***Some Important Concepts***

**Population:** The totality or collection of all experimental units which we want to study.

Example:

* All registered voters in our Country.
* All daily maximum temperatures in July for different cities in Bangladesh.
* All students in a university.

**Types of population:**

**Finite population:** A population is said to be finite if it consists of a finite or fixed number of elements.

For example: The number of vehicles crossing a bridge every day, the number of births per years and the number of words in a book are finite populations.

**Infinite population:** A population is said to be infinite if there is no limit to the number of elements it can contain.

For example: 1- Stars in sky, 2- Dots in a line, 3- Hair on head, 4- The number of germs in the body of a sick patient is perhaps something which is uncountable.

**Experimental unit:** Each individual or object of a population is called an experimental unit.

Example: An employee of a firm, a student of a class, a patient of a hospital etc.

**Sample:** A sample is a representative part of the population. When a survey is being carried out, a sample is taken from the population one wants to study.

Example:

* Some registered voters in our Country.
* Some daily maximum temperatures in July for different cities in Bangladesh.
* Some students in a university.

**Parameter**

A **parameter** is any summary number, like an average or percentage that describes the entire population.

The population mean μ (the greek letter "mu") and the population proportion *p* are two different population parameters.

**Statistic**

A **statistic** is any summary number, like an average or percentage that describes the sample.

The sample mean, x¯, and the sample proportion p^ are two different sample statistics.

**Variable:** Variable is a characteristics which vary from experimental unit to experimental unit.

Example: Age of a worker, religion of a student, income of household, gross profits of a company, gender of garment worker etc.

**Types of variable:**

1. **Quantitative variables**

When you collect quantitative data, the numbers you record represent real amounts that can be added, subtracted, divided, etc. There are two types of quantitative variables: **discrete** and **continuous**.

| **Discrete vs continuous variables** | | |
| --- | --- | --- |
| **Type of variable** | **What does the data represent?** | **Examples** |
| **Discrete variables** | Counts of individual items or values. | * Number of students in a class (1-5) * Number of defective items in a lot * Number of printing mistakes per page |
| **Continuous variables** | * Measurements of continuous or non-finite values. * If you can take any values within a certain range | * Distance (1m-5m) * Volume * Age(1y-5y) |

1. **Qualitative or Categorical variables**

Categorical variables represent groupings of some kind. They are sometimes recorded as numbers, but the numbers represent categories rather than actual amounts of things.

There are three types of categorical variables: **binary**, **nominal**, and **ordinal** variables.

| **Binary vs nominal vs ordinal variables** | | |
| --- | --- | --- |
| **Type of variable** | **What does the data represent?** | **Examples** |
| **Binary variables** | Yes/no outcomes | * Heads/tails in a coin flip * Win/lose in a football game |
| **Nominal variables** | Groups with no rank or order between them. | * Species names * Colors * Brands * Religion (1=Muslim, 2=Hindu, 3=Buddist and 4=Christian) |
| **Ordinal variables** | Groups that are ranked in a specific order. | * Exam grade of the students * Socio economic status (Poor, middle, rich) |

## ****Data:** A set of observations obtained from a particular study.**

**Example:** Daily production of a factory, age of workers, IQ scores of students etc.

## ****Data Collection****

Depending on the source, it can classify as primary data or secondary data. Let us take a look at them both.

### Primary Data

These are the [data](https://www.toppr.com/guides/economics/organisation-of-data/raw-data-classification-of-data-and-variables/) that are **collected for the first time** by an investigator for a specific purpose. Primary data are ‘pure’ in the sense that no statistical operations have been performed on them and they are original.

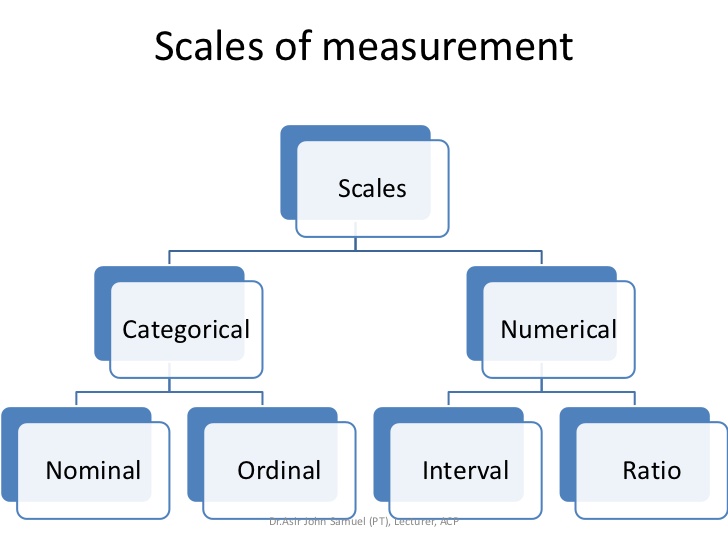
An example of primary data is the [**Census of**](https://www.toppr.com/guides/general-awareness/current-affairs/population-census/) **Bangladesh**.

**Secondary Data**

They are the data that are sourced from some place that has originally collected it. This means that this kind of data has already been collected by some researchers or investigators in the past and is available either in published or unpublished form. This information is impure as statistical operations may have been performed on them already.

An example is information available on the Government of Bangladesh, the Department of Finance’s website or in other repositories, books, journals, etc.

**Measurements:** It is a process of assigning numbers to some characteristics or variables according to some scientific rules.

****

|  |  |  |
| --- | --- | --- |
| **Levels of scales** | **Definition** | **Example** |
| Nominal scale | The scale of measurement by which we can identify and classify a qualitative variable according to different categories is called nominal scale.  (No mathematical analysis) | * Gender(1=Male, 2= Female) * Political preferences * Place of residence(Rural, Urban) * a customer survey asking “Which brand of smartphones do you prefer?” Options : “Apple”- 1 , “Samsung”-2, “OnePlus”-3. * Religion |
| Ordinal scale | The scale of measurement by which we can identify, classify and rank a qualitative variable according to different categories is called ordinal scale.  (<,>) | * Status at workplace, * Tournament team rankings, * order of product quality, * order of agreement or satisfaction   How satisfied are you with our services?  1- Very Unsatisfied  2- Unsatisfied  3- Neutral  4- Satisfied  5- Very Satisfied   * Grade of the students * Socio economic Status(Poor, middle, rich) |
| Interval scale | The scale of measurement by which we can measure a quantitative variable numerically on experimental unit with arbitrary zero as origin is called interval scale.  (<,>,+,-) | * Temperature * Marks obtained by students in an examination * Calendar year |
| Ratio scale | The level of measurement is called ratio scale when a quantitative variable is measured numerically on experimental unit with absolute zero as origin.  (<,>,+,-,×,÷) | * Age * Height * Weight * Number of printing mistakes per page in book * Number of defects of a product |

Summary –  Levels of Measurement

The four data measurement scales – nominal, ordinal, interval, and ratio –  are quite often discussed in academic teaching. Below easy-to-remember chart might help you in your statistics test.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Offers:** | **Nominal** | **Ordinal** | **Interval** | **Ratio** |
| The sequence of variables is established | – | Yes | Yes | Yes |
| Mode | Yes | Yes | Yes | Yes |
| Median | – | Yes | Yes | Yes |
| Mean | – | – | Yes | Yes |
| Difference between variables can be evaluated | – | – | Yes | Yes |
| Addition and Subtraction of variables | – | – | Yes | Yes |
| Multiplication and Division of variables | – | – | – | Yes |
| Absolute zero | – | – | – | Yes |