ENORMOUS

Information

Meaning Of Statistics:- The word Statistics has been derived from a Latin word Status or an Italian word Statistia both of these words mean a political state, originally meant information useful to the state. For example information about the size of population and armed forces. But this word has now acquired different meanings.

Firstly, It is used in the plural sense to refer to numerical facts in any field of study. These facts are collected in a systematic manner with a definite purpose in view. Often, we read about statistics of deaths and births, price statistics, agricultural statistics etc. we also use the word Data to refer to statistics in this sense.

Secondly, the word Statistics is used in the singular sense. In this sense, it refers to the science comprising methods and techniques which are used in the collection, presentation, analysis and interpretation of numerical data.

Thirdly, the word Statistics is used in a technical sense, as plural of statistic, by statistic, we mean a quantity calculated from sample observations. For example the sample mean, the sample standard deviation, sample proportion etc.

Characteristics of Statistics(as Data):- Statistics (as Data) numerically expressed facts and figures have the following characteristics.

1. Statistics are aggregate of facts, a single number however important it may be, is not statistics. For example the height of a single person is not statistics, but a collection of heights say 100 students would be statistics.
2. Statistics are affected to a great extent by multiplicity of causes. For example the heights and weights of individuals are affected by their parents height, climatic condition, nature of diet etc.
3. Statistics are numerically expressed. Qualitative expressions like poor, good, young, beautiful etc. Do not form a part of statistics.
4. Statistics are enumerated or estimated according to a reasonable standards of accuracy, which is determined by the purpose with which statistics are collected. For example the height of a group of persons can be measured accurate to the nearest centimetres, and distance between two cities can be measured accurate only to the nearest kilometres or meters etc.
5. Statistics are collected in a systematic manner. Without a proper plan, order or design cannot be accurate.
6. Statistics are collected with a definite object in view. Data collected without any purpose would be of no use.
7. Statistics are capable of being placed in relation each to other so that they are comparable i.e statistics (Data) should be homogeneous.

Limitations of Statistics:-

1. Statistical results are true only on average or in the long run because it is basically designed to measure the group characteristics and does not study the individual cases.
2. Statistics does not study qualitative problems directly. It application is limited to those facts which are numerically expressed.
3. If sufficient care is not exercised in all stages of a statistical inquiry, the results may be fallacious and misleading.
4. Only a person who has an expert knowledge of statistics can handle statistical data efficiently.
5. Statistics provides only tools for analysis. It cannot change the cause and effect relationship. For example if statistical result shows that the relationship between crime and illiteracy is very high. Then it's the job of government to improve education and not the statistics that its application has resulted such a dangerous conclusion.

Importance of Statistics:-

Statistics has come to play an important role in almost every field of human activity. The following functions and uses of statistics in most diverse fields serve to indicate its importance.

1. Statistics assists in summarizing the large sets of data in a form that is easily understandable.
2. Statistics assists in the efficient design of laboratory and field experiments as well as surveys.
3. Statistics assists in a sound and effective planning in any field of inquiry.
4. Statistics assists in drawing general conclusions and making predictions of how much of a thing will happen under given conditions.
5. Statistical techniques being powerful tools for analysing numerical data, are used in almost every branch of learning. In the biological and physical sciences, genetics, Astronomy, Physics, Geology etc. are the main areas where statistical techniques have been developed and are increasingly used.

Branches of Statistics:-

The science of Statistics may be classified into two main branches.

(a) Pure Statistics (b) Applied Statistics

(a) Pure Statistics:- It is concerned with the creation and development of theories, rules and principles applicable to different specific problems in different fields.

(b) Applied Statistics:- It is concerned with the proper application of statistical methods in different fields to investigate any problem.

This branch of Statistics is further classified into the following sub-branches.

(a) Descriptive Statistics (b) Inferential Statistics

(a) **Descriptive Statistics:-** It is that branch which deals with collection of data, its presentation in various forms such as tables, graphs and diagrams and finding averages and other measures which would describe data.

(b) **Inferential Statistics:-** It is that branch in which statistical techniques are used for analysis of data, making estimates and drawing conclusions on the basis of sample about the population.

Population and Sample: In Statistics the word population refers to the whole aggregate or totality of all individuals' members or objects, whether finite or infinite, relevant to some characteristic of interest e.g. the number of students in a college, the amount of blood in a human body, area, animals etc.

Sample is a representative part or a subset of a population.

Parameter and Statistics: A numerical characteristics, computed from the observations of the population is called population parameter and is generally denoted by Greek letters such as μ, δ, ρ etc.

A numerical characteristics computed from sample observations is called statistics and are denoted by Latin letters such as $\bar{X}$, S, r etc.

Variable:- A characteristic that varies from one individual to another or from one object to another is called a variable. For example age, weight or income are variables, as it varies from person to person.

Types of variables:- Variables are usually of two types

(a) Quantitative variable (b) Qualitative variable

(a) ***Quantitative variable:-*** A variable which is capable of assuming a numerical value. For example height of a person, weight or number of students in a class. Quantitative variables can further be divided into two types.

(i) Discrete variable (ii) Continuous variable

(i) ***Discrete variable:-*** A discrete variable is one that can take only isolated points on a number line or whole numbers. A discrete variable represents count data such as the number of children in a family, the number of chairs in a class room etc.

(ii) ***Continuous variable:-*** A continuous variable is one that can take all possible values in an interval on the number line. For example plant height, temperature at a place etc. A continuous variable represent measurement data or takes values from a set of real numbers.

(b) ***Qualitative variable:-*** A qualitative variable also known as categorical variable is one which is not capable of taking numerical measurements. An observation is made when an individual is allocated to one of several mutually exclusive categories. Observations falling in each class can only be counted. For example education, sex, eye colours, poverty, intelligence etc.

Types Of Data By Source:- The source of data means from where the data is collected. Data may be classified into two ways by source.

(i) Primary data (ii) Secondary data

(i) Primary data:- Data that have been originally collected (raw data) and have not under gone any sort of statistical treatment, are called primary data.

(ii) Secondary data:- The data that have undergone any sort of treatment by statistical methods at least once i.e. the data have been collected, classified, tabulated or presented in some form for a certain purpose, are called secondary data.

Methods Of Collection Of Primary Data :- The primary data may be collected by any of the following methods.

1. **Direct personal investigation:-** In this method, an investigator collects the information personally from the individual concerned. Since he interviews the informants himself, the information collected is generally considered quite accurate and complete. But this method may prove very costly and time consuming when the area to be covered is vast.
2. **Indirect investigation:-** Sometimes, the direct sources do not exist or the informants would not either disclose the facts at all or give intentionally wrong information. In such a situation, information is collected from third parties or witnesses supposed to know the facts directly or indirectly. This method is used when the information to be collected is complex or the informants are reluctant to disclose the true facts.
3. **Collection through questionnaires:-** In this method a standard list of questions directly related to the purpose of inquiry is prepared. This list or schedule known as questionnaire, is sent to the informants by mail requesting them to return it back to the investigator after answering all the given questions. This is the most popular method of collecting primary data. But the questions framed for this purpose should be simple, clear, relevant, and answerable and designed in a logical sequence. This method is fairly cheap, expeditious, cover vast area and have a greater scope.
4. **Collection through enumerators:-** In this method information is collected through trained enumerators who assist the informants in making the entries in the questionnaires (called schedules) correctly. This method gives the most reliable information and is considered as the best method when a large scale governmental inquiry is to be conducted. But it's too much costly.
5. **Collection through local sources:-** In this method there is no formal collection of data but the agents or local correspondents are directed to collect and to send the required information using their own judgements as to the best way of obtaining it.

Collection Of Secondary Data:- The secondary data may be obtained from:

(i) Official Sources (ii) Semi-official (iii) Publication
(iv) Technical and Trade Journals (v) Research Organisation etc.

Processing Operation:- After the brief introduction, we can now proceed with the explanation of processing operations.

1. **Editing:-** Editing of data is a process of examining the collected raw data to detect errors and omissions and to correct these when possible. As a matter of fact, editing involves a careful scrutiny of the completed questionnaires or schedules. Editing is done to assure that the data are accurate, consistent, uniformly entered, as completed as possible and have been well arranged to facilitate classification.
2. **Coding:-** Coding refers to the process of assigning numerals or other symbols to answer so that responses can be put into a limited number of categories or classes.. Such classes should be appropriate to the research problem under consideration.
3. **Classification:-** Most research studies result in a large volume of raw data which must be reduced into homogeneous groups if we are to get meaningful relationships. Therefore it is necessary to organize it into a meaningful and readily comprehensible form to make possible and facilitate further statistical analysis. That arrangement of data in groups or classes on the basis of common characteristics is called classification. Data having common characteristics are placed in one class and this way the entire data get divided into a number of groups or classes.

Types of classification:- The classification of data depends on the purpose and nature of the phenomenon involved. There are four important types of classification

(a) **Geographical or spatial classification:-** The arrangement of data on the basis of difference in geographical location that is area wise, region wise or district wise. The division of literacy rate of the four provinces of Pakistan is an example of this type.

(b) **Chronological or Temporal classification:-** The arrangement of data on the basis of difference in time is called temporal classification. For example the population of Pakistan in censuses of 1951,1961,1972,1981 and 1998 etc. The whole data is then called time series.

(c) **Quantitative classification or according to class intervals:-** The data relating to quantitative variables which are measured through some statistical units such as age, weight, production, temperature, marks are classified on the basis of class-intervals or differences in quantity(magnitude) is called quantitative classification.

(d) **Classification according to attributes (Qualitative):-** The arrangement of data on the basis of differences in quality. For example the distribution of population according to sex(male, female) or wealth (rich, middle, poor) or education (literate, illiterate) etc.

Qualitative classification may be of two types:

(i) **Simple classification:-** When the classification is done considering only one variable characteristic or attribute is known as simple classification or one way classification. It divides the data into two classes called dichotomy.

(ii) **Manifold classification:-** When the classification is done considering more than one variable characteristic or attributes, it is known as manifold classification.

Aims of classification:- The main aims of classification are:

- To reduce the large set of data to an easily understood summary.
- To bring out clearly the points of similarity and dissimilarity.
- To save mental strain by eliminating unnecessary details.
- To reflect the important aspects of the data.
- To prepare the ground for tabulation, comparison and analysis.

4. **Tabulation:-** Tabulation is the next to classification in the process of statistical investigation. A table is a systematic arrangement of data in columns and rows. The process of arranging data into rows and columns is called tabulation.

Tabulation is the last stage in the collection and compilation of data and is a sort of stepping stone to the fields of analysis and interpretation. Tabulation makes the data easy to understand, facilitate comparisons and provides an effective way to convey information to the readers.

Types of Tabulation:- Tabulation is the mechanical part of classification. Therefore the types of tables depend upon the type of classification. For instance a one way classification, the tabulation is called simple. Similarly for a two way classification and manifold classification, we can have two-way table and complex table etc.

Main parts of a table:- A statistical table has at least four parts, the Title, the Stub, the Box-head and the Body. In addition, some tables have one or more prefatory notes, a foot note and a source note.

A Brief description of these parts is given below:

1. **Title:-** Every table must have a title. A title is a heading at the top of the table describing its contents. A title usually tells us the What, Where, How and When of the data, in that order.
2. **Row Caption and Stub:-** The heading for a row is called the row caption. The portion of the table containing the row-captions is known as Stub of the table.
3. **Column captions and Box head:-** The heading of each column is called column caption, while the section of a table that contain the column captions, is referred to as Box head.
4. **Body and arrangement of data:-** The Body of a table is the most important part, which contains the entries in the appropriate cells of the table together with totals etc. Body of the table contain the entire data arranged in column and rows.
5. **Prefatory note and Foot notes:-** Both of these notes are used to explain certain characteristics of the data.

   The prefatory note appears between the Title and the Body of the table. It is usually used to throw some light about the table as a whole.

   A Footnote appears immediately below the body of the table. It is used to explain a single fact or a part of the table (column or row). A Footnote does not throw light on the table as a whole.
6. **Source note:-** Every table must have a source note, unless the data are original one. Source note is placed immediately below the table but after the footnote, if any. The source note must indicate the compiling agency, publication, date of publication, page etc. As they are used as a means of verification and reference.

The general sketch of a table is given below:

⟵ .....TITLE..... ⟶

.....Prefatory notes.....

Box head	COLUMN CAPTION				
	Units				
STUB	.....	.....B O	D Y.....	.....	

Foot note.....

Source notes.....

Frequency Distribution:- The arrangement of a set of data in a table showing the distribution of the data into classes or groups together with the number of observations in each class or group is called a frequency distribution. The number of observations falling in a particular class is called a class frequency and is denoted by f. Data presented in the form of a frequency distribution are also called grouped data otherwise ungrouped data.

Some basic terms related with frequency distribution:

1. **Class interval:-** The interval "$L - U$" defining a specific class.
   Where L is the lower class limit.(the starting point)
   And U is the upper class limit.(the finishing point)
   Class interval is also called class-limits. Both the limits i.e. upper and lower are both included in the class interval. Then this type of classification is known as "inclusive method". For example $40 - 49, 50 - 59, 60 - 69$ etc.
2. **Class boundaries:-** The precise class limits, constructed in such a way that the upper class limit "U" of a class coincides with the lower class limit "L" of the next higher class, hence there exists no gap or break or jump in between the two consecutive classes. This type of classification is known as "exclusive method" and the limits of a class are called class boundaries. For example $40 - 50, 50 - 60, 60 - 70$ etc.
3. **Class size or Class width:-** The difference between the upper and lower class boundary of a class or the difference between the two successive lower class limits or the difference between the two successive class mid-points. It is generally denoted by "c" or "h" or "i".

Construction of a frequency distribution:- The following points should serve as a guideline that should be kept in mind when constructing a frequency distribution.

I. Find "N" the total number of values in raw data.

II. Pick up the maximum and minimum values and compute the range i.e. the difference between them.

RANGE = Maximum value – Minimum value $(R = X_m - X_0)$

III. Decide on the number of classes into which the data are to be grouped. Although there is no hard and fast rule for deciding on the number of classes, but H.A Sturges has proposed an empirical rule which will approximately determine the number of classes i.e.

$$K = 1 + 3.3 \log N$$

Where "K" denotes the number of classes & "N" is the total number of observations collected.

IV. Divide the range "R" by the number of classes "K" to determine the approximate width or size of the class i.e. $h = R/K$. In case of fractional results, the next higher whole number is usually taken as the size or width of the class interval.

V. Decide where to locate the class limit of the lowest class and then the lower class boundary. The lowest class usually starts with the smallest data value or a number less than it.

VI. Determine the remaining class limits and class boundaries by adding the class size repeatedly.

VII. Distribute the data into appropriate classes this is done by using "Tally column".

VIII. Finally, total the frequency column to see that all the data have been accounted for i.e. $\sum f = N$.

EXAMPLE:- Make a frequency distribution from the following data, relating to the weight recorded to the nearest gram of 60 apples.

106,107,76,82,109,107,115,93, 187,95,111,92,86,70,126,68,130,129,139,119,
100,186,84,99,113,204,111,141,136,123,98,110,78,185,162,178,140,152,173,146,
148,90,107,181,131,75,184,104,110,80,123,115,90,158,194,118,82,115,128,125.

SOLOUTION:- Here

(i) $N = 60, X_m = 204, X_0 = 68$

(ii) $R = X_m - X_0 \Rightarrow 204 - 68 = 136$

(iii) $K = 1 + 3.3 \log N \Rightarrow 1 + 3.3 \log 60$
$= 1 + 3.3(1.7782) = 6.8 \approx 7$ Classes

(iv) $h = R/K \Rightarrow 136/7 = 19.4 \approx 20$ Units.

(v) Now let us decide to locate the lower limit of the lowest class at 65.

Weight	Tally Column	Frequency
65 – 84	𝍸 IIII	9
85 – 104	𝍸 𝍸	10
105 – 124	𝍸 𝍸 𝍸 II	17
125 – 144	𝍸 𝍸	10
145 – 164	𝍸	5
165 – 184	IIII	4
185 – 204	𝍸	5
Total		60

Cumulative frequency distribution:- The total frequency of a variable from its one end to a certain value (lower or upper class boundary of a class), called the base, is known as the cumulative frequency, less than or more than the variable.

A table that shows the cumulative frequencies is called a cumulative frequency distribution.

Generally there are two methods of constructing a cumulative frequency distribution.

1. **Less than Type:-** In this case the frequencies are added serially from top to bottom and they are represented against the upper class boundaries of the class intervals upon which they fall and indicate the number of items or values smaller than or equal to the value against which it is represented.
2. **More than Type:-** In this the frequencies are added serially from bottom to top and cumulative frequencies are represented against the lower class boundaries of the corresponding class interval. This type of cumulative frequency indicate the number of items or values having values greater than or equal to that particular value.

From the previous example, we have

Weight	Frequency	c.f (Less than Type)	c.f (More than Type)
65 – 84.5	9	9 (Below 84)	51 + 9 = 60
85 – 104.5	10	9 + 10 = 19	41 + 10 = 51
105 – 124	17	19 + 17 = 36	24 + 17 = 41
125 – 144	10	36 + 10 = 46	14 + 10 = 24
145 – 164	5	46 + 5 = 51	9 + 5 = 14
165 – 184	4	51 + 4 = 55	5 + 4 = 9
185 – 204	5	55 + 5 = 60	5 (Above 185)
Total	60		

Relative frequency Distribution:- If in a frequency distribution, the class frequencies are when divided by the total frequency, we get the relative frequencies, and the table is then relative frequency distribution.

Moreover if the relative frequencies are expressed as percentages, then it is called percentage relative frequencies.

ASSIGNMENT NO.1:- The following data represent marks obtained by 70 students

74,49,103,95,90,118,52,88,101,96,72,59,62,96,82,65,85,
105,116,91,83,99,77,104,96,84,62,58,66,100,80,54,75,56,96,
83,57,60,51,114,120,121,92,88,56,52,55,64,64,76,99,63,110,
84,104,95,97,89,78,78,54,58,61,65,67,70,74,77,81,102.

QUESTION:- Form a grouped frequency distribution indicating:

Class limits, Class Boundaries, Mid points, Cumulative frequencies and Percentage relative frequencies.

Graphic and Diagramatic Presentation of Data

After the classification and tabulation techniques that help in summarizing the collected data and presenting it in a meaningful fashion, one of the most effective and interesting way in which a statistical data may be presented is through graphs and diagrams. In which a statistical data may be presented is through graphs and diagrams. In which the data can be presented pictorially in the form of points, lines, areas and other geometrical forms. Such a visual representation can be divided into **Graphs** and **Diagrams.**

Graphs:- Graph is a representation of data by a continuous curve or straight lines showing fluctuations and trends in statistical data.

Types of Graphs:- Graphs can be divided into two main categories:

1. Graph of a time series (Historigram) 2.Graphs of Frequency distributions (Histogram)

The important graphs of a frequency distributions are:

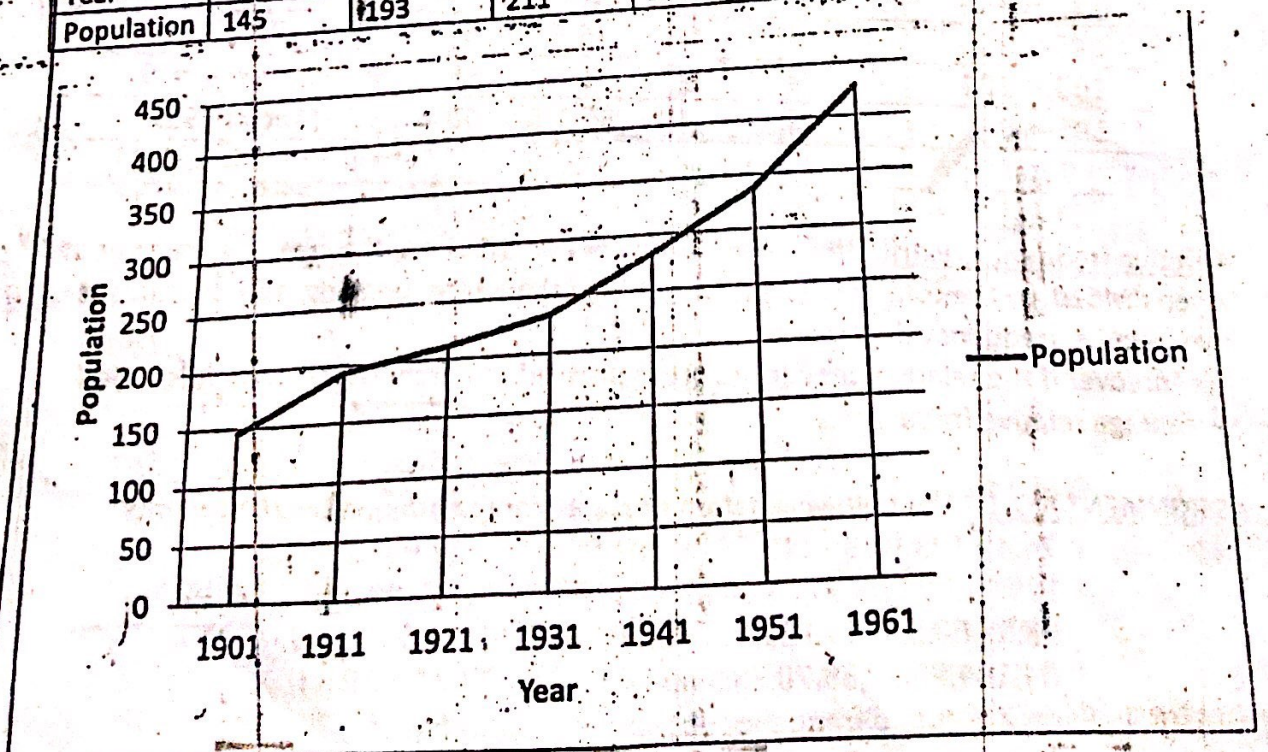
Histogram, Frequency polygon, Frequency curve and the Cumulative frequency curve or Ogive.

1. Graph of a time series (Historigram):- A collection of data with respect to its time of occurrence is called a Time series and a graph relating to this type of series spread over a period of time is called **Historigram.**

Generally, in the construction of a histogram time is taken along X-axis and the respective variable values are placed along Y-axis. The values are plotted against their respective time period and are connected by straight lines.

Example:- Data given in the following table can be shown with the help of a Historigram as follows:

Year	1901	1911	1921	1931	1941	1951	1961
Population	145	193	211	233	282	337	428



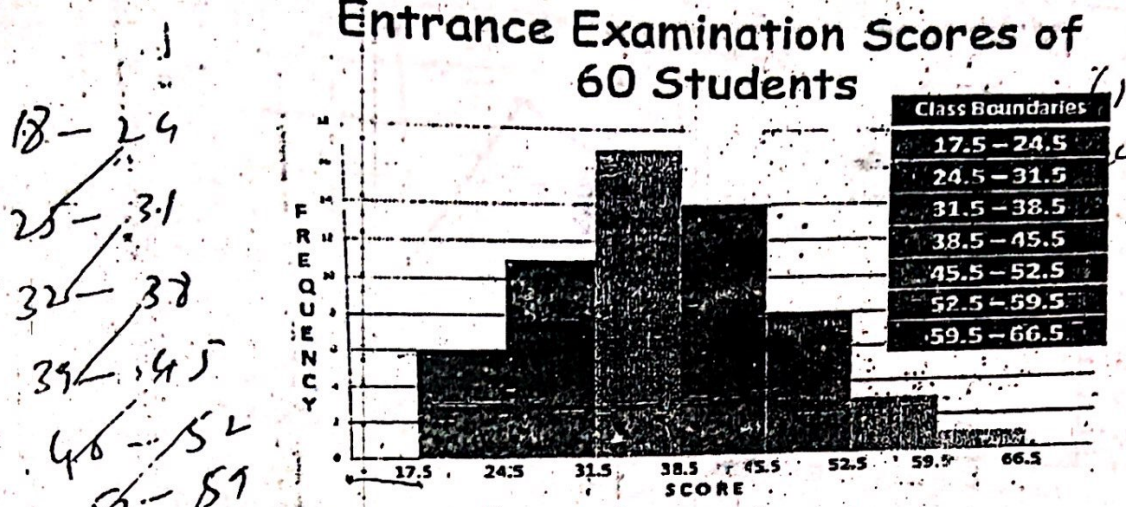
2. **Graphs of a frequency distributions:**

a. **Histogram:-** A Histogram is the visual representation of a frequency distribution of a continuous variable. The Histogram is constructed of adjacent rectangles. The heights of a rectangle represent the class frequencies and the bases extend between

successive classes boundaries (not class limits). But for unequal class interval the heights must be proportional to the ratio of frequencies to the width of the class.

Example:- Draw a Histogram for the following data.

Marks	18 – 24	25 – 31	32 – 38	39 – 45	46 – 52	53 – 59	60 – 66
No. of students	6	11	17	14	8	3	1



Histogram with unequal class interval:

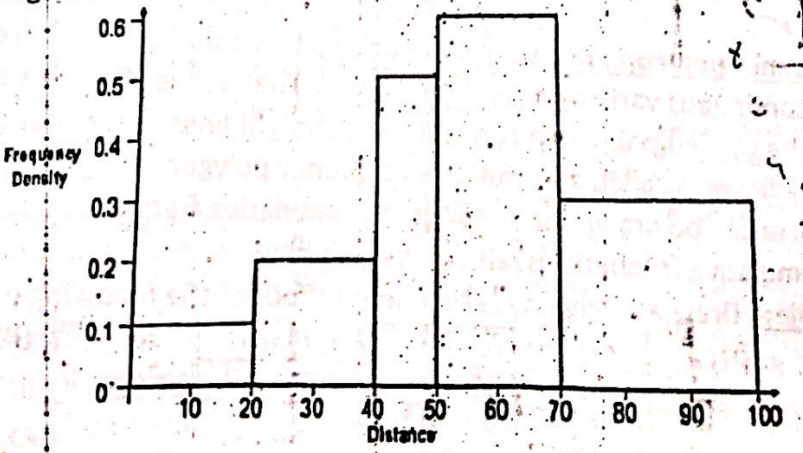
Example:- Draw a Histogram for the following data set:

Distance	Frequency	Frequency Density	
$0 < t \leq 20$	2	0.1	2 ÷ 20
$20 < t \leq 40$	4	0.2	4 ÷ 20
$40 < t \leq 50$	5	0.5	5 ÷ 10
$50 < t \leq 70$	12	0.6	12 ÷ 20
$70 < t \leq 100$	9	0.3	9 ÷ 30

Before presenting the data in the form of Histogram, unequal class intervals are adjusted as we construct each rectangle with the class interval as base and frequency density as height.

$$Fequency\ Density = \frac{Frequency\ of\ the\ class\ interval}{Class\ size\ of\ the\ interval}$$

The required Histogram is shown as under:

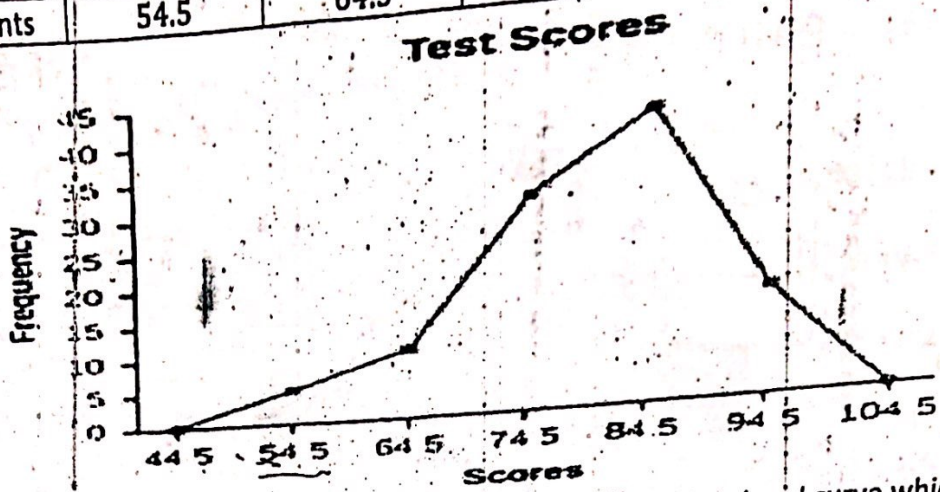


Frequency Polygon:- A Polygon is a closed figure having many sides. A frequency polygon is constructed by plotting the class mid-points along X-axis against their class frequencies (or adjusted frequencies if required) along Y-axis and then connecting these points by straight lines, extended outward to the base line, half on interval away on the both sides. A

