"Statistics are measurements, enumerations or estimates of natural or social phenomena, usually systematically arranged, analysed and presented as to exhibit important inter–relationships among them."

– A. M. Tuttle

2

Presentation of Data

Contents :

2.1 Classification

Meaning and necessity

In the previous chapter, we have seen that the statistical information consists of two types of data : quantitative data and qualitative data. Quantitative data are based on quantitative variable where as qualitative data are based on qualitative variable. Quantitative variable is of two types (i) discrete variable and (ii) continuous variable. If a variable can assume definite or countable values within the specified range, then it is called discrete variable. For example, number of children per family, number of accidents on road. If a variable x assumes the pecific values 1.2, 1.3, 1.5, etc. then it is also called discrete variable. If a variable can assume any value within the specified range, then it is called continuous variable. For example, height of person, maximum temperature on a day, etc. are the examples of continuous variable. In practice, if we have to count something for obtaining the value of variable, then it is called discrete variable. For example, for obtaining the value of number of children per family, we have to count the children, for obtaining the value of number of accident on road, we have to count the accidents, so they are examples of discrete variables. If we have to measure the observation for obtaining the value of variable or if the value of the variable is expressed along with the measuring unit, then it is called continuous variable. Height of person is measured in cm. or inches or feet, maximum temperature is measured in celsius so they are the examples of continuous variables. The data based on discrete variable are called discrete data and the data based on continuous variable are called continuous data.

The data obtained at the end of sample inquiry or population inquiry are called raw data or ungrouped data. These data are in a haphazard form so the statistical analysis becomes difficult. Hence it is necessary to arrange the data in systematic and short form. A process of arranging ungrouped or raw data into proper form is called classification of data and the data thus obtained are called classified data. For example, it is known that the daily demands of a certain commodity during a week are 12, 16, 8, 12, 8, 8, and 10 respectively. From these raw data, it is clear that there are three days during which demand of commodity is 8 units, one day during which the demand is 10 units, two days during which the demand is 12 units and one day during which the demand is 16 units. This information can be presented in the following classified table :

Table showing daily demand of commodity during a week

Demand of commodity	8	10	12	16	**Total**
Number of days	3	1	2	1	7

Thus, the method of representing raw data into systematic and short form is called classification. In statistical study, the main reasons to classify the data are as follows :

(1) To represent large data into simple, short and attractive manner.

(2) For easy comparison between the various characteristics of the data. (In classification data are distributed in different groups according to the similarities of characteristics. Hence the comparisons become simple.)

(3) To save time, money and labour. (Analysis based on raw or ungrouped data requires more time, money and labour.)

(4) To obtain information easily, regarding various characteristics of the area under study.

2.2 Types of Classification

There are two types of classifications : (i) classification of quantitative data (ii) classification of qualitative data. Let us understand them by considering the following examples:

Suppose a sample of 100 families is selected from a region and the information regarding the 'number of children per family' is obtained. 100 observations regarding 'number of children per family' are collected and these data are called ungrouped or raw data. Now, on the basis of the study, it is concluded that there are 10 families having no children, 35 families having one child, 40 families having 2 children and 15 families having three children. This is called quantitative classification and concisely it can be represented as under :

Classification of number of children in 100 families

Number of children	0	1	2	3	Total
Number of families	10	35	40	15	100

In the above example, 'number of children per family' is a numeric variable and hence it is called numerical classification or frequency distribution. In the above example, instead of 'number of children per family', if the information regarding 'core occupation of family' is collected and on the basis of it, if it is concluded that core occupation of 30 families is farming, core occupation of 25 is business, core occupation 25 is service and core occupation of remaining 20 families is labour then this is also called classification and can be represented as under :

Tabulation of 100 families with reference to their core occupation

Core occupation of family	Farming	Business	Service	Labour	Total
Number of families	30	25	25	20	100

In the above example, 'core occupation of family' is qualitative variable, so it is called classification of qualitative data or tabulation of data.

Thus, the raw or ungrouped data are mainly classified as (i) classification of quantitative data (ii) qualitative classification of data.

2.2.1 Classification of Quantitative Data

There are two types of quantitative variable (i) discrete variable and (ii) continuous variable. Classification of discrete variable is called discrete frequency distribution and classification of continuous variable is called continuous frequency distribution.

2.2.1.1 Discrete Frequency Distribution

A numeric value showing the repetition of value of an observation is called frequency (f) of that observation. A table showing various possible values of discrete variable with their respective frequencies is called discrete frequency distribution. For understanding, let us consider the following illustration:

During the month of May, information regarding number of accidents per day on a particular road of a city is as under :

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

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

We have to classify these data.

'The number of accidents per day' given here is a discrete variable (x) and we prepare its frequency distribution as under:

The minimum value of variable x is 0 and its maximum value is 5. Therefore, the possible values of variable x are 0, 1, 2, 3, 4, and 5. We read the given ungrouped data in sequence and put a tally mark ('|') against the value of the variable that we read. If four tally marks are put against any value of the variable then the fifth tally mark is put across to have a group of five tally marks (𝍸). The purpose of doing this is only to simplify the counting. Continue reading of the raw data till all the vlues are exhausted. Then tally marks are counted against each value of the variable and the respective frequency (f) is obtained. The sum of all frequencies against the value of observations must be the total number of values of the variable i.e. in this illustration $n = \Sigma f = 31$. The frequency distribution thus obtained is as under :

A discrete frequency distribution of 'number of road accidents per day' during the month of May

Number of accidents x	**Tally marks**	**Number of days (f)**
0	𝍸, ||||	9
1	𝍸, ||	7
2	𝍸, ||	7
3	||||	4
4	||	2
5	||	2
	Total	**31**

The difference between maximum and minimum value of a variable of raw data is called range (R).

i.e. Range R = Maximum value – minimum value

For the above illustration $R = 5 - 0 = 5$.

Note : When the range of discrete variable is too large then the discrete frequency distribution of that data is not advisable. In such case, inclusive continuous frequency distribution is appropriate, which we will study with continuous frequency distribution.

Illustration 1 : In a television manufacturing company, 500 television sets are produced during a week. A sample of 50 television sets is drawn and each television set is examined. The number of defects per set is given below. Prepare an appropriate frequency distribution.

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

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

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

0 1 1 0 2

'Number of defects in a television set' is a discrete variable and its maximum value is 5 and minimum value is 0. Hence, the range of the given data is $R = 5 - 0 = 5$. Therefore, the discrete frequency distribution of number of defects per television set can be obtained as shown on page no. 18.

A discrete frequency distribution of 'number of defects per television set' in a sample of 50 units

Number of defects per television set (x)	Tally marks	Number of television sets (f)
0	𝍸 𝍸 𝍸 III	18
1	𝍸 𝍸 II	12
2	𝍸 IIII	9
3	𝍸 I	6
4	III	3
5	II	2
	Total	**50**

Activity

Collect the information regarding the number of family members for 30 families residing around your residence and prepare its frequency distribution.

2.2.1.2 Continuous Frequency Distribution

A continuous frequency distribution is carried out when the variable of raw data is continuous or range of the data is large. For this, first of all the number of classes 'K' or class length 'C' is decided for the raw data. Generally, depending on the raw data, any number from 6 to 15 is decided for the number of classes. It should be noted here that this is accepted as a universal trend which may not be true for every study. Under special circumstances, the number of classes may be less than 6 or more than 15. After deciding number of classes, the class length 'C' is determined for each class. For this purpose, range R of the raw data is divided by the number of classes 'K'. Practically, the value of number of classes K and the value of class length C are decided in such a way that they are positive integers and their product value is at least the value of range R. In notations, $C \cdot K \geq R$. If the class length for different classes are equal then it is called continuous frequency distribution with equal class length. Whenever the range of raw data is too large then depending on the given information, the class length of different classes may be different. It is called frequency distribution with unequal class length. After deciding the class length, class limits are decided for each class. On the basis of class limits, continuous frequency distribution can be categorized in two ways : exclusive frequency distribution and inclusive frequency distribution. If the upper limit of any class and the lower limit of its succeeding class are same then it is called exclusive class. In exclusive class, the observation having value equal to the lower limit is included in that class but the observation having value equal to the upper limit is included in the next class. e.g. for an exclusive class 30 – 35, observation having value 30 is included in this class whereas another observation having value 35 is included in the succeeding class. If the upper limit of any class and lower limit of its succeeding class are not same then it is called inclusive class. In inclusive class,

observation having value equal to the lower limit is included in that class and the observation having value equal to the upper limit is also included in the same class, e.g. for an inclusive class 30 - 35, observation having value 35 is included in this class.

Usually, exclusive continuous frequency distribution is carried out for the continuous raw data and inclusive continuous frequency distribution is carried out for the discrete raw data having large value of range. For converting inclusive continuous frequency distribution into exclusive continuous frequency distribution, each class limit of inclusive continuous frequency distribution is replaced by class boundary point. Lower boundary point and upper boundary point of any class are calculated by using following formulae :

$$\text{Lower boundary point of a class} = \frac{\text{value of lower limit of that class} + \text{value of upper limit of previous class}}{2}$$

$$\text{Upper boundary point of a class} = \frac{\text{value of upper limit of that class} + \text{value of lower limit of succeeding class}}{2}$$

OR

Upper boundary point of a class = lower boundary point + class length

Thus, the lower boundary point of a class is an average of lower limit of that class and the upper limit of previous class. Simlarly, the upper boundary point of a class is an average of the upper limit of that class and the lower limit of succeeding class. It should be noted here that for exclusive classes, their lower limit is the lower boundary point and the upper limit is the upper boundary point. The difference between the upper boundary point and lower boundary point of a class is called class length of that class.

Class length = upper boundary point – lower boundary point

An average of class limits is called mid point or mid value of that class.

$$\text{Mid value of a class} = \frac{\text{value of upper limit} + \text{value of lower limit}}{2}$$

When mid value of a class and class lengths are known then the class boundary Points of that class can determined by using the following formula :

$$\textit{Lower boundary point} = \textit{mid value} - \frac{1}{2}\ (\textit{Class length})$$

$$\textit{Upper boundary point} = \textit{mid value} + \frac{1}{2}\ (\textit{Class length})$$

Let us consider the following illustration for classifying ungrouped continuous data into continuous frequency distribution :

Following are the ages (in years) of employees working in a school.

32	34	48	31	34	27	57	36	49	51
45	29	36	46	46	49	51	47	50	30
35	41	36	47	30	35	48	53	37	47
45	30	50	44	49	43	42	46	28	48
52	36	43	38	39	50	49	34	36	50

Suppose we have to form a continuous frequency distribution of 7 classes from the above data.

Raw data on age of 50 employees of a school are given here and 'age of employee' is a continuous variable. We have to prepare its frequency distribution.

The youngest employee is 27 years old and the eldest one is 57 years old.

Therefore the range $R = 57 - 27$

$= 30$ years

The data are to be divided into 7 classes $\therefore K = 7$

$\therefore$ class length $C = \frac{\text{Range}}{\text{number of classes}} = \frac{R}{K}$

$= \frac{30}{7}$

$= 4.29$

The value of class length is usually taken as a positive integer. Hence, we can take $C = 4$ or $C = 5$. But if we take $C = 4$, then $C \cdot K = 4 \times 7 = 28$, which is less than the value of range. ($\because$ $C \cdot K \geq R$) Therefore, $C = 4$ is not possible. Now if $C = 5$, then $C \cdot K = 5 \times 7 = 35$ which is greater than the value of range. So, $C = 5$.

After deciding $C = 5$ and $K = 7$, the lower class limits for the first class is decided in such a way that it contains the smallest value 27. So, we can take the lower limit as 25 and if we add class length 5 to it then the upper limit is 30. Therefore the first class is 25 – 30, next class is 30 – 35, and so on. Now the last class must contain the maximum value of the observation. Hence, it will be 55 – 60. It should be noted here that for the given data, frequency distribution other than this can also be made.

Now the given ungrouped data can be distributed into different classes by putting tally marks as under:

An exclusive continuous frequency distribution of 'age of 50 employees of a school'

Age of employee (years) (Exclusive classes)	**Tally marks**	**Number of employees Frequency (f)**
25 - 30	𝍸	3
30 - 35	𝍸 𝍸	8
35 - 40	𝍸 𝍸	10
40 - 45	𝍸	5
45 - 50	𝍸 𝍸 𝍸	15
50 - 55	𝍸 𝍸	8
55 - 60	𝍸	1
	Total	**50**

Illustration 2 : Figures regarding the sales (in thousand rupees) of different items in a super mall during four weeks are as follows:

228	125	100	90	115	125	230
220	130	80	95	160	180	200
200	128	120	85	185	140	265
230	135	127	100	145	150	210

Construct a frequency distribution by classifying these data into 8 classes.

The variable here is 'sales (in thousand rupees)' which is a continuous variable and the number of classes $K = 8$ is given.

$$\text{Range of given data } R = \text{Maximum value} - \text{Minimum value}$$
$$= 265 - 80$$
$$= 185$$
$$\therefore \quad \text{Class length } C = \frac{\text{Range}}{\text{Number of class}} = \frac{R}{K}$$
$$= \frac{185}{8}$$
$$= 23.125$$

$\therefore$ For the convenience, class length should be taken as 25, i.e. $C = 25$ $(C\,K \geq R)$

By taking the lower limit of the first class as 75 and the upper limit as 100, the minimum value 80 is included in the first class 75 – 100. Similarly the last class 250 – 275 includes the maximum value 265.

An exclusive continuous frequency distribution of 'sales (in thousand rupees)' during for weeks

Sales of items (thousand rupees) classes	Tally marks	Number of days Frequency (f)
75 - 100	IIII	4
100 - 125	IIII	4
125 - 150	⧸⧸⧸⧸ III	8
150 - 175	II	2
175 - 200	II	2
200 - 225	IIII	4
225 - 250	III	3
250 - 275	I	1
	Total	**28**

Illustration 3: During a season of 50 days, the data regarding the numbers of roses grown daily on the different plants in a garden are given below. Prepare a frequency distribution from it having one class as 30 – 39.

34 35 37 39 39 54 52 69 71 75 74 76 84 96 23 33 51 39
26 46 65 65 53 53 72 71 84 94 34 24 99 19 18 27 17 38
45 55 57 66 82 85 35 19 18 28 47 52 64 75

The variable 'number of roses on different rose plants of a garden' is a discrete variable.

$$\text{Range of the given data } R = 99 - 17$$
$$= 82$$

The range of this discrete variable is large, so it is advisable to prepare inclusive continuous frequency distribution. According to the given inclusive class 30 – 39, the class including the lowest value of observation 17 is 10 – 19 and the class including the highest value of observation 99 is 90 - 99.

An inclusive continuous frequency distribution of 'number of roses on different rose plants of a garden'

Number of roses (Inclusive classes)	Tally marks	Number of days Frequency (f)
10 - 19	𝍸	5
20 - 29	𝍸	5
30 - 39	𝍸 𝍸	10
40 - 49	𝍸	3
50 - 59	𝍸 𝍸	8
60 - 69	𝍸	5
70 - 79	𝍸 𝍸	7
80 - 89	𝍸	4
90 - 99	𝍸	3
	Total	**50**

Illustration 4 : By considering the inclusive frequency distribution obtained in illustration 3, find (i) Exclusive frequency distribution and (ii) Frequency distribution showing mid value of each class.

(i) In illustration 3, upper limit of the first class is 19 and the lower limit of the second class is 20 and its class length is 10. Lower boundary point of the second class is $\frac{20+19}{2}$ = 19.5 and its upper boundary point is 19.5 + 10 = 29.5. Thus, the class boundaries for the second class are 19.5 – 29.5 and class boundaries for the first class are 9.5 – 19.5.

The alternate method for calculating these is as under :

Upper limit for the first class is 19 and the lower limit for the next class is 20. The difference of these two limits (20 – 19 = 1) is divided by 2, which gives 0.5. Now subtract 0.5 from the lower limit and add 0.5 to the upper limit of a each class. We have the lower boundary point and upper boundary point of each class as under :

Number of roses (exclusive classes)	Number of days Frequency (f)
9.5 – 19.5	5
19.5 – 29.5	5
29.5 – 39.5	10
39.5 – 49.5	3
49.5 – 59.5	8
59.5 – 69.5	5
69.5 – 79.5	7
79.5 – 89.5	4
89.5 – 99.5	3
Total	**50**

(ii) Mid value of a class is obtained by taking average of the upper limit and the lower limit of that class. Hence the frequency distribution showing the mid value of each class is as under :

Number of roses **Inclusive classes**	**Mid value** $= \frac{\text{upper limit + lower limit}}{2}$	**Number of days** **Frequency (f)**
10 - 19	$\frac{10+19}{2} = 14.5$	5
20 - 29	$\frac{20+29}{2} = 24.5$	5
30 - 39	$\frac{30+39}{2} = 34.5$	10
40 - 49	$\frac{40+49}{2} = 44.5$	3
50 - 59	$\frac{50+59}{2} = 54.5$	8
60 - 69	$\frac{60+69}{2} = 64.5$	5
70 - 79	$\frac{70+79}{2} = 74.5$	7
80 - 89	$\frac{80+89}{2} = 84.5$	4
90 - 99	$\frac{90+99}{2} = 94.5$	3
	Total	**50**

Illustration 5: The following frequency distribution of unequal class length is obtained for some data. Prepare a frequency distribution stating class length and mid value of each class.

classes	0 - 20	20 - 50	50 - 70	70 - 90	90 - 100	Total
Frequency	20	30	30	15	5	100

A frequency distribution showing mid value and class length of each class is as under :

Classes	**Class length**	**Mid value**	**Frequency**
0 - 20	20 – 0 = 20	$\frac{0+20}{2} = 10$	20
20 - 50	50 – 20 = 30	$\frac{20+50}{2} = 35$	30
50 - 70	70 – 50 = 20	$\frac{50+70}{2} = 60$	30
70 - 90	90 – 70 = 20	$\frac{70+90}{2} = 80$	15
90 - 100	100 – 90 = 10	$\frac{90+100}{2} = 95$	5
		Total	**100**

Activity

For 30 families residing around your residence, collect information regarding the height of the eldest person of each family and prepare its frequency distribution.

2.2.1.3 Cumulative Frequency Distribution

Sum of the frequencies upto the value of an observation or a class is called cumulative frequency (*cf*) and its distribution is called cumulative frequency distribution.

The sum of the frequencies upto the specified value of the observation or specified upper boundary point of a class is called 'less than' type cumulative frequency of that specified value and the distribution is called 'less than' type cumulative frequency distribution.

The sum of the all frequencies of the specified value of the observation or the lower limit of the specified class and the values or classes succeeding to it is called as 'more than' type cumulative frequency. Its distribution is called 'more than' type cumulative frequency distribution.

When the cumulative frequency distribution is obtained by considering the value of a discrete variable, it is called discrete cumulative distribution, whereas the cumulative distribution obtained by considering the boundary points is called continuous cumulative frequency distribution.

Illustration 6 : A frequency distribution of number of children in 50 families of a region is as under :

Number of children (x)	0	1	2	3	**Total**
Number of families (f)	10	25	12	3	50

Obtain 'less than' type and 'more than' type cumulative frequency distributions for these data.

This distribution is for discrete variable. Therefore, 'less than' and 'more than' type cumulative frequency distributions can be obtained as under:

'less than' type discrete cumulative frequency distribution

Number of children (x)	**Number of families (f)**	**Number of children less than or equal to x ($\leq x$)**	**Cumulative frequency (cf)**
0	10	0	10 = 10
1	25	1	10 + 25 = 35
2	12	2	10 + 25 + 12 = 47
3	3	3	10 + 25 + 12 + 3 = 50
Total	**50**		

'More than' type discrete cumulative frequency distribution

Number of children (x)	Number of families (f)	Number of children x or above($\geq x$)	Cumulative frequency (cf)
0	10	0	3 + 12 + 25 + 10 = 50
1	25	1	3 + 12 + 25 = 40
2	12	2	3 + 12 = 15
3	3	3	3 = 3
Total	**50**		

Illustration 7: A frequency distribution of monthly income of 500 persons is as under. Obtain 'less than' type and 'more than' type cumulative frequency distributions :

Monthly Income (Thousand ₹)	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50	50 - 55	55 - 60	Total
Number of persons (f)	30	80	100	50	150	80	10	500

An exclusive continuous frequency distribution of monthly income of 500 persons is given here and the following 'less than' type cumulative frequency distribution is obtained from it by considering upper boundary points. It is clear from the given data that no one is earning less than ₹ 25000. For indicating this, the lower boundary point of the first class is taken as upper boundary point of previous class with frequency zero.

'less than' type cumulative frequency distribution showing the monthly income of 500 persons

monthly income lesser than upper boundary point (Thousand rupees)	'less than' type cumulative frequency (cf)
25	0 = 0
30	0 + 30 = 30
35	0 + 30 + 80 = 110
40	0 + 30 + 80 + 100 = 210
45	0 + 30 + 80 + 100 + 50 = 260
50	0 + 30 + 80 + 100 + 50 + 150 = 410
55	0 + 30 + 80 + 100 + 50 + 150 + 80 = 490
60	0 + 30 + 80 + 100 + 50 + 150 + 80 + 10 = 500

Now, by considering the lower boundary point of each class, 'more than' type cumulative frequency distribution is obtained as under. It is clear from the given data that no one is earning more than ₹ 60,000. For indicating this, the upper boundary point of the last class is taken as the lower boundary point of the next class with frequency zero.

'more than' type cumulative frequency distribution showing the monthly income of 500 persons

Income more than or equal to lower boundary point (thousand ₹)	'more than' type cumulative frequency
25	10 + 80 + 150 + 50 + 100 + 80 + 30 = 500
30	10 + 80 + 150 + 50 + 100 + 80 = 470
35	10 + 80 + 150 + 50 +100 = 390
40	10 + 80 + 150 + 50 = 290
45	10 + 80 + 150 = 240
50	10 + 80 = 90
55	10 = 10
60	0 = 0

Illustration 8: The frequency distribution of daily demand of rooms at an international hotel during 90 days is as under. Obtain 'less than' type and 'more than type' cumulative frequency distribution from it .

Demand of rooms	1 - 50	51 - 100	101 - 150	151 - 200	201 - 250	Total
Number of days	10	20	30	18	12	90

This inclusive continuous frequency distribution can be expressed as exclusive continuous frequency distribution as under :

Demand of rooms	No. of days
0.5 - 50.5	10
50.5 - 100.5	20
100.5 - 150.5	30
150.5 - 200.5	18
200.5 - 250.5	12
Total	**90**

From the above exclusive classes, the 'less than' and 'more than' cumulative frequency distributions can be obtained as under :

'less than' type cumulative frequency distribution showing the demand of rooms at the hotel during 90 days

Demand less than upper boundary	'less than' type cumulative frequency
0.5	0 = 0
50.5	0 + 10 = 10
100.5	0 + 10 + 20 = 30
150.5	0 + 10 + 20 + 30 = 60
200.5	0 + 10 + 20 + 30 + 18 = 78
250.5	0 + 10 + 20 + 30 + 18 + 12 = 90

'More than' type cumulative frequency distribution showing the demand of rooms at hotel during 90 days

Demand more than or equal to lower boundary point	'More than' type cumulative frequency
0.5	12 + 18 + 30 + 20 + 10 = 90
50.5	12 + 18 + 30 + 20 = 80
100.5	12 + 18 + 30 = 60
150.5	12 + 18 = 30
200.5	12 = 12
250.5	0 = 0

Illustration 9 : 'Less than' type cumulative frequency distribution of weight (in kg.) of 50 persons is given in the following table.

less than upper boundary point Weight (kg.)	'less than' type cumulative frequency
30	0
35	7
40	15
45	30
50	38
55	44
60	47
65	49
70	50

(i) How many persons have weight less than 45 kg. ?

(ii) How many persons have weight between 50 kg. and 65 kg. ?

(iii) Obtain the original frequency distribution.

(i) From the given table, it is clear that 30 persons have weight less than 45 kg.

(ii) Number of persons having weight less than 65 kg. = 49

Number of persons having weight less than 50 kg. = 38

$\therefore$ Number of persons having weight between 65 and 50 kg. = 49 – 38 = 11

(iii) Original distribution of weights of 50 persons is as under :

Weight (kg.) Class	Number of persons (frequency)
30 - 35	7 – 0 = 7
35 - 40	15 – 7 = 8
40 - 45	30 – 15 = 15
45 - 50	38 – 30 = 8
50 - 55	44 – 38 = 6
55 - 60	47 – 44 = 3
60 - 65	49 – 47 = 2
65 - 70	50 – 49 = 1
Total	**50**

Illustration 10 : From the following 'more than' cumulative frequency distribution,

(i) Determine the number of the persons having age 40 or more. (ii) Determine the number of the persons having age less than 40 (iii) Determine original frequency distribution.

Lower boundary point or more than it	25	30	35	40	45	50	55	60	65	70
Cumulative frequency of persons	100	96	87	70	45	25	14	6	1	0

(i) The number of persons having age 40 or more = 70

(ii) Total frequency = 100

Number of persons having age less than 40

= Total number of persons – number of persons having age 40 or more.

= 100 – 70 = 30

(iii) Original frequency distribution can be obtained as under :

Age	Number of Persons
25 - 30	100 – 96 = 4
30 - 35	96 – 87 = 9
35 - 40	87 – 70 = 17
40 - 45	70 – 45 = 25
45 - 50	45 – 25 = 20
50 - 55	25 – 14 = 11
55 - 60	14 – 6 = 8
60 - 65	6 – 1 = 5
65 - 70	1 – 0 = 1
Total	**100**

2.2.1.4 Points for preparing continuous frequency distribution :

Some important points to be considered for preparing continuous frequency distribution from ungrouped data are as under :

(1) Normally, continuous frequency distribution may consist minimum 6 and maximum 15 classes to represent the raw data. ln exceptional situation, the number of classes may be other than these, depending upon the given data.

(2) Total number of classes is denoted by K.

(3) Range R is obtained for the given raw data.

$$Range\ R = Maximum\ value\ of\ observations - Minimum\ value\ of\ observations.$$

(4) The value of class length C is determined by $C = \frac{R}{K} = \frac{\text{Range}}{\text{No. of classes}}$ such that $C \cdot K \geq R$.

(5) Depending on the class length, class limits are decided in such a way that the first class must include the value of the minimum observation of the raw data and the last class must include the maximum value of observations of the raw data. Usually, frequency distribution with equal class length is preferred but when the range of the data is large then irregular class length is preferred.

(6) When mid values of the class and class lengths of a distribution are known, class boundary points can be determined by using the following formula.

$$Lower\ boundary\ point = mid\ value - \frac{1}{2}\ (Class\ length)$$

$$Upper\ boundary\ point = mid\ value + \frac{1}{2}\ (Class\ length)$$

(7) Usually, exclusive classes are to be prepared for the continuous data. Whereas, when the range of the discrete data is large then it is a normal practice to prepare inclusive classes.

(8) Inclusive classes should be converted into exclusive classes for preparing cumulative frequency distribution.

(9) In order to prepare 'less than' type cumulative frequency distribution from the continuous frequency distribution, the lower boundary point of the first class is taken as upper boundary point of previous class with respective frequency zero whereas in 'more than' type cumulative frequencies the upper boundary point of the last class is taken as the lower boundary point of the next class and its respective frequency is also taken as zero.

Note : By the classification of the raw data into continuous frequency distribution, the approximate values are used instead of original data. e.g. if the value of one observation is 8, then it is included in the class 0 – 10. Thus, the original value 8 is lost in such a classification. But on the basis of classification of the data, we can easily get idea about the spread and other characteristics of the data.

Illustration 11: Information regarding the ages of 250 drivers of a transport company is as under. Obtain continuous frequency distribution of equal class length from it.

Mid value of age of driver	22.5	27.5	32.5	37.5	42.5	47.5
Number of drivers (Frequency f)	25	30	50	80	50	15

'Age of driver' is a continuous variable. Hence, from the given data, it is advisable to prepare exclusive continuous frequency distribution.

Difference between two successive mid values is 5. So, the class length, $C = 5$

Lower limit of the first class = Mid value $- \frac{1}{2}$(Class length)

$= 22.5 - \frac{1}{2}(5)$

$= 22.5 - 2.5$

$= 20$

Upper limit of the first class = Mid value $+ \frac{1}{2}$(Class length)

$= 22.5 + \frac{1}{2}(5)$

$= 22.5 + 2.5$

$= 25$

$\therefore$ the first class is 20–25 and similary the class limits are obtained for the remaining classes.

An exclusive continuous classification showing the age of 250 drivers

Age of Driver	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50	**Total**
Number of drivers	25	30	50	80	50	15	250

Illustration 12: Information regarding the number of errors per page in a book of 500 pages is as under. Find the inclusive frequency distribution from it.

Mid value of a class of number of errors per page	0.5	2.5	4.5	6.5	8.5
Number of pages (f)	380	100	12	6	2

Number of errors per page of a book is a discrete variable. Hence, for the given data, it is advisable to have inclusive continuous frequency distribution.

Difference between two successive mid value is 2. So, class length, $C = 2$

Lower limit of the first class = Mid value $- \frac{1}{2}$ (Class length)

$= 0.5 - \frac{1}{2}(2)$

$= 0.5 - 1.0$

$= -0.5$

Upper limit of the first class = Mid value $- \frac{1}{2}$ (Class length)

$= 0.5 + \frac{1}{2}(2)$

$= 0.5 + 1$

$= 1.5$

$\therefore$ the first class is -0.5 to 1.5 and similarly the class limits are obtained for the remaining classes.

Number of errors per page	0.5 - 1.5	1.5 - 3.5	3.5 - 5.5	5.5 - 7.5	7.5 - 9.5	**Total**
Number of pages(f)	380	100	12	6	2	500

The above exclusive continuous frequency distribution can be converted into inclusive continuous frequency distribution by adding 0.5 to the lower boundary point and subtracting 0.5 from the upper boundary point.

An inclusive continuous frequency distribution of number of errors per page of a book

Number of errors per page	0 - 1	2 - 3	4 - 5	6 - 7	8 - 9	Total
Number of pages	380	100	12	6	2	500

Note : The constant added to the lower boundary point and subtracted from the upper boundary point for converting exclusive continuous frequency distribution into inclusive continuous frequency distribution is decided on basis of the given data.

Illustration 13 : The daily record of sales of electronic equipments sold by a shopkeeper of electronic items during the month of April in as under. Prepare an exclusive continuous frequency distribution for which one of the classes is 60 - 70. Hence,

(i) Find the number of days during which the sales of equipment is maximum.

(ii) State the number of equipments sold for the maximum number of days.

54 58 52 73 57 39 46 64 49 53 75 34 57 68 51

44 34 40 82 88 80 36 85 66 58 41 62 72 80 81

The variable 'number of electronic equipments' sold is a discrete variable but it is clearly given that we have to prepare an exclusive continuous frequency distribution for which one of the classes is 60 – 70.

First class = class including of the smallest value 34 = 30 – 40

Last class = class including the maximum value 88 = 80 – 90

An exclusive continuous frequency distribution of 'number of electronic equipments' sold

Number of equipments sold Exclusive classes	Tally marks	Number of days Frequency (f)
30 - 40	IIII	4
40 - 50	𝍸	5
50 - 60	𝍸 III	8
60 - 70	IIII	4
70 - 80	III	3
80 - 90	𝍸 I	6
	Total	**30**

(i) The maximum number of equipments sold is 80 to 90 and there are 6 such days.

(ii) 50 to 60 equipments are sold on maximum 8 days.

Illustration 14: Out of 300 persons residing in a region, a sample of 30 persons is selected at random and the heights (in cm.) of these selected persons are as under :

163 148 151 162 145 152 149 158 153 149

150 152 145 141 162 168 148 158 149 141

146 155 159 150 161 153 162 160 154 165

(i) Distribute these data into 6 classes and also find the mid value of each class.

(ii) Obtain 'less than' type cumulative frequency distribution.

(iii) What is the percentage of persons having height less than 155 cm.?

(iv) Obtain 'more than' type cumulative frequency distribution.

(v) How many persons have the height between 147 to 157 cm.?

(i) 'Height of a person' is a continuous variable.

Range of data R = 168–141

= 27

Number of classes $K = 6$

Class interval $C = \frac{R}{K} = 4.5 \simeq 5$

An exclusive continuous frequency distribution of height of 30 persons selected at random from a group of 300 persons

Height (cm.) Exclusive classes	Tally marks	Number of persons Frequency (f)	Mid value
140 - 145	II	2	142.5
145 - 150	𝍸 III	8	147.5
150 - 155	𝍸 III	8	152.5
155 - 160	IIII	4	157.5
160 - 165	𝍸 I	6	162.5
165 - 170	II	2	167.5
	Total	**30**	

(ii) 'less than' type cumulative frequency distribution of heights of 30 persons selected at random from a group of 300 persons is as under :

Upper boundary point (less than height)	'less than' type cumulative frequency (cf)
140	0 = 0
145	0 + 2 = 2
150	0 + 2 + 8 = 10
155	0 + 2 + 8 + 8 = 18
160	0 + 2 + 8 + 8 + 4 = 22
165	0 + 2 + 8 + 8 + 4 + 6 = 28
170	0 + 2 + 8 + 8 + 4 + 6 + 2 = 30

(iii) From the above table, it is clear that the number of persons having height less than 155 cm. = 18

$\therefore$ the percentage of persons = $\frac{18}{30} \times 100 = 60\ \%$

(iv) 'More than' type cumulative frequency distribution can be obtained as under :

Lower boundary point (minimum height)	'More than' type cumulative frequency
140	2 + 6 + 4 + 8 + 8 + 2 = 30
145	2 + 6 + 4 + 8 + 8 = 28
150	2 + 6 + 4 + 8 = 20
155	2 + 6 + 4 = 12
160	2 + 6 = 8
165	2 = 2
170	0 = 0

(v) To determine the number of persons having height 147 to 157 cm., we will use the original frequency distribution.

147 is included in the class 145 - 150, which has frequency 8.

Thus, when the class length is 5 (145 - 150) then frequency is 8.

When the class length is 3 (147 - 150) then the frequency is $\frac{3}{5} \times 8 = 4.8$

The frequency of the class 150 - 155 is 8 and the frequency for the class 155 - 157 can be calculated as follows:

When the class length is 5 (155 - 160) then frequency is 4

$\therefore$ When the class length is 2 (155 - 157) then the frequency is = $\frac{2}{5} \times 4 = 1.6$

$\therefore$ the total number of persons having height 147 to 157 cm. = $4.8 + 8 + 1.6 = 14.4 \approx 14$

Note : If we count the number of persons having height between 147 and 157 cm. from the given ungrouped data, we find 15 persons. But according to above calculation, it is 14. The obvious reason for this difference is, by classification into continuous frequency distribution, approximate values are used instead of the original raw data.

Illustration 15 : Changes in prices of shares during a day of 40 different companies registered at Bombay stock exchange are as follows. Find an inclusive continuous frequency distribution having mid value –1 for one of the classes and regular class length as 5.

– 8	8	7	16	8	22	6	10	– 7	5
3	– 4	9	– 11	11	16	9	– 3	– 11	2
5	– 6	10	– 6	13	– 5	3	– 7	12	0
7	6	12	– 5	21	0	4	–10	14	– 2

'Changes in the price of share' is a continuous variable.

The class limits of a class having mid value – 1 and class length $C = 5$ are :

$$\text{Lower limit} = \text{Mid value} - \frac{1}{2} C$$

$$= -1 - \frac{1}{2}(5)$$

$$= -3.5$$

$$\text{Upper limit} = \text{Mid value} + \frac{1}{2} C$$

$$= -1 + \frac{1}{2}(5)$$

$$= 1.5$$

Hence that class is – 3.5 to 1.5 which is exclusive class.

Converting it into the inclusive class, we get $-3.5 + 0.5$ to $1.5 - 0.5$

$= -3$ to 1

In the given data, the minimum value is –11 and its maximum value is 22. Hence the remaining classes should be made by including these values.

An inclusive continuous frequency distribution of change in price of shares of 40 compaines

Change in the price of shares	Tally marks	Number of companies (f)
–13 to –9	III	3
–8 to –4	𝍸 III	8
–3 to 1	IIII	4
2 to 6	𝍸 III	8
7 to 11	𝍸 IIII	9
12 to 16	𝍸 I	6
17 to 21	I	1
22 to 26	I	1
	Total	**40**

Illustration 16 : The monthly salaries (in rupees) of 24 employees working in a private company are given below. Prepare an appropriate frequency distribution from it :

3000, 3500, 4200, 5600, 7500, 9100, 10600, 16200, 18100, 24000, 30000, 36000

3200, 3800, 5200, 7000, 8400, 9600, 12800, 17700, 22750, 24900, 34000, 40000

'Salary of employee' is considered as a continuous variable. Range $R = 40000 - 3000 = 37000$ is too large. Hence, we can form a frequency distribution with unequal class lengths. The class lengths of different classes will be decided by studying the given raw data.

A continuous frequency distribution with unequal class length of salary of 24 employees of a company

Salary of employee (₹) (classes)	Tally marks	Number of employees frequency (f)
3000 - 5000	𝍸	5
5000 - 10000	𝍸 II	7
10000 - 20000	𝍸	5
20000 - 25000	III	3
25000 - 45000	IIII	4
	Total	**24**

Note : Here it is possible to have frequency distribution other than these classes.

Illustration 17 : Obtain the original frequency distribution from the following frequency distribution:

(i)

Class	24 - 29	24 - 34	24 - 39	24 - 44	24 - 49	24 - 54	24 - 59	24 - 64
Cumulative frequency	3	12	30	55	78	88	95	100

(ii)

Class	10 - 90	20 - 90	30 - 90	40 - 90	50 - 90	60 - 90	70 - 90	80 - 90
Cumulative frequency	200	180	140	90	55	30	8	3

(i) 'Less than' type cumulative frequency distribution is given here. The orginal frequency distribution can be obtained from it as under :

Classes	Cumulative frequency	Frequency (f)
24 - 29	3	3
29 - 34	12	12 – 3 = 9
34 - 39	30	30 – 12 = 18
39 - 44	55	55 – 30 = 25
44 - 49	78	78 – 55 = 23
49 - 54	88	88 – 78 = 10
54 - 59	95	95 – 88 = 7
59 - 64	100	100 – 95 = 5
	Total	**100**

(ii) 'More than' type cumulative frequency distribution is given here. The original frequency distribution can be obtained from it as under :

Classes	Cumulative frequency	Frequency (f)
10 - 20	200	200 – 180 = 20
20 - 30	180	180 – 140 = 40
30 - 40	140	140 – 90 = 50
40 - 50	90	90 – 55 = 35
50 - 60	55	55 – 30 = 25
60 - 70	30	30 – 8 = 22
70 - 80	8	8 – 3 = 5
80 – 90	3	= 3
	Total	**200**

EXERCISE 2.1

1. The data regarding the number of children of 50 families residing in a certain area are given below. Prepare appropriate frequency distribution :

1	1	2	1	1	1	1	2	1	0
0	2	2	0	3	3	2	1	2	1
2	1	3	1	1	2	2	2	1	2
3	0	3	0	2	1	2	2	2	2
0	1	2	2	2	2	3	3	2	1

2. The ages (in full years) of 60 employees working in an office are registered as follows. Prepare a frequency distribution by taking class length as 5 from this information.

32	42	48	35	23	58	52	38	36	44	48	39
24	27	29	32	34	41	45	51	30	47	45	44
52	38	41	31	25	38	36	34	37	51	25	56
32	39	32	35	42	26	46	42	57	28	43	33
31	42	43	53	43	39	27	54	21	47	26	40

3. The data regarding the number of mobile phones produced during last 60 days by a mobile phone manufacturing company is given below. Distribute it into 10 classes :

699	380	625	653	452	763	385	959	485	970
749	595	1029	500	499	453	525	621	465	565
103	785	286	1060	760	355	645	775	825	235
390	399	530	540	695	999	849	550	720	430
752	389	1075	701	875	552	351	265	199	370
1025	825	783	225	603	553	503	663	385	465

Obtain 'less than' and 'more than' type cumulative frequency distribution from it.

4. Rewrite the following frequency distribution by stating class length and mid value of each class.

Class	0 - 99	100 - 299	300 - 499	500 - 749	750 - 899	900 - 999
Frequency	10	12	14	16	8	10

5. From the following frequency distribution, obtain 'less than' and 'more than' type cumulative frequency distributions.

Number of errors per page	0	1	2	3
Number of pages	140	110	120	30

6. Obtain an inclusive continuous frequency distribution from the following data.

Lower boundary point or more than that	44.5	49.5	54.5	59.5	64.5	69.5	74.5	79.5
Cumulative frequency	500	470	390	290	240	90	10	0

7. Obtain an exclusive continuous frequency distribution from the following data.

Less than weight (kg.)	30	35	40	45	50	55	60	65	70
Cumulative frequency	0	17	25	40	48	54	57	59	60

8.

Mid values	25	105	230	400	650	900	**Total**
Frequency	10	30	40	60	80	30	250
Class length	50	110	140	200	300	200	

Obtain the original frequency distribution.

9. Information regarding the number of accidents in a city during a year is as under. Find the inclusive continuous frequency distribution from it.

No. of accidents (Mid values)	11.5	21.5	31.5	41.5	51.5	**Total**
No. of days	160	120	43	40	2	365

10. From the following data, obtain class boundary points from the class limits and write the frequency distribution.

Class	1 - 1.475	1.5 - 1.975	2 - 2.475	2.5 - 2.975	3 - 3.475	3.5 - 3.975	**Total**
Frequency	5	10	20	20	10	5	70

2.2.2 Classification of Qualitative Data

The process of systematic arrangement of qualitative or attribute raw data into rows and/or columns on the basis of standards of classification is called classification of qualitative data. Classification of qualitative data is done to represent qualitative ungrouped data into short and attractive form which is also called tabulation. Usually, there are two types of classification : (i) Simple classification or table and (ii) Manifold classification or table.

2.2.2.1 Simple Classification

A classification prepared on the basis of a single attribute is called simple classification or tabulation. e.g. The employees of a bank can be classified as a manager, a clerk, a peon, a security personnel, etc. based on their designation or position in the bank.

Illustration 18: On the basis of a study of different branches of a co-operative bank of Ahmedabad city, the following information is obtained. There are 20 security personnel, 30 peons, 40 clerks and 8 managers in the bank. Express this information in a table.

In the given data, 'designation of employee' is a qualitative characteristic.

A Table showing designation of the employees working at different branches of a co-operative bank of Ahmedabad city

Designation of employee	Number of employees
Security personnal	20
Peon	30
Clerk	40
Manager	8
Total	**98**

2.2.2.2 Manifold Classification

A classification of raw data carried out by considering more than one attribute of the units under study is called manifold classification.

Illustration 19: On the basis of the study of different branches of a co-operative bank of Ahmedabad city, the following information is obtained. In this bank, out of 20 employees working as security personnal, 6 are females, out of 30 peons, 10 are females, out of 40 clerks, 25 are females and out of 8 managers, 3 are females.

Express this information in a table.

In the given data, there are two attributes (i) 'designation of employee' and (ii) 'gender of employee'.

A Table showing designation and gender of the employees working at different branches of a co-operative bank of Ahmedabad city

Designation of employee	Gender		Total
	Male	Female	
Security person	14	**6**	**20**
Peon	20	**10**	**30**
Clerk	15	**25**	**40**
Manager	5	**3**	**8**
Total	**54**	**44**	**98**

Note : The figures shown in bold letters are the values given in the description of the illustration and the remaining values can be obtained by simple calculations.

Illustration 20 : On the basis of the study of different branches of a co-operative bank of Ahmedabad city, the following information is obtained. In this bank, out of 20 employees working as security persons, 7 males are married and out of 6 females, 4 are married. 12 peons out of 20 male peons are married and 10 female peons are unmarried. Out of 40 clerks, 25 are females and of them 12 are married whereas 7 male clerks were married and out of 8 managers, all 3 female managers are unmarried and all the male managers are married. Express this information in a table.

