1. INTRODUCTION TO STATISTICS [IT2110]

By Department of Mathematics and Statistics,

Faculty of Humanities and Sciences, SLIIT

STATISTICIANS ARE MEAN AND SLIGHTLY DEVIANT

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

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.***

Definition - Statistics

- Statistics is the study of the ***collection***, ***analysis***, ***interpretation***, ***presentation***, and ***organization*** of data.

-Oxford: A Dictionary of Statistics-

- Statistics are numbers that ***summarize*** raw facts and figures in some ***meaningful*** way.

- Head First Statistics -

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

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

TERMINOLOGIES

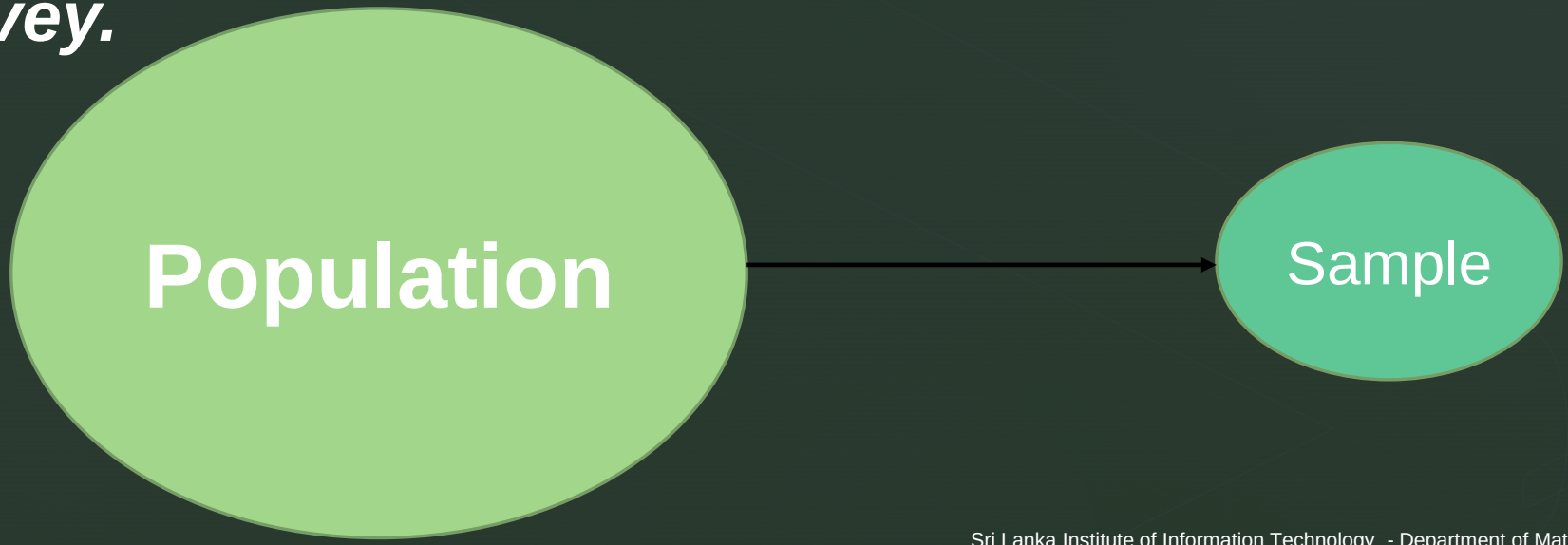
Definition - Population

- 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***.

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.

Definition - Experiment

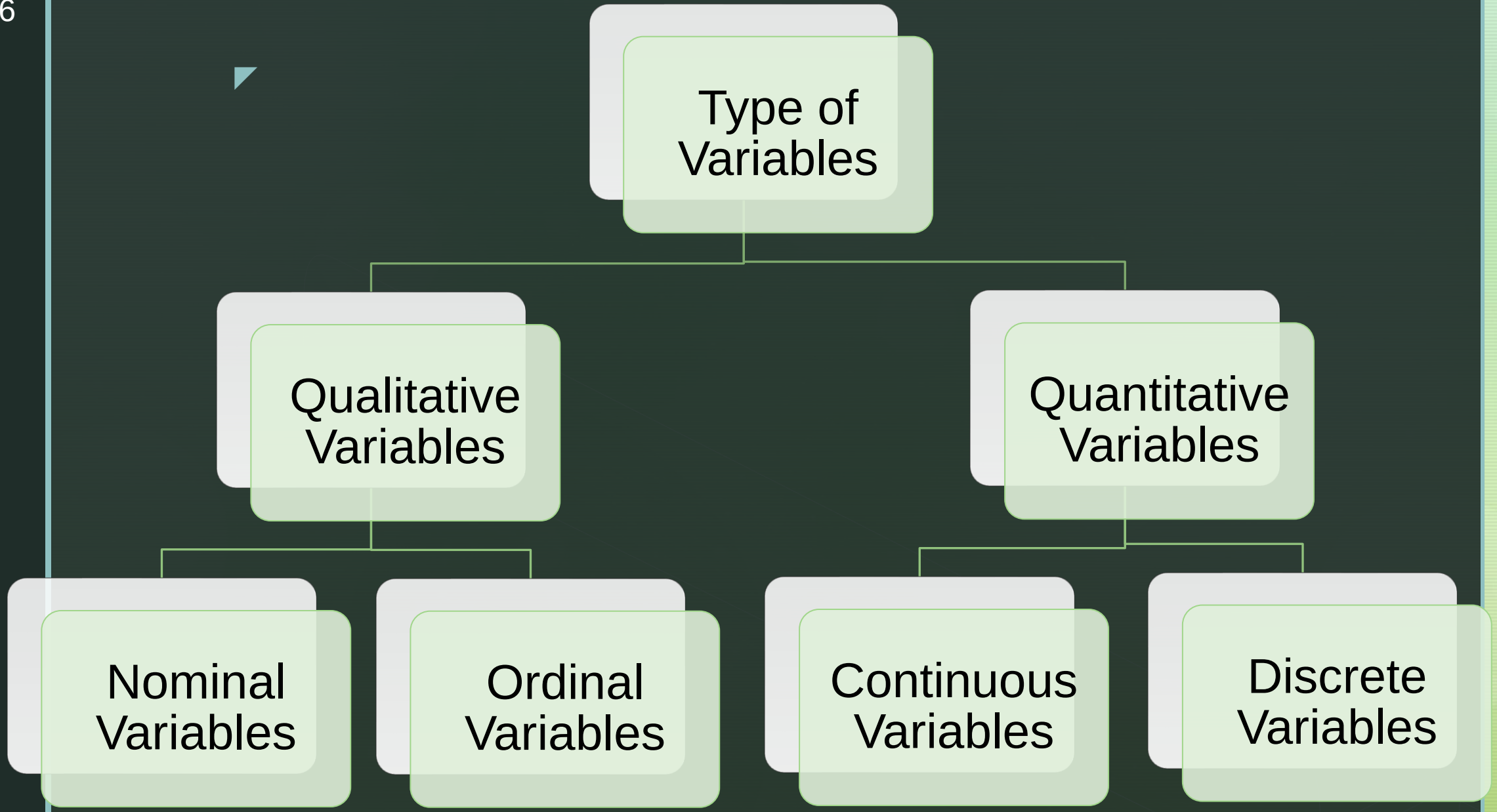
- An experiment is a planned activity whose results yield a set of data.

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
Variable	:	Weight
Summary Characteristic	:	Average Weight ➡ **Statistic**
Type of survey	:	Sample survey

TYPES OF VARIABLES


Type of Variables
Qualitative Variables
Quantitative Variables
Nominal Variables
Ordinal Variables
Continuous Variables
Discrete Variables

- **Qualitative/Attribute/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.

Qualitative Variables

- **Nominal Variables**: Categories are not naturally ordered.

  E.g. : Gender, Hair Color, Marital Status

- **Ordinal Variables**: Categories are naturally ordered.

  E.g. : Satisfaction Rating, Pain Severity, Highest education qualification.

Quantitative Variables

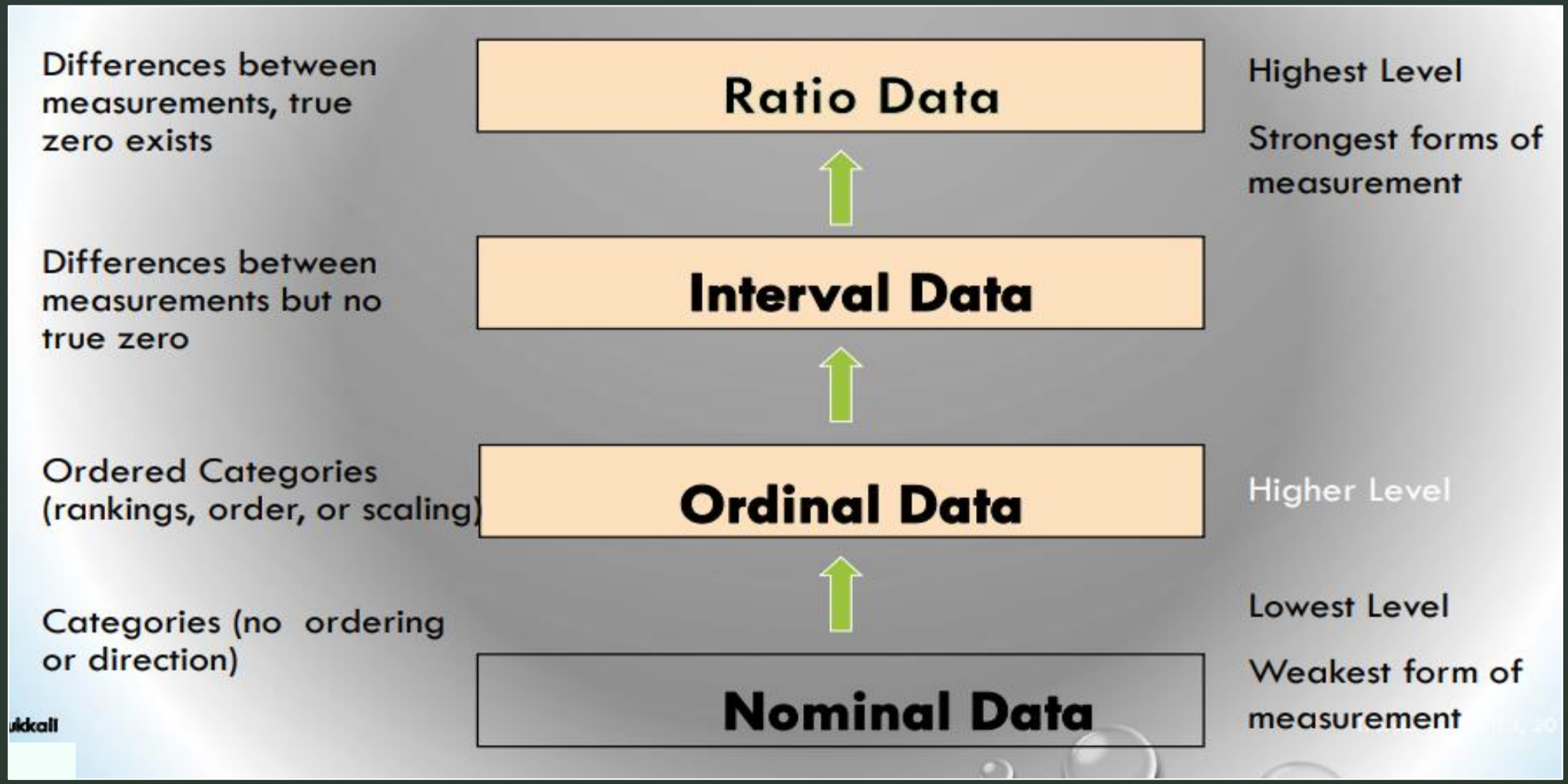
- **Discrete variables**: Distance between two values exists.

  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.

