Introduction to statistics

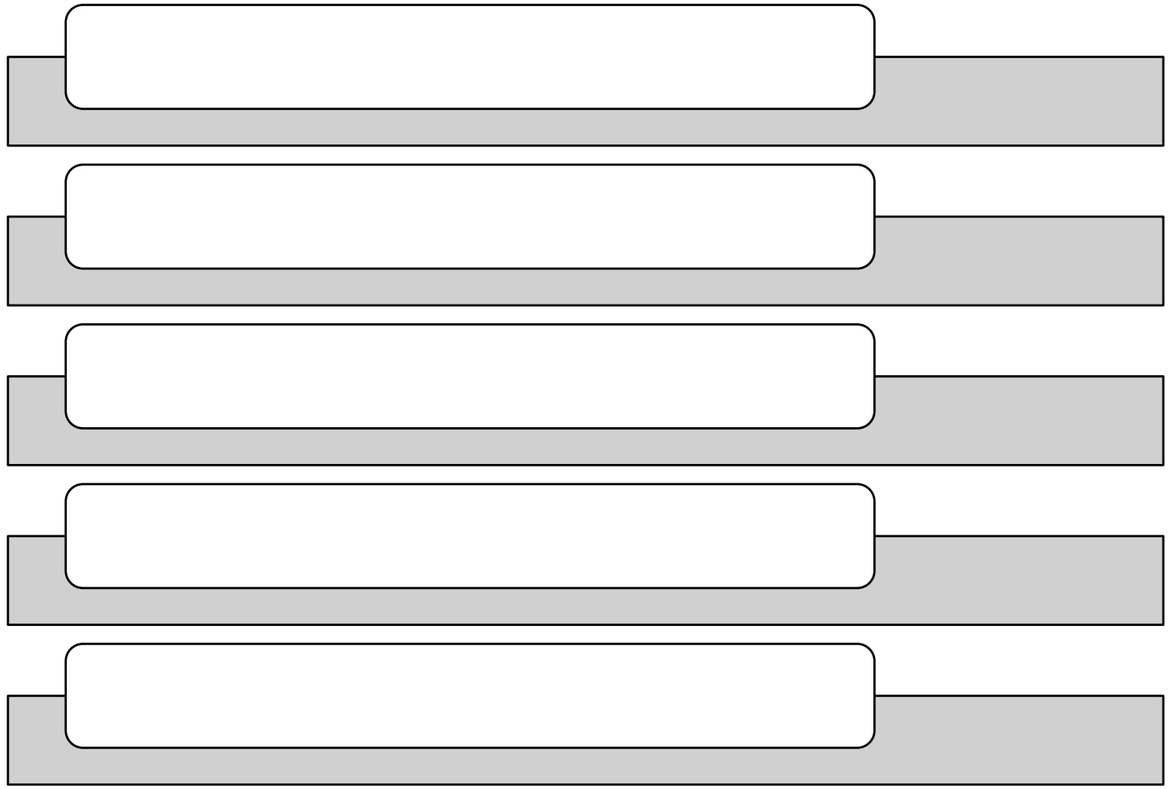
Md. Ismail Hossain Riday

What is Statistics

- Statistics is a science that deals with data

What is Statistics

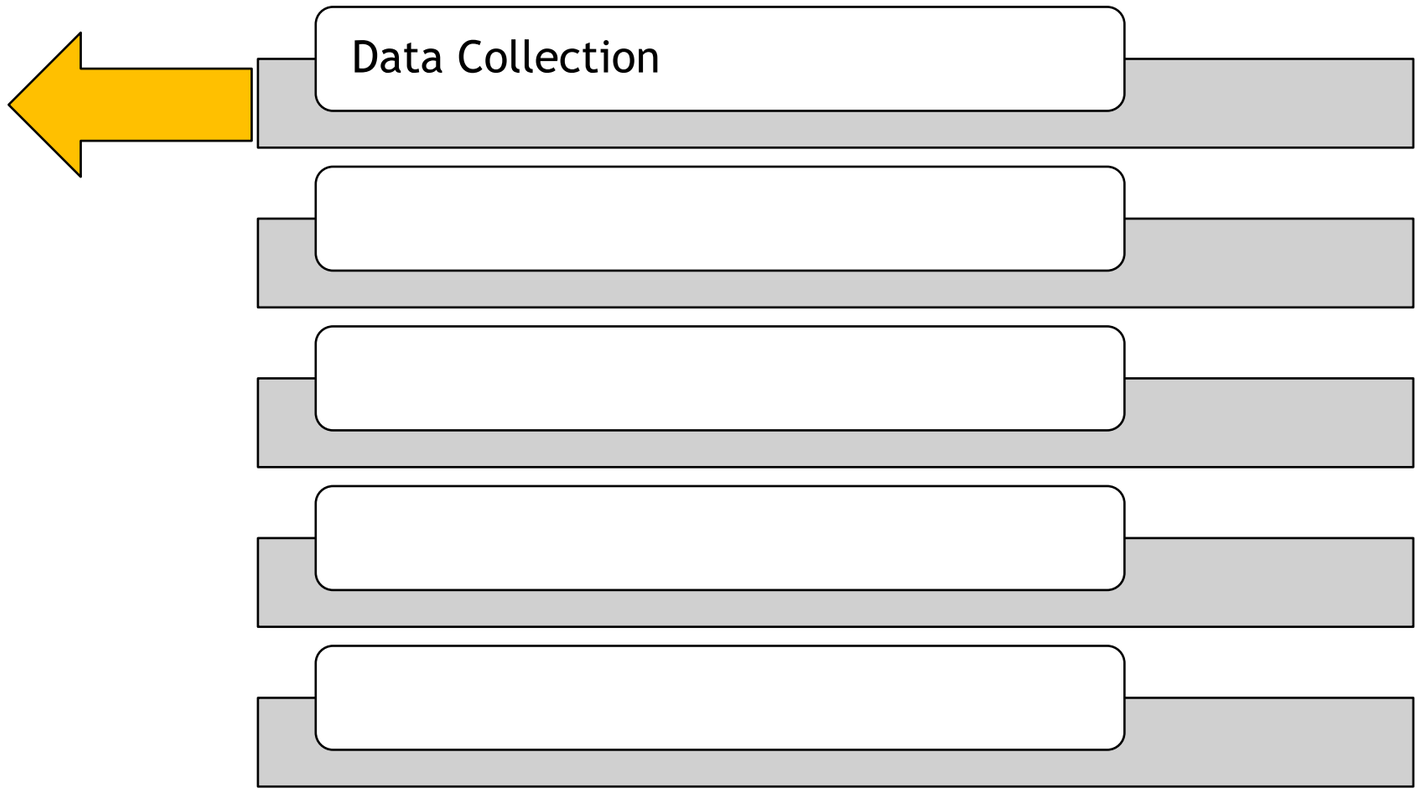
- Statistics is a science that deals with data



