STATISTICS

PRIMARY DATA

Primary data are those which are collected for the first time and they are original in character.

For the collection of primary data, the investigator may choose anyone of the following methods:

1. Direct personal observation
2. Indirect oral interview
3. Information through agencies
4. Mailed Questionaries
5. Schedules sent through enumerators.

1. DIRECT PERSONAL OBSERVATION :

Under this method, the data is collected by the investigator personally. The investigator must be a keen observer, tactful and courteous in behaviour. He asks or cross - examiners the informant and collects necessary information. It is original in character.

2. INDIRECT ORAL INTERVIEW :

When the informant is reluctant to supply information, the method of indirect oral

investigation can be followed.

* Under this method, the investigator approaches the witnesses or third parties, who are in touch with the informant.

* The enumerator interviews the people, who are directly (or) indirectly connected with the problem under study.

3. INFORMATION THROUGH AGENCIES:

* Under this method, local agents or correspondents will be appointed.

* They collect the information and transmit it to the office or person.

* They do this according to their own ways and tastes.

* This system is adopted by newspapers, periodicals, agencies etc.,

* The informants are generally called "correspondents."

4. MAILED QUESTIONARIES:

In this method, a Questionnaire consisting of a list of questions pertaining to the enquiry is

Prepared.

* This Questionnaire is sent to the respondants, who are expected to write the answers in the blank spaces.

* A covering letter is also sent along with the questionnaire, resquesting the respondants to extend their full co-operation by giving the correct replies and returning the questionnaire duty filled in time.

* To get quick and better response, the return Postage expense is borne by the investigators, by sending a self-addressed and stamped envelope.

* This method is adopted by research workers, private individuals, non-official agencies and state and central governments.

5. SCHEDULES SENT THROUGH ENUMERATORS:

* It is the most widely used method of collection of primary data.

* A number of enumerators are selected & trained.

* They are provided with standardised questionnaire.

* Specific training and instructions are given to them for filling up schedules. Each enumerator will be in charge of a certain area.

* The investigator goes to the informants along with the questionnaire to and gets replies to the Questions in the schedule. and record their answers.

* He explains clearly the object and purpose of enquiry.

* In former method, the Questionnaire is sent to the informants, whereas, in this method, the investigator carries the schedule to the informant.

* This method is used by public organisations and research institutions.

SECONDARY DATA

Secondary data are those which are already collected by someone for some purpose and are available for the present study.

The various sources of secondary data can be divided into two broad categories:

1. Published sources
2. Unpublished sources

1. PUBLISHED SOURCES:

Various Governmental, international and local

agencies publish statistical data, and chief among them are:

(a) International publications:

International agencies and international bodies publish regular and occassional reports on economic and statistical matters.

They are the I.M.F., the I.B.R.D., the I.C.A.F.E., and U.N.O., etc.,

(b) Official publications of central and state government:

Departments of the union and state governments regularly publish reports on a number of subjects. They gather additional information. some of the important publications are Census of india, the Reserve bank of India Bulletin, statistical Abstracts of states, Agricultural statistics of india, Indian trade journal, etc.,

(c) Semi - official publications:

Semi - government institutions, like municipal corporation, District Board, Panchayat, etc., publish reports,

(d) Publication of Research Institutions :

Indian statistical Institutions (I.S.I.)

Indian council of Agricultural Research (I.C.A.R.)

Indian Agricultural statistics research institute (I.A.S.R.I.) etc., Publish the findings of their research programme.

(e) Publication of commercial and Financial Institutions.

(f) Reports of various committees and commissions appointed by the governments :

(g) Journals and Newspapers :

Current and important materials on statistical and social - economic problems can be obtained from journals and newspapers.

2) UNPUBLISHED SOURCES :

There are various sources of unpublished data. They are the records maintained by various government and private offices, the researching carried out by individual research scholars in the universities or research institutes.

TYPES OF CLASSIFICATION

There are four important types of classification, They are:

a. Geographical
b. Chronological
c. Qualitative
d. Quantitative

a. GEOGRAPHICAL CLASSIFICATION

In geographical or spatial classification, the basis of classification is the geographical or locational differences between various items in the statistical data like states, districts, cities, taluks, regions, zone, area, etc.,

b. CHRONOLOGICAL CLASSIFICATION

This type of statistical data is classified according to the time of its occurence, such as years, months, weeks, days, hours, etc.,

Time series are called chronological classification.

C. QUALITATIVE CLASSIFICATION

When the data are classified according to some quality or attributes is termed as Qualitative classification (or) attribute (or) descriptive attributes.
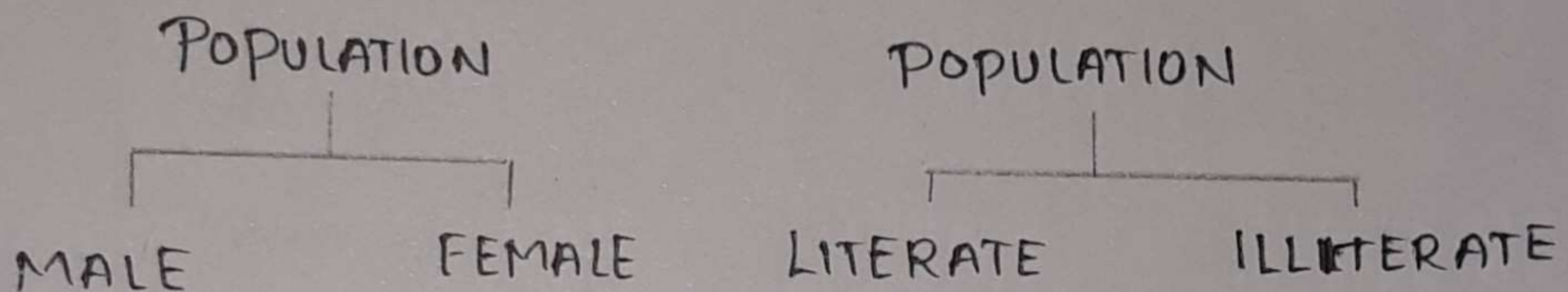
This again be classified into
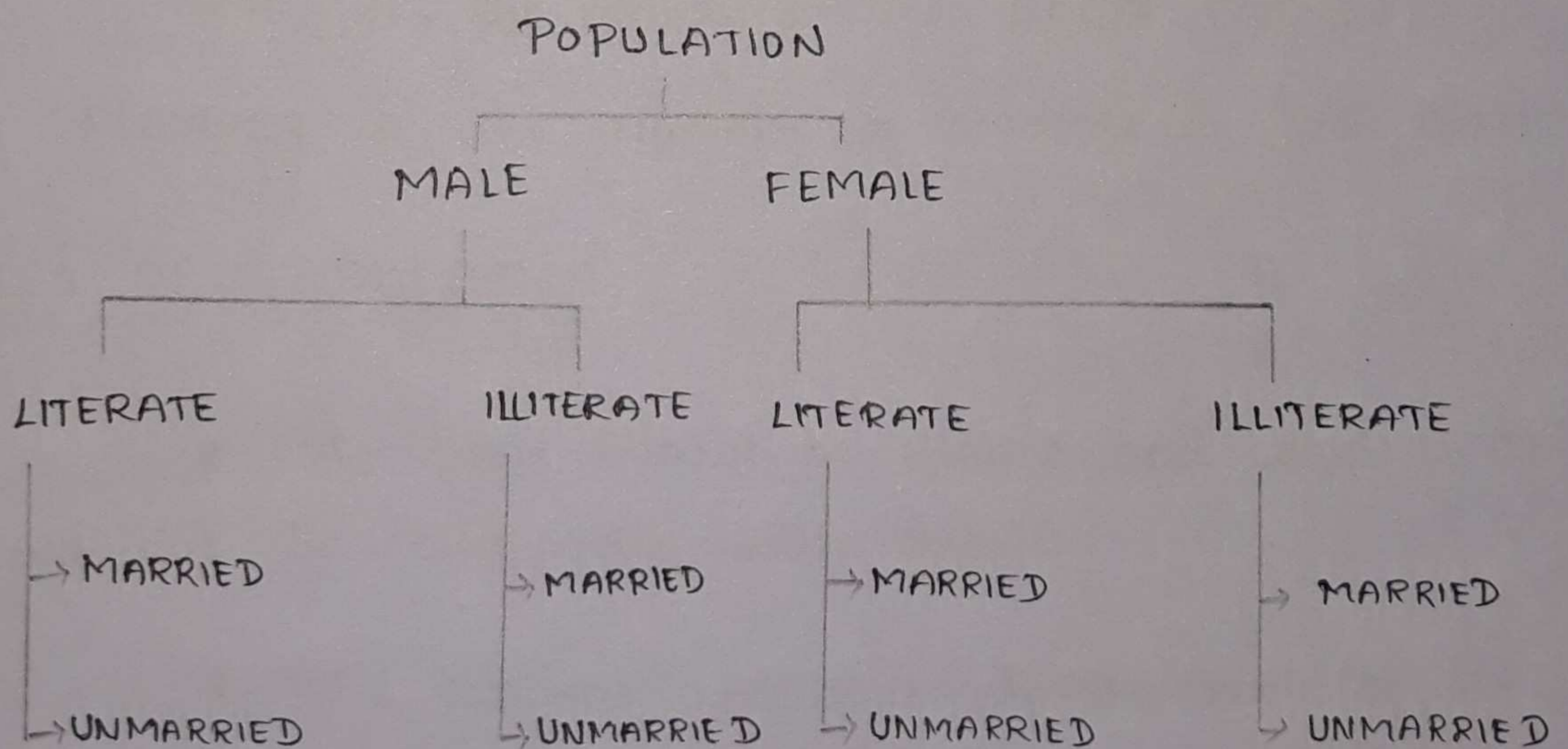* simple
* Manifold

* SIMPLE CLASSIFICATION :

This classification is normaly dichotomy or two-fold.

for example,

```
     POPULATION                    POPULATION
         |                             |
   ┌─────┴─────┐                 ┌─────┴─────┐
  MALE       FEMALE          LITERATE    ILLITERATE
```

* MANIFOLD CLASSIFICATION:

In this , the universe is classified on the basis of more than one attribute at a time.

```
                        POPULATION
                   ┌────────┴────────┐
                 MALE              FEMALE
            ┌─────┴─────┐      ┌─────┴─────┐
        LITERATE   ILLITERATE LITERATE  ILLITERATE
           │           │         │          │
        →MARRIED    →MARRIED  →MARRIED   →MARRIED
           │           │         │          │
        →UNMARRIED →UNMARRIED →UNMARRIED →UNMARRIED
```

d. QUANTITATIVE CLASSIFICATION

If the data are classified according to some characteristic which is capable of quantitative measurement like age, income, height, weight, price, production, sales, Profits, etc., it is called Quantitative classification.

MARKS	No. of STUDENTS
10 - 20	10
20 - 30	7
30 - 40	13
40 - 50	18
50 - 60	12

There are two elements, * Variable

* Frequency.

VARIABLE . ie, the marks in the above example.

FREQUENCY ie, the number of students in each class.

RULES OF TABULATION:

* The table should be simple and compact. It should not be overloaded with details.

* The captions and stubs in the table should be arranged in a systemic manner. It is easy to read the important items.

* It should suit the purpose of the investigation.

* The unit of measurements should be clearly defined and given in the tables;

* Figures may be rounded off to avoid unneccessary details in the table. But a foot-note must be given.

* Suitable approximation may be adopted.

* A miscellaneous column should be added to include unimportant items.

* A table should be complete and self-explanatory.

* A table should be attractive to draw the attention of readers.

* As it forms a basis for statistical analysis, it should be accurate and free form all sorts of errors.

* Abbrevations should be avoided.

* Do not use ditto marks that may be mistaken.

* Proper lettering will help to adjust the size of the table.

* If it is a big table, it will lose its simplicity and understanding; and in such a case break it into two or three tables.

_____

UNIT-II

DIAGRAMMATIC PRESENTATION

TYPES OF DIAGRAM:

These are various diagrammatic devices by which statistical data can be presented. We shall discuss a few of them, which are mostly used. The following are the common type of diagrams:-

(i) ONE - DIMENSIONAL DIAGRAM

In one - dimensional diagram, the length of the lines or bars is considered and the width of the bars is not taken into consideration. The term 'BARS' means a thick wide line. The following are major types:

a) LINE DIAGRAM

This is the simplest of all the diagrams on the basis of size of the figures. Heights of bars or lines are drawn. The distance between lines is kept uniform. It makes comparison easy. This diagram is not attractive; hence it is less important.

b) SIMPLE BAR DIAGRAM

A simple bar diagram can be drawn

either on horizontal or vertical bars. Bars on horizontal base are more common. A bar diagram is simple to draw and easy to understand. In business and economics, it is commonly used.

C) MULTIPLE BAR DIAGRAM

Multiple bar diagram are used to denote more than one phenomenon. Multiple bars are used for direct comparison between two values. The bars are drawn side by side.

d) SUB-DIVIDED BAR DIAGRAM

The bars is subdivided into various Parts in Proportion to the values given in the data and may be drawn on absolute figure, or percentages. each component occupies a Part of the bar proportional to its share in the total. To distinguish different components from one another, different colours or shades maybe given.

e) PERCENTAGE SUB-DIVIDED BAR DIAGRAM

This comparison is made on a relative basis and not to used to represent absolute value. The various components are expressed as percentage to total for dividing the bars these % are cumulated. In this one, the bars are all of equal height.

f) OTHER BAR DIAGRAM

* DEVIATION BAR

Deviation bar diagram is used to depict the net deviations in different values. which have both Positive or negative values. Positive are shown above the base line and negative below the base line.

* BROKEN BAR

In certain cases, we may come across data which contain very wide variations in values. Very small or very large. In order to provides adequate & reasonable shop to the smaller bars, the larger bars may be broken at the top. The value to each bar is written at the top of the bars.

TWO - DIMENSIONAL DIAGRAMS

In this, the area of the diagram represents the data, ie, the length & breadth are considered.

