DIAGRAMS

Visual representation is much better than tabular representation of the data to understand them. Depending on the nature of the data we can use one of the following diagrams .

1. One dimensional diagrams - Bar diagrams

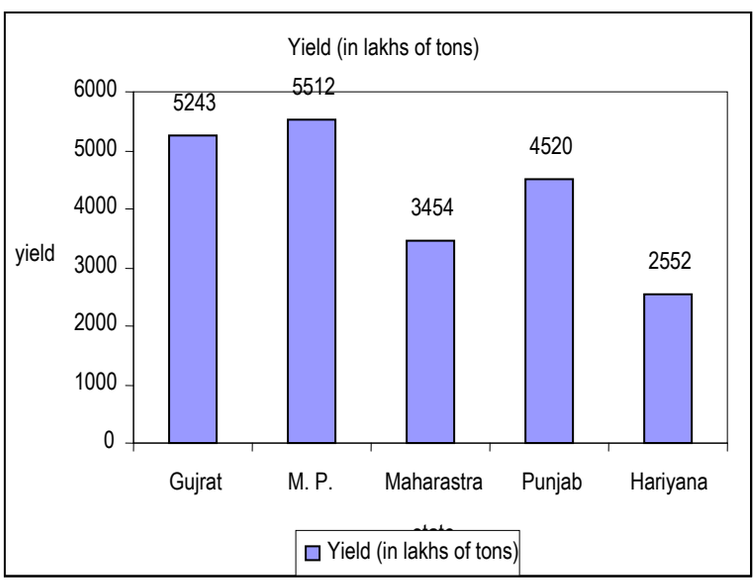
2. Two dimensional diagrams Squares ,Rectangles , Pie diagrams

1)One dimensional diagrams - Bar diagrams :- These are is the simplest and most commonly used diagrams. Variable is represented by a thick bar of uniform width leaving uniform gaps between two bars. The height of the bar is proportional to the value of the variable. Bar diagrams are classified into three types. (1) simple bar diagram,(2) subdivided bar diagram and (3) multiple bar diagram

(1)Simple bar diagram when data are about single variable, we use simple bar diagram

Ex 1.Consider the following data;

State	Yield (in lakhs of tons)
Gujrat	5243
Madhya Pradesh	5512
Maharastra	3454
Punjab	4520
Hariyana	2552

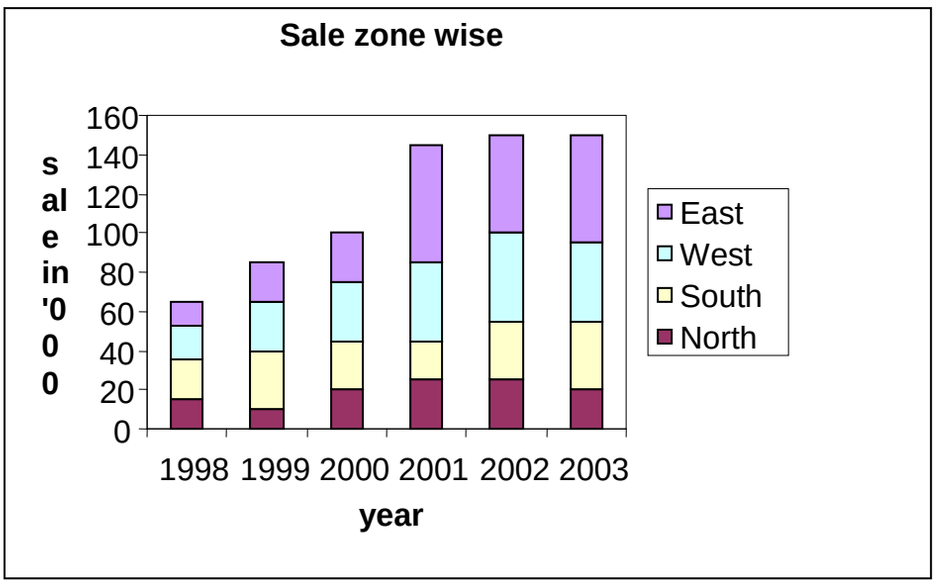


(2)Segmented bar diagram:- Sometimes variable is divided in to different components then we use segmented bar diagrams.