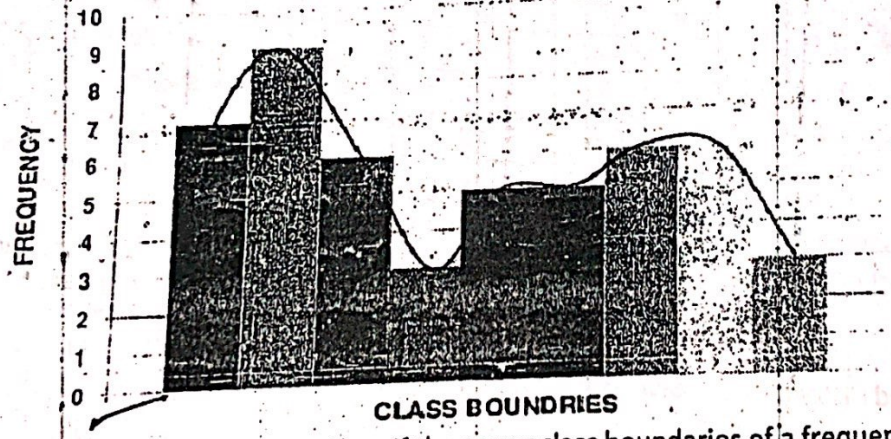
frequency polygon is also obtained by connecting mid-points of the top of the adjacent rectangles by straight lines.

Example:- Draw a frequency polygon for the following data set:

Test score	49.5 – 59.5	59.5 – 69.5	69.5 – 79.5	79.5 – 89.5	89.5 – 99.5
Frequency	5	10	30	40	15
Mid points	54.5	64.5	74.5	84.5	94.5



Frequency curve:- If the frequency polygon is smoothed out by a free hand curve which avoids the sharp turns and minimize the sudden changes then this curve is known as frequency curve.



Cumulative frequency polygon or Ogive:- If the upper class boundaries of a frequency distribution of a continuous variable are plotted against their respective class cumulative frequencies and the plotted points are connected by straight lines or free hand smooth curve the resulting figure is called a cumulative frequency polygon or Ogive.

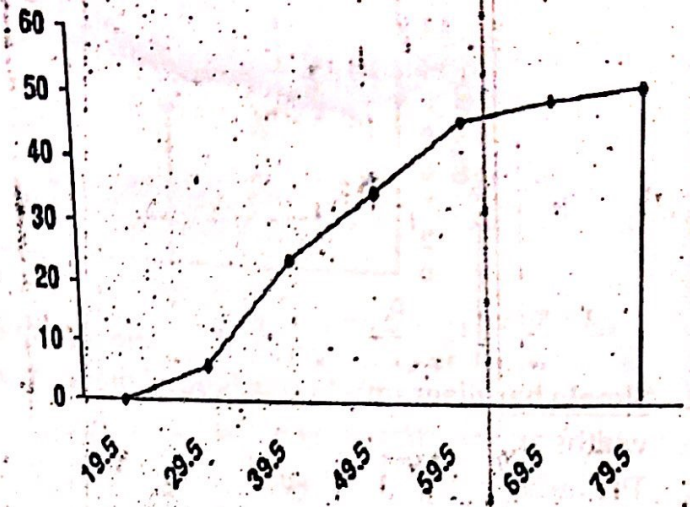
Similarly an Ogive can be drawn for a More than cumulative frequency distribution simply plotting cumulative frequencies with respect to lower class boundaries.

Example:- Draw an Ogive by Less than method for the following data:

Class Limits	Frequency
20 – 29	6
30 – 39	18
40 – 49	11
50 – 59	11
60 – 69	3
70 – 79	1
Total	50

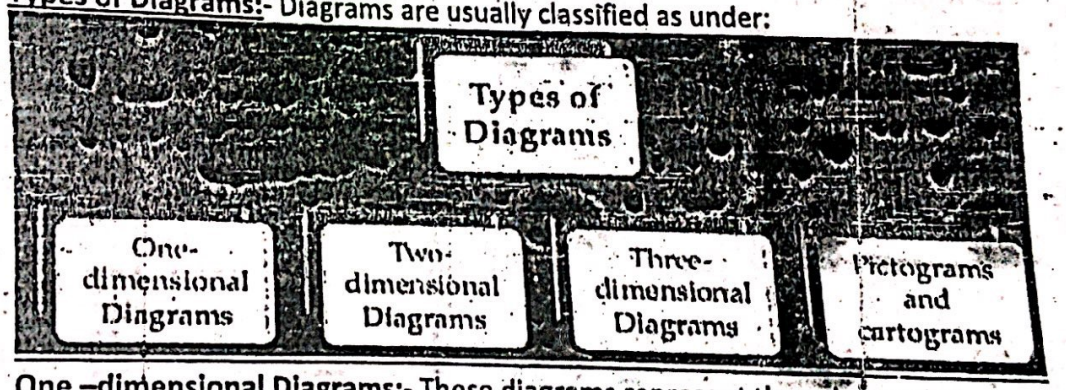
Solution:- To draw a cumulative frequency polygon, first we calculate cumulative frequencies as given in the following table.

Class Limits	Frequency	Cumulative frequency
20 - 29	6	6
30 - 39	18	24
40 - 49	11	35
50 - 59	11	46
60 - 69	3	49
70 - 79	1	50
Total	50	



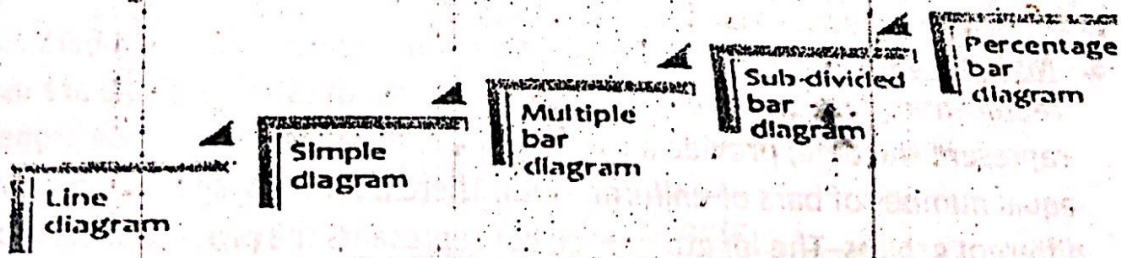
b. **DIAGRAMS:-** As we know that a diagram is also a visual representation of statistical data in the form of geometric figures like bars, circles, maps, pictorials, cartogram etc.These devices can take many attractive forms and it exhibt the result and information more quickly clearly contained in the statistical data.

Types of Diagrams:- Diagrams are usually classified as under:



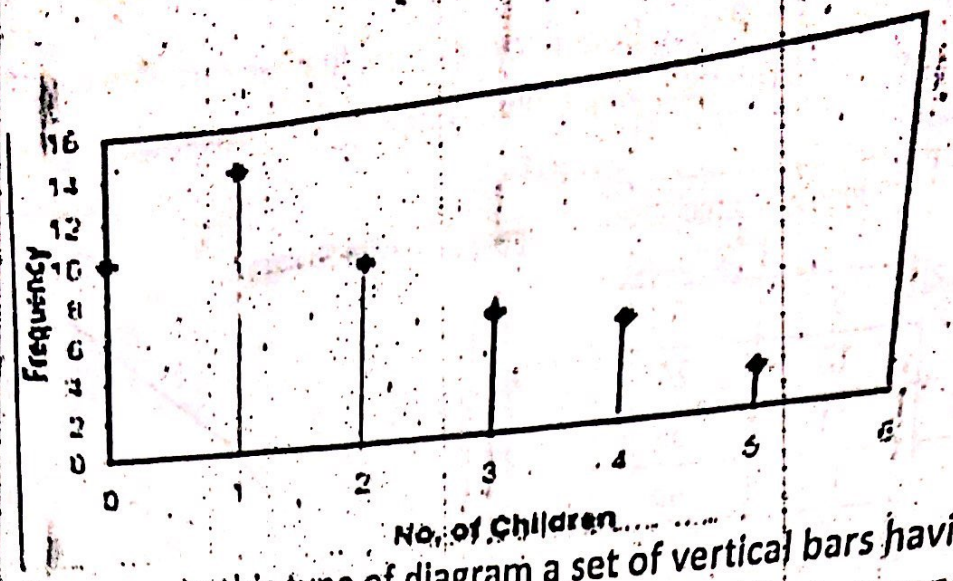
One –dimensional Diagrams:- These diagrams represent the values only by one dimension, generally by the length in the shape of bars or lines. The length of the bar or line is proportional to the different figures they represent.

One dimensional diagram is further subdivided into:



1. **Line Diagram:-** Line diagram is used in case where there are many items to be shown and there is not much of difference in their values. Such diagram is prepared by drawing a vertical line for each item according to the scale. The distance between lines is kept uniform. Line diagram makes comparison easy, but it is less attractive.

Example :

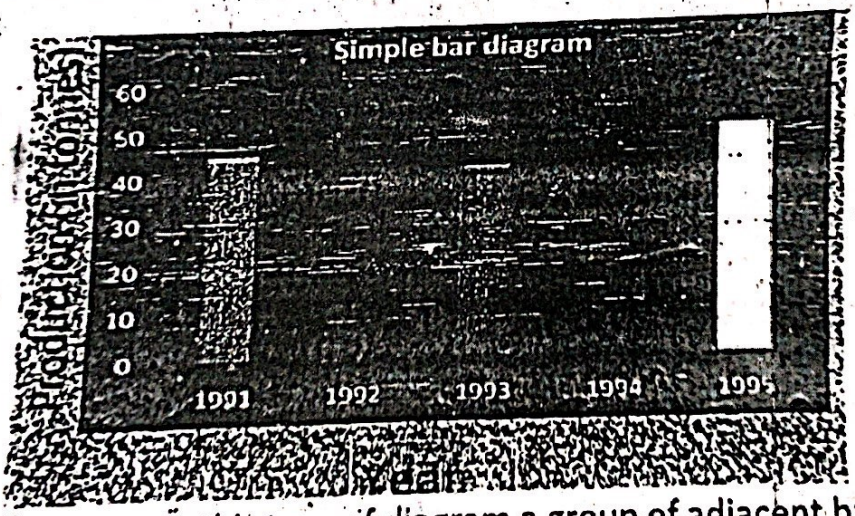


2. **Simple bar diagram:-** In this type of diagram a set of vertical bars having uniform widths and separated by equal intervals on the horizontal line are constructed. The heights of bars are taken proportional to the magnitude of the respective variable. The height of the tallest bar is adjusted in such a way that it suit to the proper scale. These bars are arranged if possible in ascending or descending order of the height. The width of the bars has no other use except to make the bars look more attractive.

Example:- Represent the following data by a bar diagram:

Year	1991	1992	1993	1994	1995
Production	45	40	42	55	50

Solution:-

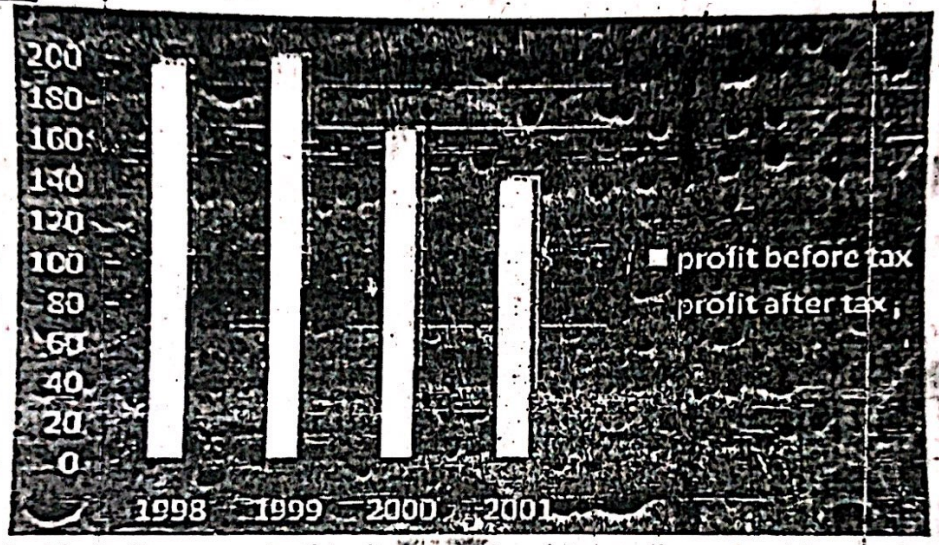


3. ***Multiple bar diagram:-*** *In this type of diagram a group of adjacent bars, representing two or more inter related characteristics of a variable are used to represent the data, provide a direct comparison among them. Each group has equal number of bars of uniform width. There is an equal space among the different groups. The length of each bar represents the proportional respective magnitude. In order to provide better understanding bars are decorated for distinction with colours, dots, shades etc.*

Example:- Draw a multiple bar diagram for the following data.

Year	Profit before tax (in lakhs of rupees)	Profit after tax (in lakhs of rupees)
1998	195	80
1999	200	87
2000	165	45
2001	140	32

