Variable:
The measurement of elements of a population having certain characteristics may vary from element to element either in magnitude or in quality. These measurable characteristics are called variables.

Thus, a measurable characteristic, which can vary from element to element with in its domain called a variable. Usually we denote the variables by capital letters and their values by small letters.

Example: Height, weight, age, SSC and HSC marks, family size, sex, etc. are some variables of 1st semester BBA students of BRAC University.

Types of Variables
There are two basic types of variables -

1. Qualitative variable (also known as categorical variable or attribute)
A qualitative variable is one for which numerical measurement is not possible. In other word when the characteristic being studied is nonnumeric, it is called a qualitative variable or an attribute.

For example: Hair color (brown, black, white etc.), religion (Muslim, Hindu, etc.), sex (male, female), home district (Dhaka, Rajshahi, Bogra etc.), occupational status (employed, unemployed, self-employed, others) etc.

An individual is simply assigned to any one of the several mutually exclusive categories on the basis of observation on the individual. The qualitative observations can neither meaningfully ordered nor physically measured, these can only be classified and then enumerated.

In dealing with the qualitative data, researchers are usually interested in how many or what proportion fall in each category.

For Example:
- What percent of students of BRAC Universities of English medium background?
- What proportion of people opted in favor of construction of the new Airport?
- How many Muslims and how many Hindus are there in Bangladesh?

2. Quantitative variable (also known as numerical variable)
A quantitative variable is one for which the resulting observations have numeric value and thus possesses a natural ordering.

Quantitative Variables are measured on a numeric or **quantitative scale.**

Example: districts population size, customer's shoe size, vehicle's speed on a highway, amount of daily sell of a shop, parents income etc. are all **quantitative variables**

Quantitative (numeric) variable is further subdivided as discrete and continuous variables.

- **Discrete variable:**
  When a variable can assume only isolated values within a given range is called discrete variable.
  Example: Number of children in a family, number of road accident in a year, number of phone call received in a phone booth, height of nails etc.

- **Continuous variable:**
  A variable is said to be a continuous variable if it can theoretically assume any value within a given range or ranges.
  Example: height of a person –since it can take any value between 5.6 feet and 5.8 feet.

Exercise:

a. Classify each variable as qualitative or quantitative:
   i. Marital status of nurses in a hospital
   ii. Time it takes to run a marathon
   iii. Weights of lobsters in a tank in a restaurant
   iv. Colors of automobiles in a shopping center parking lot
   v. Ages of people living in a personal care home

b. Classify each variable as discrete or continuous:
   i. Number of pizzas sold by Pizza Express each day
   ii. Lifetimes (in hours) of 15 iPod batteries
   iii. Weights of the backpacks of first graders on a school bus
   iv. Number of students each day who make appointments with a mathematics tutor at a local college
   v. Blood pressures of runners in a marathon

Data:

Numerical facts gathered from a statistical investigation are called a data.
In a statistical analysis the first work is to collect data the raw materials of statistics after identifying a specific problem and field of enquiry.

Data is in fact the *plural form* of *'datum'*. Single information of a phenomenon on any subject of interest is called a datum. So data is called the collection of datum.

Example: If we are interested about the height of the students of 1st semester in BBA of BU, then a single value (that is the height of a student) is called a datum, and the set of all values of height will be data.

Sources and Types of data:

Based on the sources data can be of two types.

Primary data: A data is said to be primary data if it is obtained from an investigation conducted for the first time. Thus, the data collected for the first time by the investigator as original data are known as primary data.

Secondary data:

When a statistical analysis is conducted on a data set available from a prior investigation is called a secondary data.

Example: National income data collected by the government are primary data but they become secondary data for those who use them.

Raw data:

In any statistical investigation, when data first collected usually appear in raw form where, information has been recorded merely in arbitrary order in which they happened to occur. This is known as the raw data set.

Raw data, collected for any statistical investigation, is unable to represent the summary information, which are although preliminary but necessary for analyses with advanced statistical method. So it is necessary to represent the raw data in such a way, which will enable us to extract the preliminary ideas about the variable(s) under study, to get some summary measures and also to perform further statistical analysis.

Dealing with Raw Data: How to prepare data for further Statistical operation

In the next few subsequent segments, we are going to discuss on some techniques of statistics that we usually used to condense raw data, to make the data prepared for further statistical application.

The most frequently used methods for data condensation or/and representation are

i. Classification
ii. Tabulation
iii. Graphical representation

Classification:

Classification is the process of arranging data values of a variable in groups or classes according to their affinities or of our interest. It is the first step towards further processing of a heterogeneous mass of data in to a number of homogeneous groups and subgroups by their respective characteristics.

Purpose of classification:

Classification is necessary to serve the following purpose:

i. To eliminate unnecessary details.
ii. To bring out clearly point of similarity and dissimilarity.
iii. To enable one to form mental picture of the object.
iv. To enable one to make comparisons.
v. To pin point the most significant features of the data at glance.
vi. To enable a statistical treatment of the collected data.

Principles of determining the number of classes:

Usually we determine the number of classes in the light of the following conjoined considerations

i. The number of observations of a variable.
ii. The lowest and highest value of a variable.

iii. Even distribution of the values with in classes.
iv. A regular sequence of frequencies.
v. Avoidance of extremely large or small number of classes.

Tabulation:
A statistical method of data condensation by which we can represent summary information of one or more variables, is defined as tabulation.

A statistical table is the logical listing of collected data in vertical columns and horizontal rows of numbers with sufficient explanatory and qualifying words, terms and statements in the form of titles, headings and notes which make clear the full meaning of data and their origin.

Principles of the constructions of a table:
Some of the most basic principles that one should consider in constructing table are as follows:

1. The table should be self-explanatory. The title describing the contents of the table should be clear, concise and to the point.
2. The table should be as simple as possible. Two or three tables are often preferable to a large table containing too many details and variables.
3. The specified units of measurements for the data should be given.
4. Necessary code or symbols used in table should be explained in a footnote.
5. Sources of data should be mentioned.

Frequency distribution:
The number of times a particular value of a certain variable occurs in a set of observations is called the frequency of that value and the manner in which the frequencies are distributed in the different classes is known as the frequency distribution of the values of that variable.

That is frequency distribution can be defined as a summary presentation of a number of observations of an attributes or values of a variable arrange according to their occurrence either individually (in case of discrete data) or in a range (in case of both discrete or continuous data)

Table 01: Frequency distribution of number of children per family

Number of children	Number of families
0	10
1	27
2	15
3	18
4	9

Table 02: Frequency distribution of height of trees in Sundarbans

Height of the tree (In Feet)	Number of trees
0-50	1000
50 – 100	2735
100 – 150	1589
150 – 200	1518
More than 200	719

Class limit:
Class limits are the highest and the lowest values that can be included in the class.

For example, if we consider the class 50 – 100, here 50 is the lower limit and 100 is the upper limit. In such case no values greater than 100 shall fall into that class. Similarly, no values less than 50 shall fall into that class either.

Class interval:
The difference between the upper limit and the lower limit of a class is called the class interval.
Class interval is usually denoted by c, i, h or w.
For example, the class interval of the class '50 – 100' is 50.

Class frequency:
The number of observations falling with in a particular class is called its frequency or class frequency.

Class midpoint or class mark:
The value of the variable that lies in the middle of the upper and lower limits is called mid value or midpoint of the class.
It can be obtained as follows:

Class midpoint $= \frac{l+s}{2}$; Where l = Upper limit of the class, s = Lower limit of the class

Relative frequency (also known as proportion):
Instead of presenting the frequencies in absolute terms, it is sometimes convenient to express the frequencies in percentages. The relative frequency (also known as proportion) corresponding to a class is simply the ratio of the total number of items in that class to the total number of elements in the total set. ***Multiplying relative frequency by 100 one can obtain the percentage of observation that belongs to any particular class***.

$$\text{Relative frequency} = \text{Proportion} = \frac{\text{Frequency in each class}}{\text{Total number of observations}}$$

Cumulative frequency:
The cumulative frequency corresponding to a class is the total of all frequency up to and including that class.

Example: let us consider the following table showing the distribution of mark of 27 students

Class limit	Class mid value	Frequency	Relative frequency	Cumulative frequency	Cumulative relative frequency
0 – 10	5	4	0.148	4	0.148
10 – 20	15	8	0.296	4+8	0.444
20 – 30		5		4+8+5	
30 - 40		4			
40 – 50		3			
50 – 60		2			
60 – 70	65	1			
Total					

Assignment W1D2_001

Question 1:
What is tally bar? (https://www.youtube.com/watch?v=R6m8OQAQzPk)

Question 2:
The following information, extracted from a survey of a Microfinance institution (MFI) represents the amount of loan request of 50 potential borrowers from any particular branch of that MFI.

1850	9250	6100	4500	5100	1800	6100	6500	6999	6780
3100	7475	6400	4950	8789	6100	6480	7050	9900	4790
4400	7900	6900	3865	5556	4859	6999	6780	8050	9900
5600	6600	9980	4800	8855	5550	1200	4790	6500	8050
3858	7300	8050	6200	7155	4980	8050	6480	7050	1500

For the given data construct a suitable frequency distribution table featuring the following components

i. Class mid value
ii. Tally Bars
iii. Frequency
iv. Relative frequency
v. Cumulative frequency
vi. Cumulative relative frequency

Using the aforementioned information also answer the following

a. Determine the number of loan request between tk 4000-6000
b. Determine the proportion of loan request between 4000 – 6000.
c. Determine the number of loan request below tk. 7000.
d. Determine the proportion of loan request below tk. 7000.