Ex 2. Represent the data by suitable data.

Year	North	South	West	East	Total sale
1998	15	20	18	12	65
1999	10	30	25	20	85
2000	20	25	30	25	100
2001	25	20	40	60	145
2002	25	30	45	50	150

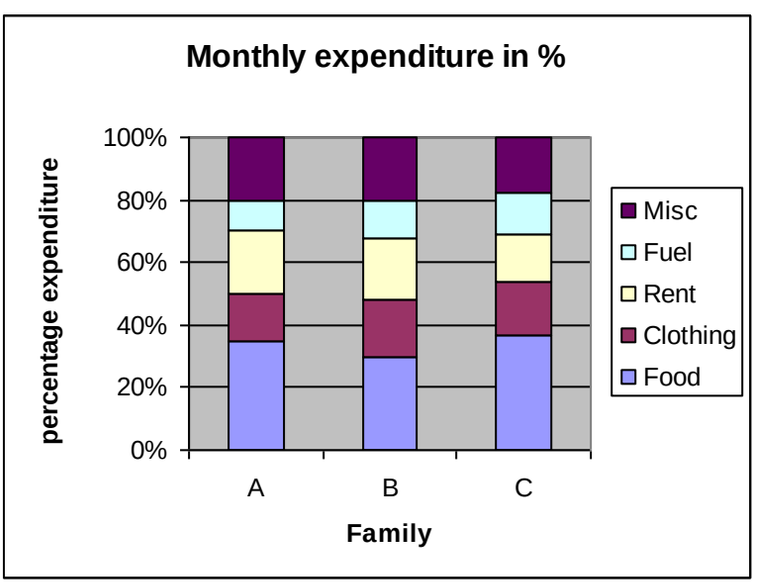
2003	20	35	40	55	150



Sometimes division of the variable is expressed in terms of percentages ,then we use segmented bar diagram for percentage and is known as Percentage bar diagram.

Ex 3): Following data represent the relative break up of monthly expenditure of three families A,B, C.

Item	Family A	Family B	Family C
Food	35	30	37
Clothing	15	18	17
Rent	20	20	15
Fuel	10	12	13
Misc	20	20	18