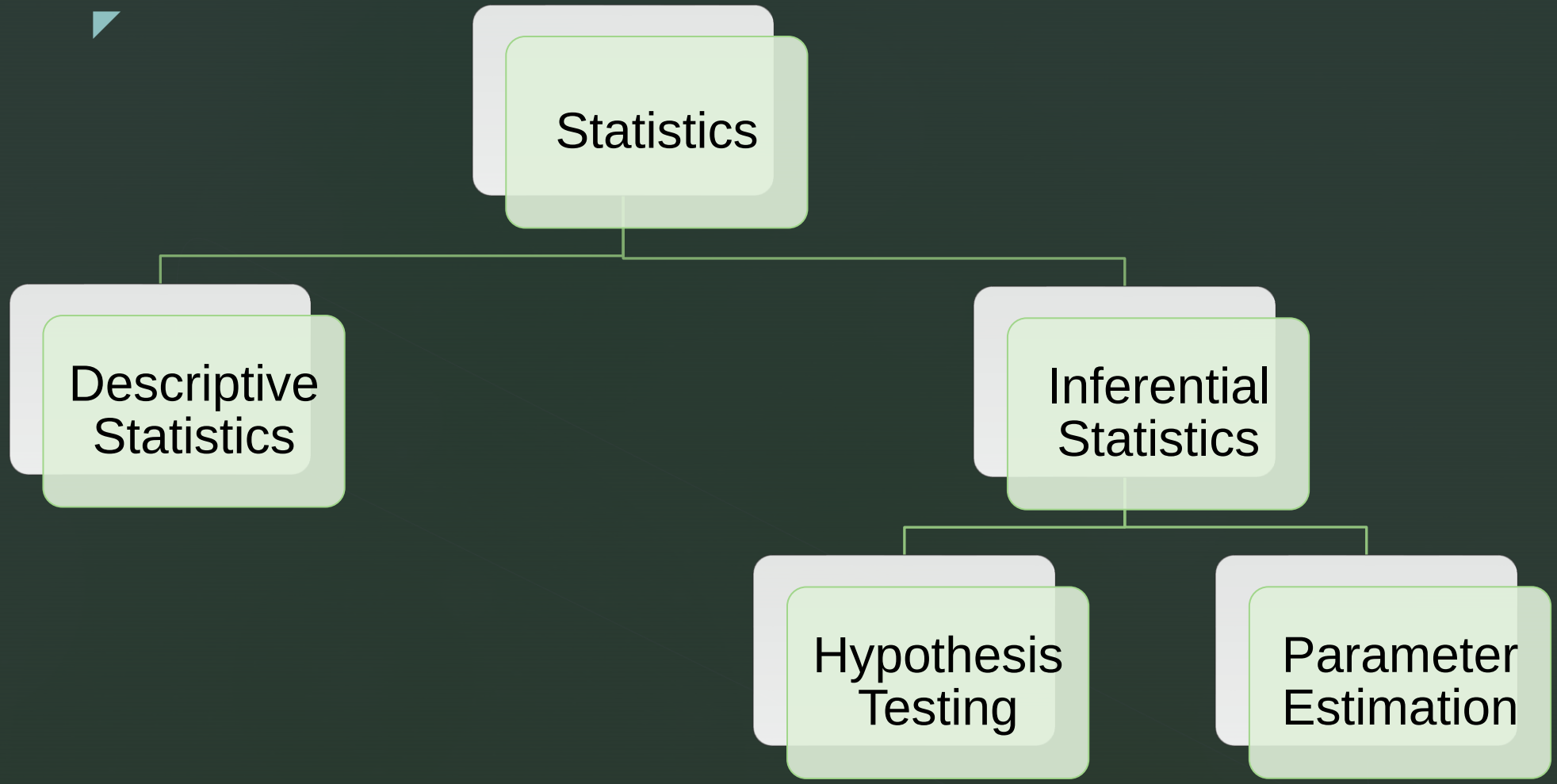
Measurement Scales



Interval Scale Vs. Ratio Scale

Interval Scale	Ratio Scale
In this scale, variables can be added and subtracted. But ratio and multiplication is not possible.	Including ratio and multiplication of variables it has all characteristics of an interval scale.
Can calculate mean, median and mode.	Can calculate mean, median and mode.
Difference between variables can be evaluated.	Difference between variables can be evaluated.
Does not have a true zero point. (Eg:- Temperature can be below zero degree Celsius and negative)	True zero point exist. (Eg:- Weight can not be zero or below zero)
Examples:- Temperature in Celsius, Temperature in Farenhite, pH Value	Examples:- Height, Weight, Temperature in Kelvin, No of sales, Income of an individual, Heart Rate

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***.

- 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.***
- In this module we will discuss how to analyze data by using ***R***.

INTRODUCTION

R Software

- Independent and Open source.
- Initially developed at University of Auckland in the mid1990s.
- Distributed under the GNU open software license.
- Developed by the user community.
- Available On: Linux, Windows and Mac.
- Latest Version: 4.4.2 - Released in 2024.
- Terminal and GUI available.
- IDEs for R: R Studio, Rattle.

THANK YOU!

Any Questions?