

Chapter 1 Introduction to Statistics

Definition of statistics

The word '*statistics*' has 3 meanings.

1. *Statistics* refers to numerical facts.
2. *Statistics* refers to the field or discipline of study.
→ Statistics is a group of methods that are used to collect, analyze, present, and interpret data and to make decisions.
3. *Statistics* refers to particular quantity that had been calculated from a sample data.
✓ Mean, variance, standard deviation etc.

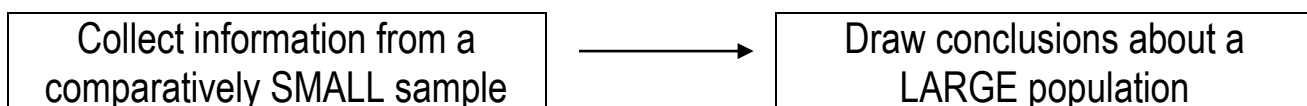
Branches of statistics

1. Descriptive Statistics
→ Consists of methods for organizing, displaying, and describing data by using tables, graphs, and summary measures.
→ Deals with the description and analysis of a given group of data.
→ Present information in a convenient, usable and comprehensible form.
2. Inductive Statistics (Inferential Statistics)
→ Consists of methods that use sample results to make decisions or predictions about a population.
→ Deals with the problems of making inferences or drawing conclusions about population based on information obtained from the samples taken from the population.
3. Probability Theory
→ Gives a measurement of the likelihood that a certain outcome will occur.
→ Acts as a link between descriptive and inferential statistics.

→ Is used to make statements about the occurrence of a certain event under uncertain conditions.

Purpose of statistics

→ Drawing conclusions about the population by studying the sample.



Population versus Sample

Population or Target Population

→ Consists of all elements -- individuals, items, or objects -- whose characteristics are being studied.

Sample

→ A portion of the population selected for study.

Example 1:

Suppose we want to study the height of the students from TARUC.

The *population* is all students from TARUC.

The *sample* is 500 students from TARUC.

Parameter

A numerical value summarizing all the data of an entire population.

Example 2:

A student is interested in finding out something about the average ringgit value of cars owned by the staff of TAR University College.

The *population* is the collection of all cars owned by all staff at TAR University College.

The *sample* is any subset of that population. For example, the cars owned by members of the mathematics and statistics department is a sample.

The *variable* is the “ringgit value” of each individual car.

The data are the set of values that correspond to the sample obtained (65000; 42000; 120000; ...)

The *parameter* is the “average” value of all cars in the population.
The *statistic* is the average value of the cars in the sample.

Example 3:

The admissions office wants to estimate the cost of textbooks for students at our college. Let the variable x be the total cost of all textbooks purchased by a student this semester. The plan is to randomly identify 50 students and obtain their total textbook costs. The average cost for 50 students will be used to estimate the average cost for all students.

- a. Describe the parameter the admissions office wishes to estimate.
- b. Describe the population.
- c. Describe the sample.
- d. Describe the statistics and how you would use the 50 data collected to calculate the statistics.

Solution

- a. Average cost of textbooks for the semester per student for all students.
- b. All students enrolled for this semester.
- c. The 50 students.
- d. The average cost of textbooks for the semester per student for the 50 students; add all 50 values and divide the total by 50.

Type of variables

1. *Quantitative variable*
 - A variable whose values can be measured numerically.
 - Examples: incomes, heights, sales, number of accidents etc.
 - Can be classified as discrete variables or continuous variables.

- *Discrete variable*
 - A variable whose values are countable and usually integer-valued.
 - Can assume only certain values with no intermediate values.
 - Example: The number of cars sold in one month etc.
- *Continuous variable*
 - A variable whose values cannot take exact value.
 - The precision depends on the instruments.
 - Assume any numerical value over a certain interval or intervals.
 - Example: The height of student etc.
 -
- 2. *Qualitative or Categorical variable*
 - A variable that cannot assume a numerical value but can be classified or ranked into two or more nonnumeric categories.
 - Examples: Beauty, intelligence, sex etc.

Types of data

1. *Quantitative data*
 - The data collected on a quantitative variable.
 - Consists of discrete data or continuous data.
- *Discrete data*
 - Can take only exact values.
 - Example of discrete variable: The number of cars sold in one month.
 - Example of discrete data: 3061, 4781, etc.
 - *Continuous data*
 - Cannot take exact values.
 - Can be given only within a certain range or measured to a certain degree of accuracy.

- Example of continuous variable: The height of student from TARC.
- Example of continuous data: 170.1 cm, 178.1 cm etc.

2. *Qualitative or Categorical data*

- The data collected on a qualitative variable.
- Consist of descriptions using 'names'.
- Examples of qualitative variable: The grade for AAMS2613.
- Examples of qualitative data: 'A', 'B', 'C' etc.