Firstly, let's consider why we need to learn statistics. As students of disciplines like CSE, BBA, or Economics, why is statistics important for us?

Well… Statistics plays a crucial role in the field of Computer Science and Engineering (CSE) in several ways. Here are some key areas where statistics is important in CSE:

a) ***Data Analysis:*** CSE involves working with large amounts of data, and statistical techniques are essential for analyzing and making sense of this data.
b) ***Machine Learning and Artificial Intelligence:*** Statistics forms the foundation of machine learning and artificial intelligence algorithms. Statistical models and techniques are used to train, validate, and optimize machine learning models. These models enable tasks such as image recognition, natural language processing, recommendation systems, and predictive analytics.
c) ***Quality Assurance and Testing***: Statistical methods play a role in quality assurance and software testing. Techniques like hypothesis testing, control charts, and statistical process control are employed to evaluate software reliability, identify defects, and assess the overall quality of software products.
d) ***Network Analysis and Security:*** Statistical analysis is used in network traffic analysis, intrusion detection, and cybersecurity. Statistical models and algorithms can help identify patterns of abnormal behavior, detect network anomalies, and develop intrusion detection systems.
e) ***Optimization and Decision Making:*** Statistics is utilized in optimization problems and decision-making processes. Statistical techniques like linear programming, regression analysis, and optimization algorithms help in solving complex problems, optimizing resource allocation, and making data-driven decisions.

In summary, statistics plays a critical role in various aspects of Computer Science and Engineering. It provides the tools and methods necessary to analyze data, develop models, evaluate performance, optimize systems, ensure quality, and make informed decisions in a wide range of CSE applications.

So, what do you think? Do we need this course or not? Obviously, we need!!!!

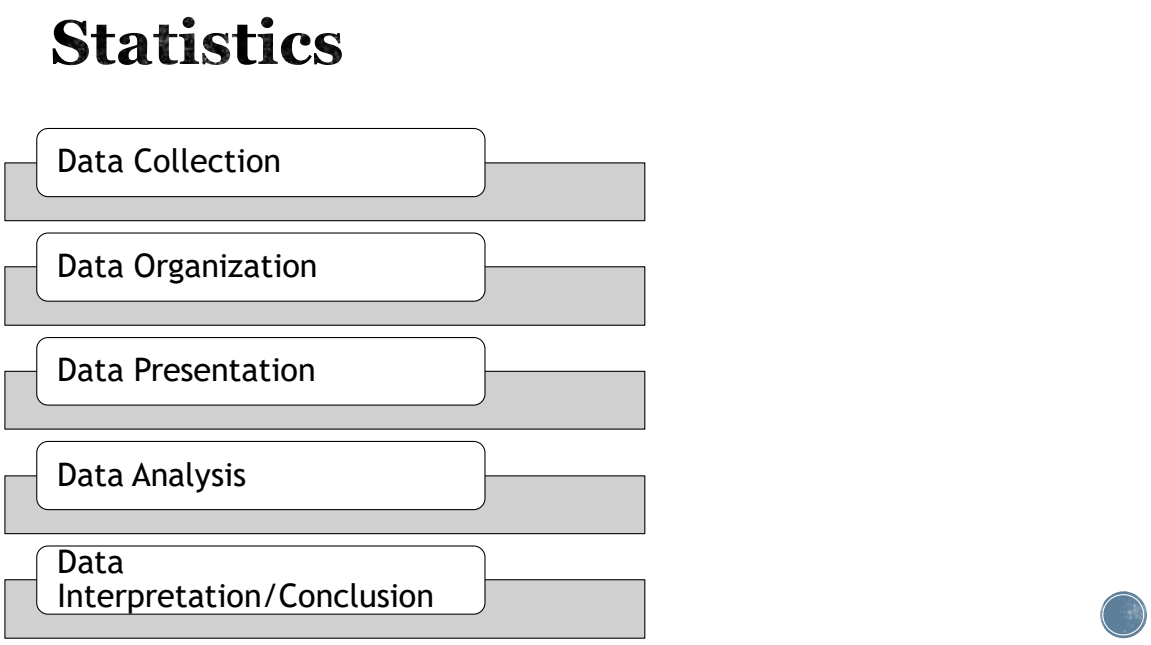
Now the question is "Sir, what is statistics?"

