

# 03. Population Vs Sample Data

Mahir Faysal Tusher

June 2025

## Introduction

Before diving into key statistical concepts such as **measures of central tendency** and **measures of dispersion**, it is essential to understand the foundational concepts of **Population** and **Sample**.

## 1 Population

- A **Population** in statistics refers to the entire group that you want to draw conclusions about.
- It includes **all possible observations** of a particular type.
- It is usually denoted by a capital letter **N**.

### Example:

Imagine an island with 100,000 residents. This entire group of 100,000 people is considered the **population**.

## 2 Sample

- A **Sample** is a *subset* of the population.
- It is used when collecting data from the entire population is impractical or impossible.
- It is denoted by a lowercase letter **n**.

### Example:

From the island's population of 100,000 people, selecting 10,000 individuals to collect weight data from would be a **sample**.

## 3 Why Sampling is Needed

- Gathering data from the entire population may be time-consuming, costly, or logistically challenging.
- People might not be available, or access might be restricted.
- Therefore, using a sample provides a practical way to make statistical inferences.

## 4 Real-World Applications

- **Exit Polls:** During elections, it is not feasible to ask every voter whom they voted for. Instead, a sample of voters is surveyed, and based on this, predictions are made.
- **Medical Studies:** Researchers test new drugs on a sample of patients and use statistical inference to generalize results to the wider population.

## 5 Key Notations

- Population size:  $N$
- Sample size:  $n$

## 6 Summary

- **Population** = Complete data set (e.g., all residents of an island)
- **Sample** = Part of the population selected for analysis
- Sampling enables us to make predictions and conclusions about the population using manageable data sets.

## Next Topic

In the next session, we will begin exploring the **Measures of Central Tendency** which include Mean, Median, and Mode.