The important types are:

* RECTANGLES

Rectangles are used when two or more magnitudes with different components have to be compared. The area of rectangular are kept to the values.

It may be two types.

a) PERCENTAGE SUB-DIVIDED RECTANGULAR DIAGRAM:-

In such a diagram, the width of rectangular is kept according to the porportional of the values components of the values are converted into percentage and rectangular divided according to them.

b) SUB-DIVIDED RECTANGLE :-

Sub-divided rectangle such diagram occured to show some related Phenomenon.

* ANGULAR OR PIE DIAGRAM:

The pie diagram ranks high in under-standing just as we divide a bar or a rectangle to show its components, a circle can also be divided into sectors. As there are 360 degrees at the centre, Proportionate sectors are cut taking the whole data equal to 360 degrees. This will be clear from the following illustration.

GRAPHIC PRESENTATION:

A graph is visual form of the presentation. Graphs are drawn on a special type of paper known as graph paper.

GENERAL RULES :-

* Every graph must have a title indicating the facts Presented by the graph.

* It is necessary to plot the independent variables on the horizontal axis and dependent variables on the vertical axis.

* Problem axis regarding the choice of a suitable scale. The choice must accomodate the whole data.

* The showing proportional relative changes in the magnitudes, the ratio or logarithmetic scale should be used.

* The graph must not be over crowded with curves.

* If more than one variable is plotted on the same graph it is necessary to distinguish them by different lines viz, dotted lined, broken lines, dots, dot with dash, thick thin dashed lines etc.,

* Index should be given to show the scales of & the meaning of different curves.

* All lettering must be horizontal.

* Source of information should be mentioned as foot-note.

DIFFERENCE BETWEEN DIAGRAMS AND GRAPHS

DIAGRAMS	GRAPHS
• Ordinary paper can be used.	• Graph paper is needed.
• It is attractive & is easily understandable.	• It needs some efforts to understand.
• It cannot be used for interpolation & extra polation technique.	• It is helpful in intrapolation & extrapolation technique.
• Median and mode cannot be estimated.	• The value of median & mode can be estimated.
• It is used for comparison only.	• It represents a mathematical relationship b/n the 500 variables.
• Data are represented by bars and rectangles.	• Data are represented by Points or lines of different kinds, dots, dashes, etc.,
• Diagrams are used for Publicity as they are attractive. They give only approximately information. TO a statistician or a research diagrams are not helpful in analysis.	• Graphs are very much useful to statisticians or researchers in analysis.

The most commonly used graphs are,

    (i) Histogram

    (ii) Frequency Polygon

    (iii) Frequency curve

    (iv) Ogives or cumulative frequency curve.

(i) HISTOGIRAM

One of the most important and useful methods of presenting frequency distribution of continuous series is known as Histogram. In this, the magnitude of the class interval is plotted along the horizontal axis and the frequency on the vertical axis. This process will give us rectangle are proportional to their frequencies. Histogram is known as "BLOCK DIAGIRAM" or "STAIRCASE CHART".

When the distribution has an unequal class interval, the frequency of each class is to be adjusted for making the adjustment, consideration the lower class interval & make adjustment to the upper class interval. If three times the class interval the lower, then divide its frequency by three times, and so on.

(ii) FREQUENCY POLYGION:

A simple method of so smoothing the histogram is to draw a frequency polygon.

This is connected by the midpoint of the top of each rectangle with the mid point adjustment rectangle by straight lines. In this, Mode can easily be find out.

(iii) FREQUENCY CURVE :

A frequency curve is drawn by smoothing the frequency polygon. It is smoothed in such a way that the sharp turns are avoided. A frequency polygon, if smoothed further, so as to minimise sudden changes, results into a continuous smooth curve known as frequency (or) smooth frequency curve. The curve should begin and end at the base line.

(iv) OGIVES OR CUMMULATIVE FREQUENCY CURVES :

When cummulative frequencies are plotted on a graph then the frequency curve obtained is called " OGIVES " or " CUMMULATIVE FREQUENCY CURVE ". Ogive determine median quartiles, percentiles, etc.,

There are two methods of constructing ogives, viz,

* Less than ogives
* More than ogives

* LESS THAN OGIVES :

In less than ogives, the less than cummulative frequencies are plotted against upper class boundaries of the respective classes. Then the points are joined by a smooth free hand curve and has the shape of an elongated s($\int$).

* MORE THAN OGIVES :

In more than ogives, the more than cummulative frequencies are plotted against the lower class boundaries of the respective classes then the points are joined by a smooth free hand curves and has the appearance of an elongated s, upside down.

One can locate the median by drawing two Ogives on less than and the other more than, for this, one has to calculate less than and more than cummulative frequencies.

As mentioned above, from the intersection Point of these ogives, on the perpendicular line touches of x-axis is the value of median.