Solution:-

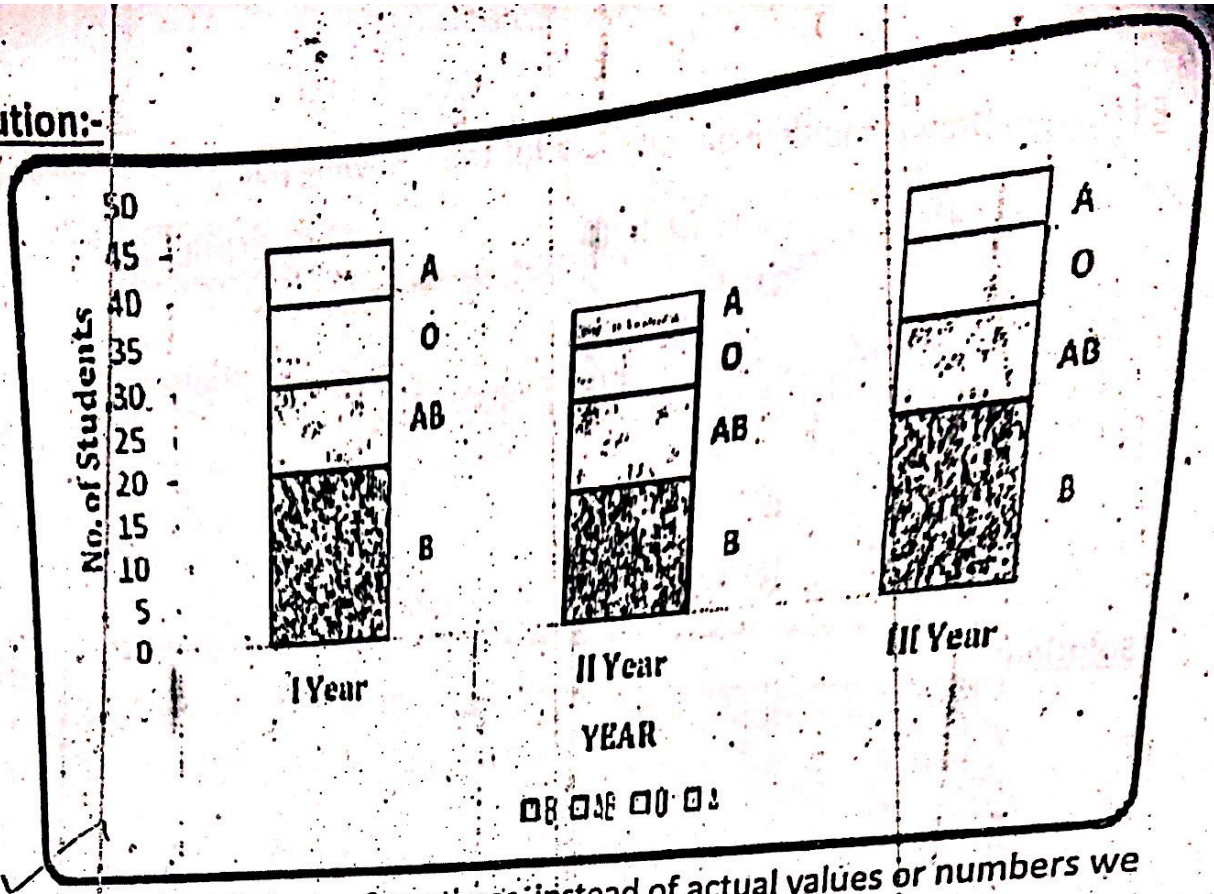


4. **Subdivided bar diagram:-** In this diagram simple equally spaced bars representing both the absolute magnitudes of a variable quantity and its components are constructed with height proportional to the total magnitude of the constituent components. These bars are further subdivided in proportion to the magnitude of their components. These components are also arranged in ascending or descending order in the highest bar and same order is followed in the other bars. The subdivisions of the bars are coloured or shaded differently to have better understanding. These bars are used when it is necessary to show the breakup of one variable in several components.

Example:- Data shown Blood group in three years of students given as under. Represent it by a subdivided bar diagram:

Blood Group.	No. of Students		
	I – Year	II – Year	III - Year
A	6	4	5
B	20	15	19
AB	10	10	10
O	9	6	8

Solution:-



5. **Percentage bar diagram:-** Sometimes, instead of actual values or numbers we are interested in the percentage (or the information itself is given in the percentages) to have an easy and better comparison. Generally the components of subdivided bar diagram are presented on percentage basis. In this diagram all the bars have equal widths and height is taken as a total of 100. These bars are subdivided in accordance with the percentage of components.

Example:- Represent the following data by a percentage bar diagram:

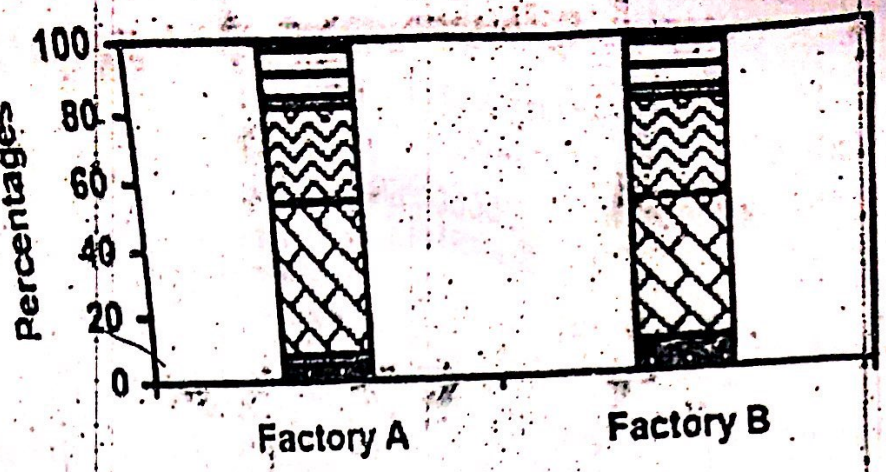
Particular	Factory A	Factory B
Selling Price	400	650
Quantity Sold	240	365
Wages	3500	5000
Materials	2100	3500
Miscellaneous	1400	2100

Solution:-

Convert the given values in percentages as follows by using the following formula:

$$Component\ Percentage = \frac{Component\ value}{Total} \times 100$$

Particulars	Factory A		Factory B	
	Rs.	%	Rs.	%
Selling Price	400	5	650	6
Quantity Sold	240	3	365	3
Wages	3500	46	5000	43
Materials	2100	28	3500	30
Miscellaneous	1400	18	2100	18
Total	7640	100	11615	100



Two-dimensional Diagrams:- In two-dimensional diagram the area represent the data and so the length and breadth have both to be taken into account. Such diagrams are also called area diagram or surface diagrams.

The important types of area diagrams are:

1. Rectangles.
2. Squares.
3. Pie-diagrams.

1. **Rectangles:-** These are generally used to represent the relative magnitude of two or more variable quantities. The area of the rectangles is kept in proportion to the total values of a variable. Rectangles are placed side by side for comparison. We may represent the figures as they are given or may convert them to percentages and then subdivide the length into various components.

Example:- Represent the following data by sub-divided percentage rectangular diagram.

Items of Expenditure	Family A (Income Rs.5000)	Family B (Income Rs 8000)
Food	2000	2500
Clothing	1000	2000
House Rent	800	1000
Fuel and lighting	400	500
Miscellaneous	800	2000
Total	5000	8000

Solution:- The items of the expenditure will be converted into percentage as shown below:

Items of Expenditure	Family A		Family B	
	Rs	Y	Rs	Y
Food	2000	40	2500	31
Clothing	1000	20	2000	25
House Rent	800	16	1000	13
Fuel and Lighting	400	8	500	6
Miscellaneous	800	16	2000	25
Total	5000	100	8000	100

Now to construct a diagram, the heights of rectangles are taken as constant 100, and the widths are taken in proportion. Then subdivide the rectangles according to cumulative percentages.