What is Statistics

- Statistics is a science that deals with data

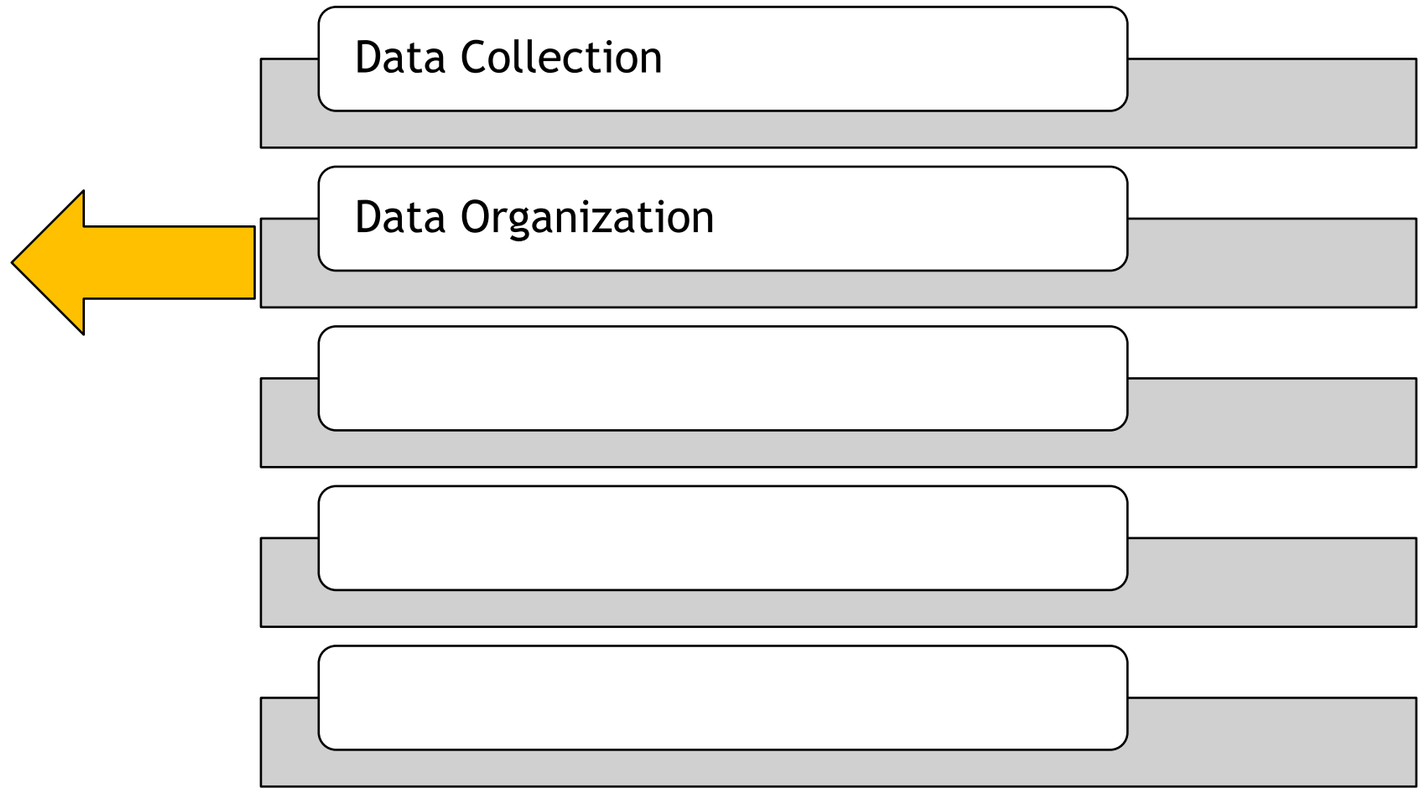
Statistics involves the process of gathering data from various sources.



