UNIT 15 GRAPHICAL PRESENTATION OF DATA-I

Structure

15.1 Introduction
Objectives
15.2 Graphical Presentation
15.3 Types of Graphs
Histogram
Frequency Polygon
Frequency Curve
Ogive
15.4 Summary
15.5 Solutions/Answers

15.1 INTRODUCTION

An important function of Statistics is to present the complex and huge data in such a way that they can easily understandable. In previous unit, we have discussed the diagrammatic presentation of the data where we have become familiar with some of the most commonly used diagrams. After discussing the diagrammatic presentation of data, we are now moving towards the graphical presentation of data. The graphs are plotted for frequency distributions and are used to interpolate/extrapolate items in a series including locating various partition values. In this unit, we shall discuss some of the most useful and commonly used graphs.

The graphical presentation can be divided into two categories

(i) Graphs for frequency distributions.

(ii) Graphs for time series.

In this unit, we will concentrate ourselves to the graphs for frequency distributions only. In this regard, we would like to discuss the most commonly used graphs for frequency distributions, i.e. Histograms, Frequency polygon, Frequency curve and Cumulative frequency curves, or Ogives.

Objectives

After studying this unit, you would be able to:

- describe the graphical presentation;
- explain the advantages of graphical presentation;
- draw the histogram for continuous frequency distribution;
- draw the frequency polygon for a frequency distribution;
- draw the frequency curves of different shapes; and
- draw the cumulative frequency curves.

15.2 GRAPHICAL PRESENTATION

A graphical presentation is a geometric image of a set of data. Graphical presentation is done for both frequency distributions and times series. Unlike diagrams, they are used to locate partition values like median, quartiles, etc, in particular, and interpolate/extrapolate items in a series, in general. They are also used to measure absolute as well as relative changes in the data. Another important feature of graphs is that if a person once sees the graphs, the figure representing the graphs is kept in his/her brain for a long time. They also help us in studying cause and affect relationship between two variables. The graph of a frequency distribution presents the huge data in an interesting and effective manner and brings to light the salient features of the data at a glance. Before closing this Sec. let us see some advantages of graphical presentation.

Advantages of Graphical Presentation

The following are some advantages of the graphical presentation:

- It simplifies the complexity of data and makes it readily understandable.
- It attracts attention of people.
- It saves time and efforts to understand the facts.
- It makes comparison easy.
- A graph describes the relationship between two or more variables.

After going through the advantages of graphical presentation of data, you were keen to know the commonly used graphs to represent the data and how these graphs are drawn. Next section will address these issues.

15.3 TYPES OF GRAPHS

Now days a large variety of graphs are in practical use. However, we shall discuss only some important graphs which are more popularly used in practice. The various types of graphs can be divided broadly under the following two heads:

(i) Graphs of Frequency Distributions

(ii) Graphs for Time Series Data

In this Sec. we will focus on graphs for frequency distributions and graphs for time series data will be discussed in next unit, i.e. Unit 16.

Graphs of Frequency Distributions

The graphical presentation of frequency distributions is drawn for discrete as well as continuous frequency distributions.

Let us first consider the frequency distribution of a discrete variable.

To represent a discrete frequency distribution graphically, we take two rectangular axes of co-ordinates, the horizontal axis for the variable and the vertical axis for the frequency. The different values of the variable are then located as points on the horizontal axis. At each of these points, a perpendicular bar is drawn to present the corresponding frequency.

Such a diagram is called a ‘Frequency Bar Diagram’. For example, if we take the frequency distribution for the number of peas per pod for 198 pods as given in Table 15.1:

Table 15.1

No of peas per pod	1	2	3	4	5	6	7
Frequency (number of pods)	14	23	66	40	26	18	11

Then, the frequency bar diagram is shown in Fig. 15.1:

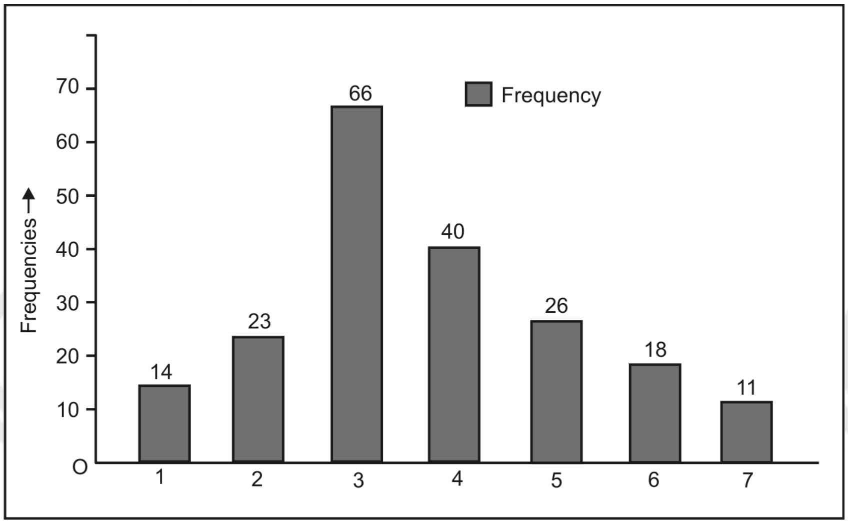


Fig. 15.1: Frequency Bar Diagram for the Frequency Distribution of Number of the Peas for 198 Pods.

Note 1: O represents origin and choice of scale used along horizontal and vertical axes depends upon given data.

Now, we take the case of frequency distribution of a continuous variable.

The following are the most commonly used graphs for continuous frequency distributions:

(i) Histogram

(ii) Frequency Polygon

(iii) Frequency Curve

(iv) Cumulative Frequency Curve or Ogives

Let us discuss these one by one:

(i) Histogram

In previous example, we have discussed how a graph is drawn for discrete frequency distribution.

For the continuous frequency distribution, a better way to represent the data graphically is to use a histogram. A histogram is drawn by constructing adjacent rectangles over the class intervals such that the length of the rectangles is proportional to the corresponding class-frequencies.

Histogram is similar to a bar diagram which represents a frequency distribution with continuous classes. The width of all bars is equal to class interval. Each rectangle is joined with the other so as to give a continuous picture.

The class-boundaries are located on the horizontal axis. If the class-intervals are of equal size, the heights of the rectangles will be proportional to the class-frequencies themselves. If the class-intervals are not of equal size, the heights of the rectangles will be proportional to the ratios of the frequencies to the width of the corresponding classes. In other words, the frequencies of the class-intervals having the least width are written as they are and the frequencies of other class intervals are written as follows:

$$\frac{\text{Given frequency}}{\text{Width of its Class-interval}} \times (\text{The least width}) \qquad \ldots (15.1)$$

Let us draw a histogram to the following frequency distribution given below in the table 15.2

Table 15.2

Class Intervals	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	2	3	13	18	9	7	6	2

Histogram for the above data is given below.

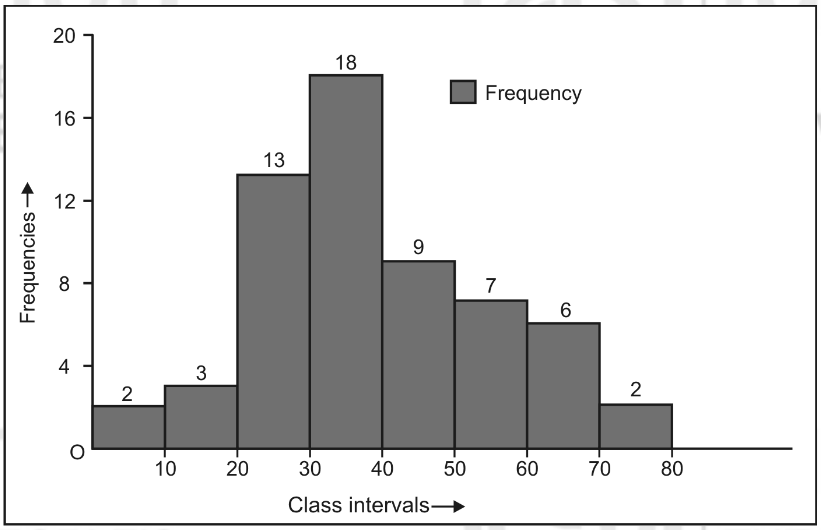


Fig. 15.2: Histogram for Frequency Distribution when Class-intervals are of Equal Width.

Now, let us consider the frequency distribution for unequal class intervals as given in the Table 15.3

Table 15.3

Class	0-10	10-20	20-30	30-40	40-70	70-80	80-100
Frequency	20	32	8	2	60	35	10

As it is a case of unequal class intervals, so we have to adjust the frequencies of the classes 40-70 and 80-100 by the formula suggested in equation 15.1. These calculations are shown in table 15.4 given below:

Table 15.4

Class Interval (CI)	Frequency	Width of (CI)	Heights of the rectangles
0-10	20	10	20
10-20	32	10	32
20-30	8	10	8
30-40	2	10	2
40-70	60	30	$(60/30)\times 10 = 20$
70-80	35	10	35
80-100	10	20	$(10/20) \times 10 = 5$

The histogram for this frequency distribution is shown in Fig. 15.3.

