Statistics Overview

Statistics is the branch of mathematics that deals with the collecting, organizing, presenting, analyzing, and interpreting of data. These steps work together to turn raw information into useful knowledge.

Steps in Statistics

Collecting Data

Gathering relevant information (e.g., asking students their test scores).

Organizing Data

Arranging the data systematically (e.g., making tables or sorting scores).

Presenting Data

Showing the data clearly using tables, charts, or graphs so others can understand it.

Analyzing Data

Calculating summaries like averages and finding patterns.

Interpreting Data

Explaining what the analysis means and drawing conclusions.

Introduction to Statistics with Frequency Distribution and Histogram

Dataset: Score Ranges and Frequencies

Score Range	Frequency
0 - 9	5
10 - 19	8
20 - 29	12
30 - 39	7
40 - 49	3

Step 1: Collecting Data

The teacher collects data on how many students scored within each range on a test.

Step 2: Organizing Data

The scores are grouped into intervals, and their frequencies (number of students) are recorded in a table as shown above.

Step 3: Presenting Data

A **Histogram** visually shows this data:

- X-axis shows the score ranges (0-9, 10-19, etc.)
- Y-axis shows the frequency (number of students)
- Each bar's height corresponds to the frequency of students who scored in that range
- Bars are adjacent to reflect continuous data intervals

Step 4: Analyzing Data

- The highest frequency is 12 in the 20-29 range, indicating most students scored in this interval.
- The frequencies decrease for the highest and lowest score ranges, showing fewer students scored very low or very high.

Step 5: Interpreting Data

- The middle score ranges (10-29) have the most students, showing a central tendency.
- The teacher can focus on helping students who scored in the lower ranges (0-9) to improve.

Plotting Histogram in Python (Matplotlib example)

In [ ]:

Population and Sample in Statistics

Population

- Population refers to the entire group or set of individuals, objects, or data points that you want to study or get information about.

- It includes every member of that group.
- The characteristics of a population (like the exact mean or standard deviation) are called **parameters**.
- **Examples:**
  - All students in a school.
  - Every book in a library.
  - All voters in a city.

Sample

- A sample is a smaller subset taken from the population.
- It is used when it's impractical or impossible to collect data from the entire population.
- The characteristics of a sample (like the sample mean or sample standard deviation) are called **statistics**.
- Samples are used to make estimates or inferences about the whole population.
- **Examples:**
  - 50 students selected randomly from a school.
  - 100 books randomly checked in a library.
  - 500 voters questioned in a city survey.

Key Differences

Feature	Population	Sample
Definition	Entire group of interest	Subset of the population
Size	Usually large or complete group	Smaller part of the population
Data Type	Parameters (true population values)	Statistics (estimates from sample)
Data Collection	Census (complete enumeration)	Sampling (partial data collection)
Use	Gives exact information	Used to make inferences about population

---

Why Use Samples?

- Populations can be too large or difficult to study completely.
- Sampling saves time and resources.
- Proper sampling methods allow us to make reliable conclusions about the population using sample data.

This explanation helps differentiate population and sample clearly for better understanding in statistics.

Types of Data in Statistics

Data is generally classified into two main types:

---

1 Qualitative Data (Categorical Data)

This data describes **qualities or categories** — not measured with numbers.

Types of Qualitative Data:

Type	Description	Example
Nominal	Categories with **no natural order**	Gender: Male, Female, Other
Ordinal	Categories **with a meaningful order**	Shirt Size: Small, Medium, Large

---

2 Quantitative Data (Numerical Data)

This data deals with **numbers and quantities** — things you can count or measure.

📌 Types of Quantitative Data:

Type	Description	Example
Discrete	**Countable** numbers (whole numbers)	Number of siblings: 0, 1, 2
Continuous	**Measurable** numbers (decimals possible)	Height: 165.5 cm, 172.3 cm

---

Simple Example

Imagine you're collecting data from students:

Student	Gender	Shirt Size	Age	Number of Books	Height (cm)
A	Male	Medium	16	4	165.2
B	Female	Small	17	6	170.5

Breakdown:

- **Gender** → Qualitative → **Nominal**
- **Shirt Size** → Qualitative → **Ordinal**
- **Age** → Quantitative → **Discrete**
- **Number of Books** → Quantitative → **Discrete**
- **Height** → Quantitative → **Continuous**

Main Type	Subtype	Description	Example
Qualitative	Nominal	No order	Colors, Gender
Qualitative	Ordinal	Ordered categories	Rank, Shirt size
Quantitative	Discrete	Countable, whole numbers	No. of students
Quantitative	Continuous	Measurable, can have decimals	Weight, Temperature

Example 3: BMI of College Students

✅ Step 1: Define the Aim

- Estimate the **average Body Mass Index (BMI)** of college students.

✅ Step 2: Define Variable & Population

- **Variable:** BMI = Weight (kg) / [Height (m)]^2
- **Population:** All college students in a university or city

✅ Step 3: Data Collection & Measurement

- **Sampling Procedure:** Stratified sampling from various departments
- **Sample Size:** 200 students
- **Measurement Method:** Weighing scale and measuring tape

✅ Step 4: Analysis Techniques

- **Descriptive:** Calculate mean BMI, SD, max, and min
- **Inferential:** Estimate average BMI using confidence interval; compare groups using t-tests

In [ ]: