

## STAT-2311 (Probability and Statistics)

### Introduction of statistics (1st chapter)

Origin of statistics: The word statistics have been derived from the latin word 'statulus' or the Italian word 'statista' or the German word 'statistik' or the French word 'statistique'. Each of which means 'political state' or a Government.

Def<sup>n</sup> of statistics: Statistics is a branch of science dealing with the collection of data, organizing, summarizing, Presenting, analysing data and drawing valid conclusions and thereafter making reasonable decisions on the basis of such analysis.

(মিথোলা তথ্যক সাজোয়া সার্থক represent করা)

### Most essential functions of statistics:

1. It simplifies much of figure.
2. It helps in determining the relationships two or more phenomenon.
3. It helps in formulating and testing suitable hypothesis.
4. It helps a central management in formulating suitable private policy.
5. It facilitates classification.
6. It helps in predicting future trends.

### Limitations of statistics:

1. Statistics deals with aggregate of items and not with individual item or measurement.
2. Statistics deals with quantitative characteristics.
3. Statistics laws hold good only for the averages.
4. It plays only an auxiliary role.
5. It can be misused.

### Importance of statistics:

1. ~~The~~ Wealth and manpower are important for development and planning.

2. Statistics are invaluable in business and commerce.
3. It helps the planner to estimate the revenue income and expenditure of the country.
4. Agriculture statistics may play a key role in agriculture development.
5. In industry, statistics is widely used to provide quality control.
6. Statistics is usually used in education and psychology too.

## Relation between computer and statistics.

Statistics is defined as the science of collecting, organizing which we call data. It is very important for a student of computer sciences.

As computer sciences also deals with organization and interpretation of numerical facts.

Most of the principles of computer sciences are based on concept of statistics.

The computer can process large amount of data quickly and accurately.

For processing the large amount of data some of the important statistical packages that have been used are- SPSS, SAS, STRATA, S-plus and MINITAB.

Population: The totality of all elements under the study or discussion is called population. The population in statistics includes all members of a defined group that we are studying or collecting information.

Example: If we measure the heights and weights of a group of person then it is called population.



There are two types: (i) Finite population (countable)  
(ii) Infinite population (uncountable)

Finite population: A population is called finite population if its elements are countable. Exp: Number of students in a university.

Infinite population: A population is called infinite population if its elements are not countable. Exp: Number of fishes in the Bay of Bengal.

Sample: A representative part of population is called sample. Or a part of population is called sample.  
(Population to represent ~~for~~)

Exp: If we measure the height and weight of ITUC student, then that of the CSE/EEE/ETE department student are sample.

## Difference between population and sample:

Population	Sample
1. The totality of all elements under the study or discussion is called Population.	1. A part of the population is called sample.
2. Population may be finite or infinite.	2. Sample must be finite.
3. Collecting data from every element of a population is not easy.	3. Collecting data is easy.
4. All registered voters in our country.	4. All registered voters in Chittagong district.

Variable: Any phenomenon which varies from individual to individual is called Variable. Variables are represented by symbols (e.g.,  $x$ ,  $y$ , or  $z$ )

Example of variable: Age, weight, height, income, expenses, country of birth etc.

There are two types: 1. qualitative (categorical)  
2. quantitative (numeric)

Qualitative Variable: A variable which cannot be expressed as numerically is called qualitative (categorical) variable.

Example: Hair colour, gender, business type etc. (অনুমান convert করা যায় না)

Quantitative Variable: A variable which can be expressed as numerically then it is called quantitative variable.

Example: Height, Age, GPA, salary etc.

(অনুমান convert করা যায়)

Two types of quantitative variable:

① Discrete

② Continuous

Discrete Variable: A variable is called discrete variable if it can take only isolated (whole) values.

Example: Number of children in a family, number of business locations, all of which measured as whole units (i.e., 1, 2, 3)

(କୃତ୍ରିମ ନୂନାଂଶ, ଅସ, ପରିମିତିର ବିଶେଷ)

Continuous variable: A variable is called continuous

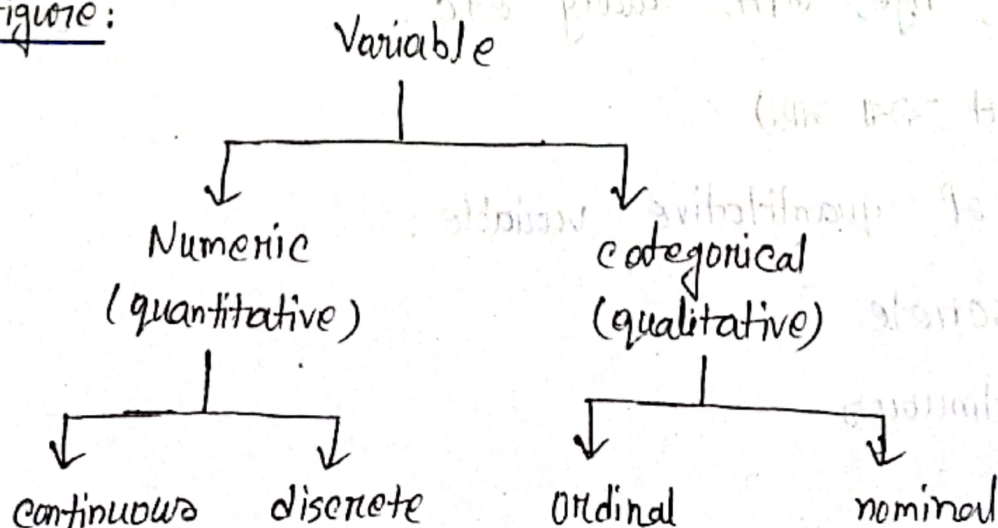
variable if it can take any values between certain limits.