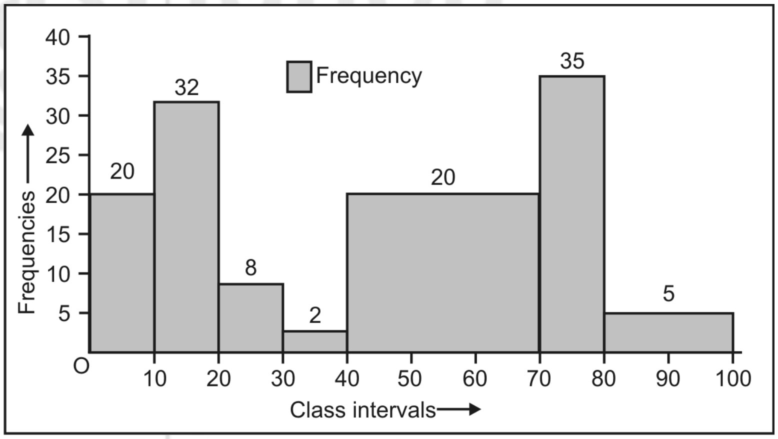


Fig. 15.3: Histogram for Frequency Distribution when Class Intervals are of Unequal Width

Note 2: Sometimes, a histogram is also used for the frequency distribution of a discrete variable. Each value of the discrete variable is regarded as the mid-point of an interval. But generally, its use is not recommended, because in discrete case each frequency actually corresponds to a single point and not to an interval.

Now, you can try the following exercises.

E1) Draw a histogram from the following data

Class Interval:	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Frequency:	3	5	10	14	24	17	14	10	3

E2) Draw a histogram for the following frequency distribution

Wage (Rs):	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
No. of Workers:	30	70	100	110	140	150	130	100	90	60

(ii) Frequency Polygon

Another method of presenting a frequency distribution graphically other than histogram is to use a frequency polygon. In order to draw the graph of a frequency polygon, first of all the mid values of all the class intervals and the corresponding frequencies are plotted as points with the help of the rectangular co-ordinate axes. Secondly, we join these plotted points by line segments. The graph thus obtained is known as frequency polygon, but one important point to keep in mind is that whenever a frequency polygon is required we take two imaginary class intervals each with frequency zero, one just before the first class interval and other just after the last class interval. Addition of these two class intervals facilitate the existence of the property that

Area under the polygon = Area of the histogram

For example, if we take the frequency distribution as given in Table 15.2 then, we have to first plot the points (5, 2), (15, 3), …, (75, 2) on graph paper along with the horizontal bars. Then we join the successive points (including the mid points of two imaginary class intervals each with zero frequency) by line segments to get a frequency polygon. The frequency polygon for frequency distribution given in Table 15.2 is shown in Fig. 15.4.

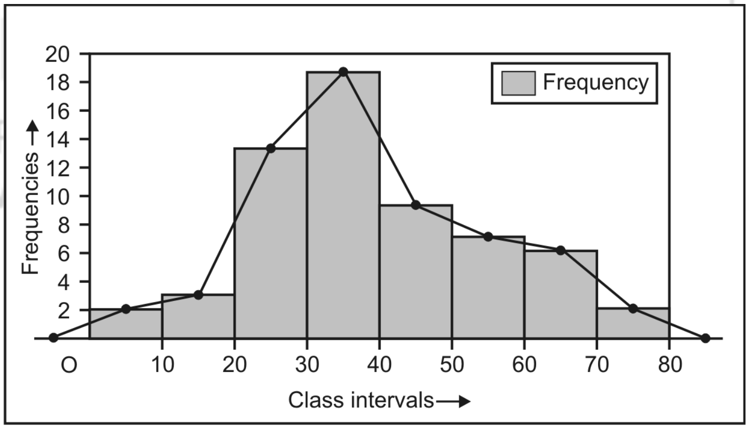


Fig. 15.4: Frequency Polygon for the Frequency Distribution given in Table 15.2.

Note 3: In some cases first class interval does not start from zero. In such situations we mark a kink on the horizontal axis, which will indicates the continuity of the scale starting from zero. Let us take an example of this type.

Example 1: Draw a frequency polygon for the following frequency distribution:

Class Interval	40-50	50-60	60-70	70-80	80-90	90-100	100-110	110-120
Frequency	4	10	11	13	18	14	11	5

Solution: Frequency polygon for the given data is shown in Fig. 15.5:

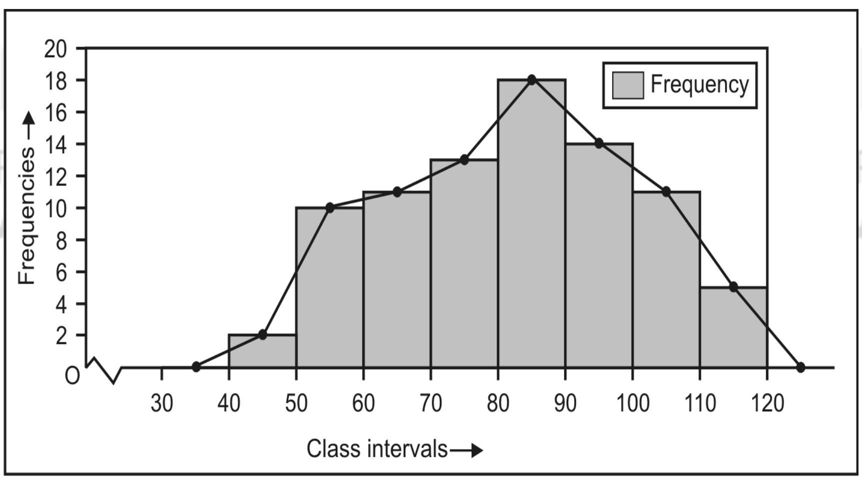


Fig. 15.5: Frequency Polygon for Continuous Frequency Distribution.

Now, you can try the following exercise.

E3) Draw a frequency polygon from the following data

Wage:	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
No of Workers:	20	50	80	110	140	150	120	150	100	80

(iii) **Frequency Curve**

In simple words frequency curve is a smooth curve obtained by joining the points (not necessary all points) of the frequency polygon such that

(a) Like frequency curve it also starts from the base line (horizontal axis) and ends at the base line.

(b) Area under frequency curve remains approximately equal to the area under the frequency polygon.

In other words, let us try to explain the concept theoretically. Suppose we draw a sample of size n from a large population. Frequency curve is the graph of a continuous variable. So theoretically continuity of the variable implies that whatever small class interval we take there will be some observations in that class interval. That is, in this case there will be large number of line segments and the frequency polygon tends to coincides with the smooth curve passing through these points as sample size (n) increases. This smooth curve is known as frequency curve.

In the following example we have drawn both frequency polygon and frequency curve to make the idea clear for you.

Example 2: Draw frequency polygon and frequency curve for the following frequency distribution.

Class Intervals	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
frequency	2	5	8	15	18	10	3	1

Solution: Frequency polygon and frequency curve for the above data is given below in Fig. 15.6.

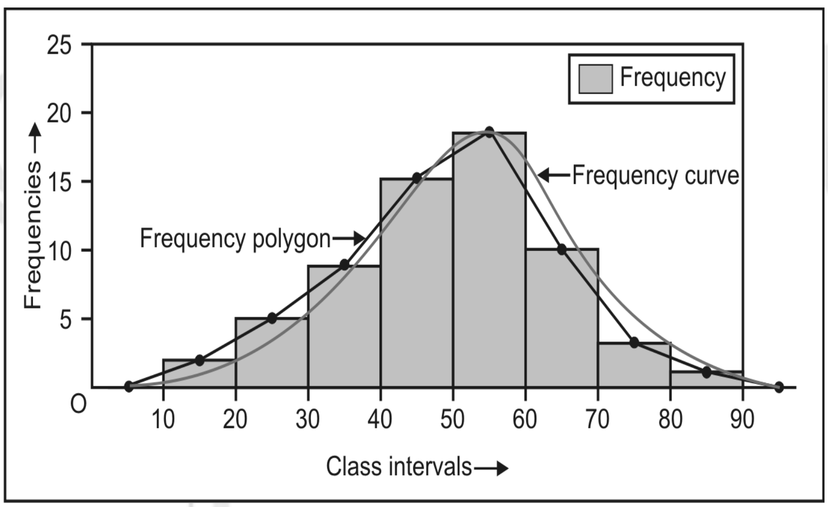


Fig. 15.6: Frequency Curve along with Frequency Polygon.

On the next page some important types of frequency curves are given which are generally obtained in the graphical presentations of frequency distributions. That is, symmetrical, positively skewed, negatively skewed, J shaped, U shaped, bimodal and multimodal frequency curves. You note that the shapes of these curves justify their names.

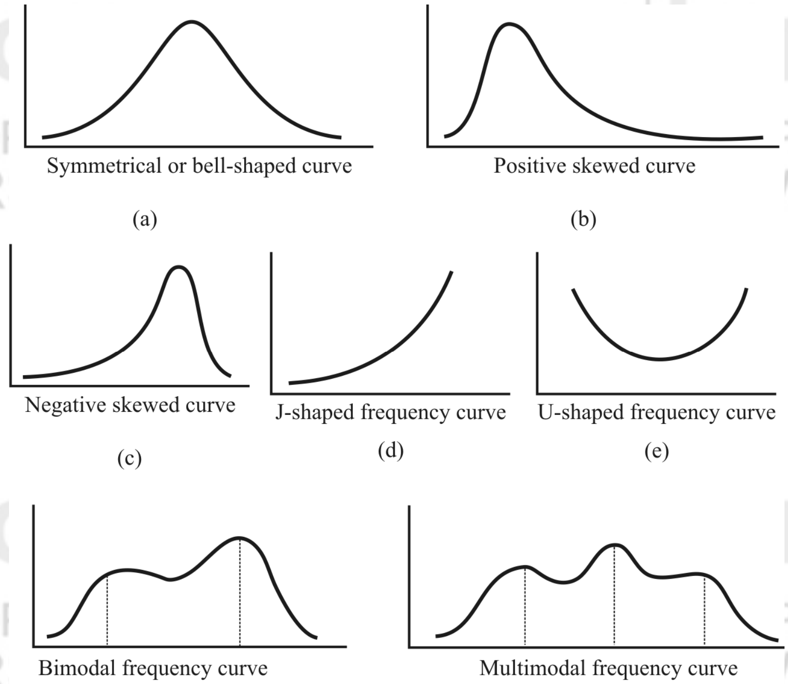


Fig. 15.7

Now, you can try the following exercise.

