Introduction to Statistics: Population, Sample, Data

Presented By
Aditya Joshi
Assistant Professor
Graphic Era Deemed to be University

Statistics

- It is a field of mathematical science and its study concern with summarizing data, interpreting data, and making decisions based on data.

- **Statistic:** The Quantity calculated in a sample to estimate a value in population is called statistic.

Def. A collection of tools and techniques that are used to convert data into meaningful information.

Def. It's a study of collecting organizing and summarizing and presenting data .

What does statistician do?

1. Collects number or data.
2. Systematically organizes or arranges the data.
3. Analyses the data: extract relevant information to provide a complete numerical description.
4. Infers general conclusions about the problem using this numerical description.

Population

- It is a universal set of all objects under study.
- It is any entire collection of people, animals, plants or things from which we may collect data.
- It is the entire group we are interested in and wish to describe or draw conclusions about.

Example: Students of Graphic Era

People living in India

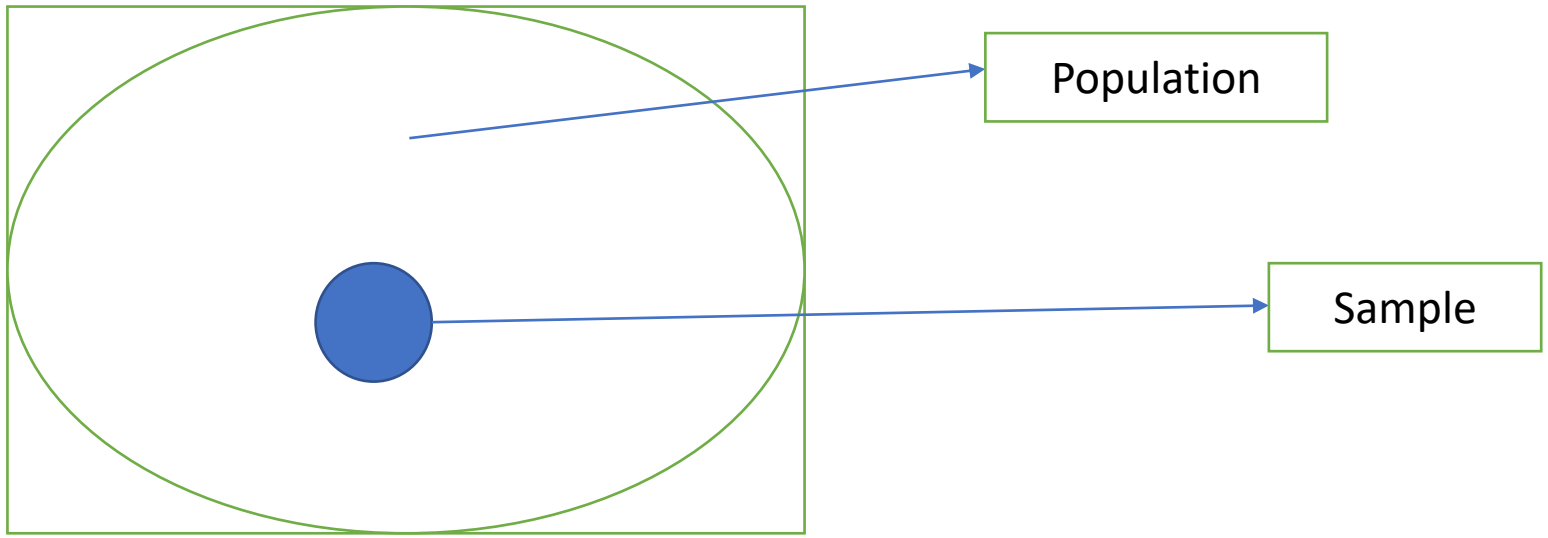
Bulbs made in factory

Different models of cell phones.