Example: time, height, temperature.

(ନୂନାଂଶ, ପରିମିତିର ଦୂରତା ଓ ତାପମାତ୍ରା)

❑ Every continuous variable are discrete but every discrete variable are not continuous.

Figure:



(କୃତ୍ରିମ ନାମ ଶାବ୍ଦ, Noun ଏବଂ ବ୍ରାହ୍ମ କାଳ କାଳ)



## Difference between qualitative variable and quantitative variable:

Qualitative Variable	quantitative Variable
<ul style="list-style-type: none"> <li>① Cannot be measured numerically.</li> <li>② It is not countable</li> <li>③ It can be measure by using nominal and ordinal scale.</li> <li>④ The value of this Variable is generally discrete.</li> <li>⑤ Algebraic expression meaningless Example:</li> <li>⑥ Skin colour, hair colour, gender etc.</li> </ul>	<ul style="list-style-type: none"> <li>① Can be measured numerically.</li> <li>② It is countable.</li> <li>③ It can be measure by using interval and ratio scale.</li> <li>④ The value of this Variable is discrete and continuous.</li> <li>⑤ Algebraic expression meaningful.</li> <li>⑥ Example: GPA, age, temperature etc.</li> </ul>

## Difference between Discrete Variable and continuous Variable:

Discrete Variable	Continuous Variable
<ul style="list-style-type: none"> <li>① A variable is called discrete if it can take only isolated values.</li> <li>② It is countable</li> <li>③ It may be finite or infinite</li> <li>④ Example: Family member, student member.</li> </ul>	<ul style="list-style-type: none"> <li>① A variable is called continuous if it can take any values between certain limits.</li> <li>② It is measurable</li> <li>③ It must be finite.</li> <li>④ Example: age, weight, salary etc.</li> </ul>

Constant: A number that is not changing. It is usually denoted by  $a, b, c$  and  $d$ .

Difference between Variable and constant:

Variable	constant
<p>① A variable is always subject to change.</p> <p>② It is denoted by <math>x, y, z, u, v</math>.</p> <p>③ Variable are qualitative and quantitative.</p> <p>④ Example: Age, weight, height etc.</p>	<p>① A constant will not change.</p> <p>② It is denoted by <math>a, b, c</math> or <math>d</math>.</p> <p>③ constant has no classification.</p> <p>④ Total number of days in a week, Number of finger in a hand.</p>

Data: A set of observations obtained from a particular enquiry is called data.

Exp: Income of workers or examination marks of a student.

There are two types of data. ① Primary data  
② Secondary data

Primary data: Data collected by the investigator himself/herself for a specific purpose. Exp: Data collected by diff. govt. public, private organizations.

Secondary data: When an investigator uses the data which has already been collected by others, such data are called secondary data.

This data can be obtained from journals, report, Books, Internet etc.

Methods of collecting primary data :

1. Through interview
2. Through questionnaire
3. Through schedule
4. Through local agent
5. Through observations
6. Through experimentation.



## Difference between primary and secondary data.

Primary data	Secondary data
<p>① Primary data are always original as it is collected by the investigator.</p>	<p>① Secondary data lacks originality. The investigator makes use of the data collected by the other agencies.</p>
<p>② Suitability of the primary data will be positive because it has been systematically collected.</p>	<p>② Secondary data may or may not suit the objects of enquiry.</p>
<p>③ Primary data are expensive and time consuming.</p>	<p>③ Secondary data are relatively cheaper.</p>
<p>④ Extra precautions are not required.</p>	<p>④ It is used with great care and caution.</p>
<p>⑤ Primary data are in shape of raw material.</p>	<p>⑤ Secondary data are usually in the shape of readymade products.</p>
<p>⑥ Possibility of personal Prejudice.</p>	<p>⑥ Possibility of lesser degree of personal Prejudice.</p>



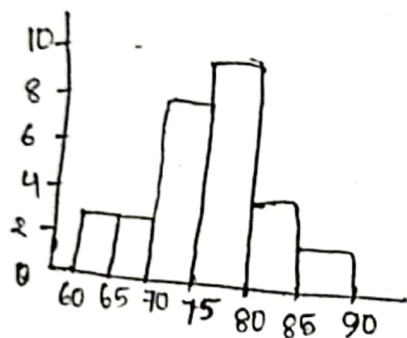
## Graphical representation of statistical data:

In addition to presentation of statistical data through tabular form, one can present the same through some visual aids known as graphs and diagrams.

### Types of Graphs and Diagrams:

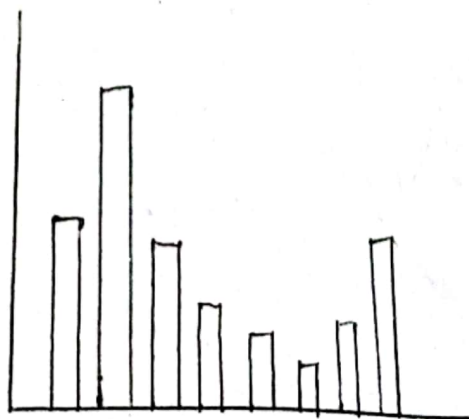
1. Histogram
2. Bar diagram
3. Frequency polygon
4. Pie diagram
5. Scatter diagram
6. Line diagram
7. Ogive
8. Stem and leaf plot.
9. Box plot.

### Histogram:



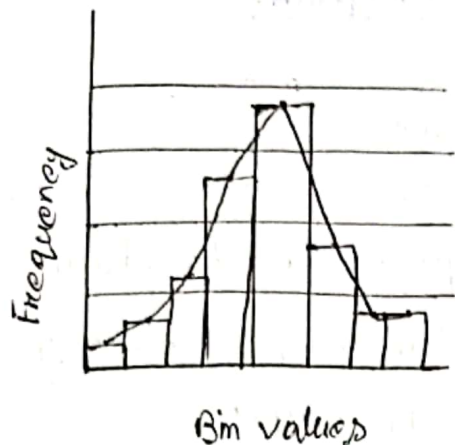
A histogram consists of tabular frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval.

### Bar chart:



A bar chart with rectangular bars with length proportional to the values that they represent. The bars can be plotted vertically or horizontally.

## Frequency polygon (সমন্বিত বহুভুজ)

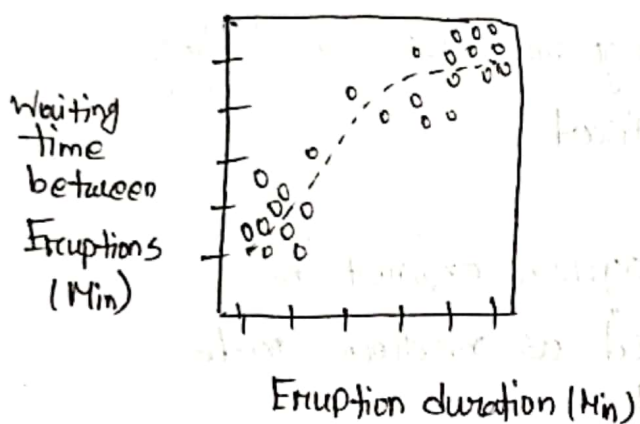


## Pie chart:

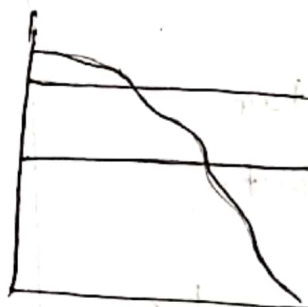


A pie chart shows percentage values as a slice of pie.

## Scatter diagram:

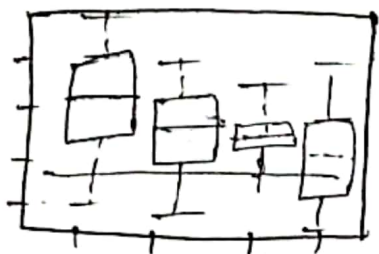


## Line chart:



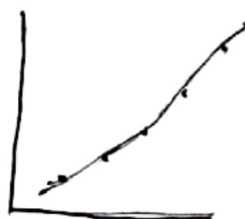
A line chart is a two dimensional scatter plot of ordered observations where the observations are connected following their order.

## Box plot:

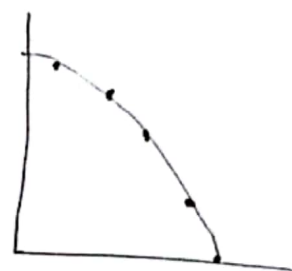


## Cumulative frequency polygon or Ogive:

upper class:



lower class



upper and lower class



## Difference between graph and diagram:

Graph	Diagram
<p>① Graph can be drawn only on plain paper.</p> <p>② Graph is not more effective and impressive.</p> <p>③ Graph doesn't have everlasting effect.</p> <p>④ Graph can be used as median, mode etc.</p> <p>⑤ Graph cannot be represented as an approximate data.</p>	<p>① Diagram can be drawn on plain paper and any sort of paper.</p> <p>② Diagram is more effective and impressive.</p> <p>③ Diagram have everlasting effect.</p> <p>④ Diagram cannot be used as median, mode etc.</p> <p>⑤ Diagram can be represented as an approximate data.</p>

