Chapter 30

STATISTICS

Statistics means collection of information in the form of numerical data, organisation, summarisation and presentation of data by tables and graphs (charts), analysing the data and drawing inferences from the data.

Raw data. *The numerical data recorded in its original form as it is measured or received is called* ***raw*** *(or* ***ungrouped****)* ***data****.*

Variable. *A quantity which is being measured in an experiment (or survey) is called a* ***variable****.*

Height, age and weight of people, income and expenditure of people, number of members in a family, marks obtained by students in a test, number of goals scored in a football match etc. are examples of variables.

Range. *The difference between the maximum and minimum values of a variable is called its* ***range****.*

Variate. *A particular value of a variable is called* ***variate****.*

Frequency. *The number of times a variates occurs in a given data is called* ***frequency*** *of that variate.*

Frequency distribution. *A tabular arrangement of given numerical data showing the frequency of the different variates is called* ***frequency distribution****, and the table itself is called* ***frequency distribution table****.*

PRESENTATION OF DATA

Suppose there are 47 employees in a Government office. They were asked how many children they have. The results were :

1, 2, 3, 1, 0, 2, 0, 1, 2, 2, 1, 3, 5, 2, 4, 0, 0, 2, 4, 1, 1, 2, 2, 0, 3, 0, 0, 2, 1, 3, 6, 0, 2, 1, 0, 3, 2, 2, 2, 1, 0, 0, 1, 1, 3, 1, 4.

Each entry in the above list is an **observation** and the collection of these observations is the **raw data**. Here the number of children is the **variable** and the numbers 0, 1, 2, 3, 4, 5, 6 are **variates**.

The data in the above form gives very little information and it is very difficult to interpret it. Let us arrange the above data in ascending or descending order of size. The above data can be written in the ascending order of size as :

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

The data arranged in this form is called **data array** or **arrayed data**. The presentation of the data in this form gives much more information :

(*i*) the number of children varies from 0 to 6.

(*ii*) most of the employees have 0, 1 or 2 children.

(*iii*) only two employees have more than 4 children.

However, the presentation of data in this form is quite tedious and time consuming, particularly when the number of observations is large.

To make the above (raw) data easily understandable, we present it in the form of a table called **frequency distribution table**. To prepare this, write the variates in the extreme left column, then we take each observation from the raw data, one at a time, and mark a stroke (|) called **tally mark** in the next column opposite to the variate. For convenience, we write tally marks in bunches of five, the fifth one crossing the four diagonally. The number of *tally marks* opposite to a variate is its frequency and it is written in the next column opposite to tally marks of the variate.

The frequency distribution table for the above (raw) data is :

Number of children (Variate)	*Tally marks*	*Frequency (No. of employees)*
0	卌 卌 |	11
1	卌 卌 ||	12
2	卌 卌 |||	13
3	卌 |	6
4	|||	3
5	|	1
6	|	1
	Total	47

*The above table is called **simple frequency distribution table**.*

Example 1. The number of goals scored by a football team in different matches is given below :

3, 1, 0, 4, 6, 0, 0, 1, 1, 2, 2, 3, 5, 1, 2, 0, 1, 0, 2, 3, 9, 2, 0, 1, 0, 1, 4, 1, 0, 2, 5, 1, 2, 2, 3, 1, 0, 0, 0, 1, 1, 0, 2, 3, 0, 1, 5, 2, 0

(*i*) Construct data array.

(*ii*) Construct simple frequency distribution table.

Solution. (*i*) Arranging the given data in ascending order, we get

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 5, 5, 5, 6, 9

(*ii*) The simple frequency distribution table for the given data is :

Number of goals scored (Variate)	*Tally marks*	*Frequency (No. of matches played)*
0	卌 卌 ||||	14
1	卌 卌 |||	13
2	卌 卌	10
3	卌	5
4	||	2
5	|||	3
6	|	1
9	|	1
	Total	49

COLUMN (OR BAR) GRAPHS

In column (or bar) graphs, columns (or bars) of equal width are drawn with various heights. The height of a column (or bar) represents the frequency of the corresponding variate. All columns (or bars) are drawn with equal spacing between them.