In the given data, there are three attributes (i) 'designation of employee' and (ii) 'gender of employee' and (iii) 'marital status of employee'

A Table showing designation, gender and marital status of the employees working at different branches of a co-operative bank of Ahmedabad city

Designation of employee	Gender								
	Male			Female			Total		
	Unmarried	married	Total	Unmarried	married	Total	Unmarried	married	Total
Security person	7	**7**	14	2	**4**	**6**	9	11	**20**
Peon	8	12	20	**10**	0	**10**	18	12	**30**
Clerk	8	**7**	15	13	**12**	**25**	21	19	**40**
Manager	0	**5**	5	**3**	0	**3**	3	5	**8**
Total	**23**	**31**	**54**	**28**	**16**	**44**	**51**	**47**	**98**

Note : The figures shown in bold letters are the given values in the description of the problem and the remaining values can be obtained by simple calculations. e.g. there are 20 security persons, of which 6 are females. Therefore, number of males = 20 – 6 =14. Also, out of 6 security female employees, 4 are married. Therefore, unmarried female security employees = 6 – 4 = 2. Out of 14 male security employees, 7 are married. Hence, the unmarried male security employees = 14 – 7 = 7. Total unmarried employees working as security =7 + 2 = 9 and total number of married security employees = 7 + 4 = 11 and the total of these is 9 + 11 = 20 which is the total number of security persons in the bank. Similarly, the calculations for the remaining employees can be carried out.

2.3 Types of Tabulation and its uses

After deciding the standards of classification the quantitative data or the qualitative data are arranged in the form of classified table. Simple table of numerical data is also called frequency distribution and frequency distribution of bi-variate data is called bi-variate frequency distribution which is not included in our syllabus.

Uses of tabulation :

(1) It represents the extensive data in simple, organized and precise manner.

(2) The time, money and labour required for the study of data under consideration is saved as unnecessary information is removed from the data.

(3) In tabulation, various characteristics to be compared are placed side by side. Hence, the comparison becomes easy.

(4) Row and/or column totals are found in the table. Therefore, any error occurring due to oversight can be detected and eventually rectified easily.

(5) The analysis of the data becomes simple and convenient if they are presented in table.

2.3.1 Guiding rules for Tabulation

In order to make the information more meaningful and to derive significant decisions easily from the table, some important guiding rules of tabulation are followed :

(1) Appropriate title should be given to the table.

(2) There should be clear and simple captions to the rows and columns.

(3) If the figures are too large then it should be represented in hundreds, thousands, lakhs or crores.

(4) The interrelated information should be shown adjacent to each other in such a way that the analysis can be done in a simple way the conclusions can be drawn easily.

(5) In order to distinguish between the main characteristics of the data in the table, it should be separated by lines.

(6) There should be a provision for indicating the totals of primary and subsidiary characteristics.

(7) Source of the data must be mentioned at the end of the table.

(8) Before preparing the final table, a rough table should be prepared.

(9) It is better to present as much data as possible in a single table, so that the comparison, computation and analysis becomes easy. But for large volume of data, it is advisable to prepare different tables instead of a single table.

For representing qualitative data, different persons can prepare different tables, but the table which satisfies the objective of the classification is called the best table. In practice, there may be more than 3 attributes to be classified through tabulation, but we have classified the data having maximum 3 attributes as per the limitation of our syllabus.

Illustration 21 : In a university, out of total 50,000 students, 35% are in commerce faculty, 30% are in arts faculty, 20% are in science faculty, 10% are in engineering faculty and remaining 5% are in medical faculty. The ratio of number of boys and girls in commerce faculty is 4:3. In arts faculty, the number of girls is double than that of boys. In science and engineering faculty, there are 60% and 70% boys respectively and in medical faculty, boys and girls are in equal numbers.

Represent the above data in appropriate table.

The given attributes are (i) faculty of students (ii) gender of students.

Number of students in commerce faculty = $50000 \times \frac{35}{100} = 17500$

The number of boys = $\frac{4}{4+3} \times 17500 = 10000$

The number of girls = $\frac{3}{4+3} \times 17500 = 7500$

Number of students in arts faculty = $50000 \times \frac{30}{100} = 15000$

The number of girls are double than that of boys.

If number of boys = x then the number of girls = $2x$

and $x + 2x = 15000$

$\therefore \quad x = 5000$

$\therefore$ Number of boys in arts faculty = 5000 and number of girls = 10000

Number of students in science faculty = $50000 \times \frac{20}{100} = 10000$

The number of boys = $10000 \times \frac{60}{100} = 6000$

$\therefore$ The number of girls = $10000 - 6000 = 4000$

Number of students in engineering faculty = $50000 \times \frac{10}{100} = 5000$

The number of boys = $5000 \times \frac{70}{100} = 3500$

$\therefore$ The number of girls = $5000 \times \frac{30}{100} = 1500$

Number of students in medical faculty = $50000 \times \frac{5}{100} = 2500$

The number of boys = $\frac{2500}{2} = 1250$

and the number of girls = $\frac{2500}{2} = 1250$

A table showing classification of 50,000 students of a university according to their faculty and gender

Faculty	Gender of student		Total
	Boys	Girls	
Commerce	10,000	7500	17500
Arts	5000	10,000	15,000
Science	6000	4000	10,000
Engineering	3500	1500	5000
Medical	1250	1250	2500
Total	**25,750**	**24,250**	**50,000**

Illustration 22: 80 members participated in a picnic organized by a school and the average contribution was ₹ 300 as expenditure. There were 60 students and each of them contributed ₹ 325. Teachers contributed little more for the picnic. There was a support staff of 10 males and contribution was not collected from them. 20% of those participated in picnic were females and 2 of them were teachers. Represent the data in a table.

There are two attributes : (1) Participant (2) Gender.

Total contribution for the picnic = $80 \times 300 = 24000$

Contribution of the students = $60 \times 325 = 19500$

Contribution of support staff = 0

Contribution of the teachers = $24000 - 19500 = 4500$

$\therefore$ Per head contribution of teacher = $\frac{4500}{10} = 450$

Table showing category and gender wise contribution of the participants of school picnic

Category of participant	Gender		Total	Per head Contribution (₹)	Total contribution (₹)
	Male	Female			
Students	46	14	**60**	325	19500
Support staff	**10**	–	**10**	–	–
Teacher	**8**	**2**	10	450	4500
Total	**64**	**16**	**80**	–	**24,000**

Activity

For 40 families residing around your residence, collect information for the adults of the family regarding their gender, education and marital status and represent it in a table.

EXERCISE 2.2

1. There were 1400 students studying in a commerce college. Among them, 855 were boys and of them, 225 boys were in the second year. In the second year, the number of boys and the number of girls are equal. Among the 550 students of the first year, the proportion of number of boys and girls is 3:2. In the third year, number of boys is three times the number of girls. Represent the above information in a table.

2. 1600 employees are working in an office. Among these employees the number of men exceeded the number of women by 15% of the total number of employees. The number of unmarried employees are 800 less than the number of married employees. The number of unmarried women is 195. Represent the above data in a suitable table.

3. Prepare a blank table by considering the following characteristics for the candidates who appear for the different jobs at a bank.

(1) Designation: Manager, clerk, cashier, peon.

(2) Marital status: Married, unmarried.

(3) Gender: Male, female.

4. Out of total 1850 women working in a factory, 549 were residing in labour area. Out of total married women of labour area, 250 had experience and 93 were inexperienced, the number of experienced and inexperienced women from other area were 87 and 400 respectively. The total number of inexperienced women was 1336 and out of them, 136 were from labour area. Out of total women, 1020 were unmarried. Among them, the number of experienced women from labour area and from other area were 163 and 14 respectively. Present these data in tabular form.

5. There were 1250 skilled and 400 unskilled workers in a private company in the year 2011. There were 220 female workers and of them, 140 were unskilled. In the year 2012, the number of skilled workers was 1475 and of them, 1300 were males. Out of 250 unskilled workers, 200 were males. In 2013, there were 1700 skilled and 50 unskilled workers. Out of total workers, 250 were females of them 240 were skilled. In the year 2014, there were 2000 workers and of them, 2% were unskilled. Out of total workers, 300 were females and of them, 10 were unskilled. Present the above data in the form of table.

2.4 Diagrams

Importance and Limitations of diagrams in statistics :

In order to understand and represent huge and complex data into simple and attractive manner, classified data are presented in the form of graphs or diagrams. As per our syllabus, we will discuss only diagrams. The main characteristics of the data presented by graphs are self-explanatory, e.g. for the forecasting of weather conditions on television, various diagrams are used. Similarly, various departments of central and state government, industrial firms, etc. publish their annual activities in various publications by using diagrams.

Importance : The importance of diagram is clear from their uses given below :

(1) Diagram represents the data in attractive, simple and concise form.

(2) The characteristics of the data expressed by diagrams are rememberd for longer time.

(3) When two or more sets of data are presented by diagrams on the same scale, the comparative study of those data becomes very simple.

(4) Diagrams have visual presentation of the data, so one can save time for studying the data.

(5) The core characteristics of the data conveyed by diagrams are easily understood by the illiterates, less educated or even by children.

(6) By using diagrams, industrialists and businessmen can make effective advertisement for their products.

(7) Diagrams are very useful for conveying the messages related to social reforms effectively,

(8) Pictorial diagrams are easy to understand irrespective of language barriers.

Limitations :

(1) If the diagram is not drawn accurately then it leads to wrong interpretations.

(2) Sometimes illusionary effect of diagrams misleads the public opinion.

(3) There is a loss of accuracy of the data by diagrammatic representation.

2.4.1 Types of Diagrams

There are three main types of diagrams :

(1) One dimensional diagram

(2) Two dimensional diagrams

(3) Pictorial diagrams

2.4.2 One Dimensional Diagram

A diagram drawn by considering only one characteristic of the data is called one dimensional diagram. We shall study the following four diagrams :

(1) Bar diagram

(2) Multiple or adjacent bar diagram

(3) Simple divided bar diagram

(4) Percentage divided bar diagram

2.4.2.1 Bar Diagram

Bar diagram is constructed by considering only one characteristic of the data. lt is used to represent the information on different places, things or time on the diagram. For this purpose, the different places, things or time are taken on x-axis on a graph paper and bars with equal widths at equal distance are drawn, with the heights proportional to the measure of respective places, things or time on y - axis with appropriate scale. The diagram formed by such bars is called bar diagram. In a bar diagram, the logical order of bars should be maintained. When the data based on different places or things are given then it should be arranged in proper order before preseting it on the graph, so that the comparative study can be done easily. When the data based on time are given then they are presented as it is on a graph paper.

Illustration 23 : Information regarding the number of students studying in a college in different faculties in a year is given below. Represent it by a suitable diagram.

Faculty	Arts	Commerce	Science	Engineering	Medical	Law
Number of students	500	1300	900	400	200	300

Only one characteristic 'Faculty' is to be expressed diagrammatically. So, we draw bar diagram. Arranging the data in descending order of number of students of faculty, the following table is obtained.

Faculty	Commerce	Science	Arts	Engineering	Law	Medical
Number of students	1300	900	500	400	300	200

By taking faculty on $x-$axis and number of students on $y-$axis, the bar diagram is as under :

Bar diagram showing the number of students in different faculties of a college

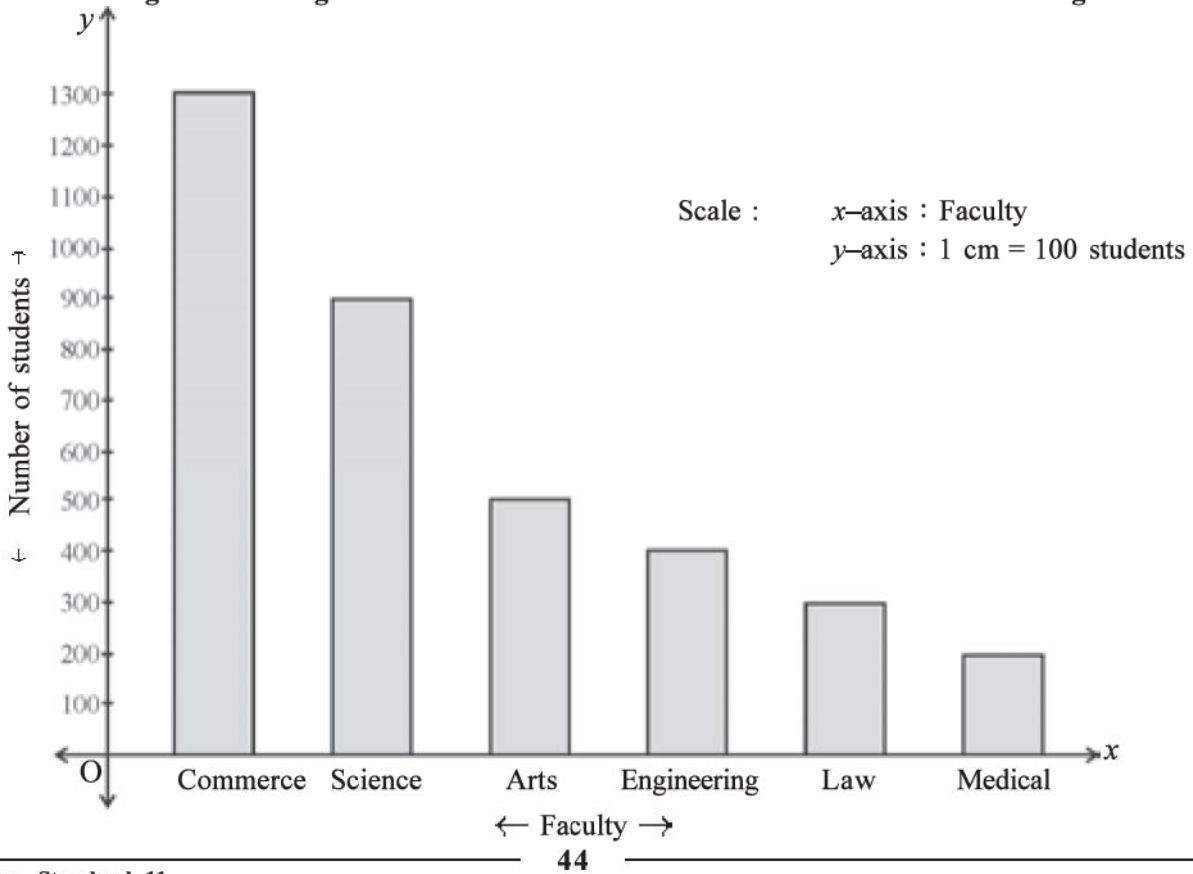
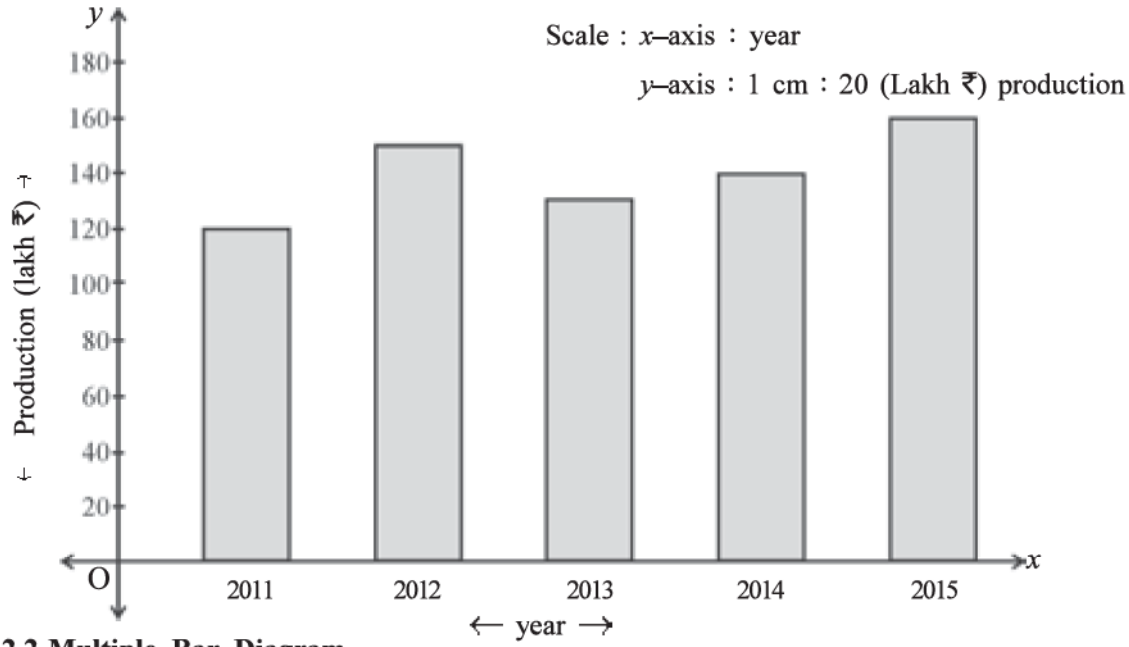


Illustration 24 : The information regarding the production (in lakh ₹) in a factory during five years is given below. Present it in a suitable diagram.

Year	2011	2012	2013	2014	2015
Production (Lakh ₹)	120	150	130	140	160

As the attirbute (production) dependent on time is to be shown on the graph, we will draw bar diagram. By taking year on x-axis and production (in lakh ₹) on y-axis, the bar diagram is as under :

Bar diagram showing the production (in lakh ₹) in a factory during five years



2.4.2.2 Multiple Bar Diagram

If the data about different places, things or time are collected on more than one mutually related characteristics then such data can be presented by using multiple bar diagrams or adjacent bar diagrams by placing the related bars close to each other. If the data are related to time then the bars are drawn in the order of time but when the data are not related to time then they are arranged in ascending or desending order by considering any one of the characteristics and then they are presented on graph.

Illustration 25 : The information regarding the number of boys and girls studying in a college in different faculties in a year is given below. Represent it by a suitable diagram.

Faculty	Science	Commerce	Arts	Engineering	Medical	Law
Number of Boys	500	700	200	300	100	100
Number of Girls	400	600	300	100	100	200

The information for two characteristics is given for each faculty. So, we draw adjacent bar diagram. First of all, we arrange the given data in descending order of number of boys in each faculty. The given data can be represented as under :

Faculty	Commerce	Science	Engineering	Arts	Medical	Law
Number of Boys	700	500	300	200	100	100
Number of Girls	600	400	100	300	100	200

The following diagram is drawn taking faculty on x-axis and the number of boys/girls on y-axis.

Multiple bar diagram showing the number of boys and girls studying in different faculties of a college

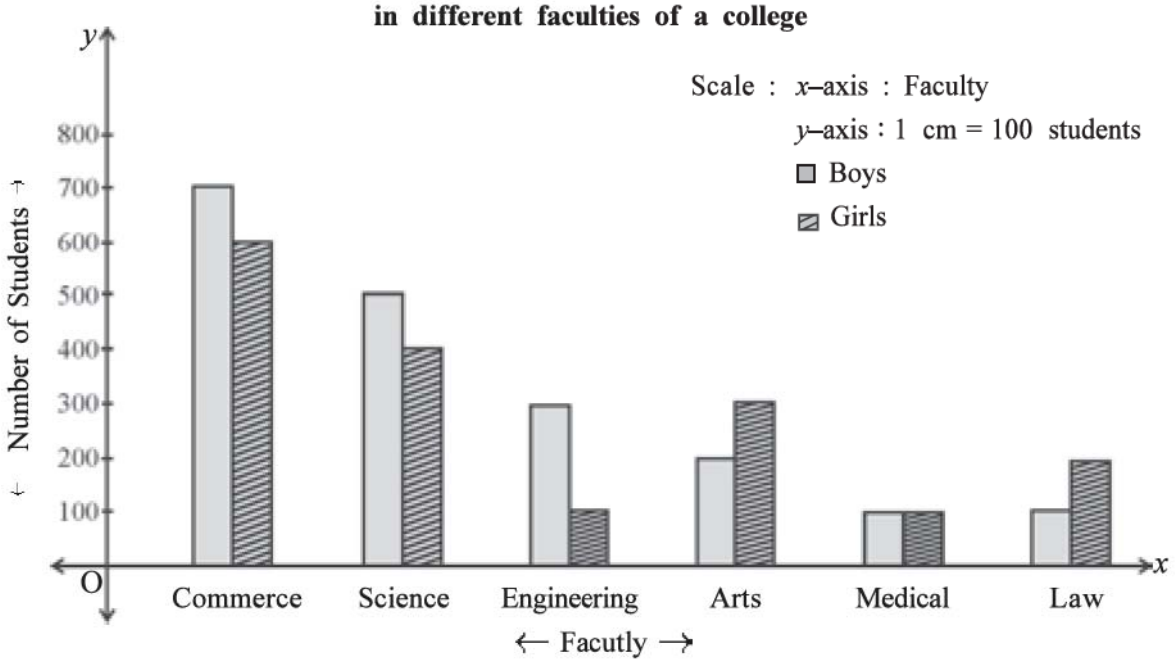
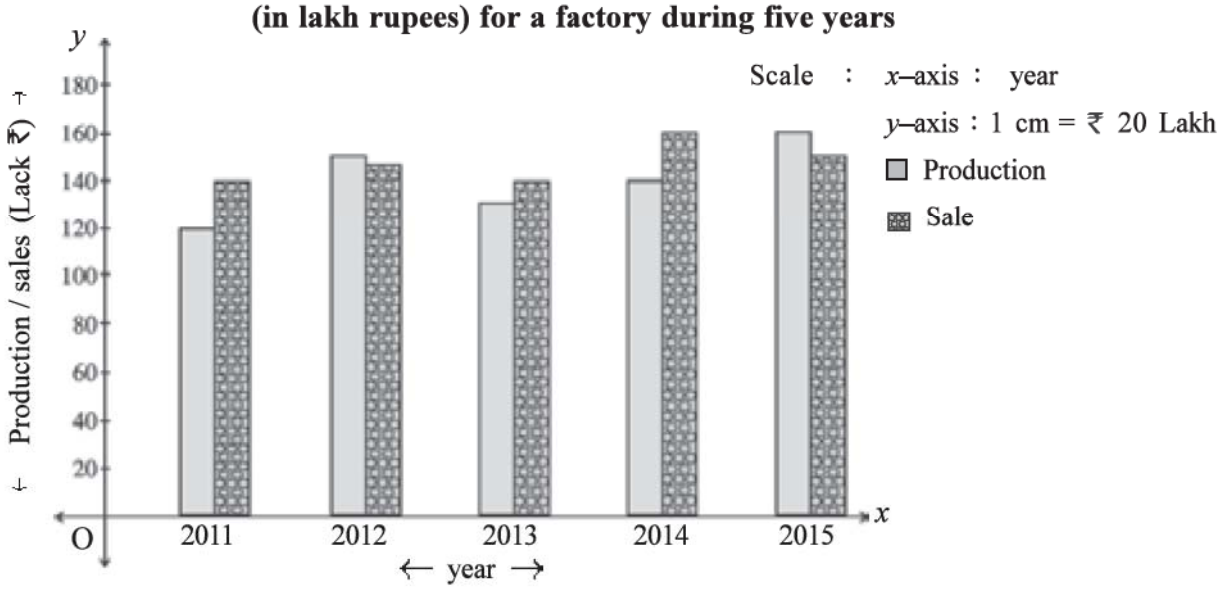


Illustration 26: The information regarding the production (in lakh ₹) and sales (in lakh ₹) for a factory during five years is given below. Present it in a suitable diagram.

Year	2011	2012	2013	2014	2015
Production (Lakh ₹)	120	150	130	140	160
Sales (Lakh ₹)	140	145	140	160	150

Year wise data regarding the production and sales for a factory are given here. So, we draw multiple bar diagram by taking years on x – axis and production/sales (in lakh ₹) on y – axis. The diagram is as follows :

Multiple bar diagram showing the production and sales (in lakh rupees) for a factory during five years



2.4.2.3 Simple divided Bar Diagram

If the data related to different places, things or times consist of several mutually related sub-data on different components are presented on bar diagram by dividing it in accordance with the sub-data, then the diagram obtained is called simple divided bar diagram. For example, if the information regarding the expenses on food, clothing, rent, fuel and miscellaneous for the livelihood of a family are given, then a bar of total expense which is divided into different segments indicated by various signs is called simple divided bar diagram. In short, when the total value is given, simple divided bar diagram is used for presenting its interrelated sub-data.

Illustration 27 : The data on monthly expenses of two different families living in a city are given below. Present it through appropriate diagram.

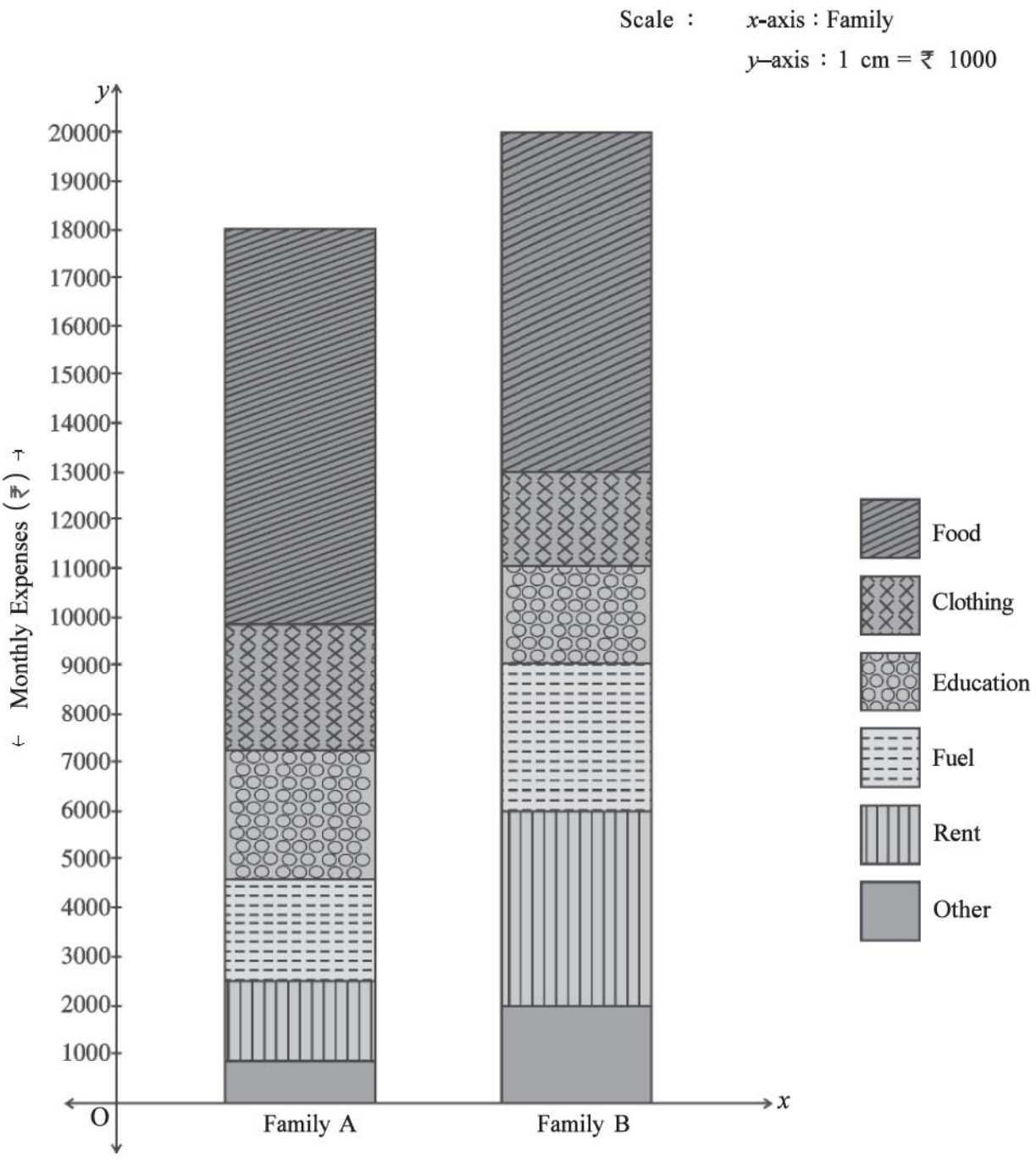
Monthly expense (₹)	Food	Clothing	Education	Fuel	Rent	Other	Total
Family A	8100	2700	2880	1800	1620	900	18,000
Family B	7000	2000	2000	3000	4000	2000	20,000

Information regarding the monthly expenses on different particulars of livelihood is given for two families. So, simple divided bar diagrams is to be drawn. By taking families on x – axis and expenses on y – axis, with appropriate scale, simple divided bar diagram is as under :

A bar of total expense is drawn on graph paper and then the different sub-data are separated by the lines according to the calculations shown in the following table.

	Family A		**Family B**	
Particulars	**Expense**	**Split-up line**	**Expense**	**Split-up line**
Food	8100	18000 – 8100 = 9900	7000	20000 – 7000 = 13000
Clothing	2700	9900 – 2700 = 7200	2000	13000 – 2000 = 11000
Education	2880	7200 – 2880 = 4320	2000	11000 – 2000 = 9000
Fuel	1800	4320 – 1800 = 2520	3000	9000 – 3000 = 6000
Rent	1620	2520 – 1620 = 900	4000	6000 – 4000 = 2000
Other	900	–	2000	–
Total	**18,000**	–	**20,000**	–

Simple divided bar diagram showing monthly expenses of two families



2.4.2.4 Percentage divided Bar Diagram

In simple divided bar diagram, sub-data are presented in an attractive manner. But the mutually related sub-data cannot be effectively compared. To overcome this limitation, percentage divided bar diagram is used. The total value is considered as 100% and percentages of sub-data calculated on the basis of it are presented in divided bar diagram.

Illustration 28 : To compare the data on expenses of the families given in illustration 27, prepare an appropriate diagram.

Here we have to compare the sub-data on the monthly expenses of two families so we draw percentage divided bar diagram.

Particulars	Family A			Family B		
	Expense	Percentage	Split-up	Expense	Percentage	Split-up
Food	8100	45	100 – 45 = 55	7000	35	100 – 35 = 65
Clothing	2700	15	55 – 15 = 40	2000	10	65 – 10 = 55
Education	2880	16	40 – 16 = 24	2000	10	55 – 10 = 45
Fuel	1800	10	24 – 10 = 14	3000	15	45 – 15 = 30
Rent	1620	9	14 – 9 = 5	4000	20	30 – 20 = 10
Other	900	5	–	2000	10	–
Total	**18,000**	**100**	–	**20,000**	**100**	–

Percentage divided bar diagram showing monthly expenses of two families

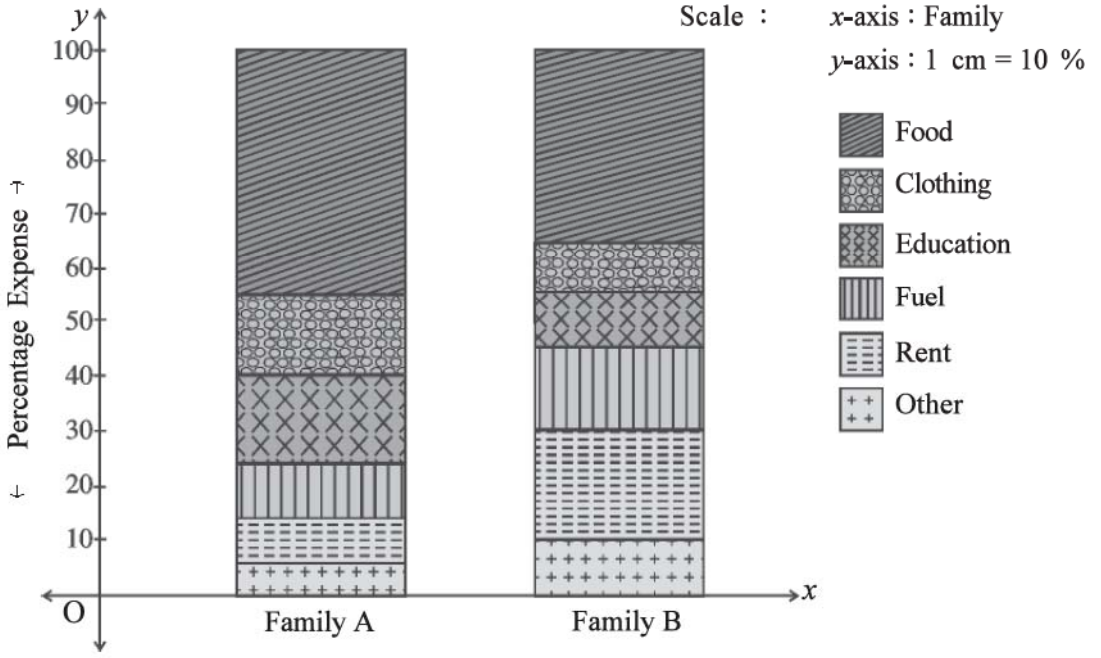


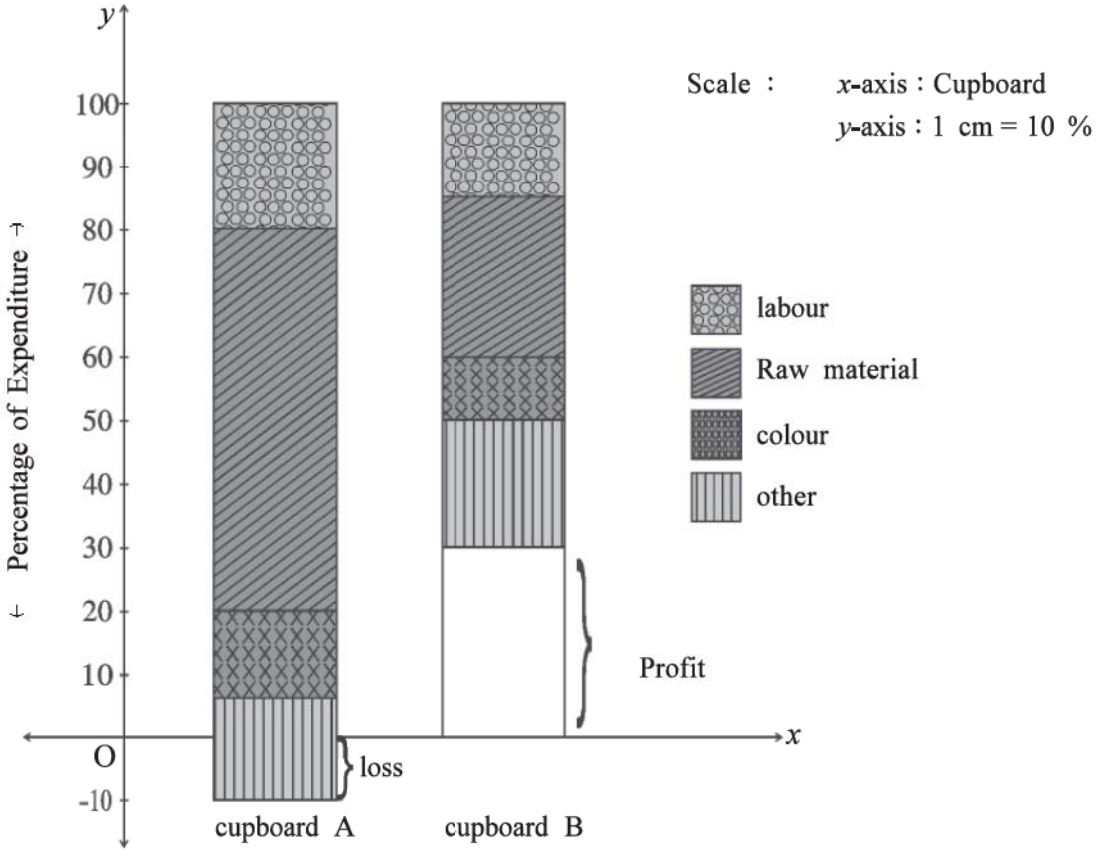
Illustration 29 : A factory produces and sells two types of cupboards, A and B. Cupboard A is not much attractive but is sturdy and hence it has more usage in industrial units. Whereas, cupboard B is good-looking, hence it is used for household purpose. The selling price of cupboard A is ₹ 5000 and that of cupboard B is ₹ 8000 per piece. The particulars about the making expenses are as follows. Draw a percentage divided bar diagram showing particulars of expenses of cupboards and the profit/loss by their sales.

Particulars of expenses	Cupboard A	Cupboard B
Labour	1000	1200
Raw material	3000	2000
Colours	700	800
Other	800	1600
Total	**5500**	**5600**

As we have to draw a diagram showing profit/loss by selling these cupboards, considering selling price of cupboard as 100 %, the percentage of the expenses are determined and are shown in percentage divided bar diagram.

Particulars	Cupboard A			Cupboard B		
	Expense	Percentage	Split-up line	Expense	Percentage	Split-up line
Labour	1000	20	100 – 20 = 80	1200	15	100 – 15 = 85
Raw material	3000	60	80 – 60 = 20	2000	25	85 – 25 = 60
Colours	700	14	20 – 14 = 6	800	10	60 – 10 = 50
Other	800	16	6 – 16 = –10	1600	20	50 – 20 = 30
Production cost	5500	110	–	5600	70	–
Selling price	5000	100	–	8000	100	–
Profit/loss	**–500**	**–10**	–	**2400**	**30**	–

A percentage divided diagram showing profit/loss by selling cupboards



2.4.3 Two Dimensional Diagram

One dimensional diagram is used to represent a single characteristic of the data. In such diagrams, either height or breadth is taken into consideration. But when the volume of the data is large, then in order to present it on a diagram, both length and breadth are taken into consideration.Thus total value is shown as an area in the diagram for which the diagrams like square, rectangle, circle or pie (sectorial) diagram are included. We shall study only circle diagrams and sectorial or pie diagrams.

2.4.3.1 Circle Diagram

When the volume of data regarding two or more places, things or time is large, then circular diagrams are used to present the data. The volume of the data is represented by area of a circle.

Total volume of data = Area of circle = πr^2

Where $\pi = 3.14$ or $\frac{22}{7}$, r = radius

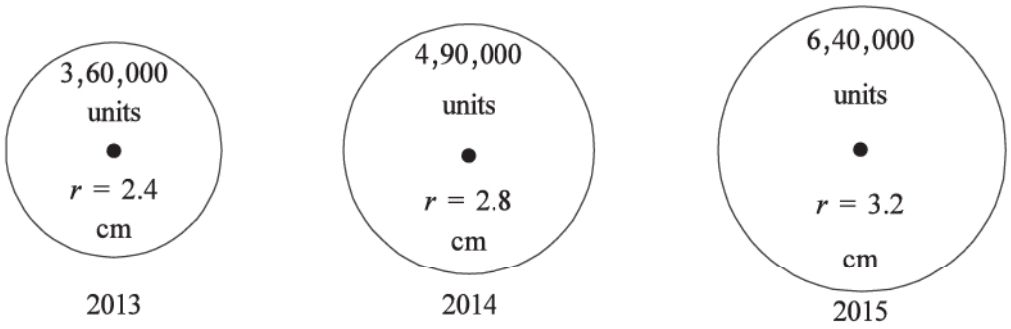
Area of a circle is proportional to the square of its radius. Therefore, in circle diagram, square roots of the volumes of different data are taken as the radii of the circle. The radii thus obtained are arranged in ascending or descending order and are drawn with centers on same line at equal distance from each other. If the data are given with respect to time then circles are drawn in order of time only. If the radius of circle is too large then it is divided by a constant and if it is too small then it is multiplied by a constant and then circle diagram is drawn.

Illustration 30 : The data on the production of an industrial unit during three years is as under. Present it by using circle diagram.

Year	Production (in units)
2013	3,60,000
2014	4,90,000
2015	6,40,000

As the given data are numerically large, radius is taken as the square root of each value which is further divided by 250.

Year	Production (Units)	Square root	Radius = square root/250
2013	3,60,000	600	2.4
2014	4,90,000	700	2.8
2015	6,40,000	800	3.2



2.4.3.2 Pie–diagram

If the data related to different places, things or times consist of several mutually related sub-data on different components are numerically large, then instead of divided bar diagram, pie diagram is used. The total volume of data is represented by a circle of suitable radius and this circle is then divided into sectors to present the sub-data. In the diagram, total volume of the data is taken as 360° and the volumes of sub-data are expressed in terms of measures of angle and are presented on circle with respective circular sectors.

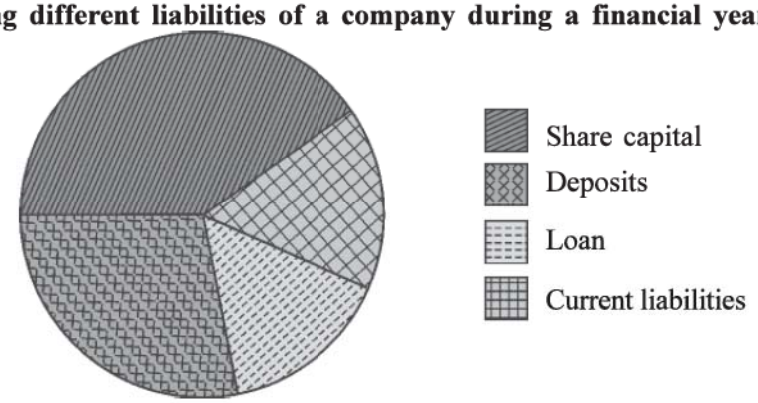
Illustration 31 : In a ledger of a company, the liabilities in the final balance sheet are as under. Present the data by a pie diagram.

Particulars	Share capital	Deposits	Loan	Current liabilities	Total
Liabilities (₹)	12,00,000	8,00,000	4,00,000	4,80,000	28,80,000

Measure of angle in degrees for the liabilities of different particulars can be calculated as under :

Particulars	Liabilities (₹)	Angle (degree)
Share capital	12,00,000	$\frac{1200000}{2880000} \times 360° = 150°$
Deposits	8,00,000	$\frac{800000}{2880000} \times 360° = 100°$
Loan	4,00,000	$\frac{400000}{2880000} \times 360° = 50°$
Current liabilities	4,80,000	$\frac{480000}{2880000} \times 360° = 60°$
Total	28,80,000	360°

Pie diagram showing different liabilities of a company during a financial year



Activity

For your family, collect information regarding monthly expenses on food, education, fuel and miscellaneous and represent it through a pie diagram.

Illustration 32 : Information regarding the annual expense on different particulars of two middle class families is given below. Draw a pie diagram for the data.

Particulars	Expenditure (₹)	
	Family A	Family B
Food	40000	50,000
Clothing	10000	20,000
Rent	25000	30,000
Education	10000	32,000
Miscellaneous	5000	28,000
Total	**90,000**	**1,60,000**

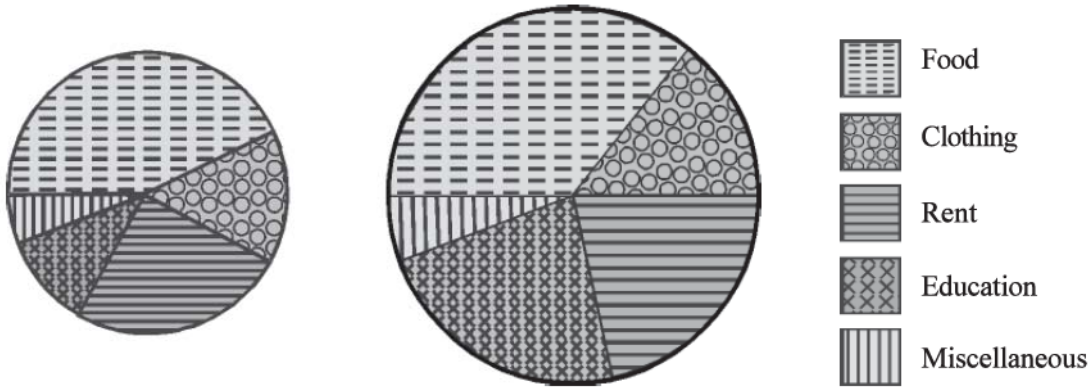
We shall draw two circles for two families with radii proportional to the square root of the total expenditure of two families. By taking total expenditure as 360° the measures of angles of sub-data are obtained in terms of degrees.

Radius for family A $= \frac{\sqrt{90000}}{100} = \frac{300}{100} = 3$ cm

Radius for family B $= \frac{\sqrt{160000}}{100} = \frac{400}{100} = 4$ cm

	Expenditure (₹)			
	Family A		**Family B**	
Particulars	**Expense**	**Degree**	**Expense**	**Degree**
Food	40,000	$\frac{40000}{90000} \times 360° = 160°$	50,000	$\frac{50000}{160000} \times 360° = 112.5°$
Clothing	10,000	$= 40°$	20,000	$= 45°$
Rent	25,000	$= 100°$	30,000	$= 67.5°$
Education	10,000	$= 40°$	32,000	$= 72°$
Miscellaneous	5000	$= 20°$	28,000	$= 63°$
Total	**90,000**	**$= 360°$**	**1,60,000**	**$= 360°$**

Pie diagram showing annaual expenditure of two families



Family A Family B

Activity

Represent the above information in percentage divided bar diagram.

2.4.4 Pictogram

A diagram in which the data are represented by appropriate pictures is called pictogram. The pictures are selected with reference to the data. For example, if the data related to the population are given then it can be shown by symbols of human. The pictures are drawn in proportion to the amount of the given data. Pictogram draws quick attention of the viewer. If the data are presented by pictograms then it can be easily understood by the less educated people and by the children. Also, pictogram has no barrier of language. The main disadvantage of this method is that it has limited usage in statistical analysis of data.

Illustration 33 : Information regarding the population of five cities is given below. Present it by a pictogram.

City	A	B	C	D	E
Population	20,000	40,000	60,000	80,000	90,000

Here, one figure of human = 20,000

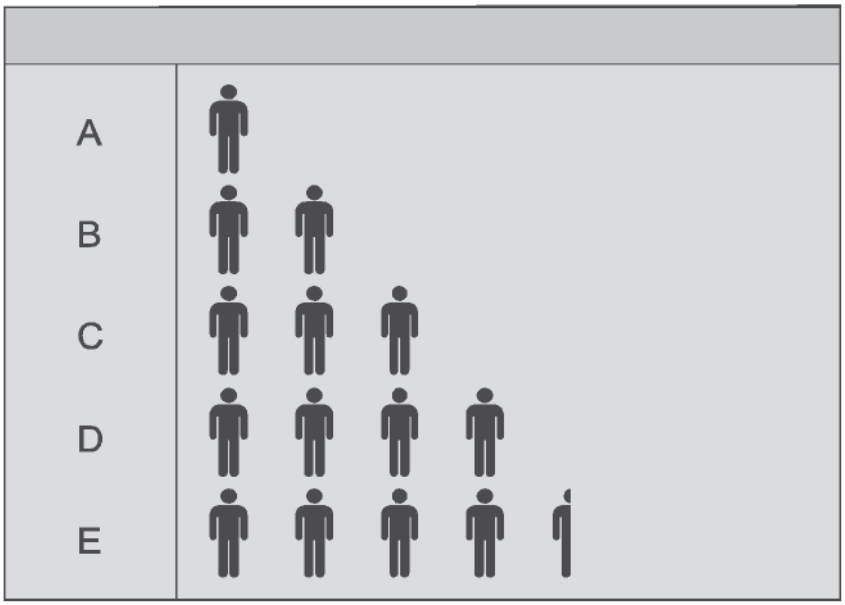
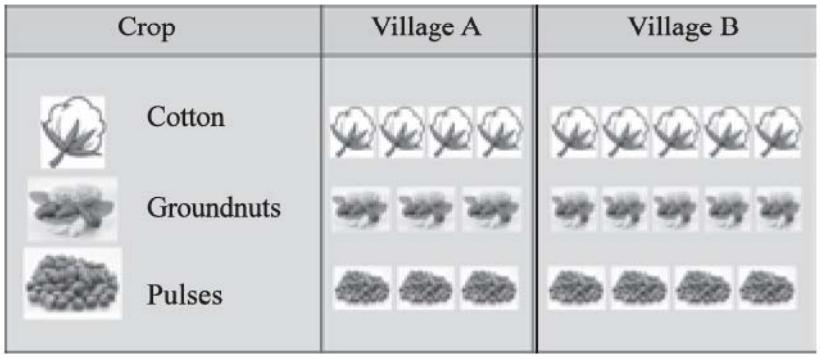


Illustration 34 : The data regarding area under cultivation of various crops at farms of two villages is given below. Present it through pictogram.

Crop	Area under Cultivation (hector)	
	Village A	Village B
Cotton	400	500
Groundnut	300	500
Pulses	300	400

By considering the symbol of the respective crop, the pictogram drawn is as under :

Pictogram showing area under cultivation in two villages



Here, one figure of crop = 100 acres of area under cultivation.

EXERCISE 2.3

1. Following is the data on the number of employees working in various Government departments. Present it with a suitable diagram.

Department	**Road transport**	**Railway**	**Income tax**	**Finance**	**Planning**
Number of employees	4000	6000	3000	2500	1500

2. The data on the profit of a company is as under. Present it by an appropriate diagram :

Year	2010	2011	2012	2013	2014	2015
Profit (crore ₹)	10	5	–2	4	8	6

3. Changes in the price of share of 5 companies at Bombay Stock Exchange during an interval of 15 days are as follows. Present them by a suitable diagram :

Company	A	B	C	D	E
Price of share (₹)	40	20	100	80	30
Price of same share after 15 days (₹)	60	30	150	60	10

4. Information regarding the birth rate and death rate of 5 countries is as under. Present it by a suitable diagram.

Country	**U.S.A.**	**Japan**	**Bharat**	**Germany**	**U.K.**
Birth rate	16.5	20.8	34.2	16.4	15.2
Death rate	10.2	12.2	20.4	10.3	12.0

5. Information regarding the age of persons living in two different regions is as under. Present it by using an appopriate diagram :

Age	Less than 15 (Child)	15 to 35 (Young)	35 to 60 (Adult)	More than 60 (Old)	Total
Region A	480	360	240	120	1200
Region B	350	250	200	200	1000

6. Draw the percentage divided bar diagram for the data given in example 5.

7. A car production company has produced the following number of cars during three years. Represent it through a circle diagram.

Year	2012	2013	2014
Production of Cars	25,600	1,02,400	1,60,000

8. The following data represent the percentage sales of copies of daily newspapers. Represent it by pie diagram.

News paper	P	Q	R	S	Total
Percentage of selling	25	23	24	28	100

9. Represent the following information by a pictogram :

Year	2010	2011	2012	2013	2014
Production of mangoes (kg.)	1,00,000	1,50,000	2,50,000	1,50,000	75,000

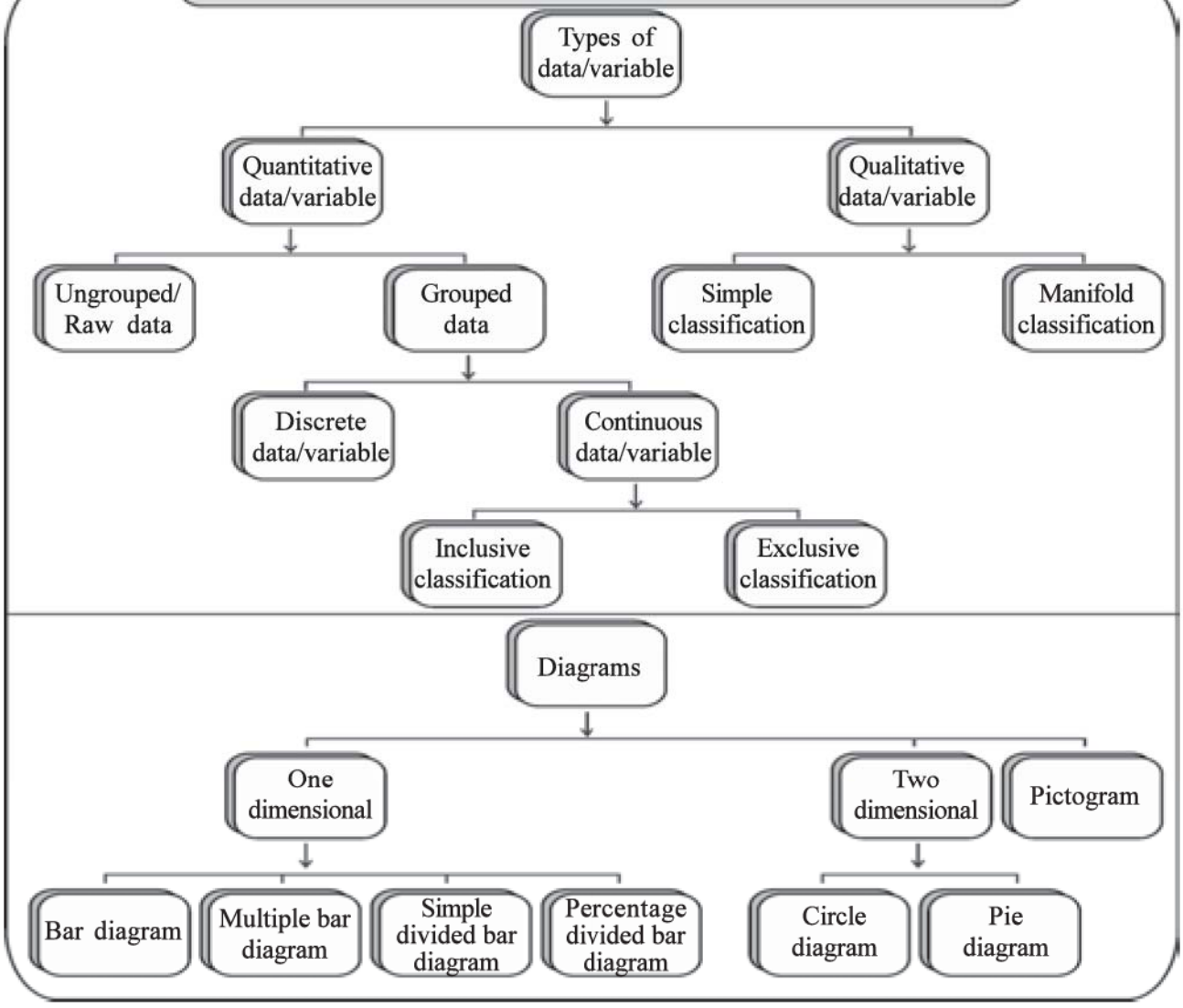
10. Details of production of electric bulb by two well-known companies are as under. Represent them by a pictogram.

Year	**Production of electric bulbs (lakh units)**	
	Company A	**Company B**
2012	50	100
2013	100	150
2014	175	200
2015	200	200

Summary

- A variable assuming definite value between two specified limits is called discrete variable.
- A variable assuming any value between two specified limits is called continuous variable.
- A process of arranging raw data into short and systemic manner is called classification.
- A numeric value indicating the repetition of the value of observation is called frequency for that observation.
- A table showing the different values of discrete variable with the respective frequencies is called discrete frequency distribution.
- When the values of observations of raw data are classified in different classes then the table is called continuous frequency distribution.
- When the range of discrete variable is large then the data are expressed by inclusive continuous frequency distribution.
- Sum of the frequencies upto the value of observation or class is called cumulative frequency and its distribution is called cumulative frequency distribution.
- While preparing cumulative frequency distribution, inclusive classes should be converted into exclusive classes.
- In exclusive classes, class limits and class boundary points are identical.
- When the range of raw data is too large then a frequency distribution with unequal class length is to be obtained.
- Diagrams are used to present classified data is simple and attractive form.
- In one dimensional diagram, the diagram is drawn by considering only one characteristic of the data.
- To express one characteristic of a data, bar diagram is used, whereas for representing more than one characteristics, multiple bar diagram is used.
- To represent various sub-data of the classified data, divided bar diagram is used and to compare them, percentage divided bar diagram is used.
- When numerically large data is given then pie diagram is used to represent it. It is also used to compare large data.
- Pictorial representation of data is called pictogram.

CHAPTER AT GLANCE



Formulae :

(1) Range of data R = Maximum value – Minimum value

(2) Class length $C = \dfrac{\text{Range}}{\text{No. of classes}}$

(3) Lower boundary point = $\dfrac{\text{lower limit of that class + upper limit of previous class}}{2}$

Upper boundary point = $\dfrac{\text{upper limit of that class + lower limit of succeeding class}}{2}$

(4) Mid point or mid value of a class = $\dfrac{\text{value of upper limit + value of lower limit}}{2}$

(5) $\textit{Lower boundary point} = \textit{mid value} - \frac{1}{2}\ (\textit{Class length})$

$\textit{Upper boundary point} = \textit{mid value} + \frac{1}{2}\ (\textit{Class length})$

EXERCISE 2

Section A

Choose the correct option for the following multiple choice questions :

1. Which of the following variables is discrete ?

(a) Height of a person (b) Weight of a commodity

(c) Area of a ground (d) Number of children per family.

2. Which of the following variables is continuous ?

(a) Number of errors per page of a book (b) Number of cars produced

(c) Number of accidents on road (d) Monthly income of a person

3. Name the method of classification of raw data related to daily demand of a product.

(a) Classification of attribute data (b) Classification of numeric data

(c) Raw distribution (d) Manifold classification

4. Name the type of classification of the data related to the occupation and education of a person living in a certain region.

(a) Tabulation (b) Classification of numeric data

(c) Raw distribution (d) Discrete frequency distribution

5. In continuous frequency distribution, what is the class length of a class ?

(a) Average of two successive lower boundary points.

(b) Average of class limits

(c) Difference between upper boundary point and lower boundary point of that class.

(d) Average of upper boundary point and lower boundary point of the class.

6. Range of an ungrouped data is 55 and it is divided into 6 classes. Then what is the class length ?

(a) 10 (b) 9 (c) 9.17 (d) 10.17

7. Inclusive classes for a distribution are 10-19.5, 20-29.5, 30-39.5. What are the exclusive class limits for the second class ?

(a) 19.5 - 29.5 (b) 19.75 - 29.75 (c) 20 - 30 (d) 19 - 29

8. A discrete variable has values 0, 1, 2, 3, 4 with the respective frequency 2, 4, 6, 8, 14. What is the value of 'more than' type cumulative frequency when the value of variable is 2 ?

(a) 28 (b) 12 (c) 34 (d) 6

9. A continuous distribution has classes 0 – 9, 10 – 19, 20 – 29, 30 – 39 with the respective frequencies 10, 20, 40, 10. What is the 'less than' type cumulative frequency for the boundary point 29.5 ?

(a) 30 (b) 50 (c) 70 (d) 80

10. For a continuous variable, classes are 1 - 1.95, 2 - 2.95, 3 - 3.95, 4 - 4.95, 5 - 5.95 then what is the lower boundary point of the second class ?

(a) 1.995 (b) 2 (c) 2.975 (d) 1.975

11. Which of the following statement is/are true ?

Statement 1: A method of representing the large and complex data in simple and attractive manner is called diagram.

Statement 2: Self-explanatory representation of main characteristics of the data is called diagram.

Statement 3: Representation of comparative study of data is called diagram.

(a) Only statement 1 is true. (b) Only statements 1 and 2 are true.

(c) Statements 1,2 and 3 are true. (d) All three statements are false.

12. The class intervals for a continuous variable are 0 - 99, 100 - 199, 200 - 299, 300 - 399, 400 - 499. What is the mid value of the second class ?

(a) 149.5 (b) 150 (c) 199.5 (d) 99.5

13. What do we call a table that shows designation, gender and marital status of employees of a company ?

(a) Simple classification (b) Classification of numeric data

(c) Manifold classification (d) Simple table

14. Which of the following diagrams is used to represent sub-data of classified information ?

(a) Bar diagram (b) Divided bar diagram

(c) Multiple bar diagram (d) Pictogram

15. Which of the following diagrams is used for comparing the sub-data of the classified data ?

(a) Pictogram (b) Pie chart

(c) Bar diagram (d) Divided bar diagram

Section B

Give answer in one sentence for the following question :

1. Define discrete variable.

2. Define continuous variable.

3. What is classification ?

4. State the types of classification.

5. Define the frequency of an observation.

6. State the method to determine number of classes on the basis of range of data and class length.

7. When should one from a frequency distribution with unequal class lengths ?

8. Define cumulative frequency.

9. Define 'less than' type cumulative frequency distribution for discrete data.

10. Define 'more than' type cumulative frequency distribution for continuous data.

11. Write a formula for finding mid value of a class.

12. Define tabulation.

13. Define manifold classification.

14. What is the characteristic of the best table to represent qualitative data ?

15. What is the main disadvantage of classification of data ?
16. In statistical study, what is the main objective of a diagram ?
17. State the types of diagrams.
18. For which type of data, multiple bar diagram is drawn ?
19. When do we draw divided bar diagram ?
20. State the main objective of percentage divided bar diagram.

Section C

Answer the following questions :

1. Define quantitative and qualitative data.
2. Define discrete frequency distribution with illustration.
3. Define continuous frequency distribution with illustration.
4. Explain the definition of inclusive continuous frequency distribution.
5. Explain the definition of exclusive continuous frequency distribution.
6. Write formulae for obtaining class boundary points from inclusive class limits.
7. Find mid values of each class of the following frequency distribution.

Class	0 - 9	10 - 24	25 - 49	50 - 74	75 - 100
Frequency	10	20	30	20	10

8. For the frequency distribution given in the above problem, find the class length of each class.
9. Prepare 'less than' type cumulative frequency distribution from the following.

Observation	10	20	30	40	50
Frequency	10	30	30	20	10

10. Demand of a certain item is classified as good, moderate and weak. On the basis of a study for entire year, it is known that the demand was moderate during 22 weeks, whereas the demand was weak during 18 weeks. Present this information in a table.
11. Complete the following table :

	Attribute A			**Attribute B**			**Total**		
Year	**Sub-data-1**	**Sub-data-2**	**Total**	**Sub-data-1**	**Sub-data-2**	**Total**	**Sub-data-1**	**Sub-data-2**	**Total**
2014	200		300	100		200			
2015		400		150	300		300		

12. Differentiate between inclusive and exclusive continuous frequency distribution.
13. State the limitations of diagram.
14. What are one dimensional diagrams? State their names.
15. Explain two dimensional diagrams in brief.

16. Represent the following data through a bar diagram :

(1)

Year	2011	2012	2013	2014	2015
Production (in crore ₹)	3.5	4.2	5.8	7.4	10.2

(2)

Stream	**Arts**	**Commerce**	**Science**	**Engineering**	**Other**
Number of students	5900	10,200	6000	4500	8000

Section D

Solve the following :

1. What is the necessity of classification in statistical study ?

2. Explain the classification of numeric data with an appropriate illustration.

3. Explain the classification of qualitative data with a suitable illustration.

4. Write a short note on 'cumulative frequency distribution'.

5. Discuss the points for constructing continuous frequency distribution.

6. State the guiding rules for the construction of a table.

7. State the uses of tabulation.

8. Obtain the original frequency distribution from the following data.

Mid value	250	350	450	550	650	750	Total
Frequency	20	80	80	40	60	20	300

9. Out of 40 persons working in an office, 60% are females and remaining 40 % are males. 50 % of males are married, where as the ratio of married and unmarried females is 5:3. Present this information in a table.

10. Information regarding the monthly income of 100 workers is given below. Obtain original frequency distribution from it.

Less than Monthly income	2400	2900	3400	3900	4400	4900	5400	5900	6400
Number of worker	0	3	12	30	55	78	88	95	100

11. Marks of 200 students in an examination are as under. Obtain the original frequency distribution.

Marks	10 - 100	20 - 100	30 - 100	40 - 100	50 - 100	60 - 100	70 - 100	80 - 100	90 - 100
No. of Students	200	180	140	90	55	30	8	2	1

12. From the data given below, obtain original frequency distribution.

Mid value	12.5	17.5	22.5	27.5	32.5	37.5	42.5	47.5
Frequency	12	18	16	22	14	10	6	2

13. There are 1000 buses used for the public transport in Ahmedabad city. Of them, 350 are used as BRTS and remaining as AMTS. Out of total 400 air conditioned buses, 250 were used as BRTS. Present this information in a suitable table.

14. Out of 1500 students of a college, 900 were boys and of them, 250 were in science stream. 250 girls were in commerce stream. Present these data in an appropriate table.
15. Explain the importance of diagrams in statistical study.
16. Write a short note on one dimensional diagrams.
17. Write a short note on two dimensional diagrams.
18. Explain pictogram with an illustration.
19. The agricultural production index numbers for two different states are as under. Present them by using suitable diagram.

Year	2011	2012	2013	2014	2015
State A	139	147	152	162	170
State B	110	115	125	140	150

20. Area (in sq.mt.) of 5 different regions is as under. Draw a pie diagram.

Region	A	B	C	D	E
Area	5	8	29	44	71

21. Production of a commodity in three different factories is as under. Present it through suitable diagram.

Factory	P	Q	R
Production (lakh ₹)	256	576	1024

Section E

Solve the following :

1. Number of mangoes received from different trees of mangoes in a farm during a season of 30 days is as under. Prepare a frequency distribution by taking class length 5.

94	96	100	104	122	107	108	106	119	120
98	123	102	125	95	125	115	104	114	109
128	112	103	92	114	101	113	118	124	118

2. The data regarding the earnings (₹) of 40 rickshaw drivers during a certain day are as follows. Prepare a frequency distribution having one class as 220 - 239 and class length 20.

285	215	200	225	255	250	235	242	298	312
328	294	266	335	330	270	315	275	245	265
210	235	275	305	332	355	307	230	348	350
310	290	264	228	236	336	356	322	215	345

3. Information on monthly water consumption (in units) of 50 residents of a region is as under. By taking one of the classes as 25 - 30, prepare exclusive continuous frequency distribution.

24	34	41	55	45	25	40	38	40	44
28	35	40	48	35	44	27	57	42	30
28	26	42	49	47	33	52	52	28	34
36	30	44	33	31	30	39	25	24	47
28	36	32	57	25	29	35	44	50	56

4. The data obtained by inquiring price of an item at 50 different shops are as under. Prepare a frequency distribution having the last class 85-90.

82	75	73	70	84	79	79	77	80	66
70	70	72	62	64	80	85	64	75	65
66	75	71	82	69	70	72	80	66	70
79	69	80	63	66	75	68	78	86	66
85	66	69	85	70	60	70	75	79	86

5. Obtain 'less than' type and 'more than' type cumulative frequency distribution from the following frequency distribution, .

Class	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59	**Total**
Frequency	3	8	10	5	15	8	1	50

6. The following data refer to the daily absence of workers in a factory during 30 days. Prepare an appropriate frequency distribution and hence obtain 'less than' type cumulative frequency distribution.

0	1	4	5	4	0	0	2	3	4	1	2	6	4	0
3	2	3	2	1	1	0	2	1	1	3	3	5	1	3

7. There were 850 students studying in higher standards of a school. The number of students in standard 10, 11 and 12 were in the proportion 8:5:4. In standard 10, the number of boys is 30% of the number of students in the school. In standard 11, the numbers of boys and girls are equal. In standard 12, the number of boys is three times the number of girls. Present the above data in a tabular form.

8. In the year 2013, there were 1200 students studying in a school and of them, 400 were girls. 50 girls were not residing in hostel. In all 600 boys were residing in hostel. In the year 2014, there is an increase of 20% in the number of boys and the number of girls increased by 30%. During this year, 260 boys and 100 girls were not residing in hostel. In the year 2015, 140 boys and 100 girls were newly admitted in the school and all of them resided with the hostel students. Present above data in a tabular form.

9. Present the following data in an appropriate tabular form. A bank receives 2000 applications as a response to the job advertisement. Of the total applicants, 50% were graduates, 40% were post graduates and remaining 10 % have professional degree. Among the graduates, 60% were males and of them, 25% were married. 40% female graduates were married. Among the post graduates, 60% were males and 40% of them were married. Among post graduate females, 50% were married. 30% of the females had professional degree and of them, 60% were married. The number of married and unmarried males having professional degree was equal.

10. The folowing table represents the number of workers of a factory according to their gender, residence and year :

	Local			**Non-local**			**Total**		
Year	**Men**	**Women**	**Total**	**Men**	**Women**	**total**	**Men**	**Women**	**Total**
2010	1200	300	1500	300	200	500	1500	500	2000
2015	2000	600	2600	300	100	400	2300	700	3000

Answer the following questions using the above table :

(1) What is the percentage increase in the total number of workers during the period of five years?

(2) Find the percentage decline in the number of non-local workers in the year 2015.

(3) Find the percentage increase in the number of men and women during the period of 5 years.

11. A mobile phone manufacturing company produces and sells two types of mobile phones. The particulars about it are given in the following table. Present it by a suitable diagram :

Particulars	Mobile A	Mobile B
Raw material	5000	6000
Assembly expense	3000	3000
Other expense	4000	4500
Total expense	12,000	13,500
Selling price	13,000	15,000

12. Information regarding the average monthly expenses (in ₹) of two families is as under. Present it through a pie diagram.

Particulars	Family A	Family B
Food	20,000	16,000
Fuel	5000	4000
Transportation	10,000	8800
House rent	15,000	18,000
Other	22,000	18,000

Section F

Solve the following :

1. A sample of 25 lenses is selected from a day's production of a company manufacturing eye lenses. The thicknesses (in millimeter) of these selected lenses are as under. Distribute these data into five classes of equal length.

1.518 1.509 1.527 1.505 1.520 1.511 1.518 1.522 1.528 1.528 1.520
1.520 1.514 1.508 1.525 1.506 1.519 1.523 1.521 1.517 1.514 1.515
1.516 1.521 1.507

If the company decides that the lenses having thicknesses less than 1.510 and more than 1.525 are considered as defective then what percent of lenses in the sample are defective ?

2. The data related to variations in the price of a share for 30 days in a share market are as under. Prepare an exclusive continuous classification having class limits of one of the classes as 18.5 - 20.5.

10.50	14.70	17.20	15.20	14.50	19.20	15.80	19.30
18.40	20.50	18.70	14.90	18.50	16.90	10.50	12.50
13.60	12.50	18.50	18.60	14.00	16.20	13.30	13.30
18.60	17.60	20.20	14.50	20.80	14.90		

On the basis of this frequency distribution, answer the following questions.

(1) What is mid value of the 4th class ?

(2) Find the number of days during which the price of share is less than ₹ 16.50.

(3) Find the number of days during which the price of share is at least ₹ 19.50.

3. Owner of a factory has decided to produce 50 mixers used as household equipment, but the daily production of mixers changes due to variation in the number of workers. A variation in production of mixers with respect to a pre-decided number of production (100 units) during 40 days is recorded as under. Prepare an exclusive continuous frequency distribution having class length 6 and mid value of one of the classes as 3. Also prepare 'less than' and 'more than' cumulative frequency distributions.

7	6	12	16	12	18	11	–5	10	3	10	7	8	
14	–10	16	–7	20	9	12	–2	0	5	–4	23	6	
–3	4	4	3	4	2	0	22	1	5	–1	5	19	6

4. The data regarding the height (in cm.) of 30 students of a school are as under. Prepare an inclusive continuous frequency distribution of 6 classes and hence prepare 'less than' and 'more than' cumulative frequency distributions.

141	145	152	150	150	159	148	163	162	151	155	148
145	162	161	152	168	153	149	148	162	158	157	160
153	149	154	165	141	149						

On the basis of it, answer the following questions :

(1) If participation in the N.C.C. activities requires a minimum height of 160 cm. then how many students are eligible to participate ?

(2) Find the number of students having height from 153 cm. to 163 cm.

(3) Find maximum height of one third of the students having minimum height.

5. The students of a university were classified according to faculty and gender. 60% of total 40,000 students were boys. The number of girls in engineering faculty was three times the number of girls in commerce faculty. 15% and 10% of the total number of university students were boys and girls respectively who belonged to medical faculty. 20% of the total number of students in the university belonged to faculty of science and among these students, the number of girls were one-seventh of the number of boys. 7% and 17% of the total number of students of arts faculty were boys and girls respectively. 3.75% of the total number of students of the university belonged to the commerce faculty and the proportion of boys and girls among them was 3:7.

Present the above data in an appropriate table.

Prof. C. R. Rao
(1920)

Prof. C.R. Rao is an Indian born, naturalized American, mathematician and statistician. He is currently professor emeritus at Penn State University and Research Professor at the University of Buffalo. Rao has been honoured by numerous colloquia, honorary degrees and festschrifts and was awarded the US National Medal of Science in 2002. The American Statistical Association has described him as "a living legend whose work has influenced not just statistics, but has had far reaching implications for fields as varied as economics, genetics, anthropology, geology, national planning, demography, biometry and medicine." The Times of India listed Rao as one of the top 10 Indian scientists of all time.

●