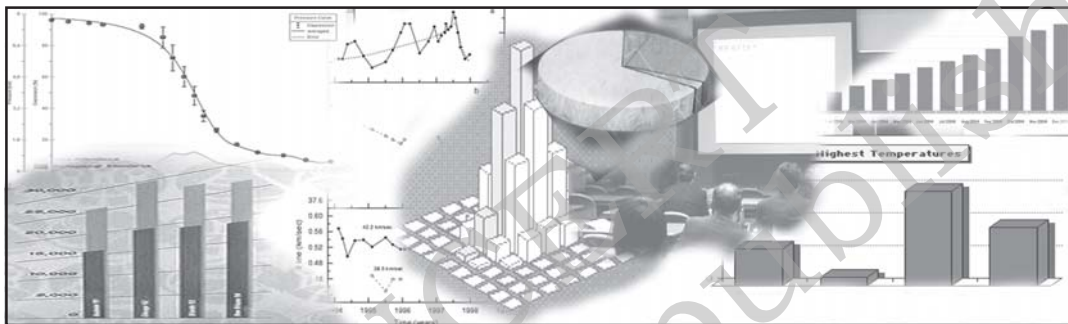


Presentation of Data



Studying this chapter should enable you to:

- present data using tables;
- represent data using appropriate diagrams.

1. INTRODUCTION

You have already learnt in previous chapters how data are collected and organised. As data are generally voluminous, they need to be put in a compact and presentable form. This chapter deals with presentation of data precisely so that the voluminous data collected could be made usable readily and are easily comprehended. There are generally three forms of presentation of data:

- Textual or Descriptive presentation
- Tabular presentation
- Diagrammatic presentation.

2. TEXTUAL PRESENTATION OF DATA

In textual presentation, data are described within the text. When the quantity of data is not too large this form of presentation is more suitable. Look at the following cases:

Case 1

In a *bandh* call given on 08 September 2005 protesting the hike in prices of petrol and diesel, 5 petrol pumps were found open and 17 were closed whereas 2 schools were closed and remaining 9 schools were found open in a town of Bihar.

Case 2

Census of India 2001 reported that Indian population had risen to 102 crore of which only 49 crore were females against 53 crore males. 74 crore people resided in rural India and only 28 crore lived in towns or cities. While there were 62 crore non-worker population against 40 crore workers in the entire country, urban population had an even higher share of non-workers (19 crores) against the workers (9 crores) as compared to the rural population where there were 31 crore workers out of a 74 crore population....

In both the cases data have been presented only in the text. A serious drawback of this method of presentation is that one has to go through the complete text of presentation for comprehension but at the same time, it enables one to emphasise certain points of the presentation.



3. TABULAR PRESENTATION OF DATA

In a tabular presentation, data are presented in rows (read horizontally) and columns (read vertically). For example see Table 4.1 below tabulating information about literacy rates. It has

3 rows (for male, female and total) and 3 columns (for urban, rural and total). It is called a 3×3 Table giving 9 items of information in 9 boxes called the "cells" of the Table. Each cell gives information that relates an attribute of gender ("male", "female" or total) with a number (literacy percentages of rural people, urban people and total). The most important advantage of tabulation is that it organises data for further statistical treatment and decision-making. Classification used in tabulation is of four kinds:

- Qualitative
- Quantitative
- Temporal and
- Spatial

Qualitative classification

When classification is done according to qualitative characteristics like social status, physical status, nationality, etc., it is called qualitative classification. For example, in Table 4.1 the characteristics for classification are sex and location which are qualitative in nature.

TABLE 4.1

Literacy in Bihar by sex and location (per cent)

Sex	Location		Total
	Rural	Urban	
Male	57.70	80.80	60.32
Female	30.03	63.30	33.57
Total	44.42	72.71	47.53

Source: Census of India 2001, Provisional Population Totals.

Quantitative classification

In quantitative classification, the data are classified on the basis of

characteristics which are quantitative in nature. In other words these characteristics can be measured quantitatively. For example, age, height, production, income, etc are quantitative characteristics. Classes are formed by assigning limits called class limits for the values of the characteristic under consideration. An example of quantitative classification is Table 4.2.

TABLE 4.2
Distribution of 542 respondents by their age in an election study in Bihar

Age group (yrs)	Nb. of respondents	Per cent
20-30	3	0.55
30-40	61	11.25
40-50	132	24.35
50-60	153	28.24
60-70	140	25.83
70-80	51	9.41
80-90	2	0.37
All	542	100.00

Source: Assembly election Patna central constituency 2005, A.N. Sinha Institute of Social Studies, Patna.

Here classifying characteristic is age in years and is quantifiable.

Activities

- Construct a table presenting data on preferential liking of the students of your class for Star News, Zee News, BBC World, CNN, Aaj Tak and DD News.
- Prepare a table of
 - heights (in cm) and
 - weights (in kg) of students of your class.

Temporal classification

In this classification time becomes the classifying variable and data are categorised according to time. Time may be in hours, days, weeks, months, years, etc. For example, see Table 4.3.

TABLE 4.3
Yearly sales of a tea shop from 1995 to 2000

Years	Sale (Rs in lakhs)
1995	79.2
1996	81.3
1997	82.4
1998	80.5
1999	100.2
2000	91.2

Data Source: Unpublished data.

In this table the classifying characteristic is year and takes values in the scale of time.

Activity

- Go to your library and collect data on the number of books in economics, the library had at the end of the year for the last ten years and present the data in a table.

Spatial classification

When classification is done in such a way that place becomes the classifying variable, it is called spatial classification. The place may be a village/town, block, district, state, country, etc.

Here the classifying characteristic is country of the world. Table 4.4 is an example of spatial classification.

TABLE 4.4
**Export from India to rest of the world in
 one year as share of total export (per cent)**

<i>Destination</i>	<i>Export share</i>
USA	21.8
Germany	5.6
Other EU	14.7
U K	5.7
Japan	4.9
Russia	2.1
Other East Europe	0.6
OPEC	10.5
Asia	19.0
Other LDCs	5.6
Others	9.5
<i>All</i>	<i>100.0</i>

(Total Exports: US \$ 33658.5 million)

Activity

- Construct a table presenting data collected from students of your class according to their native states/residential locality.

4. TABULATION OF DATA AND PARTS OF A TABLE

To construct a table it is important to learn first what are the parts of a good statistical table. When put together in a systematically ordered manner these parts form a table. The most simple way of conceptualising a table may be data presented in rows and columns alongwith some explanatory notes. Tabulation can be done using one-way, two-way or three-way classification depending upon the number of characteristics involved. A good table should essentially have the following:

(i) Table Number

Table number is assigned to a table for identification purpose. If more than one table is presented, it is the table number that distinguishes one table from another. It is given at the top or at the beginning of the title of the table. Generally, table numbers are whole numbers in ascending order if there are many tables in a book. Subscripted numbers like 1.2, 3.1, etc. are also in use for identifying the table according to its location. For example, Table number 4.5 may read as fifth table of the fourth chapter and so on. (See Table 4.5)

(ii) Title

The title of a table narrates about the contents of the table. It has to be very clear, brief and carefully worded so that the interpretations made from the table are clear and free from any ambiguity. It finds place at the head of the table succeeding the table number or just below it. (See Table 4.5).

(iii) Captions or Column Headings

At the top of each column in a table a column designation is given to explain figures of the column. This is called caption or column heading. (See Table 4.5)

(iv) Stubs or Row Headings

Like a caption or column heading each row of the table has to be given a heading. The designations of the rows are also called stubs or stub items, and the complete left column is known as

stub column. A brief description of the row headings may also be given at the left hand top in the table. (See Table 4.5).

were non-workers in 2001. (See Table 4.5).

(v) Body of the Table

Body of a table is the main part and it contains the actual data. Location of any one figure/data in the table is fixed and determined by the row and column of the table. For example, data in the second row and fourth column indicate that 25 crore females in rural India

(vi) Unit of Measurement

The unit of measurement of the figures in the table (actual data) should always be stated alongwith the title if the unit does not change throughout the table. If different units are there for rows or columns of the table, these units must be stated alongwith 'stubs' or 'captions'. If figures are large, they should be rounded up and the method

Table 4.5 Population of India according to workers and non-workers by gender and location

Location	Gender	Workers			Non-worker	Total
		Main	Marginal	Total		
Rural	Male	17	3	20	18	38
	Female	6	5	11	25	36
	Total	23	8	31	43	74
Urban	Male	7	1	8	7	15
	Female	1	0	1	12	13
	Total	8	1	9	19	28
All	Male	24	4	28	25	53
	Female	7	5	12	37	49
	Total	31	9	40	62	102

Source : Census of India 2001

Foot note : Figures are rounded to nearest crore

(Note : Table 4.5 presents the same data in tabular form already presented through case 2 in textual presentation of data)

of rounding should be indicated (See Table 4.5).

(ii) **Source Note**

It is a brief statement or phrase indicating the source of data presented in the table. If more than one source is there, all the sources are to be written in the source note. Source note is generally written at the bottom of the table. (See Table 4.5).

(iii) **Footnote**

Footnote is the last part of the table. Footnote explains the specific feature of the data content of the table which is not self explanatory and has not been explained earlier.

Activities

- How many rows and columns are essentially required to form a table?
- Can the column/row headings of a table be quantitative?

5. DIAGRAMMATIC PRESENTATION OF DATA

This is the third method of presenting data. This method provides the quickest understanding of the actual situation to be explained by data in comparison to tabular or textual presentations. Diagrammatic presentation of data translates quite effectively the highly abstract ideas contained in numbers into more concrete and easily comprehensible form.

Diagrams may be less accurate but are much more effective than tables in presenting the data.

There are various kinds of diagrams in common use. Amongst them the important ones are the following:

- (i) Geometric diagram
- (ii) Frequency diagram
- (iii) Arithmetic line graph

Geometric Diagram

Bar diagram and pie diagram come in the category of geometric diagram for presentation of data. The bar diagrams are of three types - simple, multiple and component bar diagrams.

Bar Diagram

Simple Bar Diagram

Bar diagram comprises a group of equispaced and equiwidth rectangular bars for each class or category of data. Height or length of the bar reads the magnitude of data. The lower end of the bar touches the base line such that the height of a bar starts from the zero unit. Bars of a bar diagram can be visually compared by their relative height and accordingly data are comprehended quickly. Data for this can be of frequency or non-frequency type. In non-frequency type data a particular characteristic, say production, yield, population, etc. at various points of time or of different states are noted and corresponding bars are made of the respective heights according to the values of the characteristic to construct the diagram. The values of the characteristics (measured or counted)

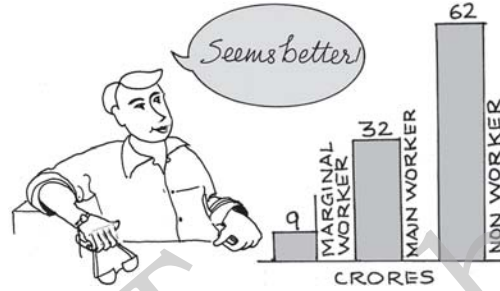
retain the identity of each value. Figure 4.1 is an example of a bar diagram.

Activity

- You had constructed a table presenting the data about the students of your class. Draw a bar diagram for the same table.

Different types of data may require different modes of diagrammatical representation. Bar diagrams are suitable both for frequency type and non-frequency type variables and attributes. Discrete variables like family size, spots on a dice, grades in an examination, etc. and attributes such as gender, religion, caste, country, etc. can be represented by bar diagrams. Bar diagrams are more convenient for non-frequency data such as income-

expenditure profile, export/imports over the years, etc.



A category that has a longer bar (literacy of Kerala) than another category (literacy of West Bengal), has more of the measured (or enumerated) characteristics than the other. Bars (also called columns) are usually used in time series data (food grain produced between 1980-2000, decadal variation in work participation

TABLE 4.6
Literacy Rates of Major States of India

Major Indian States	2001			1991		
	Person	Male	Female	Person	Male	Female
Andhra Pradesh (AP)	60.5	70.3	50.4	44.1	55.1	32.7
Assam (AS)	63.3	71.3	54.6	52.9	61.9	43.0
Bihar (BR)	47.0	59.7	33.1	37.5	51.4	22.0
Jharkhand (JH)	53.6	67.3	38.9	41.4	55.8	31.0
Gujarat (GJ)	69.1	79.7	57.8	61.3	73.1	48.6
Haryana (HR)	67.9	78.5	55.7	55.8	69.1	40.4
Karnataka (KA)	66.6	76.1	56.9	56.0	67.3	44.3
Kerala (KE)	90.9	94.2	87.7	89.8	93.6	86.2
Madhya Pradesh (MP)	63.7	76.1	50.3	44.7	58.5	29.4
Chhattisgarh (CH)	64.7	77.4	51.9	42.9	58.1	27.5
Maharashtra (MR)	76.9	86.0	67.0	64.9	76.6	52.3
Orissa (OR)	63.1	75.3	50.5	49.1	63.1	34.7
Punjab (PB)	69.7	75.2	63.4	58.5	65.7	50.4
Rajasthan (RJ)	60.4	75.7	43.9	38.6	55.0	20.4
Tamil Nadu (TN)	73.5	82.4	64.4	62.7	73.7	51.3
Uttar Pradesh (UP)	56.3	68.8	42.2	40.7	54.8	24.4
Uttaranchal (UT)	71.6	83.3	59.6	57.8	72.9	41.7
West Bengal (WB)	68.6	77.0	59.6	57.7	67.8	46.6
India	64.8	75.3	53.7	52.2	64.1	39.3

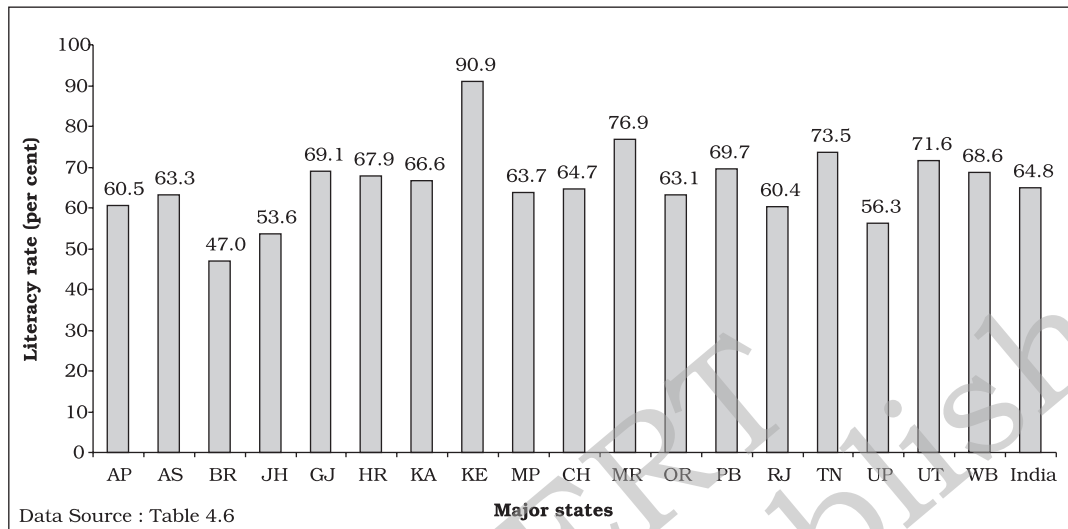


Fig. 4.1: Bar diagram showing literacy rates (person) of major states of India, 2001.

rate, registered unemployed over the years, literacy rates, etc.) (Fig 4.2).

Bar diagrams can have different forms such as multiple bar diagram and component bar diagram.

Activities

- How many states (among the major states of India) had higher female literacy rate than the national average in 2001?
- Has the gap between maximum and minimum female literacy rates over the states in two consecutive census years 2001 and 1991 declined?

Multiple Bar Diagram

Multiple bar diagrams (Fig.4.2) are used for comparing two or more sets of data, for example income and expenditure or import and export for

different years, marks obtained in different subjects in different classes, etc.

Component Bar Diagram

Component bar diagrams or charts (Fig.4.3), also called sub-diagrams, are very useful in comparing the sizes of different component parts (the elements or parts which a thing is made up of) and also for throwing light on the relationship among these integral parts. For example, sales proceeds from different products, expenditure pattern in a typical Indian family (components being food, rent, medicine, education, power, etc.), budget outlay for receipts and expenditures, components of labour force, population etc. Component bar diagrams are usually shaded or coloured suitably.

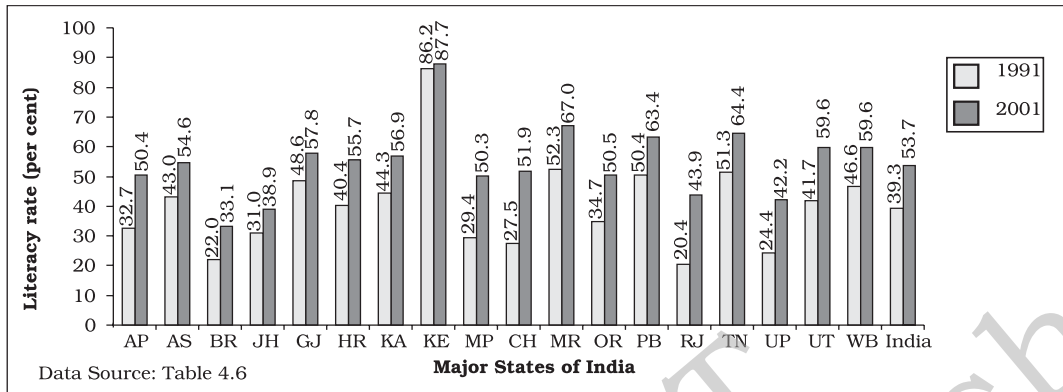


Fig. 4.2: Multiple bar (column) diagram showing female literacy rates over two census years 1991 and 2001 by major states of India.

Interpretation: It can be very easily derived from Figure 4.2 that female literacy rate over the years was on increase throughout the country. Similar other interpretations can be made from the figure like the state of Rajasthan experienced the sharpest rise in female literacy, etc.

TABLE 4.7
Enrolment by gender at schools (per cent) of children aged 6-14 years in a district of Bihar

Gender	Enrolled (per cent)	Out of school (per cent)
Boy	91.5	8.5
Girl	58.6	41.4
All	78.0	22.0

Data Source: Unpublished data

A component bar diagram shows the bar and its sub-divisions into two or more components. For example, the bar might show the total population of children in the age-group of 6-14 years. The components show the proportion of those who are enrolled and those who are not. A component bar diagram might also contain different component bars for boys, girls and the total of children in the given age group range, as shown in Figure 4.3. To construct a component bar diagram, first of all, a bar is constructed on the x-axis with

its height equivalent to the total value of the bar [for per cent data the bar height is of 100 units (Figure 4.3)]. Otherwise the height is equated to total value of the bar and proportional heights of the components are worked out using unitary method. Smaller components are given priority in parting the bar.

Pie Diagram

A pie diagram is also a component

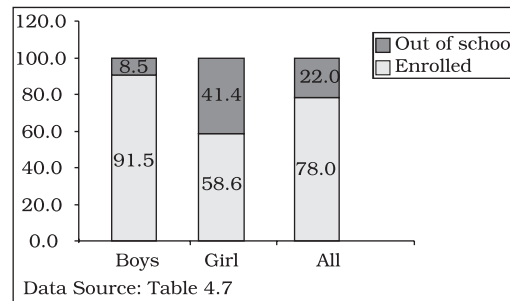
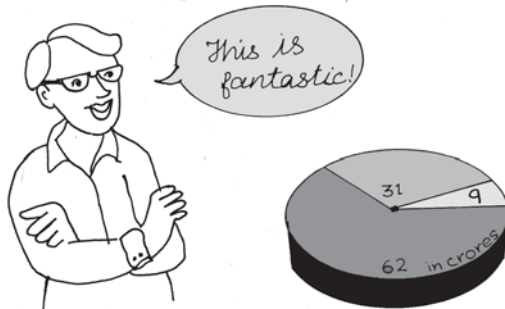


Fig. 4.3: Enrolment at primary level in a district of Bihar (Component Bar Diagram)

diagram, but unlike a component bar diagram, a circle whose area is proportionally divided among the components (Fig.4.4) it represents. It



is also called a pie chart. The circle is divided into as many parts as there are components by drawing straight lines from the centre to the circumference.

Pie charts usually are not drawn with absolute values of a category. The values of each category are first expressed as percentage of the total value of all the categories. A circle in a pie chart, irrespective of its value of radius, is thought of having 100 equal parts of 3.6° ($360^\circ/100$) each. To find out the angle, the component shall subtend at the centre of the circle, each percentage figure of every component is multiplied by 3.6° . An example of this conversion of percentages of components into angular components of the circle is shown in Table 4.8.

It may be interesting to note that data represented by a component bar diagram can also be represented equally well by a pie chart, the only requirement being that absolute values

of the components have to be converted into percentages before they can be used for a pie diagram.

TABLE 4.8
Distribution of Indian population by their working status (crore)

Status	Population	Per cent	Angular Component
Marginal Worker	9	8.8	32°
Main Worker	31	30.4	109°
Non-Worker	62	60.8	219°
All	102	100.0	360°

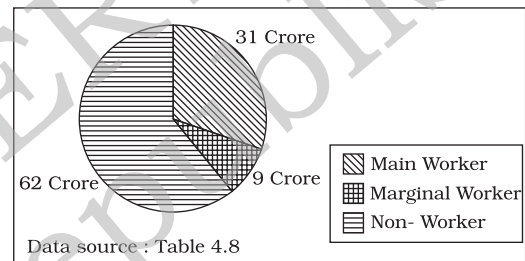


Fig. 4.4: Pie diagram for different categories of Indian population according to working status 2001.

Activities

- Represent data presented through Figure 4.4 by a component bar diagram.
- Does the area of a pie have any bearing on total value of the data to be represented by the pie diagram?

Frequency Diagram

Data in the form of grouped frequency distributions are generally represented by frequency diagrams like histogram, frequency polygon, frequency curve and ogive.

Histogram

A histogram is a two dimensional diagram. It is a set of rectangles with bases as the intervals between class boundaries (along X-axis) and with areas proportional to the class frequency (Fig.4.5). If the class intervals are of equal width, which they generally are, the area of the rectangles are proportional to their respective frequencies. However, in some type of data, it is convenient, at times necessary, to use varying width of class intervals. For example, when tabulating deaths by age at death, it would be very meaningful as well as useful too to have very short age intervals (0, 1, 2, ..., yrs/0, 7, 28, ..., days) at the beginning when death rates are very high compared to deaths at most other higher age segments of the population. For graphical representation of such data, height for area of a rectangle is the quotient of height (here frequency) and base (here width of the class interval). When intervals are equal, that is, when all rectangles have the same base, area can conveniently be represented by the frequency of any interval for purposes of comparison. When bases vary in their width, the heights of rectangles are to be adjusted to yield comparable measurements. The answer in such a situation is frequency density (class frequency divided by width of the class interval) instead of absolute frequency.

TABLE 4.9
Distribution of daily wage earners in a
locality of a town

Daily earning (Rs)	No. of wage earners (f)	Cumulative 'Less than'	Cumulative 'More than'
45-49	2	2	85
50-54	3	5	83
55-59	5	10	80
60-64	3	13	75
65-69	6	19	72
70-74	7	26	66
75-79	12	38	59
80-84	13	51	47
85-89	9	60	34
90-94	7	67	25
95-99	6	73	18
100-104	4	77	12
105-109	2	79	8
110-114	3	82	6
115-119	3	85	3

Source: Unpublished data

Since histograms are rectangles, a line parallel to the base line and of the same magnitude is to be drawn at a vertical distance equal to frequency (or frequency density) of the class interval. A histogram is never drawn for a discrete variable/data. Since in an interval or ratio scale the lower class boundary of a class interval fuses with the upper class boundary of the previous interval, equal or unequal, the rectangles are all adjacent and there is no open space between two consecutive rectangles. If the classes are not continuous they are first converted into continuous classes as discussed in Chapter 3. Sometimes the common portion between two adjacent rectangles (Fig.4.6) is omitted giving a better impression of continuity. The resulting figure gives the impression of a double staircase.

A histogram looks similar to a bar diagram. But there are more differences than similarities between the two than it may appear at the first impression. The spacing and the width or the area of bars are all arbitrary. It is the height and not the width or the area of the bar that really matters. A single vertical line could have served the same purpose as a bar of same width. Moreover, in histogram no space is left in between two rectangles, but in a bar diagram some space must be left between consecutive bars (except in multiple bar or component bar diagram). Although the bars have the same width, the width of a bar is unimportant for the purpose of comparison. The width in a histogram is as important as its height. We can have a bar diagram both for discrete and

continuous variables, but histogram is drawn only for a continuous variable. Histogram also gives value of **mode** of the frequency distribution graphically as shown in Figure 4.5 and the x-coordinate of the dotted vertical line gives the mode.

Frequency Polygon

A frequency polygon is a plane bounded by straight lines, usually four or more lines. Frequency polygon is an alternative to histogram and is also derived from histogram itself. A frequency polygon can be fitted to a histogram for studying the shape of the curve. The simplest method of drawing a frequency polygon is to join the midpoints of the topside of the consecutive rectangles of the histogram. It leaves us with the two

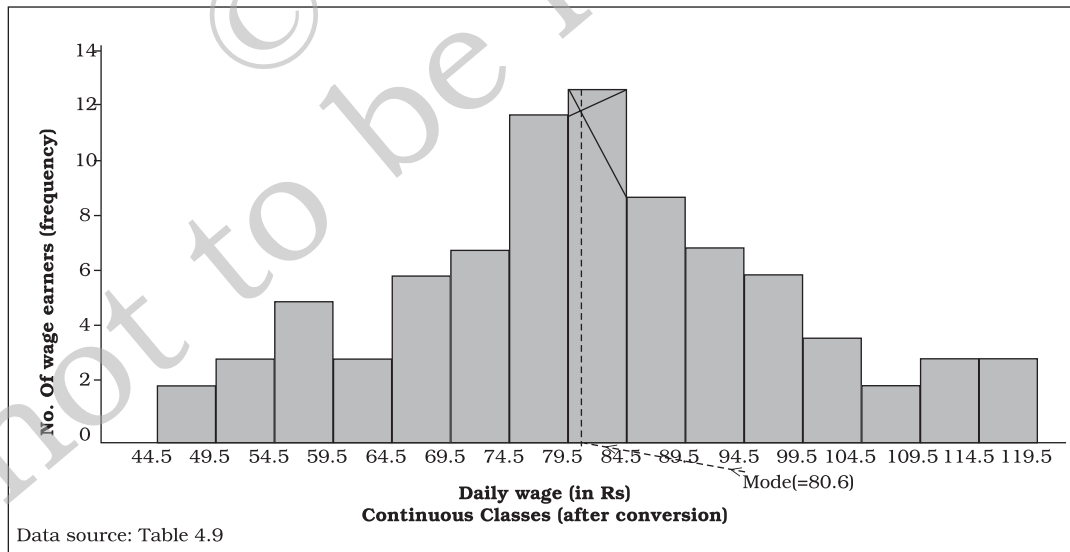


Fig. 4.5: Histogram for the distribution of 85 daily wage earners in a locality of a town.

ends away from the base line, denying the calculation of the area under the curve. The solution is to join the two end-points thus obtained to the base line at the mid-values of the two classes with zero frequency immediately at each end of the distribution. Broken lines or dots may join the two ends with the base line. Now the total area under the curve, like the area in the histogram, represents the total frequency or sample size.

Frequency polygon is the most common method of presenting grouped frequency distribution. Both class boundaries and class-marks can be used along the X-axis, the distances between two consecutive class marks being proportional/equal to the width of the class intervals. Plotting of data becomes easier if the class-marks fall on the heavy lines of the graph paper.

No matter whether class boundaries or midpoints are used in the X-axis, frequencies (as ordinates) are always plotted against the mid-point of class intervals. When all the points have been plotted in the graph, they are carefully joined by a series of short straight lines. Broken lines join midpoints of two intervals, one in the beginning and the other at the end, with the two ends of the plotted curve (Fig.4.6). When comparing two or more distributions plotted on the same axes, frequency polygon is likely to be more useful since the vertical and horizontal lines of two or more distributions may coincide in a histogram.

Frequency Curve

The frequency curve is obtained by drawing a smooth freehand curve passing through the points of the

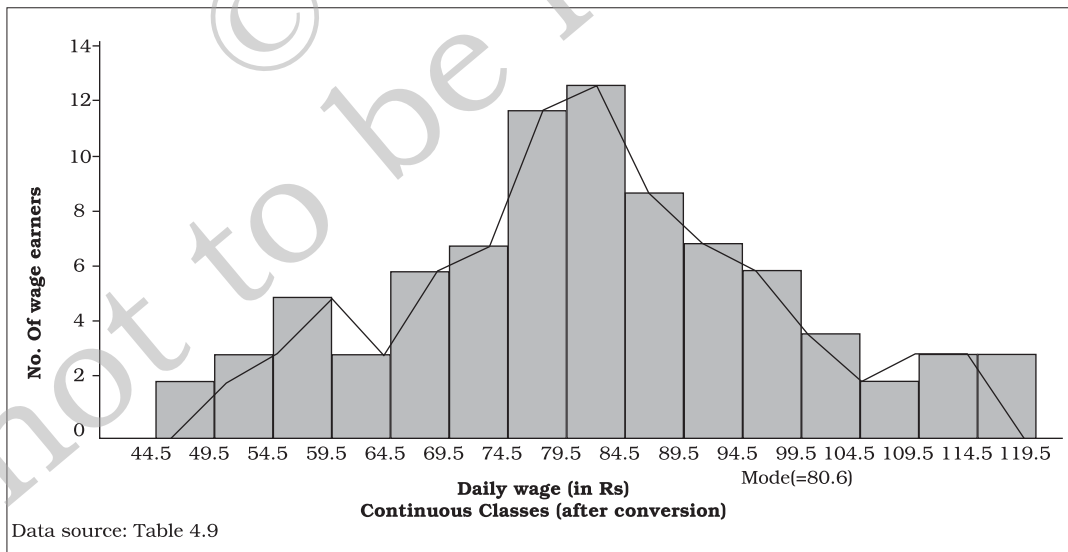


Fig. 4.6: Frequency polygon drawn for the data given in Table 4.9

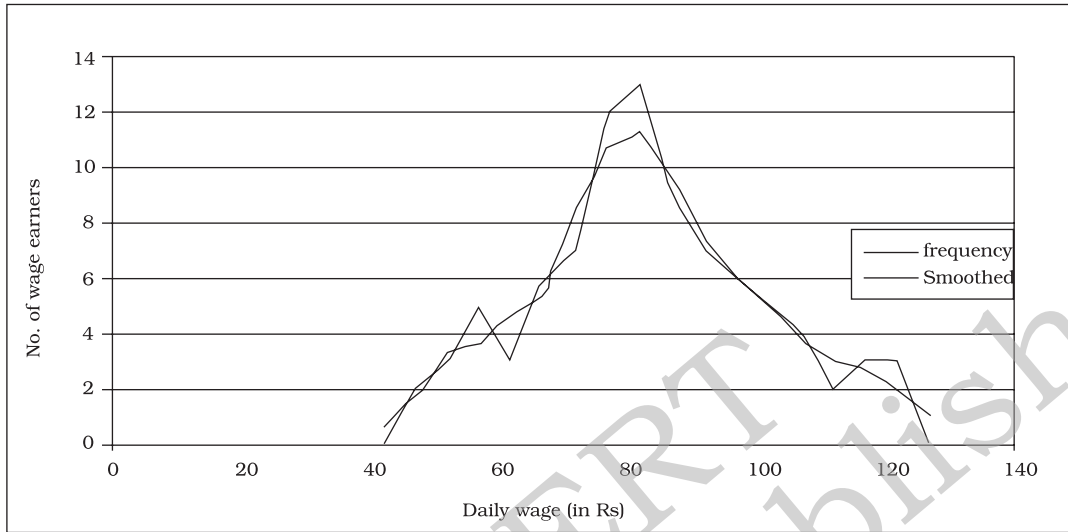


Fig. 4.7: Frequency curve for Table 4.9

frequency polygon as closely as possible. It may not necessarily pass through all the points of the frequency polygon but it passes through them as closely as possible (Fig. 4.7).

Ogive

Ogive is also called cumulative frequency curve. As there are two types of cumulative frequencies, for example less than type and more than type, accordingly there are two ogives for any grouped frequency distribution data. Here in place of simple frequencies as in the case of frequency polygon, cumulative frequencies are plotted along y-axis against class limits of the frequency distribution. For less than ogive the cumulative frequencies are plotted against the respective upper limits of the class intervals whereas for more than ogives the cumulative

frequencies are plotted against the respective lower limits of the class interval. An interesting feature of the two ogives together is that their intersection point gives the **median** Fig. 4.8 (b) of the frequency distribution. As the shapes of the two ogives suggest, less than ogive is never decreasing and more than ogive is never increasing.

TABLE 4.10
Frequency distribution of marks obtained in mathematics

Marks <i>x</i>	Number of students <i>f</i>	'Less than' cumulative frequency	'More than' cumulative frequency
0-20	6	6	64
20-40	5	11	58
40-60	33	44	53
60-80	14	58	20
80-100	6	64	6
<i>Total</i>	<i>64</i>		

