1062CH14

# STATISTICS 14

## 14.1 Introduction

In Class IX, you have studied the classification of given data into ungrouped as well as grouped frequency distributions. You have also learnt to represent the data pictorially in the form of various graphs such as bar graphs, histograms (including those of varying widths) and frequency polygons. In fact, you went a step further by studying certain numerical representatives of the ungrouped data, also called measures of central tendency, namely, *mean*, *median* and *mode*. In this chapter, we shall extend the study of these three measures, i.e., mean, median and mode from ungrouped data to that of *grouped data*. We shall also discuss the concept of cumulative frequency, the cumulative frequency distribution and how to draw cumulative frequency curves, called *ogives*.

## 14.2 Mean of Grouped Data

The mean (or average) of observations, as we know, is the sum of the values of all the observations divided by the total number of observations. From Class IX, recall that if $x_1, x_2, \ldots, x_n$ are observations with respective frequencies $f_1, f_2, \ldots, f_n$, then this means observation $x_1$ occurs $f_1$ times, $x_2$ occurs $f_2$ times, and so on.

Now, the sum of the values of all the observations $= f_1x_1 + f_2x_2 + \ldots + f_nx_n$, and the number of observations $= f_1 + f_2 + \ldots + f_n$.

So, the mean $\bar{x}$ of the data is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n}$$

Recall that we can write this in short form by using the Greek letter Σ (capital sigma) which means summation. That is,

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

which, more briefly, is written as $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$, if it is understood that $i$ varies from 1 to $n$.

Let us apply this formula to find the mean in the following example.

**Example 1 :** The marks obtained by 30 students of Class X of a certain school in a Mathematics paper consisting of 100 marks are presented in table below. Find the mean of the marks obtained by the students.

| **Marks obtained ($x_i$)** | 10 | 20 | 36 | 40 | 50 | 56 | 60 | 70 | 72 | 80 | 88 | 92 | 95 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Number of students ($f_i$)** | 1 | 1 | 3 | 4 | 3 | 2 | 4 | 4 | 1 | 1 | 2 | 3 | 1 |

**Solution:** Recall that to find the mean marks, we require the product of each $x_i$ with the corresponding frequency $f_i$. So, let us put them in a column as shown in Table 14.1.

**Table 14.1**

| **Marks obtained ($x_i$)** | **Number of students ($f_i$)** | $f_i x_i$ |
|---|---|---|
| 10 | 1 | 10 |
| 20 | 1 | 20 |
| 36 | 3 | 108 |
| 40 | 4 | 160 |
| 50 | 3 | 150 |
| 56 | 2 | 112 |
| 60 | 4 | 240 |
| 70 | 4 | 280 |
| 72 | 1 | 72 |
| 80 | 1 | 80 |
| 88 | 2 | 176 |
| 92 | 3 | 276 |
| 95 | 1 | 95 |
| **Total** | $\Sigma f_i = 30$ | $\Sigma f_i x_i = 1779$ |

Now, $$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1779}{30} = 59.3$$

Therefore, the mean marks obtained is 59.3.

In most of our real life situations, data is usually so large that to make a meaningful study it needs to be condensed as grouped data. So, we need to convert given ungrouped data into grouped data and devise some method to find its mean.

Let us convert the ungrouped data of Example 1 into grouped data by forming class-intervals of width, say 15. Remember that, while allocating frequencies to each class-interval, students falling in any upper class-limit would be considered in the next class, e.g., 4 students who have obtained 40 marks would be considered in the class-interval 40-55 and not in 25-40. With this convention in our mind, let us form a grouped frequency distribution table (see Table 14.2).

**Table 14.2**

| **Class interval** | 10 - 25 | 25 - 40 | 40 - 55 | 55 - 70 | 70 - 85 | 85 - 100 |
|---|---|---|---|---|---|---|
| **Number of students** | 2 | 3 | 7 | 6 | 6 | 6 |

Now, for each class-interval, we require a point which would serve as the representative of the whole class. *It is assumed that the frequency of each class-interval is centred around its mid-point*. So the *mid-point* (or *class mark*) of each class can be chosen to represent the observations falling in the class. Recall that we find the mid-point of a class (or its class mark) by finding the average of its upper and lower limits. That is,

$$\textbf{Class mark} = \frac{\textbf{Upper class limit + Lower class limit}}{\textbf{2}}$$

With reference to Table 14.2, for the class 10-25, the class mark is $\frac{10+25}{2}$, i.e., 17.5. Similarly, we can find the class marks of the remaining class intervals. We put them in Table 14.3. These class marks serve as our $x_i$'s. Now, in general, for the $i$th class interval, we have the frequency $f_i$ corresponding to the class mark $x_i$. We can now proceed to compute the mean in the same manner as in Example 1.

**Table 14.3**

| Class interval | Number of students ($f_i$) | Class mark ($x_i$) | $f_i x_i$ |
|---|---|---|---|
| 10 - 25 | 2 | 17.5 | 35.0 |
| 25 - 40 | 3 | 32.5 | 97.5 |
| 40 - 55 | 7 | 47.5 | 332.5 |
| 55 - 70 | 6 | 62.5 | 375.0 |
| 70 - 85 | 6 | 77.5 | 465.0 |
| 85 - 100 | 6 | 92.5 | 555.0 |
| **Total** | $\Sigma f_i = 30$ | | $\Sigma f_i x_i = 1860.0$ |

The sum of the values in the last column gives us $\Sigma f_i x_i$. So, the mean $\bar{x}$ of the given data is given by

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1860.0}{30} = 62$$

This new method of finding the mean is known as the **Direct Method**.

We observe that Tables 14.1 and 14.3 are using the same data and employing the same formula for the calculation of the mean but the results obtained are different. Can you think why this is so, and which one is more accurate? The difference in the two values is because of the mid-point assumption in Table 14.3, 59.3 being the exact mean, while 62 an approximate mean.

Sometimes when the numerical values of $x_i$ and $f_i$ are large, finding the product of $x_i$ and $f_i$ becomes tedious and time consuming. So, for such situations, let us think of a method of reducing these calculations.

We can do nothing with the $f_i$'s, but we can change each $x_i$ to a smaller number so that our calculations become easy. How do we do this? What about subtracting a fixed number from each of these $x_i$'s? Let us try this method.

The first step is to choose one among the $x_i$'s as the *assumed mean*, and denote it by '$a$'. Also, to further reduce our calculation work, we may take '$a$' to be that $x_i$ which lies in the centre of $x_1, x_2, \ldots, x_n$. So, we can choose $a = 47.5$ or $a = 62.5$. Let us choose $a = 47.5$.

The next step is to find the difference $d_i$ between $a$ and each of the $x_i$'s, that is, the **deviation** of '$a$' from each of the $x_i$'s.

i.e., $$d_i = x_i - a = x_i - 47.5$$

The third step is to find the product of $d_i$ with the corresponding $f_i$, and take the sum of all the $f_i d_i$'s. The calculations are shown in Table 14.4.

**Table 14.4**

| Class interval | Number of students ($f_i$) | Class mark ($x_i$) | $d_i = x_i - 47.5$ | $f_i d_i$ |
|---|---|---|---|---|
| 10 - 25 | 2 | 17.5 | –30 | –60 |
| 25 - 40 | 3 | 32.5 | –15 | –45 |
| 40 - 55 | 7 | 47.5 | 0 | 0 |
| 55 - 70 | 6 | 62.5 | 15 | 90 |
| 70 - 85 | 6 | 77.5 | 30 | 180 |
| 85 - 100 | 6 | 92.5 | 45 | 270 |
| **Total** | $\Sigma f_i = 30$ | | | $\Sigma f_i d_i = 435$ |

So, from Table 14.4, the mean of the deviations, $\bar{d} = \dfrac{\Sigma f_i d_i}{\Sigma f_i}$.

Now, let us find the relation between $\bar{d}$ and $\bar{x}$.

Since in obtaining $d_i$, we subtracted '$a$' from each $x_i$, so, in order to get the mean $\bar{x}$, we need to add '$a$' to $\bar{d}$. This can be explained mathematically as:

Mean of deviations, $$\bar{d} = \frac{\Sigma f_i d_i}{\Sigma f_i}$$

So, $$\bar{d} = \frac{\Sigma f_i (x_i - a)}{\Sigma f_i}$$

$$= \frac{\Sigma f_i x_i}{\Sigma f_i} - \frac{\Sigma f_i a}{\Sigma f_i}$$

$$= \bar{x} - a\frac{\Sigma f_i}{\Sigma f_i}$$

$$= \bar{x} - a$$

So, $$\bar{x} = a + \bar{d}$$

i.e., $$\bar{x} = a + \frac{\Sigma f_i d_i}{\Sigma f_i}$$

Substituting the values of $a$, $\Sigma f_i d_i$ and $\Sigma f_i$ from Table 14.4, we get

$$\bar{x} = 47.5 + \frac{435}{30} = 47.5 + 14.5 = 62.$$

Therefore, the mean of the marks obtained by the students is 62.

The method discussed above is called the **Assumed Mean Method**.

**Activity 1 :** From the Table 14.3 find the mean by taking each of $x_i$ (i.e., 17.5, 32.5, and so on) as '$a$'. What do you observe? You will find that the mean determined in each case is the same, i.e., 62. (Why ?)

So, we can say that the value of the mean obtained does not depend on the choice of '$a$'.

Observe that in Table 14.4, the values in Column 4 are all multiples of 15. So, if we divide the values in the entire Column 4 by 15, we would get smaller numbers to multiply with $f_i$. (Here, 15 is the class size of each class interval.)

So, let $u_i = \frac{x_i - a}{h}$, where $a$ is the assumed mean and $h$ is the class size.

Now, we calculate $u_i$ in this way and continue as before (i.e., find $f_i u_i$ and then $\Sigma f_i u_i$). Taking $h$ = 15, let us form Table 14.5.

**Table 14.5**

| Class interval | $f_i$ | $x_i$ | $d_i = x_i - a$ | $u_i = \frac{x_i - a}{h}$ | $f_i u_i$ |
|---|---|---|---|---|---|
| 10 - 25 | 2 | 17.5 | –30 | –2 | –4 |
| 25 - 40 | 3 | 32.5 | –15 | –1 | –3 |
| 40 - 55 | 7 | 47.5 | 0 | 0 | 0 |
| 55 - 70 | 6 | 62.5 | 15 | 1 | 6 |
| 70 - 85 | 6 | 77.5 | 30 | 2 | 12 |
| 85 - 100 | 6 | 92.5 | 45 | 3 | 18 |
| **Total** | $\Sigma f_i = 30$ | | | | $\Sigma f_i u_i = 29$ |

Let $$\bar{u} = \frac{\Sigma f_i u_i}{\Sigma f_i}$$

Here, again let us find the relation between $\bar{u}$ and $\bar{x}$.

We have, $$u_i = \frac{x_i - a}{h}$$

Therefore, $$\bar{u} = \frac{\Sigma f_i \frac{(x_i - a)}{h}}{\Sigma f_i} = \frac{1}{h}\left[\frac{\Sigma f_i x_i - a\Sigma f_i}{\Sigma f_i}\right]$$

$$= \frac{1}{h}\left[\frac{\Sigma f_i x_i}{\Sigma f_i} - a\frac{\Sigma f_i}{\Sigma f_i}\right]$$

$$= \frac{1}{h}[\bar{x} - a]$$

So, $$h\bar{u} = \bar{x} - a$$

i.e., $$\bar{x} = a + h\bar{u}$$

So, $$\bar{x} = a + h\left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right)$$

Now, substituting the values of $a$, $h$, $\Sigma f_i u_i$ and $\Sigma f_i$ from Table 14.5, we get

$$\bar{x} = 47.5 + 15 \times \left(\frac{29}{30}\right)$$

$$= 47.5 + 14.5 = 62$$

So, the mean marks obtained by a student is 62.

The method discussed above is called the **Step-deviation** method.

We note that :

- the step-deviation method will be convenient to apply if all the $d_i$'s have a common factor.
- The mean obtained by all the three methods is the same.
- The assumed mean method and step-deviation method are just simplified forms of the direct method.
- The formula $\bar{x} = a + h\bar{u}$ still holds if $a$ and $h$ are not as given above, but are any non-zero numbers such that $u_i = \frac{x_i - a}{h}$.

Let us apply these methods in another example.

**Example 2 :** The table below gives the percentage distribution of female teachers in the primary schools of rural areas of various states and union territories (U.T.) of India. Find the mean percentage of female teachers by all the three methods discussed in this section.

| **Percentage of female teachers** | 15 - 25 | 25 - 35 | 35 - 45 | 45 - 55 | 55 - 65 | 65 - 75 | 75 - 85 |
|---|---|---|---|---|---|---|---|
| **Number of States/U.T.** | 6 | 11 | 7 | 4 | 4 | 2 | 1 |

**Source :** *Seventh All India School Education Survey conducted by NCERT*

**Solution :** Let us find the class marks, $x_i$, of each class, and put them in a column (see Table 14.6):

**Table 14.6**

| **Percentage of female teachers** | **Number of States /U.T. ($f_i$)** | $x_i$ |
|---|---|---|
| 15 - 25 | 6 | 20 |
| 25 - 35 | 11 | 30 |
| 35 - 45 | 7 | 40 |
| 45 - 55 | 4 | 50 |
| 55 - 65 | 4 | 60 |
| 65 - 75 | 2 | 70 |
| 75 - 85 | 1 | 80 |

Here we take $a = 50$, $h = 10$, then $d_i = x_i - 50$ and $u_i = \frac{x_i - 50}{10}$.

We now find $d_i$ and $u_i$ and put them in Table 14.7.

**Table 14.7**

| Percentage of female teachers | Number of states/U.T. ($f_i$) | $x_i$ | $d_i = x_i - 50$ | $u_i = \frac{x_i - 50}{10}$ | $f_i x_i$ | $f_i d_i$ | $f_i u_i$ |
|---|---|---|---|---|---|---|---|
| 15 - 25 | 6 | 20 | –30 | –3 | 120 | –180 | –18 |
| 25 - 35 | 11 | 30 | –20 | –2 | 330 | –220 | –22 |
| 35 - 45 | 7 | 40 | –10 | –1 | 280 | –70 | –7 |
| 45 - 55 | 4 | 50 | 0 | 0 | 200 | 0 | 0 |
| 55 - 65 | 4 | 60 | 10 | 1 | 240 | 40 | 4 |
| 65 - 75 | 2 | 70 | 20 | 2 | 140 | 40 | 4 |
| 75 - 85 | 1 | 80 | 30 | 3 | 80 | 30 | 3 |
| **Total** | **35** | | | | **1390** | **–360** | **–36** |

From the table above, we obtain $\Sigma f_i = 35$, $\Sigma f_i x_i = 1390$,

$$\Sigma f_i d_i = -360, \quad \Sigma f_i u_i = -36.$$

Using the direct method, $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1390}{35} = 39.71$

Using the assumed mean method,

$$\bar{x} = a + \frac{\Sigma f_i d_i}{\Sigma f_i} = 50 + \frac{(-360)}{35} = 39.71$$

Using the step-deviation method,

$$\bar{x} = a + \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right) \times h = 50 + \left(\frac{-36}{35}\right) \times 10 = 39.71$$

Therefore, the mean percentage of female teachers in the primary schools of rural areas is 39.71.

**Remark :** The result obtained by all the three methods is the same. So the choice of method to be used depends on the numerical values of $x_i$ and $f_i$. If $x_i$ and $f_i$ are sufficiently small, then the direct method is an appropriate choice. If $x_i$ and $f_i$ are numerically large numbers, then we can go for the assumed mean method or step-deviation method. If the class sizes are unequal, and $x_i$ are large numerically, we can still apply the step-deviation method by taking $h$ to be a suitable divisor of all the $d_i$'s.

**Example 3 :** The distribution below shows the number of wickets taken by bowlers in one-day cricket matches. Find the mean number of wickets by choosing a suitable method. What does the mean signify?

| **Number of wickets** | 20 - 60 | 60 - 100 | 100 - 150 | 150 - 250 | 250 - 350 | 350 - 450 |
|---|---|---|---|---|---|---|
| **Number of bowlers** | 7 | 5 | 16 | 12 | 2 | 3 |

**Solution :** Here, the class size varies, and the $x_i$'s are large. Let us still apply the step-deviation method with $a = 200$ and $h = 20$. Then, we obtain the data as in Table 14.8.

**Table 14.8**

| **Number of wickets taken** | **Number of bowlers ($f_i$)** | $x_i$ | $d_i = x_i - 200$ | $u_i = \frac{d_i}{20}$ | $u_i f_i$ |
|---|---|---|---|---|---|
| 20 - 60 | 7 | 40 | –160 | –8 | –56 |
| 60 - 100 | 5 | 80 | –120 | –6 | –30 |
| 100 - 150 | 16 | 125 | –75 | –3.75 | –60 |
| 150 - 250 | 12 | 200 | 0 | 0 | 0 |
| 250 - 350 | 2 | 300 | 100 | 5 | 10 |
| 350 - 450 | 3 | 400 | 200 | 10 | 30 |
| **Total** | **45** | | | | **–106** |

So, $\bar{u} = \frac{-106}{45}$. Therefore, $\bar{x} = 200 + 20\left(\frac{-106}{45}\right) = 200 - 47.11 = 152.89$.

This tells us that, on an average, the number of wickets taken by these 45 bowlers in one-day cricket is 152.89.

Now, let us see how well you can apply the concepts discussed in this section!

## Activity 2 :

Divide the students of your class into three groups and ask each group to do one of the following activities.

1. Collect the marks obtained by all the students of your class in Mathematics in the latest examination conducted by your school. Form a grouped frequency distribution of the data obtained.
2. Collect the daily maximum temperatures recorded for a period of 30 days in your city. Present this data as a grouped frequency table.
3. Measure the heights of all the students of your class (in cm) and form a grouped frequency distribution table of this data.

After all the groups have collected the data and formed grouped frequency distribution tables, the groups should find the mean in each case by the method which they find appropriate.

## EXERCISE 14.1

**1.** A survey was conducted by a group of students as a part of their environment awareness programme, in which they collected the following data regarding the number of plants in 20 houses in a locality. Find the mean number of plants per house.

| **Number of plants** | 0 - 2 | 2 - 4 | 4 - 6 | 6 - 8 | 8 - 10 | 10 - 12 | 12 - 14 |
|---|---|---|---|---|---|---|---|
| **Number of houses** | 1 | 2 | 1 | 5 | 6 | 2 | 3 |

Which method did you use for finding the mean, and why?

**2.** Consider the following distribution of daily wages of 50 workers of a factory.

| **Daily wages (in ₹)** | 500 - 520 | 520-540 | 540 - 560 | 560 - 580 | 580-600 |
|---|---|---|---|---|---|
| **Number of workers** | 12 | 14 | 8 | 6 | 10 |

Find the mean daily wages of the workers of the factory by using an appropriate method.

**3.** The following distribution shows the daily pocket allowance of children of a locality. The mean pocket allowance is Rs 18. Find the missing frequency $f$.

| **Daily pocket allowance (in ₹)** | 11 - 13 | 13 - 15 | 15 - 17 | 17 - 19 | 19 - 21 | 21 - 23 | 23 - 25 |
|---|---|---|---|---|---|---|---|
| **Number of children** | 7 | 6 | 9 | 13 | $f$ | 5 | 4 |

**4.** Thirty women were examined in a hospital by a doctor and the number of heartbeats per minute were recorded and summarised as follows. Find the mean heartbeats per minute for these women, choosing a suitable method.

| **Number of heartbeats per minute** | 65 - 68 | 68 - 71 | 71 - 74 | 74 - 77 | 77 - 80 | 80 - 83 | 83 - 86 |
|---|---|---|---|---|---|---|---|
| **Number of women** | 2 | 4 | 3 | 8 | 7 | 4 | 2 |

**5.** In a retail market, fruit vendors were selling mangoes kept in packing boxes. These boxes contained varying number of mangoes. The following was the distribution of mangoes according to the number of boxes.

| **Number of mangoes** | 50 - 52 | 53 - 55 | 56 - 58 | 59 - 61 | 62 - 64 |
|---|---|---|---|---|---|
| **Number of boxes** | 15 | 110 | 135 | 115 | 25 |

Find the mean number of mangoes kept in a packing box. Which method of finding the mean did you choose?

**6.** The table below shows the daily expenditure on food of 25 households in a locality.

| **Daily expenditure (in ₹)** | 100 - 150 | 150 - 200 | 200 - 250 | 250 - 300 | 300 - 350 |
|---|---|---|---|---|---|
| **Number of households** | 4 | 5 | 12 | 2 | 2 |

Find the mean daily expenditure on food by a suitable method.

**7.** To find out the concentration of $SO_2$ in the air (in parts per million, i.e., ppm), the data was collected for 30 localities in a certain city and is presented below:

| **Concentration of $SO_2$ (in ppm)** | **Frequency** |
|---|---|
| 0.00 - 0.04 | 4 |
| 0.04 - 0.08 | 9 |
| 0.08 - 0.12 | 9 |
| 0.12 - 0.16 | 2 |
| 0.16 - 0.20 | 4 |
| 0.20 - 0.24 | 2 |

Find the mean concentration of $SO_2$ in the air.

**8.** A class teacher has the following absentee record of 40 students of a class for the whole term. Find the mean number of days a student was absent.

| **Number of days** | 0 - 6 | 6 - 10 | 10 - 14 | 14 - 20 | 20 - 28 | 28 - 38 | 38 - 40 |
|---|---|---|---|---|---|---|---|
| **Number of students** | 11 | 10 | 7 | 4 | 4 | 3 | 1 |

**9.** The following table gives the literacy rate (in percentage) of 35 cities. Find the mean literacy rate.

| **Literacy rate (in %)** | 45 - 55 | 55 - 65 | 65 - 75 | 75 - 85 | 85 - 95 |
|---|---|---|---|---|---|
| **Number of cities** | 3 | 10 | 11 | 8 | 3 |

## 14.3 Mode of Grouped Data

Recall from Class IX, a mode is that value among the observations which occurs most often, that is, the value of the observation having the maximum frequency. Further, we discussed finding the mode of ungrouped data. Here, we shall discuss ways of obtaining a mode of grouped data. It is possible that more than one value may have the same maximum frequency. In such situations, the data is said to be multimodal. Though grouped data can also be multimodal, we shall restrict ourselves to problems having a single mode only.

Let us first recall how we found the mode for ungrouped data through the following example.

**Example 4 :** The wickets taken by a bowler in 10 cricket matches are as follows:

2 6 4 5 0 2 1 3 2 3

Find the mode of the data.

**Solution :** Let us form the frequency distribution table of the given data as follows:

| **Number of wickets** | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| **Number of matches** | 1 | 1 | 3 | 2 | 1 | 1 | 1 |

Clearly, 2 is the number of wickets taken by the bowler in the maximum number (i.e., 3) of matches. So, the mode of this data is 2.

In a grouped frequency distribution, it is not possible to determine the mode by looking at the frequencies. Here, we can only locate a class with the maximum frequency, called the **modal class**. The mode is a value inside the modal class, and is given by the formula:

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$$

where $l$ = lower limit of the modal class,

$h$ = size of the class interval (assuming all class sizes to be equal),

$f_1$ = frequency of the modal class,

$f_0$ = frequency of the class preceding the modal class,

$f_2$ = frequency of the class succeeding the modal class.

Let us consider the following examples to illustrate the use of this formula.

**Example 5 :** A survey conducted on 20 households in a locality by a group of students resulted in the following frequency table for the number of family members in a household:

| **Family size** | 1 - 3 | 3 - 5 | 5 - 7 | 7 - 9 | 9 - 11 |
|---|---|---|---|---|---|
| **Number of families** | 7 | 8 | 2 | 2 | 1 |

Find the mode of this data.

**Solution :** Here the maximum class frequency is 8, and the class corresponding to this frequency is 3 – 5. So, the modal class is 3 – 5.

Now

modal class = 3 – 5, lower limit ($l$) of modal class = 3, class size ($h$) = 2

frequency ($f_1$) of the modal class = 8,

frequency ($f_0$) of class preceding the modal class = 7,

frequency ($f_2$) of class succeeding the modal class = 2.

Now, let us substitute these values in the formula :

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$$

$$= 3 + \left(\frac{8-7}{2 \times 8 - 7 - 2}\right) \times 2 = 3 + \frac{2}{7} = 3.286$$

Therefore, the mode of the data above is 3.286.

**Example 6 :** The marks distribution of 30 students in a mathematics examination are given in Table 14.3 of Example 1. Find the mode of this data. Also compare and interpret the mode and the mean.

**Solution :** Refer to Table 14.3 of Example 1. Since the maximum number of students (i.e., 7) have got marks in the interval 40 - 55, the modal class is 40 - 55. Therefore,

the lower limit ( $l$ ) of the modal class = 40,

the class size ( $h$) = 15,

the frequency ( $f_1$) of modal class = 7,

the frequency ( $f_0$) of the class preceding the modal class = 3,

the frequency ( $f_2$) of the class succeeding the modal class = 6.

Now, using the formula:

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h,$$

we get

$$\text{Mode} = 40 + \left(\frac{7-3}{14-6-3}\right) \times 15 = 52$$

So, the mode marks is 52.

Now, from Example 1, you know that the mean marks is 62.

So, the maximum number of students obtained 52 marks, while on an average a student obtained 62 marks.

**Remarks :**

1. In Example 6, the mode is less than the mean. But for some other problems it may be equal or more than the mean also.

2. It depends upon the demand of the situation whether we are interested in finding the average marks obtained by the students or the average of the marks obtained by most

of the students. In the first situation, the mean is required and in the second situation, the mode is required.

**Activity 3 :** Continuing with the same groups as formed in Activity 2 and the situations assigned to the groups. Ask each group to find the mode of the data. They should also compare this with the mean, and interpret the meaning of both.

**Remark :** The mode can also be calculated for grouped data with unequal class sizes. However, we shall not be discussing it.

## EXERCISE 14.2

1. The following table shows the ages of the patients admitted in a hospital during a year:

| **Age (in years)** | 5 - 15 | 15 - 25 | 25 - 35 | 35 - 45 | 45 - 55 | 55 - 65 |
|---|---|---|---|---|---|---|
| **Number of patients** | 6 | 11 | 21 | 23 | 14 | 5 |

Find the mode and the mean of the data given above. Compare and interpret the two measures of central tendency.

2. The following data gives the information on the observed lifetimes (in hours) of 225 electrical components :

| **Lifetimes (in hours)** | 0 - 20 | 20 - 40 | 40 - 60 | 60 - 80 | 80 - 100 | 100 - 120 |
|---|---|---|---|---|---|---|
| **Frequency** | 10 | 35 | 52 | 61 | 38 | 29 |

Determine the modal lifetimes of the components.

3. The following data gives the distribution of total monthly household expenditure of 200 families of a village. Find the modal monthly expenditure of the families. Also, find the mean monthly expenditure :

| **Expenditure (in ₹)** | **Number of families** |
|---|---|
| 1000 - 1500 | 24 |
| 1500 - 2000 | 40 |
| 2000 - 2500 | 33 |
| 2500 - 3000 | 28 |
| 3000 - 3500 | 30 |
| 3500 - 4000 | 22 |
| 4000 - 4500 | 16 |
| 4500 - 5000 | 7 |

**4.** The following distribution gives the state-wise teacher-student ratio in higher secondary schools of India. Find the mode and mean of this data. Interpret the two measures.

| Number of students per teacher | Number of states / U.T. |
|---|---|
| 15 - 20 | 3 |
| 20 - 25 | 8 |
| 25 - 30 | 9 |
| 30 - 35 | 10 |
| 35 - 40 | 3 |
| 40 - 45 | 0 |
| 45 - 50 | 0 |
| 50 - 55 | 2 |

**5.** The given distribution shows the number of runs scored by some top batsmen of the world in one-day international cricket matches.

| Runs scored | Number of batsmen |
|---|---|
| 3000 - 4000 | 4 |
| 4000 - 5000 | 18 |
| 5000 - 6000 | 9 |
| 6000 - 7000 | 7 |
| 7000 - 8000 | 6 |
| 8000 - 9000 | 3 |
| 9000 - 10000 | 1 |
| 10000 - 11000 | 1 |

Find the mode of the data.

**6.** A student noted the number of cars passing through a spot on a road for 100 periods each of 3 minutes and summarised it in the table given below. Find the mode of the data :

| Number of cars | 0 - 10 | 10 - 20 | 20 - 30 | 30 - 40 | 40 - 50 | 50 - 60 | 60 - 70 | 70 - 80 |
|---|---|---|---|---|---|---|---|---|
| Frequency | 7 | 14 | 13 | 12 | 20 | 11 | 15 | 8 |

## 14.4 Median of Grouped Data

As you have studied in Class IX, the median is a measure of central tendency which gives the value of the middle-most observation in the data. Recall that for finding the median of ungrouped data, we first arrange the data values of the observations in ascending order. Then, if $n$ is odd, the median is the $\left(\frac{n+1}{2}\right)$th observation. And, if $n$ is even, then the median will be the average of the $\frac{n}{2}$th and the $\left(\frac{n}{2}+1\right)$th observations.

Suppose, we have to find the median of the following data, which gives the marks, out of 50, obtained by 100 students in a test :

| **Marks obtained** | 20 | 29 | 28 | 33 | 42 | 38 | 43 | 25 |
|---|---|---|---|---|---|---|---|---|
| **Number of students** | 6 | 28 | 24 | 15 | 2 | 4 | 1 | 20 |

First, we arrange the marks in ascending order and prepare a frequency table as follows :

**Table 14.9**

| **Marks obtained** | **Number of students (Frequency)** |
|---|---|
| 20 | 6 |
| 25 | 20 |
| 28 | 24 |
| 29 | 28 |
| 33 | 15 |
| 38 | 4 |
| 42 | 2 |
| 43 | 1 |
| **Total** | **100** |

Here $n = 100$, which is even. The median will be the average of the $\frac{n}{2}$th and the $\left(\frac{n}{2}+1\right)$th observations, i.e., the 50th and 51st observations. To find these observations, we proceed as follows:

**Table 14.10**

| Marks obtained | Number of students |
|---|---|
| 20 | 6 |
| upto 25 | 6 + 20 = 26 |
| upto 28 | 26 + 24 = 50 |
| upto 29 | 50 + 28 = 78 |
| upto 33 | 78 + 15 = 93 |
| upto 38 | 93 + 4 = 97 |
| upto 42 | 97 + 2 = 99 |
| upto 43 | 99 + 1 = 100 |

Now we add another column depicting this information to the frequency table above and name it as *cumulative frequency column*.

**Table 14.11**

| Marks obtained | Number of students | Cumulative frequency |
|---|---|---|
| 20 | 6 | 6 |
| 25 | 20 | 26 |
| 28 | 24 | 50 |
| 29 | 28 | 78 |
| 33 | 15 | 93 |
| 38 | 4 | 97 |
| 42 | 2 | 99 |
| 43 | 1 | 100 |

From the table above, we see that:

50th observaton is 28 (Why?)

51st observation is 29

So, Median = $\frac{28 + 29}{2} = 28.5$

**Remark :** The part of Table 14.11 consisting Column 1 and Column 3 is known as *Cumulative Frequency Table*. The median marks 28.5 conveys the information that about 50% students obtained marks less than 28.5 and another 50% students obtained marks more than 28.5.

Now, let us see how to obtain the median of grouped data, through the following situation.

Consider a grouped frequency distribution of marks obtained, out of 100, by 53 students, in a certain examination, as follows:

**Table 14.12**

| Marks | Number of students |
|---|---|
| 0 - 10 | 5 |
| 10 - 20 | 3 |
| 20 - 30 | 4 |
| 30 - 40 | 3 |
| 40 - 50 | 3 |
| 50 - 60 | 4 |
| 60 - 70 | 7 |
| 70 - 80 | 9 |
| 80 - 90 | 7 |
| 90 - 100 | 8 |

From the table above, try to answer the following questions:

How many students have scored marks less than 10? The answer is clearly 5.

How many students have scored less than 20 marks? Observe that the number of students who have scored less than 20 include the number of students who have scored marks from 0 - 10 as well as the number of students who have scored marks from 10 - 20. So, the total number of students with marks less than 20 is 5 + 3, i.e., 8. We say that the cumulative frequency of the class 10 -20 is 8.

Similarly, we can compute the cumulative frequencies of the other classes, i.e., the number of students with marks less than 30, less than 40, . . ., less than 100. We give them in Table 14.13 given below:

**Table 14.13**

| Marks obtained | Number of students (Cumulative frequency) |
|---|---|
| Less than 10 | 5 |
| Less than 20 | 5 + 3 = 8 |
| Less than 30 | 8 + 4 = 12 |
| Less than 40 | 12 + 3 = 15 |
| Less than 50 | 15 + 3 = 18 |
| Less than 60 | 18 + 4 = 22 |
| Less than 70 | 22 + 7 = 29 |
| Less than 80 | 29 + 9 = 38 |
| Less than 90 | 38 + 7 = 45 |
| Less than 100 | 45 + 8 = 53 |



The distribution given above is called the *cumulative frequency distribution of the less than type*. Here 10, 20, 30, . . . 100, are the upper limits of the respective class intervals.

We can similarly make the table for the number of students with scores, more than or equal to 0, more than or equal to 10, more than or equal to 20, and so on. From Table 14.12, we observe that all 53 students have scored marks more than or equal to 0. Since there are 5 students scoring marks in the interval 0 - 10, this means that there are 53 – 5 = 48 students getting more than or equal to 10 marks. Continuing in the same manner, we get the number of students scoring 20 or above as 48 – 3 = 45, 30 or above as 45 – 4 = 41, and so on, as shown in Table 14.14.

**Table 14.14**

| Marks obtained | Number of students (Cumulative frequency) |
|---|---|
| More than or equal to 0 | 53 |
| More than or equal to 10 | 53 – 5 = 48 |
| More than or equal to 20 | 48 – 3 = 45 |
| More than or equal to 30 | 45 – 4 = 41 |
| More than or equal to 40 | 41 – 3 = 38 |
| More than or equal to 50 | 38 – 3 = 35 |
| More than or equal to 60 | 35 – 4 = 31 |
| More than or equal to 70 | 31 – 7 = 24 |
| More than or equal to 80 | 24 – 9 = 15 |
| More than or equal to 90 | 15 – 7 = 8 |

The table above is called a *cumulative frequency distribution of the more than type*. Here 0, 10, 20, . . ., 90 give the lower limits of the respective class intervals.

Now, to find the median of grouped data, we can make use of any of these cumulative frequency distributions.

Let us combine Tables 14.12 and 14.13 to get Table 14.15 given below:

**Table 14.15**

| Marks | Number of students ($f$) | Cumulative frequency (cf) |
|---|---|---|
| 0 - 10 | 5 | 5 |
| 10 - 20 | 3 | 8 |
| 20 - 30 | 4 | 12 |
| 30 - 40 | 3 | 15 |
| 40 - 50 | 3 | 18 |
| 50 - 60 | 4 | 22 |
| 60 - 70 | 7 | 29 |
| 70 - 80 | 9 | 38 |
| 80 - 90 | 7 | 45 |
| 90 - 100 | 8 | 53 |

Now in a grouped data, we may not be able to find the middle observation by looking at the cumulative frequencies as the middle observation will be some value in a class interval. It is, therefore, necessary to find the value inside a class that divides the whole distribution into two halves. But which class should this be?

To find this class, we find the cumulative frequencies of all the classes and $\frac{n}{2}$. We now locate the class whose cumulative frequency is greater than (and nearest to) $\frac{n}{2}$. This is called the *median class*. In the distribution above, $n = 53$. So, $\frac{n}{2} = 26.5$. Now 60 – 70 is the class whose cumulative frequency 29 is greater than (and nearest to) $\frac{n}{2}$, i.e., 26.5.

Therefore, 60 – 70 is the **median class**.

After finding the median class, we use the following formula for calculating the median.

$$\text{Median} = l + \left(\frac{\frac{n}{2} - \text{cf}}{f}\right) \times h,$$

where $l$ = lower limit of median class,

$n$ = number of observations,

cf = cumulative frequency of class preceding the median class,

$f$ = frequency of median class,

$h$ = class size (assuming class size to be equal).

Substituting the values $\frac{n}{2} = 26.5$, $l = 60$, cf = 22, $f = 7$, $h = 10$

in the formula above, we get

$$\text{Median} = 60 + \left(\frac{26.5 - 22}{7}\right) \times 10$$

$$= 60 + \frac{45}{7}$$

$$= 66.4$$

So, about half the students have scored marks less than 66.4, and the other half have scored marks more than 66.4.

**Example 7 :** A survey regarding the heights (in cm) of 51 girls of Class X of a school was conducted and the following data was obtained:

| Height (in cm) | Number of girls |
|---|---|
| Less than 140 | 4 |
| Less than 145 | 11 |
| Less than 150 | 29 |
| Less than 155 | 40 |
| Less than 160 | 46 |
| Less than 165 | 51 |

Find the median height.

**Solution :** To calculate the median height, we need to find the class intervals and their corresponding frequencies.

The given distribution being of the *less than type*, 140, 145, 150, . . ., 165 give the upper limits of the corresponding class intervals. So, the classes should be below 140, 140 - 145, 145 - 150, . . ., 160 - 165. Observe that from the given distribution, we find that there are 4 girls with height less than 140, i.e., the frequency of class interval below 140 is 4. Now, there are 11 girls with heights less than 145 and 4 girls with height less than 140. Therefore, the number of girls with height in the interval 140 - 145 is 11 – 4 = 7. Similarly, the frequency of 145 - 150 is 29 – 11 = 18, for 150 - 155, it is 40 – 29 = 11, and so on. So, our frequency distribution table with the given cumulative frequencies becomes:

**Table 14.16**

| Class intervals | Frequency | Cumulative frequency |
|---|---|---|
| Below 140 | 4 | 4 |
| 140 - 145 | 7 | 11 |
| 145 - 150 | 18 | 29 |
| 150 - 155 | 11 | 40 |
| 155 - 160 | 6 | 46 |
| 160 - 165 | 5 | 51 |

Now $n = 51$. So, $\frac{n}{2} = \frac{51}{2} = 25.5$. This observation lies in the class 145 - 150. Then,

$l$ (the lower limit) = 145,

cf (the cumulative frequency of the class preceding 145 - 150) = 11,

$f$ (the frequency of the median class 145 - 150) = 18,

$h$ (the class size) = 5.

Using the formula, Median $= l + \left(\frac{\frac{n}{2} - \text{cf}}{f}\right) \times h$, we have

$$\text{Median} = 145 + \left(\frac{25.5 - 11}{18}\right) \times 5$$

$$= 145 + \frac{72.5}{18} = 149.03.$$

So, the median height of the girls is 149.03 cm.

This means that the height of about 50% of the girls is less than this height, and 50% are taller than this height.

**Example 8 :** The median of the following data is 525. Find the values of $x$ and $y$, if the total frequency is 100.

| Class interval | Frequency |
|---|---|
| 0 - 100 | 2 |
| 100 - 200 | 5 |
| 200 - 300 | $x$ |
| 300 - 400 | 12 |
| 400 - 500 | 17 |
| 500 - 600 | 20 |
| 600 - 700 | $y$ |
| 700 - 800 | 9 |
| 800 - 900 | 7 |
| 900 - 1000 | 4 |

**Solution :**

| Class intervals | Frequency | Cumulative frequency |
|---|---|---|
| 0 - 100 | 2 | 2 |
| 100 - 200 | 5 | 7 |
| 200 - 300 | $x$ | $7 + x$ |
| 300 - 400 | 12 | $19 + x$ |
| 400 - 500 | 17 | $36 + x$ |
| 500 - 600 | 20 | $56 + x$ |
| 600 - 700 | $y$ | $56 + x + y$ |
| 700 - 800 | 9 | $65 + x + y$ |
| 800 - 900 | 7 | $72 + x + y$ |
| 900 - 1000 | 4 | $76 + x + y$ |

It is given that $n = 100$

So, $76 + x + y = 100$, i.e., $x + y = 24$ (1)

The median is 525, which lies in the class 500 – 600

So, $l = 500$, $f = 20$, $\text{cf} = 36 + x$, $h = 100$

Using the formula : $$\text{Median} = l + \left(\frac{\frac{n}{2} - \text{cf}}{f}\right)h,$$ we get

$$525 = 500 + \left(\frac{50 - 36 - x}{20}\right) \times 100$$

i.e., $525 - 500 = (14 - x) \times 5$

i.e., $25 = 70 - 5x$

i.e., $5x = 70 - 25 = 45$

So, $x = 9$

Therefore, from (1), we get $9 + y = 24$

i.e., $y = 15$

Now, that you have studied about all the three measures of central tendency, let us discuss **which measure would be best suited for a particular requirement**.

The mean is the most frequently used measure of central tendency because it takes into account all the observations, and lies between the extremes, i.e., the largest and the smallest observations of the entire data. It also enables us to compare two or more distributions. For example, by comparing the average (mean) results of students of different schools of a particular examination, we can conclude which school has a better performance.

However, extreme values in the data affect the mean. For example, the mean of classes having frequencies more or less the same is a good representative of the data. But, if one class has frequency, say 2, and the five others have frequency 20, 25, 20, 21, 18, then the mean will certainly not reflect the way the data behaves. So, in such cases, the mean is not a good representative of the data.

In problems where individual observations are not important, and we wish to find out a 'typical' observation, the median is more appropriate, e.g., finding the typical productivity rate of workers, average wage in a country, etc. These are situations where extreme values may be there. So, rather than the mean, we take the median as a better measure of central tendency.

In situations which require establishing the most frequent value or most popular item, the mode is the best choice, e.g., to find the most popular T.V. programme being watched, the consumer item in greatest demand, the colour of the vehicle used by most of the people, etc.

**Remarks :**

1. There is a empirical relationship between the three measures of central tendency :

$$\textbf{3 Median = Mode + 2 Mean}$$

2. The median of grouped data with unequal class sizes can also be calculated. However, we shall not discuss it here.

## EXERCISE 14.3

1. The following frequency distribution gives the monthly consumption of electricity of 68 consumers of a locality. Find the median, mean and mode of the data and compare them.

| Monthly consumption (in units) | Number of consumers |
|---|---|
| 65 - 85 | 4 |
| 85 - 105 | 5 |
| 105 - 125 | 13 |
| 125 - 145 | 20 |
| 145 - 165 | 14 |
| 165 - 185 | 8 |
| 185 - 205 | 4 |

2. If the median of the distribution given below is 28.5, find the values of $x$ and $y$.

| Class interval | Frequency |
|---|---|
| 0 - 10 | 5 |
| 10 - 20 | $x$ |
| 20 - 30 | 20 |
| 30 - 40 | 15 |
| 40 - 50 | $y$ |
| 50 - 60 | 5 |
| **Total** | 60 |

3. A life insurance agent found the following data for distribution of ages of 100 policy holders. Calculate the median age, if policies are given only to persons having age 18 years onwards but less than 60 year.

| Age (in years) | Number of policy holders |
|---|---|
| Below 20 | 2 |
| Below 25 | 6 |
| Below 30 | 24 |
| Below 35 | 45 |
| Below 40 | 78 |
| Below 45 | 89 |
| Below 50 | 92 |
| Below 55 | 98 |
| Below 60 | 100 |

**4.** The lengths of 40 leaves of a plant are measured correct to the nearest millimetre, and the data obtained is represented in the following table :

| Length (in mm) | Number of leaves |
|---|---|
| 118 - 126 | 3 |
| 127 - 135 | 5 |
| 136 - 144 | 9 |
| 145 - 153 | 12 |
| 154 - 162 | 5 |
| 163 - 171 | 4 |
| 172 - 180 | 2 |

Find the median length of the leaves.

(**Hint :** The data needs to be converted to continuous classes for finding the median, since the formula assumes continuous classes. The classes then change to 117.5 - 126.5, 126.5 - 135.5, . . ., 171.5 - 180.5.)

**5.** The following table gives the distribution of the life time of 400 neon lamps :

| Life time (in hours) | Number of lamps |
|---|---|
| 1500 - 2000 | 14 |
| 2000 - 2500 | 56 |
| 2500 - 3000 | 60 |
| 3000 - 3500 | 86 |
| 3500 - 4000 | 74 |
| 4000 - 4500 | 62 |
| 4500 - 5000 | 48 |

Find the median life time of a lamp.

**6.** 100 surnames were randomly picked up from a local telephone directory and the frequency distribution of the number of letters in the English alphabets in the surnames was obtained as follows:

| Number of letters | 1 - 4 | 4 - 7 | 7 - 10 | 10 - 13 | 13 - 16 | 16 - 19 |
|---|---|---|---|---|---|---|
| Number of surnames | 6 | 30 | 40 | 16 | 4 | 4 |

Determine the median number of letters in the surnames. Find the mean number of letters in the surnames? Also, find the modal size of the surnames.

**7.** The distribution below gives the weights of 30 students of a class. Find the median weight of the students.

| Weight (in kg) | 40 - 45 | 45 - 50 | 50 - 55 | 55 - 60 | 60 - 65 | 65 - 70 | 70 - 75 |
|---|---|---|---|---|---|---|---|
| Number of students | 2 | 3 | 8 | 6 | 6 | 3 | 2 |

## 14.5 Graphical Representation of Cumulative Frequency Distribution

As we all know, pictures speak better than words. A graphical representation helps us in understanding given data at a glance. In Class IX, we have represented the data through bar graphs, histograms and frequency polygons. Let us now represent a cumulative frequency distribution graphically.

For example, let us consider the cumulative frequency distribution given in Table 14.13.

Recall that the values 10, 20, 30, . . ., 100 are the upper limits of the respective class intervals. To represent the data in the table graphically, we mark the upper limits of the class intervals on the horizontal axis ($x$-axis) and their corresponding cumulative frequencies on the vertical axis ($y$-axis), choosing a convenient scale. The scale may not be the same on both the axis. Let us now plot the points corresponding to the ordered pairs given by (upper limit, corresponding cumulative frequency), i.e., (10, 5), (20, 8), (30, 12), (40, 15), (50, 18), (60, 22), (70, 29), (80, 38), (90, 45), (100, 53) on a graph paper and join them by a free hand smooth curve. The curve we get is called a **cumulative frequency curve**, or an **ogive** (of the less than type). (See Fig. 14.1)

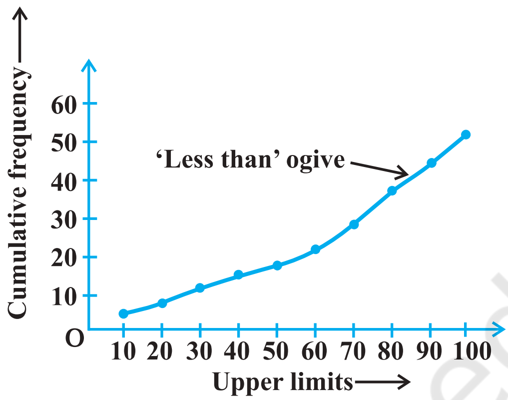



Fig. 14.1

> The term 'ogive' is pronounced as 'ojeev' and is derived from the word **ogee**. An ogee is a shape consisting of a concave arc flowing into a convex arc, so forming an S-shaped curve with vertical ends. In architecture, the ogee shape is one of the characteristics of the 14th and 15th century Gothic styles.

Next, again we consider the cumulative frequency distribution given in Table 14.14 and draw its ogive (of the more than type).

Recall that, here 0, 10, 20, . . ., 90 are the lower limits of the respective class intervals 0 - 10, 10 - 20, . . ., 90 - 100. To represent 'the more than type' graphically, we plot the lower limits on the $x$-axis and the corresponding cumulative frequencies on the $y$-axis. Then we plot the points (lower limit, corresponding cumulative frequency), i.e., (0, 53), (10, 48), (20, 45), (30, 41), (40, 38), (50, 35), (60, 31), (70, 24), (80, 15), (90, 8), on a graph paper, and join them by a free hand smooth curve. The curve we get is a *cumulative frequency curve,* or an *ogive* (*of the more than type*). (See Fig. 14.2)

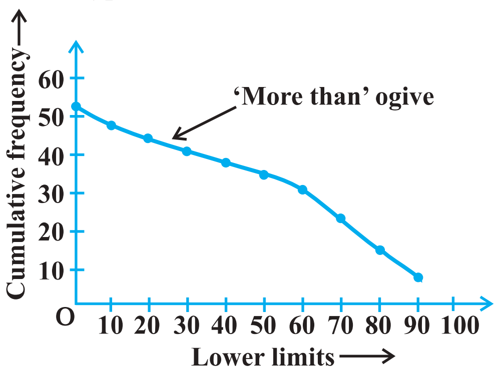



Fig. 14.2

**Remark :** Note that both the ogives (in Fig. 14.1 and Fig. 14.2) correspond to the same data, which is given in Table 14.12.

Now, are the ogives related to the median in any way? Is it possible to obtain the median from these two cumulative frequency curves corresponding to the data in Table 14.12? Let us see.

One obvious way is to locate $\frac{n}{2} = \frac{53}{2} = 26.5$ on the $y$-axis (see Fig. 14.3). From this point, draw a line parallel to the $x$-axis cutting the curve at a point. From this point, draw a perpendicular to the $x$-axis. The point of intersection of this perpendicular with the $x$-axis determines the median of the data (see Fig. 14.3).

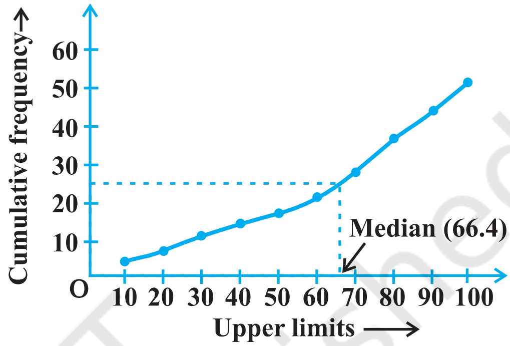



Fig. 14.3

Another way of obtaining the median is the following :

Draw both ogives (i.e., of the less than type and of the more than type) on the same axis. The two ogives will intersect each other at a point. From this point, if we draw a perpendicular on the $x$-axis, the point at which it cuts the $x$-axis gives us the median (see Fig. 14.4).

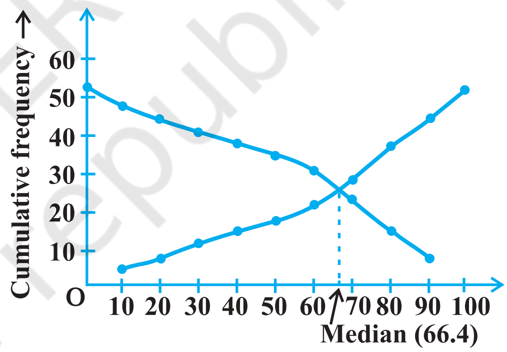



Fig. 14.4

**Example 9 :** The annual profits earned by 30 shops of a shopping complex in a locality give rise to the following distribution :

| Profit (Rs in lakhs) | Number of shops (frequency) |
|---|---|
| More than or equal to 5 | 30 |
| More than or equal to 10 | 28 |
| More than or equal to 15 | 16 |
| More than or equal to 20 | 14 |
| More than or equal to 25 | 10 |
| More than or equal to 30 | 7 |
| More than or equal to 35 | 3 |

Draw both ogives for the data above. Hence obtain the median profit.

**Solution :** We first draw the coordinate axes, with lower limits of the profit along the horizontal axis, and the cumulative frequency along the vertical axes. Then, we plot the points (5, 30), (10, 28), (15, 16), (20, 14), (25, 10), (30, 7) and (35, 3). We join these points with a smooth curve to get the 'more than' ogive, as shown in Fig. 14.5.

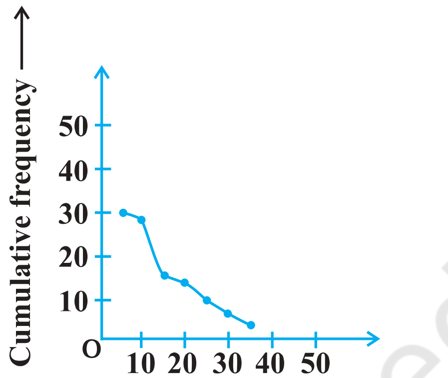



Fig. 14.5

Now, let us obtain the classes, their frequencies and the cumulative frequency from the table above.

**Table 14.17**

| Classes | 5 - 10 | 10 - 15 | 15 - 20 | 20 - 25 | 25 - 30 | 30 - 35 | 35 - 40 |
|---|---|---|---|---|---|---|---|
| **No. of shops** | 2 | 12 | 2 | 4 | 3 | 4 | 3 |
| **Cumulative frequency** | 2 | 14 | 16 | 20 | 23 | 27 | 30 |

Using these values, we plot the points (10, 2), (15, 14), (20, 16), (25, 20), (30, 23), (35, 27), (40, 30) on the same axes as in Fig. 14.5 to get the 'less than' ogive, as shown in Fig. 14.6.

The abcissa of their point of intersection is nearly 17.5, which is the median. This can also be verified by using the formula. Hence, the median profit (in lakhs) is ₹ 17.5.

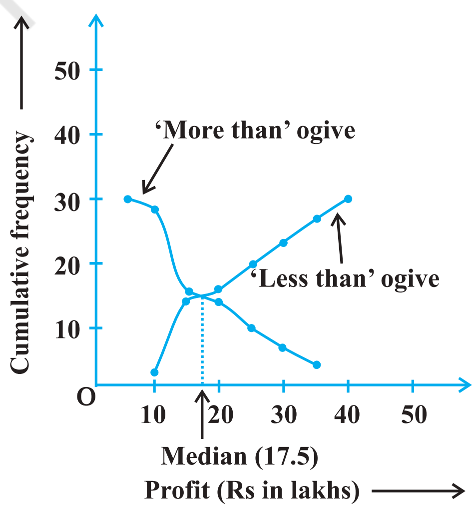



Fig. 14.6

**Remark :** In the above examples, it may be noted that the class intervals were continuous. For drawing ogives, it should be ensured that the class intervals are continuous. (Also see constructions of histograms in Class IX)

## EXERCISE 14.4

1. The following distribution gives the daily income of 50 workers of a factory.

| **Daily income (in ₹)** | 100 - 120 | 120 - 140 | 140 - 160 | 160 - 180 | 180 - 200 |
|---|---|---|---|---|---|
| **Number of workers** | 12 | 14 | 8 | 6 | 10 |

Convert the distribution above to a less than type cumulative frequency distribution, and draw its ogive.

2. During the medical check-up of 35 students of a class, their weights were recorded as follows:

| **Weight (in kg)** | **Number of students** |
|---|---|
| Less than 38 | 0 |
| Less than 40 | 3 |
| Less than 42 | 5 |
| Less than 44 | 9 |
| Less than 46 | 14 |
| Less than 48 | 28 |
| Less than 50 | 32 |
| Less than 52 | 35 |

Draw a less than type ogive for the given data. Hence obtain the median weight from the graph and verify the result by using the formula.

3. The following table gives production yield per hectare of wheat of 100 farms of a village.

| **Production yield (in kg/ha)** | 50 - 55 | 55 - 60 | 60 - 65 | 65 - 70 | 70 - 75 | 75 - 80 |
|---|---|---|---|---|---|---|
| **Number of farms** | 2 | 8 | 12 | 24 | 38 | 16 |

Change the distribution to a more than type distribution, and draw its ogive.

## 14.6 Summary

In this chapter, you have studied the following points:

1. The mean for grouped data can be found by :

(i) the direct method : $\bar{x} = \dfrac{\Sigma f_i x_i}{\Sigma f_i}$

(ii) the assumed mean method : $\bar{x} = a + \frac{\Sigma f_i d_i}{\Sigma f_i}$

(iii) the step deviation method : $\bar{x} = a + \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right) \times h$,

with the assumption that the frequency of a class is centred at its mid-point, called its class mark.

**2.** The mode for grouped data can be found by using the formula:

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$$

where symbols have their usual meanings.

**3.** The cumulative frequency of a class is the frequency obtained by adding the frequencies of all the classes preceding the given class.

**4.** The median for grouped data is formed by using the formula:

$$\text{Median} = l + \left(\frac{\frac{n}{2} - \text{cf}}{f}\right) \times h,$$

where symbols have their usual meanings.

**5.** Representing a cumulative frequency distribution graphically as a cumulative frequency curve, or an ogive of the less than type and of the more than type.

**6.** The median of grouped data can be obtained graphically as the $x$-coordinate of the point of intersection of the two ogives for this data.

## A NOTE TO THE READER

For calculating mode and median for grouped data, it should be ensured that the class intervals are continuous before applying the formulae. Same condition also apply for construction of an ogive. Further, in case of ogives, the scale may not be the same on both the axes.

1062CH15

# PROBABILITY 15

*The theory of probabilities and the theory of errors now constitute a formidable body of great mathematical interest and of great practical importance.*

– R.S. Woodward

## 15.1 Introduction

In Class IX, you have studied about experimental (or empirical) probabilities of events which were based on the results of actual experiments. We discussed an experiment of tossing a coin 1000 times in which the frequencies of the outcomes were as follows:

Head : 455 Tail : 545

Based on this experiment, the empirical probability of a head is $\frac{455}{1000}$, i.e., 0.455 and that of getting a tail is 0.545. (Also see Example 1, Chapter 15 of Class IX Mathematics Textbook.) Note that these probabilities are based on the results of an actual experiment of tossing a coin 1000 times. For this reason, they are called *experimental* or *empirical probabilities*. In fact, experimental probabilities are based on the results of actual experiments and adequate recordings of the happening of the events. Moreover, these probabilities are only 'estimates'. If we perform the same experiment for another 1000 times, we may get different data giving different probability estimates.

In Class IX, you tossed a coin many times and noted the number of times it turned up heads (or tails) (refer to Activities 1 and 2 of Chapter 15). You also noted that as the number of tosses of the coin increased, the experimental probability of getting a head (or tail) came closer and closer to the number $\frac{1}{2}$. Not only you, but many other persons from different parts of the world have done this kind of experiment and recorded the number of heads that turned up.

For example, the eighteenth century French naturalist Comte de Buffon tossed a coin 4040 times and got 2048 heads. The experimental probabilility of getting a head, in this case, was $\frac{2048}{4040}$ , i.e., 0.507. J.E. Kerrich, from Britain, recorded 5067 heads in 10000 tosses of a coin. The experimental probability of getting a head, in this case, was $\frac{5067}{10000} = 0.5067$. Statistician Karl Pearson spent some more time, making 24000 tosses of a coin. He got 12012 heads, and thus, the experimental probability of a head obtained by him was 0.5005.

Now, suppose we ask, 'What will the experimental probability of a head be if the experiment is carried on upto, say, one million times? Or 10 million times? And so on?' You would intuitively feel that as the number of tosses increases, the experimental probability of a head (or a tail) seems to be settling down around the number 0.5 , i.e., $\frac{1}{2}$, which is what we call the *theoretical probability* of getting a head (or getting a tail), as you will see in the next section. In this chapter, we provide an introduction to the theoretical (also called classical) probability of an event, and discuss simple problems based on this concept.

## 15.2 Probability — A Theoretical Approach

Let us consider the following situation :

Suppose a coin is tossed *at random*.

> When we speak of a coin, we assume it to be 'fair', that is, it is symmetrical so that there is no reason for it to come down more often on one side than the other. We call this property of the coin as being 'unbiased'. By the phrase 'random toss', we mean that the coin is allowed to fall freely without any *bias* or *interference*.

We know, in advance, that the coin can only land in one of two possible ways — either head up or tail up (we dismiss the possibility of its 'landing' on its edge, which may be possible, for example, if it falls on sand). We can reasonably assume that each outcome, head or tail, is *as likely to occur as the other.* We refer to this by saying that *the outcomes* head and tail, *are equally likely*.

For another example of equally likely outcomes, suppose we throw a die once. For us, a die will always mean a fair die. What are the possible outcomes? They are 1, 2, 3, 4, 5, 6. Each number has the same possibility of showing up. So the *equally likely outcomes* of throwing a die are 1, 2, 3, 4, 5 and 6.

Are the outcomes of every experiment equally likely? Let us see.

Suppose that a bag contains 4 red balls and 1 blue ball, and you draw a ball without looking into the bag. What are the outcomes? Are the outcomes — a red ball and a blue ball equally likely? Since there are 4 red balls and only one blue ball, you would agree that you are more likely to get a red ball than a blue ball. So, the outcomes (a red ball or a blue ball) are *not* equally likely. However, the outcome of drawing a ball of any colour from the bag is equally likely. So, all experiments do not necessarily have equally likely outcomes.

However, in this chapter, from now on, we will **assume that all the experiments have equally likely outcomes.**

In Class IX, we defined the experimental or empirical probability P(E) of an event E as

$$P(E) = \frac{\text{Number of trials in which the event happened}}{\text{Total number of trials}}$$

The empirical interpretation of probability can be applied to every event associated with an experiment which can be repeated a large number of times. The requirement of repeating an experiment has some limitations, as it may be very expensive or unfeasible in many situations. Of course, it worked well in coin tossing or die throwing experiments. But how about repeating the experiment of launching a satellite in order to compute the empirical probability of its failure during launching, or the repetition of the phenomenon of an earthquake to compute the empirical probability of a multi-storeyed building getting destroyed in an earthquake?

In experiments where we are prepared to make certain assumptions, the repetition of an experiment can be avoided, as the assumptions help in directly calculating the exact (theoretical) probability. The assumption of equally likely outcomes (which is valid in many experiments, as in the two examples above, of a coin and of a die) is one such assumption that leads us to the following definition of probability of an event.

The **theoretical probability** (also called **classical probability**) of an event E, written as P(E), is defined as

$$\mathbf{P(E)} = \frac{\textbf{Number of outcomes favourable to E}}{\textbf{Number of all possible outcomes of the experiment}},$$

where we assume that the outcomes of the experiment are *equally likely*.

We will briefly refer to theoretical probability as probability.

This definition of probability was given by Pierre Simon Laplace in 1795.

Probability theory had its origin in the 16th century when an Italian physician and mathematician J.Cardan wrote the first book on the subject, *The Book on Games of Chance*. Since its inception, the study of probability has attracted the attention of great mathematicians. James Bernoulli (1654 – 1705), A. de Moivre (1667 – 1754), and Pierre Simon Laplace are among those who made significant contributions to this field. Laplace's *Theorie Analytique des Probabilités,* 1812, is considered to be the greatest contribution by a single person to the theory of probability. In recent years, probability has been used extensively in many areas such as biology, economics, genetics, physics, sociology etc.

**Pierre Simon Laplace (1749 – 1827)**

Let us find the probability for some of the events associated with experiments where the equally likely assumption holds.

**Example 1 :** Find the probability of getting a head when a coin is tossed once. Also find the probability of getting a tail.

**Solution :** In the experiment of tossing a coin once, the number of possible outcomes is two — Head (H) and Tail (T). Let E be the event 'getting a head'. The number of outcomes favourable to E, (i.e., of getting a head) is 1. Therefore,

$$P(E) = P\text{ (head)} = \frac{\text{Number of outcomes favourable to E}}{\text{Number of all possible outcomes}} = \frac{1}{2}$$

Similarly, if F is the event 'getting a tail', then

$$P(F) = P(\text{tail}) = \frac{1}{2} \quad \text{(Why ?)}$$

**Example 2 :** A bag contains a red ball, a blue ball and a yellow ball, all the balls being of the same size. Kritika takes out a ball from the bag without looking into it. What is the probability that she takes out the

(i) yellow ball? (ii) red ball? (iii) blue ball?

**Solution :** Kritika takes out a ball from the bag without looking into it. So, it is equally likely that she takes out any one of them.

Let Y be the event 'the ball taken out is yellow', B be the event 'the ball taken out is blue', and R be the event 'the ball taken out is red'.

Now, the number of possible outcomes = 3.

(i) The number of outcomes favourable to the event Y = 1.

So, $$P(Y) = \frac{1}{3}$$

Similarly, (ii) $P(R) = \frac{1}{3}$ and (iii) $P(B) = \frac{1}{3}$.

**Remarks :**

1. An event having only one outcome of the experiment is called an *elementary event.* In Example 1, both the events E and F are elementary events. Similarly, in Example 2, all the three events, Y, B and R are elementary events.

2. In Example 1, we note that : P(E) + P(F) = 1

In Example 2, we note that : P(Y) + P(R) + P(B) = 1

Observe that **the sum of the probabilities of all the elementary events of an experiment** is 1. This is true in general also.

**Example 3 :** Suppose we throw a die once. (i) What is the probability of getting a number greater than 4 ? (ii) What is the probability of getting a number less than or equal to 4 ?

**Solution :** (i) Here, let E be the event 'getting a number greater than 4'. The number of possible outcomes is six : 1, 2, 3, 4, 5 and 6, and the outcomes favourable to E are 5 and 6. Therefore, the number of outcomes favourable to E is 2. So,

$$P(E) = P(\text{number greater than } 4) = \frac{2}{6} = \frac{1}{3}$$

(ii) Let F be the event 'getting a number less than or equal to 4'.

Number of possible outcomes = 6

Outcomes favourable to the event F are 1, 2, 3, 4.

So, the number of outcomes favourable to F is 4.

Therefore, $$P(F) = \frac{4}{6} = \frac{2}{3}$$

Are the events E and F in the example above elementary events? No, they are **not** because the event E has 2 outcomes and the event F has 4 outcomes.

**Remarks :** From Example 1, we note that

$$P(E) + P(F) = \frac{1}{2} + \frac{1}{2} = 1 \qquad (1)$$

where E is the event 'getting a head' and F is the event 'getting a tail'.

From (i) and (ii) of Example 3, we also get

$$P(E) + P(F) = \frac{1}{3} + \frac{2}{3} = 1 \qquad (2)$$

where E is the event 'getting a number >4' and F is the event 'getting a number ≤ 4'.

Note that getting a number *not* greater than 4 is same as getting a number less than or equal to 4, and vice versa.

In (1) and (2) above, is F not the same as 'not E'? Yes, it is. We denote the event 'not E' by $\overline{E}$.

So, $\quad P(E) + P(\text{not } E) = 1$

i.e., $\quad P(E) + P(\overline{E}) = 1$, which gives us $P(\overline{E}) = 1 - P(E)$.

**In general, it is true that for an event E,**

$$\mathbf{P(\overline{E}) = 1 - P(E)}$$

The event $\overline{E}$, representing 'not E', is called the **complement** of the event E. We also say that E and $\overline{E}$ are **complementary** events.

Before proceeding further, let us try to find the answers to the following questions:

(i) What is the probability of getting a number 8 in a single throw of a die?

(ii) What is the probability of getting a number less than 7 in a single throw of a die?

**Let us answer (i) :**

We know that there are only six possible outcomes in a single throw of a die. These outcomes are 1, 2, 3, 4, 5 and 6. Since no face of the die is marked 8, so there is no outcome favourable to 8, i.e., the number of such outcomes is zero. In other words, getting 8 in a single throw of a die, is *impossible*.

So, $\quad P(\text{getting } 8) = \frac{0}{6} = 0$

That is, the probability of an event which is *impossible* to occur is 0. Such an event is called an **impossible event**.

**Let us answer (ii) :**

Since every face of a die is marked with a number less than 7, it is *sure* that we will always get a number less than 7 when it is thrown once. So, the number of favourable outcomes is the same as the number of all possible outcomes, which is 6.

Therefore, $\quad P(E) = P(\text{getting a number less than 7}) = \frac{6}{6} = 1$

So, the probability of an event which is *sure* (or *certain*) to occur is 1. Such an event is called a **sure event** or a **certain event**.

**Note :** From the definition of the probability P(E), we see that the numerator (number of outcomes favourable to the event E) is always less than or equal to the denominator (the number of all possible outcomes). Therefore,

$$\mathbf{0 \le P(E) \le 1}$$

Now, let us take an example related to playing cards. Have you seen a deck of playing cards? It consists of 52 cards which are divided into 4 suits of 13 cards each— spades (♠), hearts (♥), diamonds (♦) and clubs (♣). Clubs and spades are of black colour, while hearts and diamonds are of red colour. The cards in each suit are ace, king, queen, jack, 10, 9, 8, 7, 6, 5, 4, 3 and 2. Kings, queens and jacks are called *face cards*.

**Example 4 :** One card is drawn from a well-shuffled deck of 52 cards. Calculate the probability that the card will

(i) be an ace,

(ii) not be an ace.

**Solution :** Well-shuffling ensures *equally likely* outcomes.

(i) There are 4 aces in a deck. Let E be the event 'the card is an ace'.

The number of outcomes favourable to E = 4

The number of possible outcomes = 52 (Why ?)

Therefore, $\quad P(E) = \frac{4}{52} = \frac{1}{13}$

(ii) Let F be the event 'card drawn is not an ace'.

The number of outcomes favourable to the event F = 52 – 4 = 48 (Why?)

The number of possible outcomes = 52

Therefore, $P(F) = \frac{48}{52} = \frac{12}{13}$

**Remark :** Note that F is nothing but $\overline{E}$. Therefore, we can also calculate P(F) as follows: $P(F) = P(\overline{E}) = 1 - P(E) = 1 - \frac{1}{13} = \frac{12}{13}$.

**Example 5 :** Two players, Sangeeta and Reshma, play a tennis match. It is known that the probability of Sangeeta winning the match is 0.62. What is the probability of Reshma winning the match?

**Solution :** Let S and R denote the events that Sangeeta wins the match and Reshma wins the match, respectively.

The probability of Sangeeta's winning = P(S) = 0.62 (given)

The probability of Reshma's winning = P(R) = 1 – P(S)

[As the events R and S are complementary]

= 1 – 0.62 = 0.38

**Example 6 :** Savita and Hamida are friends. What is the probability that both will have (i) different birthdays? (ii) the same birthday? (ignoring a leap year).

**Solution :** Out of the two friends, one girl, say, Savita's birthday can be any day of the year. Now, Hamida's birthday can also be any day of 365 days in the year.

We assume that these 365 outcomes are equally likely.

(i) If Hamida's birthday is different from Savita's, the number of favourable outcomes for her birthday is 365 – 1 = 364

So, P (Hamida's birthday is different from Savita's birthday) = $\frac{364}{365}$

(ii) P(Savita and Hamida have the same birthday)

= 1 – P (both have different birthdays)

$= 1 - \frac{364}{365}$ [Using $P(\overline{E}) = 1 - P(E)$]

$= \frac{1}{365}$

**Example 7 :** There are 40 students in Class X of a school of whom 25 are girls and 15 are boys. The class teacher has to select one student as a class representative. She writes the name of each student on a separate card, the cards being identical. Then she puts cards in a bag and stirs them thoroughly. She then draws one card from the bag. What is the probability that the name written on the card is the name of (i) a girl? (ii) a boy?

**Solution :** There are 40 students, and only one name card has to be chosen.

(i) The number of all possible outcomes is 40

The number of outcomes favourable for a card with the name of a girl = 25 (Why?)

Therefore, P (card with name of a girl) = P(Girl) = $\frac{25}{40} = \frac{5}{8}$

(ii) The number of outcomes favourable for a card with the name of a boy = 15 (Why?)

Therefore, P(card with name of a boy) = P(Boy) = $\frac{15}{40} = \frac{3}{8}$

**Note :** We can also determine P(Boy), by taking

$$P(\text{Boy}) = 1 - P(\text{not Boy}) = 1 - P(\text{Girl}) = 1 - \frac{5}{8} = \frac{3}{8}$$

**Example 8 :** A box contains 3 blue, 2 white, and 4 red marbles. If a marble is drawn at *random* from the box, what is the probability that it will be

(i) white? (ii) blue? (iii) red?

**Solution :** Saying that a marble is drawn at random is a short way of saying that all the marbles are equally likely to be drawn. Therefore, the

number of possible outcomes = 3 +2 + 4 = 9 (Why?)

Let W denote the event 'the marble is white', B denote the event 'the marble is blue' and R denote the event 'marble is red'.

(i) The number of outcomes favourable to the event W = 2

So, P(W) = $\frac{2}{9}$

Similarly, (ii) P(B) = $\frac{3}{9} = \frac{1}{3}$ and (iii) P(R) = $\frac{4}{9}$

Note that P(W) + P(B) + P(R) = 1.

**Example 9 :** Harpreet tosses two different coins simultaneously (say, one is of ₹1 and other of ₹ 2). What is the probability that she gets *at least* one head?

**Solution :** We write H for 'head' and T for 'tail'. When two coins are tossed simultaneously, the possible outcomes are (H, H), (H, T), (T, H), (T, T), which are all *equally likely*. Here (H, H) means head up on the first coin (say on ₹ 1) and head up on the second coin (₹ 2). Similarly (H, T) means head up on the first coin and tail up on the second coin and so on.

The outcomes favourable to the event E, 'at least one head' are (H, H), (H, T) and (T, H). (Why?)

So, the number of outcomes favourable to E is 3.

Therefore, $\qquad P(E) = \frac{3}{4}$

i.e., the probability that Harpreet gets at least one head is $\frac{3}{4}$.

**Note :** You can also find P(E) as follows:

$$P(E) = 1 - P(\bar{E}) = 1 - \frac{1}{4} = \frac{3}{4} \qquad \left(\text{Since } P(\bar{E}) = P(\text{no head}) = \frac{1}{4}\right)$$

Did you observe that in all the examples discussed so far, the number of possible outcomes in each experiment was finite? If not, check it now.

There are many experiments in which the outcome is any number between two given numbers, or in which the outcome is every point within a circle or rectangle, etc. Can you now count the number of all possible outcomes? As you know, this is not possible since there are infinitely many numbers between two given numbers, or there are infinitely many points within a circle. So, the definition of (theoretical) probability which you have learnt so far cannot be applied in the present form. What is the way out? To answer this, let us consider the following example :

**Example 10* :** In a musical chair game, the person playing the music has been advised to stop playing the music at any time within 2 minutes after she starts playing. What is the probability that the music will stop within the first half-minute after starting?

**Solution :** Here the possible outcomes are all the numbers between 0 and 2. This is the portion of the number line from 0 to 2 (see Fig. 15.1).


**Fig. 15.1**

* Not from the examination point of view.

Let E be the event that 'the music is stopped within the first half-minute'.

The outcomes favourable to E are points on the number line from 0 to $\frac{1}{2}$.

The distance from 0 to 2 is 2, while the distance from 0 to $\frac{1}{2}$ is $\frac{1}{2}$.

Since all the outcomes are equally likely, we can argue that, of the total distance of 2, the distance favourable to the event E is $\frac{1}{2}$.

So, $$P(E) = \frac{\text{Distance favourable to the event E}}{\text{Total distance in which outcomes can lie}} = \frac{\frac{1}{2}}{2} = \frac{1}{4}$$

Can we now extend the idea of Example 10 for finding the probability as the ratio of the favourable area to the total area?

**Example 11* :** A missing helicopter is reported to have crashed somewhere in the rectangular region shown in Fig. 15.2. What is the probability that it crashed inside the lake shown in the figure?

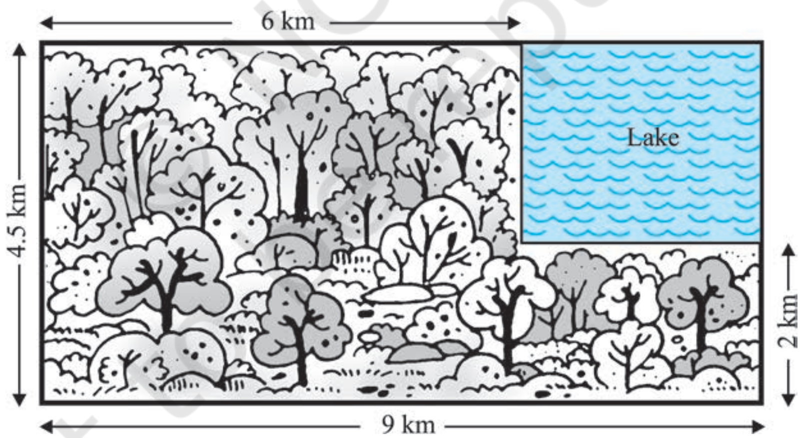



**Fig. 15.2**

**Solution :** The helicopter is equally likely to crash anywhere in the region.

Area of the entire region where the helicopter can crash

$$= (4.5 \times 9)\ \text{km}^2 = 40.5\ \text{km}^2$$

* Not from the examination point of view.

Area of the lake = (2.5 × 3) km$^2$ = 7.5 km$^2$

Therefore, P (helicopter crashed in the lake) = $\frac{7.5}{40.5} = \frac{75}{405} = \frac{5}{27}$

**Example 12 :** A carton consists of 100 shirts of which 88 are good, 8 have minor defects and 4 have major defects. Jimmy, a trader, will only accept the shirts which are good, but Sujatha, another trader, will only reject the shirts which have major defects. One shirt is drawn at random from the carton. What is the probability that

(i) it is acceptable to Jimmy?

(ii) it is acceptable to Sujatha?

**Solution :** One shirt is drawn at random from the carton of 100 shirts. Therefore, there are 100 equally likely outcomes.

(i) The number of outcomes favourable (i.e., acceptable) to Jimmy = 88 (Why?)

Therefore, P (shirt is acceptable to Jimmy) = $\frac{88}{100}$ = 0.88

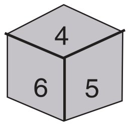
(ii) The number of outcomes favourable to Sujatha = 88 + 8 = 96 (Why?)

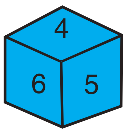
So, P (shirt is acceptable to Sujatha) = $\frac{96}{100}$ = 0.96

**Example 13 :** Two dice, one blue and one grey, are thrown at the same time. Write down all the possible outcomes. What is the probability that the sum of the two numbers appearing on the top of the dice is

(i) 8? (ii) 13? (iii) less than or equal to 12?

**Solution :** When the blue die shows '1', the grey die could show any one of the numbers 1, 2, 3, 4, 5, 6. The same is true when the blue die shows '2', '3', '4', '5' or '6'. The possible outcomes of the experiment are listed in the table below; the first number in each ordered pair is the number appearing on the blue die and the second number is that on the grey die.




| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | (1, 1) | (1, 2) | (1, 3) | (1, 4) | (1, 5) | (1, 6) |
| 2 | (2, 1) | (2, 2) | (2, 3) | (2, 4) | (2, 5) | (2, 6) |
| 3 | (3, 1) | (3, 2) | (3, 3) | (3, 4) | (3, 5) | (3, 6) |
| 4 | (4, 1) | (4, 2) | (4, 3) | (4, 4) | (4, 5) | (4, 6) |
| 5 | (5, 1) | (5, 2) | (5, 3) | (5, 4) | (5, 5) | (5, 6) |
| 6 | (6, 1) | (6, 2) | (6, 3) | (6, 4) | (6, 5) | (6, 6) |

**Fig. 15.3**

Note that the pair (1, 4) is different from (4, 1). (Why?)

So, the number of possible outcomes = $6 \times 6 = 36$.

(i) The outcomes favourable to the event 'the sum of the two numbers is 8' denoted by E, are: (2, 6), (3, 5), (4, 4), (5, 3), (6, 2) (see Fig. 15.3)

i.e., the number of outcomes favourable to E = 5.

Hence, $$P(E) = \frac{5}{36}$$

(ii) As you can see from Fig. 15.3, there is no outcome favourable to the event F, 'the sum of two numbers is 13'.

So, $$P(F) = \frac{0}{36} = 0$$

(iii) As you can see from Fig. 15.3, all the outcomes are favourable to the event G, 'sum of two numbers $\leq 12$'.

So, $$P(G) = \frac{36}{36} = 1$$

## EXERCISE 15.1

**1.** Complete the following statements:

(i) Probability of an event E + Probability of the event 'not E' = ________ .

(ii) The probability of an event that cannot happen is ________ . Such an event is called ________ .

(iii) The probability of an event that is certain to happen is ________ . Such an event is called ________ .

(iv) The sum of the probabilities of all the elementary events of an experiment is ________ .

(v) The probability of an event is greater than or equal to ________ and less than or equal to ________ .

**2.** Which of the following experiments have equally likely outcomes? Explain.

(i) A driver attempts to start a car. The car starts or does not start.

(ii) A player attempts to shoot a basketball. She/he shoots or misses the shot.

(iii) A trial is made to answer a true-false question. The answer is right or wrong.

(iv) A baby is born. It is a boy or a girl.

**3.** Why is tossing a coin considered to be a fair way of deciding which team should get the ball at the beginning of a football game?

**4.** Which of the following cannot be the probability of an event?

(A) $\frac{2}{3}$ (B) $-1.5$ (C) 15% (D) 0.7

**5.** If P(E) = 0.05, what is the probability of 'not E'?

**6.** A bag contains lemon flavoured candies only. Malini takes out one candy without looking into the bag. What is the probability that she takes out

(i) an orange flavoured candy?

(ii) a lemon flavoured candy?

**7.** It is given that in a group of 3 students, the probability of 2 students not having the same birthday is 0.992. What is the probability that the 2 students have the same birthday?

**8.** A bag contains 3 red balls and 5 black balls. A ball is drawn at random from the bag. What is the probability that the ball drawn is (i) red ? (ii) not red?

**9.** A box contains 5 red marbles, 8 white marbles and 4 green marbles. One marble is taken out of the box at random. What is the probability that the marble taken out will be (i) red ? (ii) white ? (iii) not green?

**10.** A piggy bank contains hundred 50p coins, fifty ₹ 1 coins, twenty ₹ 2 coins and ten ₹ 5 coins. If it is equally likely that one of the coins will fall out when the bank is turned upside down, what is the probability that the coin (i) will be a 50 p coin ? (ii) will not be a ₹ 5 coin?

**11.** Gopi buys a fish from a shop for his aquarium. The shopkeeper takes out one fish at random from a tank containing 5 male fish and 8 female fish (see Fig. 15.4). What is the probability that the fish taken out is a male fish?

**Fig. 15.4**

**12.** A game of chance consists of spinning an arrow which comes to rest pointing at one of the numbers 1, 2, 3, 4, 5, 6, 7, 8 (see Fig. 15.5 ), and these are equally likely outcomes. What is the probability that it will point at

(i) 8 ?

(ii) an odd number?

(iii) a number greater than 2?

(iv) a number less than 9?

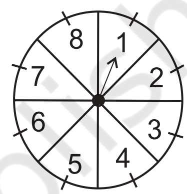



**Fig. 15.5**

**13.** A die is thrown once. Find the probability of getting

(i) a prime number; (ii) a number lying between 2 and 6; (iii) an odd number.

**14.** One card is drawn from a well-shuffled deck of 52 cards. Find the probability of getting

(i) a king of red colour (ii) a face card (iii) a red face card

(iv) the jack of hearts (v) a spade (vi) the queen of diamonds

**15.** Five cards—the ten, jack, queen, king and ace of diamonds, are well-shuffled with their face downwards. One card is then picked up at random.

(i) What is the probability that the card is the queen?

(ii) If the queen is drawn and put aside, what is the probability that the second card picked up is (a) an ace? (b) a queen?

**16.** 12 defective pens are accidentally mixed with 132 good ones. It is not possible to just look at a pen and tell whether or not it is defective. One pen is taken out at random from this lot. Determine the probability that the pen taken out is a good one.

**17.** (i) A lot of 20 bulbs contain 4 defective ones. One bulb is drawn at random from the lot. What is the probability that this bulb is defective?

(ii) Suppose the bulb drawn in (i) is not defective and is not replaced. Now one bulb is drawn at random from the rest. What is the probability that this bulb is not defective ?

**18.** A box contains 90 discs which are numbered from 1 to 90. If one disc is drawn at random from the box, find the probability that it bears (i) a two-digit number (ii) a perfect square number (iii) a number divisible by 5.

**19.** A child has a die whose six faces show the letters as given below:

| A | B | C | D | E | A |
|---|---|---|---|---|---|

The die is thrown once. What is the probability of getting (i) A? (ii) D?

**20*.** Suppose you drop a die at random on the rectangular region shown in Fig. 15.6. What is the probability that it will land inside the circle with diameter 1m?

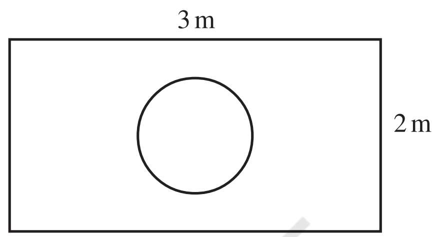



**Fig. 15.6**

**21.** A lot consists of 144 ball pens of which 20 are defective and the others are good. Nuri will buy a pen if it is good, but will not buy if it is defective. The shopkeeper draws one pen at random and gives it to her. What is the probability that

(i) She will buy it ?

(ii) She will not buy it ?

**22.** Refer to Example 13. (i) Complete the following table:

| **Event : 'Sum on 2 dice'** | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Probability** | $\frac{1}{36}$ | | | | | | $\frac{5}{36}$ | | | | $\frac{1}{36}$ |

(ii) A student argues that 'there are 11 possible outcomes 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12. Therefore, each of them has a probability $\frac{1}{11}$. Do you agree with this argument? Justify your answer.

**23.** A game consists of tossing a one rupee coin 3 times and noting its outcome each time. Hanif wins if all the tosses give the same result i.e., three heads or three tails, and loses otherwise. Calculate the probability that Hanif will lose the game.

**24.** A die is thrown twice. What is the probability that

(i) 5 will not come up either time? (ii) 5 will come up at least once?

[**Hint :** Throwing a die twice and throwing two dice simultaneously are treated as the same experiment]

* Not from the examination point of view.

**25.** Which of the following arguments are correct and which are not correct? Give reasons for your answer.

(i) If two coins are tossed simultaneously there are three possible outcomes—two heads, two tails or one of each. Therefore, for each of these outcomes, the probability is $\frac{1}{3}$.

(ii) If a die is thrown, there are two possible outcomes—an odd number or an even number. Therefore, the probability of getting an odd number is $\frac{1}{2}$.

## EXERCISE 15.2 (Optional)*

**1.** Two customers Shyam and Ekta are visiting a particular shop in the same week (Tuesday to Saturday). Each is equally likely to visit the shop on any day as on another day. What is the probability that both will visit the shop on (i) the same day? (ii) consecutive days? (iii) different days?

**2.** A die is numbered in such a way that its faces show the numbers 1, 2, 2, 3, 3, 6. It is thrown two times and the total score in two throws is noted. Complete the following table which gives a few values of the total score on the two throws:

Number in first throw

Number in second throw (rows)

| + | 1 | 2 | 2 | 3 | 3 | 6 |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 3 | 4 | 4 | 7 |
| 2 | 3 | 4 | 4 | 5 | 5 | 8 |
| 2 | | | | | 5 | |
| 3 | | | | | | |
| 3 | | | 5 | | | 9 |
| 6 | 7 | 8 | 8 | 9 | 9 | 12 |

What is the probability that the total score is

(i) even? (ii) 6? (iii) at least 6?

**3.** A bag contains 5 red balls and some blue balls. If the probability of drawing a blue ball is *double* that of a red ball, determine the number of blue balls in the bag.

**4.** A box contains 12 balls out of which $x$ are black. If one ball is drawn at random from the box, what is the probability that it will be a black ball?

If 6 more black balls are put in the box, the probability of drawing a black ball is now double of what it was before. Find $x$.

* These exercises are not from the examination point of view.

**5.** A jar contains 24 marbles, some are green and others are blue. If a marble is drawn at random from the jar, the probability that it is green is $\frac{2}{3}$. Find the number of blue balls in the jar.

## 15.3 Summary

In this chapter, you have studied the following points :

1. The difference between experimental probability and theoretical probability.
2. The theoretical (classical) probability of an event E, written as P(E), is defined as

$$P(E) = \frac{\text{Number of outcomes favourable to E}}{\text{Number of all possible outcomes of the experiment}}$$

where we assume that the outcomes of the experiment are equally likely.

3. The probability of a sure event (or certain event) is 1.
4. The probability of an impossible event is 0.
5. The probability of an event E is a number P(E) such that

$$0 \le P(E) \le 1$$

6. An event having only one outcome is called an elementary event. The sum of the probabilities of all the elementary events of an experiment is 1.
7. For any event E, P (E) + P ($\bar{E}$) = 1, where $\bar{E}$ stands for 'not E'. E and $\bar{E}$ are called complementary events.

### A NOTE TO THE READER

The experimental or empirical probability of an event is based on what has actually happened while the theoretical probability of the event attempts to predict what will happen on the basis of certain assumptions. As the number of trials in an experiment, go on increasing we may expect the experimental and theoretical probabilities to be nearly the same.