Probability & Statistics (IT2110)

LECTURE 01

Module Introduction

Introduction

Course Delivery

- Lectures – 2 hours
- Tutorials – 1 hour
- Labs – 2 hours

Courseweb Enrollment Key : IT2110

Module Outline – Please refer courseweb

Course Evaluation

Continuous Assessments – 50%

- Midterm Examination – 30%
- Assignment – 20%

Final Examination – 50%

Contact Information

Ms. Malika Lakmali (Lecturer–in-charge) – malika.l@sliit.lk

Ms. Nilushi Dias – nilushi.d@sliit.lk

Course Content

- Sampling Methods
- Exploratory Data Analysis / Descriptive Statistics
- Probability
- Random Variables (R.V.s) and Probability Distributions
- Continuous Probability Distributions
- Statistical Inference
- Chi-squared Test
- Regression
- Introduction to Time Series

References

Recommended Readings:

Advance Engineering Mathematics - HK Dass

The Exploration and Analysis of Data by Roxy Peck and Jay L Devore (e book)

Introductory Statistics by Prem S Mann (e book)

Introduction to Statistics

Introduction to Statistics

What is Statistics

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

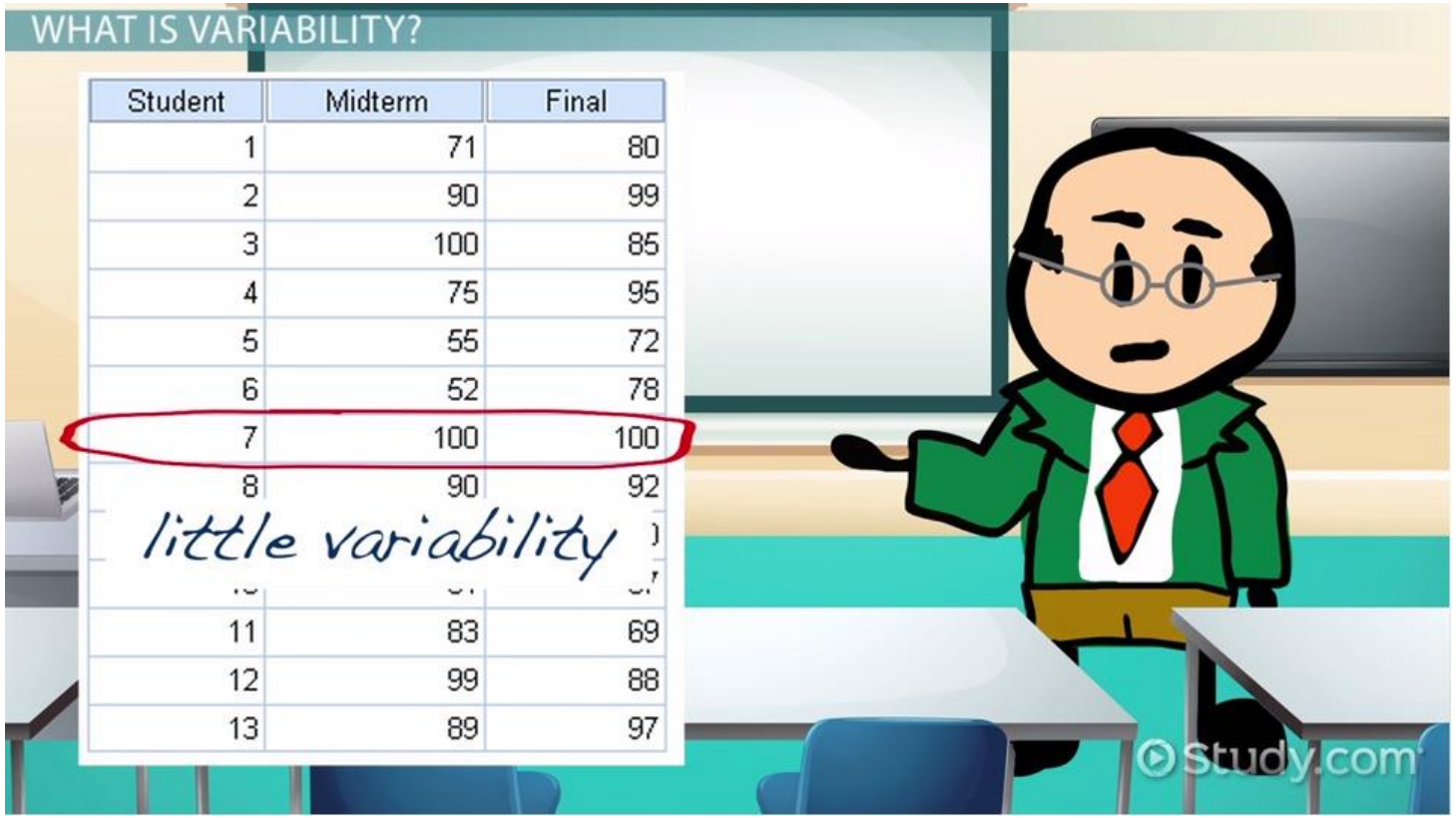
– Oxford : A Dictionary of Statistics-

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

- Head First Statistics -

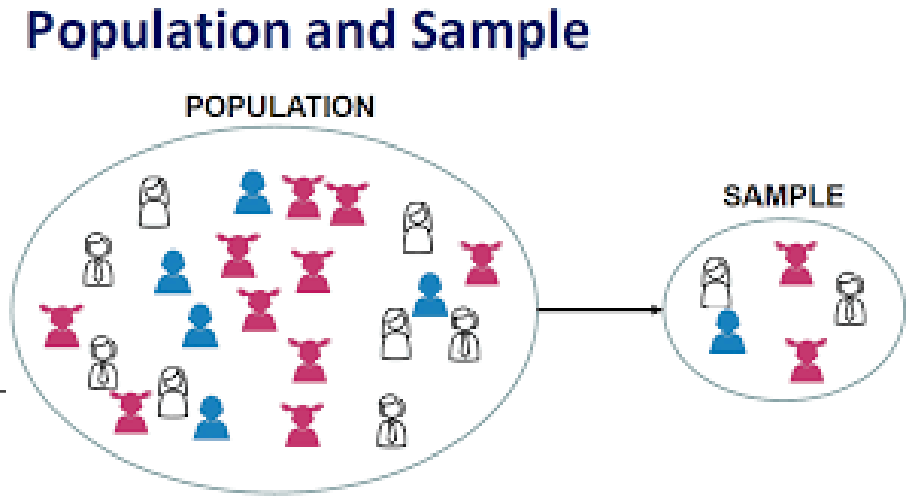
What is Statistics

- Statistics is the study of uncertainty
- Statistics can identify variability in data



Terminology

Population vs Sample



POPULATION

Collection of all items of interest to our study

Population Size – N

Parameters

Surveys done with the whole population is Census Survey

SAMPLE

A subset of the population

Sample Size – n

Statistics

Surveys done with the Sample is a sample Survey

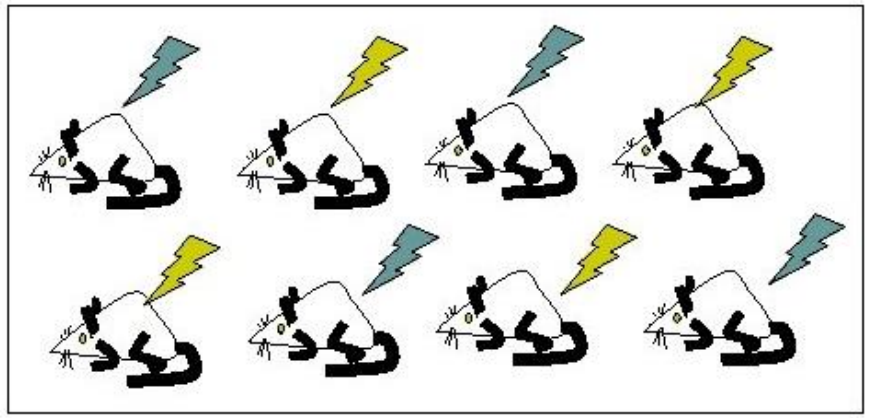
Population Vs Sample

Terminology

Experiment

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

The animal as the experimental unit



N=8
n=4

Animals individually treated. May be individually housed or grouped

Terminology

Variable

Variable is a characteristic/property of each individual in the population or a sample.
Examples :- Age, Gender,Temperature etc.

Capital letters are used to denote variables.

Terminology

Data

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.

Parameter vs Statistics

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.

STATISTICS

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.

Example

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

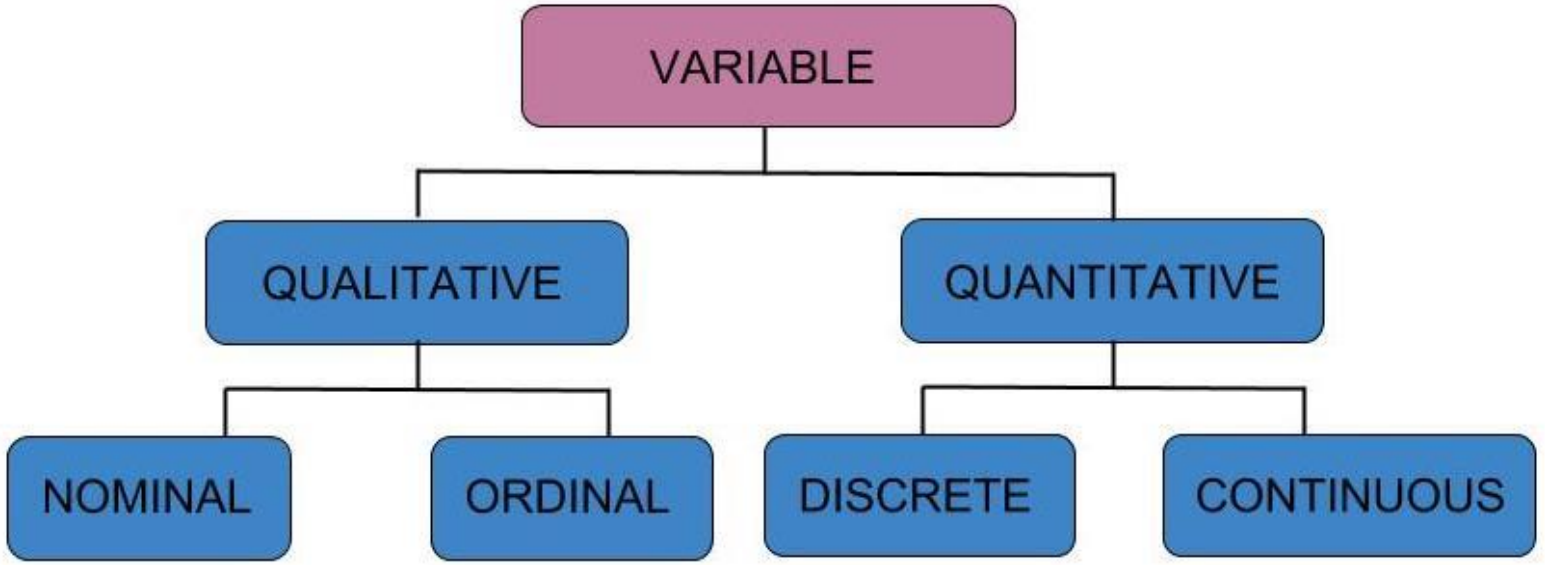
Sample : All first year students in computing faculty

Variable : Weight

Summary Characteristic : Average Weight

Type of survey : Sample survey

Types of Variables



Types of Variables

Variables

QUALITATIVE/CATEGORICAL

A variable that categorizes and describes an element.

E.g. : Hair color, Gender, Marital status, Highest education qualification.

QUANTITATIVE/NUMERICAL

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.

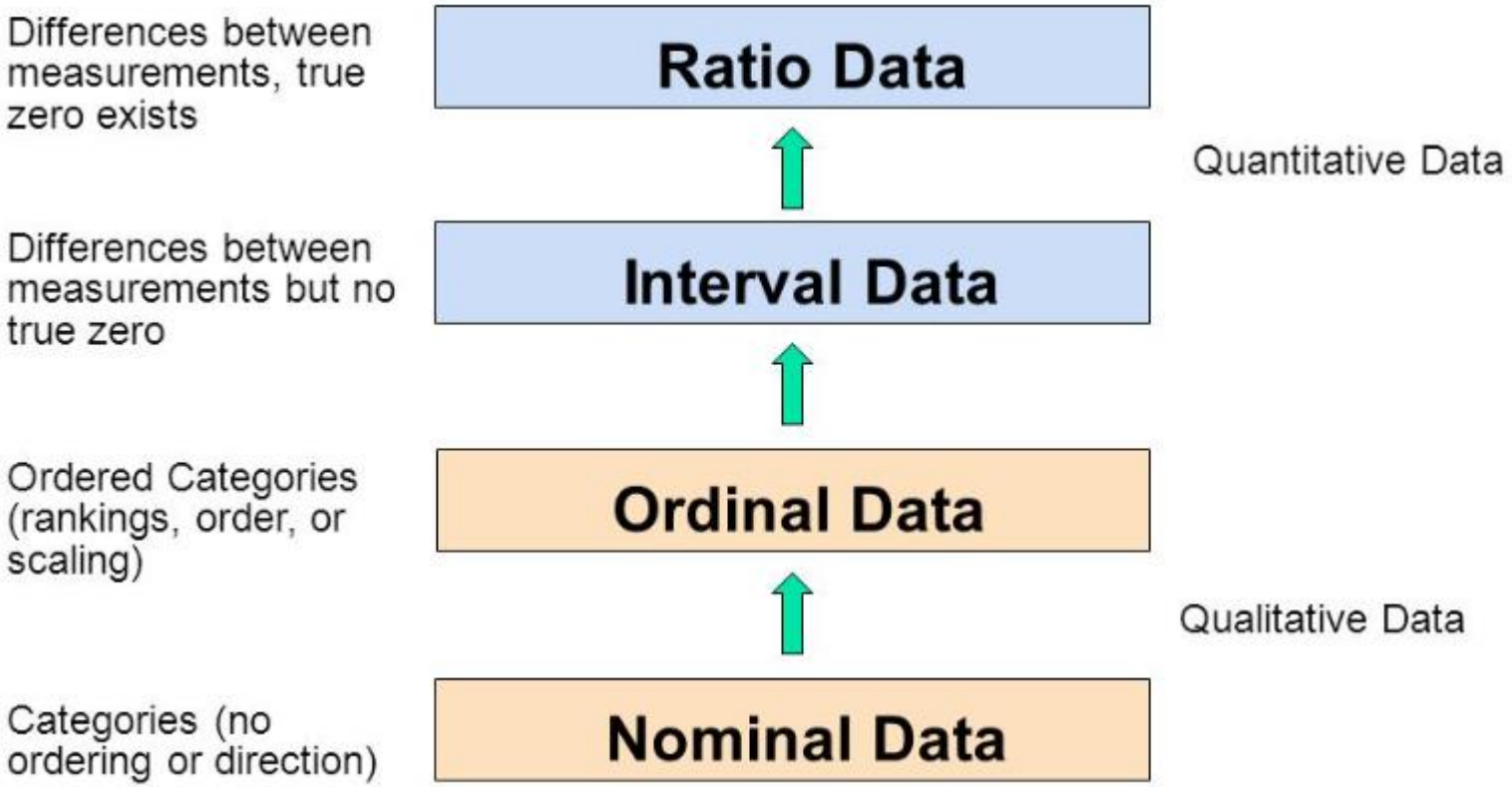
E.g. : Age in years, No of children in a family, Number of accidents in a junction within an hour

Continuous variables:

This will contain any value within a given range.

E.g.:- Temperature, Heartbeat of a patient etc.

Measurement Scale



Measurement Scales

Scale	*Order*	*Distance*	*True Zero*	*Examples*
Nominal	no	no	no	Color, Gender, Ethnicity, Country
Ordinal	yes	no	no	Rating scales, Rank orders
Interval	yes	yes	no	Time of day, Year, IQ, Likert scales
Ratio	yes	yes	yes	Age, Height, Weight, Rates

Why?

Interval Vs Ratio Scale

INTERVAL SCALE

In this scale, variables can be added and subtracted. But ratio and multiplication is not possible.

Can calculate mean, median and mode.

Difference between variables can be evaluated.

Does not have a true zero point. (Eg:- Temperature can be below zero degree Celsius and negative)

Examples:- Temperature in Celsius, Temperature in Farenhite, pH Value

RATIO SCALE

Including ratio and multiplication of variables it has all characteristics of an interval scale.

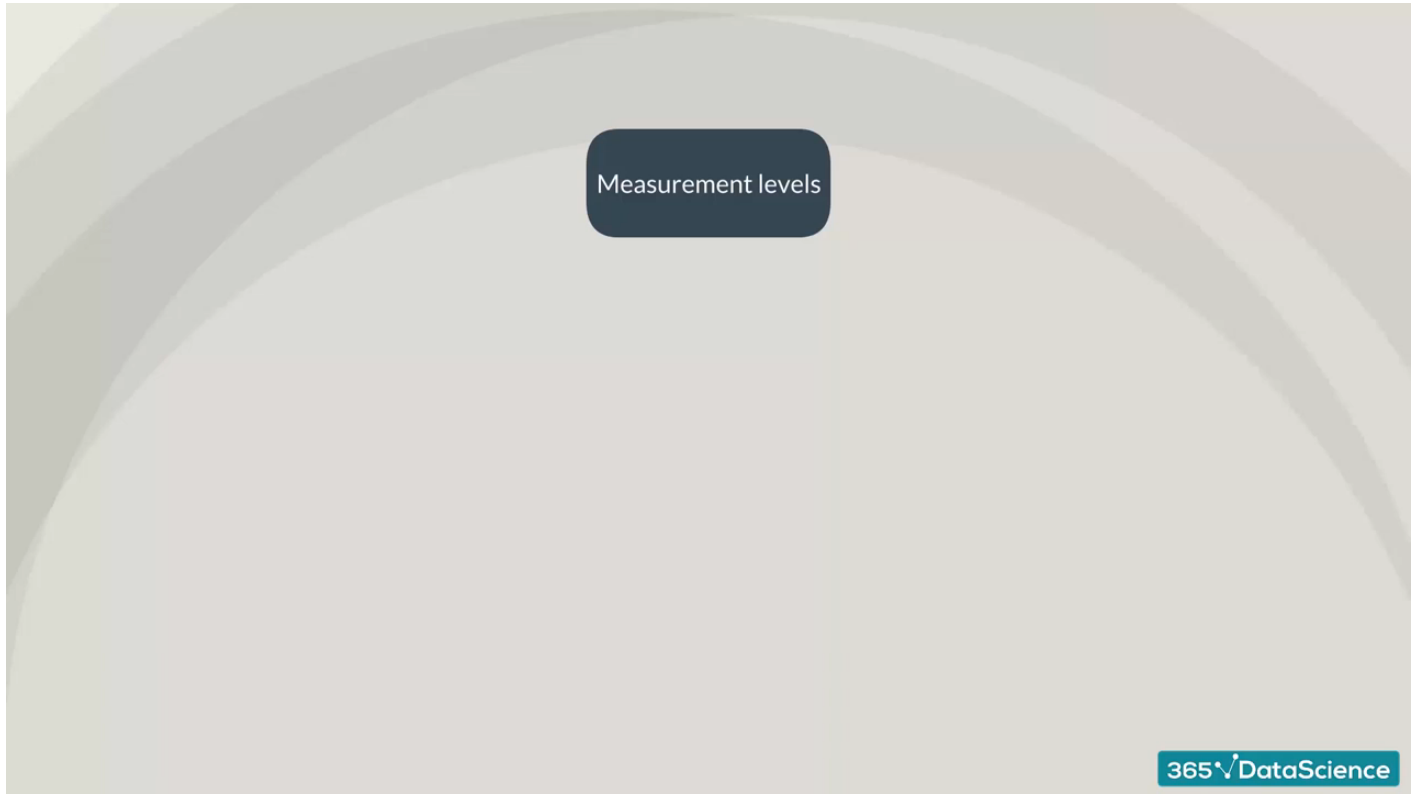
Can calculate mean, median and mode.

Difference between variables can be evaluated.

True zero point exist. (Eg:- Weight can not be zero or below zero)

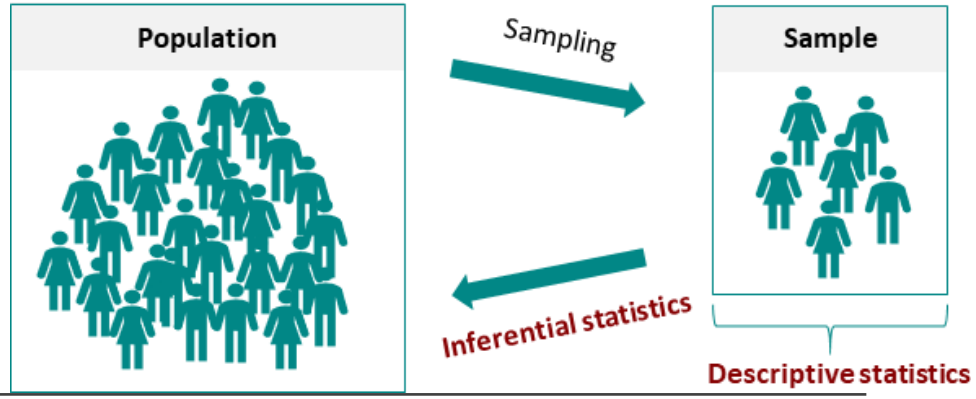
Examples:- Height, Weight, Temperature in Kelvin, No of sales, Income of an individual, Heart Rate

Types of Variables



Areas of Statistics

Areas of Statistics



INFERENTIAL STATISTICS

Inferences (conclusions) are made on the population based on the sample data

- Hypothesis Testing
- Parameter Estimation

DESCRIPTIVE STATISTICS

Explores sample data or describes data

There are two methods

- Numerical Methods
- Graphical Methods

Areas of Statistics

Statistical Packages

Statistical Packages

To analyze data, statistical packages are used.

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 to R

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 (OS X).

Latest Version: 3.6.2 (Dark and Stormy Night) - Released 12/12/2019.

Terminal and GUI available.

IDEs for R: R Studio, Rattle.

Thank You

QUESTIONS?