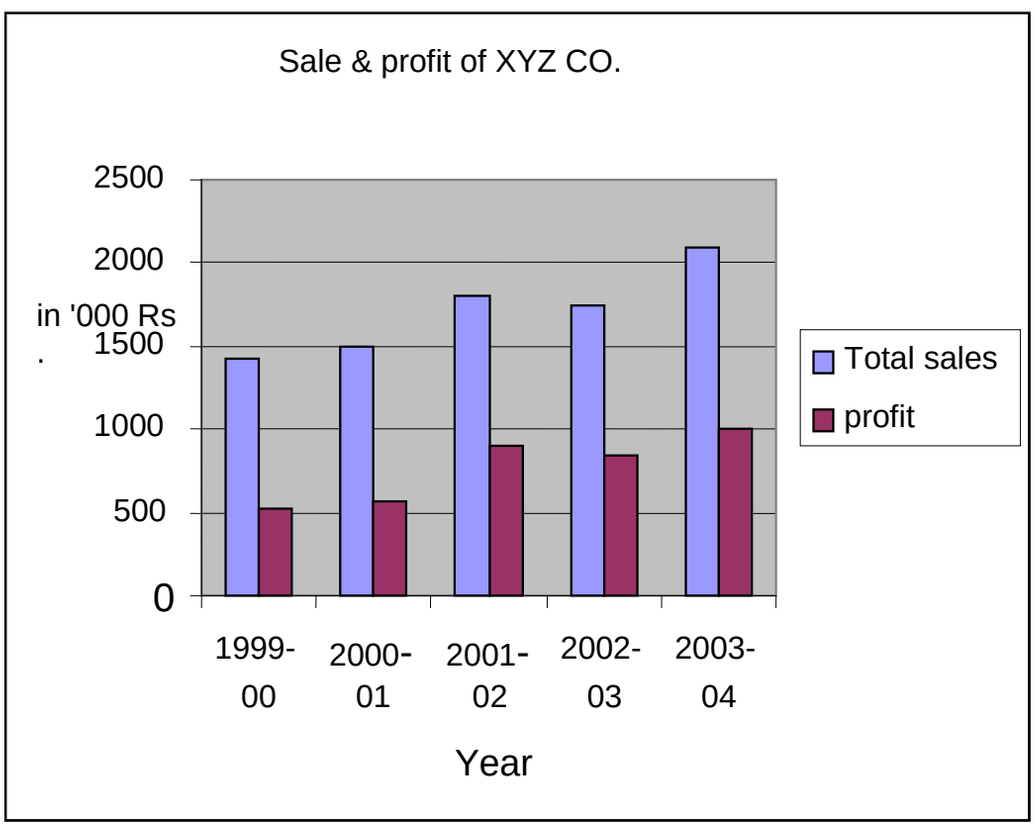


(3)Multiple bar diagram. To represent two or more sets of values related with same group , we use multiple bar diagram.

Ex 4). Following data represent sale & profit of a company.

Sale and Profit of XYZ Company. (all figures are in '000 Rs.)

Year	Total sales	profit
1999-00	1420	520
2000-01	1500	560
2001-02	1800	900
2002-03	1750	850
2003-04	2100	1000



2 Two dimensional diagrams :In two dimensional diagrams area is proportional to the value of the variable. Commonly used two dimensional diagrams are squares rectangles and Pie diagram.

i) Square: A square is used to compare the values of the characteristic which differ widely from one another. The area of the square is proportional to the

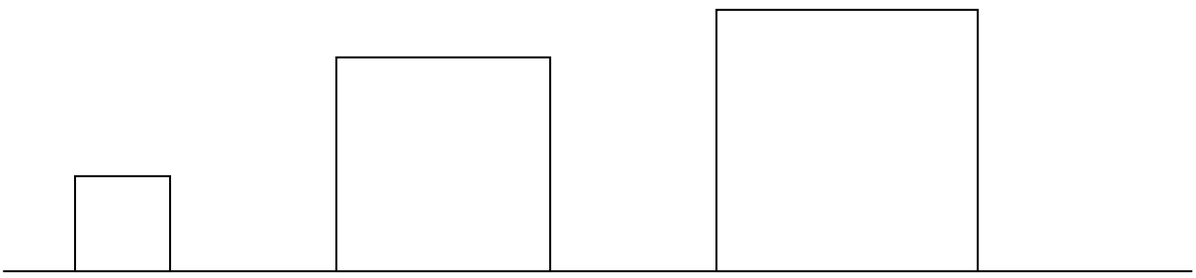
value of the variable. Hence side of the square is proportional to the square root of the value of the variable.

Ex 5:_Represent the following data by suitable diagram.

Year	2002	2006	2008
Export (in million Rs)	1600	6400	10000

Ans Side =√ (export)

1 c.m. = 40 units

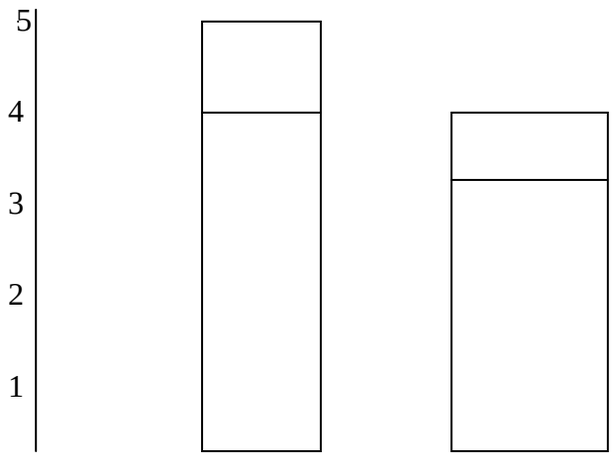


2002 2006 2008

ii) Rectangles: When we have the data on two characteristics relating to different items and the product of these values have a significance. Then we use rectangle. One characteristic is taken along height and other along width. Eg data is with respect to the selling price and number of units sold of a certain commodity, then the product represent revenue generated from that commodity.

Ex6: Represent the data by suitable diagram.

	Product A	Product B
No. of units Sold	10	15
Selling price (in' 000 Rs)	5	4
Cost price(in '000 Rs)	4	3.5



0 10 units 15 units

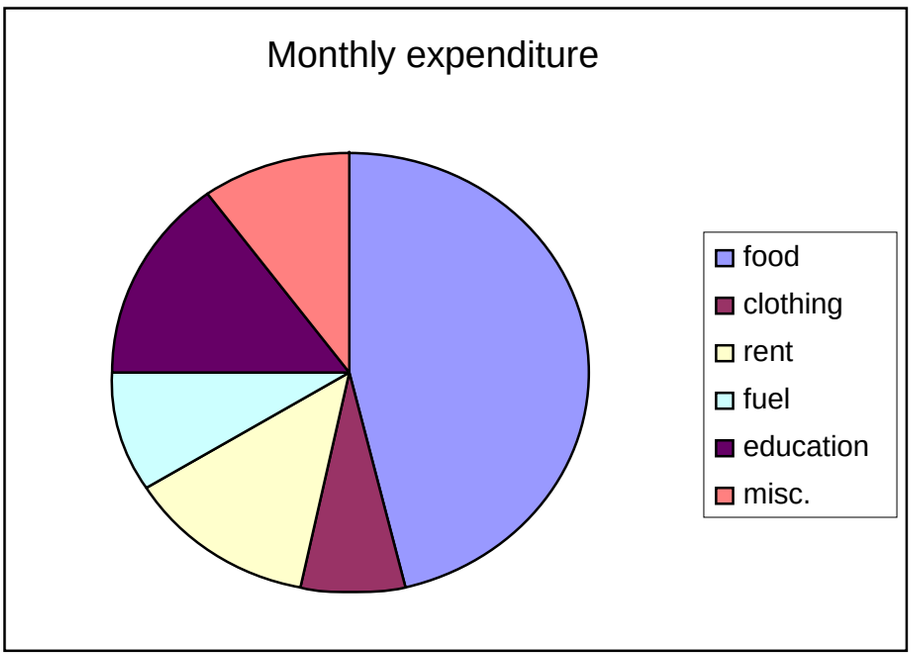
Pie diagram : When the number of components of a variable are large, the segmented bar diagram fails to give proper visual representation. In such case we use pie diagram.

A pie diagram consists of a circle divided into number of sectors representing different components of a variable. The areas of these sectors are proportional to the values of the components. Since area of the sector is proportional to its angle. We consider value of the component proportional to angle. We determine angle for each sector by using formula Angle of the sector = 360*(value of the component / total of all components).

Ex 7)The following data represents monthly expenditure of a family on different items

Item	Expenditure	Angle in degree
Food	4600	360*(4600/10000) = 165.6
Clothing	700	360*(700/10000) = 25.2
Rent	1300	360*(1300/10000) = 46.8
Fuel & lights	900	360*(900/10000) = 32.4
Education	1500	360*(1500/10000) = 54.0
Miscellaneous	1000	360*(1000/10000) = 36.0
Total	10000	360.0

Pie diagram



Merits (✓) and demerits (×) of Diagrams

- ✓ The diagrams are visuals and hence easy to understand the data.
- ✓ They provide the information instantly.
- ✓ The diagrams are much more attractive than the numerical data.
- ✓ Diagrams can be easily remembered than the numerical figures.
- ✓ Comparisons can be done more easily using diagrams.
- × The diagrams cannot represent exact values.
- × The diagrams cannot give us detailed information.
- × Diagrams are supplement to tabular representation but not alternative to it.
- × For further statistical analysis diagrams are not useful.

CLASSIFICATION

When the data is collected the data has to be classified. Classification implies grouping of relevant items together into classes. Characteristics or properties of one class differ from any other class.