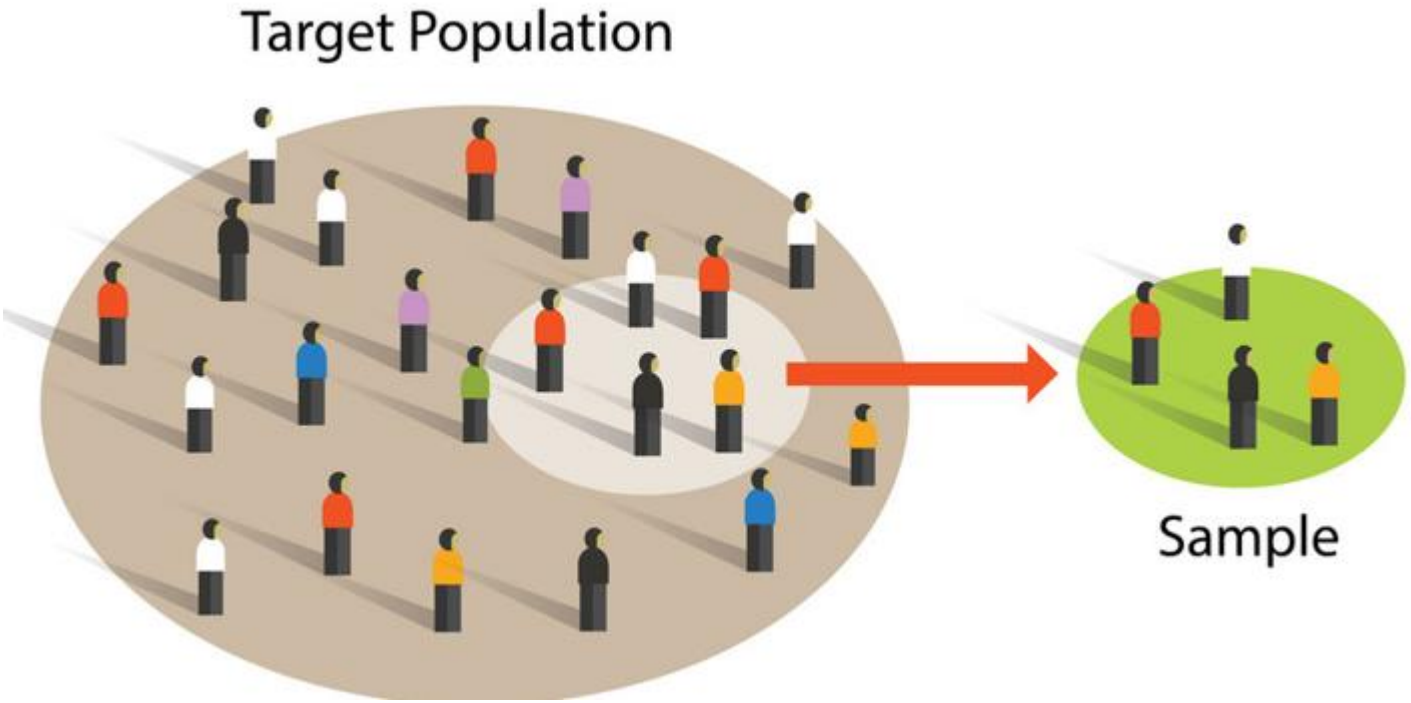
A **population** is a collection of data whose properties are analyzed. The population is the **complete** collection to be studied, it includes all subjects of interest.

Sample(subset of population)

A **Sample** is a group of units selected form a larger group (the population) by studying the sample it is hoped to draw valid conclusion about the larger group.



The term population is used in statistics to represent all possible measurements or outcomes that are of interest to us in a particular study.



Sample Size

Sample size is the number of observations used for calculating estimates of a given population.

E.g. If we interviewed 30 random students at a given university to see if they are interested in data science **“30 students”** would be our Sample Size.

E.g. if you want to know the average height of girls aged 15-30 in Graphic Era.

Population: Every one in that age range.

Sample: selections made form the population.

Data

A collection of facts or information.

E.g. Restaurants in India.

Types of cars.

Height of all students in your class.

Age of all students in Graphic Era.

Example: find the height of all students in your class. Organizes and summarizes the data. (compare, Average etc)

How many of you like dominos, KFC, Mc Donald's, Pizza Hut

Different type of Data

1. Primary Data
2. Secondary Data
3. Qualitative Data
4. Quantitative Data

Primary Data

Original data that has been collected specially for the purpose in mind.

It means when an authorized organization, investigator or an enumerator collects the data for the first time from the original source.

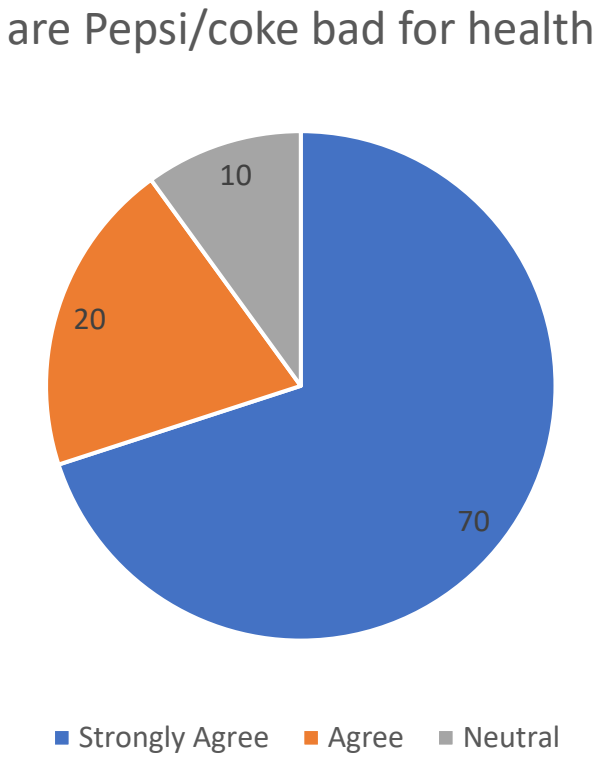
Data collected this way is called primary data

Eg. Questionnaire

Survey information

Survey : are Pepsi/coke bad for health

1. Strongly Agree
2. Agree
3. Neutral



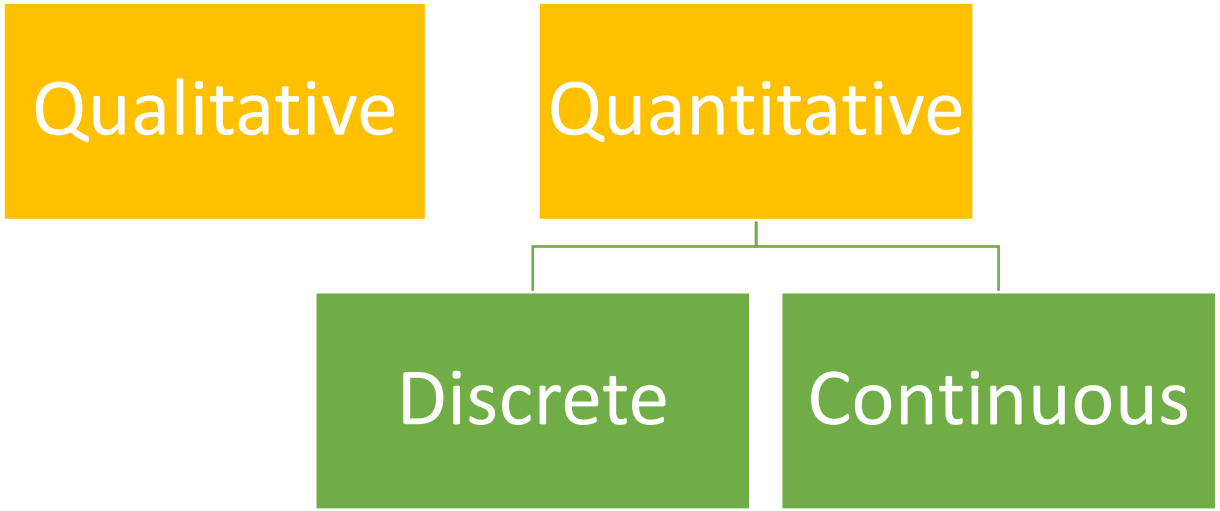
Secondary Data

Secondary data is a data that has been collected from another purpose. When we use statistical method or analysis in primary data then it is called secondary data.

Secondary data is the data that has been reused.

- Data from book
- New paper
- Magazine
- internet

Other Types of Data



Qualitative Data

It measures a quality or
Characteristic of an each experimental unit it is also called **categorical data**

E.g. Hair Color (Black, Brown, White, Grey)
Make of Car(Honda, Tata, Ford, Toyota)
Gender(Male, Female, Trans)
Place of Birth(Dehradun, Nainital, Pauri, Almora)

Quantitative Data

- Quantitative Data is a numerical measurement in terms of numbers.

E.g. Temperature= 26 degrees

Height=1.8 meters

Length =2.5 feet

Age= 20 years

Note: Quantitative Data always are associated with a scale measure (degree/feet/years)

Quantitative data measure a numerical quantity on each experimental unit.

e.g. for each orange tree, the number of oranges is measured.

For a particular day, the number of cars entering in CS/IT parking

Time until a light bulb burns out (4 months)

Qualitative vs. Quantitative

Qualitative > quality	Quantitative > quantity
Deals with descriptions	Deals with numbers
Data can be observed but not measured	Data which can be measured
Color, texture, smells, taste, appearance, beauty etc.	Length, height, area, volume, weight, speed, time, temperature, humidity, sound levels, cost, members, ages etc.

Types of Quantitative Data

- Discrete Data: it can only have specific value when we use to count something (in whole numbers)

1,2,4,5, …..

Rolling a dice {1,2,3,4,5,6}

Number of questions in exam {5,7,14}

No of students {20000}

- Continuous Data: it can take any value in an interval , it can be measured,

[1,10]

1<=x<=10

1, 1.3, 1.56, 6.89,10

E.g. Temperature

Check for discrete or continuous

- The speed of a car
- The number of dogs in society
- Your body weight
- The number of books in your library
- The height of student
- Number of students in MCA 1
- Your exact age
- No of goals scored by hockey team