Statistics

Statistics (a plural word) mean data.

Statistics (a singular word with same spelling) is a branch of applied mathematics that teaches us how to collect, organize, summarize and analyze data, and how to interpret the results. It helps us take informed decisions in the presence of uncertainty and variation.

Population and Sample

Population is the collection of all individuals or items under consideration in a study. We use *N* to represent the size of the population (number of items in the population).

- If the objective of a study is to know the average age of first year students of Dhaka University, then the population consists of all first-year students of Dhaka University.

- If the objective of a study is to know the average age of first year students of Bangladeshi universities, then the population consists of all first-year students of all universities in Bangladesh.

Sample is representative part of the population from which information is collected. We use *n* to represent the size of the sample.

Parameter and Statistic

Parameter is a population characteristic (a value that could be calculated if all the values in the population were known). For example, population mean, population percentage, etc. are parameters.

Statistic is a sample characteristic (a value calculated from the sample). For example, sample mean, sample percentage, etc. are statistics.

Problem:
You are interested in the average monthly income of Dhaka University graduates. You selected 100 graduates of Dhaka University at random and recorded their monthly incomes. The average of these 100 numbers is 33.7 thousand. Answer the following questions in this context.

a. What is the population?
b. What is the sample?
c. What is the parameter?
d. What is the statistic?

Solution:

a. All DU graduates.
b. 100 DU graduates selected at random.
c. Average monthly income of all DU graduates.
d. Average monthly income of 100 DU graduates.

Problem:
You are interested in the percentage of DU students that support tuition fee hike. You selected 100 students at random and 13 of them supported tuition fee hike.

a. What is the population?
b. What is the sample?
c. What is the parameter?
d. What is the statistic?

Solution:
Do it yourself

Two branches of Statistics:

- **Descriptive Statistics:** Discusses how to collect, organize, summarize and analyze data.
- **Inferential Statistics:** Discusses how to make conclusion about population based on the results obtained from sample data.

Data and Variable

A **variable** is any characteristic whose value may change from one object to another in the population.

Two types of variable

1. **Qualitative/Categorical variable:** Gender of respondents, Political affiliation (Democrat, Republican, Independent), Political ideology (Liberal, Moderate, Conservative), smoking status (yes/no).

2. **Quantitative/Numerical variable:** Age, height, weight, household size.

Two types of quantitative variable

a. **Discrete variable:** A numerical variable is discrete if its set of possible values either is finite or else can be listed in an infinite sequence (one in which there is a first number, a second number, and so on).
   A discrete variable almost always results from counting, in which case possible values are 0, 1, 2, 3, … or some subset of these integers.

   Example: Household size, Number of persons known by the respondent with HIV.

b. **Continuous variable:** A numerical variable is continuous if its possible values consist of an entire interval on the number line.
   Continuous variables arise from making measurements.

   Example: Weight of individuals, Age, Income.

Scales of measurement

Qualitative variables:

- **Nominal:** Unordered (unranked) categorization
  Example: gender/sex, eye color.

- **Ordinal:** Ordered categories
  Example: income (high, medium, low).

Quantitative variables:

- **Interval:** Numeric values, but zero is arbitrary. Therefore, we can only add or subtract, but cannot take ratios (cannot multiply or divide).
  Example: temperature (0 does not imply "no temperature"), calendar year.

- **Ratio:** Numeric values with a meaningful zero point.
  Example: height, income.

Problem: Classify (categorical/discrete/continuous) the following variables. In which scale (nominal/ordinal/interval/ratio) they are measured?

1. Practices exercise (Never, Sometimes, Regular)
2. College major (Education, Anthropology, Physics, etc.)
3. Political ideology (Liberal, Moderate, Conservative)
4. Age
5. Temperature (Low, medium, high)
6. Marital status (never married, married, divorced, widowed)