The basic objectives for classification are

- To condense the huge data into few classes so that similarities and dissimilarities in the data are easily recognized.
- To help comparison.
- To highlight important features and to pinpoint most significant features.
- To do away with unimportant features.
- To present data in a form from which further statistical analysis is feasible.

In general we classify the data using following categories.

1) Geographical classification (Area-wise classification) :It is with geographical areas or regions

Ex 8:CROP YIELD OF WHEAT IN THE YEAR 2003

State	Yield (in lakhs of tons)
Gujrat	5243
Madhya Pradesh	5512
Maharastra	3454
Punjab	4520
Hariyana	2552

2) Chronological classification (time series data.) Here units are classified with respect to time variable.

If we can represent the data time wise , then data is known as time series data.

Ex9 : RESULT OF T.Y.B.SC. FOR 2000-2004

Year	I Class	II Class	Pass Class	Fail	Total
2000	80	65	40	15	200
2001	75	75	50	20	220
2002	55	55	55	15	180
2003	100	65	55	20	240
2004	90	80	30	25	225

Quantitative data (frequency distribution.) : To prepare a frequency distribution , first we have to decide classes, then number of observations falling in each class is to be found. The number of observations belonging to each class is called frequency of that class. The frequency distribution is a tabular arrangement of data by classes together with class frequencies.

If we have N observations and these are to be grouped in to classes of equal width then we have to follow following steps

Number of classes=K= 1+3.332log N

K should be integer .

Locate maximum and minimum of the data then width of class is given by

Width of the class = d =(max-min)/K

Write K classes with width d in such a way that minimum value get included in the first class and maximum value in the last class. Class limits should be round figures, and class intervals should be non-overlapping and must include all observations.

Ex 10.Consider the following data which gives us the weights

52.5, 59.5, 49.5, 52.9, 57.4, 52.9, 64.7, 51.8, 61.3, 71.4, 50.7, 73.5, 58.7, 61.8, 62.8, 56.6, 69.0, 56.4, 62.8, 47.8, 55.4, 69.9, 48.1, 51.2, 62.5, 57.1, 64.3, 45.6, 64.8, 60.9, 57.2, 56.8, 50.5, 63.4, 49.2, 61.2, 56.6, 67.6, 61.7, 45.1.

Number of observations =N=40 Maximum=73.5 minimum=45.1

Number of classes=K= 1+3.332log N=6.33 = 6 (appr)

Width of the class=(max-min)/K=(73.5-45.1)/6=4.7=5 (appr)

So data has to be classified into 6 classes each class of width 5.

So we can write classes as 45-50, 50-55, 55-60…

Class mark = Mid point of the Class = (Lower limit + Upper limit)/2

Class	Class mark	Tally mark	Frequency
45-50	47.5	𝍸 |	6
50-55	52.5	𝍸 ||	7
55-60	57.5	𝍸 𝍸	10
60-65	62.5	𝍸 𝍸 ||	12
65-70	67.5	|||	3
70-75	72.5	||	2

Bivariate classification

Many a times we have to study two variables simultaneously. Eg height and weight, income and expenditure, Age of Husband and age of wife, marks scores in two subjects….. In such cases our data consists of paired observations (Xi ,Yi) i= 1, 2,3,……..n. The frequency distribution results in cross classification which gives bi variate frequency distribution.. the bivariate frequency table is two way table.

Ex.11 The following data gives marks scored in two test (X,Y) by 20 students

(10,12) (12 ,11) (7 , 10) (15 ,19) (17 ,21) (12 , 8) (16 ,10) (14 ,14) (22 ,18) (16 ,14) (15 ,16) (22 ,20) (19 ,15) (7 ,18) (11 ,11) (12 ,18) (10 ,10) (5 ,13) (11 ,7) (10 ,10)

Taking classes 0-5, 5-10, 10-15…….. bi vaiate table can be prepared.

Y -> X|	5-10	10-15	15-20	20-25	Total
5-10	-	2	1	-	3
10-15	2	6	1	-	9
15-20	-	2	3	1	6
20-25	-	-	1	1	2
Total	2	10	6	2	20

The total column at the extreme left is the marginal frequency column for the variate X

Similarly total row at the bottom is the marginal frequencies for variate Y.