E4) Draw a frequency curve from the following frequency distribution

Class Intervals:	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9
Frequency:	3	9	20	25	18	12	6	3

(iv) Cumulative Frequency Curves

In sub Sec. 13.2.2 of the Unit 13 of this course you have already studied the concept of cumulative frequency, cumulative frequency distribution, less than cumulative frequency, more than cumulative frequency, less than cumulative frequency distribution, more than cumulative frequency distribution, etc. Here our aim is to study graphical presentation of less than and more than cumulative frequency distributions, which are known as less than frequency curve (or less than ogive) and more than frequency curve (or more than ogive) respectively.

For drawing less than cumulative frequency curve (or less than ogive), first of all the cumulative frequencies are plotted against the values (upper limits of the class intervals) up to which they correspond and then we simply join the points by line segments, curve thus obtained is known as less than ogive. Similarly, more than frequency curve (more than ogive) can be obtained by plotting more than cumulative frequencies against lower limits of the class intervals. As we have already mentioned within brackets that less than cumulative frequency curve and more than cumulative frequency curve are also called **less than ogive** and **more than ogive** respectively.

In other words we may define less than ogive and more than ogive as follow:

Less Than Ogive: If we plot the points with the upper limits of the classes as abscissae and the cumulative frequencies corresponding to the values less then the upper limits as ordinates and join the points so plotted by line segments, the curve thus obtained is nothing but known as "less than cumulative frequency curve" or "less than ogive". It is a rising curve.

More Than Ogive: If we plot the points with the lower limits of the classes as abscissae and the cumulative frequencies corresponding to the values more than the lower limits as ordinates and join the points so plotted by line segments, the curve thus obtained is nothing but known as "more than cumulative frequency curve" or "more than ogive". It is a falling curve.

Let us draw both the ogives ('less than' and 'more than') for the following frequency distribution of the weekly wages of number of workers given in Table 15.5.

Table 15.5

Weekly wages	0-10	10-20	20-30	30-40	40-50
No. of workers	45	55	70	40	10

Before drawing the ogives, we make a cumulative frequency distribution as given in table 15.6

Table 15.6

Weekly wages	No. of workers	Less than Cumulative frequency distribution		More than Cumulative frequency distribution	
		Wages Less than	Number of workers	Wages More than	Number of workers
0-10	45	10	45	0	220
10-20	55	20	100	10	175
20-30	70	30	170	20	120
30-40	40	40	210	30	50
40-50	10	50	220	40	10

From above data, we construct both the ogives as shown in Fig. 15.8 and Fig. 15.9:

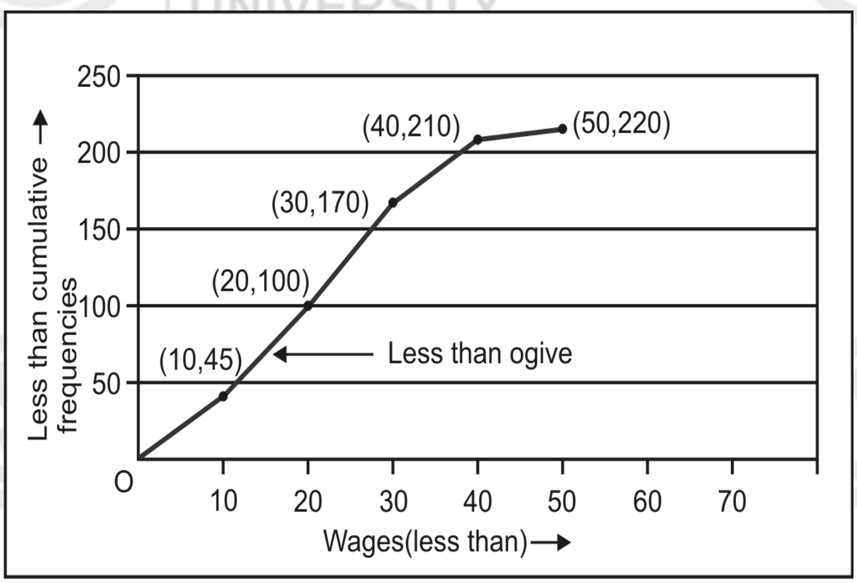


Fig. 15.8: Less Than Ogive.

