CM 1110 Fundamentals of Mathematics and Statistics

Dr. Priyanga D. Talagala

2020-01-31

Contents

Course Syllabus

Pre-requisites

None

Learning Outcomes

On successful completion of this module, students will be able to apply fundamental concepts in Mathematics and Statistics for real world problem solving.

Outline Syllabus

- Number Systems
- Sequence and Series
- Introduction to Logic
- Boolean Algebra
- Differentiation and Integration
- Descriptive Statistics
- Sets and Relations
- Probability
- Correlation and Regression

Method of Assessment

- Mid-semester examination
- End-semester examination

0.1 Lecturer

Dr. Priyanga Dilini Talagala

Schedule

Lectures: TBA

Consultation times: TBA

Chapter 1

Number Systems

Chapter 2

Sequence and Series

Chapter 3

Introduction to Logic

Chapter 4

Boolean Algebra

Chapter 5

Differentiation and Integration

Chapter 6

Descriptive Statistics

6.1 Introduction to Statistics

6.1.1 Some Basic Terminologies Used in Statistics

i Population

- The set of **all** possible elements in the universe of interest to the researcher

ii Sample

- A Sample is a **subset** (a portion or part) of the population of interest
- The sample must be a representative of the population of interest

iii Element

- Element is an **entity or object** which the information is collected.
- *Eg: Student, household, farm, company, tomato plant*

iv Variable

- A variable is **a feature characteristic which has different 'values' or categories for different elements** (items/subjects/individuals)
- *Eg: Gender of client, brand of mobile phones, risk level, number of emails received per day, age of client, income of client*

v Data

- Data are **measurements or facts** that are collected from a statistical unit/entity of interest
- We collect data on variables
- Data are raw numbers or facts that must be processed (analysed) to get useful information.
- We get information from data.
- *Eg:*

Variable: *Age (in years) of client*

Data: *21, 45, 18, 32, 30, 22, 23, 27*

Information:

The mean age is 27.25 years

The minimum age is 18 years

The range of ages is 18-45

The percentage of clients below 25 years of age: 50%

vi Statistic

- **Characteristic** of a **sample**
- The value which calculated based on sample data

vii Parameter

- **Characteristic** of a **population**
- The value which calculated based on population data

viii Census

- When a researcher **gathers data from the whole population for a given measurement,** it is called a census

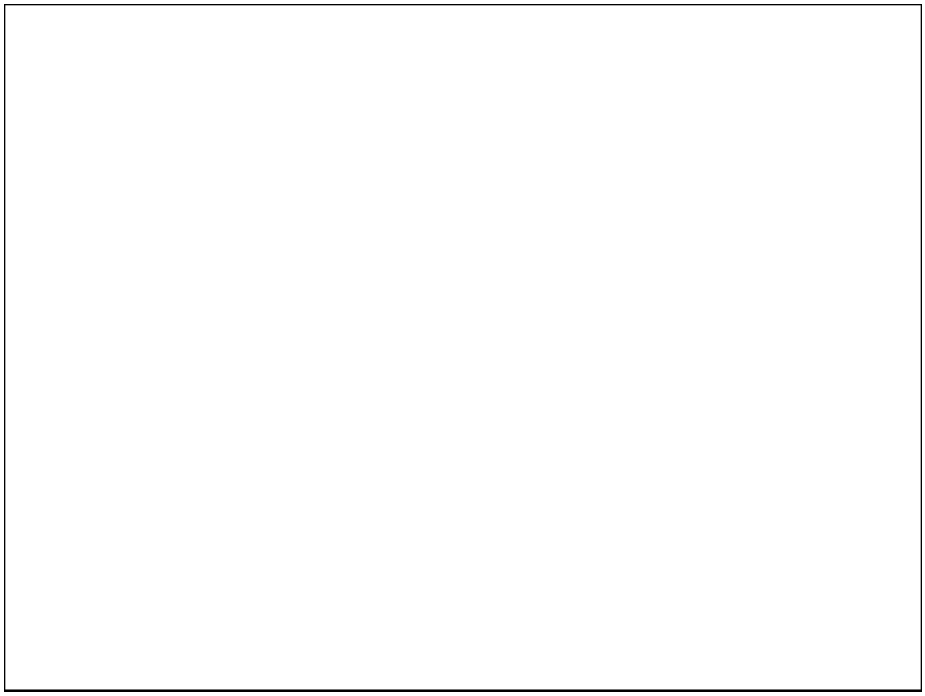
ix Sampling

- When a researcher **gathers data from a sample of the population for a given measurement,** it is called sampling
- The process of selecting a sample is also called sampling

Why take a sample instead of studying every member of the population ?