Example 2. 120 students of class VIII of a certain school use different modes of travel to school as given below :

Mode of travel	Car	School Bus	Bicycle	Walking	Others
Number of students (Frequency)	24	48	18	10	20

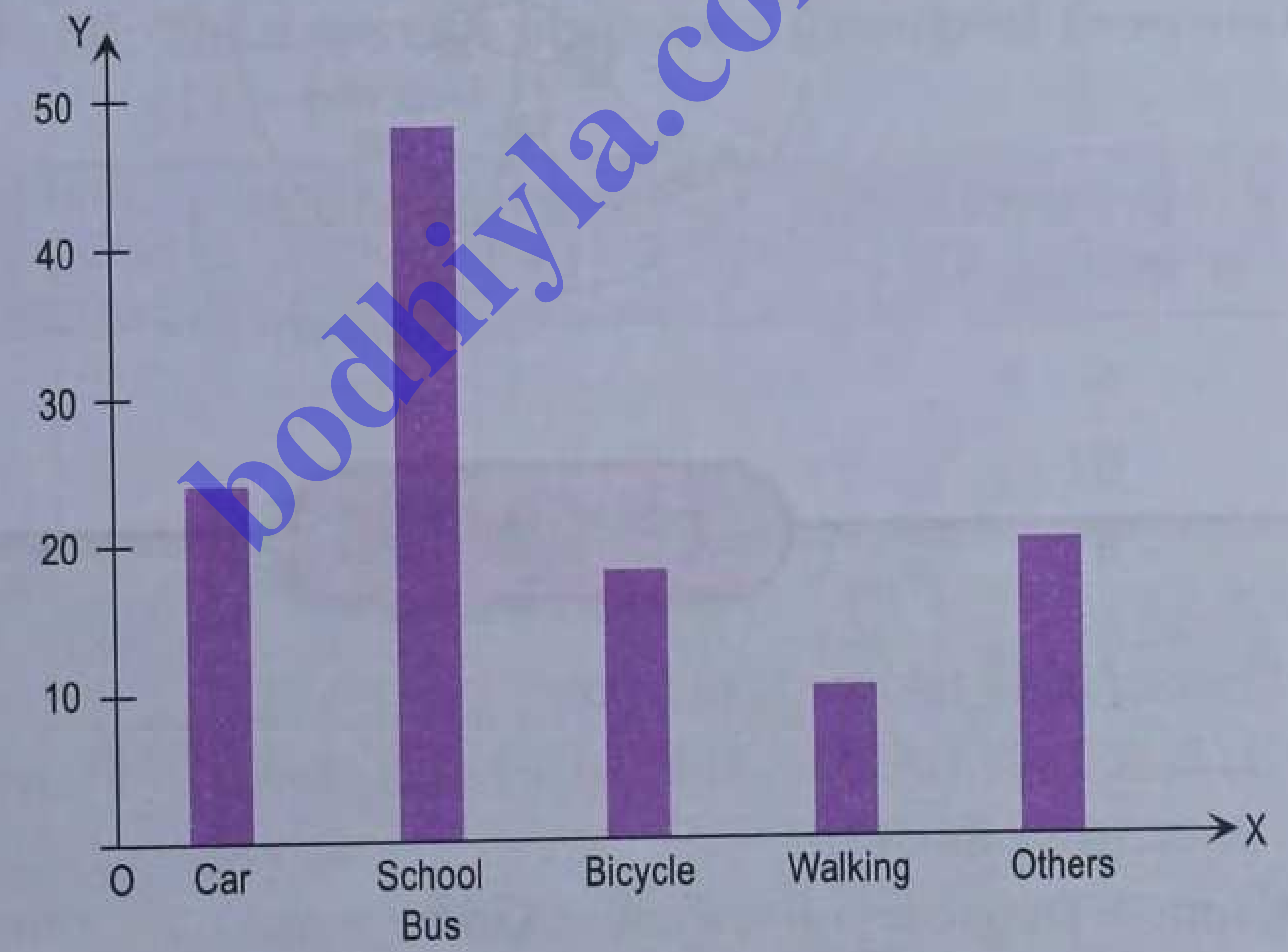
Represent the above data by a column graph.

Solution. Take the mode of travel along x-axis and the number of students (frequency) along y-axis.

Choose 1 small division = 1 student on y-axis.

Draw columns of equal width with equal spacing between them and heights corresponding to the numbers of students taking different modes of travel to school.

The column graph for the given data is shown below :



PIE CHART

The circle graphs are commonly called **pie charts** or **pie graphs**. The pie charts represent numerical data by various sectors of a circle. As the total angle of a circle is 360°, the angle of the sector corresponding to an item is

$$\text{Angle of sector} = \frac{\text{Value of item}}{\text{Sum of values of all items}} \times 360°$$

Example 3. The following table shows the number of students in various classes in a hobby school :

Hobby	Computers	Painting	Pottery	Paper cutting	Glass work
Students	180	150	27	75	108

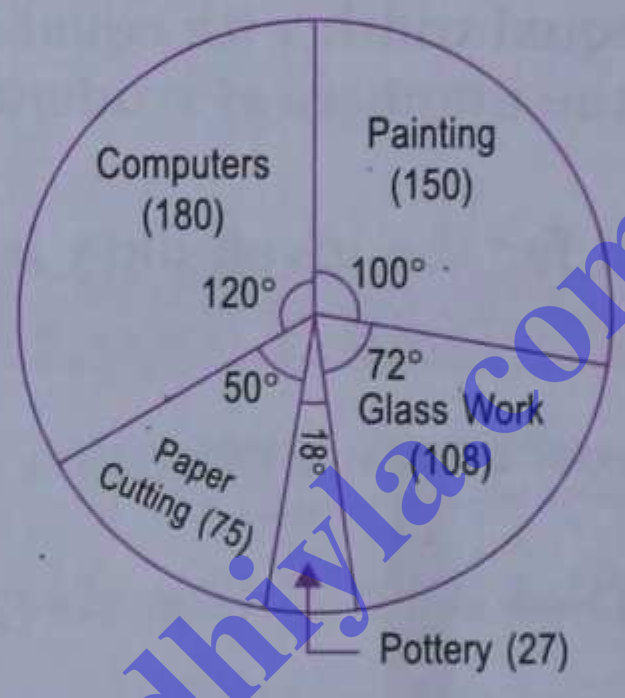
Represent the above data by a pie chart.

Solution. To construct a pie chart, the angles for various sectors can be calculated as :

Total number of students = 180 + 150 + 27 + 75 + 108 = 540.

Hobby	*Number of students*	*Angle*
Computers	180	$\frac{180}{540} \times 360° = 120°$
Painting	150	$\frac{150}{540} \times 360° = 100°$
Pottery	27	$\frac{27}{540} \times 360° = 18°$
Paper cutting	75	$\frac{75}{540} \times 360° = 50°$
Glass Work	108	$\frac{108}{540} \times 360° = 72°$
Total	540	360°

Hence, the corresponding pie chart is :·

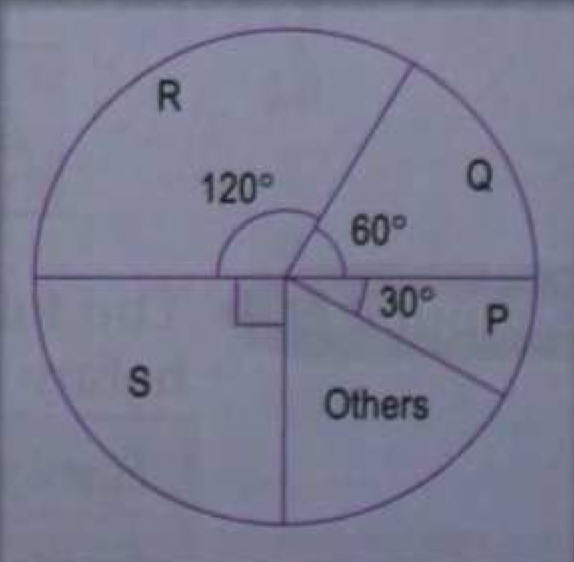


Exercise 30.1

1. The number of rooms in 25 houses is as below :
   4, 3, 2, 6, 4, 3, 1, 2, 5, 3, 2, 3, 4, 3, 5, 1, 6, 1, 3, 4, 2, 3, 4, 3, 5
   (*a*) Construct data array for this.
   (*b*) Construct simple frequency distribution table.
   (*c*) Draw column graph to represent the above data.

2. A family spends its monthly income in the following manner :
   Rent 20%, Food 30%, Education 15%, Miscellaneous 25% and Savings 10%.
   Represent this data by a
   (*a*) column graph (*b*) pie chart.

3. A total of 12000 soap cakes were sold in a year. Four main brands were P, Q, R, S. Use adjoining pie chart to find the number of soap cakes sold by the four main brands.



4. The following table shows the population of five major metros of India in 1981 :

City	*Population (in lakhs)*
Bangaluru	29
Chennai	43
Delhi	57
Kolkata	92
Mumbai	82

(*a*) Draw a column graph to represent above information.

(*b*) Which city had maximum population? Explain with the help of above diagram.

GROUPING OF DATA

Let us suppose the pocket money of 25 students in a class is (in rupees per week) :
49, 35, 22, 27, 30, 27, 25, 27, 13, 36, 25, 12, 40, 41, 29, 22, 11, 28, 32, 34, 38, 37, 30,12, 28.

A glance at the data will convince you that frequency table will have too many rows. In such a case, it is more useful to form small groups of values. Here values vary between 11 (minimum) and 49 (maximum). So, we can try forming **groups** like 10 – 20, 20 – 30, 30 – 40, 40 – 50. Then we can prepare a **grouped frequency distribution table**.

Pocket money (₹ per week)	*Tally marks*	*Number of students (Frequency)*
10 – 20	||||	4
20 – 30	卌 卌	10
30 – 40	卌 |||	8
40 – 50	|||	3

The groups like 10 – 20, 20 – 30,... are also called **classes**. In the class 10 – 20, the **lower class limit** is 10 and **upper class limit** is 20. The difference between the two is called **class interval** or **class size** or **class width**. In this case, the class size is 20 – 10 = 10. The mid-point of class interval is called **class mark**. Here the class mark for class 10 – 20 is $\frac{10+20}{2}$ = 15, class mark for class 20 – 30 is $\frac{20+30}{2}$ = 25, and so on. Notice that size of all classes is same (10) but class marks are different. Also notice that value 30 goes into class 30 – 40 and not into class 20 – 30. Thus, lower class limit is included in class but upper class limit is not included.

Formation of classes from a given raw data

We condense the given raw data into classes (or groups) as follows :

1. Find the range *i.e.* the difference between the maximum and minimum values of the variable.
2. Decide about the number of classes (usually between 5 to 10).
3. As far as possible, classes should be of same size.
4. Size (width) of a class is the natural number immediately greater than the quotient obtained by dividing the range by the number of classes.
5. Include the lower limit in the class but exclude the upper limit

Example 1. The weights (in kg) of 38 students of class VIII of a certain school are given below :

36, 38, 42, 46, 37, 49, 51, 53, 44, 37, 41, 46, 44, 33, 32, 40, 42, 41, 39, 50, 33, 38, 42, 47, 49, 34, 34, 38, 42, 43, 36, 35, 43, 44, 48, 42, 37, 44

Construct a grouped frequency distribution table.

Solution. Here maximum value = 53 and minimum value = 32,

$\therefore$ range = 53 – 32 = 21

Let us form 5 classes each of size 5.

Since we want to include 53 in the last class, 55 can be taken as the upper limit of the last class. Hence, the classes can be taken as

30–35, 35–40, 40–45, 45–50, 50–55.

The grouped frequency distribution table for the given data is :

Classes	*Tally mark*	*Frequency*
30–35	卌	5
35–40	卌 卌	10
40–45	卌 卌 ||||	14
45–50	卌 |	6
50–55	|||	3
	Total	38

HISTOGRAMS

When a grouped frequency distribution table is given, usually we use a histogram to represent the data in visual form. A histogram consists of a set of adjacent rectangles. We mark the class limits on the x-axis. Correspond to each class, we take the breadth of the rectangle as equal to class size and mark it along x-axis, the end points corresponding to the class limits marked on the x-axis. The length of the rectangle is taken as equal to the frequency of that class and marked on the y-axis.

Note that bar chart is used for a data array or a simple frequency table. Histogram is used for grouped data. Also, there are some gaps between bars, whereas no such gap is there between adjacent rectangles in a histogram. In a bar graph, the breadth of a rectangle has no significance; whereas in a histogram, the breadth of a rectangle is meaningful and it represents the class size. The following examples will make this clear.

Example 2. Draw a histogram to represent the following data :

Class-intervals	10–20	20–30	30–40	40–50	50–60
Frequency	3	5	10	8	2

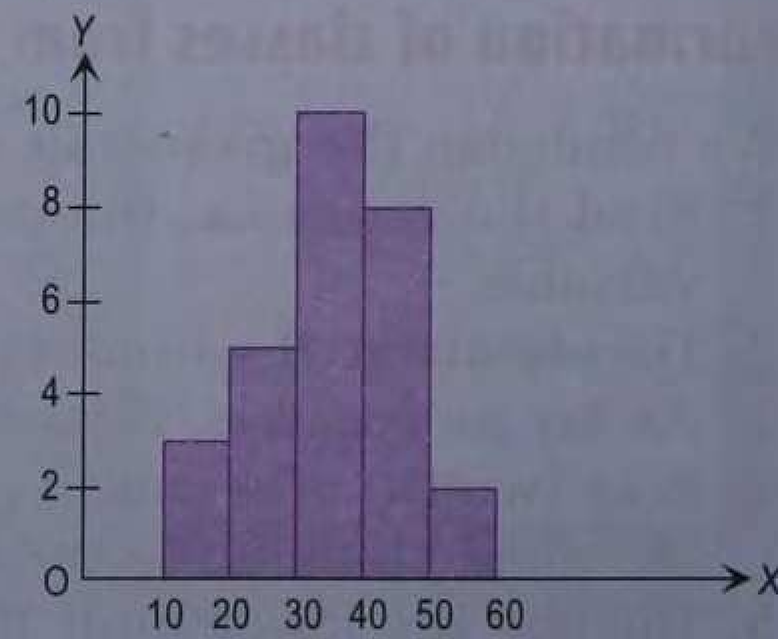
Solution. *Steps*

(*i*) Take 1 cm on x-axis = 10 units.

(*ii*) Take 1 cm on y-axis = 2 frequency.

(*iii*) Construct rectangles corresponding to given data.

The required histogram is shown in the adjoining figure.



Example 4. Draw a histogram for the following data :

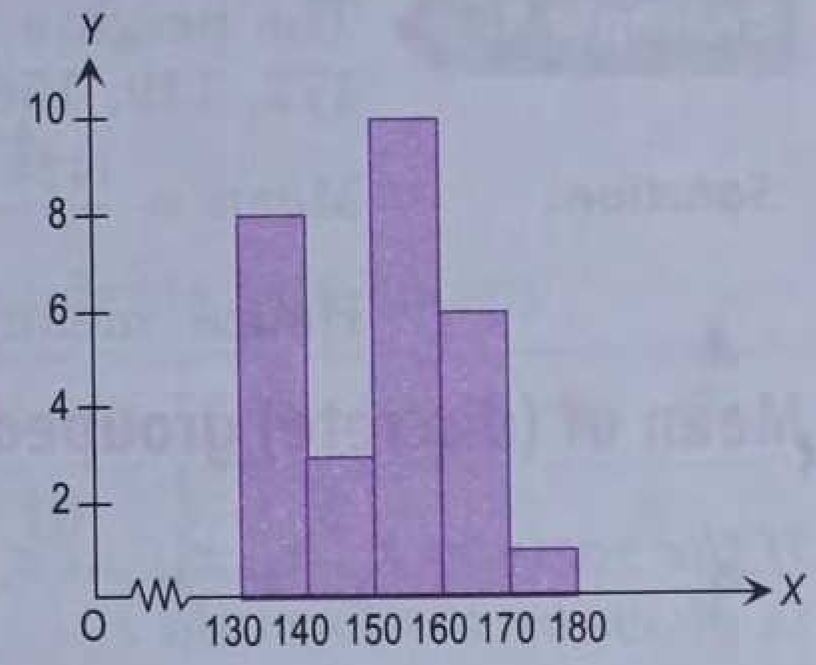
Height (in cm)	130 – 140	140 – 150	150 – 160	160 – 170	170 – 180
No. of students	8	3	10	6	1

Solution. *Steps*

(*i*) Since the scale on x-axis starts at 130, a break (kink or zig-zag curve) is shown near the origin along x-axis to indicate that the graph is drawn to scale beginning at 130 and not on the origin itself to save space.

(*ii*) Take 1 cm on x-axis = 10 cm (height).

(*iii*) Take 1 cm on y-axis = 2 (no. of students).



(*iv*) Construct rectangles corresponding to the given data.

The required histogram is shown in the adjoining figure.

Exercise 30.2

1. A teacher gave a surprise test of 50 marks to the students of his class. The results were :
48, 42, 38, 37, 25, 23, 10, 12, 15, 28, 6, 3, 35, 37, 40, 20, 28, 27, 37, 40, 30, 40, 41, 14, 17.
   (*a*) Construct data array.
   (*b*) Prepare grouped frequency distribution using groups 0 – 10, 10 – 20,...
   (*c*) What is the second class? Mention its width, lower limit, upper limit and class mark.
   (*d*) What is the fourth class? Mention its width, lower limit, upper limit and class mark.
2. The weights of 40 students in a class are (in kilograms) :
25, 16, 30, 18, 17, 14, 18, 19, 22, 23, 18, 24, 24, 24, 18, 19, 22, 26, 21, 20, 23, 17, 31, 17, 16, 22, 23, 20, 20, 19, 18, 24, 23, 22, 21, 27, 24, 23, 22, 20.
   (*a*) Construct data array.
   (*b*) Prepare a grouped frequency distribution table using suitable classes.
3. In a test of 50 marks, 20 students scored marks as 2, 8, 10, 13, 14, 14, 20, 23, 23, 23, 28, 28, 31, 31, 34, 34, 38, 39, 39, 45. Prepare a grouped frequency distribution table with classes as 0 – 10, 10 – 20,... and draw the histogram.
4. The weights of 29 patients in a hospital were recorded as follows :

Weight (in kg)	50 – 55	55 – 60	60 – 65	65 – 70	70 – 75	75 – 80
Number	7	4	4	9	2	3

Draw a histogram to represent this data visually.

MEAN

The arithmetic average of a number of observations/items is called **mean**.

Mean of raw (ungrouped) data

The mean of n observations (items) x_1, x_2, x_3, ..., x_n is given by the formula :

$$\text{Mean} = \frac{\text{sum of observations}}{\text{number of observations}} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\Sigma x_i}{n}$$

Here Greek letter Σ (read as sigma) represents the sum.

Thus, if Virat scores 70, 45, 38, 120, 96, 18 runs in 6 innings, his average score or mean score is

$$\frac{70+45+38+120+96+18}{6}=\frac{387}{6} = 64{\cdot}5.$$

Example 1. The heights (in cm) of 7 students of class VIII are 142, 153, 166, 161, 172, 149, 156. Find their mean height.

Solution. Mean = $\frac{142+153+166+161+172+149+156}{7} = \frac{1099}{7} = 157.$

Hence, mean height = 157 cm.

Mean of (discrete) grouped data

If the variates $x_1, x_2, x_3, ..., x_n$ have frequencies $f_1, f_2, f_3, ..., f_n$ respectively, then their mean is given by the formula :

$$\text{Mean} = \frac{f_1x_1+f_2x_2+f_xx_3+...+f_nx_n}{f_1+f_2+f_3+...+f_n}=\frac{\Sigma f_i x_i}{\Sigma f_i} \text{ or } \frac{\Sigma f x}{\Sigma f}$$

For example, consider the following frequency distribution :

Marks obtained x_i	0	3	4	5	6	7	8	9
No. of students f_i	2	4	2	6	1	3	5	2

Here $\Sigma f_i = 2 + 4 + 2 + 6 + 1 + 3 + 5 + 2 = 25$

$\Sigma f_i x_i = 0 \times 2 + 3 \times 4 + 4 \times 2 + 5 \times 6 + 6 \times 1 + 7 \times 3 + 8 \times 5 + 9 \times 2$

$= 0 + 12 + 8 + 30 + 6 + 21 + 40 + 18 = 135$

∴ Mean $= \frac{\Sigma f_i x_i}{\Sigma f_i}=\frac{135}{25} = 5{\cdot}4.$

Example 2. The following table shows the daily pocket money of 44 students of class VIII of a school :

Pocket money (in ₹)	3	5	8	10	15	17
Number of students	4	7	10	12	8	3

Find the mean pocket money per day.

Solution. To calculate the mean pocket money, we construct the following table :

Pocket money (in ₹) x_i	*Number of students* f_i	$f_i x_i$
3	4	12
5	7	35
8	10	80
10	12	120
15	8	120
17	3	51
Total	44	418

Mean $= \frac{\Sigma f_i x_i}{\Sigma f_i}=\frac{418}{44}=\frac{19}{2} = 9{\cdot}5$

Mean pocket money per day = ₹ 9·50.

Mean of grouped data when frequency distribution is given in the form of classes

In this case, we take x as the class mark and use the formula :

$$\text{Mean} = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

where f_i is the frequency of the class and x_i is its class mark.

Example 3. The following table shows the weights of students in a class :

Weight (in kg)	30–36	36–42	42–48	48–54	54–60
No. of students	8	3	10	4	1

Find the mean weight.

Solution. To find the mean weight, we construct the table as under :

Classes	*Class mark* x_i	*Frequency* f_i	$f_i x_i$
30–36	33	8	264
36–42	39	3	117
42–48	45	10	450
48–54	51	4	204
54–60	57	1	57
	Total	26	1092

$$\text{Mean} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1092}{26} = 42$$

Hence, the mean weight = 42 kg.

MEDIAN

When the observations (items) are arranged in ascending or descending order, the middle observation (item) is called ***median.***

Here we shall consider only simple data.

(*i*) If the number of observations is odd, then there will be only one middle term and the middle term is the median.

(*ii*) If the number of observations is even, then there are two middle terms and the average of these two terms is the median.

Example 4. Find the median of the following data :

15, 22, 8, 16, 0, 10, 16.

Solution. Arranging the given data in ascending order, we get

0, 8, 10, 15, 16, 16, 22.

Total number of observations (items) = 7 (odd).

The middle observation is 15.

Hence, the median of the given data = 15.

Example 5. Find the median of the following data :
2, 2, 0, 4, 12, 10, 6, 8

Solution. Arranging the given data in ascending order, we get
0, 2, 2, 4, 6, 8, 10, 12.
Total number of observations (items) = 8 (even).
There are two middle items 4 and 6.
Their average = $\frac{4+6}{2}$ = 5.
Hence, the median of the given data = 5.

MODE

The observation (variate) which appears maximum number of times in the given data is called ***mode.***
In a simple grouped data, mode is the observation which has maximum frequency.

Example 6. Find the mode of the following data :
3, 5, 1, 2, 4, 6, 0, 2, 2, 3

Solution. In the given data, 2 is repeated more number of times than any other number.
∴ Mode = 2.

Example 7. The marks obtained (out of 50) by 60 students in a class test are given below:

Marks obtained	13	17	25	32	38	42	44	47	48	49
No. of students	7	15	12	3	6	2	1	2	1	1

Find the mode of the above distribution.

Solution. In the given distribution, the variate 17 has maximum frequency.
∴ Mode = 17.

Exercise 30.3

1. Find the mean of the following data :
   (*i*) 40, 30, 30, 0, 26, 60
   (*ii*) 1, 2, 3, 4, 5, 6, 7
   (*iii*)

x	2	3	4	5	10
f	3	2	6	7	2

   (*iv*)

x	0–20	20–40	40–60	60–80
f	15	6	12	7

2. Find the median of the following data :
   (*i*) 3, 1, 5, 6, 3, 4, 5
   (*ii*) 3, 1, 5, 6, 3, 4, 5, 6

3. Find the mode of the following data :

(*i*) 3, 1, 5, 6, 3, 4, 5, 3

(*ii*)

Marks	15	17	20	22	25
No. of students	6	17	12	18	8

4. Calculate the mean, the median and the mode of the following numbers :

1, 3, 2, 6, 2, 3, 1, 3

5. Calculate the mean, the median and the mode of the following numbers :

3, 7, 2, 5, 3, 4, 1, 5, 3, 6

6. Find the mean and the mode of the following distribution :

No. of goals	0	1	2	3	4	5
No. of matches	1	4	7	6	9	3

Summary

- Data recorded as it is measured or received is called raw data.
- If we arrange raw data in increasing or decreasing order, we get a data array.
- The number of times a particular observation occurs is called its frequency.
- A table showing the frequency of various observations is called frequency distribution table.
- Tally marks (|) are useful for counting observations. We write tally marks in bunches of five like 𝍸 .
- Data can be represented diagrammatically by column graphs, horizontal bar charts, pie charts etc.
- In column graphs, columns of equal width are drawn with equal spacing between them. The height of a column represents the frequency of the corresponding variate (observation).
- In pie charts, the circle is subdivided to show the comparison between different values. The angle of the sector corresponding to an item is given by

$$\text{Angle of sector} = \frac{\text{Value of item}}{\text{Sum of values of all items}} \times 360°$$

- If number of observations is large, then grouping of data is done. Groups or classes are formed. The frequency table so prepared is called grouped frequency distribution table.
- The lower value of a class is called lower class limit and upper value is called upper class limit. Difference between these two values is called class interval or class size or class width.
- Mid-point of class interval is called class mark.
- Histograms are used to represent grouped data. Histogram consists of a set of adjacent rectangles. The height of rectangles correspond to the frequency of the class and the breadth of the rectangle corresponds to the class size.
- Arithmetic average of a number of observations is called its mean.
- For simple data, mean = $\frac{\text{sum of observations}}{\text{number of observations}} = \frac{\Sigma x}{n}$

- If the variates $x_1, x_2, x_3, ..., x_n$ have frequencies $f_1, f_2, f_3, ..., f_n$ then mean = $\frac{\Sigma f_i x_i}{\Sigma f_i}$.
- For grouped data, mean = $\frac{\Sigma f_i x_i}{\Sigma f_i}$, where f_i is the frequency of the ith class and x_i is its class mark.
- Median is the middle item when the observations are arranged in ascending (increasing) or descending (decreasing) order of magnitude. If the number of items is even, there are two middle items : the average of these two is the median.
- The observation which appears maximum times in given data is called mode. In tabulated data, mode is the observation which has maximum frequency.

Check Your Progress

1. The marks obtained (out of 50) by 30 students in a class were :

28, 31, 45, 03, 05, 18, 35, 46, 49, 17, 10, 28, 31, 36, 40, 44, 47, 13, 19, 25, 24, 31, 38, 32, 27, 19, 25, 28, 48, 15.

(*a*) Construct data array.

(*b*) Construct grouped frequency distribution table using groups 0–10, 10–20, 20–30, 30–40, 40–50.

(*c*) What is the third class?

(*d*) What is the lower class limit of third class?

(*e*) What is the upper class limit of third class?

(*f*) What is the class interval of third class?

(*g*) What is the class mark of third class?

2. Draw a histogram for the career record of a batsman :

Score	0–25	25–50	50–75	75–100	100–125	125–150
No. of innings	8	7	4	1	4	0

3. A boy scored the following marks in various class tests, each test being marked out of 20 marks :

15, 17, 16, 7, 10, 12, 14, 16, 19, 12, 16.

(*i*) What are his mean marks?

(*ii*) What are his median marks?

(*iii*) What are his modal marks?

4. In an examination, the marks obtained by 50 students are given in the table below. Calculate the mean marks.

Marks obtained	0–10	10–20	20–30	30–40	40–50
Number of students	8	14	13	10	5

5. Find the mean and the mode of the following grouped data :

Height (in cm)	65	66	67	68	69	70	71	72	73
No. of plants	1	4	5	7	11	10	6	4	2