

# Statistics

---

**Dr. Shikha Deep**



# Chapter Goals

---

**After completing this chapter, you should be able to:**

- Explain how decisions are often based on incomplete information
- Explain key definitions:
  - ◆ Population vs. Sample
  - ◆ Parameter vs. Statistic
  - ◆ Descriptive vs. Inferential Statistics
- Explain the difference between Descriptive and Inferential statistics

# Dealing with Uncertainty

---

**Everyday decisions are based on incomplete Information.**

**Consider:**

- Will the job market be strong when I graduate?
- Will the price of Yahoo stock be higher in six months than it is now?
- Will interest rates remain low for the rest of the year if the federal budget deficit is as high as predicted?

# What is Statistics ??

---

**Numbers and data are used to assist decision making.**

**Statistics** is a tool to help process, summarize, analyze, and interpret numerical data.

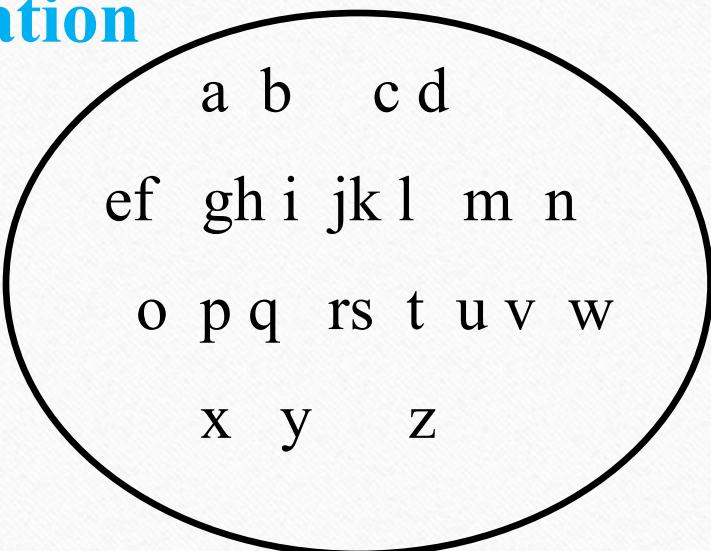
# Key Definitions

---

- A **population** is the collection of all items of interest or under investigation  
 $N$  represents the population size
- A **sample** is an observed subset of the population  
 $n$  represents the sample size
- A **parameter** is a specific characteristic of a population
- A **statistic** is a specific characteristic of a sample

# Population vs. Sample

## Population



Values calculated using population data are called **Parameters**

## Sample



Values computed from sample data are called **Statistics**

# Examples of Populations

---

- Names of all registered voters in the United States
- Incomes of all families living in Daytona Beach (City in Florida)
- Annual returns of all stocks traded on the New York Stock Exchange
- Grade point averages of all the students in your university

# Random Sampling

---

Simple Random Sampling is a procedure in which

- each member of the population is chosen strictly by chance,
- each member of the population is equally likely to be chosen,
- every possible sample of  $n$  objects is equally likely to be chosen

The resulting sample is called a Random Sample.

# Parameter and Statistic

---

- In a study of household incomes in a small town of **1000 households**, one might conceivably obtain the income of every household.
- However, it is probably very expensive and time consuming to do this.
- Therefore, a better approach might be to obtain the data from a portion of the households (let's say **125 households**).

# Parameter and Statistic

---

- In this scenario, the incomes of the 1000 households are referred to as the population and the incomes of the randomly selected 125 households are referred to as a sample.
- In this household incomes example, is the **average (mean)** income of all 1000 households a parameter or statistic? Is the average (mean) income of the 125 households a parameter or a statistic?

# Parameter and Statistic

---

- In this household incomes example, is the average (mean) income of all 1000 households a parameter or statistic? Is the average (mean) income of the 125 households a parameter or a statistic?
- Solution: The average (or mean) of the 1000 households is a **parameter**, whereas the average (mean) income of the 125 households is a **statistic**.

# **Descriptive and Inferential Statistics**

---

**Two branches of Statistics:**

## **1) Descriptive statistics**

- Graphical and Numerical procedures to summarize and process data

## **2) Inferential statistics**

- Using data to make Predictions, Forecasts and Estimates to assist Decision Making

# Descriptive Statistics

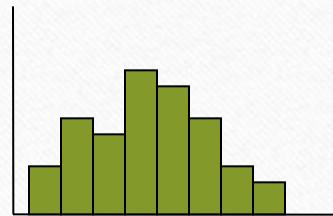
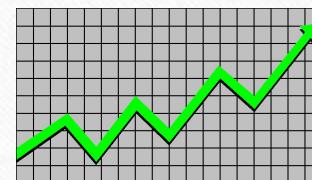
- Collect data

e.g., Survey



- Present data

e.g., Tables and graphs



- Summarize data

e.g., Sample mean =  $\frac{\sum X_i}{n}$

# Inferential Statistics

- **Estimation**

e.g., Estimate the population mean weight using the sample mean weight

- **Hypothesis Testing**

e.g., Test the claim that the population mean weight is 140 pounds



# Numeric Examples of Statistics

---

## Example 1: Village Food Consumption

Data: The weekly food consumption (in kg) for 10 households is: 25, 30, 22, 28, 30, 26, 35, 29, 30, 31.

**Task: Calculate the mean, median, and mode of the data.**

**Mean:**

$$\text{Mean} = (25 + 30 + 22 + 28 + 30 + 26 + 35 + 29 + 30 + 31) / 10 = 286 / 10 = 28.6 \text{ kg}$$

# Continued...

---

## **Median:**

Arrange the data in ascending order: 22, 25, 26, 28, 29, 30, 30, 30, 31, 35.

Median is the average of the 5th and 6th values:  $\text{Median} = (29 + 30) / 2 = 29.5 \text{ kg}$

## **Mode:**

The most frequent value is 30 kg (appears 3 times).

# Continued...

---

## Example 2: Farmer's Planting Decision (Probability)

Data: The farmer has data from the last 10 years showing that in 6 out of 10 years, it rained within the first two weeks of April.

**Task:** Calculate the probability that it will rain in the first two weeks of April this year.

Probability Calculation:

Probability (Rain) = Number of years it rained / Total number of years =  $6 / 10 = 0.6$  or 60%

# Continued...

---

## Example 3: School Test Scores

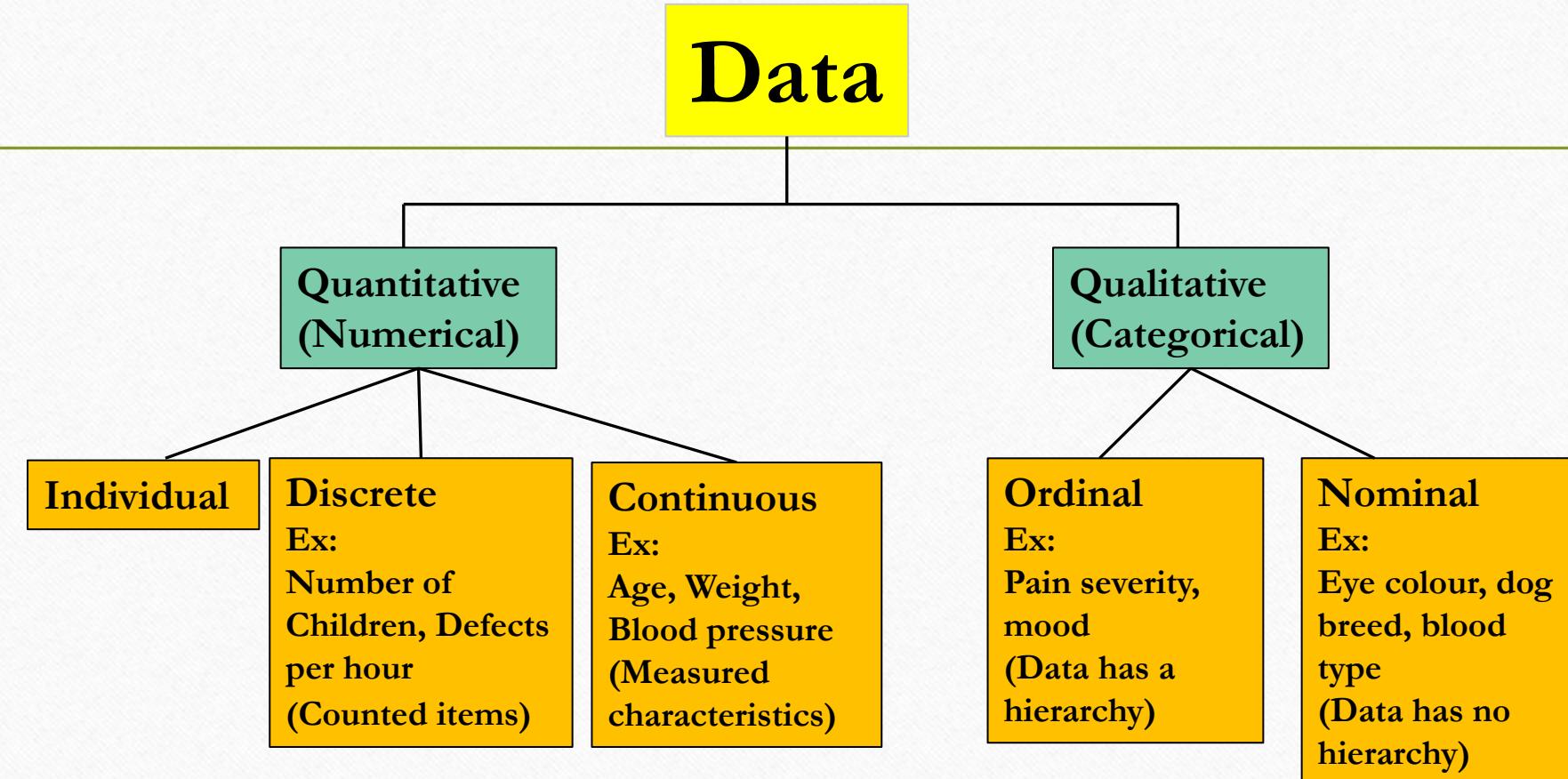
Data: Test scores of 20 students: 55, 60, 65, 70, 75, 80, 85, 55, 60, 65, 70, 75, 80, 85, 55, 60, 65, 70, 75, 80.

**Task:** Create a frequency distribution table and draw a histogram.

Frequency distribution is:

Interval	Frequency
55-59	3
60-64	3
65-69	3
70-74	3
75-79	3
80-85	5

# Types of Data



# Presentation of Data