- Prohibitive cost of census
- Destruction of item being studied may be required
- Not possible to test or inspect all members of a population being studied.

6.1.2 Branches of Statistics

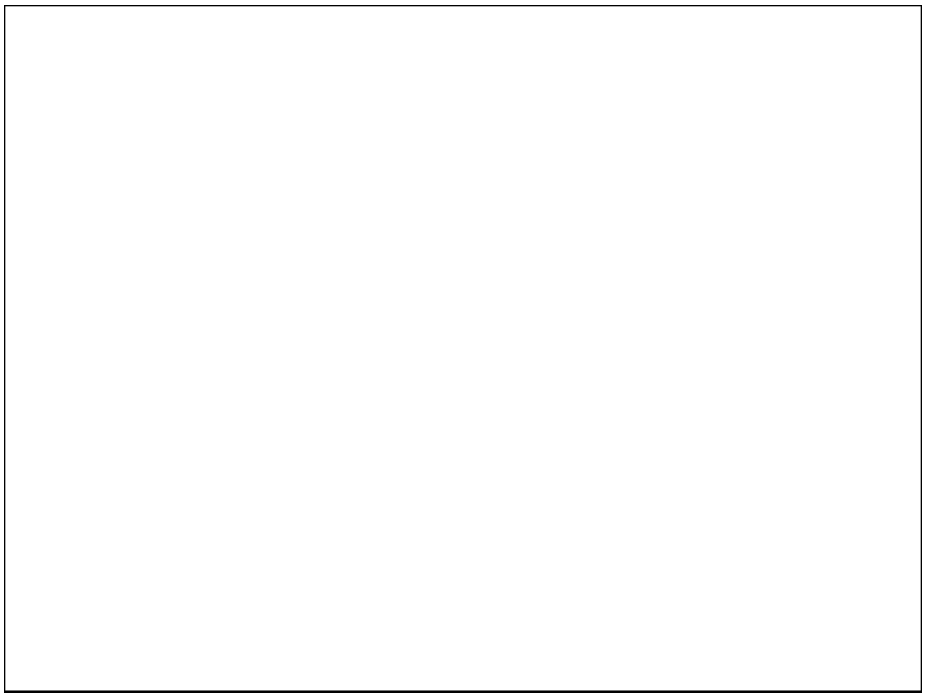


i Descriptive Statistics

- Descriptive statistics consists of organizing, summarizing and presenting data in an informative way.
- The main purpose of descriptive statistics is to provide an overview of the data collected.
- Descriptive statistics describes the data collected through frequency tables, graphs and summary measures (mean, variance, quartiles, etc.).

ii Inferential Statistics

- In inferential statistics sample data are used to draw inferences (i.e. derive conclusions) or make predictions about the populations from which the sample has been taken.
- This includes methods used to make decisions, estimates, predictions or generalizations about a population based on a sample.
- This includes point estimations, interval estimation, test of hypotheses, regression analysis, time series analysis, multivariate analysis, etc.



6.1.3 Types of Variables

6.1.3.1 Qualitative / Quantitative Variables

i Qualitative variable (Categorical variable)

- The characteristic is a quality.
- The data are categories.
- They cannot be given numerical values.
- However, it may be given a numerical label
- Qualitative variables are sometimes referred as categorical variables.
- *Eg:*

Gender:

Age group:

Education level:

A/L stream:

Degree type:

Hair colour:

FIT student batch:

Undergraduate level:

Grade that you can obtain for CM 1110/ CM1130

ii Quantitative variable

- The characteristic is a quantity
- The data are numbers
- Quantitative data require numeric values that indicate how much or how many.
- They are obtained by counting or measuring with some scale
- *Eg:*

Number of family members:

Number of emails received per day:

Weight of a student:

Age:

Credit balance in the SIM card:

Time remaining in class:

Temperature:

Marks

6.1.3.2 Discrete/ Continuous Variables

- Quantitative variables can be classified as either discrete or continuous.

i Discrete Variables

- Quantitative
- Usually the data are obtained by counting
- There are impossible values between any two possible values
- *Eg:*

Number of family members:

Number of emails received per day:

ii Continuous Variables

- Quantitative
- Usually, the data are obtained by measuring with a scale

- There are no impossible values between any two possible values.(any value between any two possible values is also a possible value)
- i.e a continuous variable can take any value within a specified range.
- *Eg:*

Weight of a student:

Age:

Credit balance in the SIM card:

Time remaining in class:

Temperature:

Marks

6.1.4 Scales of Measurements

- There are four levels of measurements called, **nominal, ordinal, interval and ratio.**
- Each levels has its own rules and restrictions
- Different levels of measurement contains different amount of information with respect to whatever the data are measuring

i Nominal Scale

- Qualitative
- No order or ranking in categories.
- These categories have to be mutually exclusive, i.e. it should not be possible to place an individual or object in more than one category
- A name of a category can be substituted by a number, but it will be mere label and have no numerical meaning

ii Ordinal Scale

- Qualitative
- Categories can be ordered or ranked
- A name of a category can be substituted by a number, but such a sequence does not indicate absolute quantities.
- Difference between any two numbers on the scale does not have a numerical meaningful.
- It cannot be assumed that the differences between adjacent numbers on the scale are equal.

iii Interval Scale

- Quantitative
- Data can be ordered or ranked
- There is no absolute zero point. Zero is only an arbitrary point with which other values can compare
- Difference between two numbers is a meaningful numerical value
- Ration of two numbers is not a meaningful numerical value.

iv Ratio Scale

- Quantitative
- Highest level of measurement
- There exist an absolute zero point (It has a true zero point)
- Ratio between different measurements is meaningful

6.2 Presentation of Data

The sinking of the Titanic is one of the most infamous shipwrecks in history.

On April 15, 1912, during her maiden voyage, the widely considered "unsinkable" RMS Titanic sank after colliding with an iceberg. Unfortunately, there

weren't enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew

[1]

Here's a quick summary of our variables:

Variable Name	Description
PassengerID	Passenger ID (just a row number, so obviously not useful for prediction)
Survived	Survived (1) or died (0)
Pclass	Passenger class (first, second or third)
Name	Passenger name
Gender	Passenger Gender
Age	Passenger age
SibSp	Number of siblings/spouses aboard
Parch	Number of parents/children aboard
Ticket	Ticket number
Fare	Fare
Cabin	Cabin
Embarked	Port of embarkation (S = Southampton, C = Cherbourg, Q = Queenstown)

6.2.1 Tabular Presentations of Data

Raw Data

- Raw data are collected data that have not been organized numerically
- Eg: Passenger age

```
##   PassengerId Pclass                                          Name    Sex  Age
## 1         892      3                              Kelly, Mr. James   male 34.5
## 2         893      3              Wilkes, Mrs. James (Ellen Needs) female 47.0
## 3         894      2                     Myles, Mr. Thomas Francis   male 62.0
## 4         895      3                              Wirz, Mr. Albert   male 27.0
## 5         896      3 Hirvonen, Mrs. Alexander (Helga E Lindqvist) female 22.0
## 6         897      3                    Svensson, Mr. Johan Cervin   male 14.0
##   SibSp Parch  Ticket    Fare Cabin Embarked
## 1     0     0  330911  7.8292              Q
```

[1] Data source: https://www.kaggle.com/varimp/a-mostly-tidyverse-tour-of-the-titanic

```
## 2     1     0  363272  7.0000             S
## 3     0     0  240276  9.6875             Q
## 4     0     0  315154  8.6625             S
## 5     1     1 3101298 12.2875             S
## 6     0     0    7538  9.2250             S
```

```
##  [1] 34.5 47.0 62.0 27.0 22.0 14.0 30.0 26.0 18.0 21.0   NA 46.0 23.0 63.0 47.0
## [16] 24.0 35.0 21.0 27.0 45.0 55.0  9.0   NA 21.0 48.0 50.0 22.0 22.5 41.0   NA
## [31] 50.0 24.0 33.0   NA 30.0 18.5   NA 21.0 25.0   NA
```

An array

- An array is an arrangement of raw numerical data in ascending or descending order of magnitude.
- Eg: Passenger age

```
##  [1]  9.0 14.0 18.0 18.5 21.0 21.0 21.0 21.0 22.0 22.0 22.5 23.0 24.0 24.0 25.0
## [16] 26.0 27.0 27.0 30.0 30.0 33.0 34.5 35.0 41.0 45.0 46.0 47.0 47.0 48.0 50.0
## [31] 50.0 55.0 62.0 63.0
```

Frequency Table (Frequency Distributions)

- A frequency table (frequency distribution) is a listing of the values a variable takes in a data set, along with how often (frequently) each value occurs
- frequency can be recorded as a
  - frequency or count: the number of times a value occurs, or
  - percentage frequency: the percentage of times a value occurs
- Percentage frequency can be calculated as,

Chapter 7

Sets and Relations

Chapter 8

Probability

Chapter 9

Correlation and Regression