

CHAPTER - 14

STATISTICS

Exercise 14.1

Question 1:

Give five examples of data that you can collect from day to day life.

Answer 1:

In our day to day life, we can collect the following data.

1. Number of females per 1000 males in various states of our country
2. Weights of students of our class
3. Production of wheat in the last 10 years in our country
4. Number of plants in our locality
5. Rainfall in our city in the last 10 years

Question 2:

Classify the data in Q1 above as primary or secondary data.

Answer 2:

The information which is collected by the investigator himself with a definite objective in his mind is called as primary data whereas when the information is gathered from a source which already had the information stored, it is called as secondary data. It can be observed that the data in 1, 3, and 5 is secondary data and the data in 2 and 4 is primary data.

Exercise 14.2

Question 1:

The blood groups of 30 students of Class VIII are recoded as follows:

A, B, O, O, AB, O, A, O, B, A, O, B, A, O, O, A, AB, O, A, A, O, O, AB, B, A, O, B, A, B, O.

Represent this data in the form of a frequency distribution table. Which is the most common, and which is the rarest, blood group among these students?

Answer 1:

It can be observed that 9 students have their blood group as A, 6 as B, 3 as AB, and 12 as O.

Therefore, the blood group of 30 students of the class can be represented as follows.

Blood group	Number of students
A	9
B	6
AB	3
O	12
Total	30

It can be observed clearly that the most common blood group and the rarest blood group among these students is O and AB respectively as 12 (maximum number of students) have their blood group as O, and 3 (minimum number of students) have their blood group as AB.

Question 2:

The distance (in km) of 40 engineers from their residence to their place of work were found as follows:

5 3 10 20 25 11 13 7 12 31 19 10 12 17 18 11 32 17 16 2 7 9 7 8 3 5 12 15 18 3
12 14 2 9 6 15 15 7 6 12

Construct a grouped frequency distribution table with class size 5 for the data given above taking the first interval as 0 - 5 (5 not included). What main feature do you observe from this tabular representation?

Answer 2:

It is given that a grouped frequency distribution table of class size 5 has to be constructed. Therefore, the class intervals will be 0 - 5, 5 - 10, 10 - 15, 15 - 20... By observing the data given as above, a grouped frequency distribution table can be constructed as follows.

Distance (in km)	Tally mark	Number of engineers
0 - 5		5
5 - 10		11
10 - 15		11
15 - 20		9
20 - 25		1
25 - 30		1
30 - 35		2
Total		40

It can be observed that there are very few engineers whose homes are at more than or equal to 20 km distance from their work place. Most of the engineers have their workplace up to 15 km distance from their homes.

Question 3:

The relative humidity (in %) of a certain city for a month of 30 days was as follows:

98.1 98.6 99.2 90.3 86.5 95.3 92.9 96.3 94.2 95.1 89.2 92.3 97.1 93.5 92.7 95.1
97.2 93.3 95.2 97.3 96.2 92.1 84.9 90.2 95.7 98.3 97.3 96.1 92.1 89

- (i) Construct a grouped frequency distribution table with classes 84 - 86, 86 - 88
- (ii) Which month or season do you think this data is about?
- (iii) What is the range of this data?

Answer 3:

(i). A grouped frequency distribution table of class size 2 has to be constructed. The class intervals will be 84 – 86, 86 – 88, and 88 – 90...

By observing the data given above, the required table can be constructed as follows.

Relative humidity (in %)	Number of days (frequency)
84 – 86	1
86 – 88	1
88 – 90	2
90 – 92	2
92 – 94	7
94 – 96	6
96 – 98	7
98 – 100	4
Total	30

(ii). It can be observed that the relative humidity is high. Therefore, the data is about a month of rainy season.

(iii). Range of data = Maximum value – Minimum value
= 99.2 – 84.9 = 14.3

Question 4:

The heights of 50 students, measured to the nearest centimeters, have been found to be as follows:

150 154 165 168 161 154 162 150 151 164 171 165 158 154 156 172 160 170 159
161 170 162 165 166 168 165 164 152 153 156 158 162 160 161 173 166 161 159
162 167 168 159 158 153 154 159

(i) Represent the data given above by a grouped frequency distribution table, taking the class intervals as 160 - 165, 165 - 170, etc.

(ii) What can you conclude about their heights from the table?

Answer 4:

(i) A grouped frequency distribution table has to be constructed taking class intervals 160 - 165, 165 - 170, etc. By observing the data given above, the required table can be constructed as follows.

Height (in cm)	Number of students (frequency)
150 - 155	12
155 - 160	9
160 - 165	14
165 - 170	10
170 - 175	5
Total	50

(ii) It can be concluded that more than 50% of the students are shorter than 165 cm.

Question 5:

A study was conducted to find out the concentration of sulphur dioxide in the air in parts per million (ppm) of a certain city. The data obtained for 30 days is as follows:

0.03 0.08 0.08 0.09 0.04 0.17 0.16 0.05 0.02 0.06 0.18 0.20 0.11 0.08 0.12 0.13
0.22 0.07 0.08 0.01 0.10 0.06 0.09 0.18 0.11 0.07 0.05 0.07 0.01 0.04

- (i) Make a grouped frequency distribution table for this data with class intervals as 0.00 - 0.04, 0.04 - 0.08, and so on.
- (ii) For how many days, was the concentration of sulphur dioxide more than 0.11 parts per million?

Answer 5:

(i) Taking class intervals as 0.00, 0.04, 0.08, and so on, a grouped frequency table can be constructed as follows.

Concentration of SO ₂ (in ppm)	Number of days (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
Total	30

The number of days for which the concentration of SO₂ is more than 0.11 is the number of days for which the concentration is in between 0.12 – 0.16, 0.16 – 0.20, 0.20 – 0.24.

(ii) Required number of days = 2 + 4 + 2 = 8

Therefore, for 8 days, the concentration of SO₂ is more than 0.11 ppm.

Question 6:

Three coins were tossed 30 times simultaneously. Each time the number of heads occurring was noted down as follows:

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

Prepare a frequency distribution table for the data given above.

Answer 6:

By observing the data given above, the required frequency distribution table can be constructed as follows.

Number of heads	Number of times (frequency)
0	6
1	10
2	9
3	5
Total	30

Question 7:

The value of π up to 50 decimal places is given below:

3.14159265358979323846264338327950288419716939937510

- (i) Make a frequency distribution of the digits from 0 to 9 after the decimal point.
- (ii) What are the most and the least frequently occurring digits?

Answer 7:

- (i) By observation of the digits after decimal point, the required table can be constructed as follows.

Digit	Frequency
0	2
1	5
2	5
3	8
4	4
5	5
6	4
7	4
8	5
9	8
Total	50

(ii) It can be observed from the above table that the least frequency is 2 of digit 0, and the maximum frequency is 8 of digit 3 and 9. Therefore, the most frequently occurring digits are 3 and 9 and the least frequently occurring digit is 0.

Question 8:

Thirty children were asked about the number of hours they watched TV programmes in the previous week. The results were found as follows:

1 6 2 3 5 12 5 8 4 8 10 3 4 12 2 8 15 1 17 6 3 2 8 5 9 6 8 7 14 12

(i) Make a grouped frequency distribution table for this data, taking class width 5 and one of the class intervals as 5 - 10.

(ii) How many children watched television for 15 or more hours a week?

Answer 8:

(i) Our class intervals will be 0 - 5, 5 - 10, 10 - 15.....

The grouped frequency distribution table can be constructed as follows.

Hours	Number of children
0 - 5	10
5 - 10	13
10 - 15	5
15 - 20	2
Total	30

(ii) The number of children who watched TV for 15 or more hours a week is 2 (i.e., the number of children in class interval 15 - 20).

Question 9:

A company manufactures car batteries of a particular type. The lives (in years) of 40 such batteries were recorded as follows:

2.6 3.0 3.7 3.2 2.2 4.1 3.5 4.5 3.5 2.3 3.2 3.4 3.8 3.2 4.6 3.7 2.5 4.4 3.4 3.3 2.9 3.0
4.3 2.8 3.5 3.2 3.9 3.2 3.2 3.1 3.7 3.4 4.6 3.8 3.2 2.6 3.5 4.2 2.9 3.6

Construct a grouped frequency distribution table for this data, using class intervals of size 0.5 starting from the intervals 2 – 2.5.

Answer 9:

A grouped frequency table of class size 0.5 has to be constructed, starting from class interval 2 – 2.5.

Therefore, the class intervals will be 2 – 2.5, 2.5 – 3, 3 – 3.5...

By observing the data given above, the required grouped frequency distribution table can be constructed as follows.

Lives of batteries (in hours)	Number of batteries
2 – 2.5	2
2.5 – 3.0	6
3.0 – 3.5	14
3.5 – 4.0	11
4.0 – 4.5	4
4.5 – 5.0	3
Total	40

Exercise 14.3

Question 1:

A survey conducted by an organisation for the cause of illness and death among the women between the ages 15 – 44 (in years) worldwide, found the following figures (in %):

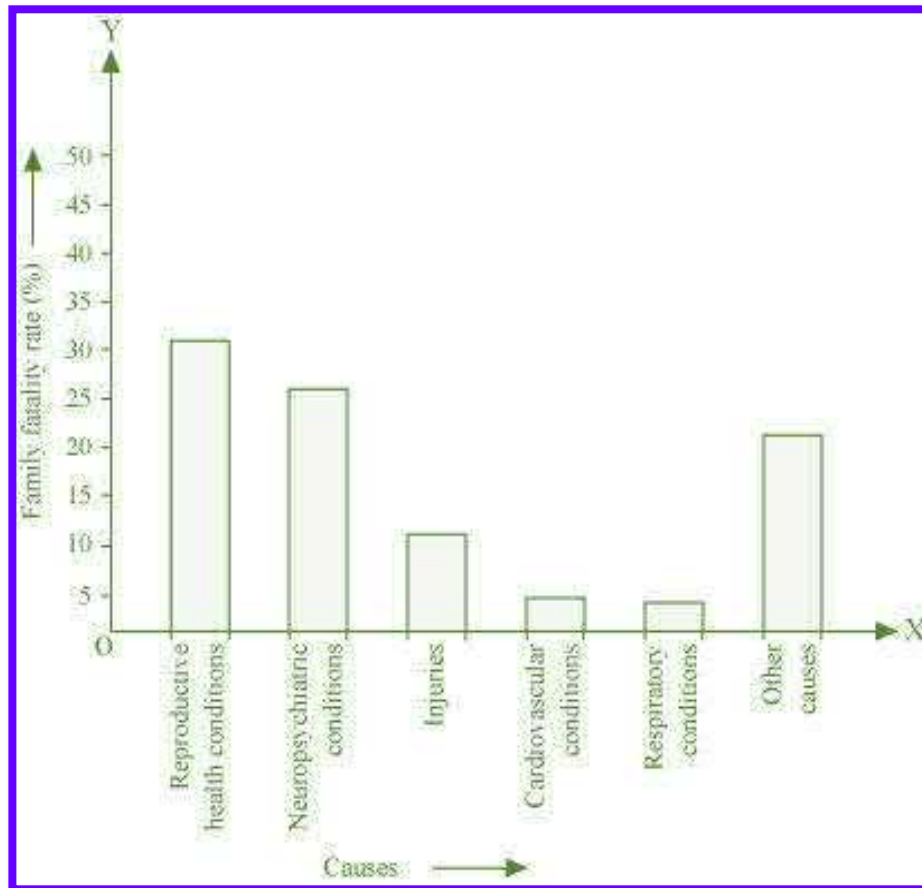
S.No.	Causes	Female fatality rate (%)
1.	Reproductive health conditions	31.8
2.	Neuropsychiatric conditions	25.4
3.	Injuries	12.4
4.	Cardiovascular conditions	4.3
5.	Respiratory conditions	4.1
6.	Other causes	22.0

(i) Represent the information given above graphically.

(ii) Which condition is the major cause of women's ill health and death worldwide? Try to find out, with the help of your teacher, any two factors which play a major role in the cause in (ii) above being the major cause.

Answer 1:

(i) By representing causes on x-axis and family fatality rate on y-axis and choosing an appropriate scale (1 unit = 5% for y axis), the graph of the information given above can be constructed as follows.



All the rectangle bars are of the same width and have equal spacing between them.

(ii) Reproductive health condition is the major cause of women's ill health and death worldwide as 31.8% of women are affected by it.

(iii) The factors are as follows.

1. Lack of medical facilities
2. Lack of correct knowledge of treatment

Question 2:

The following data on the number of girls (to the nearest ten) per thousand boys in different sections of Indian society is given below.

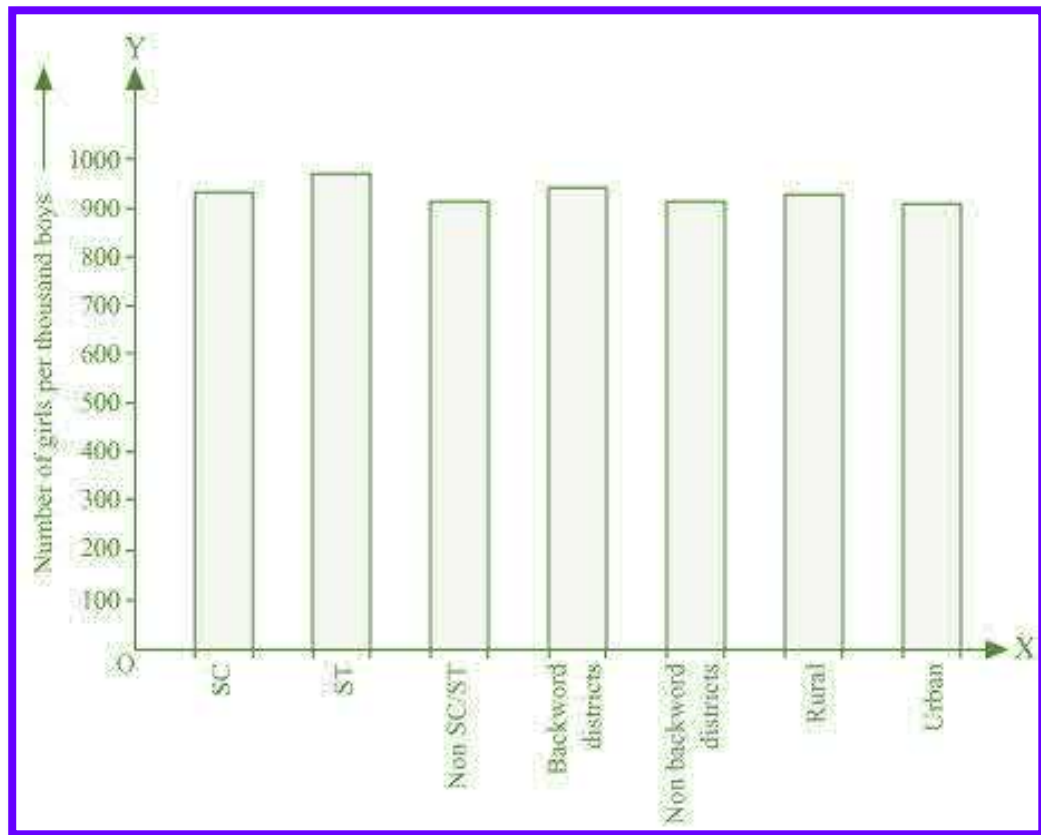
Section	Number of girls per thousand boys
Scheduled Caste (SC)	940
Scheduled Tribe (ST)	970
Non SC/ST	920
Backward districts	950
Non-backward districts	920
Rural	930
Urban	910

(i) Represent the information above by a bar graph.

(ii) In the classroom discuss what conclusions can be arrived at from the graph.

Answer 2:

(i) By representing section (variable) on x-axis and number of girls per thousand boys on y-axis, the graph of the information given above can be constructed by choosing an appropriate scale (1 unit = 100 girls for y-axis)



Here, all the rectangle bars are of the same length and have equal spacing in between them.

(ii) It can be observed that maximum number of girls per thousand boys (i.e., 970) is for ST and minimum number of girls per thousand boys (i.e., 910) is for urban. Also, the number of girls per thousand boys is greater in rural areas than that in urban areas, backward districts than that in non-backward districts, SC and ST than that in non SC/ST.

Question 3:

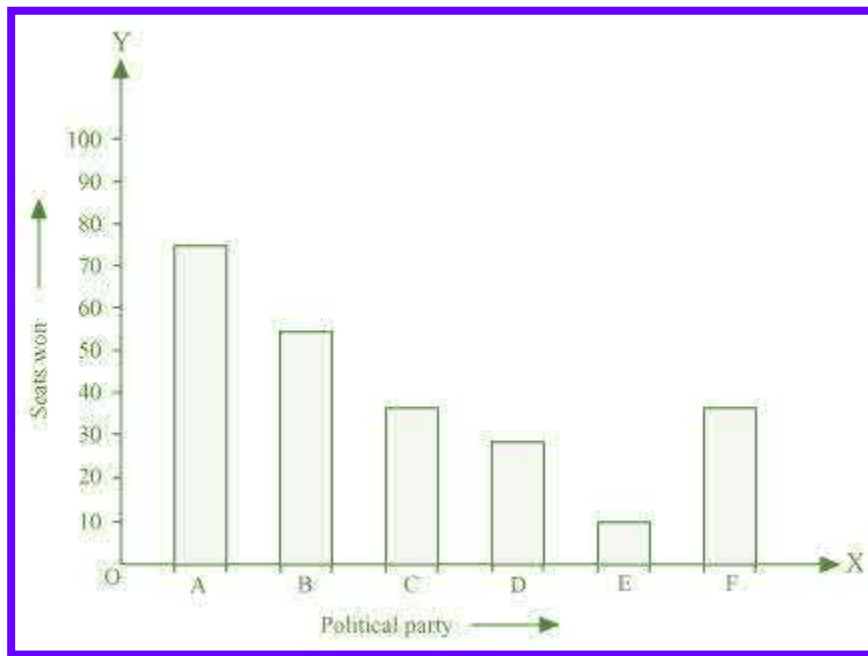
Given below are the seats won by different political parties in the polling outcome of a state assembly elections:

Political Party	A	B	C	D	E	F
Seats Won	75	55	37	29	10	37

- (i) Draw a bar graph to represent the polling results.
- (ii) Which political party won the maximum number of seats?

Answer 3:

- (i) By taking polling results on x-axis and seats won as y-axis and choosing an appropriate scale (1 unit = 10 seats for y-axis), the required graph of the above information can be constructed as follows.



Here, the rectangle bars are of the same length and have equal spacing in between them.

- (ii) Political party 'A' won maximum number of seats.

Question 4:

The length of 40 leaves of a plant are measured correct to one millimetre, and the obtained data is represented in the following table:

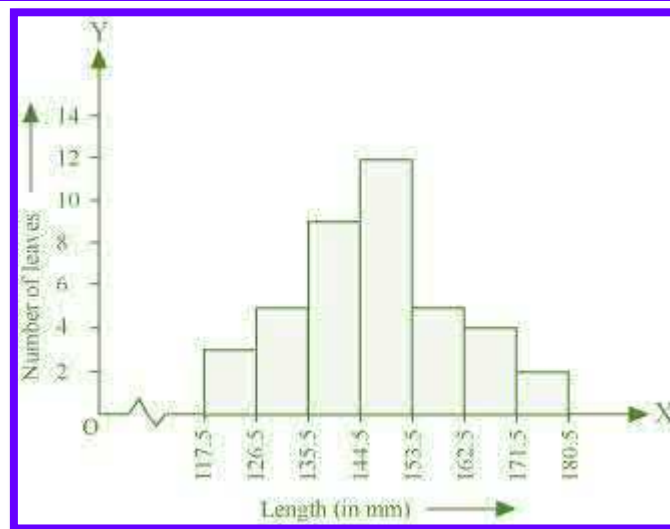
Length (in mm)	Number of leaves
118 – 126	3
127 – 135	5
136 – 144	9
145 – 153	12
154 – 162	5
163 – 171	4
172 – 180	2

- (i) Draw a histogram to represent the given data.
- (ii) Is there any other suitable graphical representation for the same data?
- (iii) Is it correct to conclude that the maximum number of leaves are 153 mm long?
Why?

Answer 4:

- (i) It can be observed that the length of leaves is represented in a discontinuous class interval having a difference of 1 in between them. Therefore, $1/2 = 0.5$ has to be added to each upper class limit and also have to subtract 0.5 from the lower class limits so as to make the class intervals continuous.

Length (in mm)	Number of leaves
117.5 – 126.5	3
126.5 – 135.5	5
135.5 – 144.5	9
144.5 – 153.5	12
153.5 – 162.5	5
162.5 – 171.5	4
171.5 – 180.5	2



Taking the length of leaves on x-axis and the number of leaves on y-axis, the histogram of this information can be drawn as above. Here, 1 unit on y-axis represents 2 leaves.

(ii) Other suitable graphical representation of this data is frequency polygon.

(iii) No, as maximum number of leaves (i.e., 12) has their length in between 144.5 mm and 153.5 mm. It is not necessary that all have their lengths as 153 mm.

Question 5:

The following table gives the life times of neon lamps:

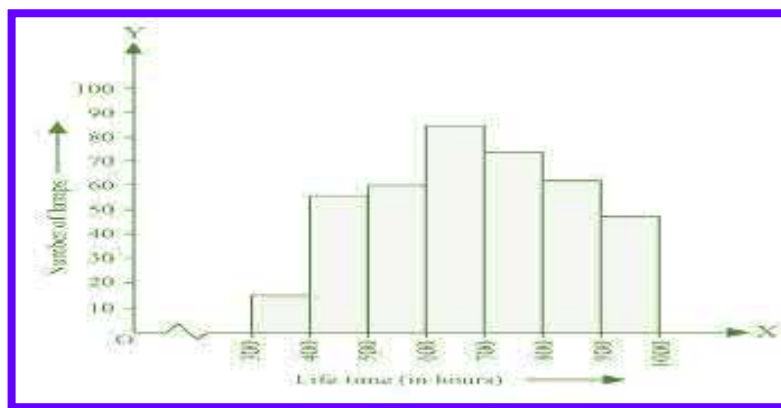
Length (in hours)	Number of lamps
300 – 400	14
400 – 500	56
500 – 600	60
600 – 700	86
700 – 800	74
800 – 900	62
900 – 1000	48

(i) Represent the given information with the help of a histogram.

(ii) How many lamps have a lifetime of more than 700 hours?

Answer 5:

(i) By taking life time (in hours) of neon lamps on x-axis and the number of lamps on y-axis, the histogram of the given information can be drawn as follows.



Here, 1 unit on y-axis represents 10 lamps.

- (ii) It can be concluded that the number of neon lamps having their lifetime more than 700 is the sum of the number of neon lamps having their lifetime as 700 – 800, 800 – 900, and 900 – 1000.

Therefore, the number of neon lamps having their lifetime more than 700 hours is 184.
(74 + 62 + 48 = 184)

Question 6:

The following table gives the distribution of students of two sections according to the mark obtained by them:

Section A		Section B	
Marks	Frequency	Marks	Frequency
0 – 10	3	0 – 10	5
10 – 20	9	10 – 20	19
20 – 30	17	20 – 30	15
30 – 40	12	30 – 40	10
40 – 50	9	40 – 50	1

Represent the marks of the students of both the sections on the same graph by two frequency polygons. From the two polygons compare the performance of the two sections.

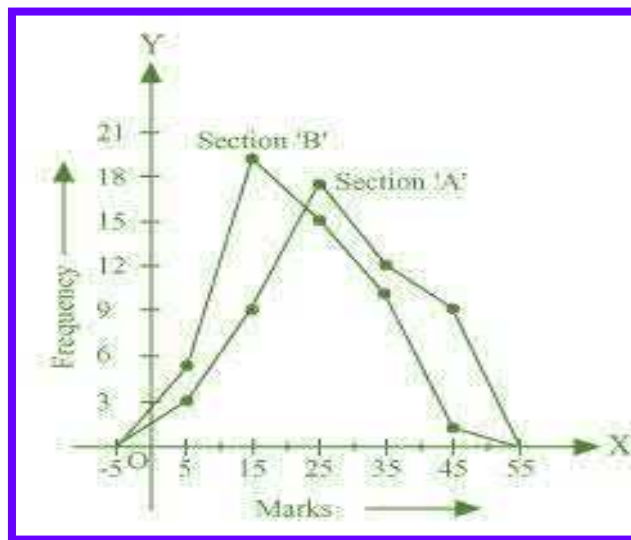
Answer 6:

We can find the class marks of the given class intervals by using the following formula.

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

Section A			Section B		
Marks	Class marks	Frequency	Marks	Class marks	Frequency
0 – 10	5	3	0 – 10	5	5
10 – 20	15	9	10 – 20	15	19
20 – 30	25	17	20 – 30	25	15
30 – 40	35	12	30 – 40	35	10
40 – 50	45	9	40 – 50	45	1

Taking class marks on x-axis and frequency on y-axis and choosing an appropriate scale (1 unit = 3 for y-axis), the frequency polygon can be drawn as follows.



It can be observed that the performance of students of section 'A' is better than the students of section 'B' in terms of good marks.

Question 7:

The runs scored by two teams A and B on the first 60 balls in a cricket match are given below:

Number of balls	Team A	Team B
1 – 6	2	5
7 – 12	1	6
13 – 18	8	2
19 – 24	9	10
25 – 30	4	5
31 – 36	5	6
37 – 42	6	3
43 – 48	10	4
49 – 54	6	8
55 – 60	2	10

Represent the data of both the teams on the same graph by frequency polygons.
[Hint: First make the class intervals continuous.]

Answer 7:

It can be observed that the class intervals of the given data are not continuous.

There is a gap of 1 in between them. Therefore, $1/2 = 0.5$ has to be added to the upper class limits and 0.5 has to be subtracted from the lower class limits.

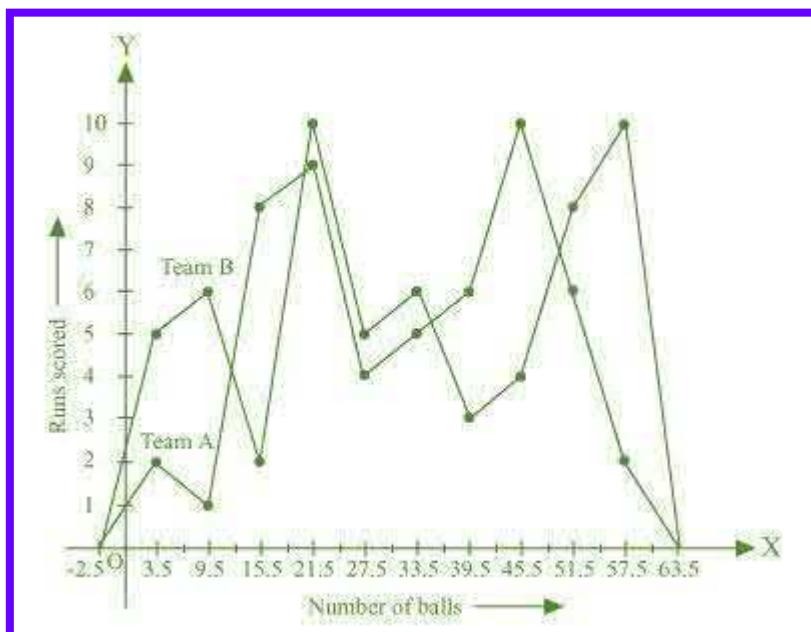
Also, class mark of each interval can be found by using the following formula.

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

Continuous data with class mark of each class interval can be represented as follows.

Number of balls	Class mark	Team A	Team B
0.5 – 6.5	3.5	2	5
6.5 – 12.5	9.5	1	6
12.5 – 18.5	15.5	8	2
18.5 – 24.5	21.5	9	10
24.5 – 30.5	27.5	4	5
30.5 – 36.5	33.5	5	6
36.5 – 42.5	39.5	6	3
42.5 – 48.5	45.5	10	4
48.5 – 54.5	51.5	6	8
54.5 – 60.5	57.5	2	10

By taking class marks on x-axis and runs scored on y-axis, a frequency polygon can be constructed as follows.



Question 8:

A random survey of the number of children of various age groups playing in 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.

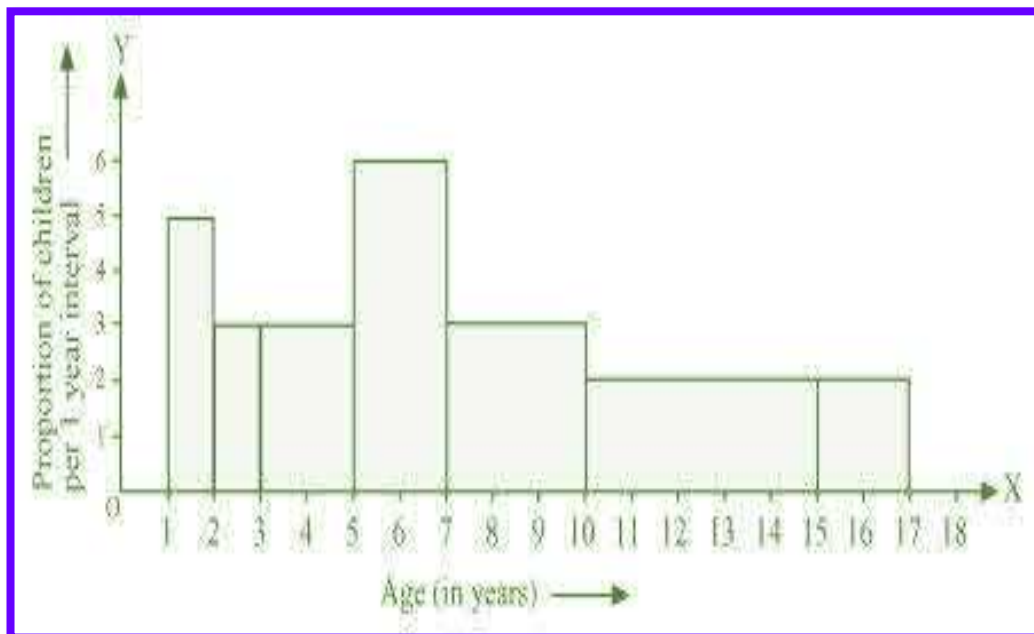
Answer 8:

Here, it can be observed that the data has class intervals of varying width. The proportion of children per 1 year interval can be calculated as follows.

Age (In years)	Frequency (Number of children)	Width of classes	Length of rectangle
1 – 2	5	1	$\frac{5 \times 1}{1} = 5$
2 – 3	3	1	$\frac{3 \times 1}{1} = 3$
3 – 5	6	2	$\frac{6 \times 1}{2} = 3$

5 – 7	12	2	$\frac{12 \times 1}{2} = 6$
7 – 10	9	3	$\frac{9 \times 1}{3} = 3$
10 – 15	10	5	$\frac{10 \times 1}{5} = 2$
15 – 17	4	2	$\frac{4 \times 1}{2} = 2$

Taking the age of children on x-axis and proportion of children per 1 year interval on y-axis, the histogram can be drawn as follows.



Question 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 surname lie.

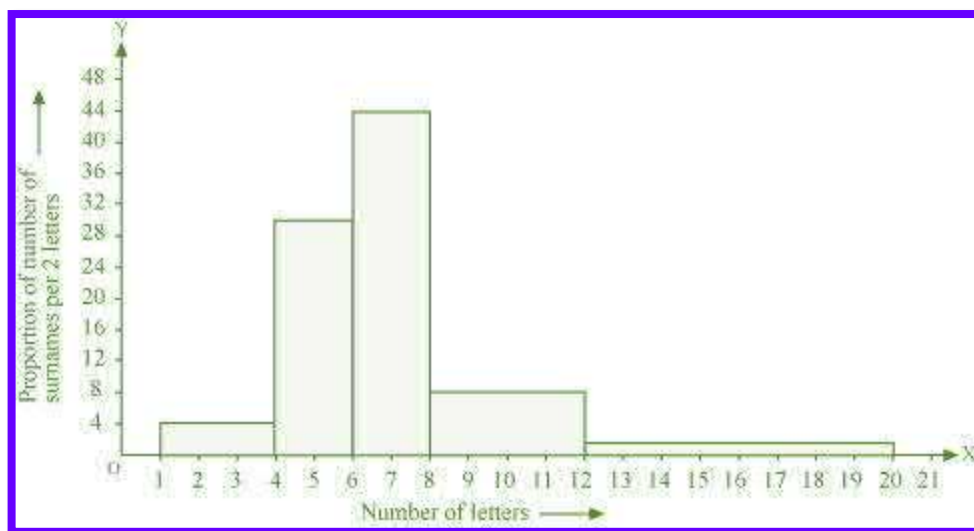
Answer 9:

(i) Here, it can be observed that the data has class intervals of varying width. The proportion of the number of surnames per 2 letters interval can be calculated as follows.

Number of letters	Frequency (Number of surnames)	Width of class	Length of rectangle
1 – 4	6	3	$\frac{6 \times 2}{3} = 4$
4 – 6	30	2	$\frac{30 \times 2}{2} = 30$
6 – 8	44	2	$\frac{44 \times 2}{2} = 44$

8 – 12	16	4	$\frac{16 \times 2}{4} = 8$
12 – 20	4	8	$\frac{4 \times 2}{8} = 1$

By taking the number of letters on x-axis and the proportion of the number of surnames per 2 letters interval on y-axis and choosing an appropriate scale (1 unit = 4 students for y axis), the histogram can be constructed as follows.



(ii) The class interval in which the maximum number of surnames lies is 6 – 8 as it has 44 surnames in it i.e., the maximum for this data.

Exercise 14.4

Question 1:

The following number of goals was 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.

Answer 1:

The number of goals scored by the team is

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

$$\begin{aligned}\text{Mean of data} &= \frac{\text{Sum of all observations}}{\text{Total number of observations}} \\ \text{Mean score} &= \frac{2+3+4+5+0+1+3+3+4+3}{10} \\ &= \frac{28}{10} = 2.8 \\ &= 2.8 \text{ goals}\end{aligned}$$

Arranging the number of goals in ascending order,

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

The number of observations is 10, which is an even number. Therefore,

$$\text{Median} = \frac{5^{\text{th}} \text{ term} + 6^{\text{th}} \text{ term}}{2} = \frac{3 + 3}{2} = \frac{6}{2} = 3$$

Mode of data is the observation with the maximum frequency in data.

Therefore, the mode score of data is 3 as it has the maximum frequency as 4 in the data.

Question 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.

Answer 2:

The marks of 15 students in mathematics test are

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

$$\begin{aligned}\text{Mean of data} &= \frac{\text{Sum of all observation}}{\text{Total number of observation}} \\ &= \frac{41+39+48+52+46+62+54+40+96+52+98+40+42+52+60}{15} \\ &= \frac{822}{15} = 54.8\end{aligned}$$

Arranging the scores obtained by 15 students in an ascending order,

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

As the number of observations is 15 which is odd, therefore, the median of data will

be $\frac{(15+1)^{th}}{2} = 8^{th}$ observation whether the data is arranged in an ascending or descending order.

Therefore, median score of data = 52

Mode of data is the observation with the maximum frequency in data. Therefore, mode of this data is 52 having the highest frequency in data as 3.

Question 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

Answer 3:

It can be observed that the total number of observations in the given data is 10 (even number). Therefore, the median of this data will be the mean of $10/2$ i.e., 5th and $(10/2)+1$ i.e., 6th term.

$$\text{Therefore, median of data} = \frac{5^{\text{th}} \text{ observation} + 6^{\text{th}} \text{ observation}}{2}$$

$$\Rightarrow 63 = \frac{x + x + 2}{2}$$

$$\Rightarrow 63 = \frac{2x + 2}{2}$$

$$\Rightarrow 63 = x + 1$$

$$\Rightarrow x = 62$$

Question 4:

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

Answer 4:

Arranging the data in an ascending order,

14, 14, 14, 14, 17, 18, 18, 18, 22, 23, 25, 28

It can be observed that 14 has the highest frequency, i.e. 4, in the given data. Therefore, mode of the given data is 14.

Question 5:

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

Salary (in Rs)	Number of workers
3000	16
4000	12
5000	10
6000	8
7000	6
8000	4
9000	3
1000	1
Total	60

Answer 5:
Using the table, the mean can be calculated as follows.

Salary (in Rs) (x_i)	Number of workers (f_i)	fixi
3000	16	$3000 \times 16 = 48000$
4000	12	$4000 \times 12 = 48000$
5000	10	$5000 \times 10 = 50000$
6000	8	$6000 \times 8 = 48000$
7000	6	$7000 \times 6 = 42000$
8000	4	$8000 \times 4 = 32000$
9000	3	$9000 \times 3 = 27000$
10000	1	$10000 \times 1 = 10000$
Total	$\sum f_i = 60$	$\sum f_i x_i = 305000$

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

$$\begin{aligned} \text{Mean salary} &= \frac{305000}{60} \\ &= 5083.33 \end{aligned}$$

Therefore, mean salary of 60 workers is Rs 5083.33.

Question 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.

Answer 6:

When any data has a few observations such that these are very far from the other observations in it, it is better to calculate the median than the mean of the data as median gives a better estimate of average in this case.

- (i) Consider the following example – the following data represents the heights of the members of a family.

154.9 cm, 162.8 cm, 170.6 cm, 158.8 cm, 163.3 cm, 166.8 cm, 160.2 cm

In this case, it can be observed that the observations in the given data are close to each other. Therefore, mean will be calculated as an appropriate measure of central tendency.

- (ii) The following data represents the marks obtained by 12 students in a test.

48, 59, 46, 52, 54, 46, 97, 42, 49, 58, 60, 99

In this case, it can be observed that there are some observations which are very far from other observations. Therefore, here, median will be calculated as an appropriate measure of central tendency.