8. A random survey of the number of children of various age groups playing in a park was found as follows:

Age (in years)	Number of children
1 - 2	5
2 - 3	3
3 - 5	6
5 - 7	12
7 - 10	9
10 - 15	10
15 - 17	4

Draw a histogram to represent the data above.

9. 100 surnames were randomly picked up from a local telephone directory and a frequency distribution of the number of letters in the English alphabet in the surnames was found as follows:

Number of letters	Number of surnames
1 - 4	6
4 - 6	30
6 - 8	44
8 - 12	16
12 - 20	4

(i) Draw a histogram to depict the given information.

(ii) Write the class interval in which the maximum number of surnames lie.

14.5 Measures of Central Tendency

Earlier in this chapter, we represented the data in various forms through frequency distribution tables, bar graphs, histograms and frequency polygons. Now, the question arises if we always need to study all the data to 'make sense' of it, or if we can make out some important features of it by considering only certain representatives of the data. This is possible, by using measures of central tendency or averages.

Consider a situation when two students Mary and Hari received their test copies. The test had five questions, each carrying ten marks. Their scores were as follows:

Question Numbers	1	2	3	4	5
Mary's score	10	8	9	8	7
Hari's score	4	7	10	10	10

Upon getting the test copies, both of them found their average scores as follows:

$$\text{Mary's average score} = \frac{42}{5} = 8.4$$

$$\text{Hari's average score} = \frac{41}{5} = 8.2$$

Since Mary's average score was more than Hari's, Mary claimed to have performed better than Hari, but Hari did not agree. He arranged both their scores in ascending order and found out the middle score as given below:

Mary's Score	7	8	(8)	9	10
Hari's Score	4	7	(10)	10	10

Hari said that since his middle-most score was 10, which was higher than Mary's middle-most score, that is 8, his performance should be rated better.

But Mary was not convinced. To convince Mary, Hari tried out another strategy. He said he had scored 10 marks more often (3 times) as compared to Mary who scored 10 marks only once. So, his performance was better.

Now, to settle the dispute between Hari and Mary, let us see the three measures they adopted to make their point.

The average score that Mary found in the first case is the *mean*. The 'middle' score that Hari was using for his argument is the *median*. The most often scored mark that Hari used in his second strategy is the *mode*.

Now, let us first look at the mean in detail.

The **mean** (or **average**) of a number of observations is the sum of the values of all the observations divided by the total number of observations.

It is denoted by the symbol $\overline{x}$, read as 'x bar'.

Let us consider an example.

Example 10 : 5 people were asked about the time in a week they spend in doing social work in their community. They said 10, 7, 13, 20 and 15 hours, respectively.

Find the mean (or average) time in a week devoted by them for social work.

Solution : We have already studied in our earlier classes that the mean of a certain number of observations is equal to $\dfrac{\text{Sum of all the observations}}{\text{Total number of observations}}$. To simplify our working of finding the mean, let us use a variable x_i to denote the ith observation. In this case, i can take the values from 1 to 5. So our first observation is x_1, second observation is x_2, and so on till x_5.

Also $x_1 = 10$ means that the value of the first observation, denoted by x_1, is 10. Similarly, $x_2 = 7$, $x_3 = 13$, $x_4 = 20$ and $x_5 = 15$.

$$\text{Therefore, the mean } \bar{x} = \frac{\text{Sum of all the observations}}{\text{Total number of observations}}$$

$$= \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$

$$= \frac{10 + 7 + 13 + 20 + 15}{5} = \frac{65}{5} = 13$$

So, the mean time spent by these 5 people in doing social work is 13 hours in a week.

Now, in case we are finding the mean time spent by 30 people in doing social work, writing $x_1 + x_2 + x_3 + \ldots + x_{30}$ would be a tedious job. We use the Greek symbol Σ (for the letter Sigma) for *summation*. Instead of writing $x_1 + x_2 + x_3 + \ldots + x_{30}$, we write $\sum_{i=1}^{30} x_i$, which is read as 'the sum of x_i *as* i varies from 1 to 30'.

So, $$\bar{x} = \frac{\sum_{i=1}^{30} x_i}{30}$$

Similarly, for n observations $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Example 11 : Find the mean of the marks obtained by 30 students of Class IX of a school, given in Example 2.

Solution : Now, $$\bar{x} = \frac{x_1 + x_2 + \ldots + x_{30}}{30}$$

$$\sum_{i=1}^{30} x_i = 10 + 20 + 36 + 92 + 95 + 40 + 50 + 56 + 60 + 70 + 92 + 88$$
$$80 + 70 + 72 + 70 + 36 + 40 + 36 + 40 + 92 + 40 + 50 + 50$$
$$56 + 60 + 70 + 60 + 60 + 88 = 1779$$

So, $$\bar{x} = \frac{1779}{30} = 59.3$$

Is the process not time consuming? Can we simplify it? Note that we have formed a frequency table for this data (see Table 14.1).

The table shows that 1 student obtained 10 marks, 1 student obtained 20 marks, 3 students obtained 36 marks, 4 students obtained 40 marks, 3 students obtained 50 marks, 2 students obtained 56 marks, 4 students obtained 60 marks, 4 students obtained 70 marks, 1 student obtained 72 marks, 1 student obtained 80 marks, 2 students obtained 88 marks, 3 students obtained 92 marks and 1 student obtained 95 marks.

So, the total marks obtained = (1 × 10) + (1 × 20) + (3 × 36) + (4 × 40) + (3 × 50) + (2 × 56) + (4 × 60) + (4 × 70) + (1 × 72) + (1 × 80) + (2 × 88) + (3 × 92) + (1 × 95)

= $f_1x_1 + \ldots + f_{13}x_{13}$, where f_i is the frequency of the ith entry inTable 14.1.

In brief, we write this as $\sum_{i=1}^{13} f_i x_i$.

So, the total marks obtained = $\sum_{i=1}^{13} f_i x_i$

= 10 + 20 + 108 + 160 + 150 + 112 + 240 + 280 + 72 + 80 + 176 + 276 + 95

= 1779

Now, the total number of observations

$$= \sum_{i=1}^{13} f_i$$

$$= f_1 + f_2 + \ldots + f_{13}$$

= 1 + 1 + 3 + 4 + 3 + 2 + 4 + 4 + 1 + 1 + 2 + 3 + 1

= 30

So, the mean $\bar{x}$ = $\dfrac{\text{Sum of all the observations}}{\text{Total number of observations}} = \left(\dfrac{\sum_{i=1}^{13} f_i x_i}{\sum_{i=1}^{13} f_i}\right)$

$$= \frac{1779}{30} = 59.3$$

This process can be displayed in the following table, which is a modified form of Table 14.1.

Table 14.12

Marks (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
	$\sum_{i=1}^{13} f_i = 30$	$\sum_{i=1}^{13} f_i x_i = 1779$

Thus, in the case of an ungrouped frequency distribution, you can use the formula

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

for calculating the mean.

Let us now move back to the situation of the argument between Hari and Mary, and consider the second case where Hari found his performance better by finding the middle-most score. As already stated, this measure of central tendency is called the *median*.

The **median** is that value of the given number of observations, which divides it into exactly two parts. So, when the data is arranged in ascending (or descending) order the median of ungrouped data is calculated as follows:

(i) When the number of observations (n) is odd, the median is the value of the $\left(\frac{n+1}{2}\right)^{th}$ observation. For example, if $n = 13$, the value of the $\left(\frac{13+1}{2}\right)^{th}$, i.e., the 7th observation will be the median [see Fig. 14.9 (i)].

(ii) When the number of observations (n) is even, the median is the mean of the $\left(\frac{n}{2}\right)^{th}$ and the $\left(\frac{n}{2}+1\right)^{th}$ observations. For example, if $n = 16$, the mean of the values of the $\left(\frac{16}{2}\right)^{th}$ and the $\left(\frac{16}{2}+1\right)^{th}$ observations, i.e., the mean of the values of the 8th and 9th observations will be the median [see Fig. 14.9 (ii)].

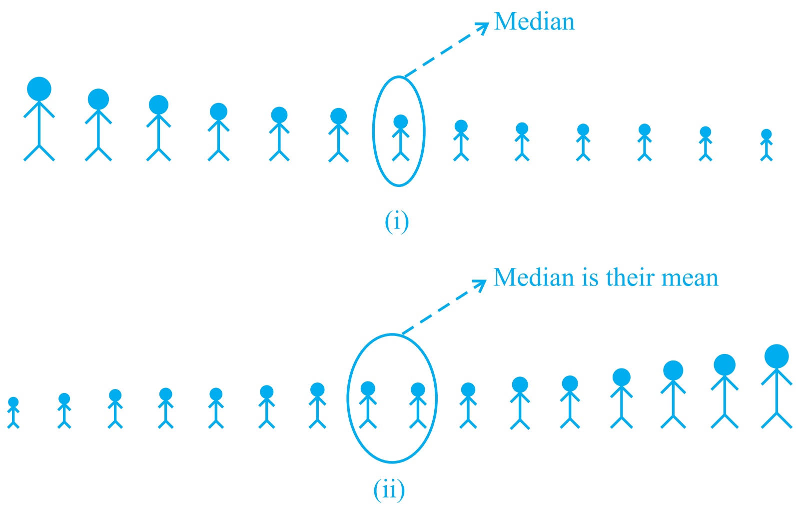

Fig. 14.9

Let us illustrate this with the help of some examples.

Example 12 : The heights (in cm) of 9 students of a class are as follows:

155 160 145 149 150 147 152 144 148

Find the median of this data.

Solution : First of all we arrange the data in ascending order, as follows:

144 145 147 148 149 150 152 155 160

Since the number of students is 9, an odd number, we find out the median by finding the height of the $\left(\frac{n+1}{2}\right)$th $=\left(\frac{9+1}{2}\right)$th = the 5th student, which is 149 cm.

So, the median, i.e., the medial height is 149 cm.

Example 13 : The points scored by a Kabaddi team in a series of matches are as follows:

17, 2, 7, 27, 15, 5, 14, 8, 10, 24, 48, 10, 8, 7, 18, 28

Find the median of the points scored by the team.

Solution : Arranging the points scored by the team in ascending order, we get

2, 5, 7, 7, 8, 8, 10, 10, 14, 15, 17, 18, 24, 27, 28, 48.

There are 16 terms. So there are two middle terms, i.e. the $\frac{16}{2}$th and $\left(\frac{16}{2}+1\right)$th, i.e., the 8th and 9th terms.

So, the median is the mean of the values of the 8th and 9th terms.

i.e, the median $=\frac{10+14}{2}=12$

So, the medial point scored by the Kabaddi team is 12.

Let us again go back to the unsorted dispute of Hari and Mary.

The third measure used by Hari to find the average was the *mode*.

The **mode** is that value of the observation which occurs most frequently, i.e., an observation with the maximum frequency is called the mode.

The readymade garment and shoe industries make great use of this measure of central tendency. Using the knowledge of mode, these industries decide which size of the product should be produced in large numbers.

Let us illustrate this with the help of an example.

Example 14 : Find the mode of the following marks (out of 10) obtained by 20 students:

4, 6, 5, 9, 3, 2, 7, 7, 6, 5, 4, 9, 10, 10, 3, 4, 7, 6, 9, 9

Solution : We arrange this data in the following form :

2, 3, 3, 4, 4, 4, 5, 5, 6, 6, 6, 7, 7, 7, 9, 9, 9, 9, 10, 10

Here 9 occurs most frequently, i.e., four times. So, the mode is 9.

Example 15 : Consider a small unit of a factory where there are 5 employees : a supervisor and four labourers. The labourers draw a salary of ` 5,000 per month each while the supervisor gets ` 15,000 per month. Calculate the mean, median and mode of the salaries of this unit of the factory.

Solution : Mean = $\frac{5000 + 5000 + 5000 + 5000 + 15000}{5} = \frac{35000}{5} = 7000$

So, the mean salary is ` 7000 per month.

To obtain the median, we arrange the salaries in ascending order:

5000, 5000, 5000, 5000, 15000

Since the number of employees in the factory is 5, the median is given by the $\left(\frac{5+1}{2}\right)$th $= \frac{6}{2}$th = 3rd observation. Therefore, the median is ` 5000 per month.

To find the mode of the salaries, i.e., the modal salary, we see that 5000 occurs the maximum number of times in the data 5000, 5000, 5000, 5000, 15000. So, the modal salary is ` 5000 per month.

Now compare the three measures of central tendency for the given data in the example above. You can see that the mean salary of ` 7000 does not give even an approximate estimate of any one of their wages, while the medial and modal salaries of ` 5000 represents the data more effectively.

Extreme values in the data affect the mean. This is one of the weaknesses of the mean. So, if the data has a few points which are very far from most of the other points, (like 1,7,8,9,9) then the mean is not a good representative of this data. Since the median and mode are not affected by extreme values present in the data, they give a better estimate of the average in such a situation.

Again let us go back to the situation of Hari and Mary, and compare the three measures of central tendency.

Measures of central tendency	Hari	Mary
Mean	8.2	8.4
Median	10	8
Mode	10	8

This comparison helps us in stating that these measures of central tendency are not sufficient for concluding which student is better. We require some more information to conclude this, which you will study about in the higher classes.

EXERCISE 14.4

1. The following number of goals were scored by a team in a series of 10 matches:

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

Find the mean, median and mode of these scores.

2. In a mathematics test given to 15 students, the following marks (out of 100) are recorded:

41, 39, 48, 52, 46, 62, 54, 40, 96, 52, 98, 40, 42, 52, 60

Find the mean, median and mode of this data.

3. The following observations have been arranged in ascending order. If the median of the data is 63, find the value of x.

29, 32, 48, 50, x, $x+2$, 72, 78, 84, 95

4. Find the mode of 14, 25, 14, 28, 18, 17, 18, 14, 23, 22, 14, 18.

5. Find the mean salary of 60 workers of a factory from the following table:

Salary (in `)	Number of workers
3000	16
4000	12
5000	10
6000	8
7000	6
8000	4
9000	3
10000	1
Total	60

6. Give one example of a situation in which

(i) the mean is an appropriate measure of central tendency.

(ii) the mean is not an appropriate measure of central tendency but the median is an appropriate measure of central tendency.

14.6 Summary

In this chapter, you have studied the following points:

1. Facts or figures, collected with a definite purpose, are called data.

2. Statistics is the area of study dealing with the presentation, analysis and interpretation of data.

3. How data can be presented graphically in the form of bar graphs, histograms and frequency polygons.

4. The three measures of central tendency for **ungrouped data** are:

(i) Mean : It is found by adding all the values of the observations and dividing it by the total number of observations. It is denoted by $\bar{x}$.

So, $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$. For an ungrouped frequency distribution, it is $\bar{x} = \dfrac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$.

(ii) Median : It is the value of the middle-most observation (s).

If n is an odd number, the median = value of the $\left(\dfrac{n+1}{2}\right)^{\text{th}}$ observation.

If n is an even number, median = Mean of the values of the $\left(\dfrac{n}{2}\right)^{\text{th}}$ and $\left(\dfrac{n}{2}+1\right)^{\text{th}}$ observations.

(iii) Mode : The mode is the most frequently occurring observation.