

UNIT-1

Concept of statistical population and sample from a population

Population/universe: A population is any group of individuals with at least one characteristic in common that distinguishes it from other individuals. For instance, if we want to know the mean weight of all 25 year old females in the Kashmir, then the population is all females who are 25 years old and living in the Kashmir.

Definition:

According to Simpson and Kafka, “ A universe or population may be defined as an aggregate of items possessing a common trait or traits”

Population could be enormous in size. There could be millions or even billions of individuals in the population. The number of individuals in the population is called population size (N). The individual units of the population are called items or elements.

Types of population:

1. Finite and infinite population: A population containing a finite number of items is known as finite population. For example the number of workers in a factory, students in a college etc. on the other hand an infinite population consists of an unlimited number of elements and are not countable. For example, the number of stars in the sky.
2. Real or Hypothetical population: Real population is that in which the items actually exist. For example students of a college, the books in the library etc. on the other hand, Hypothetical population consists of imaginary items. For example population of the throws of a die or a coin, thrown infinite number of times.

Remark: Population can refer to things as well as people.

Sample: A sample is only a small portion or subset of the population. The number of individuals in a sample is called sample size(n). Samples can be quite large or quite small. In theory, one individual from a population constitutes a sample.

Remark: To make our study less expensive in terms of time and resources, we usually study a subset(sample)of the population.

Data: Data can be defined as a collection of facts or information from which conclusions may be drawn.

Collection of data is an important thing in statistical data analysis. Data can be collected from sources or through observation ,surveys or by doing experiments.

Types of Data:

Primary data: The data which are originally collected by an investigator or agency for the first time for any statistical investigation and used by then in the statistical analysis are known as primary data. Such

data are original in character and are large in number. It is called primary because it is collected from the very source where the information is generated.

Definition: According to Horace Sacrist, “ By Primary data are meant those data which are original, that is, those in which little or no grouping has been made, the instance being recorded or itemized as encountered. They are essentially raw material.”

A) **Methods of collecting primary Data:**

Following methods can be employed for the collection of primary Data:

1. **Direct Personal Investigation.**
2. **Indirect Oral Investigation.**
3. **Information from local sources or Correspondents.**
4. **Mailed questionnaire Method.**
5. **Schedules filled by enumerators.**
6. **Telephone interviews.**

1 Direct Personal investigation: As the names implies under this method, the investigator personally studies the phenomenon in hand and obtains first hand information. He has to visit the various places and meet different people from whom the required data is to be collected. Since this method requires personal contact with the informants, it is necessary that the investigator is tactful, polite and courteous. He should talk to informants in the language which they understand and should ensure them that the information obtained shall not be used to their disadvantage. The data thus obtained is first hand and original in character.

Merits:

1. **General Accuracy:** This method generally ensures greater accuracy because of personal contact and supervision of the investigator.
2. **More response:** This method generally gets higher degree of response from the respondents.
3. **Supplementary information:** Besides the required information, the investigator can collect various other kinds of supplementary information.
4. **Less Delay:** Where the field of enquiry is small , the work can be finished in much less time as compared with the questionnaire method.

Demerits:

1. **Not suitable for vast geographical spread:** This method is confined only to intensive and localized investigation. Where the field of enquiry is vast, this method proves to be highly expansive and time consuming.

2. **Personal bias:** There are chances of personal bias of the investigator as he is the sole incharge of data collection. If he is biased, he can draw altogether different conclusions and thus makes a misuse of statistics.
3. **Costly:** This method needs greater time, large funds and more energy as compared to other methods.

Suitability of the method:

This method is suitable only where the scope of investigation is narrow and greater accuracy is desired.

2 Indirect Oral Investigation:

This method is used when the information required is complex and sensitive or there is reluctance or indifference on the part of the informants to give information. Under this method, instead of directly approaching the informants, the investigator interviews several other persons who are directly or indirectly in touch with information sought. For example, if a fire has broken out in a certain area, the investigator may contact several persons who may be there at the time of fire and record their statements and from that try to get a clue of the cause of the fire.

Generally, while adopting this method, a list of questions is prepared and these questions are put to various persons from whom the information is sought. It is generally adopted by enquiry committees and commissions appointed by the Govt. The success of this method largely depends upon the correctness of the information supplied by the informant. The informant, therefore should:

- i) Possess full facts of the problem.
- ii) Not be prejudiced.

Correspondents then collect the required information and transmit it to the investigator at regular intervals for further processing. This method is generally used by newspapers, radio and televisions:

Merits:

1. Since the enumerators contact the informants personally, they can exercise their intelligence, skill, tact etc. to extract correct and relevant information by cross examination of the informants.
2. This method is less expensive and requires less time for conducting the enquiry as compared with the method of direct personal investigation.

Demerits:

1. **Lack of reliability:** The data obtained by this method are not very reliable because of the possibility of personal bias and prejudice of the correspondents.
2. **Lack of precision:** This method cannot be adopted where a high degree of precision is required.

Suitability of the method:

This method is suitable in such cases where indirect sources of information are required.

3. Information from Local sources or correspondents:

Under this method, local agents, popularly known as 'correspondents' are appointed in different parts of the area from where information is to be collected. Necessary instructions are given to these correspondents. The correspondents then collect the required information and transmit it to the investigator at regular intervals for further processing. A correspondent is a person who helps in collecting statistical information. Generally he is a paid employee. This method is generally used by news papers, radio and televisions.

Merits:

1. **Useful for vast area:** This method is quite useful where the field of investigation is very vast and the information is to be collected from different parts of the country.
2. **Economical:** This method is quite economical and time saving.

Demerits:

1. **Lack of reliability:** The data obtained by this method are not very reliable because of the possibility of personal bias and prejudice of the correspondents.
2. **Lack of precision:** This method cannot be adopted where a high degree of precisions is required.

Suitability of method:

This method is employed where information is to be collected from vast geographical spread.

Mailed Questionnaire method:

When this method is employed , a questionnaire containing a number of questions pertaining to the enquiry in hand is prepared. The questionnaire is then sent to the informant by post. The necessary instructions as to how this questionnaire is to be filled are also sent. The informant after filling up the questionnaire sends it back to the investigator.

In this method, it is essential that the purpose, type and nature of enquiry must be clear to the informants. The questions should be simple, clear, a few and easy to understand and should not offend the feelings of the informants. It is also necessary that they are given an assurance that their answers will be kept confidential.

Merits:

1. **Of all the methods of collecting the information, the mailed questionnaire method is by far the most economical method in terms of time, money and manpower provided the respondents supply the information in time.**

2. Collection of original Data: Under this method original data are collected.

Demerits:

- 1. Less response: Most informants do not return the questionnaire. The informants are least interested in investigation hence there is lack of response**
- 2. Limited scope: This method is applicable only if the informant is educated. It has no application for uneducated informants.**

Suitability of the method:

This method can be used where the field of investigation is very vast and the investigator wants quick results at the low cost. However, this method presumes that informants are literate so that they can fill up questionnaire.

4. Schedules filled by enumerators:

Under this method, instead of mailing the questionnaire, it is sent to the informants through trained enumerators. The enumerators explain the objective of the enquiry to the respondents. This helps in quick and correct filling of the blanks. The schedule in this method is more exhaustive than used for mailed questionnaire. It is quite suitable for extensive investigation and is widely used in 'almost all Governmental enquiries'.

The enumerators, also known as local agents, must be polite, hardworking and well trained. They can be either paid persons or honorary.

Merits:

- 1. Greater accuracy: This method ensures greater accuracy because by cross examination, the investigator can get correct answers and the meaning of every question can be explained very clearly.**
- 2. More useful for illiterate informants: This method can be more useful where informants are illiterate and can't fill up questionnaire themselves.**

Demerits:

- 1. Costly method: This method cannot be used frequently as it is very much expensive.**
- 2. Skillful enumerators: Its success depends upon the proper selection and training of investigators. If they are untrained and careless, correct information cannot be collected .**

Suitability of the method :

This method is suitable when area to be covered is very wide , data is required regularly and the degree of accuracy desired is not very high.

Secondary data:

Secondary data are those which had already been collected or analyzed by some other agency. These data are available in public reports, private records and other sources,. These data can be had either from published or unpublished sources. These are explained as below:

(a) **Published sources:** The main published sources of secondary data are:

1. **Government publications:** Government at the center, state or local level publish a lot of statistical material. These publications can be mentioned as Statistical Abstract Of India (annual), Monthly Abstract Of Statistics , Annual Survey Of Industries ,Abstract Of Agricultural Statistics , Indian Trade Journal, Labour Gazette etc.
2. **Reports of committees and commissions:** Reports of special committees and commissions appointed by the government contain valuable data. As such the reports of National Labour Commission, Monopolies Commission, Finance Commission, National Agricultural Commission etc . are the storehouse of data.
3. **Publications and reports of Trade Associations and Chambers of Commerce:** In India , Indian Cotton mills Federation ,Jute Mills Association, Federation of Indian Chambers of Commerce, etc. publish statistics regarding trade and commerce. Stock exchanges, trade unions, banks etc. also collect and publish statistics.
4. **Journals and papers:** Newspapers like Economic Times, Financial Express etc. and Journals like Commerce, Eastern Economist, Indian Labour Gazette , Journal of Industry and Trade, Monthly Statistics of Trade etc. collect and publish statistics of their subjects.
5. **Research work done by scholars in the universities and institutions:** This is also an important source of statistical information . institutions like National Council Applied Economic Research , Foundation of Scientific And Economic Research , etc . also publish very valuable statistics.
6. **Publication of international bodies:** official publications of different international organization like U.N.O., E.C.A.F.E., I.M.F. etc . contain valuable international statistics.
7. **Market reviews and reports.**
8. **Articles published by distinguished authorities on this subject.**

(B) **Unpublished sources:**

All statistical material is not always published. This includes records maintained by various government and private offices , research studies done by some institutions, scholars etc. such sources can also be used in case of needs.

QUALITATIVE AND QUANTITATIVE DATA:

(A) **Qualitative data:** Practically it is difficult to present qualitative information in a quantitative form as they represent a specific quality such as intelligence, honesty , beauty, criminality etc. such attributes, at the most can be compared or ranked or this type of information can be tabulated as presence or absence of qualitative variable. For example data relating to intelligent and dull husband _ wives can be presented in a following manner

	Intelligent husband	Dull husband	Total
Intelligent wives	40	160	200
Dull wives	110	190	300

Total	150	350	500
--------------	------------	------------	------------

Data in this table show the data regarding 500 intelligent and dull husbands and wives. Here 110 husbands are intelligent but their wives are dull.

(B) Quantitative data : When data are expressed independently through different quantities, mutually exclusive of each other, it is known as quantitative data.

Quantitative data is generally expressed in following two forms:

(a) Discrete variable (b) Continuous variable

(a) Discrete variable: Discrete variable is one where the variates differ from each other by definite amount . Each class has a quantity in an integral form along with frequency element. Such a classification is not made on the basis of fractional values. An example of discrete is given as below:

No. of children	No. of families
0	5
1	7
2	20
3	15
4	4
5	6
6	5

(b) Continuous variable: A quantity which can assume all fractional or integral values within its specified range is known as continuous variable.

An example of continuous variable is given as below:

Weight (in abs)	No. of students
40 - 50	10
50 - 60	15
60 – 70	25
70 - 80	30
80 - 90	15
90 – 100	5
Total	100

DIAGRAMMATIC PRESENTATION OF DATA:

Diagrammatic presentation is a geometric technique of presenting numerical information through bar diagrams , rectangles, pie diagrams, pictograms or cartograms.

Objective of Diagrammatic Presentation:

Diagrammatic presentation has following objectives:

1. To make mass complex data comprehensive.
2. To have eye-catching and long –lasting effect upon reader’s mind.
3. Making vast data comparable.
4. To make quick interpretation of data.
5. Making the data universally acceptable.
6. To make the statistical information attractive and entertaining.

Merits, advantages or utility of diagrammatic presentation

A diagram helps us to visualize the whole meaning of numerical complex data at-a glance. It gives a quick and comparable summarization of vastly spread quantitative information . The main merits of such a presentation are given as follows :

1. Makes data simple and intelligible: Statistical diagram simplifies the complex data and reveals almost all the characteristics of entire data.
2. Attractive and impressive: Such an expression has an immediate effect upon human mind because numerical figures are presented through attractive diagrams. It has , therefore , greater publicity and propaganda value in modern age .
3. Makes comparison possible: Diagrams presents data in a simple, clear and effective manner and facilities comparison of values , trends and relationship.
- 4 .Lasting effects: Diagrams are eye-catching, therefore have a great memorizing effect upon human mind.
- 5.More informative: Statistical diagram clearly indicates trends and tendencies therefore considered as more informative than statistical series.
- 6.Universal utility: Diagrammatic presentation capture human mind easily , therefore it has universal acknowledgement.

Limitation of diagrams:

- 1.Misuse: Diagrammatic presentation of facts can easily be misused by the biased attitudes of investigator that may lead to misleading conclusions.
2. Rough idea: Diagrams portray only a rough idea of statistical facts .It lacks the characteristics of mathematical or exact conclusions.
- 3.Limited use: This is not an approach of analyzing statistical facts. An ordinary man cannot make correct decisions from them.
4. Misleading conclusion: A slight mistake in selecting a suitable diagram and its scale may result into misleading conclusions.

General Rules for Constructing Diagrams:

Following rules must be followed for the construction of diagrams:

1. **Attractive and Effective:** The main purpose of diagrams is to make a visual impact on the viewer therefore it is essential that they are made neat, clean and attractive.
2. **Proper size:** A proper size or proportion between height and width of the diagram should be maintained according to the availability of space.
3. **Proper title:** Each diagram should be given a suitable heading to indicate the subject matter and various facts depicted in the diagram.
4. **Proper scale:** The scale must be chosen with a great caution so that the size of diagram is reasonable. It must be mentioned on the top of the figure.
5. **Proper shades and colours:** Different colours, shades, dots, dashes, dotted lines, dashed lines, must be used while drawing diagrams so that they become elegant, attractive and eye-catching.
6. **Footnotes:** If necessary, the footnotes may be given at the left hand bottom of the diagram to explain certain points or facts, not otherwise covered in the title.
7. **Sources of data:** Sources of data must be appended at the bottom of the diagram so as to make it more reliable. Each diagram should also be given a number for ready reference and comparative study.
8. **Accuracy:** A diagram must be drawn accurately according to the data given. Diagrams should not be made attractive and impressive at the cost of accuracy of the original data.
9. **Simplicity:** Diagrams should be made as simple as possible so that they are easily understood even by a layman who does not have any mathematical or statistical background.
10. **Economy:** An indicator of ideal diagram is that how quickly and easily effect the brain of its reader. It should bear a minimum cost and labour-energy in its construction.

TYPES OF DIAGRAM:

Following are the main types of diagram, used in displaying the numerical facts:

1. One Dimensional Diagrams or Bar Diagrams.
2. Two Dimensional Diagrams.

One Dimensional Diagrams or Bar Diagrams

One dimensional diagrams are those diagrams in which only length of diagrams is considered. It can be drawn in the form of line or in various types of bars. Usually it is famous with the title of 'bar diagrams'.

A Bar usually means thick-wide lines or some rectangular lengths, neglecting the role of breadth. While drawing bar diagrams following rules must be kept in mind.

1. All the bars drawn on a single scale should be uniform.

2. Proper and uniform spacing should be given between different bars.
3. A suitable scale should be selected so as to cover all the given values.
4. Magnitude of each bar should be written on the top of each bar.
5. All the bars should be made on same base line.
6. Bars can be drawn vertically or horizontally, but vertical bars should be preferred.

Types of bars:

Bars can be drawn in following six ways.

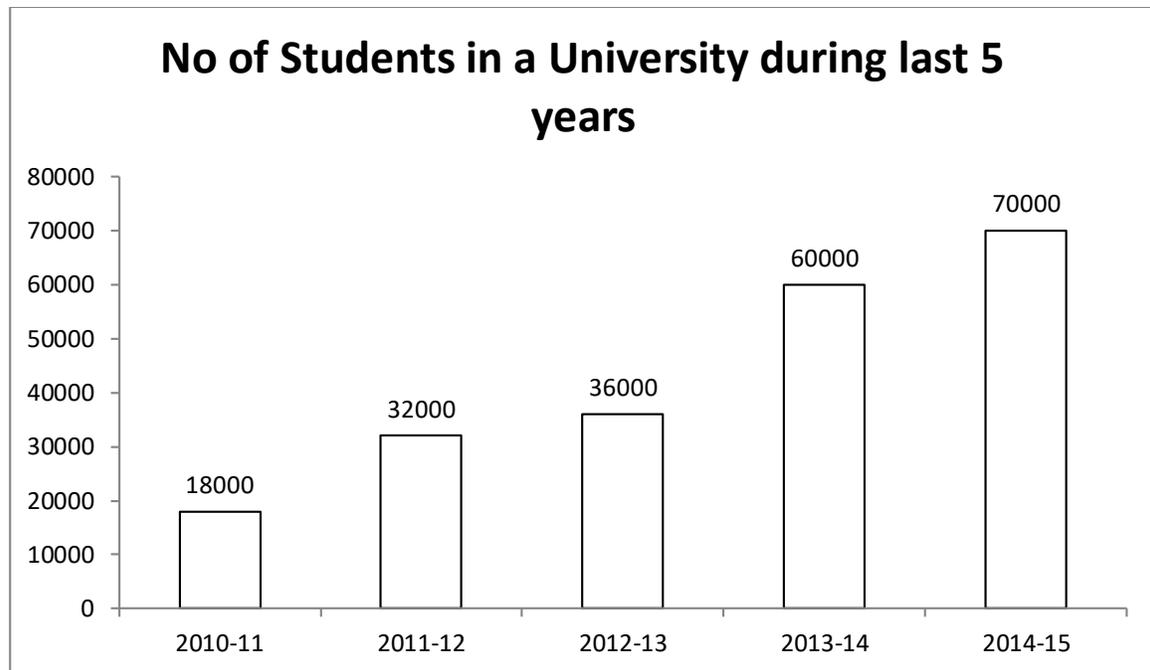
1. Simple bar diagrams
2. Multiple bar diagrams
3. Sub divided bar diagrams
4. Percentage bar diagrams
5. Broken bar diagrams
6. Due-directional bar diagrams.

Bar Diagram:

Bar Diagrams are one of the easiest and the most commonly used devices of presenting most of the business and economic data. They consist of a group of equidistant rectangles, one for each group or category of the data in which the values or the magnitudes are represented by the length or height of the rectangles, the width of the rectangles being arbitrary and immaterial. These diagrams are called one-dimensional because in such diagrams only one dimension viz, height (or length) of the rectangles is taken into account to present the given values.

Illustration: The following data represents the no of students in a university during the last 5 years ending 2014-15 as below:

Year:	2010-11	2011-12	2012-13	2013-14	2014-15
No. of Students in 000's:	18	32	36	60	70



GRAPHIC PRESENTATION OF DATA

Graphic presentation of data is a geometric expression of a statistical information .

Graphic presentation represents data in the form of curves or lines on a graph paper . It supplements the study of tabulation and a unique statistical tool to impress the masses.

MERITS OR UTILITY OR IMPORTANCE OF GRAPHIC PRESENTATION:

1. **Graphs are attractive, interesting and effective:** As compared to the tabulation of data graphic presentation of data is more attractive, interesting and effective as it makes the statistical information within the reach of a common man.
2. **Saving of time and labour:** Graph makes the statistical series visible at a glance by which it saves a lot of time and labour of the statistician and the reader.
3. **Presentation of time series:** Graphic presentation can present the time series in an effective manner. It helps in analyzing its past and future trends.

4. **Helpful in comparison:** Graphs are useful in presenting the statistical information in a comparative form. They bring out hidden fact and relationship which are further useful for analytical thinking.
5. **Location of averages and other measures:** With the help of graphs , we can determine positional averages, dispersion through Lorenz curve and correlation through graphic presentation of variables.
6. **Forecasting and projections:** Graphs help in determining the secular trend in a time series and also to project the future result.
7. **Useful for investigation:** Graphic form is also useful device to suggest the investigator for directing his investigation. He can easily check the problem at every step of investigation for its further correcting.

Limitations of graphs

1. A graphic curve simply shows the tendency of a variable to rise or fall. It does not provide any information about actual values.
2. There is a possibility of drawing wrong conclusions from graphic presentation. Two different scales will show different fluctuations in the data.
3. Graphs are difficult for a layman to understand . Accordingly these are of little value to them .
4. It is very difficult to show minor differences with graphs.
5. Graphic presentation requires skilful and expert investigator to use them.

Difference between diagrams and graphs

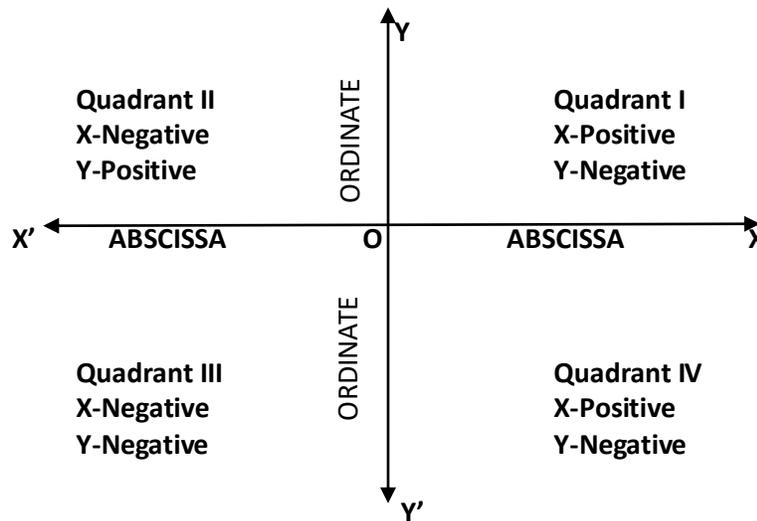
Diagrams and graphs both present the data in a simple , attractive and effective way, yet both the techniques have certain points of differences.

Following points clarify the difference between the two:

1. **Use of paper:** For graph ,graph paper is used which helps in studying mathematical relationship between two variables, whereas diagrams are generally constructed on plain paper for making a simple comparison.
2. **Forecasting:** Graphic presentation helps in making forecasting about future events whereas diagram simply presents the given the given facts.
3. **Accuracy:** Graphs present the facts more minutely and precisely, on the other hand diagrams generally provide an approximate and rough idea of numerical information.
4. **Statistical tool:** Diagram is simply a presentation of statistical fact, it can't be further used for any analytical study . On the other hand graphs are a useful device for locating averages dispersion, correlation , interpolation and time series analysis.
5. **Construction:** The construction of graphs is easier as compared to diagrams.
6. **Bars, squares and lines:** Diagrams are constructed with the help of bars , rectangles and squares where as graphs are constructed with the help of lines on a graph paper.

Construction of a Graph:

The system of graphics was introduced by William Playfair of Scotland and Sir Willard C. Brinton. A brief technique of constructing a graph is given below:



Graphs are drawn on a special type of paper known as graph paper which has a fine network of horizontal and vertical lines; the thick lines for each division of a centimeters or an inch measure and thin lines for small parts of the same. On this paper, two simple lines are firstly drawn which intersect each other at right angles. The lines are known as coordinate axis and the point of intersection is known as point of origin or the zero point. The horizontal line is called the axis of X or 'abscissa' and the vertical line, the axis of Y or 'ordinate'. The alternate appellations are X – axis and Y – axis respectively. It is shown through the adjoining diagram.

In the above figure, O is the point of origin, XOX' is the axis of X or the abscissa and YOY', the axis of Y or the ordinate. Both negative and positive values can be depicted on the graphs. Distances measured towards the right or upward from the origin are positive and those measured towards the left or downwards are negative.

The whole plotting area is divided in four quadrants.

In Quadrant 1, both the values of X and Y are positive,

In Quadrant II, X is negative and Y is positive.

In Quadrant III, both X and Y is negative,

In Quadrant IV, X is positive and Y is negative

However in business statistics Quadrant 1 i.e. YOX is generally used for depicting statistical information. Moreover, it is convenient to take the 'independent variable' on the horizontal scale and the 'dependent' on the vertical scale.

General Rules For Constructing Graphs:

1. **Neatness:** Graphs are visual aids for the presentation of statistical data, therefore they must be made neat and clean.
2. **Heading and footnotes:** Each Graph should be given a suitable title and required footnotes must be given to make them self – explanatory.
3. **Structural framework:** The position of axis should be chosen so that graph gives an attractive and proportionate get up.
4. **Choice of scale:** A proper care should be taken while choosing a scale between X – axis and Y – axis. So that entire data is accommodated on the given space without crowding.
5. **False Base line:** False base line is a method of magnifying minor fluctuations on a graph paper taking a higher value as origin on X – axis or on Y – axis. A cut is made between zero and lowest value to indicate the value of the origin.
6. **Use of different lines:** In business statistics different type of statistics are used, therefore, such statistics must be distinguished from each other by drawing different lines such as thick lines (-), broken lines (- - -) , dotted lines (.....) etc.
7. **Ratio or logarithmic scale:** In order to display proportional or relative changes in the magnitudes, the ratio or logarithmic scale should be used instead of natural or arithmetic scale.
8. **To draw a line:** While drawing graphs points must be plotted and then join with each other. It should make a line or curve. Such a line must be drawn in a uniform way.
9. **Index:** If there are more than one variable plotted on a graph, an index should be made to identify different lines drawn on it.
10. **Joining lines:** On graph paper, all the points should be joined with each other to show them in a line. Such a joining should be identical on a graph.

Types of Graphs:

1. Time series graphs or Historigram.
2. Frequency Distribution graphs.

Most commonly used graphs for charting the frequency distribution are following:

- | | |
|---------------------|--|
| 1. Histograms | 2. Frequency polygon |
| 2. Frequency curves | 4. Ogives or cumulative frequency curves |

Histograms:

Definition:

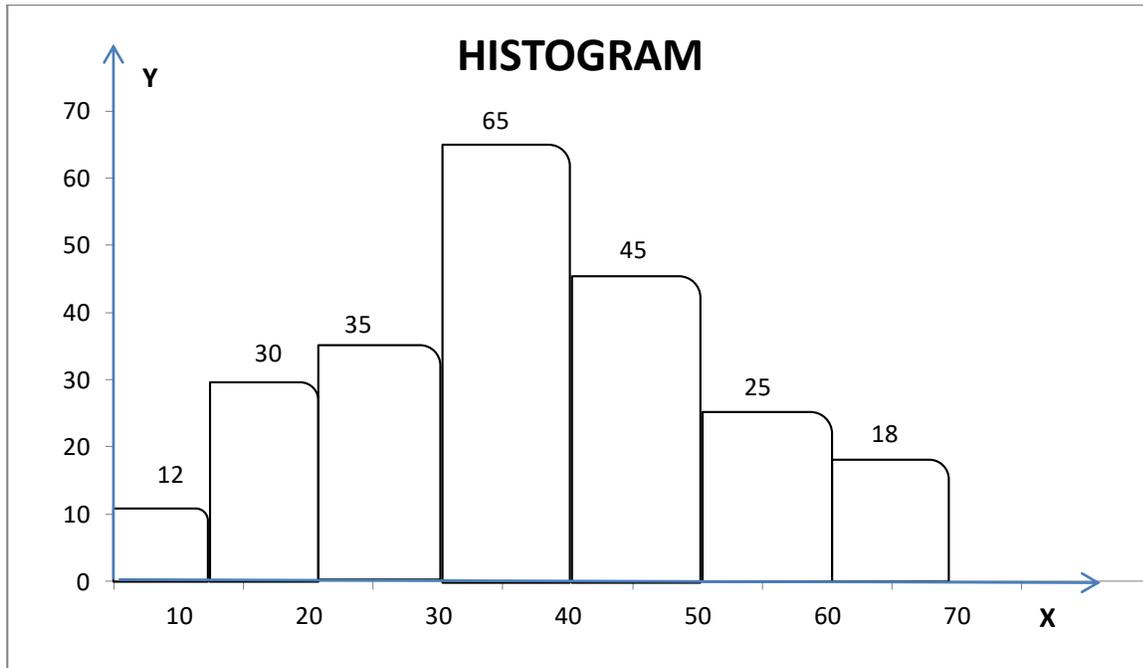
A histogram is a graphical representation of a numerical data. It is an estimate of the probability distribution of a continuous variable (quantitative variable) and was first introduced by Karl Pearson. It is a two dimensional as both height and breadth are taken into account. So it is constructed on natural scale by taking variable on X- axis with equal differences representing each class and frequency on vertical axis so as to depict adjacent vertical rectangles conveniently.

Illustration: Draw histogram for the following frequency distribution:

Marks: 0-10 10-20 20-30 30-40 40-50 50-60 60-70

No. of Students: 12 30 35 65 45 25 18

Solution:



Ogives or cumulative frequency curve:

Definition:

“Ogive, pronounced as ogive is a graphic presentation of cumulative frequency curve. Such a curve may show a falling or rising trend on the graph paper.

Ogives are constructed from cumulative frequency table, which are generally made in ‘less than’ or ‘more than’ form, therefore, there are following two types of ogives:

1. Less than Ogives
2. More than ogives

Less than Ogives: This consists in plotting the less than cumulative frequencies against the upper class boundaries of the respective classes. These points are joined with free hand to give less than ogive.

Illustration : Following distribution shows the marks obtained by a class of 100 students:

Marks	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70

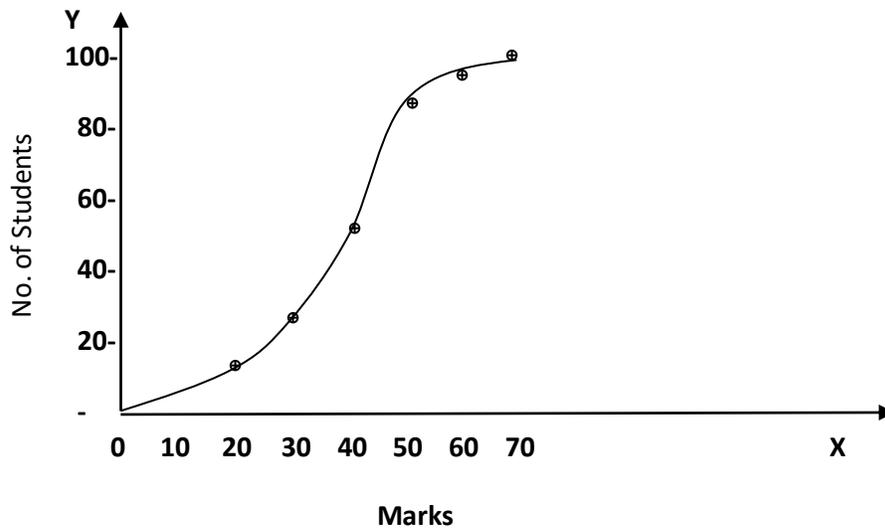
Frequency	10	15	30	32	8	5
-----------	----	----	----	----	---	---

Construct less than ogive

Solution: First of all make cumulative frequency table from data given:

Cumulative frequency table

Marks	No. of students
Less than 20	10
Less than 30	25
Less than 40	55
Less than 50	87
Less than 60	95
Less than 70	100



More than Ogive: In more than Ogive, 'more than' cumulative frequencies are plotted against the lower boundary of the respective classes. It has a decreasing or downward sloping curve left to right.

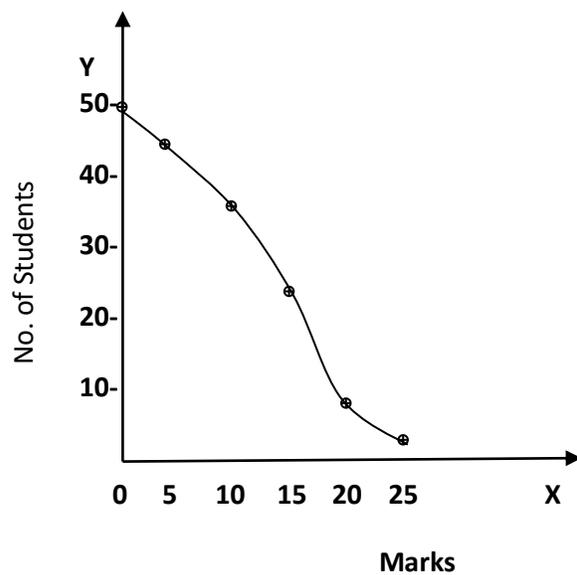
Illustration : From the following data construct more than ogive.

Marks	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
-------	-------	--------	---------	---------	---------	---------

No. of students	5	7	15	16	5	2
-----------------	---	---	----	----	---	---

Solution: Let us make 'more than' cumulative frequency:

Marks	No. of Students
More than 0	50
More than 5	45
More than 10	38
More than 15	23
More than 20	7
More than 25	2



Sumaya Showkat
Faculty Member
GDC Boys Anantnag