Marginal frequency table for X

Marks in test I	No. of students
5-10	3
10-15	9
15-20	6
20-25	2
Total	20

Marginal frequency table for Y

Marks in test II	No. of students
5-10	2
10-15	10
15-20	6
20-25	2
Total	20

Stem and leaf display: stem and leaf display is used to represent ungrouped data. The steps involved for arranging the data is as follows:

1. Divide each value of the observation into two parts. One part consisting of one or more digit as stem and rest digits as leaf.
2. The stem values are listed on the left of the vertical line and each leaf value corresponding to the stem is written in horizontal line to the right of the stem in the increasing order.

The stem and leaf display gives us the ordered data and the shape of the distribution..

Ex 12) data values are 42, 53, 65, 63, 61, 77, 47, 56, 74, 60, 64, 68, 45, 55, 57, 82, 42, 35, 39, 51, 65, 55, 33, 76, 70, 50, 52, 54, 45, 46, 25, 36, 59, 63, 83.

Stem	leaves
2	5
3	3, 5, 6, 9.
4	2, 2, 5, 5, 6, 7, 9
5	0, 1, 2, 3, 3, 4, 5, 5, 6, 7
6	0, 1, 3, 4, 5, 5, 8
7	0, 4, 6, 7
8	2,3

GRAPHS

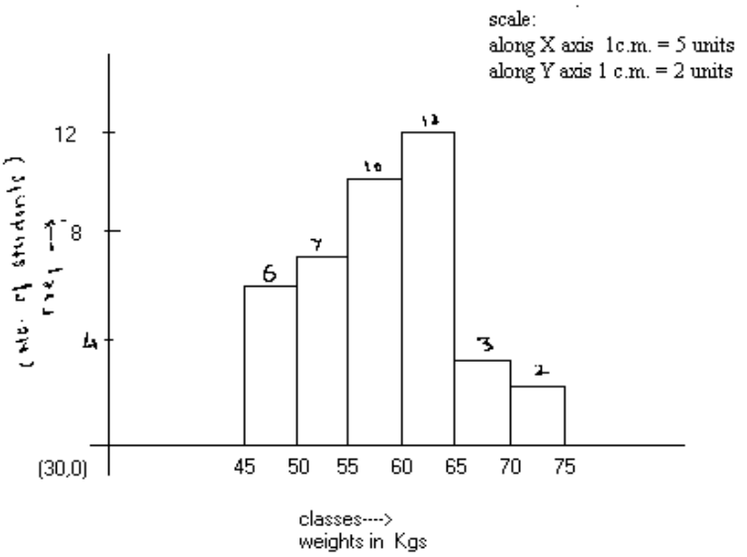
A frequency distribution can be represented graphically by

1. Histogram 2. Frequency curve 3. Frequency polygon.

1. Histogram In this diagram we represent each class by vertical bar (rectangle)

To construct histogram classes should be continuous. If width of the classes are equal then we take frequency along Y- axis .Variable is taken along X axis Then for each class interval bar of height corresponding to frequency is drawn.

For the above data (Ex 10) histogram is



If classes are of unequal width then we calculate frequency density for each class.

Frequency density=(freq /width)*constant

We take frequency density along Y axis. Variable is taken along X axis. Then for each class interval bar of height corresponding to frequency density is drawn

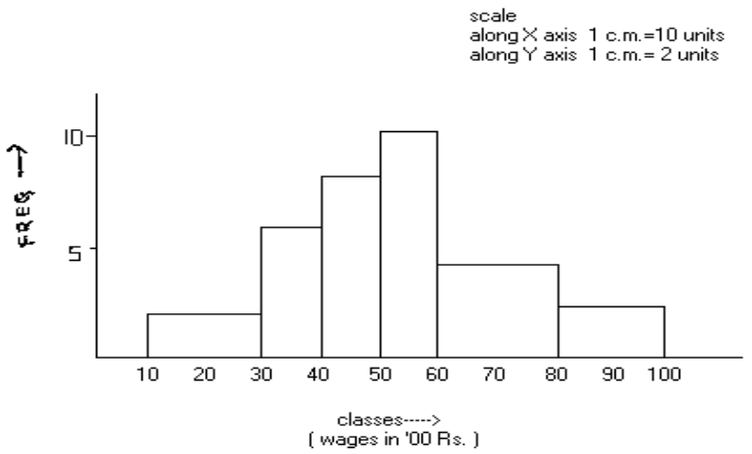
Ex 13) Following data represents monthly wages of workers in a factory.