What is Statistics

- Statistics is a science that deals with data

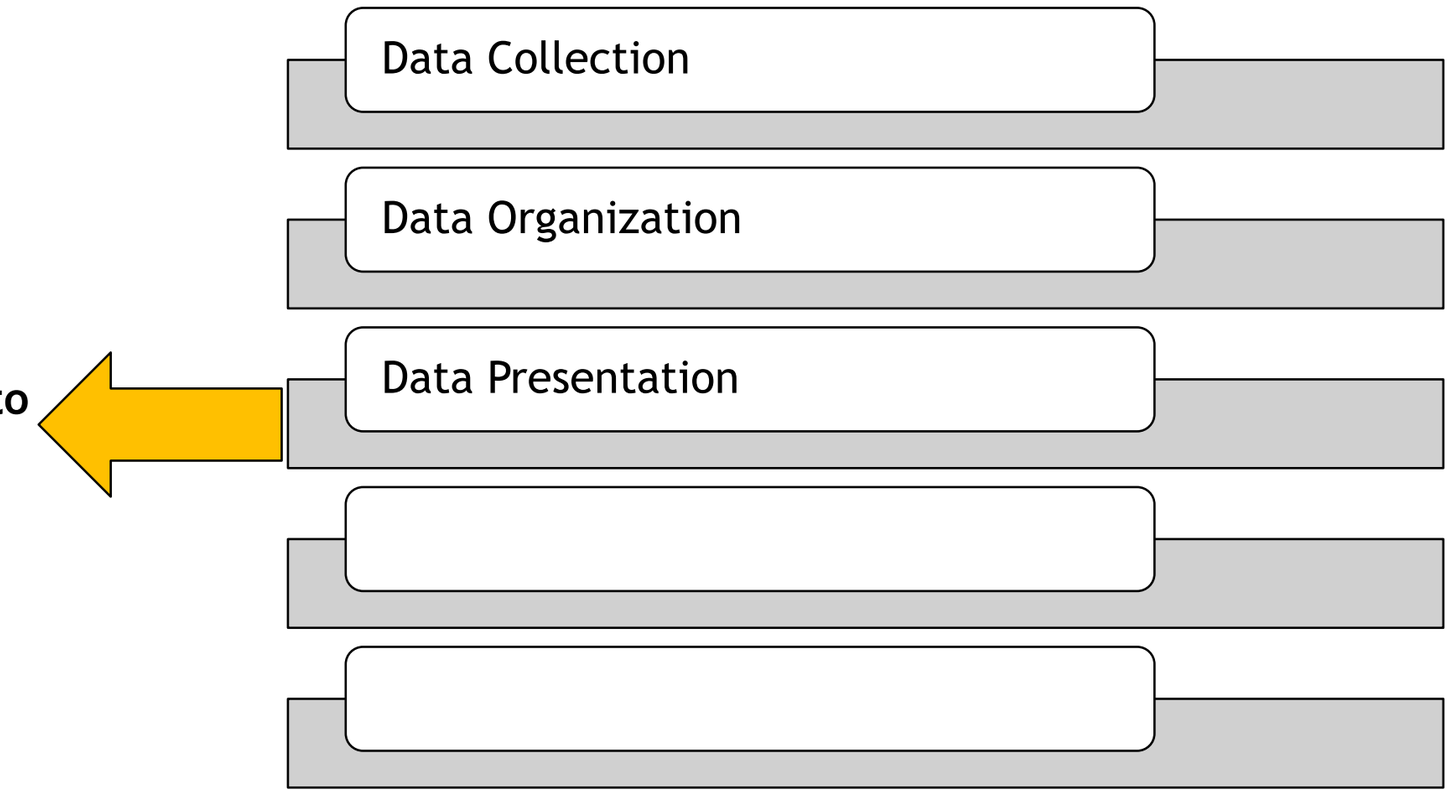
Once the data is collected, it needs to be organized in a systematic manner.



What is Statistics

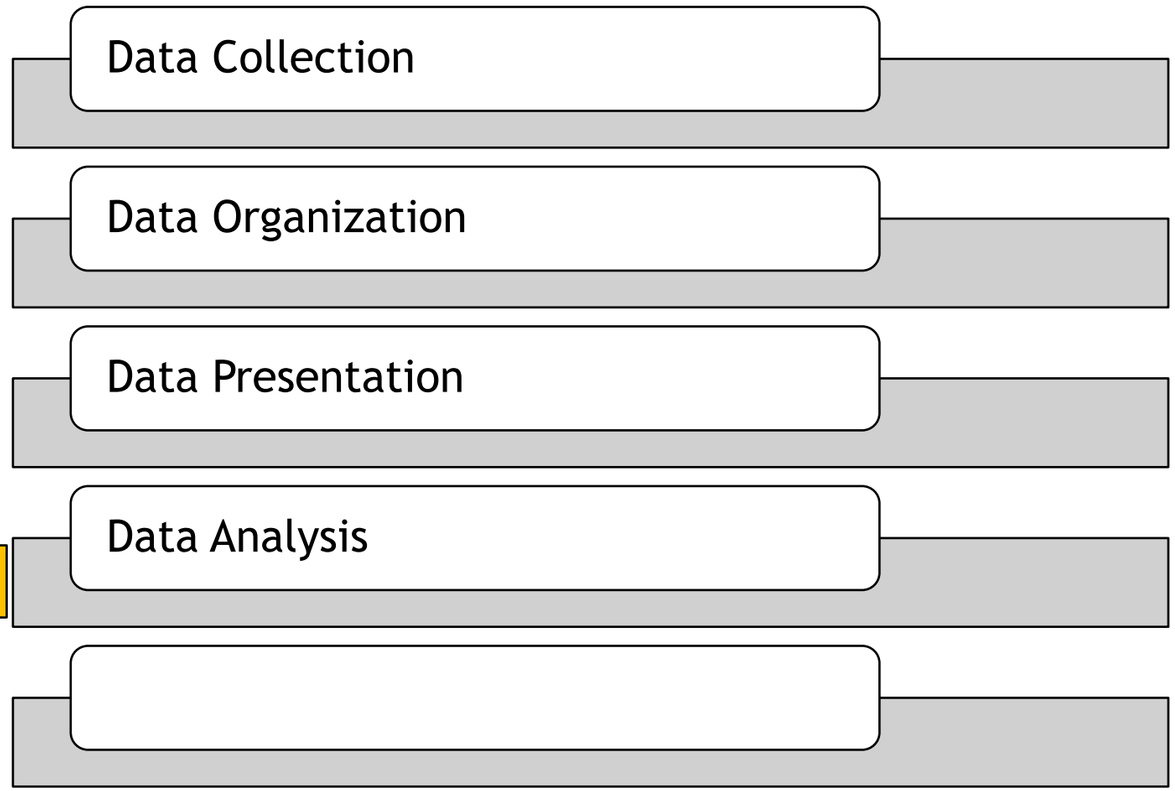
- Statistics is a science that deals with data

After organizing data, it's important to present it in a meaningful way.



What is Statistics

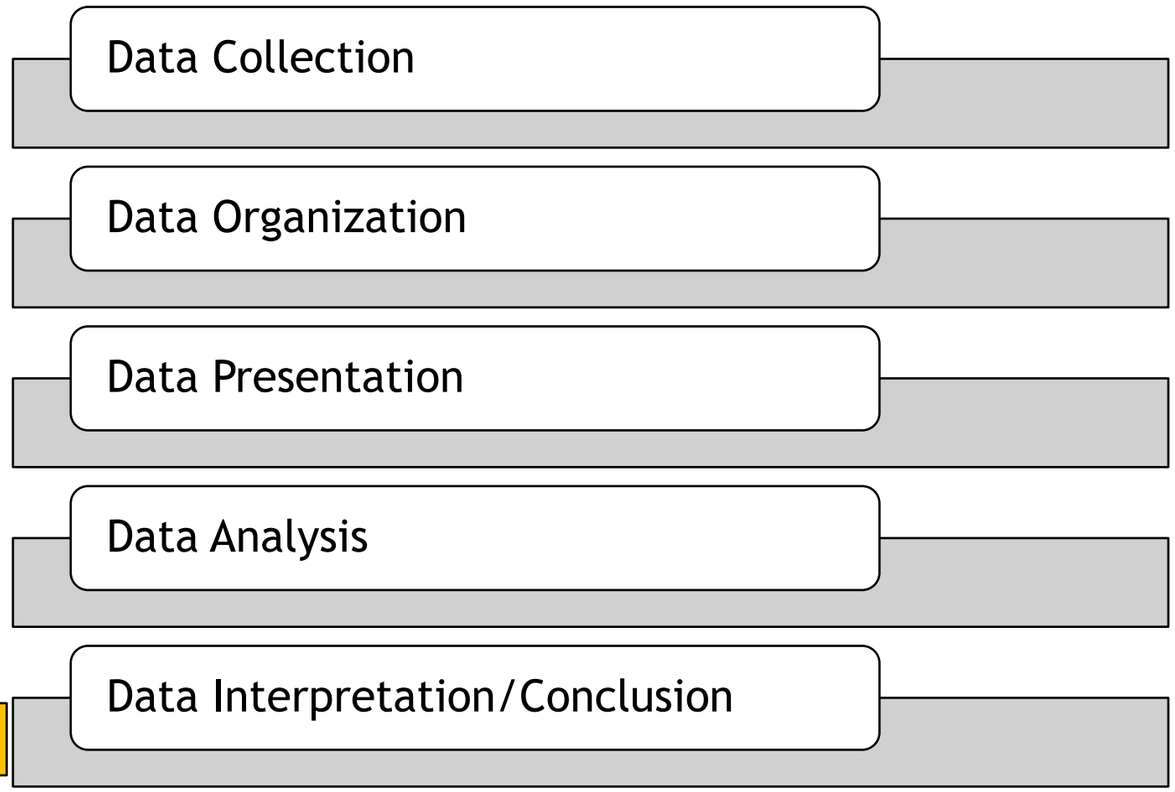
- Statistics is a science that deals with data



Statistical analysis involves applying various mathematical and statistical techniques to the collected data.

What is Statistics

- Statistics is a science that deals with data



Once the analysis is done, the results need to be interpreted.

What is Statistics

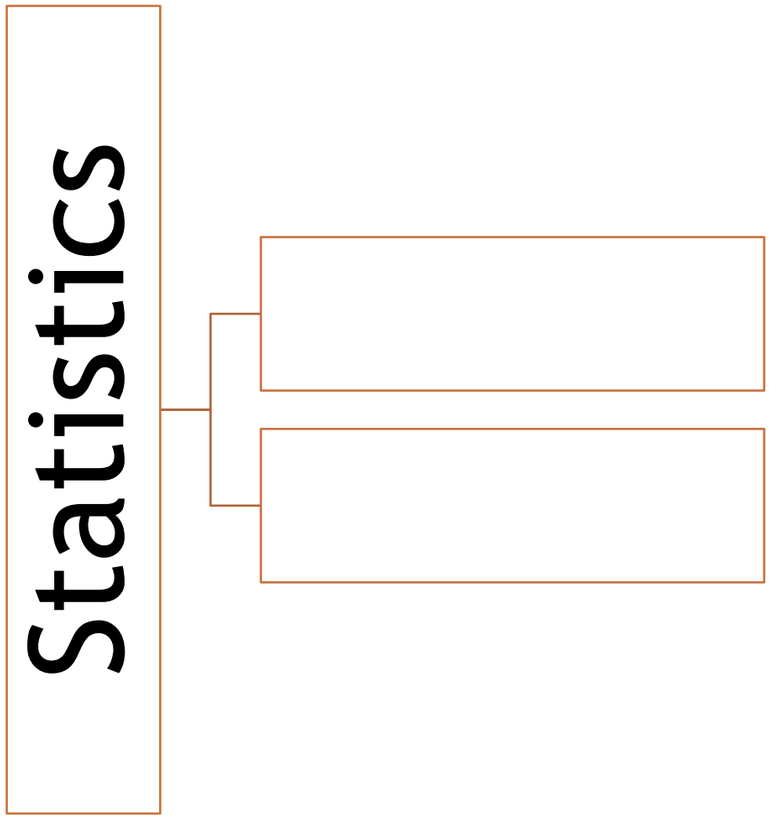
- Statistics is the science that deals with the collection, organization, summarization/presentation, analysis, and interpretation of data to assist in making more effective and reasonable decisions.

- Example: Child malnutrition status, Monthly expenditure of citizens of a city, Relationship of crime with space and time, Number of active users in a day of a website, average lifetime of the people of a country etc.

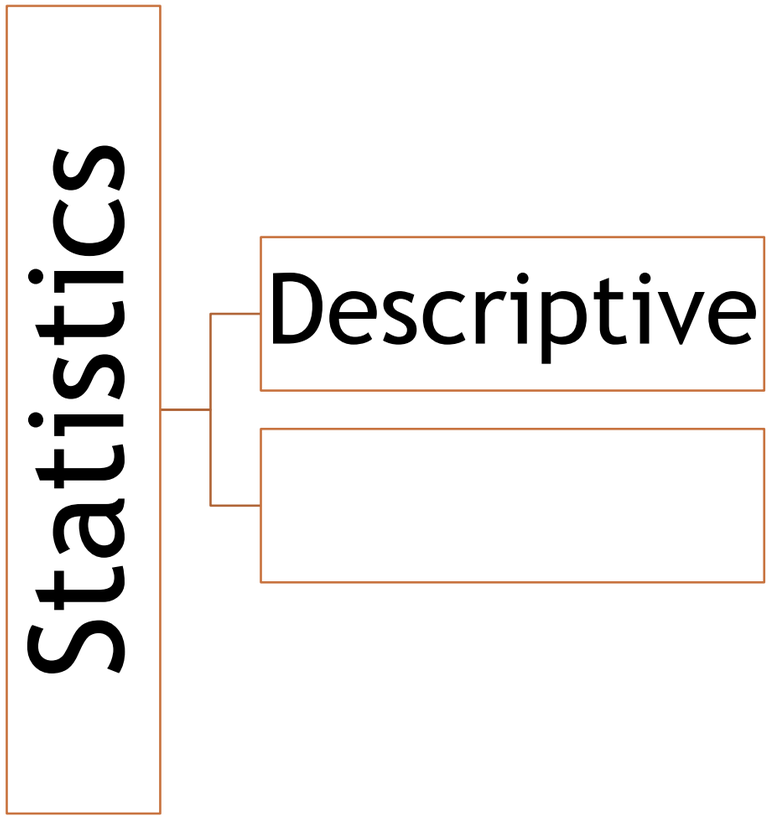
Statistics

- Data Collection
- Data Organization
- Data Presentation
- Data Analysis
- Data Interpretation/Conclusion

Types of Statistics

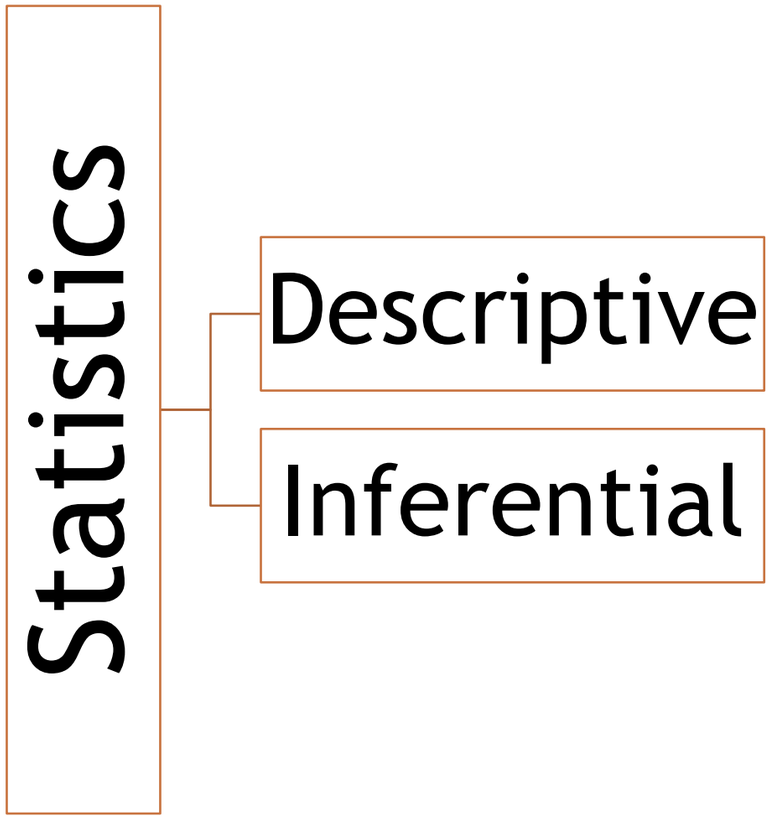


Types of Statistics



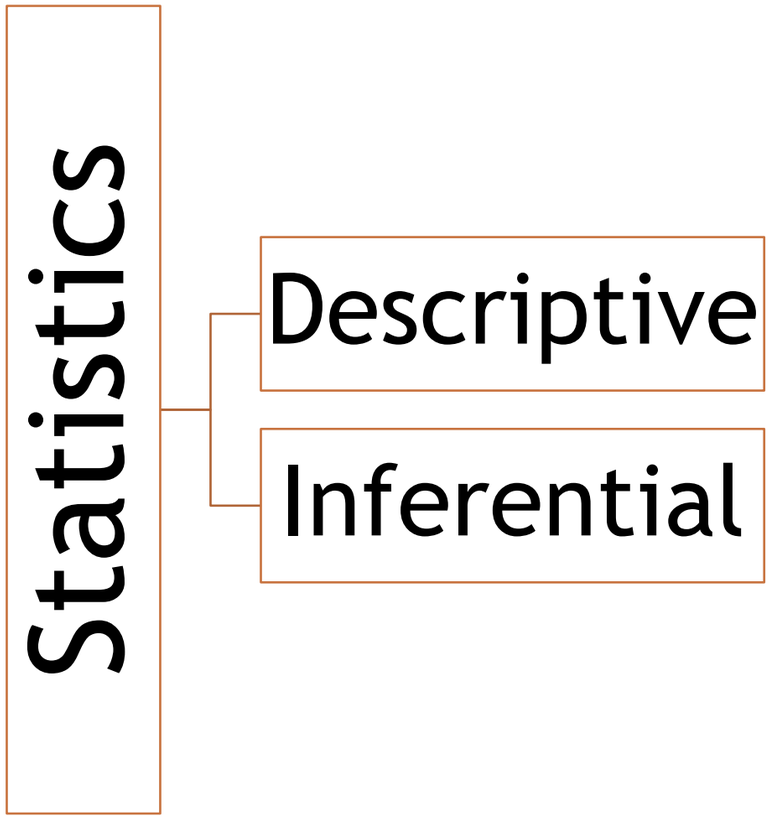
Which used to summarize, organize, and present a set of data/observations in a meaningful way (e.g., tables, graphs, numerical summaries).

Types of Statistics



Work with sample data to make a conclusion about the population.

Types of Statistics

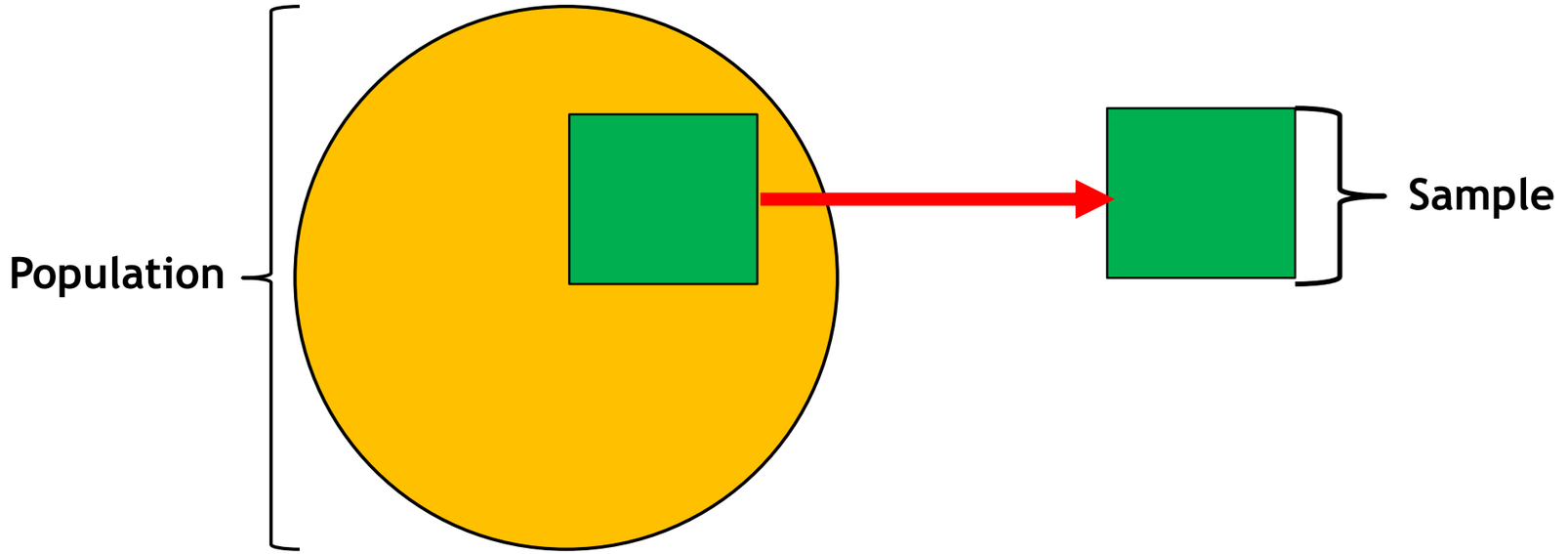


Work with sample data to make a conclusion about the population.

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

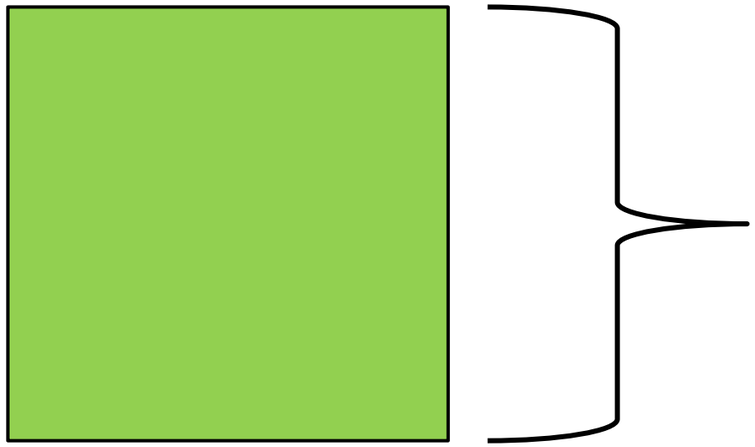
Descriptive Statistics	Inferential Statistics
Describe and summarize the main characteristics of Data	Make conclusion about population based on sample data
Applicable to both populations and samples	Applicable to only for samples
Uses measures of central tendency, measures of dispersion, and graphical 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”

Census & Survey

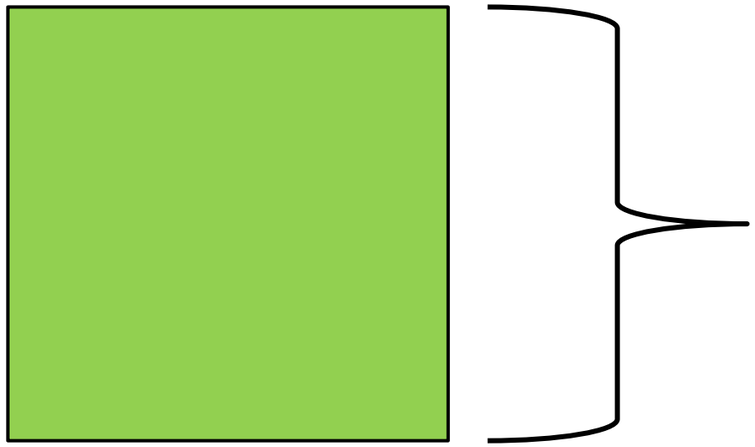
- To collect data about an entire population is called “Census”.
- To collect data from the part of the population is called “Survey”



Collects data from this entire box is “???”

Census & Survey

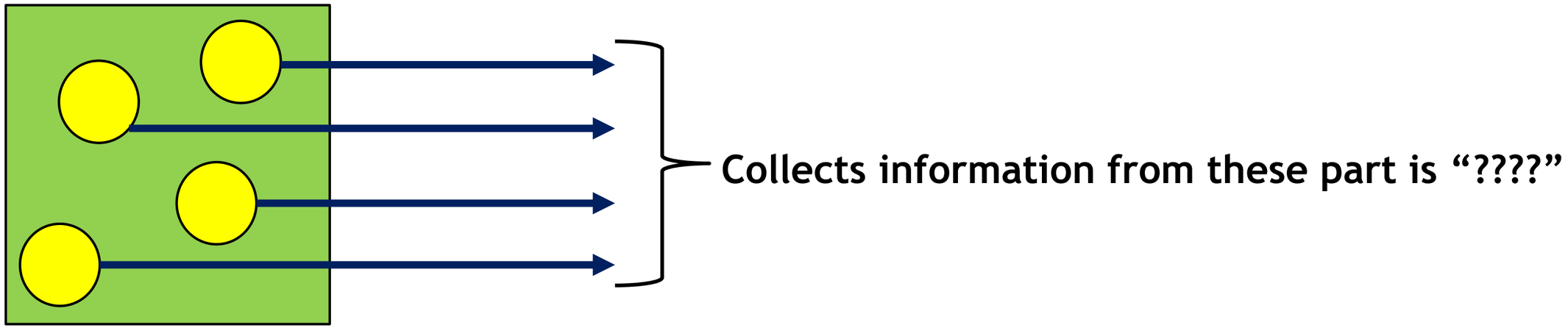
- To collect data about an entire population is called “Census”.
- To collect data from the part of the population is called “Survey”



Collects data from this entire box is “Census”

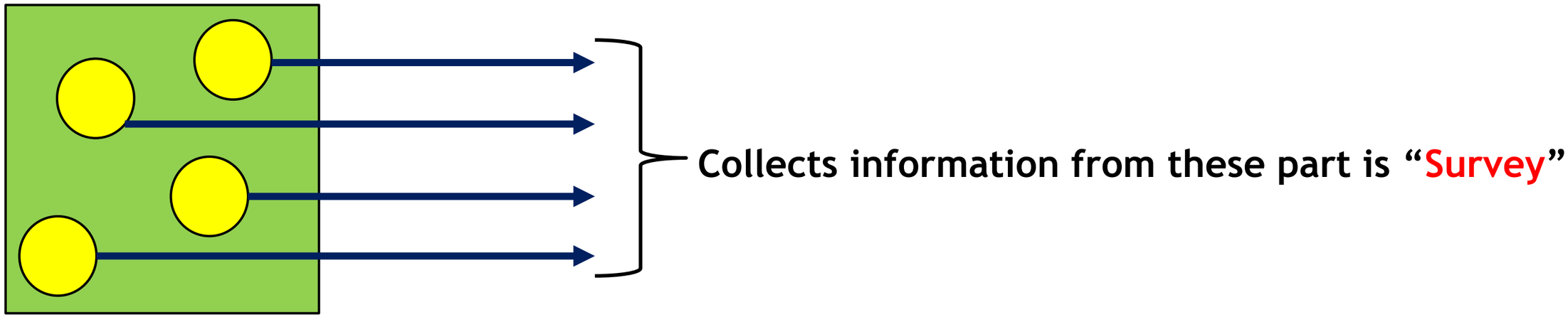
Census & Survey

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



Census & Survey

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



Parameter & Statistic

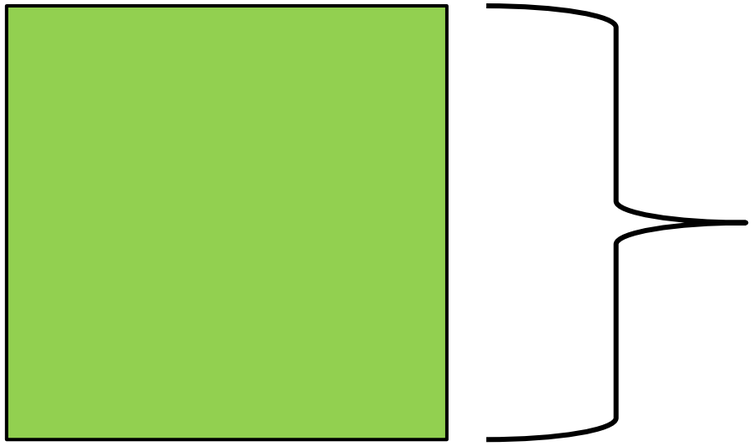
- A parameter is a numerical value that describes a characteristic (mean, variance etc.) of a population.

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

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.

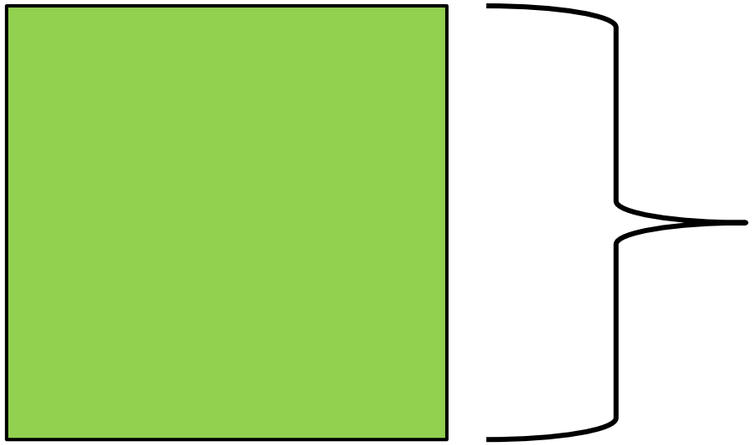


Characteristics of this entire box is “???”

Parameter & 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.

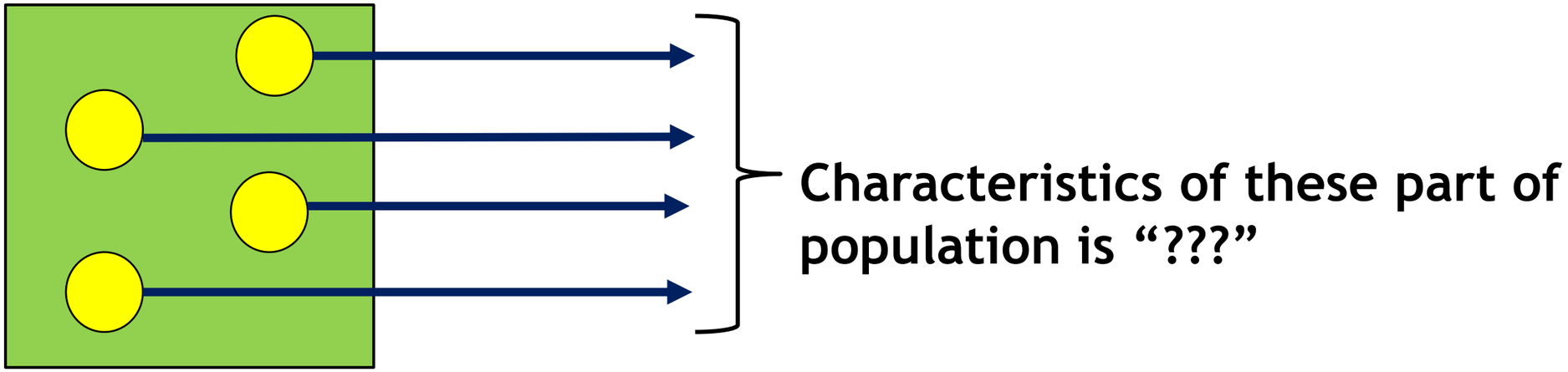


Characteristics of this entire box is "Parameter"

Parameter & Statistic

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

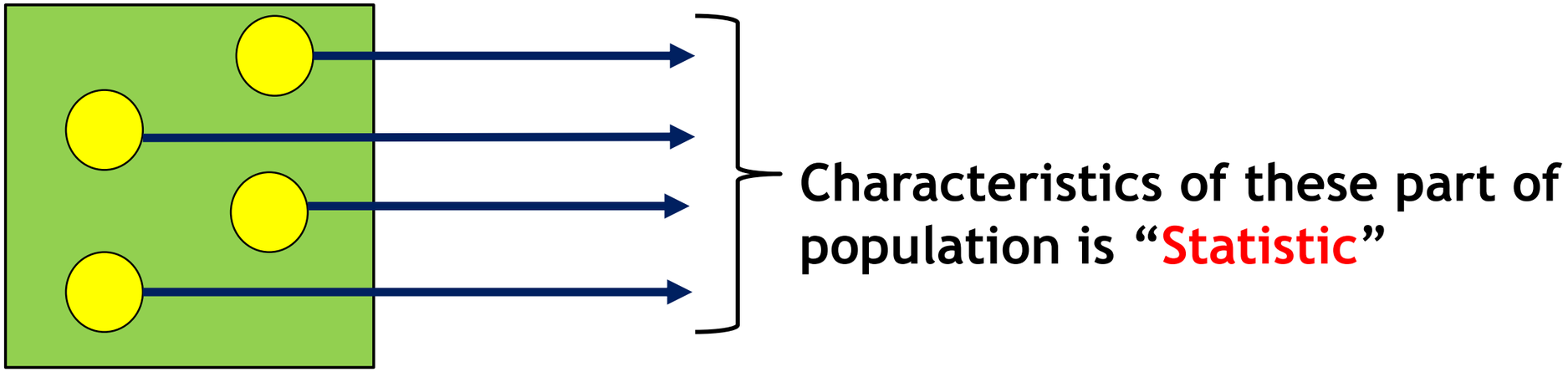
- A statistic is a numerical value that describes a characteristic (mean, variance etc.) of a sample.



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.



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)

Sources of Data

- There are two sources of getting statistical data:

1. Primary data (Fresh and First time)

2. Secondary data (Has already been collected by someone)

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

Scales of Measurements

Interval	Ratio
Must be quantitative	Must be quantitative
Can be ranked	Can be ranked
Zero is not absolute	Zero is absolute
Can be find difference but not ratios	Can be find both difference and ratios

Scales of Measurements

Levels	Property			Example
	Order	Difference	Ratio	
Nominal				Gender
Ordinal				Wealth index
Interval				Temperature
Ratio				Person’s age

Scales of Measurements

Levels	Property			Example
	Order	Difference	Ratio	
Nominal	No	No	No	Gender
Ordinal	Yes	No	No	Wealth index
Interval	Yes	Yes	No	Temperature
Ratio	Yes	Yes	Yes	Person’s age

Scales of Measurements

- Identify the scale of measurement for a variable that measures a person's level of education as "High School," "Bachelor's Degree," "Master's Degree," and "Ph.D".

Problem!!

- In a health survey, a researcher wants to find the hemoglobin level of the students of BRACU. For this purpose, 400 students were randomly selected and their hemoglobin levels were recorded.

I. What is the population?
II. What is the sample?
III. What is the variable being measured?
IV. What is the nature of variable?

Solution

i. Population: The population in this scenario is the entire group of students at BRACU (BRAC University).

ii. Sample: The sample is the subset of students from BRACU who were randomly selected and had their hemoglobin levels recorded. In this case, the sample size is 400 students.

iii. Variable being measured: The variable being measured is the "hemoglobin level" of the students.

iv. Nature of variable: The hemoglobin level is a continuous numerical variable. It can take any value within a certain range and can be measured with precision, making it a quantitative variable and Ratio level of measurement.

Problem!!

- In a customer satisfaction survey, a company wants to measure the waiting time experienced by customers in their service centers. They randomly select 200 customers and record their waiting times.

I. What is the population?
II. What is the sample?
III. What is the variable being measured?
IV. What is the nature of variable?


ANY
QUESTIONS
?