UNIT 2 TABULATION AND GRAPHICAL REPRESENTATION OF DATA*

Structure

2.0 OBJECTIVES

After going through this Unit, you will be able to familiarize yourself with

- stages of statistical inquiry after data have been collected;
- methods of organizing (classification and arrangement) and condensing statistical data;
- concepts of frequency distribution and its various types; and
- different methods of presentation of statistical data such as tables, graphs, diagrams, pictograms, etc.

---

* Adapted from IGNOU study material of EEC 13: Elementary Statistical Methods and Survey Techniques, Units 1, 2 and 3 written by Avatar Singh with modifications by Kaustuva Barik.

2.1 INTRODUCTION

In the preceding Unit, we discussed the methods of collection of data either by a statistical survey (or inquiry) or from some secondary source. Data collected either from census or sample inquiry, that is from primary source, are always hotchpotch and in rudimentary form. To start with, they are contained in hundreds and thousands of questionnaires. To make a head and tail out of them, they must be organised (i.e., classified and arranged), and condensed or summarised. For this purpose we can use various methods like preparing master sheets in which various information are recorded directly from the questionnaires. From these sheets small summary tables can be prepared manually. Now-a-days computers can be used for organisation and condensation of data more swiftly, efficiently and in much less time. Some computer softwares are available which help us to construct various types of graphs and diagrams.

Data can be summarized numerically also. Here we use summary measures like measures of central tendency of first order or degree (like Arithmetic, Geometric and Harmonic Means, Mode and Median); measures of central tendency of second order or degree, also called dispersion measures (like Range, Quartile Deviation, Mean Deviation, and Standard Deviation); measures of association in bivariate analysis (like Correlation and Regression), Index Numbers, etc. In this Unit we plan to discuss how data can be summarized using tables and graphs. Numerical summarization will be discussed subsequently in Units 3 and 4). It must be born in mind that a good summarization and presentation of data is not undertaken for its own sake. It is not an end in itself. In fact it sets the stage for useful analysis and interpretation of data. Again, a good presentation helps us to highlight significant facts and their comparisons. Figures can be made to speak out thereby making possible their intelligent use.

In this Unit we plan to concentrate on organising and condensing data in the form of simple array (ascending and descending order), frequency array and continuous frequency distributions, etc.; and presentation of statistical data in the form of tables and graphs.

2.3 ARRANGEMENT OF DATA

The mass of collected data is often voluminous, unintelligible and boring. It seems totally uninteresting and is not easily interpretable. For example, if you are provided with monthly income figures of 1000 families in a village it is difficult for you to infer anything. But if you are told that the average monthly income of the village is Rs. 2540, it is quite interesting and you are in a position to compare it with other figures.

The first step in the analysis and interpretation of data is its classification and tabulation. The process of arranging data into groups according to their common characteristics is known as its classification.

On the other hand tabulation implies a systematic presentation of data in rows and columns according to some salient features or characteristics.

In Unit 1, a questionnaire was prepared on Family Planning. Suppose this questionnaire was used to collect information from 50 families of C-III block of XYZ Colony, New Delhi. Let us assume that it produced the following types of information as given in Tables 2.1 and 2.2. *Can we make any head or tail out of it*?

Table 2.1
Number of Children per family in C-III block, XYZ Colony, New Delhi

2	0	1	5	3	1	2	1	0	2
4	3	2	2	3	4	1	0	2	3
1	4	2	3	1	2	5	4	1	3
2	1	3	2	3	4	1	2	3	1
4	5	2	1	1	0	3	2	0	2

Table 2.2
Monthly Income of 50 families of C-III block, XYZ Colony, New Delhi

547	622	691	684	567	586	680	578	583	578
708	544	528	540	730	541	720	698	763	633
640	637	598	631	618	692	600	650	604	640
646	654	689	736	731	844	798	712	772	820
678	663	800	692	700	781	658	798	709	720

As pointed out earlier, to make any head or tail out of the mass of raw data, such as presented above, we have to classify and arrange it. This can be done either by forming a simple array or a frequency array (discrete frequency distribution) or a continuous frequency distribution. Sections 2.3.1, 2.3.2 and 2.3.3 attempt to explain this aspect.

2.2.1 Simple Array

It is an arrangement of given raw (univariate) data in ascending or descending order. In the ascending order, the observations are arranged in increasing order of magnitude. For example, numbers 3,5,7,8,9,10 are arranged in ascending order. In descending order, it is the reverse. For example, the numbers 10,9,8,7,6,5,3 are in descending order.

We can prepare both types of simple arrays from Table 2.1. In the following table, the figures have been arranged in ascending order. From the arrangement, it is clear that the lowest value is 0 and the highest one is 5.

Table 2.3
Number of Children per Family in C-III Block of XYZ Colony, New Delhi
Simple Array -- Ascending Order

0	0	0	0	0	1	1	1	1	1
1	1	1	1	1	1	1	2	2	2
2	2	2	2	2	2	2	2	2	2
2	3	3	3	3	3	3	3	3	3
3	4	4	4	4	4	4	5	5	5

After this arrangement the same figures make some sense.

The possible conclusions that can be drawn from this arrangement of data are that five families are issueless, twelve families have one child each, fourteen have two children each, ten families posses three children each, six families four children each and three families have five children each.

2.2.2 Frequency Array or Discrete Frequency Distribution

Here different observations are not repeatedly written as in simple array like 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, etc. We count the number of times (i.e., frequency) an observation repeats itself. For example, in Table 2.3 the observation 4 is repeated 6 times. Thus the frequency of 4 above is 6. The frequency array, for the simple array given in Table 2.3, will look like as given below in Table 2.4.

A frequency array is a statistical table in which various observations are arranged in order of their magnitude along with their respective frequencies.

Table 2.4

Number of Children:	0	1	2	3	4	5	Total
Number of Families:	5	12	14	10	6	3	50

When the number of observations is large enough, the counting process is often undertaken by the use of *tally bars*. In this method, all possible values of the variable are written in a column. For every observation, a tally bar denoted by (|) is noted against its corresponding values. Every fifth repetition is marked by crossing the previous four bars as (𝍸). In this way, we get blocks of five which simplify counting at the end. Thus a number or an observation repeated fourteen times will be marked as (𝍸 𝍸 IIII). Note that after representing each observation by a bar on the *tally sheet*, the same will be ticked (√) or crossed (×) so that it is not duplicated. The data of Table 2.1 is rewritten in the form of frequency distribution as shown in Table 2.5 below.

Table 2.5
Frequency Distribution of Number of Children per Family

No. of Children	Tally Sheet	Frequency
0	~~IIII~~	5
1	~~IIII~~ ~~IIII~~ II	12
2	~~IIII~~ ~~IIII~~ IIII	14
3	~~IIII~~ ~~IIII~~	10
4	~~IIII~~ I	6
5	III	3
Total		**50**

2.3.3 Continuous or Grouped Frequency Distribution

Numbers like 1, 2, 3, 4, 5, 20, 40, etc. are discrete numbers and are used where no value between the two consecutive numbers is possible. As in the case of the number of children, it will be impossible as well as funny to say that a particular family has 2.083 or 2.1 or 2.75 number of children. The family can have either 2 or 3 children and not a fraction of a child. Out of the two examples of raw data mentioned in Section 2.3, the number of children (Table 2.1) is an example of discrete data while monthly income (Table 2.2) is an example of continuous variable giving continuous data.

In this Section we propose to illustrate the construction of continuous or grouped frequency distribution from the raw data of Table 2.2 on monthly income of the 50 families.

To construct a grouped frequency distribution, the range of the given data, i.e., the difference of the highest and the lowest observations, is divided into various mutually exclusive and exhaustive sub-intervals, also known as class-intervals. The frequency of each class interval is then counted and written against it.

Table 2.6
Frequency Distribution of Monthly Income of Families

Monthly Income (Rs.)	Tally Sheet	No. of Families (Frequency)
500 - 550	~~IIII~~	5
550 - 600	~~IIII~~ I	6
600 - 650	~~IIII~~ ~~IIII~~	10
650 - 700	~~IIII~~ ~~IIII~~ II	12
700 - 750	~~IIII~~ IIII	9
750 - 800	~~IIII~~	5
800 - 850	III	3
Total		**50**

We have just completed an exercise where the variable "*income of the family*" has been grouped in order to reduce it to a manageable form called grouped data or *Continuous Frequency Distribution*. However, prior to the construction of any grouped frequency distribution, it is very important to find answers to the following questions:

1. What should be the number of class intervals?
2. What should be the width of each class interval?
3. How will the class limits be designated?

1. What should be the number of class intervals?

Though there is no hard and fast rule regarding the number of classes to be formed, yet their number should be neither too small nor too large. If the number of classes is too small, i.e., width of each class is large, there is likelihood of greater loss of information due to grouping. On the other hand, if the number classes is very large, the distribution may appear to be too fragmented and may not reveal any pattern of behaviour of the variable. Based on experience, it has been observed that the minimum number of classes should not be less than 5 or 6 and in any case, there should not be more than 20 classes.

Mr. Sturges has given a formula to determine the number of classes, that is,

Number of classes = $1 + 3.322 \times \log_{10} N$,

where N is the total number of observations.

In our example of raw data on incomes of 50 families, the number of classes can be calculated as under:

Number of classes $= 1 + 3.322 \times \log_{10} 50 = 1 + 3.322 \times 1.6990$

$= 1 + 5.644 = 6.644 \approx 7.$

2. What should be the width of each class interval?

As far as possible, all the class intervals should be of equal width. However, when a frequency distribution, based on equal class intervals, does not reveal a regular pattern of behaviour of observations, it might become necessary to re-group the observations into class intervals of unequal width. By a regular pattern of behaviour we mean that there are no classes, with possible exclusion of extreme classes, where there are nil or very few observations while there is concentration of observations in their adjoining classes.

$$Width\ of\ a\ Class = \frac{Largest\ Observation - Smallest\ Observation}{Number\ of\ Class\ Intervals}$$

The approximate width of a class can be determined by the following formula:

However, the final decision, regarding width of class intervals, should also take into account the following points.

i) As far as possible, the width should be a multiple of 5, because it is easy to grasp numbers like 5, 10, 15, ... etc.

ii) It should be convenient to find the mid-value of a class.

iii) The observations in a class should be uniformly distributed.

3. How will the class limits be designated?

The smallest and the largest observations of a class interval are known as class limits. These are also termed as the lower and upper limits of a class, respectively. Since the mid-value of a class, which is used to compute mean, standard deviation, etc., is obtained from the class limits, it is very necessary to define these limits in an unambiguous manner. The following points should be kept in mind while defining class limits:

a) It is not necessary that the lower limit of the first class be exactly equal to the smallest observation of the data.
   In fact it can be less than or equal to the smallest observation. Similarly, the upper limit of the last class may be greater than or equal to the largest observation of the data.
b) It is convenient to have the lower limit of a class either equal to zero or some multiple of 5 or 10.
c) The chosen class limits should be such that the observations in a class are uniformly distributed.

The class limits can be defined in either of the following methods:

i) *Exclusive Method*, and ii) *Inclusive Method*.

i) Exclusive Method: In this method, the upper limit of a class is taken to be equal to the lower limit of the following class. In order to keep various class intervals as mutually exclusive, it is decided that the observations with magnitude greater than or equal to lower limit but less than the upper limit of a class are included in it. For example, the class 500 - 550 shall include all observations with magnitude greater than or equal to 500 but less than 550. An observation with magnitude equal to 550 will be included in the next class, i.e., the class 550 - 600.

The major benefit of exclusive class intervals is that it ensures continuity of data because the upper limit of one class is the lower limit of the next class. In our example on monthly income (Table 2.6), there are 5 families whose income lies between Rs. 500 to Rs. 550, i.e., Rs. 500 to 549 and 6 families whose income lies between Rs. 550 to Rs. 600, i.e., Rs. 550 to 599, and so on. Based on this presumption we can rewrite this frequency distribution in the form of Table 2.7 also.

Table 2.7
Exclusive Class Intervals

Monthly Income (Rs.)	*Number of Families* (Frequency)
500 but less than 550	5
550 but less than 600	6
600 but less than 650	10
650 but less than 700	12
700 but less than 750	9
750 but less than 800	5
800 but less than 850	3
Total	**50**

ii) **Inclusive Method:** In this method, all the observations with magnitude greater than or equal to the lower limit but less than or equal to the upper limit of a class is included in it. Now observe Table 2.8. Income of Rs. 549 is included in the class 500 to 549 so that an income of Rs. 550 automatically goes to the next class of 550 to 599. Since the upper limit of one class is not equal to the lower limit of the following class, this saves us from the confusion whether Rs. 550 goes to (500 to 549) or (550 to 599) class.

Table 2.8
Inclusive Class Intervals

Monthly Income (Rs.)	Number of Families (Frequency)
500 - 549	5
550 - 599	6
600 - 649	10
650 - 699	12
700 - 749	9
750 - 799	5
800 - 849	3
Total	**50**

The *choice between exclusive and inclusive methods* depends upon whether we are dealing with continuous variable like income, heights, weights, etc. or a discrete variable like number of children in a family. For a continuous variable it is desirable to construct frequency distribution by the exclusive method because, as we have seen earlier, it ensures continuity. For a discrete variable like number of children in a family or number of students getting first division, the frequency distributions should be constructed by using inclusive type of class intervals.

Mid-Value of a Class

In exclusive type of class intervals, the mid-value or class mark of a class is defined as the arithmetic mean of its lower and upper limits. However, in case of inclusive class intervals, there is a gap between the upper limit of a class and the lower limit of the following class. This gap is eliminated by adding half of the gap to the upper limit and subtracting half of the gap from the lower limit. The new class limits, thus obtained, are known as *class boundaries*. The class boundaries of the inclusive class intervals in Table 2.8 are given in Table 2.9.

Table 2.9

Monthly Income (Rs.)	No. of Families (Frequency)
499.5 - 549.5	5
549.5 - 599.5	6
599.5 - 649.5	10
649.5 - 699.5	12
699.5 - 749.5	9
749.5 - 799.5	5
799.5 - 849.5	3
Total	**50**

2.2.4 Various Forms of Frequency Distributions

Here we propose to introduce the meaning of the following frequency distributions:

a. Open Ended Frequency Distribution
b. A Frequency Distribution with Unequal Class Width
c. Cumulative Frequency Distribution
d. Relative Frequency Distribution

a) **Open End Frequency Distribution**

Open-end frequency distribution is one which has at least one of its ends open. Either the lower limit of the first class or upper limit of the last class or both are not specified. The words "below" or "less than" and "above" or "more than" are used. In the former the value extends to $-\infty$ and in the latter to $+\infty$. Example of such a frequency distribution is given in Table 2.10.

Table 2.10

Open-end Class Frequency

Class	Frequency
Below 25	1
25 - 30	3
30 - 40	5
40 - 50	2
50 and above	1
Total	**12**

Table 2.11

Unequal Class Frequency

Class	Frequency
20 - 25	1
25 - 30	3
30 - 40	5
40 - 55	2
55 - 60	1
Total	**12**

b) **A Frequency Distribution with Unequal Class Width**

The classes of a frequency distribution may or may not be of equal width. A frequency distribution with unequal class width is reproduced in Table 2.11. Here, the width of 1st, 2nd and 5th classes is 5, while that of 3rd is 10 and that of 4th is 15. As we will see in Unit 5, *mode* is not representative value in such types of series and hence not defined.

c) **Cumulative Frequency Distribution**

Suppose that, with reference to data given in Table 2.6, we ask the following questions:

i) How many families have their monthly income less than or equal to Rs. 700?
ii) How many families have their monthly income greater than or equal to Rs. 600?

The answers to the above questions can be easily obtained by forming an appropriate cumulative frequency distribution. To answer the first question, we need to form a "less than type" cumulative frequency distribution while a "greater than type" cumulative frequency distribution is required for answering the second question. These distributions are given in Tables 2.12 and 2.13 respectively.

Table 2.12

"Less-than type" Cumulative Frequency Distribution

Monthly Income (Rs.)	**Frequencies**		
	Simple		Cumulative
Less than 550	5		5
Less than 600	6	5+6	11
Less than 650	10	5+6+10	21
Less than 700	12	5+6+10+12	33
Less than 750	9	5+6+10+12+9	42
Less than 800	5	5+6+10+12+9+5	47
Less than 850	3	5+6+10+12+9+5+3	50

Table 2.13

"More-than type" Cumulative Frequency Distribution

Monthly Income (Rs.)	**Frequencies**		
	Simple		Cumulative
More than 500	5	3+5+9+12+10+6+5	50
More than 550	6	3+5+9+12+10+6	45
More than 600	10	3+5+9+12+10	39
More than 650	12	3+5+9+12	29
More than 700	9	3+5+9	17
More than 750	5	3+5	8
More than 800	3		3

d) **Relative Frequency Distribution**

So far we have expressed the frequency of a value or that of a class as the number of times an observation is repeated. We can also express these frequencies as a *fraction* or a *percentage* of the total number of observations. Such frequencies are known as *the relative frequencies*. Table 2.14 demonstrates the construction of relative frequency distribution.

Table 2.14

Relative Frequency Distribution of Monthly Income of 50 Families

Class	Frequency	Relative Frequency	
		As a fraction	As a percentage
500 - 549	5	5 ÷ 50 = 0.10	0.10 × 100 = 10
550 - 599	6	6 ÷ 50 = 0.12	0.12 × 100 = 12
600 - 649	10	10 ÷ 50 = 0.20	0.20 × 100 = 20
650 - 699	12	12 ÷ 50 = 0.24	0.24 × 100 = 24
700 - 749	9	9 ÷ 50 = 0.18	0.18 × 100 = 18
750 - 799	5	5 ÷ 50 = 0.10	0.10 × 100 = 10
800 - 849	3	3 ÷ 50 = 0.06	0.06 × 100 = 6
Total	**50**	**1.00**	**100**

From Table 2014 it is clear that the sum of relative frequencies should be either 1 (in the case of fraction) or 100 (in the case of percentages).

Check Your Progress 1

1. Distinguish between the following, giving at least two points of distinction.

   a) Discrete and continuous frequency distributions
   b) Simple and cumulative frequency distributions
   c) Exclusive and inclusive class intervals
   d) Simple and frequency array

   ....
   ....
   ....
   ....
   ....
   ....

2. Explain the following terms giving examples:

   a) Ungrouped data
   b) Class mark
   c) Open end classes
   d) Class limits
   e) Class boundaries
   f) Class frequencies
   g) Tally bar
   h) Relative frequencies

   ....
   ....
   ....
   ....
   ....
   ....

3. Build a hypothetical frequency distribution on monthly pocket money of 20 students belonging to the lower middle class of a college. Prepare a relative frequency distribution from it.

   ....
   ....
   ....
   ....
   ....

4. What points are to be kept in mind while taking decisions for preparing a frequency distribution in respect of:

   a) The number of classes, and
   b) Width of the class interval?

   ....

   ....

   ....

   ....

5. Construct less than and more than type cumulative frequency distributions from the following data:

Class:	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
Frequency:	5	8	10	12	8	7

   ....

   ....

   ....

   ....

   ....

6. Construct a relative frequency distribution for the data given in question 5.

   ....

   ....

   ....

   ....

2.3 TABULATION OF DATA

Good presentation of data is as important as their satisfactory collection and arrangement. In fact, satisfactory collection and arrangement of data must be followed by good presentation. However, good presentation is not an end in itself. It may be necessary for satisfactory analysis and interpretation. A satisfactory presentation helps us in more than one ways. *Firstly,* it helps to highlight significant facts contained in the data. *Secondly*, it facilitates the comparison of data. *Finally*, it facilitates the easy understanding and an intelligent use of statistical information.

We will discuss presentation of statistical data under three heads.

i) Formal tables

ii) Graphic methods which will include line graphs, histograms, frequency polygon and curves, and cumulative frequency curves.

iii) Geometric forms, pictures and statistical maps, which will include pie diagrams, bar diagrams, area and volume diagrams, etc.

In this Section we concentrate on tabular forms of presentation.

2.3.1 Meaning and Types of Tables

A table or a statistical table is a systematic arrangement of related statistical data in columns and rows, with a given predetermined and a well decided objective. A row of a table represents a horizontal while a column represents a vertical arrangement of data. To explain the nature of information given in a table, its rows and columns are designated by appropriate stubs and captions (or headings or sub-headings) respectively. Presentation of data in a tabular form should be simple, planned, unambiguous and logical.

Table 2.15 is based on hypothetical figures of exports and imports of country X with countries A, B, C, and D for three years 1995, 1996 and 1997.

Table 2.15

Imports and Exports of X with A, B, C and D during 1995 - 1997

(in Crore of Rupees)

Country	1995		1996		1997@	
	Imports	Exports	Imports	Exports	Imports	Exports
A	60	70	65	75	70	65
B	50	60	60	65	65	60
C	40	30	40	40	42	50
D	45	42	60	55	63	55
Total	**195**	**202**	**225**	**235**	**240**	**230**

Note : @ Figures are quick estimates.
Source : Trade Bulletin, 1998, Ministry of Foreign Trade of X.

In this table it is clear that the purpose is to show the imports and exports of country X vis-à-vis the rest of the world. Note that a particular entry of the table refers to a column and a row. For example, an entry at the intersection of second row and third column indicates that in 1996 country X imported good and services worth Rs. 60 crores from country B. This figure then can be compared with other import and export figures to seek important interpretations.

Types of Tables

Basically, we have two types of tables:

i) Reference tables or general purpose tables

ii) Text tables or special purpose tables.

i) *Reference Tables* are a general purpose tables and are a store of information with the aim of presenting detailed statistical information. From these tables, we can derive our information (i.e., secondary source). Tables presented by different government departments, ministries, Reserve Bank of India, Economic Surveys, etc. are reference tables and are a routine work of these departments. Another important example is the Population Census tables prepared by the Registrar General of India giving detailed information on the demographic features of India.

ii) Students are advised to consult the latest issue of "Economic Survey" which is issued every year along with the union budget of India. Prepare from it a table on exports and imports of India to USA, UK, Russia, Canada and Germany for three or four years.

iii) *Text Tables* are the special type of tables. They are smaller in size and are prepared from the reference tables. Their aim is to analyse only a particular aspect to bring out a specific point or to answer a particular question. For example from the Population Census tables we may pick out information on the number of people in Bombay and Delhi who speak different languages (mother tongue), profess different religions and come from different states of India. Similarly from various publications of Reserve Bank of India, we may be able to extract information, in tabular form, on money supply, rate of interest and bank rate for the last ten years or so.

Tables can be simple and one way, like the tables given in Section 2.2, where we deal with only one variable, say, income. Alternatively, it is called a univariate frequency distribution. In addition to this, we can have two-way or multi-way tables where we deal with two or more related characteristics (for example, Table 2.15).

2.3.2 Parts of a Table

Parts or *elements* of a table vary from table to table depending upon the nature of data and purpose of tabulation. Yet some points are common. These are:

1. **Table number** is required for the identification of a table particularly when there are more than one tables in a particular analysis. Table number is always mentioned in the centre at the top.

2. **Title of the table** gives the indication of the type of information contained in the body of the table. It is said that the *title is to the table what heading is to an essay*. Next to the table number, we mention the title of the table. Its purpose is to answer the questions like:

   a) *What* is in the table?

   b) *Where* is it in the table?

   c) *When* did a particular information occur?

   d) *How* has a particular information been arranged?

In respect of a sample of a table on exports and imports, (Table 2.15), these questions will be answered as below:

a) The table contains values of exports and imports of country X.

b) Information contained in the body of the table shows exports (sales to) and imports (purchases from) four countries A, B, C and D.

c) These exports and imports occurred in 1995, 1996 and 1997.

d) Information on exports and imports has been arranged according to year and countries.

Dos and Don'ts of the Title

Do not opt for long sentences. Title should be brief and to the point. Present the title in bold letters and/or in capital letters. Expressions used should not convey more than one meaning.

Avoid the expressions like 'Table Presents ....' or 'A Detailed Comparison of Data Relating to ....', etc. It should be like a telegraphic message.

3. **Head note**, also called prefactory note, is written just below the title. It shows contents and unit of measurement like (rupees crore) or (lakh tonnes) or (thousand bales). It should be written in brackets and should appear on right side top just below the title. However, every table does not need a head note, like number of students in each class.

4. **Stubs** are used to designate rows. They appear on the left hand column of the table. Stubs consist of two parts:

   a) *Stub head* describes the nature of stub entry.

   b) *Stub entry* is the description of row entries.

5. **Captions**, also called box heads, designate the data presented in the columns of the table. It may contain more than one column heads, and each column head may be sub-divided in more than one sub-head. For example, we can divide the students of a college into hostelers and non-hostlers and then again into males and females. This will help us to know the number of male hostelers in, say, first year, second year and third year.

6. **Main body of the table**, also called *field* of the table, is its most important and bulky part. It contains the relevant numerical information about which a hint is already contained in the title of the table. In our example of Table 2.15 the title amply suggests that the body of the table contains numerical information on exports and imports of country X for a period of three years.

7. **Foot Note** is a qualifying statement put just below the table (at the bottom). Its purpose is to caution about the limitations of the data or certain omissions. For example in Table 2.15, the foot note reads that "figures are quick estimates" implying that these figures are not final. Similarly in the latest population census data the foot note may be "Excluding the State of Jammu and Kashmir".

8. **Source of data** may be the last part of a table, yet it is important one. It speaks about the authenticity of the data quoted. It also offers opportunity to the reader to check the data if he so desires and get more of it.

Taking all these points into consideration, the format of a hypothetical table is presented below:

Table No. 2.16
(---------------TITLE-----------------)
Head note:

(in rupees crore)

Stub Head	←--------------- Caption ---------------→			
	Column Head I		Column Head II	
	Sub-head	Sub-head	Sub-head	Sub-head
Stub Entries	MAIN	BODY	OF THE	TABLE
Totals				

Foot note(s): Source:

2.3.3 Importance of Tables

Numerical information arranged in tabular form has distinct advantage over other forms of presentation. First, tabulated data are easy to understand and interpret. Secondly, one can make quick comparison between different characteristics, for example, 'Are imports greater than exports over all the three years?' or 'Are exports increasing?' Thirdly, it opens doors for further investigations. Fourthly, they have a more lasting impression on human mind than the textual statements. Needless to say, that the statistical tables are used extensively in almost all fields of human inquiry.

Check Your Progress 2

1. Distinguish between
   a) Caption, stub head and stub entries
   b) One-way and two-way tables
   c) Reference tables and text tables
   d) Column entry and row entry
   e) Head note and foot note

2. Comment on the statement: "Title is to the table what heading is to an essay".

3. Enumerate the various parts of a Statistical table.

....

....

....

....

4. Make a sketch of a two-way table to show the following information:
   a) Division of college according to Ist Year, 2nd Year and 3rd Year students
   b) Hosteler and non-hostelers
   c) Male and female students

   Take hypothetical information.

....

....

....

....

2.4 GRAPHICAL PRESENTATION OF DATA

Besides formal tables, statistical data can also be presented in the form of various types of graphs. Graphs are a useful way of conveying information very quickly and briefly. With the same ease and efficiency, they help in comparing data over time and space. They are visual aids and have a powerful impact on the people. It is often said, "*a picture is worth a thousand words*". They attract a reader's attention to what they are supposed to convey about the data. Further, they may help us to estimate some values at a glance, and serve as a pictorial check on the accuracy of our solutions.

However, graphical presentation of data, although useful in different ways mentioned above, is only one method of describing data. This cannot and is not a substitute for other forms of presentation as well as further statistical analysis. In the following, we discuss some of the graphical methods of presentation.

2.4.1 Line Graphs

Although there are four quadrants on a plane, in economics we usually draw our diagrams only in the first quadrant where both the quantities measured on X-axis and Y-axis are positive. Economic quantities like price, quantity demanded and supplied, national income, consumption, production and host of other such variable are non-negative (≥ 0).

Let us take a demand schedule and plot it on the graph. The resultant curve on joining different points, assuming continuity, will give us line graph expressing relation between price and quantity demanded. Such a line graph in Economics is called a *demand curve*. Note that price is measured on Y-axis and quantity demanded on X-axis. The demand curve for data given in Table 2.17 is given in Fig. 2.1.

Table 2.17

Demand Schedule

Price of X (Rs.)	Quantity of X demanded
5	16
10	12
15	8
20	4
25	2
30	1

Table 2.18

Time Series Data

Year	Production of Steel(tons)
1990	10
1991	25
1992	20
1993	40
1994	50
1995	45
1996	60

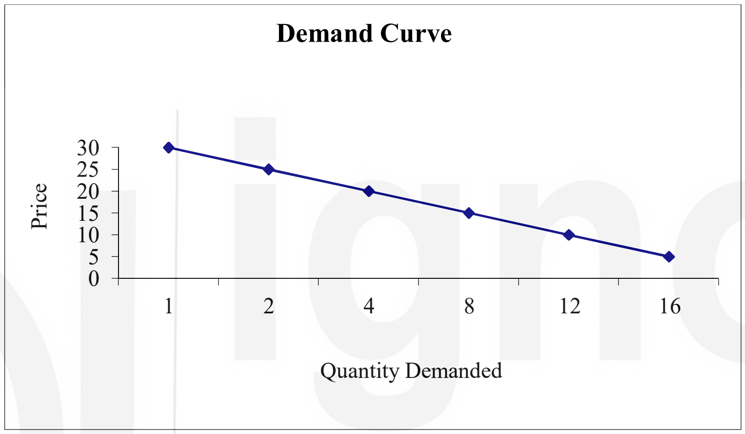


Fig. 2.1

A line graph may be used to show changes in some economic variable, say, steel production over time. In other words, if out of the two variables, one happens to be time (months, years, etc.), we get a line graph over time or simply *time series graph* or *historigram*. A time series expresses behaviour of an economic variable over time. An example of time series data is given in Table 2.18. Measuring years on X-axis and steel production on Y-axis, we can plot time series data on a graph, as shown in Fig.2.2.

Historigram of Production of Steel

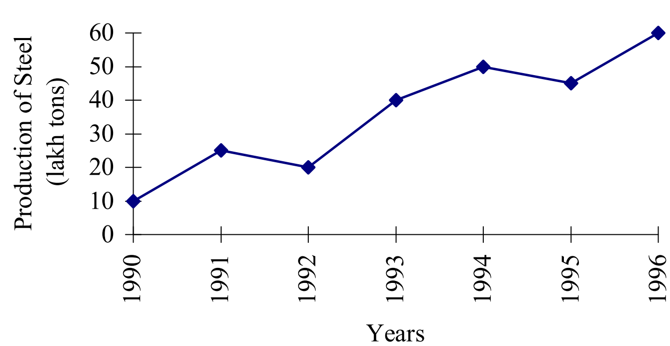


Fig. 2.2

2.4.2 Histogram, Frequency Polygon and Frequency Curve

Histogram (do not confuse with historigram discussed earlier) is a very common type of graph for displaying classified data. It is a set of rectangles erected vertically. It has the following features:

a) It is a rectangular diagram.

b) Since the rectangles are drawn with specified width and height, histogram is a two dimensional diagram. The width of a rectangle equals the class interval and height

$$= \frac{\text{Class frequency} \times \text{Width of the shortest class interval in the data}}{\text{Width of the class interval}}$$

c) The area of each rectangle is proportional to the frequency of the respective class.

Construction of Histogram

To plot a histogram of the frequency distribution given in Table 2.6 on a graph paper, we mark off class intervals like 500 - 550, 550 - 600, etc. on the horizontal axis. Similarly, we mark off frequencies on the vertical axis. Since all the classes are of equal width, the height of each rectangle is taken to be equal to the frequency of the respective class. The histogram is shown in Fig. 2.3.

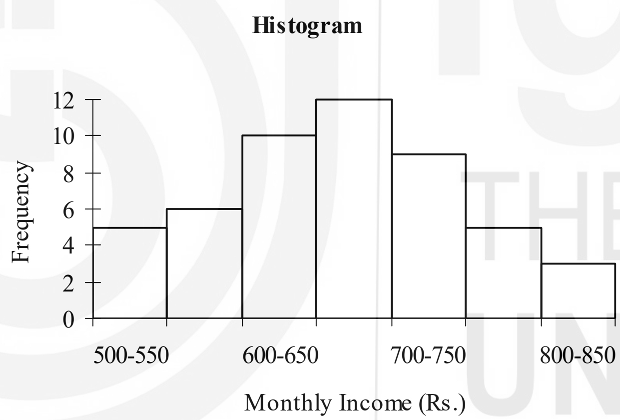


Fig. 2.3

Advantages of histogram are:

1. The width of various rectangles show the nature of classes in the distribution, i.e., whether of equal width or not.
2. Area of a rectangle shows the proportion of the class frequency in the total.

Frequency Polygon

Frequency Polygon has been derived from the word “polygon” which means many sides. In statistics, it means a graph of a frequency distribution. A frequency polygon is obtained from a histogram by joining the mid-points of the top of various rectangles with the help of straight lines, as shown in Fig. 2.4. In order that total area under the polygon remains equal to the area under histogram, two arbitrary classes, each with zero frequency, are added on both ends, as shown below.

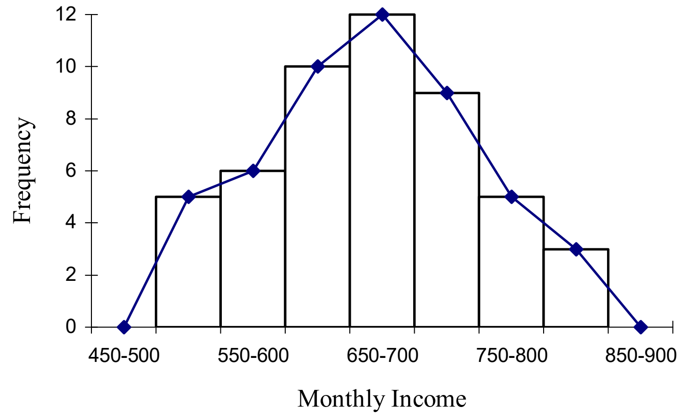


Fig. 2.4: Frequency Polygon

Frequency Curve

If the points, obtained in case of frequency polygon are joined with the help of a smooth curve, we get a frequency curve as shown in Fig. 2.5.

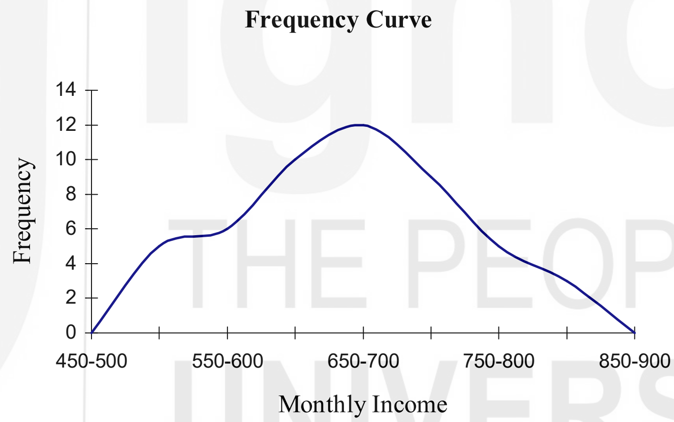


Fig. 2.5

2.4.3 Cumulative Frequency Curve – Ogive

The graph of a cumulative frequency distribution is known as cumulative frequency curve or ogive. Since a cumulative frequency distribution can be of 'less than' or 'greater than' type, accordingly, we can have 'less than' or 'greater than' type of ogives.

Ogives can be used to locate, graphically, certain partition values. We can also determine the percentage of observations lying between given limits. The ogives for the cumulative frequency distributions given in Tables 2.12 and 2.13 are drawn in Fig. 2.6.

Note that to draw a less than type ogive, we add a class interval of 'less than 500' with frequency equal to zero. Similarly, we add a class interval of 'more than 900' with frequency zero for the construction of a greater than type ogive.

'Less than' and 'More than' type Ogives

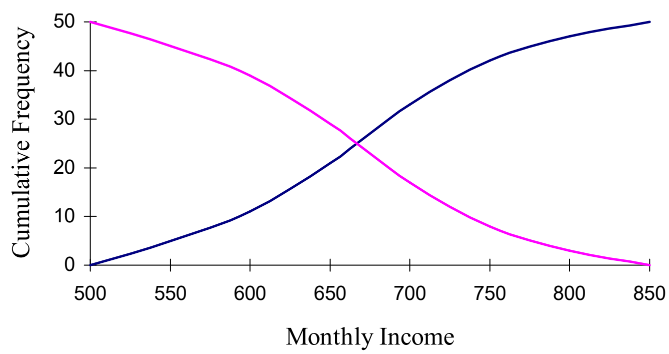


Fig 2.6

2.5 DIAGRAMMATIC PRESENTATION OF DATA

A diagram is a visual form for the presentation of statistical data. Diagram refers to bars, squares, circles, maps, pictorials, cartograms, etc. Diagrams are different from graphs as the former are used only for presentation while the later can be used for analysis in addition to the presentation of data.

2.5.1 One Dimensional Diagrams

These are also known as *bar diagrams*. A *bar* is defined as a *thick line*, often made thicker to attract the attention of a reader. The height of the bar highlights the value of the variable with *width presenting nothing*. Therefore, it has nothing to do with the area of the bar. It is different from the histogram where both the width as well as the height of the bar are important. Further, the bars of the bar diagram are separated from one another so that the gap between the successive bars is same, whereas in histogram they are placed adjacent to one another with out gap. Finally, in histogram the bars are always vertically placed whereas in bar diagram they can be placed both vertically as well as horizontally. Let us take a simple example to demonstrate the construction of a bar diagram.

Table 2.19
Number of students in four zones of a country

Zone	No. of students (lakhs)
North	6
South	10
East	2
West	4

The bar diagram of the above data is drawn in Fig. 2.7. To make the bar diagram beautiful we can either colour the bars or shade them in different ways. This is left to the aesthetic taste of the investigator.

Bar Diagram

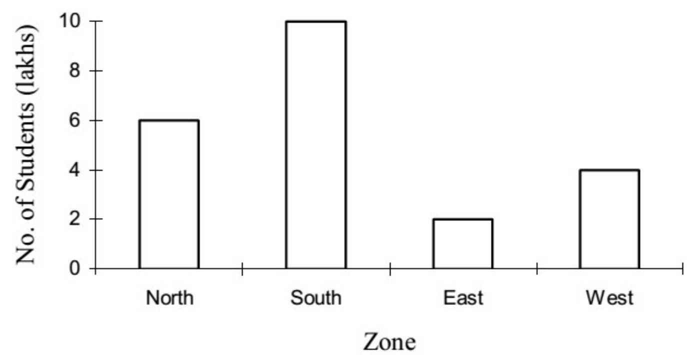


Fig. 2.7

i) Sub-divided or Component Bar Diagram

A sub-divided bar diagram is used when it is desired to represent the comparative values of different components of a phenomenon. In this diagram, the bars, corresponding to each phenomenon, is divided into various components. The portion of the bar occupied by each component denotes its share in the total. The sub-divisions of different bars must always be done in the same order and these should be distinguished from each other by using different colours or shades. A sub-divided bar diagram for the hypothetical data on sales of TV sets, given in Table 2.20 is drawn in Fig. 2.8.

Table 2.20

Zone-wise sale of TV sets (1995-1997)

Zone	Number of T.V. Sets sold (lakhs)		
	1995	1996	1997
North	12	20	28
South	8	9	15
East	5	7	10
West	6	8	11
Total	31	44	64

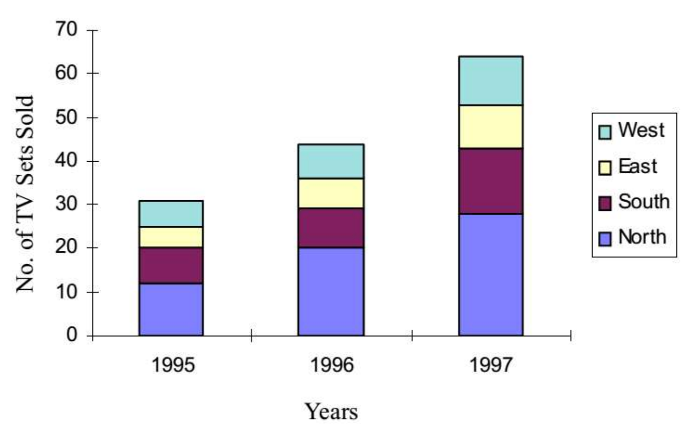


Fig. 2.8: Sub-divided or Component Bar Diagram

ii) Multiple Bar Diagram

This diagram is used when comparisons are to be shown between two or more sets of data. A set of bars for a period, place or a related phenomenon are drawn side by side without gap. Different bars are distinguished by different shades or colours. A multiple bar diagram for the hypothetical data given in Table 2.21 is drawn in Fig. 2.9.

Table 2.21
Total revenue, total cost and profit of M/S XYZ (1990-92)

(Rupees thousand)

Year	Total Revenue	Total cost	Profit
1990	30	25	5
1991	40	35	5
1992	50	40	10

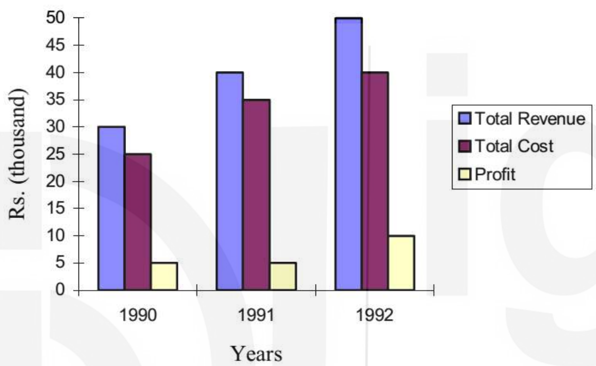


Fig. 2.9

2.5.2 Two Dimensional Diagrams or Area Diagrams

In the case of one dimensional diagrams only the height of the bar is important, and the width can be chosen according to convenience or aesthetic taste of the investigator. But in the case of two dimensional diagrams, area is more important. That is why they are also known as *Area diagrams*. There are three types of area diagrams.

a. *Rectangles*, where area equals width (or base) multiplied by the length (or height) of the rectangle.

b. *Squares* where area equals square of side (or base).

c. *Circles* where area equals πr^2, with π =22/7 and r = radius.

Let us consider data on, say, average salaries of three categories of University teachers, and prepare all the three types of area diagrams.

Table 2.22
Average Salaries of University Teachers as on 1/1/1998

Class of Teachers	Average Salaries (Rs.)
Professors	25,000
Readers	16,000
Lecturers	9,000

a) For drawing rectangles, a common base of, say, 100 is taken. Accordingly, the heights can be determined as:

1. Salary of Rs.25,000 = 100 (base) × 250 (height)
2. Salary of Rs.16,000 = 100 (base) × 160 (height)
3. Salary of Rs. 9,000 = 100 (base) × 90 (height)

Now take a scale of 2 cm = 100, so that the first rectangle has dimensions of 2 cm. × 5 cm, the second one has the dimensions of 2 cm × 3.2 cm and the third one has the dimensions of 2 cm × 1.8 cm. After this, we are in a position to draw the rectangles as area diagrams (Fig. 2.10).

Average Salaries of University Teachers (Rs.)

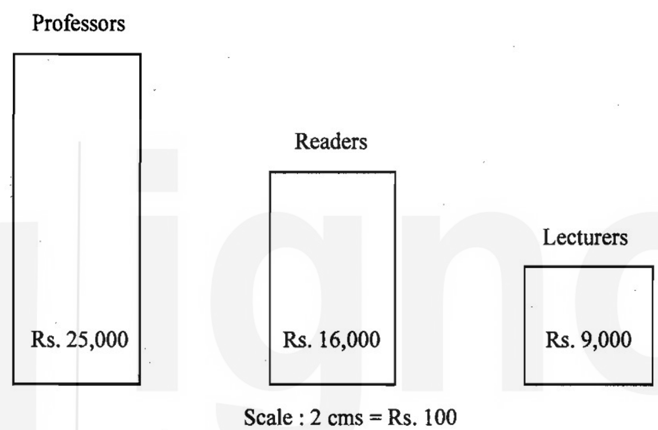


Fig. 2.10

b) For drawing squares, we find the square root of various incomes. We have,

1. $\sqrt{25,000} = 158.114$
2. $\sqrt{16,000} = 126.491$
3. $\sqrt{9000} = 94.868$

Choose a scale 1 cm = 50 so that the first square has each side approximately equal to 3.2 cm. (since 158.114/50 ≅ 3.2), second has the side of 2.53 cm. and the third has the side of 1.9 cm. The relevant squares are drawn in Fig. 2.11.

Average Salary of University Teachers (Rs.)

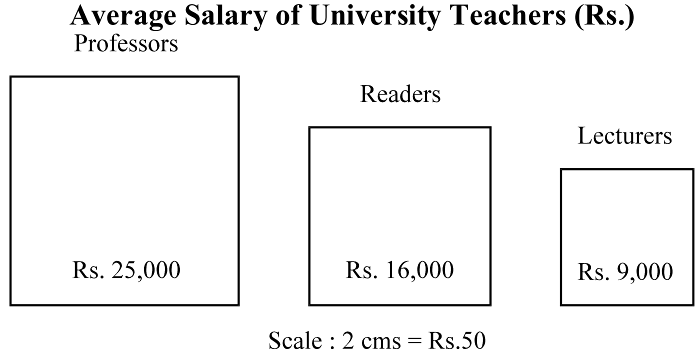
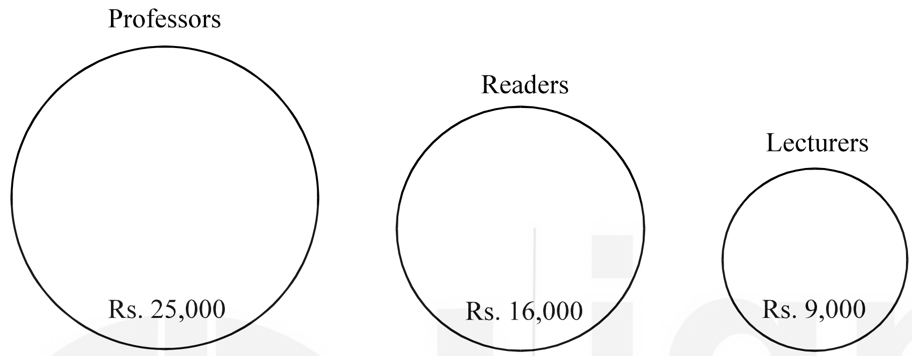


Fig. 2.11

c) For drawing Circles we take the squares of their radii in the ratio of areas, i.e., 25000: 16000: 9000 or 25: 16: 9. This is based on the property of the circles that area of a circle is proportional to the square of its radius. Let r_1, r_2 and r_3 denote the radii of the three circles, then we can write $r_1^2 : r_2^2 : r_3^2 = 25:16:9$ or $r_1 : r_2 : r_3 = 5:4:3$. Taking 1 cm = 2.5 units, the radii of the three circles will be 2.0, 1.6 and 1.2 centimetres respectively. Let us draw the required circles.

Average Salary of University Teachers (Rs.)



Scale : 1 cms = 2.5 unit.

Fig. 2.12

2.5.3 Pie Diagram or Pie Chart

It is also known as angular diagram. It is used to represent percentage break downs of the given data. For example the exports of a country to different countries and continents of the world can be expressed into ratios or percentages. These ratios or percentages can then be converted into angles by the formula

$$\frac{Share\ of\ the\ sub-division}{Total} \times 360°$$

Table 2.23

Exports of X to A, B, C and D in 1990

Country	Exports	Percentage Share	Degree
A	300	(300 × 100) ÷ 800 = 37.50	$(37.5 \times 360^0) \div 100 = 135^0$
B	250	(250 × 100) ÷ 800 = 31.25	$(31.25 \times 360^0) \div 100 = 12.5^0$
C	150	(150 × 100) ÷ 800 = 18.75	$(18.75 \times 360^0) \div 100 = 67.5^0$
D	100	(100 × 100) ÷ 800 = 12.50	$(12.5 \times 360^0) \div 100 = 45^0$
Total	800	100	360^0

Pie Diagram Representing Exports of X

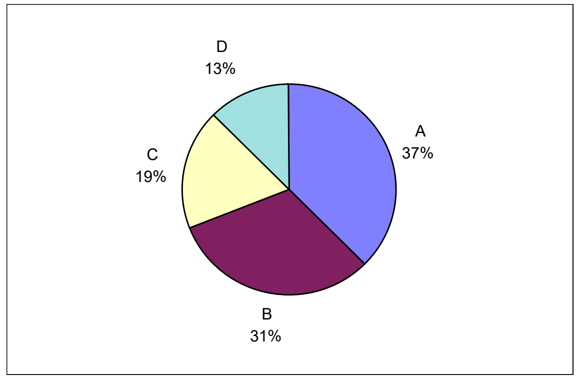


Fig. 2.13

Steps in the construction of Pie diagram

1. Find the total of all components.
2. Find ratio or percentage of the share of sub-division to the total and multiply by 360^0 to get the angle corresponding to the each sub-division.
3. Draw a circle of a suitable size.
4. Use protractor to draw different angles at the centre. Preferably start with the largest one.
5. Shade the different segments with different colours or shades.
6. Write the components with percentage values in the marked, shaded or coloured areas.

2.5.4 Three Dimensional Diagrams

These diagrams are not very popular and are used very rarely. Since these diagrams are three dimensional (involving length, breadth and width), they denote volumes. They can take the form of boxes, cubes, blocks, spheres and cylinders. They are very useful when the variations in magnitudes of the observations are very marked. Here we will explain only the presentation of data by cubes for which we take the following steps:

i) Find cube-root of each figure.
ii) Take a convenient scale, preferably in centimeters.
iii) Draw cubes, dimensions of which are calculated below for an example consisting of two classes of families: Poor and Very Rich.

Table 2.24

	Income class	Income (Rs.)	Cube-root	Side of cube
1.	Poor	216	$\sqrt[3]{216} = 6$	1.5 cms.
2.	Very Rich	3375	$\sqrt[3]{3375} = 15$	3.75 cms.

Scale: 1 cm. = 4 units.

iv) Now draw two cubes with sides equal to 1.5 cms. and 3.75 cms. respectively.

Income levels of Poor and Very Rich people (Rs.)

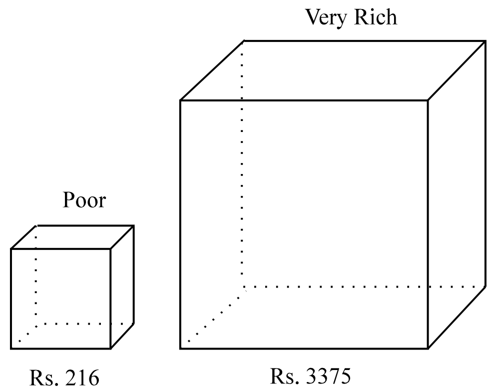


Fig. 2.14

2.5.5 Pictograms and Statistical Maps

These are also known as catrograms. Pictures are more attractive to laymen than other forms of graphic presentations.

But these are not suitable everywhere. It may suit cases involving population of people of a state or number of vehicles in a metropolitan city like Delhi or Bombay. For showing population of human beings, we draw human figures. Here also we have a scale. We may represent 1 lakh people by one human figure so that a population of three and half lakhs is shown by drawing 3 1/2 human figures, as given in Fig. 2.15.

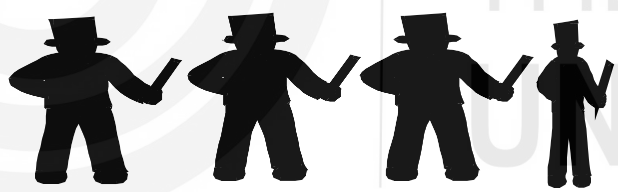


Fig. 2.15

Pictograms suffer from a defect that they present only approximate values. For more accurate presentations bar diagrams are preferable.

Check Your Progress 3

1) Distinguish between the following giving at least two points of distinction.

a) Histogram and historigram. b) Histogram and bar diagram.

c) Histogram and frequency polygon. d)"Less-than"and"More-than" ogives.

e) Pie diagram and circle.

....

....

....

2) Prepare a sub-divided bar chart and a pie diagram from the following data.

Academic Year	Expenditure on Books				
	Economics	Commerce	Maths	Languages	Total
1996 - 97	5200	10000	5000	4800	25000
1997 - 98	8000	14000	7000	6000	35000

....

....

....

....

....

3) Explain the following terms:

a. Line graph

b. Bar diagram

c. Sub-divided or component bar diagram

d. Multiple bar diagram

e. Area diagram

f. Volume diagram

....

....

....

....

....

4) Fill in the blanks with a suitable word out of those given in brackets:

a) A pie diagram is also called ________ diagram. (bar, angular, multiple bar).

b) In the case of vertical bars, the variable is measured on the ________. (X-axis, plane, Y- axis).

c) Bar diagrams, rectangles, squares, circles and pie charts are ________ forms of presenting data. (geometric, arithmetic, horizontal).

d) By joining the mid-points of the top of each rectangle of a histogram, we get ________. (an Ogive, a frequency curve, a frequency polygon)

e) Graph of "more-than" cumulative frequency distribution is also called "more - than" ________. (Ogive, frequency polygon, frequency curve)

f) The caption of a table labels data presented in the ________ of a table. (rows, columns, foot-note)

5) Are the following statements true or false? If false, what should be the correct statement?

a) A picture is worth a thousand words.

b) Squares and circles are examples of area diagrams.

c) We can have only vertical bar to present some data having one variable.

d) The graph of an ordinary frequency distribution is called ogive.

e) A time series graph is known as historigram.

f) Histogram is same as bar diagram.

....

....

....

....

....

2.6 LET US SUM UP

Collected data are unorganised and complex mass of figures. To draw some meaningful conclusions, they must be arranged in an orderly manner. This can be done in many ways, such as by forming simple and frequency array, discrete and continuous frequency distributions, etc.

Sometimes, it serves a useful purpose to form what is called "*less-than*" *or "more-than*" cumulative frequency distributions. Former is formed by successive totaling of frequencies from above and the latter by successive totaling from below.

After collection and condensation of data, good presentation of data is important. A good presentation helps to highlight important points of the data and makes possible useful comparisons and their intelligent use. This can be done through formal tables; line graphs; histograms, frequency polygon and frequency curves; "less-than" and "more-than" ogives; geometric forms – one, two and three dimensional diagrams such as bar diagrams, rectangles, squares, circles, cubes and pie diagrams; statistical maps. While using diagrams, their limitations must always be kept in mind. Diagrams give only a vague idea of the problem and can portray only a limited number of characteristics. Unlike a graphic presentation, the main limitation of a diagrammatic presentation is that it cannot be used as a tool of analysis. The level of accuracy of a graphic method is often lower than that of mathematical method.

2.7 ANSWER OR HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1. (a) See Sub-Section 2.2.2 and 2.2.3.
   (b) See Sub-Section 2.2.4.
   (c) See Sub-Section 2.2.3.
   (d) See Sub-Section 2.2.1 and 2.2.2.

2. You may give examples from your surrounding. For exact meaning of the terms refer to Section 2.2.
3. In the text we have converted the monthly income data in Table 2.2 to a frequency distribution in Table 2.6. From this you can take a clue.
4. Refer to Sub-Section 2.2.3.
5. Refer to Sub-Section 2.2.4(c).
6. Refer to Sub-Section 2.2.4(d).

Check Your Progress 2

1. Refer to Table 2.16 and Sub-Section 2.3.2 for different parts of a table.
2. Refer to Sub-Section 2.3.2(2).
3. Refer to Table 2.16.
4. It can be presented in more than one ways. We have given one below. Try another.

Division of Students of XY College

Year	Hostelers		Non-Hostelers	
	Male	Female	Male	Female
First Year				
Second Year				
Third Year				

Check Your Progress 3

1. (a) See Sub-Sections 2.4.1 and 2.4.2
   (b) See Sub-Sections 2.4.2 and 2.5.1
   (c) See Sub-Section 2.4.2
   (d) See Sub-Section 2.4.3
   (e) See Sub-Section 2.5.2 and 2.5.3

2. Refer to Sub-Sections 2.5.1 and 2.5.3

3. (a) See Sub-Section 2.4.1
   (b) See Sub-Section 2.5.1
   (c) See Sub-Section 2.5.1
   (d) See Sub-Section 2.5.1
   (e) See Sub-Section 2.5.2
   (f) See Sub-Section 2.5.4

4. (a) angular
   (b) y-axis
   (c) geometric
   (d) a frequency polygon
   (e) ogive
   (f) columns

5. True: 1, 2, 5.
   False: 3, 4, 6.