IT1152 - Essentials of Statistics

Dr. M. Kayanan

Senior Lecturer
Deprtment of Physical Science
University of Vavuniya
Sri Lanka

kayanan@vau.ac.lk

Session 1
Introduction to Statistics

Introduction

- Statistics is concerned with scientific methods for **collecting**, **organizing**, **summarizing**, **presenting** and **analyzing** data as well as deriving valid **conclusions** and making reasonable **decisions** on the basis of this analysis.
  Two areas of statistics:
  - **Descriptive Statistics:** collection, presentation, and description of sample data.
  - **Inferential Statistics:** making decisions and drawing conclusions about populations.

Some important terms in statistics

- **Population**: The population is a complete set of all possible observations of the type which is to be investigated.
  Some examples are: total number of students studying in a school or college, total number of books in a library, total number of houses in a village or town.
- **Sample**: A finite subset of statistical individuals defined in a population is called a sample. The number of units in a sample is called the sample size.

Some important terms in statistics

- **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.
  Any statistical data can be classified under two categories depending upon the sources utilized.
  These categories are: **Primary data** and **Secondary data**.

Categories of data

- **Primary data** is the one, which is collected by the investigator himself for the purpose of a specific inquiry or study. Such data is original in character and is generated by survey conducted by individuals or research institution or any organization.
  The primary data can be collected by the following methods:
  - Direct personal interviews.
  - Indirect Oral interviews.
  - Information from correspondents.
  - Mailed questionnaire method.
- **Secondary data** are those data which have been already collected and analyzed by some earlier agency for its own use; and later the same data are used by a different agency.

Classifications of data

- Statistical data are classified in respect of their characteristics. Broadly there are four basic types of classification namely
  - Chronological classification
  - Geographical classification
  - Qualitative classification
  - Quantitative classification
- In **Chronological classification**, the collected data are arranged according to the order of time expressed in years, months, weeks, etc. The data is generally classified in ascending order of time.
- In **Geographical classification**, the data are classified according to geographical region or place. For example, the paddy production in the different district in Sri Lanka, wheat production in different countries, etc.

Classifications of data

- In **Qualitative or Categorical classification**, the data are classified based on the some attributes or quality like sex, literacy, religion, employment etc. Such attributes cannot be measured along with a scale.
- **Quantitative or Numerical classification** refers to the classification of data according to some characteristics that can be measured with a scale, such as height, weight, etc.

Classifications of data

- **Qualitative or Categorical data** can be broken into two levels: nominal and ordinal.
  Nominal data are categorical data that lack an ordering scheme. For example, gender, colors, etc.
  Ordinal data are categorical data that have an ordering scheme. For example, grades, income level (low, mid, high), etc.

Classifications of data

- Quantitative or Numerical data can be classified as either discrete or continuous.
- A set of data is **discrete** if there are only a finite number of values possible or if there is a space on the number line between each two possible values.
  Discrete data are usually associated with some sort of count. For example, the number of students in a class, the number of correct answers on a student's quiz.
- **Continuous data** make up the rest of numerical data. This is a type of data that is usually associated with some sort of physical measurement. For example, height, weight, etc.