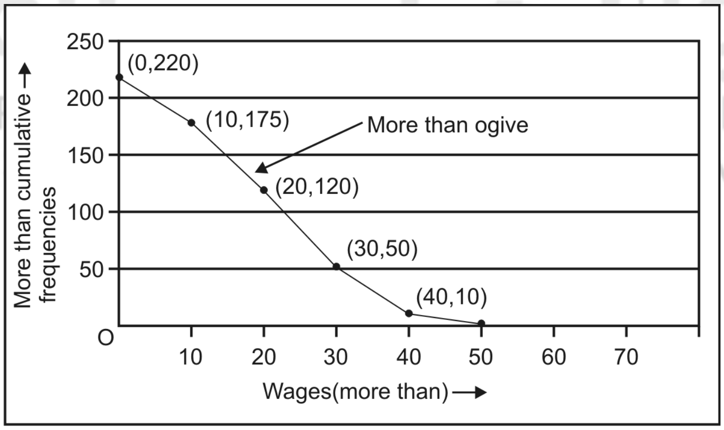


Fig. 15.9: More Than Ogive.

For "less than ogive" as shown on previous page in Fig. 15.8, we have plotted the points (10, 45), (20,100), (30, 170), (40, 210), (50, 220) and then joined them by line segments. Similarly, for "more than ogive" as shown above in Fig. 15.9, we have plotted the points (0, 220), (10, 175), (20, 120), (30, 50), (40, 10), and then joined them by line segments.

If we want to obtain a partition value, using ogives, we draw dotted horizontal line through that value at y-axis which corresponds to the partition value and then from the point, where it meets the less then ogive, we draw a dotted vertical line and let it meets the x-axis. The abscissa of the point, where it meets the x-axis is the required partition value. For example, suppose we want to find first quartile, then we draw a dotted horizontal line starting from y-axis at a point corresponding to N/4 and let it meets the "less than ogive". From that point at "less than ogive", we draw a dotted vertical line and let it meets the x-axis. The abscissa corresponding to this point is the first quartile. Similarly, for finding median or second quartile, we start drawing dotted horizontal line from y-axis at a point corresponding to N/2 and then we proceed as described above. Similarly, for third quartile 3N/4 is taken in place of N/2. In this way, we may find any partition value.

Note 4: Median may also the obtained by drawing dotted vertical line through the point of inter section of both the ogives, when drawn on a single figure.

Now, you can try the following exercises.

E5) Draw two ogives from the following data

Class:	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Frequency:	3	6	10	13	20	18	15	9	6

Hence find median. Compare your result by calculating median by direct calculatios.

E6) Draw less than ogive from the following frequency distribution of marks of 90 students

Marks:	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79
No. of Students:	7	11	19	8	20	14	8	3

Hence find Q_1, Q_2 and Q_3.

E7) Draw the more than ogive for the following frequency distribution of the weekly wages of workers:

Weekly wages:	0-10	10-20	20-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of Workers:	5	15	20	30	45	35	25	15	10

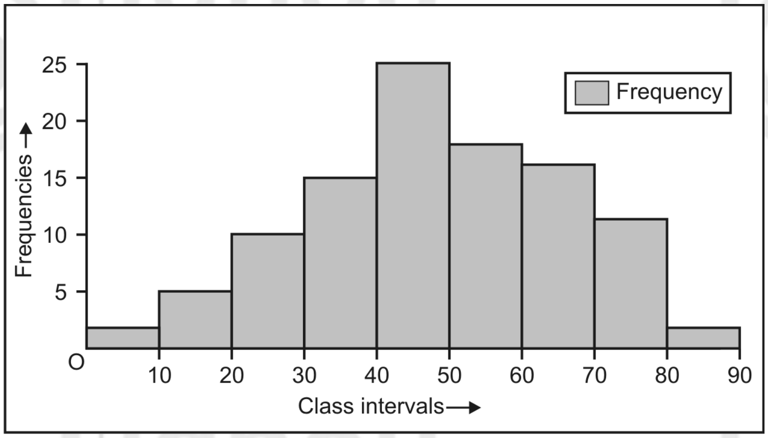
15.4 SUMMARY

In this unit we have discussed:

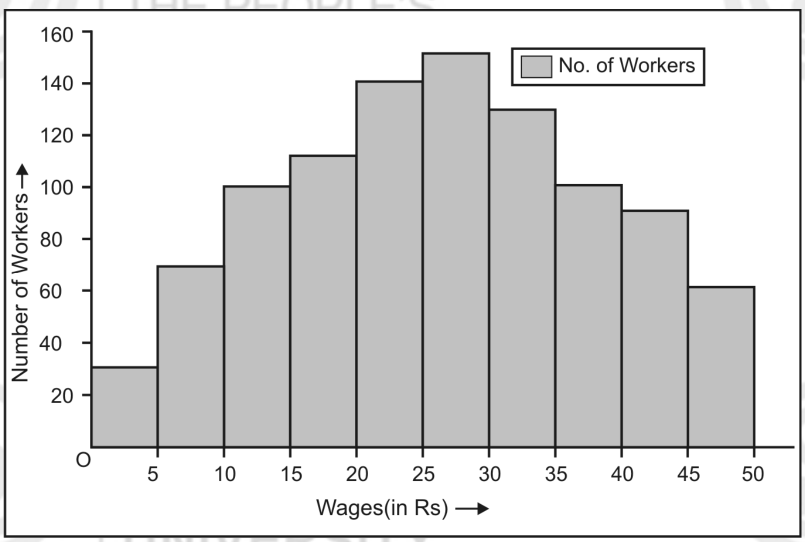
1) Various types of graphical presentation of data.
2) Way of drawing histogram for continuous frequency distributions.
3) Frequency polygon for a frequency distribution.
4) Frequency curves of different shapes, and
5) Way of drawing cumulative frequency graphs or ogives.

15.5 SOLUTIONS/ANSWERS

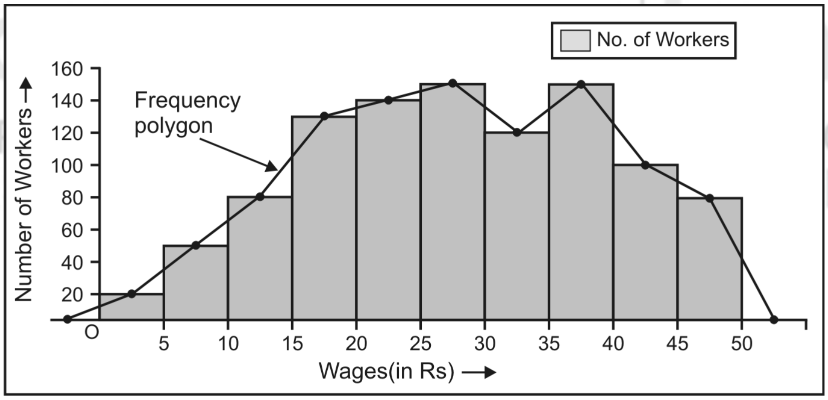
E1) Histogram of the given data is given below**:**



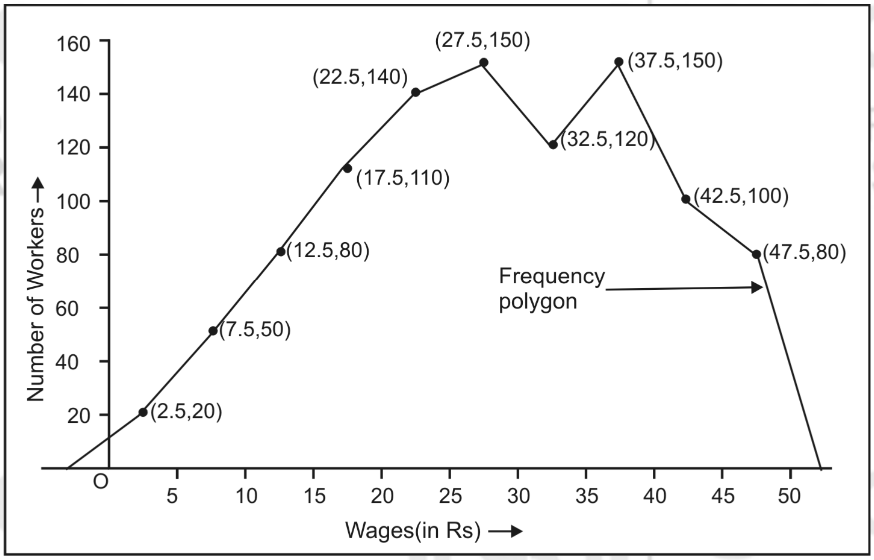
E2) Histogram of the given data is given below**:**



E3) Frequency polygon of the given data by first drawing histogram is given on the next page.

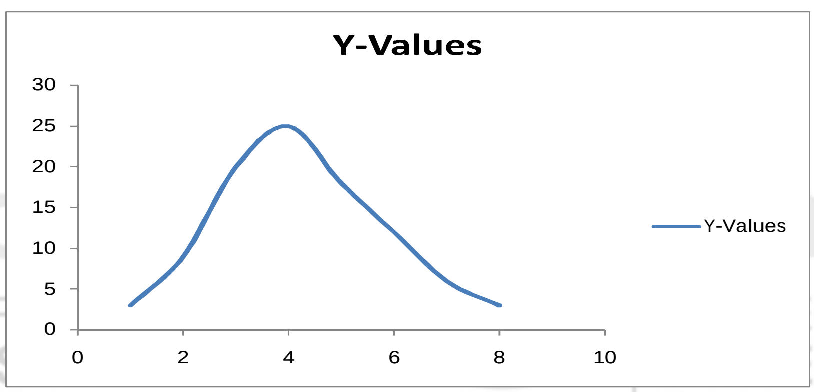


Frequency polygon can also be drawn without histogram as given below:

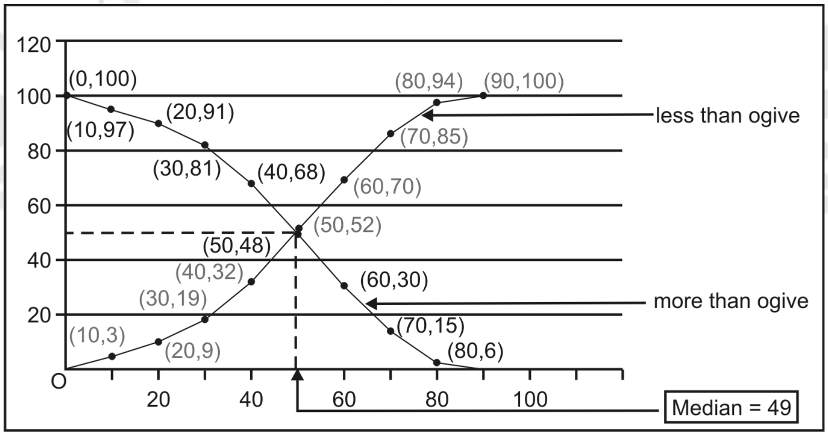


Note 5: We can draw frequency polygon by first drawing a histogram or we can draw directly.

E4) Frequency curve for the given data is given below.



E5) Two ogives for the given data are given on the next page.



Calculations of Median through direct calculation:

Given data are

Class Interval (CI)	Frequency (f)	Cumulative frequency (cf)
0-10	3	3
10-20	6	9
20-30	10	19
30-40	13	32
40-50	20	52
50-60	18	70
60-70	15	85
70-80	9	94
80-90	6	100
Total	100	

Here $N = \sum f$ = sum of all frequencies =100

$\therefore \frac{N}{2} = \frac{100}{2} = 50$ and hence median class is 40-50.

So, in usual notations we have

l = lower limit of the median class = 40,
h = width of the median class = $50 - 40 = 10$
c = cumulative frequency preceding the median class = 32
f = frequency of the median class = 20

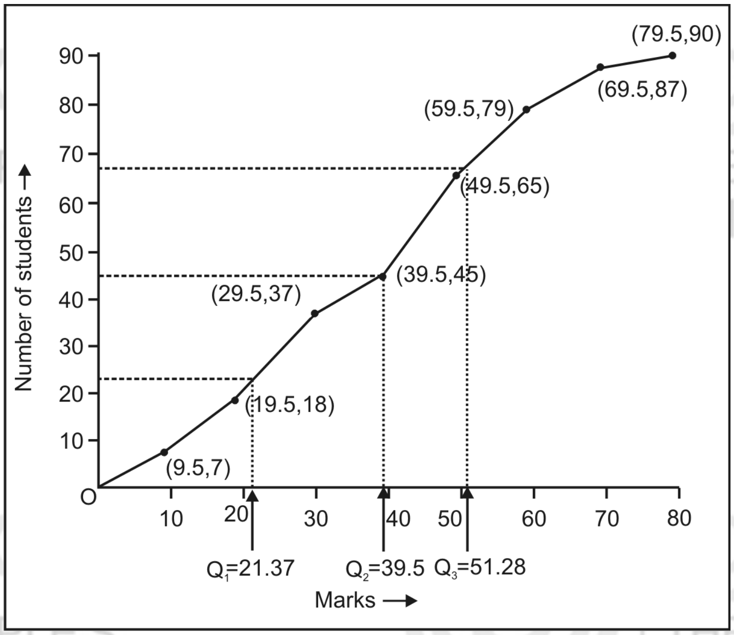
$$\text{Now, Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right) = 40 + \frac{10}{20}\left(\frac{100}{2} - 32\right) = 40 + \frac{1}{2}(50 - 32)$$

$$= 40 + \frac{1}{2}(18) = 40 + 9 = 49$$

Thus we see that median obtained by using two ogives graphically and by direct calculation are same.

E6) Less than ogive for the given frequency distribution is given below.

To find $Q_1 : \frac{N}{4} = \frac{90}{4} = 22.5$ first of all we have to draw a dotted horizontal line starting from y-axis at 22.5 $\left(\text{as} \frac{N}{4} = 22.5 \right)$ and then from the point where it meets the less ogive, we shall draw dotted vertical line, the value corresponding to the point where it meets the horizontal axis is the value of Q_1 as shown in figure. Similarly, values of Q_2 and Q_3 can be obtained as shown in the figure below:



E7) More than ogive for the given frequency distribution is given below: