Introduction to Statistics

Md. Aminul Islam Shazid

Outline

Introduction

What is Statistics?

- Collecting data
- Gaining insights from data
- Making decisions based on the insights gained from the data

Definition

Statistics can be defined as the art and science of:

- collecting, cleaning and organizing data
- summarizing and analyzing data
- presenting the summary or the analysis
- interpreting the analysis results
- gaining insights through analysis of data
- and finally, drawing valid conclusions and making sound decisions through the use of data.

Why Statistics is Necessary

-

Popuplation and Sample

- **Population** is the collection/set of all items or individuals of interest in a given study
- **Sample** is a *representative* portion of the population

For example:

- A study may target all the people in Bangladesh. However, it is unfeasible to collect information of everyone in the country in a timely or cost-effective effective way
- Therefore, data is collected from only a small portion of people from *all over the country*, this is called sampling. The individuals in a sample are usually selected randomly

Census and Survey

Parameter and Statistic

- A **parameter** is a characteristic or function of every objects or individuals in a population. For a fixed population, it is a fixed (but, usually unknown) value
- A **statistic** is a characteristic or function of every objects or individuals in a sample.
- A **statistic** is used to *estimate* a **parameter**

Parameter and Statistic (cont.)

- For a fixed population, the value of a parameter is fixed (but usually unknown)
- However, due to randomization, different samples can include different individuals from a population
- Therefore, the value of a statistic can vary across different samples

Parameter and Statistic (cont.)

For example:

- Suppose the goal is to find the average height of the students of a class
- The population average is a fixed value and it is unknown unless data is collected from everyone in the class
- If the heights of a some students are collected as a random sample, then we can estimate the population average using the sample average
- If another sample is collcted, the same individuals as the first sample may not be selected, therefore, the estimate shall be different from the first estimate

Types of Statistics

- **Descriptive statistics:** Methods for organizing, summarizing and presenting data in an informative way. For example:
  - A hypothetical customer survey finds that 50% of the customers are satisfied with a product
- **Inferential statistic:** Methods for using sample data to make predictions, test hypotheses, and generalize conclusions about a larger population. For example:
  - A study finds association between smoking and cancer

Variable

- Variable means something that can vary
- It is a characteristic that can vary across individuals or objects or items or cases of a phenomenon
- For example:
    - Age
    - Gender
    - Socioeconomic status
    - Temperature

Types of Variables

Variables:
- Qualitative
- Quantitative:
    - Discreet
    - Continuous

Qualitative vs Qualitative

- **Qualitative:** Qualitative variables describe qualities and are categorical. These are non‑numerical and descriptive values that represent attributes or categories. For example:
    - Name of a person
    - Gender
    - Hair colour
- **Quantitative:** Quantitative variables measure quantities with numbers. These are numeric data that can be counted or measured, allowing for mathematical calculations. For example:
    - Height
    - Temperature
    - Number of students in a class

Discreet vs Continuous

- **Discreet:** *Countable*, usually whole numbers. Finite number of possible values. For example:
  - Number of students in a class
- **Continuous:** *Measurable*, can have fractional values. Can take values in a given range. Infinite number of possible values in any range. For example:
  - Height