# 1. INTRODUCTION TO STATISTICS

*BY SLIIT MATHEMATICS UNIT*

*FACULTY OF HUMANITIES AND SCIENCES*




STATISTICIANS ARE MEAN AND SLIGHTLY DEVIANT


"I can prove it or disprove it! What do you want me to do?"

- Statistics is the ***study*** of ***uncertainty***.

- We need statistics to identify the ***variability*** in data.

# APPLICATIONS OF STATISTICS

- Statistics can be applied in any field. Following are some examples for such applications.
  - ***Engineering and Sciences***
  - ***Medical Sciences***
  - ***Education***
  - ***Business Analytics***
  - ***Social Sciences***
  - ***Machine Learning***
  - ***Quality Control***
  - ***Actuarial Sciences etc.***

- You can analyze data by using some statistical package.
- It allows you to analyze data easily and precisely.
- Most commonly used statistical packages are ***SPSS, SAS, Minitab, R, E-views and Matlab.***

# TERMINOLOGY

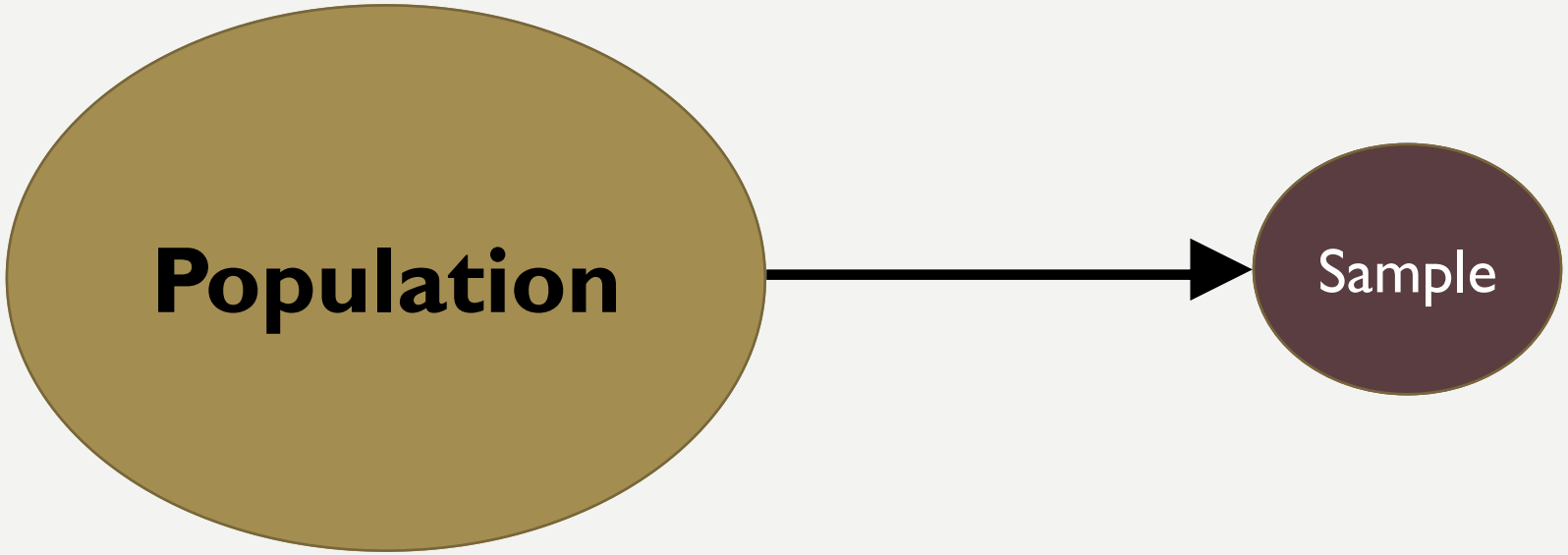
# DEFINITION - POPULATION

- This is a collection of set of individuals or objects where researcher is interested about drawing inferences.
- Population can be finite or infinite.
- If you are going to collect data from all the individuals in the population, then it is known as a ***census survey***.

# DEFINITION - SAMPLE

- A ***sub set*** of the population.

- If you are going to collect data from a part of the population (sample), then it is known as a ***sample survey.***



# DEFINITION - VARIABLE

- Variable is a ***characteristic/property*** of each individual in the population or a sample.

- ***Examples*** :- Age, Gender, Temperature etc.

- We usually use capital letters to denote variables.

# DEFINITION – DATA (SINGULAR)

- The value of the variable associated with one element of a population or sample.

- This value may be a number, a word, or a symbol.

# DEFINITION - PARAMETER

- Parameter is a ***summary characteristic*** about the individuals in the ***population***.

- Parameter is always related with the population.

- ***Examples*** :- Population mean (μ), Population variance (σ^2), Population proportion (P) etc.

# DEFINITION - STATISTIC

- Statistic is a ***summary characteristic*** about the individuals in the ***sample***.

- Statistic is always related with the sample.

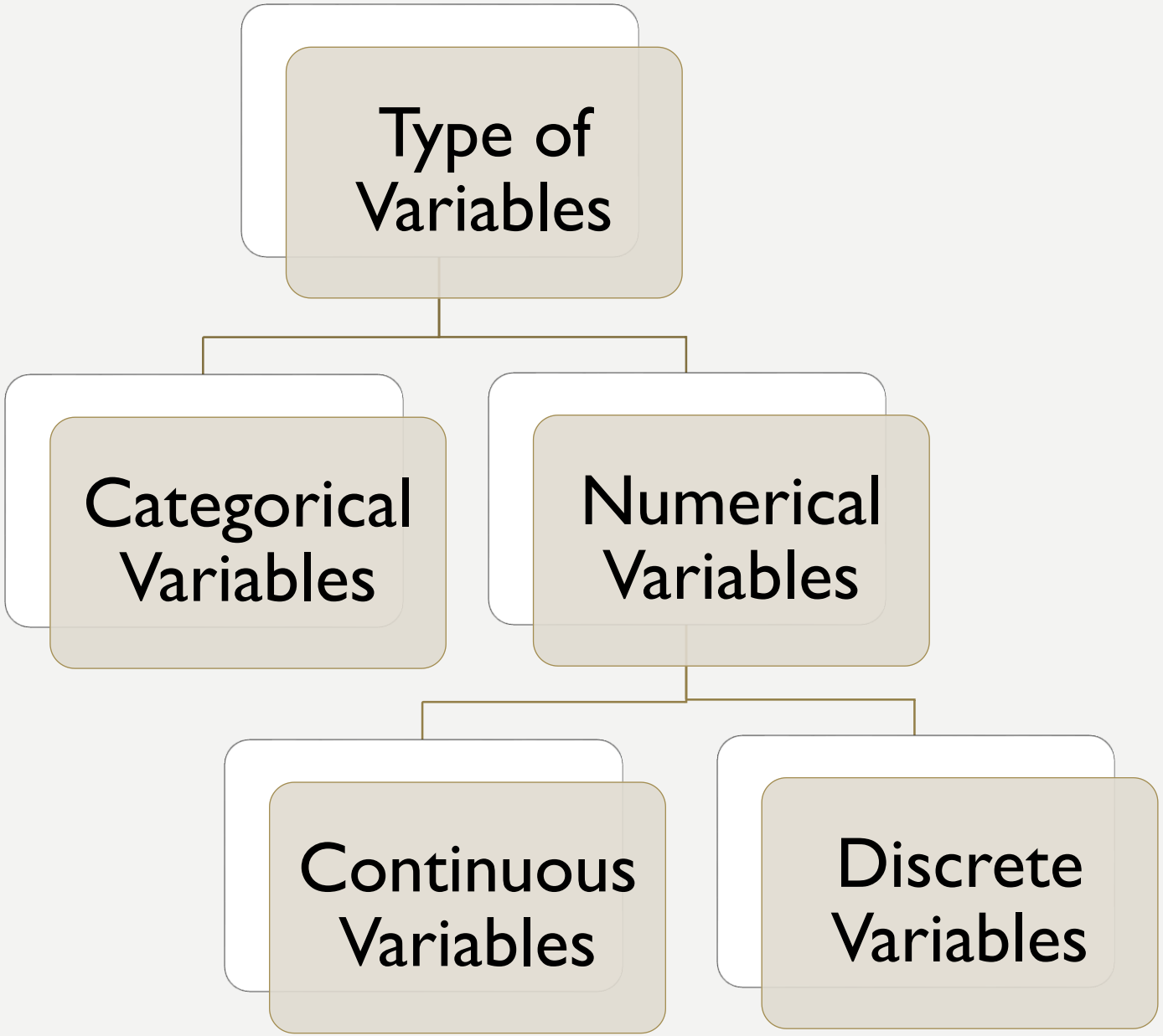
- ***Examples*** :- Sample mean ($\bar{x}$), sample variance (s^2), sample proportion (p) etc.

# EXAMPLES…

- A researcher is interested in finding the average weight of a first year student in SLIIT. He collected data from all first year students in computing faculty.

| | | |
|---|---|---|
| ***Population*** | : | All the first year students in SLIIT |
| ***Sample*** | : | All first year students in computing faculty |

# TYPES OF VARIABLES


Type of Variables
Categorical Variables
Numerical Variables
Continuous Variables
Discrete Variables

- **Qualitative/Categorical Variables** :

  A variable that categorizes and describes an element. E.g. : Hair color, Gender, Marital status, Highest education qualification.

- **Quantitative/Numerical Variables** :

  A variable that quantifies an element. E.g. : Marks for statistics, Age, Temperature, Time taken to travel to SLIIT from home.

# Numerical Variables

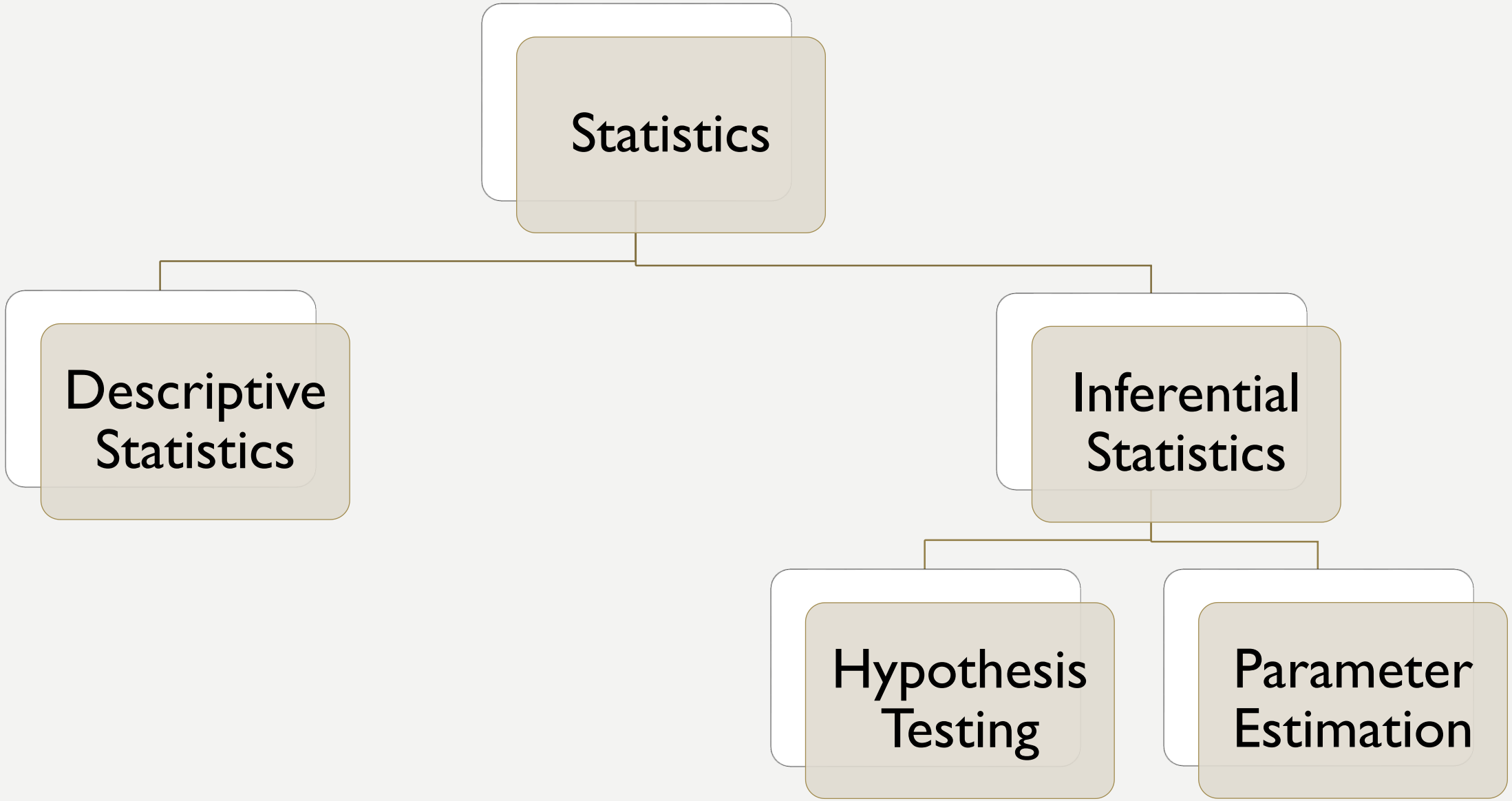
- **Discrete variables**: Distance between two values exists.

  E.g. : Age in years, No of children in a family, Number of accidents in a junction within an hour

- **Continuous variables**: This will contain any value within a given range.

  E.g.:- Temperature, Heart beat of a patient etc.

# AREAS OF STATISTICS


Statistics
Descriptive Statistics
Inferential Statistics
Hypothesis Testing
Parameter Estimation

- ***Descriptive Statistics*** :- This is also known as ***preliminary analysis*** / ***explanatory analysis***. This will give you a rough idea about the ***behavior of data***. It describes how the each of the variables behave. There are ***two methods*** that you can use under descriptive statistics. They are,
  - ***Graphical Methods***
  - ***Numerical Methods***
- ***Inferential Statistics*** :- This is ***drawing conclusions*** about population parameters by using sample statistics. Under this there are two main areas namely, ***parameter estimation*** and ***hypothesis testing***.

1) Determine whether the following variables are categorical or numerical

a) Color of a randomly selected beetle
b) Birth weight
c) Hometown
d) Life of a battery measured in hours
e) Gender
f) Hair color
g) Rainfall in mm
h) Grade for Maths
i) Method of payment (Cash, Cheque, Debit card & Credit card)
j) Eye color of children

2) To determine if the class understood the homework assignment, the math teacher checks the top 3 papers in the pile of collected homework. The teacher finds that all students understood the homework assignment.

   a) Identify population
   b) Identify sample
   c) Identify variable
   d) Identify type of the variable

# THANK YOU!

**Any questions?**