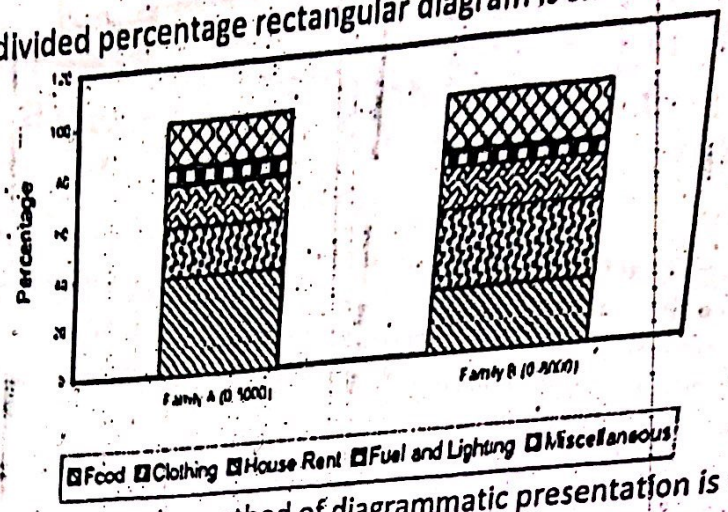
The proportion of widths can be taken as:

Family A: Family B

5000 : 8000

1 : 1.6

Then the subdivided percentage rectangular diagram is shown below:



2. **Squares:-** The rectangular method of diagrammatic presentation is difficult to use where the values of items vary widely. The method of drawing a square diagram is very simple. One has to take the square root of the values of various items that are to be shown in the diagrams and then select a suitable scale to draw the squares.

Example:-

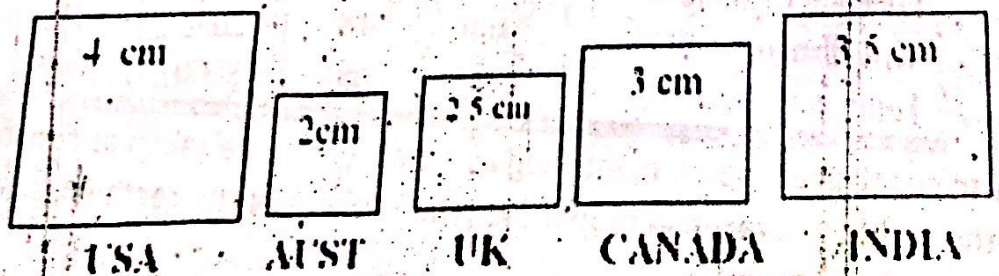
Yield of rice in Kgs per acre of five countries are

Country	U.S.A	Australia	U.K	Canada	India
Yield of rice in Kgs per acre	6400	1600	2500	3600	4900

Represent the above data by Square diagram.

Solution:- To draw the square diagram we calculate as follows:

Country	Yield	Square root	Side of the square in cm
U.S.A	6400	80	4
Australia	1600	40	2
U.K.	2500	50	2.5
Canada	3600	60	3
India	4900	70	3.5



3. **Pie Diagram or Circular Diagram:-** A Pie diagram, also known as sector diagram is a device consisting of a circle divided into sectors or Pie-shaded pieces whose areas are proportional to the various parts into which the whole quantity is divided. To construct a pie chart, draw a circle of any convenient radius. As a circle consists of 360°, the whole quantity to be displayed is equated to 360. The categories of the variable quantities are calculated into angles by using the formula:

$$Angle = \frac{Component\ part}{Whole\ quantity} \times 360°$$

Then divide the circle into different sectors by constructing angles at the centre by mans of a protector and draw the corresponding radii.

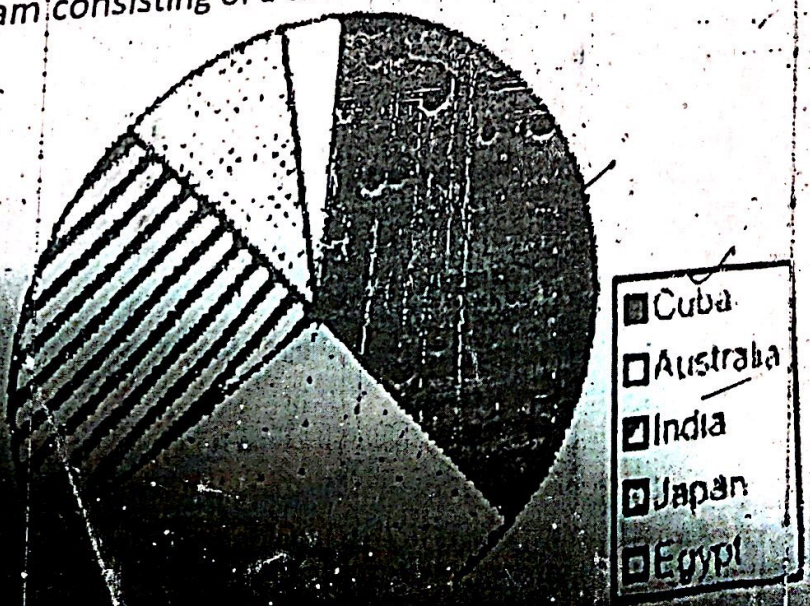
Example:- Production of sugar in quintals of various countries are given below. Represent it by a pie diagram.

Country	Production of Sugar (in quintals)
Cuba	62
Australia	47
India	35
Japan	16
Egypt	6

Solution:- The values are expressed in terms of degree as follows:

Country	Production of Sugar	
	In Quintals	In Degrees
Cuba	62	134
Australia	47	102
India	35	76
Japan	16	35
Egypt	6	13
Total	166	360

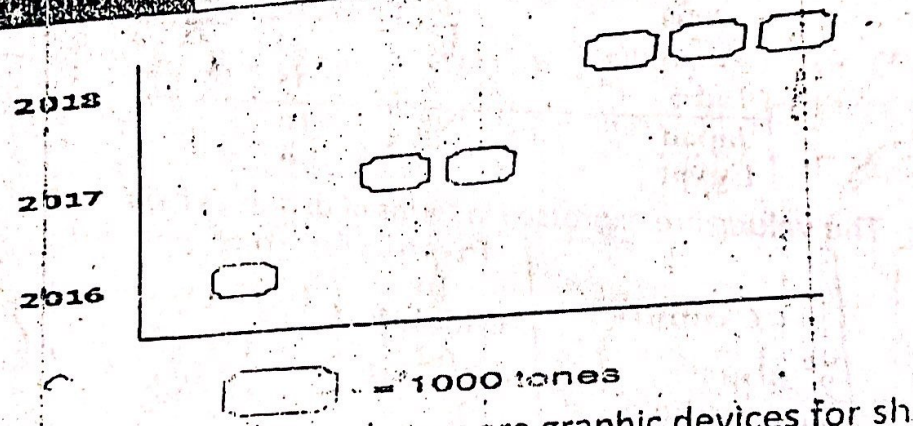
The Pie diagram consisting of a circle divided into five sectors is drawn below:



Pictogram:- In this diagram pictures are used to represent the data. Each picture represents a fixed quantity of the variable and the number of pictures gives the total magnitude of the variable. A partly completed picture shows the corresponding fraction of block of units represented by an example picture.

Example:- Sales of fish (in tons) in last three years are:

Year	2016	2017	2018
Sales	1000	2000	3000



Cartogram:- Cartograms or statistical maps are graphic devices for showing the numerical information on geographical basis. Cartograms are simple to use and easy to understand. They are generally used when regional or geographic comparison is to be made. Population density, mineral resources, cultural or industrial production in different regions etc. can be represented by cartograms.

****