Wages (in '00Rs) :10-30 30-40 40-50 50-60 60-80 80-100

No. of Workers: 4 6 8 10 8 4

Here classes are of unequal width, so to draw histogram we have to calculate frequency density for each class.

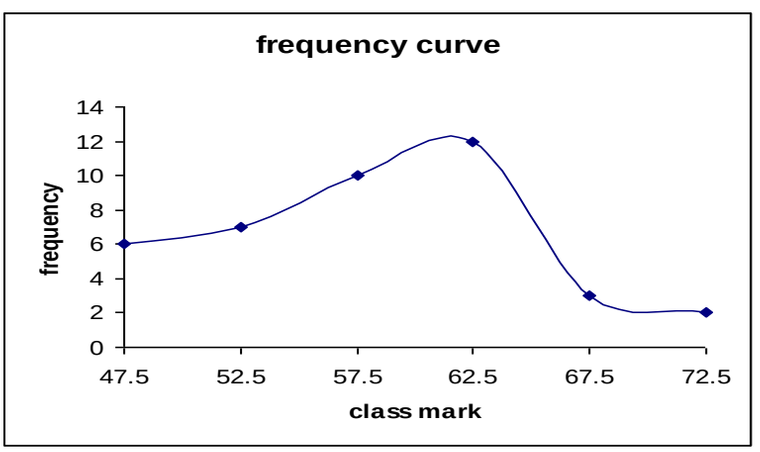
Wages (in ,00 Rs.)	No. of workers freq	Class width d	Freq. density $= \frac{freq}{d} * 10$
10-30	4	20	2
30-40	6	10	6
40-50	8	10	8
50-60	10	10	10
60-80	8	20	4
80-100	4	20	2



2) Frequency curve

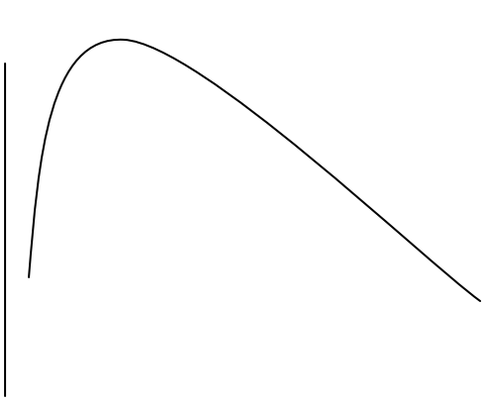
To construct frequency curve we take the variable along X-axis and frequency along Y-axis. Class mark (x) along with corresponding frequency (f) are plotted on the graph. Then these points are joined by a smooth curve . Such curve is known as 'frequency curve'.

Frequency curve for the data in Ex 10.

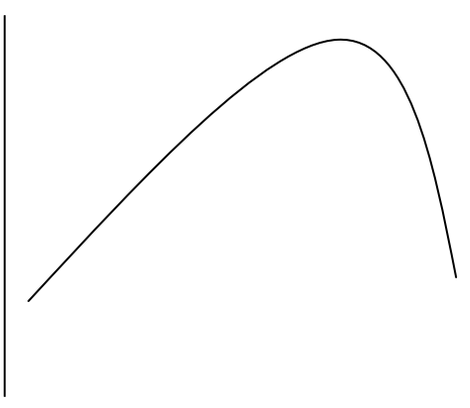


By observing frequency curve we can classify it in one of the following curves.

