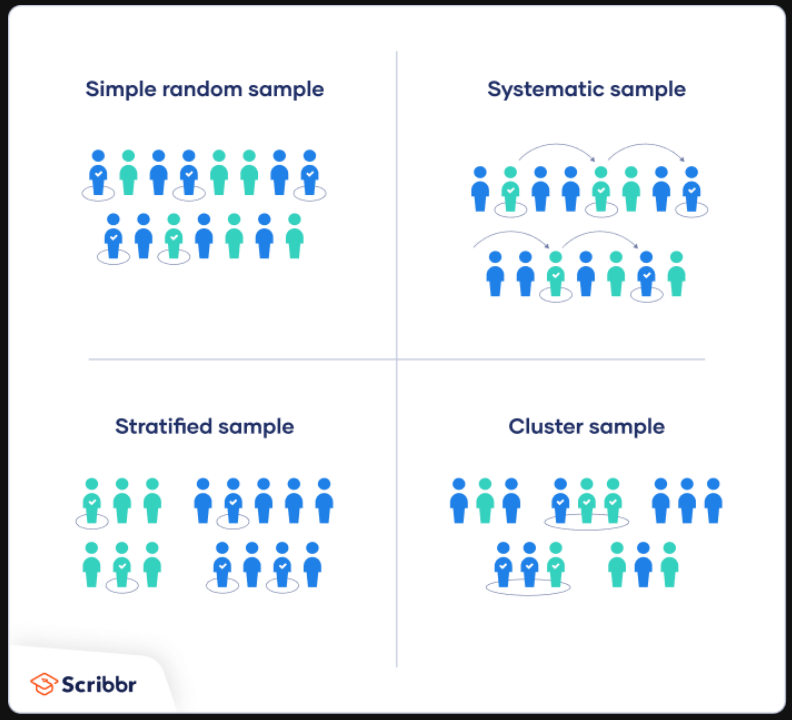
Types of Sampling

2. Stratified Sampling: Divide the population into homogeneous strata. Then, randomly sample from within each stratum



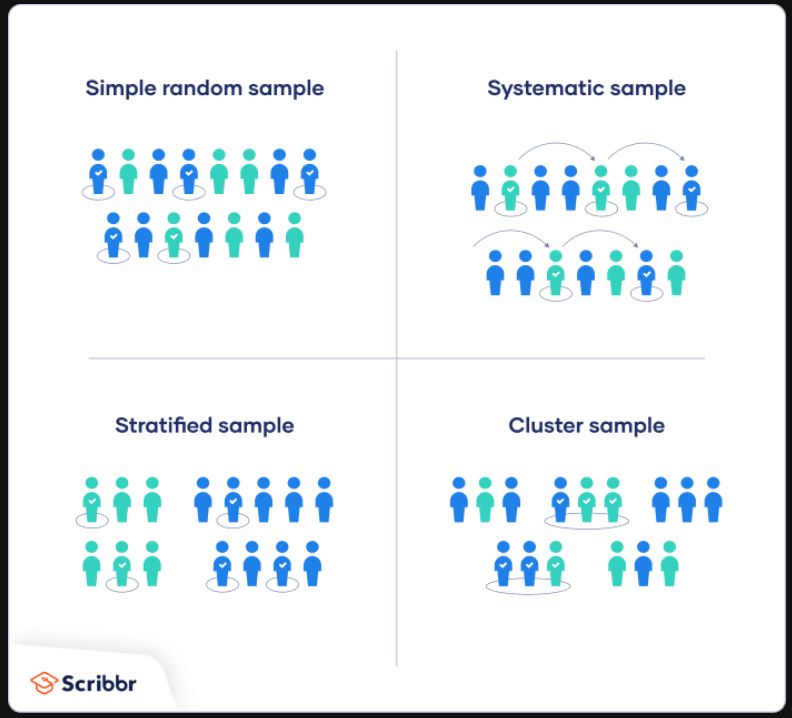
Types of Sampling

3. Cluster sampling: Divide the population into clusters. Then, randomly sample a few clusters and sample all observations within these clusters



Types of Sampling

4. Systematic Sampling: Select observation at a regular interval determined in advance.



Types of Statistics

- Descriptive Statistics
- Inferential Statistics

Data Collection Process

- Specify the objective of the study

Data Collection Process

- Specify the objective of the study
- Identify the variable(s) of interest

Data Collection Process

- Specify the objective of the study
- Identify the variable(s) of interest
- Collect the data

Types of variables

- Numerical(quantitative): take on numerical values (add, subtract, average, etc, with these values)

Types of variables

- Numerical(quantitative): take on numerical values (add, subtract, average, etc, with these values)
  - Continuous: take on an infinite number of values within some given range(amount of water in a container)

Types of variables

- Numerical(quantitative): take on numerical values (add, subtract, average, etc, with these values)
  - Continuous: take on an infinite number of values within some given range(amount of water in a container)
  - discrete: take on a specific set of numeric values (num of subscribers)

Types of variables

- Categorical(qualitative): take on limited distinct categories that can be identified with numbers but can't do arithmetic operations.

Types of variables

- Categorical(qualitative): take on limited distinct categories that can be identified with numbers but can't do arithmetic operations.
  - Nominal: Gender

Types of variables

- Categorical(qualitative): take on limited distinct categories that can be identified with numbers but can't do arithmetic operations.
  - Nominal: Gender
  - Ordinal: Level of Education