

By studying this lesson you will be able to

- find the mean of grouped data.

Grouped Data

Data collected through a survey on the number of family members that reside in each of the houses in a certain housing scheme is given below.

4, 5, 2, 7, 4, 3, 6, 8, 9, 5, 5, 4, 4, 6, 3
 8, 4, 5, 6, 4, 6, 5, 5, 4, 2, 4, 5, 3, 5, 7
 5, 5, 7, 5, 3, 5, 7, 5, 4, 5, 6, 4, 4, 6, 4

The greatest value of this data set is 9 and the least value is 2. The value that is obtained when the least value is subtracted from the greatest value is defined as the **range**.

$$\therefore \text{The range of the given data set} = 9 - 2 \\ = 7$$

When the range of a data set is small, as in the above case, the information related to the data can be tabulated as follows. A table of the following form is called a frequency distribution.

Number of family members residing in the house	Frequency (Number of families)
2	2
3	4
4	12
5	14
6	6
7	4
8	2
9	1

Let us consider another example.

Information on the marks obtained for mathematics at the term test by the grade 10 students of a certain school is given below.

25" 12" 65" 40" 32" 84" 52" 65" 32" 09
 70" 53" 67" 56" 65" 48" 20" 17" 08" 43
 52" 68" 73" 25" 39" 42" 61" 22" 37" 45
 36" 65" 24" 53" 46" 18" 39" 54" 26" 35
 27" 94" 59" 87" 72

In this case, the greatest value is 94 while the least value is 8.

Accordingly, the range of the data set = $94 - 8$

$$= 86$$

Since the range of this data set is large, if we tabulate this information under each value from 8 to 94, we will obtain a very long table.

When the range of the data set is large, as in the above case, it is convenient to divide the data into groups and tabulate the information.

Let us now see how a set of data is separated into groups (class intervals).

A frequency distribution prepared by separating the above data into class intervals is given below.

Class Interval	Frequency
8 - 16	3
17 - 25	7
26 - 34	4
35 - 43	8
44 - 52	5
53 - 61	6
62 - 70	7
71 - 79	2
80 - 88	2
89 - 97	1

What is meant by stating that the frequency of the class interval 8 - 16 is 3, is that there are 3 values (data) that lie in the interval from 8 to 16.

The greatest frequency of this distribution is 8. It corresponds to the class interval 35 - 43. This class interval is named the **modal class**.

A frequency distribution such as the above which is expressed using class intervals is called a **grouped frequency distribution**.

When preparing a grouped frequency distribution, the class intervals are formed so that there are about 10 class intervals.

Observe that all the class intervals in this example are of the same size. That is, the **size** of each class interval is 9.

The initial class interval of this distribution is 8 - 16, and the next class interval is 17 -25. The relevant data are test marks. Since there are no marks between 16 and 17, the class intervals have been organized such that when the first class interval ends with 16, the next class interval starts with 17.

Now let us consider how the mean of a grouped frequency distribution such as the above distribution is found. For this, the mid-value of each class interval has to be found first.

26.1 Mid-value of a class interval

Let us find the mid-value of the class interval 8 - 16 in the above example.

It can be found as follows; $\frac{8 + 16}{2} = 12$.

Accordingly, the mid-value of the class interval 8 – 16 is 12.

The mid-value of a class interval is found by adding together the upper value and the lower value of the class interval and then dividing this sum by 2. The mid-value of every class interval can be found in this manner.

The mid-value of a class interval is considered as the representative value of the data in that class interval when calculations are performed.

Class Interval	Mid-value	Frequency
8 - 16	12	3
17 - 25	21	7
26 - 34	30	4
35 - 43	39	8
44 - 52	48	5
53 - 61	57	6
62 - 70	66	7
71 - 79	75	2
80 - 88	84	2
89 - 97	93	1

A grouped frequency distribution prepared with the data collected on the ages of the staff members in a certain office is given below.

Age of the staff member (Years)	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50	50 - 55	55 - 60
Number of staff members	5	3	3	5	4	2	2	1

Let us recall how the class intervals were written in the example discussed earlier on the mathematics marks of the grade 10 students. The first class interval was written as 8 – 16 and the second class interval was written as 17 – 25. Since there were no marks between 16 and 17, it was appropriate to separate the class intervals in this manner. However in this example, the first class interval is written as 20 – 25, and the second class interval is written as 25 – 30. That is, the second class interval starts with the same value that the first class interval ends. The reason for this is because the data collected here are the ages of people. Since there can be people whose ages lies between 25 and 26 years, the second class interval should commence with the value with which the previous class interval ended.

The ages 20 or greater but less than 25 belong to the class interval 20 – 25. Accordingly the age 25 years belongs to the class interval 25 – 30 years.

Data such as length, weight, etc., Which can take any value within a range are called continuous data. Data such as no.of students, marks, etc., Which take only integral values are called discrete data.

Given below is the grouped frequency distribution of the ages of the staff members, together with the mid-values of the class intervals.

Class Interval	Mid-value	Frequency
20 - 25	22.5	5
25 - 30	27.5	3
30 - 35	32.5	3
35 - 40	37.5	5
40 - 45	42.5	4
45 - 50	47.5	2
50 - 55	52.5	2
55 - 60	57.5	1

Exercise 26.1

1. The marks obtained by several grade 10 students of a certain school has been grouped and tabulated as follows.

Class Interval	Mid-value	Frequency
11 - 20	15.5	1
21 - 30		7
31 - 40		9
41 - 50		8
51 - 60		10
61 - 70		7
71 - 80		4
81 - 90		2
91 - 100		2

- (i) Complete the mid-value column.
(ii) What is the size of each class interval?
(iii) What is the modal class?
2. The data (height to the nearest centimeter) obtained by measuring the heights of the children in a certain class is given below

Class Interval	Mid-value	Frequency
140 - 145		5
145 - 150		8
150 - 155		15
155 - 160		7
160 - 165		8
165 - 170		6

- (i) Copy the table and complete the mid-value column.
(ii) By using the table, find the number of children in the class whose heights are less than 150 cm.
(iii) To which class interval do the heights of the most number of students belong?

3. A grouped frequency distribution prepared using the information on the number of students who were present in a certain school during the first term is given below.

Class Interval (Number who were present)	Mid-value	Frequency (Number of days)
531 - 550		4
551 - 570		10
571 - 590		21
591 - 610		12
611 - 630		10

- (i) Copy the table and complete the mid-value column.
 - (ii) On how many days were there less than 591 students present?
 - (iii) On how many days were there more than 570 students present?
 - (iv) How many days was the school in session during the given term?
4. The information obtained from a test conducted to determine the lifetime of a certain type of light bulb is given below.

Time the bulb remained lit (hours)	Mid-value	Number of bulbs
100 - 200		5
200 - 300		12
300 - 400		25
400 - 500		30
500 - 600		16
600 - 700		12

- (i) Copy the table and complete the mid-value column.
- (ii) How many bulbs burned out in less than 400 hours?
- (iii) How many bulbs were used to conduct this test?
(Assume that the number of hours each bulb that was used remained lit was between 100 hours and 700 hours)

26.2 Calculating the mean of grouped data

When calculating the mean of grouped data, the mid-value of a class interval is taken as the value which represents the whole class interval. Let us now consider how the mean of grouped data is calculated using the mid-values.

Example 1

The following grouped frequency distribution gives the marks that 40 students received out of 25 on a mathematics test.

Class Interval (Marks)	04 - 08	08 - 12	12 - 16	16 - 20	20 - 24
Frequency	3	7	15	11	4

By using the above information, let us construct a table which contains a mid-value column and a column consisting of the product of the mid-values and the corresponding frequencies. Here x denotes the mid-value and f denotes the frequency.

Class Interval	Mid-value x	Frequency f	fx
04 - 08	6	3	18
08 - 12	10	7	70
12 - 16	14	15	210
16 - 20	18	11	198
20 - 24	22	4	88
		$\Sigma f = 40$	$\Sigma fx = 584$

Here, Σf denotes the total number of children, fx denotes the product of the mid-value x and the relevant frequency f , and Σfx denotes the sum of the values in the fx column. Then the mean is defined by $\frac{\Sigma fx}{\Sigma f}$.
That is,

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f}$$

$\therefore$ For the above grouped frequency distribution,

$$\begin{aligned} \text{Mean} &= \frac{\Sigma fx}{\Sigma f} \\ &= \frac{584}{40} \\ &= \underline{\underline{14.6}} \end{aligned}$$

Therefore, the mean mark obtained by the students is 14.6

Exercise 26.2

1. A frequency distribution prepared using the data on the quantity of beans that was brought to a certain vegetable collection centre on a certain day by 40 farmers is given below. This information was obtained through a survey conducted on the quantities of vegetables that are brought to the vegetable collection centre by farmers.

Mass (kg)	14 - 18	18 - 22	22 - 26	26 - 30	30 - 34
Number of Farmers	3	7	15	11	4

- (i) Calculate the mean quantity of beans that was brought by the farmers.
(ii) Accordingly, what is the quantity of beans that can be expected to be brought to this centre during a 10 day period?
2. Information on the number of shirts produced each day of a month by a certain garment factory is given in the following frequency distribution.

Number of shirts	01 - 15	16 - 30	31 - 45	46 - 60	61 - 75
Number of days	4	8	6	8	4

- (i) Calculate the mean number of shirts that is produced per day according to the above information.
(ii) Based on the mean, find the number of shirts that can be expected to be produced during three months.
3. A frequency distribution of the marks received for an assignment by 30 students in a certain class is given below.

Class Interval	1 - 10	11 - 20	21 - 30	31 - 40	41 - 50
Frequency	2	9	13	4	2

- (i) What is the size of a class interval?
(ii) What is the modal class?
(iii) Find the mean mark obtained by a student in the class.

4. The age groups that the teachers serving in a certain educational division belong to are given in the following table.

Age (Years)	21 - 26	26 - 31	31 - 35	36 - 41	41 - 46	46 - 51	51 - 56
Frequency	11	32	51	40	27	18	6

- (i) How many teachers are serving in this educational division?
(ii) To which age group does the most number of teachers belong?
(iii) Calculate the mean age of a teacher serving in this educational division based on the given information.

5. The information obtained by measuring the circumference of the tree trunks stacked in a certain lorry is given below.

Circumference of a tree trunk (cm)	0 - 25	25 - 50	50 - 75	75 - 100	100 - 125
Frequency	8	10	12	20	18

- (i) Find the modal class of this distribution.
(ii) Using the above information, calculate the mean circumference of a tree trunk that was stacked in the lorry.

26.3 Calculating the mean using the assumed mean

The class intervals of a grouped frequency distribution may sometimes contain large mid-values. In such situations, finding the mean using the above method may not be easy. Let us consider a more suitable method of finding the mean of a distribution of this type, through an example.

That is, by considering a simple example, let us explain how the mean is calculated using the assumed mean.

Example 1

The following table contains data on the number of units of water that was consumed during a month by 70 families who receive water from a certain water scheme.

Class interval	12 - 14	15 - 17	18 - 20	21 - 23	24 - 26	27 - 29
Number of families	5	9	11	26	11	8

Calculate the mean number of water units that was consumed by a family, to the nearest whole number.

Let us first find the mid-values which represent each of the class intervals.

Let us now assume that the mid-value 22 of the class interval 21 – 23 is the mean. That is, let us take 22 to be the assumed mean. Now let us find the deviation of each mid-value from the assumed mean by subtracting the assumed mean from each mid-value.

That is, $\boxed{\text{deviation} = \text{mid-value} - \text{assumed mean}}$

Let us now tabulate this information. We denote the deviation by d .

Class Interval	Mid-value x	Deviation d	Frequency f	fx
12 - 14	13	-9	5	-45
15 - 17	16	-6	9	-54
18 - 20	19	-3	11	-33
21 - 23	22	0	26	0
24 - 26	25	3	11	33
27 - 29	28	6	8	48
			$\Sigma f = 70$	$\Sigma fd = 81 - 132$ $= -51$

Here Σf denotes the total number of families, fd the product of the deviation and the corresponding frequency, and Σfd the sum of the values in the fd column.

The mean is obtained by,

$$\text{Mean} = \text{Assumed Mean} + \text{Mean of the Deviations}$$

Accordingly, for the above example,

$$\begin{aligned}
 \text{Mean} &= 22 + \left(\frac{-51}{70} \right) \\
 &= 22 - 0.728 \\
 &= 21.272 \\
 &\approx \underline{\underline{21}}
 \end{aligned}$$

It is easy to find the deviations by taking the assumed mean to be either the mid-value of the modal class or the mid-value of the median class.

If the assumed mean is denoted by A and the deviations are denoted by d , then the mean of the frequency distribution is given by $A + \frac{\Sigma fd}{\Sigma f}$

That is, $\text{Actual mean} = A + \frac{\Sigma fd}{\Sigma f}$

Exercise 26.3

1. Information on the ages of 100 viewers of a certain television programme is given in the following table.

Age (Years)	5 - 15	15 - 25	25 - 35	35 - 45	45 - 55	55 - 65	65 - 75
Number of viewers	7	16	25	31	14	5	2

- What is the modal class of the above frequency distribution?
 - Find the number of viewers whose ages are less than 25 years, as a percentage of the total number of viewers.
 - Find the mean age of a viewer of this programme to the nearest whole number, by taking the mid-value of the class interval 35 - 45 as the assumed mean.
2. The following table has been prepared using the number of days of leave that the staff of a private organization took during a year.

Number of days of leave	0 - 6	6 - 12	12 - 18	18 - 24	24 - 30	30 - 36	36 - 42
Number of staff	5	15	20	11	8	6	5

- What is the modal class of this distribution?
 - If gifts are to be given to those who took less than 6 days leave, what percentage of the total staff will receive gifts?
 - Find the mean number of days of leave taken by a staff member of this organization by using the mid-value of the class interval 18 - 24 as the assumed mean.
 - According to the answer to (iii) above, how many 'man days' of work can the organization expect to lose during a year?
3. A distribution of the marks obtained by 240 students at an examination is given below.

Interval of marks	0 - 8	9 - 17	18 - 26	27 - 35	36 - 44	45 - 53	54 - 62	63 - 71	72 - 80
Frequency	15	18	39	39	48	33	23	14	11

- (i) To which class interval does the greatest number of students belong?
- (ii) By taking the mid-value of the modal class as the assumed mean, find the mean mark of a student.
- (iii) If the 30% who have obtained the lowest marks are to be given remedial lessons, determine the mark below which a student would be selected to follow these lessons.
- (iv) If the top 20% are to be awarded distinctions, above which mark should this selection be made?

4. A table with information on the amount of rice that was sold during a period of 90 days at a cooperative store is given below.

Amount of rice sold in a day (kg)	151 - 175	176 - 200	201 - 225	226 - 250	251 - 275	276 - 300	301 - 325	326 - 350	351 - 375
Number of days	5	7	7	10	21	16	10	8	6

- (i) Write down the modal class of this distribution.
- (ii) By taking the mid-value of the modal class as the assumed mean, calculate to the nearest kilogramme, the mean amount of rice sold in a day during this period.
- (iii) If this pattern of sales is expected to continue during the next two months too, estimate the amount of rice that should be stored to be sufficient for the next 60 days.
- (iv) What is the probability of the amount of rice sold during a day in this period exceeding 300 kg?

5. Frequency distributions of the marks obtained by two groups of 100 students each for a mathematics test is given in the following table.

Class Interval	1 - 10	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60	61 - 70	71 - 80	81 - 90
Number of students in group A	4	8	18	24	16	14	10	4	2
Number of students in group B	7	9	17	26	14	15	8	3	1

- (i) What is the maximum mark that a student may have obtained in this test?
 - (ii) Find the mean mark of a student in each group by taking the mid-value of the class interval 41 – 50 as the assumed mean.
 - (iii) Thereby determine which group of students performed better at the test.
6. A frequency distribution containing information on the number of electricity units used during a certain month by hundred households is given below.

Number of electricity units	31 - 40	41 - 50	51 - 60	61 - 70	71 - 80	81 - 90	91 - 100
Number of households	5	12	26	34	18	3	2

- (i) What is the modal class of this distribution?
 - (ii) Find the mean number of electricity units used by a household during this month, by taking the assumed mean to be the mid-value of the class interval 61 – 70.
 - (iii) When the number of electricity units used lies in the interval 61 – 90, the Electricity Board charges Rs.14 per unit. Accordingly, what is the income that the Electricity Board can expect to earn from these 100 households during that month?
7. Information on the monthly telephone bills of the customers of a certain private telephone company, obtained by conducting a survey in a particular region is given below.

Monthly telephone bill (Rs.)	100-250	250-400	400-550	550-700	700-850	850-1000	1000-1150	1150-1300
Number of customers	2	5	7	15	20	10	8	3

- (i) What is the modal class of this distribution?
- (ii) Find the mean monthly telephone bill of a customer in this region by taking the mid-value of the class interval 550 – 700 as the assumed mean.
- (iii) According to the above mean, what is the income that the company can expect to earn in a month from 1000 customers who use this type of telephone service?
- (iv) If those whose monthly bill exceed Rs. 850 are given an opportunity to win a prize through a special draw, show that more than 30% of the customers in this group are entitled to this.

8. The following table contains information on the speed of the vehicles that passed a certain location during a period of two hours. (The interval 30 – 40 denotes the interval of speeds which are greater than 30 kmh^{-1} but less than or equal to 40 kmh^{-1} .)

Speed (kmh^{-1})	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90
Number of Vehicles	5	7	12	16	15	3	2

- (i) What is the modal class of this distribution?
- (ii) If those who exceed the speed limit of 70 kmh^{-1} are taken to courts, find the percentage of drivers who are taken to courts for exceeding this speed limit.
- (iii) Find the mean speed of a vehicle which passes this location by taking the mid-value of the class interval 50 – 60 as the assumed mean.
- (iv) What is that distance that can be covered during a period of two hours by travelling at the mean speed?