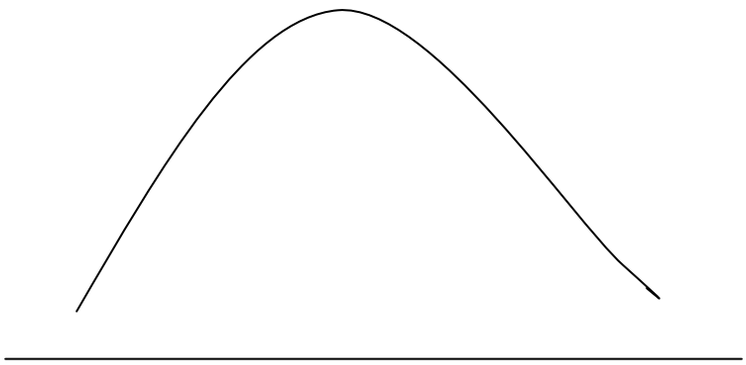
1)Positively skewed curve 2) Negatively skewed curve 3) Symmetric curve or zero skewed curve.



Positively skewed curve



Negatively skewed curve



Symmetric curve

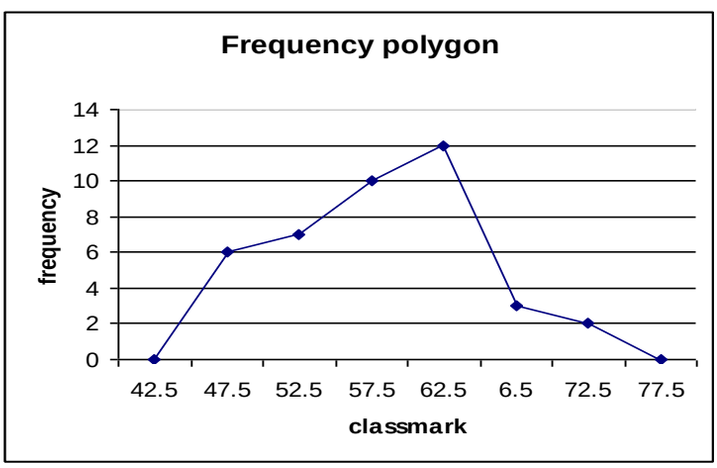
i) If peak is on the right side the curve is said to be positively skewed curve

ii)If peak is on the left hand side the curve is said to be negatively skewed curve

iii) If peak is at the centre the curve is said to be symmetric curve

3) **Frequency polygon**

Instead of joining these points by smooth curve if we join consecutive points by a line segment , then the diagram is known as frequency polygon. Here we have to add two points one preceding to the least and other succeeding the highest value and plot them with frequency zero, so frequency polygon is a closed figure. Frequency polygon for the data in Ex 1.



Ogive

To draw ogive we have to consider cumulative frequencies. An ogive is a graph that represents the cumulative frequencies for the classes in a frequency distribution. It shows how many of values of the data are below or above certain boundary. Ogives are two types 1) Ogive less than type 2) ogive more than or equal to type.

1) While drawing less than type ogive we have to consider class boundaries along X-axis and corresponding less than type cumulative frequencies along Y-axis. Less than

cumulative frequency represents the number of observations less than upper class boundary of the class interval. This is an increasing curve.

2) While drawing more than or equal to type ogive we have to consider class boundaries along X-axis and corresponding more than or equal to type cumulative frequencies along Y-axis. more than or equal to cumulative frequency represents the number of observations more than or equal to lower class boundary of the class interval. This is decreasing curve.

In both cases points are plotted on the graph and are joined by a smooth curve .

Ogives are useful in many ways-

1) The point of intersection of less than and more than ogive gives us the median. We can locate other partition values.
2) Ogives can be used to compare two or more distributions.
3) Ogives can be used to determine as well as to show the number of proportion cases below and above given value.

Ex14 For the following data draw ogives

Class	20-25	25-30	30-35	35-40	40-45	45-50
frequency	12	24	43	38	22	11

Class boundary	20	25	30	35	40	45	50
Cum freq less than type	0	12	36	79	117	139	150
Cum freq more than or equal to type	150	138	114	71	33	11	0