2 PRESENTATION OF DATA

2.1 Classification We have discussed the collection of data in the last chapter. Collected data are usually available in a form which is not easy to comprehend. For example, it we have before us the marks obtained by 1000 college students at their Matriculation Examination, it would be difficult to tell, simply by looking at the marks, as to how many students have marks between 300 and 400, between 400 and 500, and so on. To get a clear picture of the situation, the data must be presented in a form which is easy to understand. As a first step, we may arrange the data into classes or categories having similar characteristics. For example, we may arrange the marks into groups of 50 marks each, e.g. 300 to 349, 350 to 399, 400 to 449 and so on.

Classification *"The process of arranging data into classes or categories according to some common characteristics present in the data is called* ***classification****.* According to L.R. Conor, *classification is the process of arranging things (either actually or notionally) in groups or classes according to their resemblances and affinities."*

Classification may be easily compared with the process of sorting letters in a Post Office. If we happen to go to the Sorting Room of a Post Office, we will find small pigeon-holes for different *cities* or different *localities* of the city. The letters sorted are placed in the pigeon-hole of the city or locality which the letter is addressed to. In the similar way, when we classify data, we arrange the items with the same characteristic in one group or class. For example, in classifying the population of a country by religion, we may arrange all Muslims in one group, Christians in the other and so on.

Although data can be classified by many characteristics, there are four important bases for classification of data. These bases are

(i) *Qualitative* when data are classified by attributes, e.g. sex, religion, marital status, etc.

(ii) *Quantitative* when data are classified by quantitative characteristics, e.g. height, weight, income, etc.

(iii) *Geographical* when data are classified by geographical regions or locations. For example, the population of a country may be classified by Provinces, Divisions, Districts or Towns.

(iv) *Chronological or temporal* when data are arranged by their time of occurrence. An arrangement of data by their time of occurrence is called a *time series*.

2.2 Types of Classification Data may be classified by one, two, three or more characteristics at a time. When data are classified by one characteristic, classification is said to be *one-way*. For example, the population of a country may be classified by religion as Muslims, Christians, Hindus, etc. When we classify data by two characteristics at a time, classification is said to be *two-way*. Classification will be *three-way* when data are classified by three characteristics. Data classified by many characteristics give rise to a *many-way* classification.

In classifying qualitative data, we may divide a characteristic into two sub-classes, one possessing the characteristic and the other not possessing it. This is called *two-fold division* or *dichotomy* (which means cutting into two). For example, if we are studying the literacy of the population, we may divide the population into two categories, literate and illiterate. When we divide a characteristic into three sub-classes, it is called a *three-fold division*, or *trichotomy*. When we divide a characteristic into many sub-classes, it is called a *manifold division*. An example of manifold division is to divide the religion into the categories 'Muslims', 'Christians', 'Budhists' and 'Others'.

2.3 Tabulation A table is a systematic arrangement of data into vertical columns and horizontal rows. The process of arranging data into rows and columns is called *tabulation*. Tabulation may be simple, double, treble, or complex depending upon the type of classification. When tabulation corresponds to one way-classification, it is called *simple tabulation*. Tabulation is *double* when it corresponds to two-way classification. When tabulation corresponds to many-way classification, it is called *complex tabulation*. Tabulation of data on population of a country classified by one characteristic, e.g. religion or marital status, is an example of simple tabulation. Tabulation of data classified by religion and sex or religion and marital status is an example of double tabulation. An example of complex tabulation is the presentation of data on the population of a country classified by age, sex, religion, marital status, etc.

2.4 Construction of Statistical Tables *A statistical table has at least four parts the title, the stub, the box-head and the body. In addition, some tables have a prefatory note, a foot-note and a source note.* Table 2.1 shows the table for the following text.

"Although the area of Punjab is less than that of Baluchistan, the former has a larger population. According to 1972 Census of Population, the total population of Punjab was enumerated to be 37,508 thousand of which 19,942 thousand were males and 17,566 thousand were females. During the same census, the population of Baluchistan (all areas including Gawadar) was enumerated to be 2,405 thousand of which 1,272 thousand were males and 1,133 thousand were females. During the 1961 Census, the population of Punjab was enumerated to be 25,581 thousand of which 13,643 thousand were males and 11,938 thousand were females. During the same census the population of Baluchistan was enumerated to be 1,161 thousand of which 640 thousand were males and 521 thousand were females."

Table 2.1

POPULATION OF PUNJAB AND BALUCHISTAN PROVINCES BY SEX FOR 1961 AND 1972 CENSUSES } Title

Prefatory Note → (Figures in thousands)

	Punjab			Baluchistan[1]		
Census	Male	Female	Total	Male	Female	Total
1961	13,643	11,938	25,581	640	521	1,161
1972	19,942	17,566	37,508	1,272	1,133	2,405

Stub — Box Head — Body

1. All areas including Gawadar → Foot note

Source: Population Census Reports, 1961 and 1972.

Title. Every table must have a title. A title is a heading at the top of the table describing its contents. A title usually tells us the *what, where, how* and *when* of the data in that order. Title is usually in capitals throughout. If the title requires two or more lines, it is arranged to form an inverted pyramid.

Boxhead and Stub The headings for various columns and rows are called *column captions* and *row captions* respectively. The portion of the table containing column captions is called the *boxhead* and that containing row captions is called the *stub*. Following points may be kept in view:

(i) The column and row captions should be brief and clearly defined.

(ii) They should be arranged in the natural order; usually with the most important item in the first column/row. Accordingly, totals may be placed in the first column/row if they are most important, otherwise in the last column/row.

(iii) The more important characteristic may be arranged column-wise rather than row-wise and a characteristic with large number of items may be placed in a row because it is easier to run the eye down a vertical column than across a horizontal row.

(iv) In tabulating long columns of figures, spaces should be left after every five or ten rows.

(v) If the numbers tabulated have more than four or five significant digits, they should be grouped in threes or fours. For example, one should write 45 231 724 instead of 45231724.

(vi) Width of column captions may be made roughly proportional to the size of the numbers inserted. Lengthy captions may be broken into two or more lines using hyphens.

(vii) Width of the stub should be determined by the longest row caption. A very long row caption may be broken into two or more lines to save space. An appropriate heading must be given to stub describing its contents.

Prefatory Notes and Foot Notes Both the prefatory note and the foot-note are used to explain certain characteristics of the data. The prefatory note appears between the title and the body of the table. It is usually used to throw some light about the table as a whole. In Table 2.1, the prefatory note shows that the figures in the table are in thousands.

A foot-note appears immediately below the body of the table. It is used to explain a single fact or a part of the table, e.g. a column or a row. A foot-note does not throw light on the table as a whole.

The foot-notes in the body of table should be indicated by lower case alphabets enclosed in parentheses or by symbols such as *, †, ‡, etc. (but never by a number). Foot-notes for column captions or row captions may be indicated by numbers provided the captions are not of numerical nature, e.g. years, ages, etc.

Source Note Every table must have a source note unless the data are original. Source note is placed immediately below the table but after the foot-note, if any.

Body and Arrangement of Data. The set of entries in the appropriate cells of the table together with totals, etc., forms the *body* of the table. This is the most important part because it contains the entire data arranged in columns and rows.

Arrangement of data is determined by two considerations (i) the basis of classification; and (ii) the purpose of the table. Data may be arranged either (i) alphabetically (ii) geographically, i.e. according to location; (iii) according to magnitude, i.e. ascending or descending order; (iv) historically, i.e. according to time of occurrence; (v) by customary classes, e.g. males, females, children etc. or (vi) by progressive arrangement in which final figures develop from those given earlier.

Spacing and Ruling The purpose of spacing and ruling is to enhance the effectiveness and beauty of the table and to help the reader across a long line so that he does not slip to the line above or below. It can also help the reader to break the body of the table into logical categories. Some of the standard practices in spacing and ruling are as follows:

(i) Double space every fifth line.

(ii) Close the table with rulings on top, bottom and sides.

(iii) Vertical lines may be used to separate columns in the body of the table. No horizontal lines should be drawn in the body of the table except to set off tables or to separate a table into distinct parts.

(iv) Thick or double rulings may be used for separating the title, box-head, stub, etc. while a thin or single line may be used to separate the columns and the parts under captions.

Use of Zeros Zeros should not be used in a table. When no cases have been found to exist or when the value of an item is zero, this fact may be indicated by means of dots (...) or short dashes (--).

Size and Face of Type The type used can be of great help in making a table effective. Through the use of bold face type, attention can be directed to the important facts while in case of less important items lighter type can be used. Size of the type can be varied in a similar manner. In view of the above rules, the title of a table should be in large and bold-face type.

2.5 Frequency Distribution We have discussed the classification and tabulation of data. An important method of summarizing and organizing quantitative data is the formation of a frequency distribution. A frequency distribution is a tabular arrangement of data in which various items are arranged into classes and the number of items falling in each class (called *class frequency)* is stated. Data presented in the form of a frequency distribution are also called *grouped data.* Data which have not been arranged in a systematic order are called *raw data* or *ungrouped data.*

As an example of raw data, let us consider the weights of 120 students at the Punjab University (recorded to the nearest pound) as given below:

154	141	122	130	131	174	165	156	168	182
205	171	146	158	143	151	178	147	164	167
138	139	141	176	168	171	192	124	155	158
198	122	120	110	155	166	175	207	162	218
130	133	151	152	175	166	131	141	150	164
139	154	172	133	196	132	188	173	142	144
165	132	191	190	134	150	158	136	169	152
134	159	185	135	168	186	135	140	140	187

188	140	145	146	155	172	140	144	142	150
159	144	163	162	160	157	153	145	154	145
142	148	142	143	154	143	152	165	131	144
142	146	146	150	178	152	161	173	162	171

It is difficult to draw any meaningful conclusions from such data. For example, it is difficult to tell simply by looking at the above data as to how many students have weights below or above 150 pounds, between 150 and 200 pounds and so on. It is, therefore, necessary to arrange the data in such a way as their main features are clear. An arrangement of raw numerical data in ascending or descending order is called an <u>*array*.</u> The array for the above data is given below:

110	120	122	122	124	130	130	131	131	131
132	132	133	133	134	134	135	135	136	138
139	139	140	140	140	140	141	141	141	142
142	142	142	142	143	143	143	144	144	144
144	145	145	145	146	146	146	146	147	148
150	150	150	150	151	151	152	152	152	152
153	154	154	154	154	155	155	155	156	157
158	158	158	159	159	160	161	162	162	162
163	164	164	165	165	165	166	166	167	168
168	168	169	171	171	171	172	172	173	173
174	175	175	176	178	178	182	183	185	186
187	188	190	191	192	196	198	205	207	218

From the array, it is easier to answer certain questions. For example, we can now easily tell how many students have weights below or above 150 pounds, between 150 and 200 pounds, etc. Still it is difficult to look at 120 observations and obtain an accurate idea as to how these observations are distributed. We may, therefore, arrange them in a better form. For example, the data may be arranged into classes as shown in Table 2.2.

Table 2.2

Weight (lb)	Number of Students
110 – 119	1
120 – 129	4
130 – 139	17
140 – 149	28
150 – 159	25
160 – 169	18
170 – 179	13
180 – 189	6
190 – 199	5
200 – 209	2
210 – 219	1

By arranging the raw data in the above form we have distributed the data into classes and determined the number of items belonging to each class (class frequency). Such an arrangement of data by classes together with their corresponding class frequencies is called a *frequency distribution* or a <u>*frequency table*</u>. By organizing data into a frequency distribution, we lose much of the details of the original data but gain simplicity and get a clear picture of the distribution of data.

2.5.1 Class Limits and Class Boundaries In Table 2.2, we see that each class is described by two numbers. These numbers are called *class limits*; the smaller number is called the *lower class limit* and the larger number the *upper class limit*. For example, in Table 2.2, the class limits of the first class are 110 and 119; 110 is the lower class limit and 119 is the upper class limit.

Class limits are not always exactly what they look like. We know that measurements are seldom exact; mostly they are approximate. A weight of 110 pounds means a weight lying "between 109.5 and 110.5 pounds" and a weight of 119 pounds means a weight lying "between 118.5 and 119.5 pounds." When the lower class limit is given as 110 pounds, the true lower class limit is, therefore, 109.5 pounds and when the upper class limit is given as 119 pounds, the true upper class limit is 119.5 pounds. Thus if the weights have been recorded to the nearest pound, the class 110 – 119 includes all measurements from 109.5 to 119.5 pounds. The values 109.5 and 119.5 which describe the true class limits of a class are called *true class limits* or *class boundaries*. The smaller number 109.5 is the *lower class boundary* and the larger number 119.5 is the *upper class boundary*.

In practice, the class boundaries are obtained by adding the upper class limit of one class to the lower class limit of the next higher class and then dividing by 2. Class boundaries for the class limits in Table 2.2 are given in Table 2.3 below:

Table 2.3

Weight (lb)	Number of Students	Class Boundaries	Class Mark
110 – 119	1	109.5 – 119.5	114.5
120 – 129	4	119.5 – 129.5	124.5
130 – 139	17	129.5 – 139.5	134.5
140 – 149	28	139.5 – 149.5	144.5
150 – 159	25	149.5 – 159.5	154.5
160 – 169	18	159.5 – 169.5	164.5
170 – 179	13	169.5 – 179.5	174.5
180 – 189	6	179.5 – 189.5	184.5
190 – 199	5	189.5 – 199.5	194.5
200 – 209	2	199.5 – 209.5	204.5
210 – 219	1	209.5 – 219.5	214.5

2.5.2 Open-end Classes Sometimes frequency tables are formed in which a class has either no lower class limit or no upper class limit. Such a class is called an *open-end class*. An example of open-end classes is given below.

Age Group	Population
Below 5	6457785
5 – 9	6471715
10 – 14	3808462
15 – 19	3533457
20 – 24	3083245
25 and above	16087775

2.5.3 The Class Mark or Midpoint The class mark or the midpoint is that value which divides a class into two equal parts. It is obtained by adding the lower and upper class limits or class boundaries of a class and dividing the resulting total by 2. Thus the class mark of the class 110 – 119 is (110 + 119)/2 = 114.5.

2.5.4 Size of Class Interval The size of the class interval (also called the *class width* or *class length*) is the difference between the upper class boundary and the lower class boundary (not the difference between the class limits). If all the class intervals of a frequency distribution are of equal size, the common width is denoted by h. In such a case, the size of the class interval is also equal to the difference between the two successive lower or upper class limits. It can also be determined from the class marks. The class interval for the data in Table 2.3 is 119.5 – 109.5 = 120 – 110 = 124.5 – 114.5 = 10.

2.6 Formation of a Frequency Distribution Following steps are involved in the formation of a frequency distribution:

(i) *Determine the greatest and the smallest numbers* in the raw data and find the range, i.e. the difference between the greatest and the smallest numbers. In the example of weights of 120 students, the greatest number is 218 and the smallest number is 110. Thus the range is 218 – 110 = 108.

(ii) *Decide on the number of classes*[1]. There are no hard and fast rules for this purpose. For most types of data, it is al-right to have 5 to 20 classes. If we have less than 5 classes, it will result in too much information being lost. On the other hand, if we have more than 20 classes, computations become unnecessarily lengthy. In the example of weights of 120 students, the number of classes is 11.

(iii) *Determine the approximate class interval size* by dividing the range by the desirable number of classes. In case of fractional results, the next higher whole number may be used as class interval size. In the example of weights of 120 students, the class interval size is 108/11 = 9.8, or say, 10. A number used as class interval size should be easy to work with. Class interval sizes that are in tens or multiples of tens or in units or exact decimal values of units are easier to use. Class interval sizes of 1, 2, 3, 4, 5, 10, 20, 50, etc. are in most common use.

(iv) *Decide what should be the lower class limit (or the lower class boundary) of the lowest class.* The lower class limit should cover the smallest value in the raw data. The lower class limits which end with numbers 0 or 5, e.g. 0, 5, 10, 15, 20, etc. are commonly used.

(v) *Find the upper class boundary by adding the class interval size to the lower class boundary and then determine the upper class limit.* The remaining lower and upper class limits may be determined by adding the class interval size repeatedly until the largest measurement is enclosed in the final class.

---

1 *H.A. Sturges has given the following rule to determine the approximate number of classes. $m = 1 + 3.3 \log N$ where m stands for number of classes and N is the total number of values in the data. This rule, however, gives quite a large number of classes for very small data and a very few classes for very large data. For example if N = 10, m = 1 + 3.3 log 10 = 1 + 3.3(1) = 4.3 or 5 classes. But for N = 1000, m = 1 + 3.3 log 1000 = 1 + 3.3(3) = 1 + 9.9 = 10.9 or 11 classes.*

(vi) *Distribute the values in the raw data into classes* and determine the number cases falling in each class i.e. *the class frequencies*. There are two methods arranging observations into their proper classes.

(a) *By Listing the Actual Values.* Each observation is listed in its pro class. Table 2.4(a) illustrates the tabulation of weight measurements 120 students. This is called an *entry table*. We can also obtain the array the data by arranging the numbers of each class into an array as in Ta 2.4(b).

Table 2.4(a)

Weight (lb)	Entry
110 – 119	110
120 – 129	122, 124, 122, 120
130 – 139	130, 131, 138, 139, 130, 133, 131, 139, 133, 132, 132, 134, 136, 134, 135, 135, 131
140 – 149	141, 146, 143, 147, 141, 141, 142, 144, 140, 140, 140, 145, 146, 140, 144, 142, 144, 145, 145, 142, 148, 142, 143, 143, 144, 142, 146, 146
150 – 159	154, 156, 158, 151, 155, 158, 155, 151, 152, 150, 154, 150, 158, 152, 159, 155, 150, 159, 157, 153, 154, 154, 152, 150, 152
160 – 169	165, 168, 164, 167, 168, 166, 162, 166, 164, 165, 169, 168, 163, 162, 160, 165, 161, 162
170 – 179	174, 171, 178, 176, 171, 175, 175, 172, 173, 172, 178, 173, 171
180 – 189	182, 183, 185, 186, 187, 188
190 – 199	192, 198, 196, 191, 190
200 – 209	205, 207
210 – 219	218

Table 2.4(b)

Weight (lb)	Entry
110 – 119	110
120 – 129	120, 122, 122, 124
130 – 139	130, 130, 131, 131, 131, 132, 132, 133, 133, 134, 134, 135, 135, 136, 138, 139, 139
140 – 149	140, 140, 140, 140, 141, 141, 141, 142, 142, 142, 142, 142, 143, 143, 143, 144, 144, 144, 144, 145, 145, 145, 146, 146, 146, 146, 147, 148
150 – 159	150, 150, 150, 150, 151, 151, 152, 152, 152, 152, 153, 154, 154, 154, 154, 155, 155, 155, 156, 157, 158, 158, 158, 159, 159
160 – 169	160, 161, 162, 162, 162, 163, 164, 164, 165, 165, 165, 166, 166, 167, 168, 168, 168, 169
170 – 179	171, 171, 171, 172, 172, 173, 173, 174, 175, 175, 176, 178, 178
180 – 189	182, 183, 185, 186, 187, 188
190 – 199	190, 191, 192, 196, 198
200 – 209	205, 207
210 – 219	218

Table 2.5

Weight (lb)	Tally Marks	Frequency
110 – 119	/	1
120 – 129	////	4
130 – 139	~~////~~ ~~////~~ ~~////~~ //	17
140 – 149	~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ///	28
150 – 159	~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~	25
160 – 169	~~////~~ ~~////~~ ~~////~~ ///	18
170 – 179	~~////~~ ~~////~~ ///	13
180 – 189	~~////~~ /	6
190 – 199	~~////~~	5
200 – 209	//	2
210 – 219	/	1

(b) *By Using Tally Marks.* If the data are not arranged in order of magnitude, the easiest way of tabulating data is by recording a stroke (called *tally mark)* opposite the appropriate class for each observation. The first four strokes are recorded vertically (/) and the fifth one is recorded diagonally (~~////~~) so as to distinguish a set of five. The class frequencies are then written in the frequency column. Table 2.5 illustrates the tabulation of data on the weights of 120 students.

Example 2.2(d) The waist measurements in inches of 40 children aged 5 to 10 years are given below:

19.3	16.9	17.9	17.3	15.8	18.5	17.1	19.5	20.4	18.7
22.3	17.5	18.4	13.9	18.8	16.8	14.9	19.5	19.4	16.3
17.8	23.4	17.4	19.0	21.8	18.8	18.5	18.2	16.1	18.3
17.5	17.4	18.6	16.9	16.5	18.2	20.5	20.5	17.5	19.1

Group these data into a frequency distribution taking 1.0 as the class interval size. e.g. 13.5 – 14.4, 14.5 – 15.4, etc. Also determine the class boundaries.

Solution The frequency distribution of waist measurements is given in Table 2.6.

Table 2.6

Waist Measurement (in)	Tally Mark	Frequency	Class Boundary
13.5 – 14.4	/	1	13.45 – 14.45
14.5 – 15.4	/	1	14.45 – 15.45
15.5 – 16.4	///	3	15.45 – 16.45
16.5 – 17.4	~~////~~ ///	8	16.45 – 17.45
17.5 – 18.4	~~////~~ ////	9	17.45 – 18.45
18.5 – 19.4	~~////~~ ~~////~~	10	18.45 – 19.45
19.5 – 20.4	///	3	19.45 – 20.45
20.5 – 21.4	//	2	20.45 – 21.45
21.5 – 22.4	//	2	21.45 – 22.45
22.5 – 23.4	/	1	22.45 – 23.45

2.6.1 Frequency Distribution of Discrete Data As already mentioned, the cla limits in discrete data are the true class limits and there are no class boundari because discrete data are not in fractions. If class interval size is one we usually ta single values.

Example 2.3 The following figures give the number of children born to 50 women a certain locality upto the age of 40 years.

1. 5. 1, 0, 2, 5, 9, 2, 6, 3, 5, 7, 8, 4, 6, 8, 9, 10, 9, 3,
5. 7, 9, 9, 4, 5, 4, 5, 5, 7, 3, 4, 2, 3, 4, 6, 3, 4, 2, 5,
6, 4, 0, 5. 6, 8, 5, 4, 7, 6

Make a frequency distribution taking class interval size of 1.

(B.I.S.E., Gujranwala 20(

Solution The data are discrete. We, therefore, group the data by sizes, i.e. t number of children as shown below:

Number of Children	Tally Mark	Number of Women
0	//	2
1	//	2
2	////	4
3	~~////~~	5
4	~~////~~ ///	8
5	~~////~~ ~~////~~	10
6	~~////~~ /	6
7	////	4
8	///	3
9	~~////~~	5
10	/	1
Total		50

2.7 Cumulative Frequency Distribution The total frequency of all classes less than the upper class boundary of a given class is called the *cumulative frequency* of that class. For example, the cumulative frequency of the class 120 – 129 in Table 2.5 is 1 + 4 = 5. This means that 5 students have weights less than 129.5 pounds. Similarly, the cumulative frequency of the class 130 – 139 in Table 2.5 is 1 + 4 + 17 = 22, which means that 22 students have weights less than 139.5 pounds.

A table showing the cumulative frequencies is called a *cumulative frequency distribution, cumulative frequency table* or simply a *cumulative distribution.* Table 2.7 gives a cumulative frequency distribution of the weight distribution of 120 students in Table 2.5. Cumulative frequency distribution of this type is called a "*less than" cumulative frequency distribution.*

Sometimes we consider a cumulative frequency distribution of all values greater than or equal to the lower class boundary of each class interval. Such a distribution is called an "*or more" cumulative frequency distribution or decumulative frequency distribution.* Table 2.8 gives an "*or more*" cumulative frequency distribution for the weight distribution of 120 students in.

Table 2.7 **Table 2.8**

Weight (lb)	Cumulative Frequency	Weight (lb)	Cumulative Frequency
Less than 109.5	0	109.5 or more	119 + 1 = 120
Less than 119.5	1	119.5 or more	115 + 4 = 119
Less than 129.5	1 + 4 = 5	129.5 or more	98 + 17 = 115
Less than 139.5	5 + 17 = 22	139.5 or more	70 + 28 = 98
Less than 149.5	22 + 28 = 50	149.5 or more	45 + 25 = 70
Less than 159.5	50 + 25 = 75	159.5 or more	27 + 18 = 45
Less than 169.5	75 + 18 = 93	169.5 or more	14 + 13 = 27
Less than 179.5	93 + 13 = 106	179.5 or more	8 + 6 = 14
Less than 189.5	106 + 6 = 112	189.5 or more	3 + 5 = 8
Less than 199.5	112 + 5 = 117	199.5 or more	1 + 2 = 3
Less than 209.5	117 + 2 = 119	209.5 or more	1
Less than 219.5	119 + 1 = 120	219.5 or more	0

2.8 Relative Frequency Distribution The frequency of a class divided by the total frequency is called the *relative frequency* of that class. It is generally expressed as a percentage. For example, the relative frequency of the class 160–169 in Table 2.2 is (18/120) × 100 = 15%. Clearly the sum of the relative frequencies of all the classes is 1 or 100%.

Table 2.9

Weight (lb)	Relative Frequency
110 – 119	1/120 = 0.0083 or 0.83%
120 – 129	4/120 = 0.0333 or 3.33%
130 – 139	17/120 = 0.1417 or 14.17%
140 – 149	28/120 = 0.2333 or 23.33%
150 – 159	25/120 = 0.2084 or 20.84%
160 – 169	18/120 = 0.15 or 15%
170 – 179	13/120 = 0.1083 or 10.83%
180 – 189	6/120 = 0.05 or 5%
190 – 199	5/120 = 0.0417 or 4.17%
200 209	2/120 = 0.0167 or 1.67%
210 219	1/120 = 0.0083 or 0.83%

2.8.1 Relative Cumulative Frequency Distribution The cumulative frequency of a class divided by the total frequency is called the *relative cumulative frequency*. It is generally expressed as percentage. It is also called *percentage cumulative frequency*. For example, the relative cumulative frequency of weight less than 159.5 from Table 2.7 is (75/120) × 100 = 62.5% which means that 62.5% of the students have weight less than 159.5 pounds.

A table showing relative cumulative frequencies is called the *relative cumulative frequency distribution* or *percentage cumulative frequency distribution.*

Table 2.10

Weight (lb)	Relative Cumulative Frequency	Weight (lb)	Relative Cumulative Frequency
Less than 109.5	0%	Less than 169.5	$\frac{93}{120}$ = 0.7750 or 77.5%
Less than 119.5	$\frac{1}{120}$ = 0.0083 or 0.83%	Less than 179.5	$\frac{106}{120}$ = 0.8833 or 88.33%
Less than 129.5	$\frac{5}{120}$ = 0.0417or 4.17%	Less than 189.5	$\frac{112}{120}$ = 0.9333 or 93.33%
Less than 139.5	$\frac{22}{120}$ = 0.1833 or 18.33%	Less than 199.5	$\frac{117}{120}$ = 0.9750 or 97.5%
Less than 149.5	$\frac{50}{120}$ = 0.4167 or 41.67%	Less than 209.5	$\frac{119}{120}$ = 0.9917 or 99.17%
Less than 159.5	$\frac{75}{120}$ = 0.6250 or 62.5%	Less than 219.5	$\frac{120}{120}$ = 1 or 100%

Example 2.2(a) The following table gives the daily wages in Rupees of 65 employees of a company. With reference to this table, determine (i) the lower limit of the sixth class (ii) the upper limit of the fourth class (iii) the class mark of the third class (iv) the class boundaries of the fifth class (v) The size of the fifth-class interval (vi) The frequency of the third class (vii) the relative frequency of the third class (viii) The class interval having the largest frequency (ix) The percentage of employees

earning less than Rs.280 per day (x) The percentage of employees earning less than Rs.300 per day but at least Rs.260 per day

Wages (Rs.)	250-259	260-269	270-279	280-289	290-299	300-309	310-319
Number of Employees	8	10	16	14	10	5	2

Solution (i) Rs.300 (ii) Rs.290

(iii) The class mark of the third class = (270 + 279)/2 = Rs.274.50

(iv) Lower class boundary of fifth class = Rs.289.50.

Upper class boundary of fifth class = Rs.299.50.

(v) Size of fifth-class interval = 290 – 280 = 299 – 289 = Rs 10

= 299.50 – 289.50 = Rs.10.

In this case all class intervals have the same size (Rs.10)

(vi) 16 (vii) 16/56 = 0.246 = 24.6% (viii) 270 – 279

(ix) Total number of employees earning less than Rs 280 per day = 16 + 10 + 8 = 34. Percentage of employees earning less than Rs.280 per day = 34/65 = 52.3%.

(x) Number of employees earning less than Rs.300 but at least Rs.260 per day = 10 + 14 + 16 + 10 = 50 and the corresponding percentage = (50/65)100 = 76.9%.

Example 2.2(b) The following table gives the hourly wages (in Rupees) of 100 unskilled workers of a factory. With reference to the table. determine (i) class boundaries (ii) class marks (iii) relative frequencies (iv) less than cumulative frequencies (v) or more cumulative frequencies (vi) relative cumulative frequencies. (B.I.S.E., Lahore 2005)

Hourly Wage (Rs.)	30-49	50-69	70-89	90-109	110-129	130-149
No. of workers	4	20	23	35	10	8

Solution

(i) 29.5 – 49.5, 49.5 – 69.5, 69.5 – 89.5, 89.5 – 109.5, 109.5 – 129.5, 129.5 – 149.5. (ii) 39.5, 59.5. 79.5, 99.5, 119.5, 139.5 (iii) 0.04, 0.20, 0.23, 0.35, 0.10, 0.08, or 4%, 20%, 23%, 35%, 10%, 8%.

(iv) 4, 24, 47, 82, 92, 100. (v) 8, 18, 53, 76, 96, 100. (vi) 0.04, 0.24, 0.47, 0.82, 0.92, 1.00 or 4%, 24%, 47%, 82%, 92%, 100%.

Example 2.2(c) If the class marks in a frequency distribution of the weights of students are 128, 137, 146, 155, 164, 173 and 182 pounds, find (i) the class-interval size (ii) the class boundaries assuming that the weights were measured to the nearest pound.

Solution (i) Class-interval size = common difference between successive class marks = 137 – 128 = 147 – 137 = ... = 182 – 173 = 9 pounds (ii) Since the class intervals all have equal size, the class boundaries are midway between the class marks and these have the values (128 + 137)/2, (137 + 146)/2,..., (173 + 182)/2 or 132.5, 141.5.

150.5,...,177.5 pounds. The first class boundary is 132.5 – 9 = 123.5 and the last class boundary is 177.5 + 9 = 186.5, since the common class-interval size is 9 pounds. Thus all the class boundaries are given by 123.5, 132.5, 141.5, 150.5, 159.5, 168.5, 177.5, 186.5 pounds.

2.9 Bivariate Frequency Distribution So far we have considered frequency distributions which involved only one variable. Such frequency distributions are called *univariate frequency distributions* because they involve only one variable. We can also construct a distribution taking two variables at a time. Suppose we have the heights in inches and weights in pounds of 50 students at a certain college as given below:

Height (in)	60	62	61	70	64	60	65	65	73	71
Weight (lb)	100	105	104	115	110	102	110	108	119	118
Height (in)	61	60	63	64	67	68	69	64	66	62
Weight (lb)	109	108	107	112	115	117	117	111	113	104
Height (in)	63	67	71	70	68	68	71	64	63	68
Weight (lb)	108	108	116	110	114	116	119	107	108	105
Height (in)	73	69	64	67	67	64	62	67	62	64
Weight (lb)	119	107	115	111	114	108	105	117	105	107
Height (in)	65	66	67	68	61	64	65	67	66	69
Weight (lb)	108	116	118	115	104	108	109	113	113	115

Suppose we want to form a frequency distribution from these data taking a class interval of size 3 inches for heights and a class interval of size 5 pounds for weights. We arrange the class limits for heights in columns and those of weights in rows as shown in Table 2.11. Since each pair of values relates to two variables, a tally mark will be marked in a cell lying at the intersection of appropriate classes of the two variables. For example, the tally mark for the height 60 inches and weight 100 pounds will be marked at the intersection of the classes 60–62 for heights and 100–104 for weights. Similarly, the tally mark for the pair 65 inches and 110 pounds will lie in the cell determined by the classes 63–65 for heights and 110–114 for weights.

Such a frequency distribution is called a *bivariate frequency distribution* because it involves two variables. It is also called a *bivariate frequency table* or simply *bivariate distribution* or *bivariate table*. The rules for construction of a bivariate frequency distribution are the same as for a univeriate frequency distribution.

Table 2.11

Weight (lb)	Height (in)				
	60 – 62	63 – 65	66 – 68	69 – 71	72 – 74
100 – 104	𝍸				
105 – 109	𝍸	𝍸 𝍸	//	/	
110 – 114		////	𝍸 /	/	
115 – 119		/	𝍸 //	𝍸 /	//

Table 2.12

Weight (lb)	Height (in)					
	60 – 62	63 – 65	66 – 68	69 – 71	72 – 74	Total
100 – 104	5	–	–	–	–	5
105 – 109	5	10	2	1	–	18
110 – 114	–	4	6	1	–	11
115 – 119	–	1	7	6	2	16
Total	10	15	15	8	2	50

Table 2.12 is the bivariate frequency distribution of heights and weights of 50 students at a certain college.

2.10 Graphs Data can be effectively presented by means of *graphs* (also known as *line charts*). A graph consists of curves or straight lines. Graphs bring to light the salient features of the data at a glance and render comparison of two or more statistical series easy. They provide a very good method of showing fluctuations and trends in statistical data. For example, by looking at a graph showing the yield of a crop for a number of years, one can easily visualize the rise and fall in the yield. Graphs can also be used to make predictions and forecasts. Certain partition values can also be located graphically. The only disadvantage of graphs is that they do not convey very accurate information. Further, too much details spoil the graph.

2.10.1 Important Rules for Drawing Graphs

(i) Select a suitable scale so that the graph gives the true impression of the data to be represented.

(ii) Every graph must have a clear and comprehensive title.

(iii) The source of the data must be given.

(iv) The independent variable should always be taken along the horizontal axis (*X-axis*) and the dependant variable along the vertical axis (*Y-axis*).

(v) The vertical scale (*Y-axis*) should always start at zero. If the first item of the data is quite large, a scale-break should be shown between zero and the next number.

(vi) Axes should be clearly labelled. Labels should clearly state the variable and the units, e.g. production and tons, imports and exports and their value in rupees etc.

(vii) More than one curve must be differentiated by different colours or lines.

(viii) The graph should give a smart look. It should not be over crowded with too many curves.

2.11 Graphs of Frequency Distributions The important graphs of frequency distributions are histogram. frequency polygon, frequency curve and cumulative frequency polygon or ogive.

2.11.1 Histogram A *histogram* consists of a set of adjacent rectangles having bases along the *X*-axis, with centres at the class marks (i.e. marked off by class boundaries) and areas proportional to the class frequencies. If the class interval sizes are equal.

the heights of the rectangles are also proportional to the class frequencies and are taken numerically equal to the class frequencies. If the class interval sizes are not equal, then the heights of the rectangles have to be adjusted.

To draw a histogram (for equal class interval sizes), class boundaries (not class limits) are marked along the *X-axis* and rectangles are constructed with width proportional to class interval size and heights proportional to class frequencies. The resulting graph will be a histogram. Figure 2.1 gives the histogram for the frequency distribution of weights of 120 students in Table 2.3.

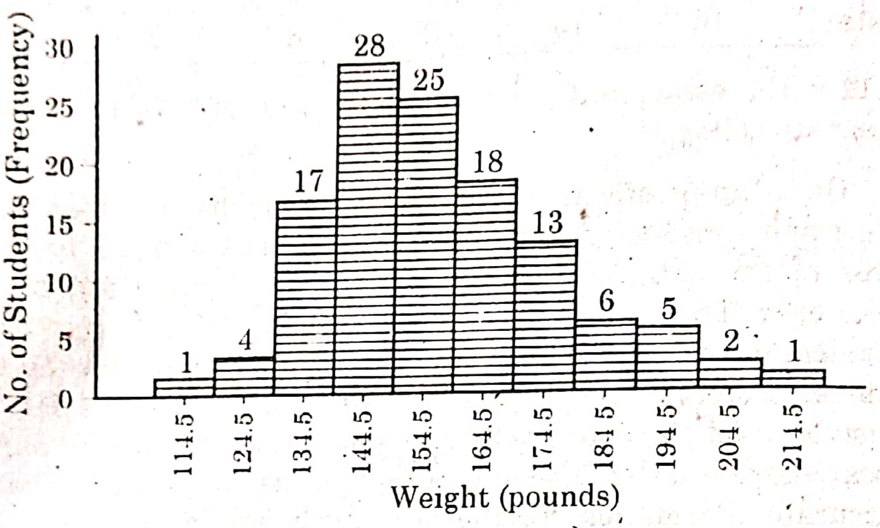


Fig. 2.1 Histogram for the Frequency Distribution of Weights of 120 Students

To adjust the heights of rectangles in a frequency distribution with unequal class interval sizes, each class frequency is divided by its class interval size as shown in Example 2.4 below.

Example 2.4 Draw a histogram for the following frequency distribution.

Class	10 – 11	12 – 14	15 – 19	20 – 29	30 – 34	35 – 39	40 – 42
Frequency	4	12	25	60	25	15	6

Solution The frequency distribution has unequal class interval sizes. Frequencies are adjusted by dividing them by the respective class interval size as shown in the following table. Figure 2.2 gives the required histogram.

Class	Class mark	Class Boundary	Frequency	Class Interval Size	Adjusted Frequency
10 – 11	10.5	9.5 – 11.5	4	2	4/2 = 2
12 – 14	13.0	11.5 – 14.5	12	3	12/3 = 4
15 – 19	17.0	14.5 – 19.5	25	5	25/5 = 5
20 – 29	24.5	19.5 – 29.5	60	10	60/10 = 6
30 – 34	32.0	29.5 – 34.5	25	5	25/5 = 5
35 – 39	37.0	34.5 – 39.5	15	5	15/5 = 3
40 – 42	41.0	39.5 – 42.5	6	3	6/3 = 2

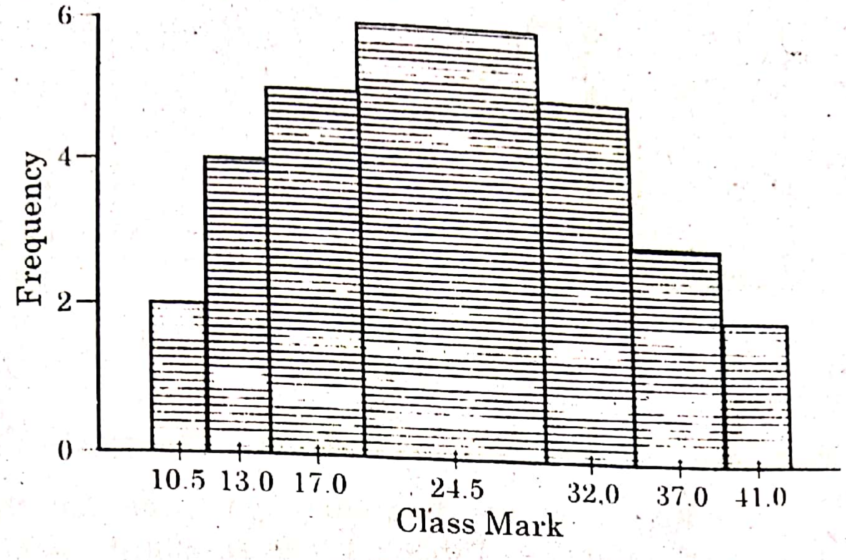


Fig. 2.2 Histogram for Unequal Class Interval Sizes

2.11.2 Frequency Polygon A frequency polygon is a many–sided closed figure. It is constructed by plotting the class frequencies against their corresponding class marks (mid-points) and then joining the resulting points by means of straight lines. A frequency polygon can also be obtained by joining the mid-points of the tops of rectangles in the histogram.

To construct a frequency polygon, we mark the class marks along the *X-axis* and class frequencies along the *Y-axis*. Points are obtained by plotting the class frequencies against their corresponding class marks. The points so formed are joined by means of straight lines. The ends of the graph so drawn do not meet the *X-axis*. We know that a polygon is a many–sided closed figure. We, therefore, add extra classes at both ends of the frequency distribution with zero frequencies. By doing so, the polygon forms a closed figure.

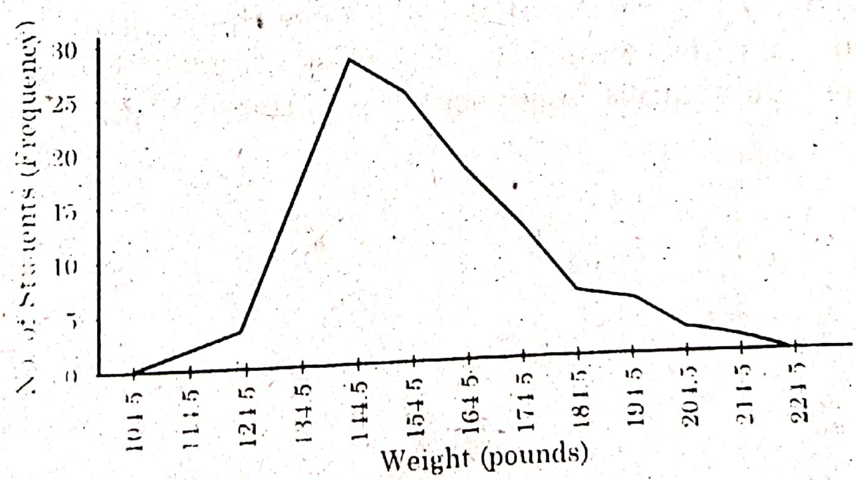


Fig. 2.3(a) - Frequency Polygon for the Frequency Distribution of Weights of 120 Students

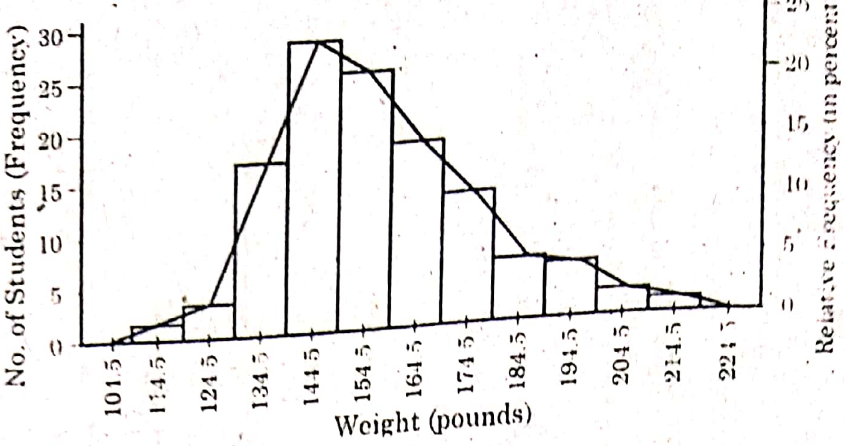


Fig. 2.3(b) Frequency Polygon for the Frequency Distribution of Weights of 120 Students

Figures 2.3(a) and 2.3(b) show the frequency polygons for the frequency distribution of weights of 120 students in Table 2.3. Figure 2.3(a) shows a frequency polygon drawn by plotting the class marks and their corresponding class frequencies while Figure 2.3(b) shows the frequency polygon obtained by joining the midpoints of the tops of the rectangles of the histogram.

2.11.3 Relative Frequency Histogram and Relative Frequency Polygon Graphic representation of relative frequency distribution can be obtained from the histogram or frequency polygon simply by changing the *Y-axis* from frequency to relative frequency on a graph. The resulting graphs are called *relative frequency histogram* or *percentage histogram* and *relative frequency polygon* or *percentage frequency polygon* respectively.

2.11.4 Cumulative Frequency Polygon or Ogive A graph showing the cumulative frequencies plotted against the upper class boundaries is called a *cumulative frequency polygon* or *an ogive*. The graphs corresponding to a "less than" and an "or more" cumulative frequency distributions are called *"less than"* and *"or more" ogives* respectively. Figure 2.4(a) shows a "less than" ogive for the "less than" cumulative frequency distribution in Table 2.7 while Figure 2.4(b) gives an "or more" ogive for the "or more" cumulative frequency distribution in Table 2.8.

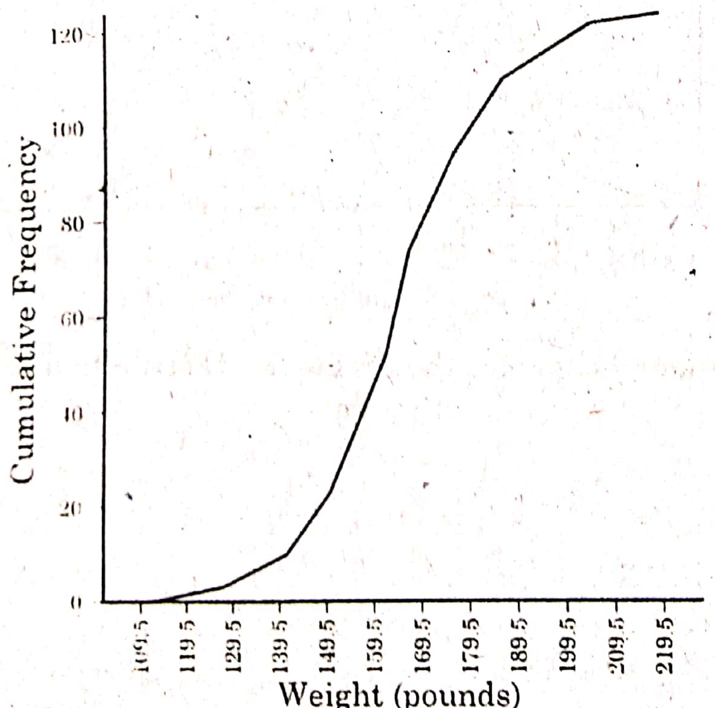


Fig. 2.4(a) Ogive for the Cumulative Frequency Distribution of Weights of 120 Students

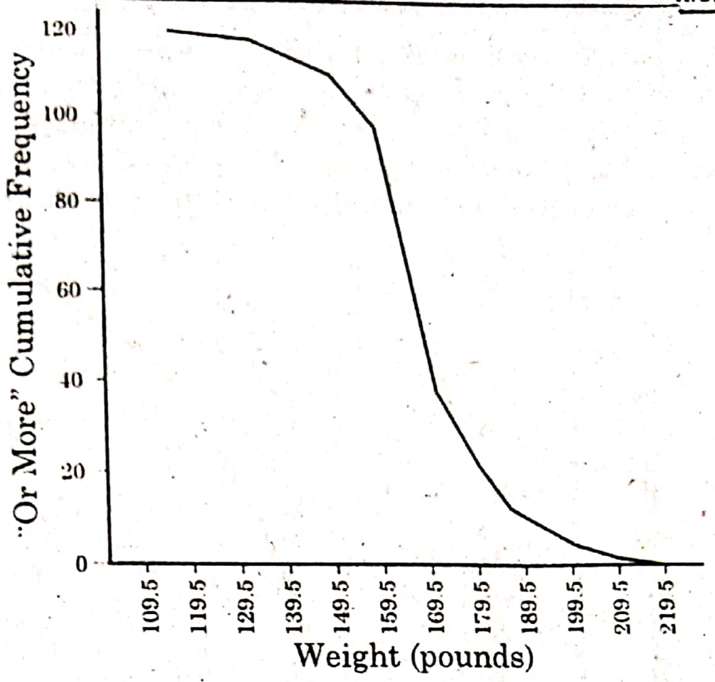
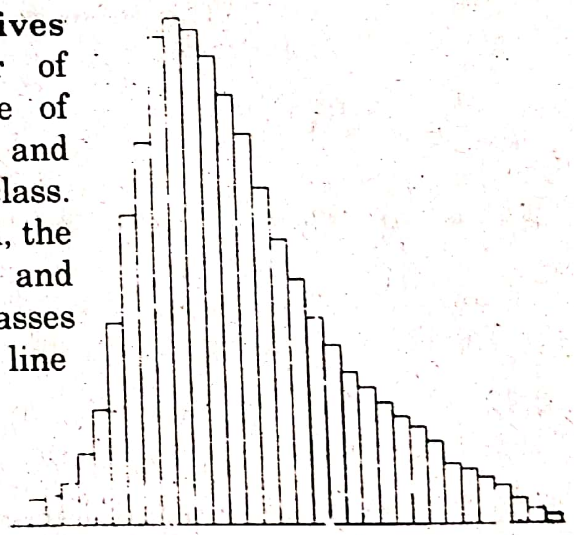


Fig. 2.4(b) Ogive for the "Or More" Cumulative Frequency Distribution of Weights of 120 Students

If we use relative cumulative frequencies in place of cumulative frequencies, the resulting graph is called a *relative cumulative frequency polygon* or *percentage ogive*.

2.11.5 Frequency Curves and Smoothed Ogives



For a population containing a large number of observations, it is theoretically possible in case of continuous data to choose class intervals very small and still have some observations falling within each class. Thus if we could make our class intervals very small, the rectangles in Fig. 2.1 would become narrower and narrower. Likewise the number of cases in the classes would become smaller and smaller. In such case the line

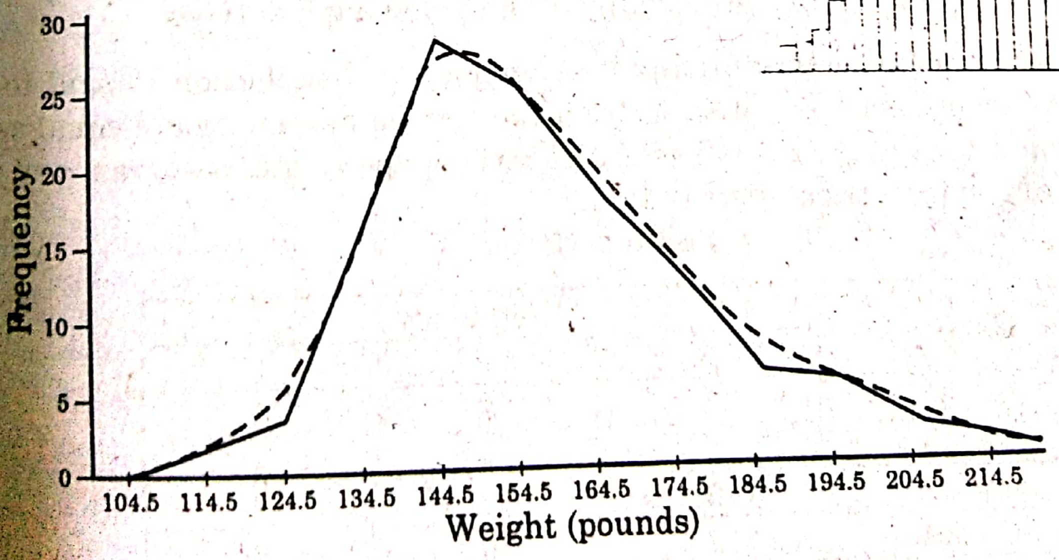


Fig. 2.5(a) Frequency Curve for the Frequency Distribution of Weights of 120 Students

connecting the tops of rectangles in Fig. 2.1 would come closer and closer to a smoothed curve as in the given figure. The smoothed curve which we should get if we study enough cases is called a *frequency curve*. Similarly, smoothed cummulative frequency polygons or smoothed ogives are obtained by smoothing the cumulative frequency polygons or ogives. It is usually easier to smooth an ogive than a frequency polygon.

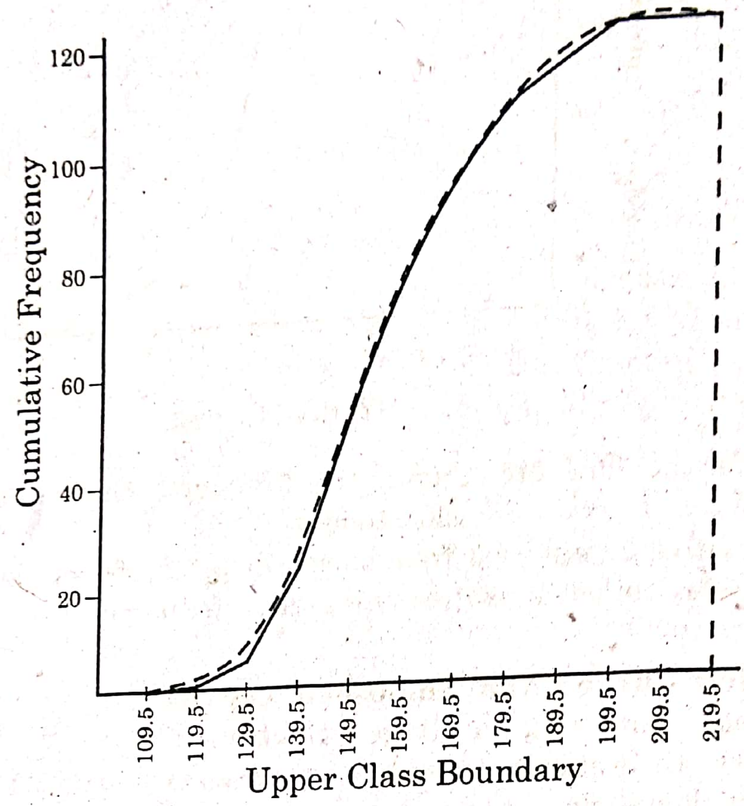


Fig. 2.5(b) Smoothed Ogive for the Cumulative Frequency Distribution of Weights of 120 Students

Figure 2.5(a) shows a frequency curve for the distribution of weights of 120 students in Table 2.3. while Fig. 2.5(b) shows a smoothed ogive for the same data (Table 2.7).

2.12 Common Types of Frequency Distributions and Curves The frequency distributions occurring in practice are usually of the following five types.

2.12.1 The Symmetrical Distributions A frequency distribution is said to be *symmetrical* if the frequencies equidistant from the central maximum are equal as in column A of Table 2.13. Figure 2.6 shows the general shape of the histogram and the frequency curve of a symmetrical distribution.

Table 2.13

Class	Frequency		
	A	B	C
0 – 9	2	2	1
10 – 19	5	5	4
20 – 29	9	12	7
30 – 39	12	9	9
40 – 49	9	7	12
50 – 59	5	4	5
60 - 69	2	1	2

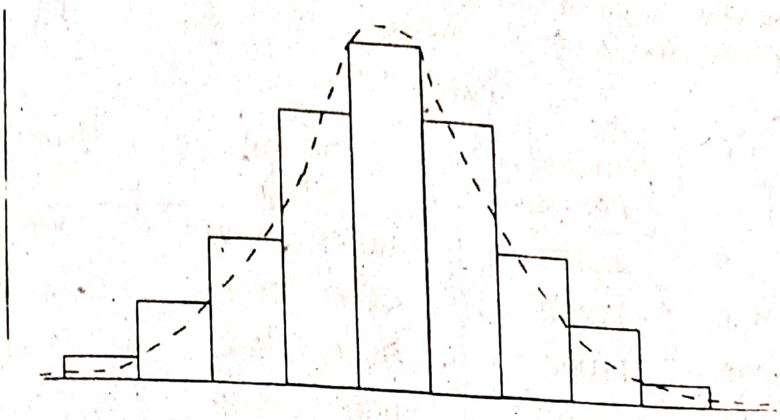


Fig. 2.6 Symmetrical Distribution

2.12.2 The Moderately Skewed or Asymmetrical Distributions A frequency distribution is said to be *skewed* when it departs from symmetry, i.e. when the frequencies tend to pile up in one end or the other end of a distribution as shown in columns B and C of Table 2.13. The distribution with frequencies in column B is said to be *positively skewed* while the distribution with frequencies in column C is said to be *negatively skewed*. Figure 2.7(a) shows the histogram and the frequency curve of a positively skewed distribution while Figure 2.7(b) shows the histogram and the frequency curve of a negatively skewed distribution.

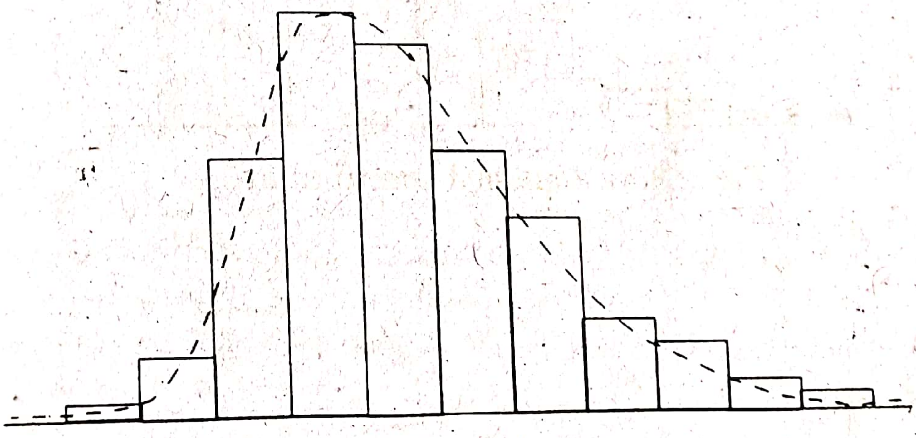


Fig. 2.7(a) Positively Skewed Distribution

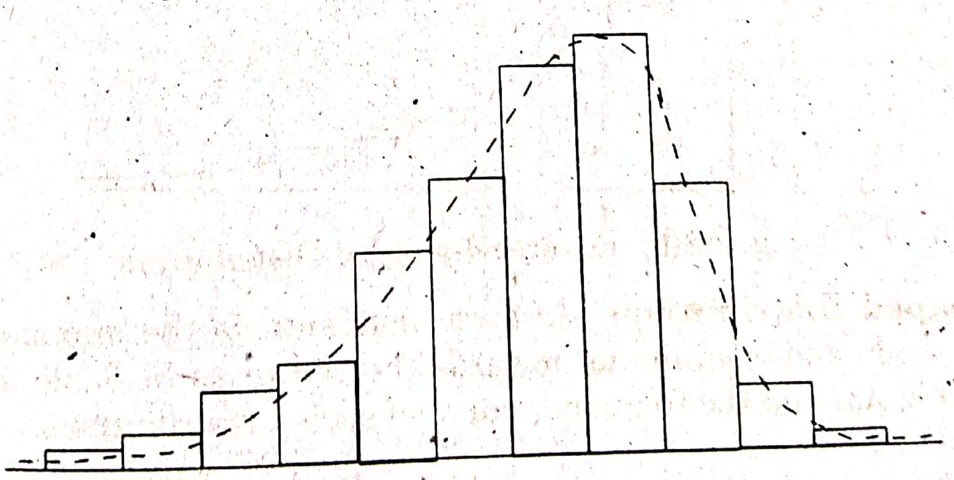


Fig. 2.7(b) Negatively Skewed Distribution

2.12.3 Extremely Skewed or J-Shaped Distributions This type of distribution has the maximum frequency at one end as in Table 2.14. Figure 2.8(a) shows the shape of the histogram and the frequency curve of a *J-shaped* distribution while

Figure 2.8(b) shows the shape of a *reverse J-shaped* distribution. Such distributions are common in economic and medical statistics.

Table 2.14

Income (Rs.)	Number of Persons	Income (Rs.)	Number of Persons
Below 2000	21560	6000 – 6999	2875
2000 – 2999	17250	7000 – 7999	2230
3000 – 3999	11090	8000 – 8999	1090
4000 – 4999	7956	9000 – 9999	975
5000 – 5999	3940	10000 and above	760

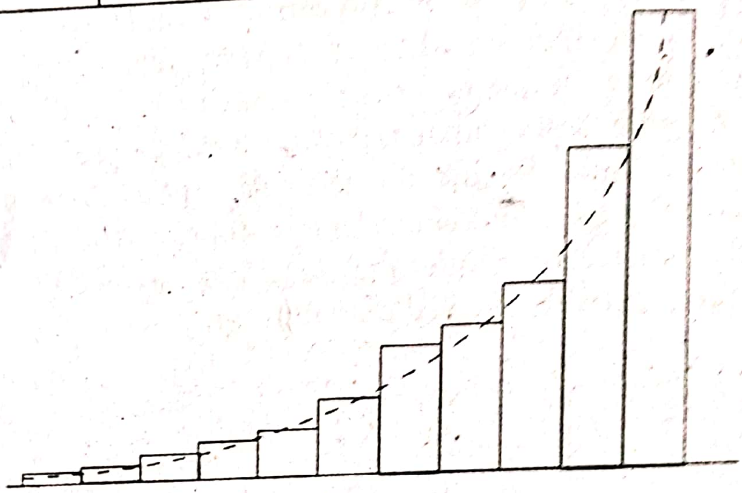


Fig. 2.8(a) J-shaped Distribution

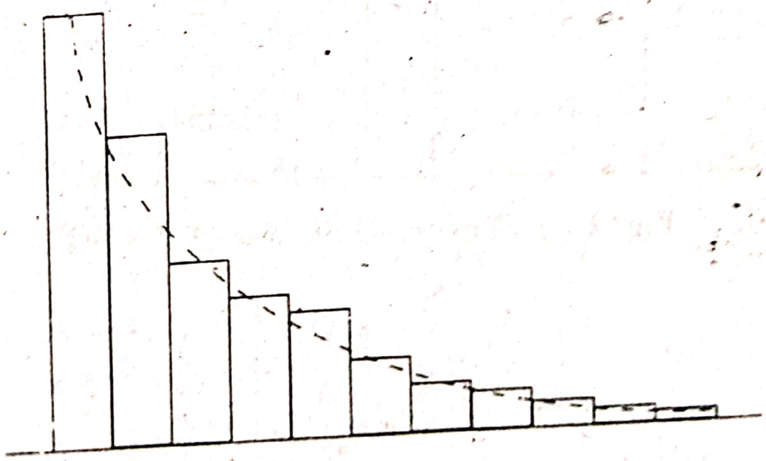


Fig. 2.8(b) Reverse J-shaped Distribution

2.12.4 U-shaped Distributions In such distributions the maximum frequencies occur at both ends and a minimum towards the centre as in Table 2.15. Figure 2... shows the histogram and the frequency curve of such a distribution.

Table 2.15

Age Group in Years	Death Rate (per 1000 population)
Below 5	45
5 – 9	30
10 – 14	18
15 – 39	12
40 – 49	24
Above 50	40

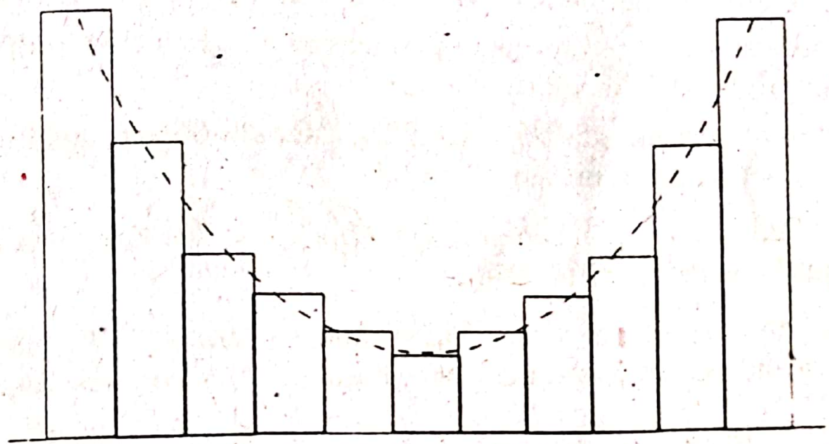


Fig. 2.9 U-shaped Distribution

2.12.5 Multi-modal Distributions Frequency distributions with more than one maximum frequency are called *multimodal distributions*. A distribution with two maxima is called a *bimodal distribution* as shown in Figure 2.10. The curve of a bimodal distribution has two humps while the curve of a multi-modal distribution has more than two humps.

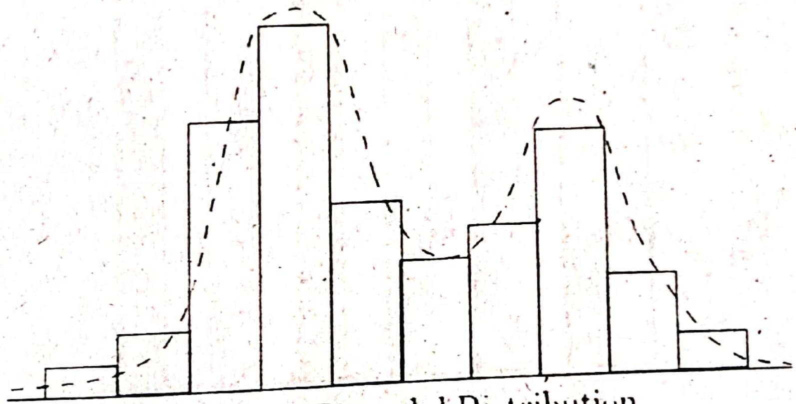


Fig. 2.10 Bi-modal Distribution

2.13 Charts or Diagrams Like graphs, charts or diagrams give visual representations of magnitudes, groupings, trends and patterns in the data. Diagrams also show comparisons between two or more sets of data.

The diagrams should be clear and easy to read and understand. Too much information should not be shown in the same diagram otherwise it may become confusing. Each diagram should include a brief and self-explanatory title dealing with the subject matter. The intervals on the vertical as well as the horizontal axi should be of equal size.

Diagrams are more suitable to illustrate the discrete data while continuou data are better represented by graphs.

2.13.1 Simple Bar Chart This chart consists of vertical or horizontal bars of equa width. The length of the bars is taken proportional to the magnitude of the value represented. The width of the bars has no significance. It is taken simply to make th charts attractive. Following rules are observed in drawing a simple bar chart.

(i) Vertical bars are used to represent data classified on quantitative c chronological basis while horizontal bars are used to represent data classifie on qualitative or geographical basis.

(ii) The bars should neither be very short and wide nor very long and narrow.

(iii) Bars should be separated by spaces which are not less than half the width of bar and greater than the width of a bar.

(iv) If the data do not relate to time, then they should be arranged in ascending c descending order of magnitude.

Example 2.5 Draw a simple bar chart to represent the production of wheat i Pakistan during the years 1981 to 1986.

Solution The data relate to time. Thus there is no need to arrange them i ascending or descending order of magnitude. Figure 2.11 shows the required chart.

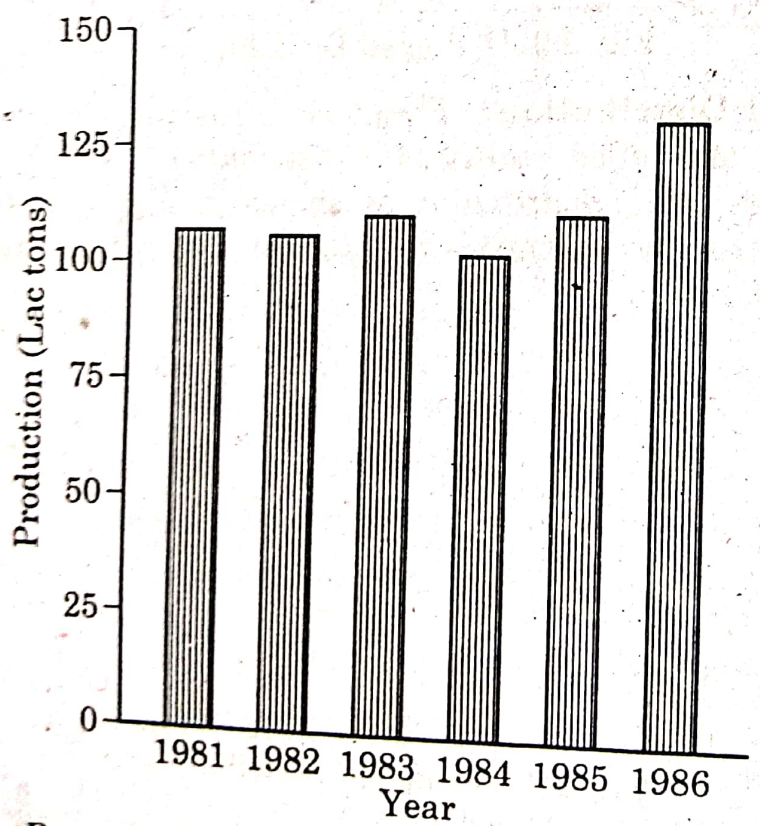


Fig. 2.11 Simple Bar Chart Showing Production of Wheat in Pakistan for the years 1981 to 1986

Example 2.6 Draw a simple bar chart or a simple bar diagram to represent the population of five Asian Countries in 1988.

Country	China	India	Indonesia	Japan	Pakistan
Population (million)	1088	816	175	123	106

Source: World Development Report, 1990.

Solution The data are already arranged in descending order of magnitude. Figure 2.12 gives the required chart.

Year	1981	1982	1983	1984	1985	1986
Production (Lac tons)	115	113	124	109	117	139

Source; Federal Bureau of Statistics.

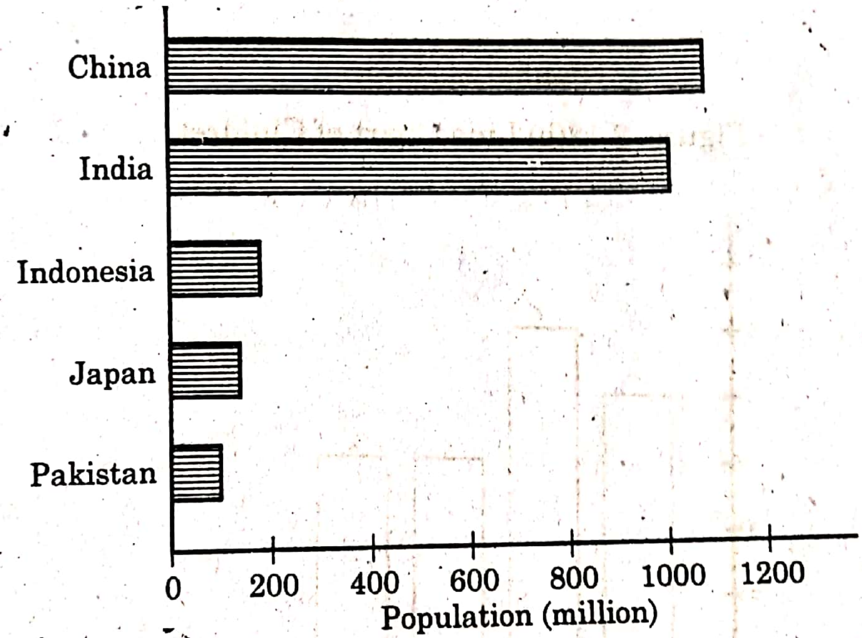


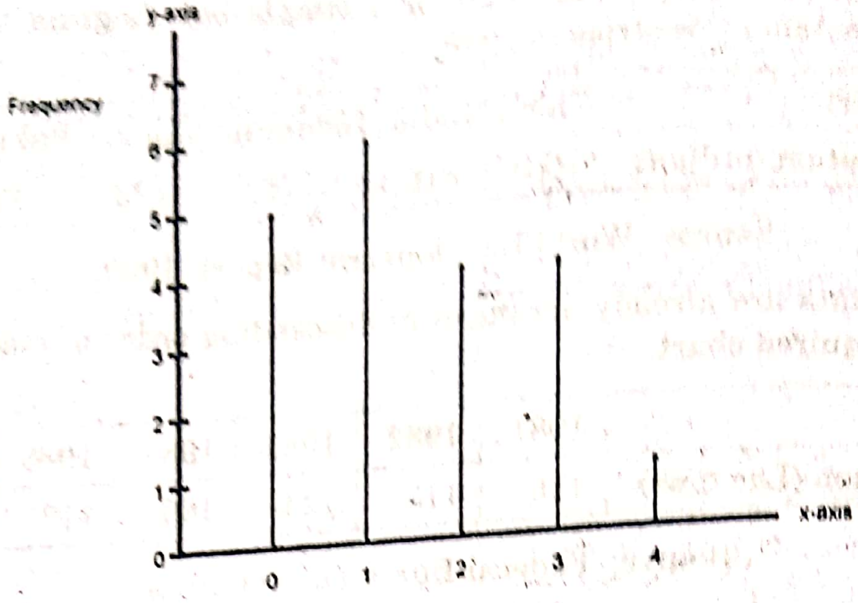
Fig. 2.12(a) Simple Bar Chart Showing Population of Five Asian Countries

Line Chart or Bar diagram To get an impression of the distribution of a discrete or categorical data set, it is usual to represent it by a line chart or bar diagrams. To construct a bar diagram or line chart the values of the variable or categories are taken along x-axis and a bar with height equal to its frequency is drawn on each category.

Example Make a line chart or bar diagram of the following data.

The number of children	0	1	2	3	4
The number of families	5	6	4	4	1

Solution To construct a bar diagram or line chart, the number of children are taken along the x-axis. The number of children vary from 0 to 4. S owe mark the x-axis with 0, 1, 2, 3 and 4. The value 0 has frequency 5, so a bar of height 5 is drawn along x-axis at point 0 on x-axis; similarly a bar of a bar of height 4 is drawn along y-axis at point 2 and 3 on x-axis and finally a bar of height 1 is drawn on point 4 It is shown in figures 2.12(b) (c).



Number of children

Figure: 2.12(b) Line Chart of Children

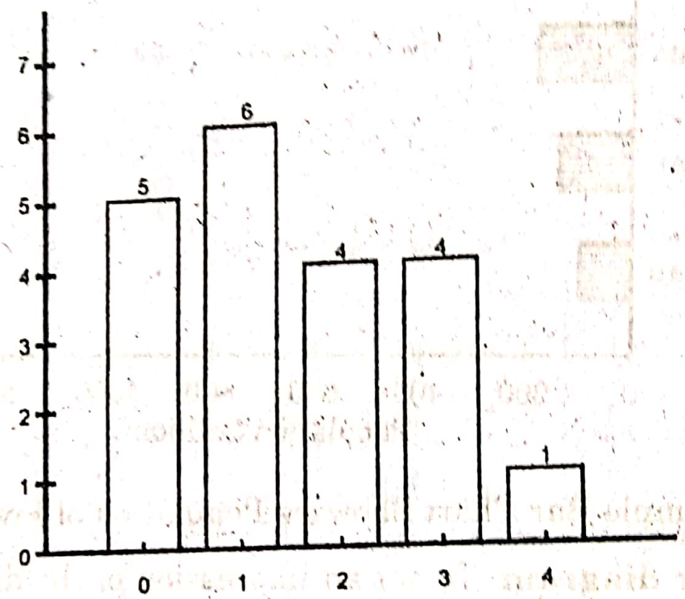


Fig: 2.12(c) Bar Diagram

The gaps between the bars in the bar chart emphasize the gaps between values that the discrete variable can take.

2.13.2 Multiple Bar Chart This chart is simply an extension of simple bar ch In this chart, grouped bars are used to represent related sets of data. For examp we may represent the imports and exports of a country for a number of years means of multiple bar charts taking groups of two bars each – one represent imports and the other representing exports. Each bar in a group is shaded coloured differently for the sake of distinction.

Example 2.7 Draw a multiple bar chart to represent the imports and exports of Pakistan (value in crores of Rupees) for the years 1982-83 to 1987-88.

Years	Value (Crores of Rupees)	
	Imports	Exports
1982 – 83	6815	3444
1983 – 84	7671	3733
1984 – 85	8978	3798
1985 – 86	9095	4959
1986 – 87	9243	6335
1987 – 88	11138	7844

Source: State Bank of Pakistan.

Solution Figure 2.13 gives the required chart.

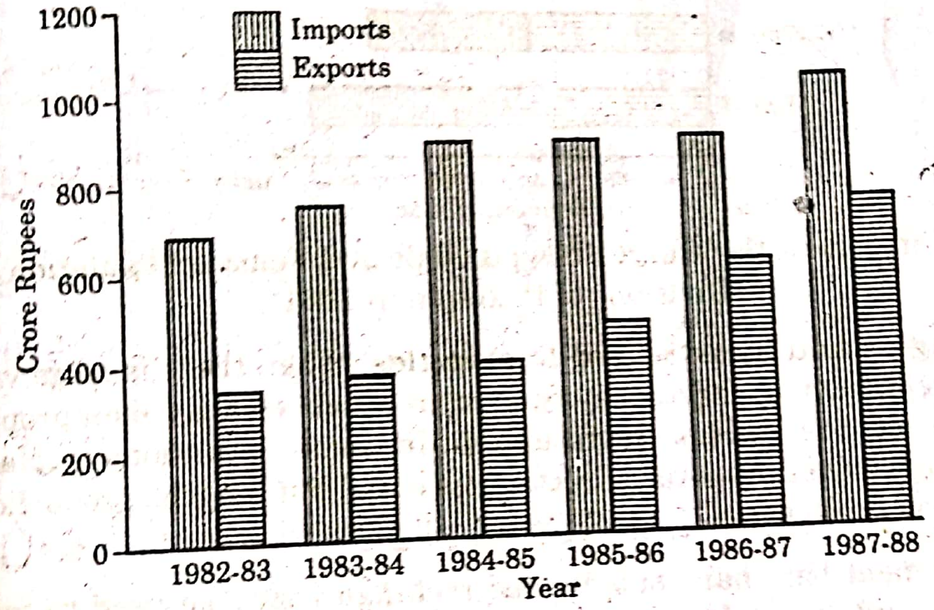


Fig. 2.13 Multiple Bar Chart Showing Imports and Exports of Pakistan for 1982-83 to 1987-88

2.13.3 Component Bar Chart This chart consists of horizontal or vertical bars which are sub-divided into two or more parts. It is called component bar chart because the bars show the various component parts. This chart is used when it is desired to present data which are sub-divisions of totals. In this chart, bars are drawn with lengths proportional to the totals. The bars are then sub-divided into parts in the ratio of their components. The component parts are shaded or coloured differently so as to distinguish different parts. For example, if we want to represent the male and female population of a number of countries or districts, we will draw bars with lengths proportional to the total populations. The males and females will be shown in each bar by sub-dividing the bars proportionately.

Example 2.8 Draw a component bar chart to represent the male and female population in Lac of five divisions of the Punjab in 1981.

Division	Male	Female	Both sexes
Rawalpindi	23	21	44
Bahawalpur	24	23	47
Multan	39	36	75
Gujranwala	39	37	76
Lahore	44	43	87

Source: Population Census Report, 1981.

Solution Figure 2.14 gives the required chart with data arranged according to si

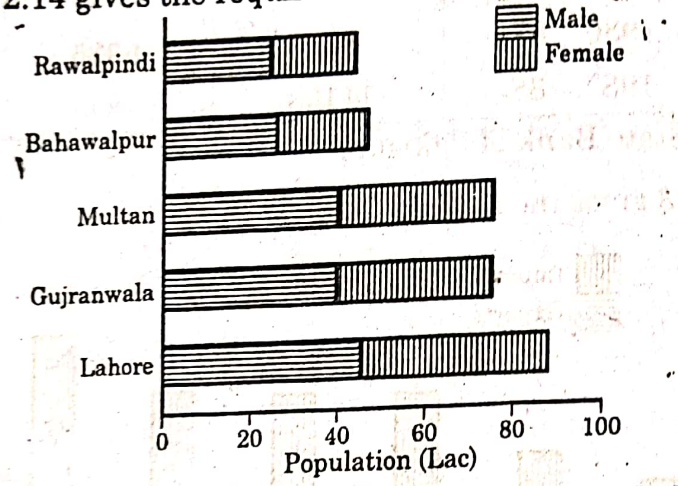


Fig. 2.14 Component Bar Chart Showing Male and Female Population of Five Divisions of Pakistan in 1981

2.13.4 Rectangles and Sub-divided Rectangles When there is large variation the data to be represented, charts may be used in which area is taken proportional the magnitude of the values represented. In such situations rectangles a constructed taking either breadths of rectangles equal but lengths proportional to t size of the values or vice versa.