Statistics is a science that deals with data. When it deals with data, it follows five steps or stages. What are they?

1. Collection: Statistics involves the process of gathering data from various sources.
2. Organization: Once the data is collected, it needs to be organized in a systematic manner.
3. Summarization/Presentation: After organizing data, it's important to present it in a meaningful way.
4. Analysis: Statistical analysis involves applying various mathematical and statistical techniques to the collected data.
5. Interpretation: Once the analysis is done, the results need to be interpreted.

For example, suppose you want to know the average height of students in your class. You can measure the height of each student and record the data. Then you can use descriptive statistics to calculate the mean, median, mode, range, and standard deviation of the data. These are numerical summaries that tell you something about the distribution of the data. You can also use graphs or charts to display the data visually. For example, you can use a histogram or a box plot to show the shape and spread of the data.

Now suppose you want to know if the average height of students in your class is different from the average height of students in another class. You can describe the data by using statistical methods and procedures and making conclusions based on the analyzed data.



Look this picture carefully. For any kind of work, any type of work we need data collection and organization. After that, you will get two different scenarios.

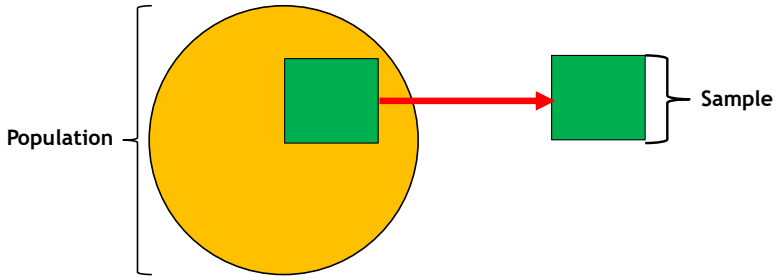
a) Presentation/Summarization: Here we just summarize and describe our data. In statistical term, it is called "Descriptive"
b) Analysis: Here we try to extract some analytical output of my data. In statistical term, it is called "Inference".

From this, we can say that "Two types of statistics"

1. Descriptive statistics: Which used to summarize, organize, and present a set of data/observations in a meaningful way (e.g., tables, graphs, numerical summaries).
2. Inferential Statistics: Work with sample data to make a conclusion about the population.

Population & Sample

- A population is the entire collection of individuals, objects.
- A small but representative part of the population is called sample.

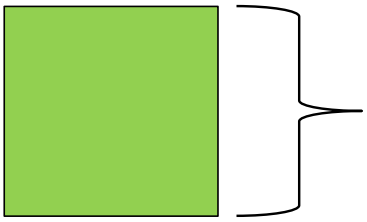


Descriptive vs Inference

Descriptive Statistics	Inferential Statistics
Describe and summarize the main characteristics of Data	Make conclusion about population based on sample data
Applicable to both populations and samples	Applicable to only for samples
Uses measures of central tendency, measures of dispersion, and graphical representations to summarize and present the data	Utilizes statistical techniques such as hypothesis testing, confidence intervals, and regression analysis

Census & Survey

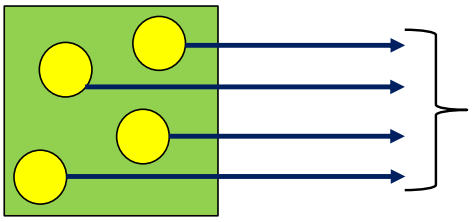
- To collect data about an entire population is called “Census”.
- To collect data from the part of the population is called “Survey”



Collects data from this entire box is “Census”

Census & Survey

- To collect data about an entire population is called “Census”.
- To collect data from the part of the population is called “Survey”

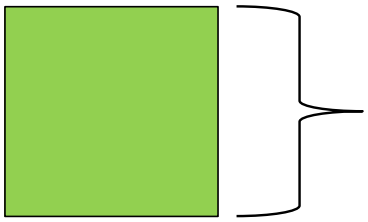


Collects information from these part is “Survey”

Parameter & Statistic

- A parameter is a numerical value that describes a characteristic (mean, variance etc.) of a population.

- A statistic is a numerical value that describes a characteristic (mean, variance etc.) of a sample.

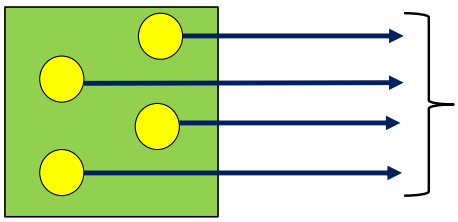


Characteristics of this entire box is "Parameter"

Parameter & Statistic

- A parameter is a numerical value that describes a characteristic (mean, variance etc.) of a population.

- A statistic is a numerical value that describes a characteristic (mean, variance etc.) of a sample.



Characteristics of these part of population is "Statistic"

Variable

For example, Gender is a variable

For example, Height is a variable

- If the values of a characteristics vary
- From person to person
- From object to object
- From phenomenon to phenomenon

Types of Variable

- Two types of variable:

1. Qualitative variable (Values can not be measured numerically)
2. Quantitative variable (Values can be measured numerically)
    a. Discrete variable (Countable values)
    b. Continuous variable (Any values within a range)

Data

- Data are "some information"
- That has been "collected" from field
- Translated into a form that is efficient for processing.

Types of Data

- Two types of data:

1. Qualitative data (Values can not be measured numerically)
2. Quantitative data (Values can be measured numerically)
    a. Discrete data (Countable values)
    b. Continuous data (Any values within a range)

Levels of measurements

Scales of Measurements

Data measurements

- Refer to the different ways in which variables or data can be categorized or measured.

- Four measurements
  1. Nominal
  2. Ordinal
  3. Interval
  4. Ratio

Scales of Measurements

Nominal	Ordinal
Must be categorical/qualitative	Must be categorical/qualitative
Can't be find differences	Can't be find differences
Can't be find ratios	Can't be find ratios
Can't be ranked	Can be ranked

Interval Level of Measurement

Definition: The interval level of measurement deals with numerical values where the intervals between numbers are meaningful. However, there is no true zero point (i.e., zero does not mean the absence of the quantity).

Key Characteristics:

Equal Intervals: The difference between two values is meaningful.
No True Zero: Zero is arbitrary and does not indicate the absence of the attribute.

Examples:

Temperature in Celsius or Fahrenheit:
The difference between 20°C and 30°C is the same as between 30°C and 40°C.
However, 0°C does not mean 'no temperature'; it's just another point on the scale.

Calendar Years:

The difference between the years 2000 and 2010 is the same as between 2010 and 2020.
The year 0 does not mean there was no time before it.

Ratio Level of Measurement

Definition: The ratio level of measurement also deals with numerical values, but unlike the interval level, it has a true zero point, meaning zero indicates the absence of the quantity.

Key Characteristics:

Equal Intervals: Just like the interval level, the difference between values is meaningful.
True Zero: Zero indicates the absence of the attribute, allowing for the calculation of ratios.

Examples:

Height and Weight:

A height of 0 cm means no height.

If one person weighs 60 kg and another 30 kg, the first person is twice as heavy.

Income:

An income of $0 means no income.

If someone earns $1000 and another person earns $500, the first person earns twice as much.

Simple Comparison:

Interval Example: Temperature (Celsius/Fahrenheit) - You can say it's 10°C hotter or colder, but 0°C doesn't mean there is no temperature.

Ratio Example: Weight - You can say one person weighs twice as much as another, and 0 kg means no weight at all.