Like component bar chart, sub-divided rectangles may be used to compare t total magnitude and its components. Sub-divided rectangles are constructed wi length equal to 100 and width proportional to the size of the values. The componer are expressed as percentages of their respective totals and are shown by differe shades in the rectangles. Sub-divided rectangles are generally drawn to compare t budgets of various families.

Example 2.10 The following table gives expenditure on different items for t families *A* and *B*. Represent the data by sub-divided rectangles.

Item of Expenditure	Family *A*	Family *B*
Food	60	105
Clothing	15	75
House Rent	18	60
Education	15	45
Miscellaneous	12	15
Total	120	300

Solution To construct sub-divided rectangles, we need percentages of expenditure on different items. For family *A*, the percentage of expenditure on food is $\frac{60}{120} \times 100 =$ 50%. The percentages for clothing, house rent, education and miscellaneous are 12.5%, 15%, 12.5% and 10% respectively. The corresponding percentages for family *B* are 35%, 25%, 20%, 15% and 5%.

Figure 2.16 shows the requisite sub-divided rectangles. Here the length of the rectangles has been taken equal to 100 and their breadth has been taken in the proportion 120 : 300 or 2 : 5.

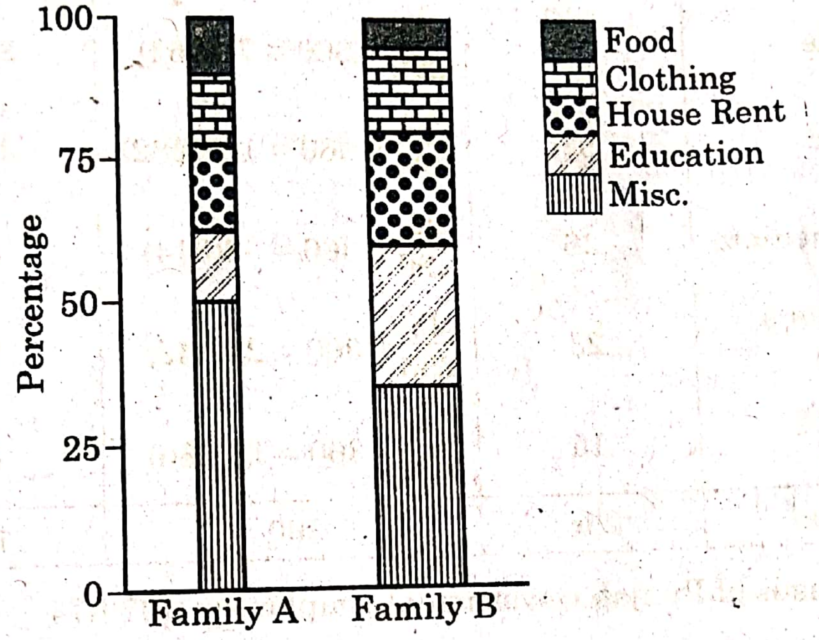


Fig. 2.16 Subdivided Rectangles Showing Budgets of Families *A* and *B*.

2.13.6 Pie Chart Like the component bar chart, pie chart can be used to compare the relation between the whole and its components. The only difference between component bar chart and pie chart is that in case of component bar chart the length of the bars are used while in case of a pie chart the area of the sector of a circle is used. In this chart, circles are drawn with radii proportional to the square root of the quantities to be represented because the area of a circle is given by $2\pi r^2$.

The sectors are shaded and coloured differently. The titles describing each component part should be written in each sector if space permits otherwise a key should be used.

To construct a pie chart, we draw a circle with some suitable radius (square root of the total). We know that a circle consists of 360°. To show the component parts by sectors, we calculate the angles for each sector by the formula:

$$\frac{\text{Component part}}{\text{Total}} \times 360$$

The circle is divided into different sectors by constructing angles at the centre by means of a protractor. The arrangement of the sectors is generally anti-clockwise.

Example 2.11 Draw a pie chart to show the distribution of Punjab Government Employees by their academic qualifications.

Academic Qualification	Number of Employees (thousand)	Angle of Sector (degrees)	Percentage
No education	47	$\frac{47}{296} \times 360 = 57$	15.9
Primary	25	$\frac{25}{296} \times 360 = 30(87)$	8.4
Middle	63	$\frac{63}{296} \times 360 = 77(164)$	21.3
Matric	97	$\frac{97}{296} \times 360 = 118(282)$	32.8
Intermediate	26	$\frac{26}{296} \times 360 = 32(314)$	8.8
Bachelor's Degree	23	$\frac{23}{296} \times 360 = 28(342)$	7.8
Master's Degree	15	$\frac{15}{296} \times 360 = 18(360)$	5.0
Total	296	360	100

Source: Census of Punjab Government Employees, 1973-74.

Solution The angles have been shown calculated in the third column of the above table alongwith cumulative angles in parentheses. The corresponding percentages are also given in the table. Figure 2.17 shows the requisite pie chart. The required dividing lines can be drawn by using a protractor.

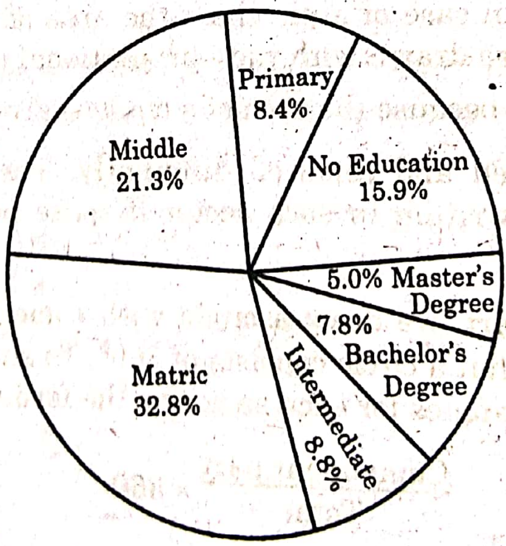


Fig. 2.17 Pie Chart Showing Punjab Government Employees by Academic Qualifications

Example 2.12 Construct pie chart to compare the budgets of two families *A* and *B*.

Item of Expenditure	Family *A*	Family *B*
Food	600	1050
Clothing	150	750
House Rent	180	600
Education	150	450
Miscellaneous	120	150
Total	1200	3000

Solution Calculation of angles of sectors for Families *A* and *B* have been shown in the following table. Cumulative angles have been shown in parentheses for convenience of plotting. The corresponding percentages are also shown in the table. Figure 2.18 shows the requisite pie charts for families A and B. The circles have been drawn with radii proportional to square root of 1200 and 3000.

Item of Expenditure	Family A		Family B	
	Angle of Sector	Percentage	Angle of Sector	Percentage
Food	$\frac{600}{1200} \times 360 = 180°$	50%	$\frac{1050}{3000} \times 360 = 126°$	35%
Clothing	$\frac{150}{1200} \times 360 = 45°(225°)$	12.5%	$\frac{750}{3000} \times 360 = 90°(216°)$	25%
House Rent	$\frac{180}{1200} \times 360 = 54°(279°)$	15%	$\frac{600}{3000} \times 360 = 72°(288°)$	20%
Education	$\frac{150}{1200} \times 360 = 45°(324°)$	12.5%	$\frac{450}{3000} \times 360 = 54°(342°)$	15%
Misc.	$\frac{120}{1200} \times 360 = 36°(360°)$	10%	$\frac{150}{3000} \times 360 = 18°(360°)$	5%

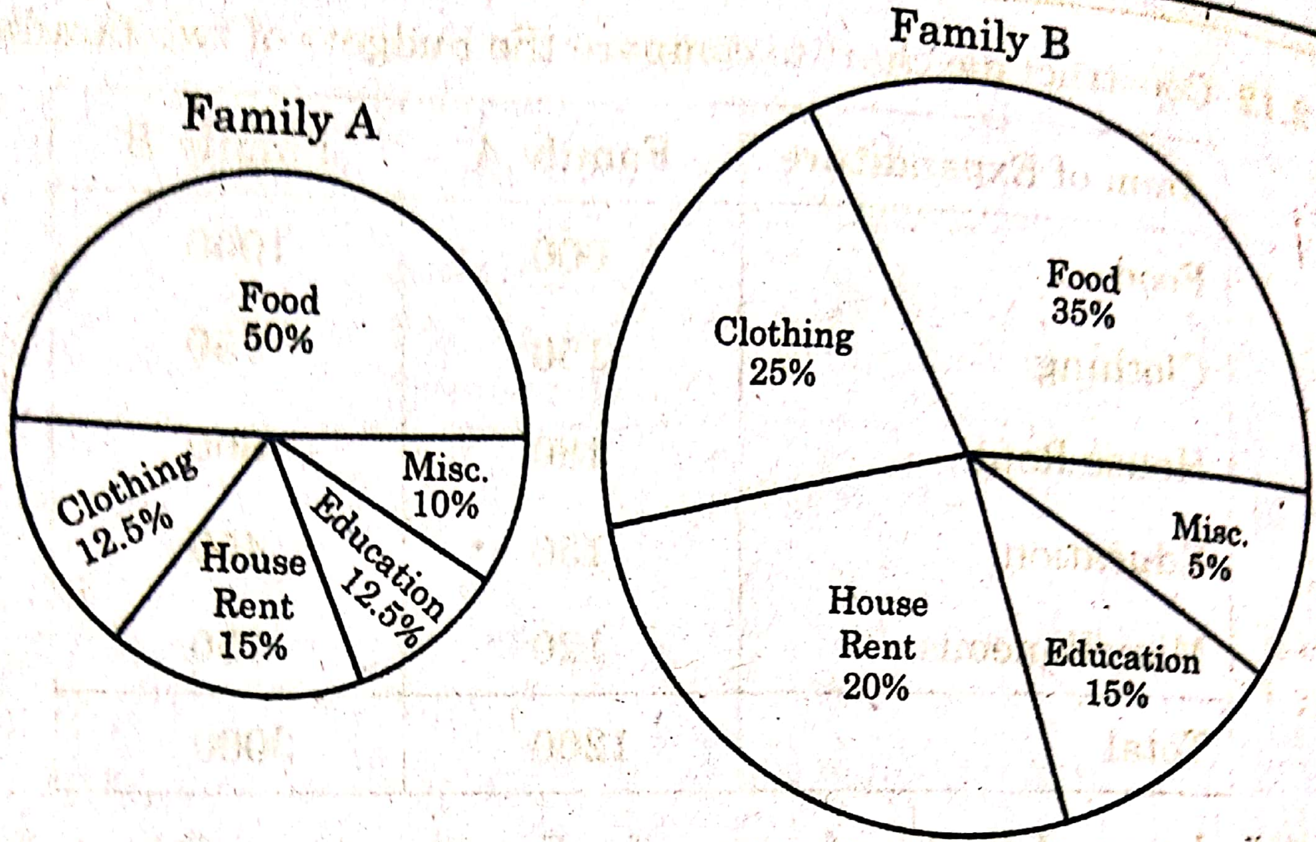


Fig. 2.18 Pie Charts Showing Budgets of Families A and B