cuemath
THE MATH EXPERT

Get better at Math. Get better at everything.

Come experience the Cuemath methodology and ensure your child stays ahead at math this summer.

Adaptive Platform | **Interactive Visual Simulations** | **Personalized Attention**

For Grades 1 - 10

LIVE online classes by trained and certified experts.

Get the Cuemath advantage

Book a FREE trial class

Chapter 14 - Statistics

Exercise 14.1 (Page 239 of Grade 09 NCERT Textbook)

Q1. Give five examples of data that you can collect from your day to day life.

Difficulty Level:
Easy

Solution:
Think of data we can collect in our day to day life, in and around where we live

1) Number of houses that own a four-wheeler in our housing society.
2) Monthly grocery expenses of our family.
3) Number of people who have used e-services of the state govt over a year.
4) Number of students who have enrolled for math Olympiad in your class.
5) Population increase in percentage over the past five years in our city.

Q2. Identify the data in 14.1(1) as primary or secondary.

Difficulty Level:
Easy

Reasoning:
Think how you have collected data:

1) Did you collect the data all by yourself investigating around? If yes, primary.
2) Did you obtain the data from a source where the information has been stored already? If yes, secondary.

Solution:
Based on the observation, data in 2 and 4 are primary and the data in 1, 3 and 5 are secondary.

Chapter 14 - Statistics

Exercise 14.2 (Page 245 of Grade 09 NCERT Textbook)

Q1. The blood groups of 30 students of Class VIII are recorded 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?

Difficulty Level:
Easy

What is the known/given?
The blood group of 30 students of class VIII.

What is the unknown?
The most common blood group and the rarest blood group.

Reasoning:
By drawing frequency distribution table, we can check which group is occurring most and which group is occurring least time.

Solution:
Let us sort the data using a table to make the data more easily understandable and its important features visible.

This table is known as frequency distribution table.
'Frequency' of a particular data refers to the number of times the data value occurs.
In our case frequency refer to the number of students having the same blood group.
∴The blood group of 30 students can be shown as follows:

BLOOD GROUP	NUMBER OF STUDENTS (frequency)
A B O AB	9 6 12 3
Total	**30**

So, we can see easily now, from the table that:

- The most common (the most frequently occurring) blood group is 'O'.
- The rarest blood group (least frequently occurring) is 'AB'.

Q2. 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 features do you observe from this tabular representation?

Difficulty Level:
Easy

What is the known/given?

- The distance in km of 40 engineers from their residence to work.
- Class size of 5
- Class interval as 0-5 (5 not included)

What is the unknown?
Main features of the tabular representation to be drawn from the given data.

Reasoning:
By drawing frequency distribution, we can observe main features.

Solution:
Since the given data is large, we need to group the data into classes of size each 5 and construct a table which will make the important features of the data visible.

This table is known as grouped frequency distribution table.

The class intervals will be 0-5, 5-10, 10-15 and so on.

The distance (in km) of 40 engineers from their residence to their place work can be represented as follows (with the help of tally marks)

DISTANCE (in km)	TALLY MARKS	NO OF ENGINEERS (frequency)
0-5	𝍸	5
5-10	𝍸 𝍸 |	11
10-15	𝍸 𝍸 |	11
15-20	𝍸 | | | |	9
20-25	|	1
25-30	|	1
30-35	||	2
Total		**40**

The following features can be observed from the table.

- 5 engineers have their houses below 5 km distance.
- A majority of engineers (36) have their houses below 20 km distance.
- Only a few engineers (4) have their houses at 20 km and above distance.

Q3. 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

1) Construct a grouped frequency distribution table with classes 84 – 86, 86 – 88, etc.
2) Which month or season do you think this data is about?
3) What is the range of this data?

Difficulty Level:
Easy

What is the known/given?

- The relative humidity (in %) of a certain city over a month of 30 days.
- Class size is 2, so the class intervals will be 84-86, 86-88, 88-90, 90-92 and so on

What is the unknown?

- Constructing a grouped frequency distribution table
- The month / season that the data talks about.
- Range of this data [which is the difference of the highest and the lowest values in the data]

Reasoning:
By drawing frequency distribution table, we can observe data.
Range of data = maximum value – minimum value

Solution:
Construct a grouped frequency distribution table with class size of 2.
The relation humidity (in %) of a certain city for a month can be represented as follows.

RELATIVE HUMIDITY (in %)	NO 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**

The following features can be observed:

- The relative humidity was 92% and above, over a period of 24 days ($\frac{4}{5}$ th of a month)
- Since the relative humidity % is very high, it must be a data from a month of rainy season.

Range of data = maximum value – minimum value
= 99.2 – 84.9
= 14.3

Q4. The heights of 50 students, measured to the nearest centimeters, have been found to be as follows:

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

Difficulty Level:
Easy

What is the known/given?

- The heights of 50 students in cm.
- Class size is 5 [from the given instruction of class intervals to be taken as 160-165, 165-170, 170-175 …. and so on]

What is the unknown?

- Constructing a grouped frequency distribution table.
- Making inferences and drawing conclusions from the data of "heights of students".

Reasoning:
According to class interval we can check how many numbers lie in between, like this we can draw frequency distribution table and can be conclude about height from table.

Solution:
A grouped frequency distribution table with a class size of 5, for the heights of 50 students can be constructed as follows

HEIGHT (in cm)	NO OF STUDENTS (frequency)
150-155	12
155-160	9
160-165	14
165-170	10
170-175	5
TOTAL	**50**

Since '150' is the lowest value, we started the 1st class interval at 150.

The following features / conclusions can be drawn from the table:

- A majority of 70% of the students (35 students) are below 165 cm height.
- Only a 10% of the students (5 students) are at a height of 170cm or above.

Q5. 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?

Difficulty Level:
Easy

What is the known/given?

- Concentration of sulphur dioxide in the air (in ppm) of a certain city observed over 30 days
- Class intervals of '0.00 – 0.04', '0.04 - 0.08', and so on. Hence class size is 0.04.

What is the unknown?
Constructing a grouped frequency table for the given data which will help us to find out for how many days was the concentration of sulphur dioxide more than 0.11 ppm.

Reasoning:
According to class interval we can check how many numbers lie in between, like this we can draw frequency distribution table and can be conclude about no. of days in which Sulphur dioxide more than 0.11 parts per million from table.

Solution:
A grouped frequency distribution table with a class size of '0.04' needs to be constructed for the given data.

CONCENTRATION OF SULPHUR DI OXIDE (in ppm)	NUMBER OF DAYS (frequency)
0.0-0.04	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**

From this table we can see that the number of days during which the concentration of Sulphur dioxide is more than 0.11 ppm, falls over three class intervals, 0.12-0.16, 0.16-0.20 and 0.20-0.24.

So, $2+4+2 \Rightarrow 8$
8 days had a concentration of Sulphur dioxide more than 0.11 ppm.

Q6. 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
1	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.

Difficulty Level:
Very Easy

What is the known/given?

- Number of times three coins were tossed simultaneously.
- Number of 'heads' that occurred during each toss.

What is the unknown?

- Constructing a required frequency distribution table.

Reasoning:
We can count in each case and can be written in front of them, like this we can draw frequency distribution table.

Solution:
By observing the given data, we can prepare an ungrouped frequency distribution table as follows.

NUMBER OF HEADS	TALLY MARKS	FREQUENCY OF OCCURRENCE
0	𝍸 |	6
1	𝍸 𝍸	10
2	𝍸 ||||	9
3	𝍸	5
TOTAL		**30**

Q7. 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?

Difficulty Level:
Easy

What is the known/given?
- The value of 'π' up to 50 decimal places.

What is the unknown?
- Constructing a required frequency distribution table
- And finding the most and the least frequently occurring digits.

Reasoning:
We can count each digit and can be written in front of them, like this we can draw frequency distribution table and can be concluded about most or least frequently occurring digits.

Solution:
We can represent the data by constructing a simple, ungrouped frequency distribution table as follows.

DIGITS	TALLY MARKS	FREQUENCY OF OCCURRENCE
0	𝍸	2
1	𝍸	5
2	𝍸	5
3	𝍸 |	8
4	||||	4
5	𝍸	5
6	||||	4
7	||||	4
8	𝍸	5
9	𝍸 |	8
TOTAL		**50**

It can be easily observed from the table that

- The most frequently occurring digits are 3 and 9 with a max frequency of 8
- The least frequently occurring digit is '0' with a lowest frequency of 2.

Q8. 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?

Difficulty Level:

Medium

What is the known/given?

- Number of hours of watching TV programmes by 30 children in a weak.
- Class width '5'
- A class interval, such as 5-10

What is the unknown?

- A grouped frequency distribution table with a class width of '5'
- Number of children who watched television for 15 or more hours per week.

Solution:

A grouped frequency distribution table can be constructed as follows with the class intervals of 0-5, 5-10, 10-15, 15-20 and so on.

HOURS	NUMBERS OF CHILDREN (frequency)
0-5	10
5-10	13
10-15	5
15-20	2
TOTAL	**30**

We can observe from the table, that the number of children who watched television for 15 or more hours a week is 2 (which falls under the class interval '15-20')

Q9. 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 interval 2-2.5.

Difficulty Level:

Medium

What is the known/given?

- The life span (in years) of 40 car batteries manufactured by a company.
- Class size is '0.5'
- A class interval such as 2-2.5

What is the unknown?

- Constructing a grouped frequency distribution table with a class size of 0.5.

Solution:

The required grouped frequency distribution table can be constructed as follows, with class intervals of '2-2.5', '2.5-3', '3-3.5', '3.5-4', and so on.

LIFE SPAN OF BATTERIES (in years)	NUMBER OF BATTERIES (frequency)
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**

Chapter 14 - Statistics

Exercise 14.3 (Page 258 of Grade 09 NCERT Textbook)

Q1. A survey conducted by an organization 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?
(iii) 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.

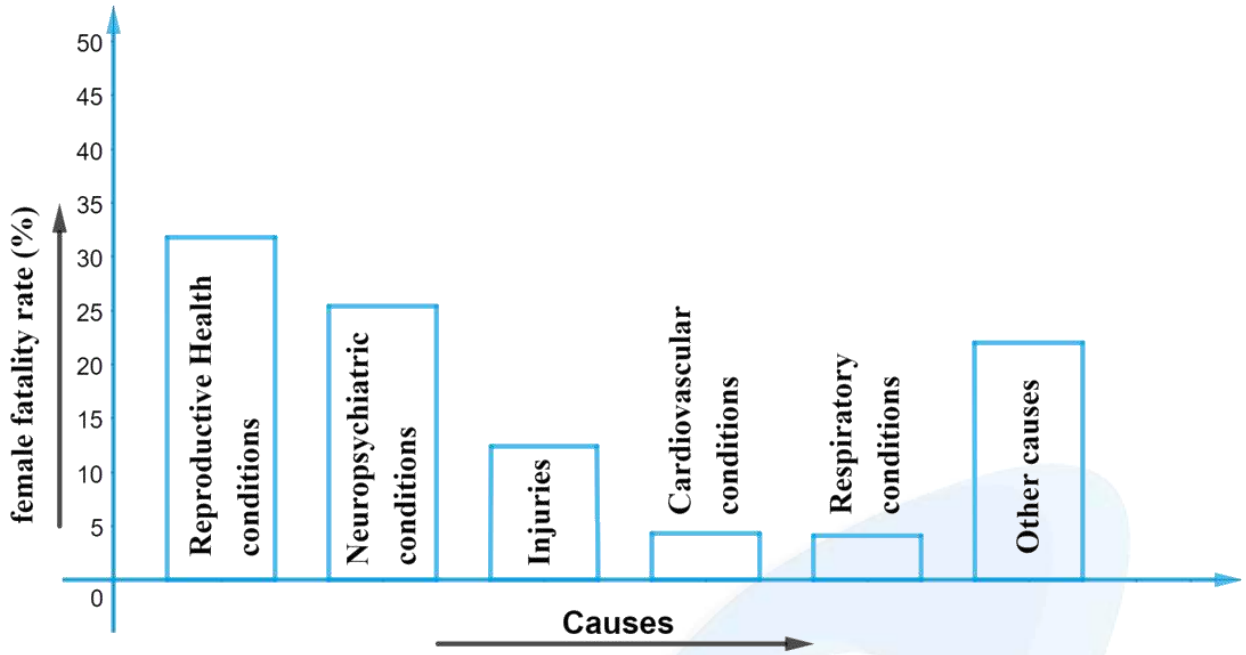
Difficulty Level:
Easy

What is the known/given?
Female fatality rate (%)

What is the Unknown?
(i) Represent the given information graphically
(ii) Major cause of women's ill health and death worldwide.
(iii) Factors that play a major role in the major cause of women's illness and death.

Reasoning:
The above data can be represented graphically using a bar graph as below
(i) We will represent the causes on the x-axis, maintaining equal widths for all bars and with equal gaps in between. One cause will be represented by one unit.
(ii) We will represent the fatality rate on the y-axis with a scale of unit as 5% as the max value.

Solution:



It can be visualized from the graph that

(i) Fatality rate due to reproductive health condition is more than double that of fatality rate owing to 'injuries.

Reproductive health condition can be considered as the major cause of women's ill health worldwide.

(ii) The factors contributing to (i) can be as follows:
- (a) Lack of awareness / knowledge about reproductive health conditions.
- (b) Lack of emergency medical facilities.
- (c) Knowledge about treatment.
- (d) Affordability of treatment cost.

Q2. 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.

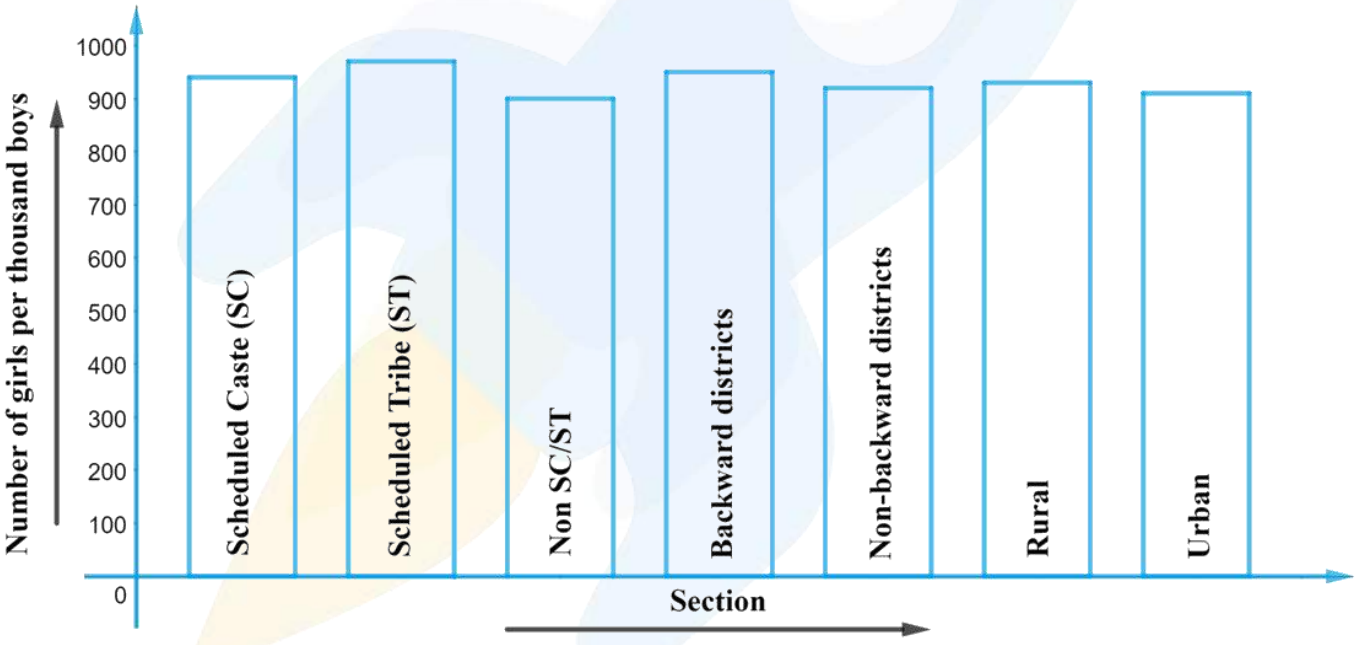
Difficulty Level:
Easy

What is known?
Number of girls / thousands of boys

What is the Unknown:
(i) Graphical representation of the given information.
(ii) Conclusions that can be arrived from the graph.

Reasoning:
The above data can be represented graphically using a bar graph as below:
(i) Represent the 'section' on the x-axis and 'number of girls per thousand boys on the y-axis.
(ii) We will select a scale of 1unit=100 girls for y-axis as the max value did not exceed low.
(iii) We will maintain equal width for all the bars in x-axis and also maintain equal gaps in between the bars.

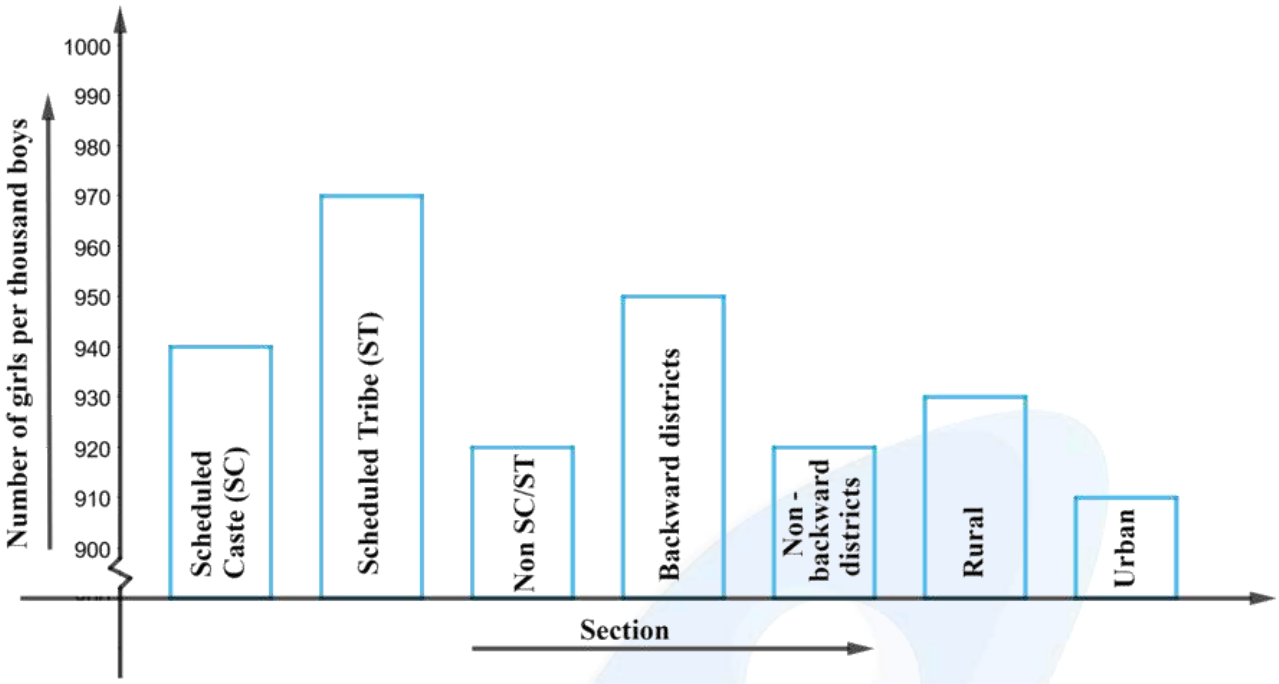
Solution:



It can be observed from the graph that:
(i) The number of girls per thousand boys is highest for ST category.
(ii) The number of girls per thousand boys is lowest for urban category.
(iii) The 'number of given / 1000' boys are higher in rural area than in urban.
(iv) The 'number of given / 1000' boys are also higher in backward districts them in non-backward districts.

Point to note:
(i) The variations in values occur mostly in the range of 900-970.
(ii) The values are same until 900.

(iii) Hence to make the variations more visible and distinct and for a better study of the graph, we can keep the starting value on the y-axis as '900' with a scale of 1unit as '10' and re construct the graph as



shown below:

As you can see, the above graph shows a clear picture of the data and helps us to arrive at conclusions easily.

Q3. 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?

Difficulty Level:
Easy

What is known/given?
The seats won by different political parties in the polling outcome of a state assembly elections.

What is the Unknown?
(i) Bar graph to represent the polling results.
(ii) The political party which won the maximum number of seats.

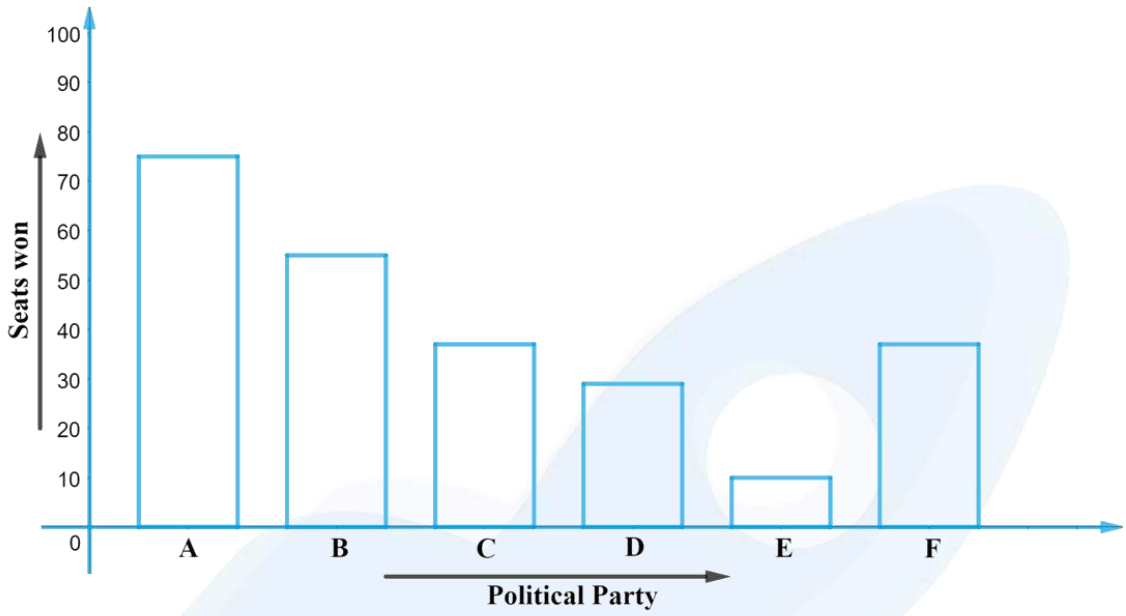
Reasoning:

The given data can be represented through a bar graph by:

(i) Taking 'Political Party' on *x*-axis and seats won on *y*-axis with a scale of 1unit = 10 seats on *y*-axis.

(ii) Maintaining equal width for all the bars in x-axis with equal gaps in between them.

Solution:



From the graph, it can be observed that the political party 'A' won the maximum number of results.

Q4. 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. [Hint: First make the class intervals continuous]

(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?

Difficulty Level:
Medium

What is known/given?
The length of 40 leaves of a plant are measured correct to one millimeter.

What is unknown?
(i) A histogram to represent the given data
(ii) Another suitable way of graphical representation for the same data.
(iii) Whether the maximum number of leaves are 153mm long.

Reasoning:
(i) From the given data, it can be observed that the length of leaves are represented in discontinuous class intervals, with a difference of 1 unit in between them.
(ii) To make the class intervals continuous, we can add 0.5

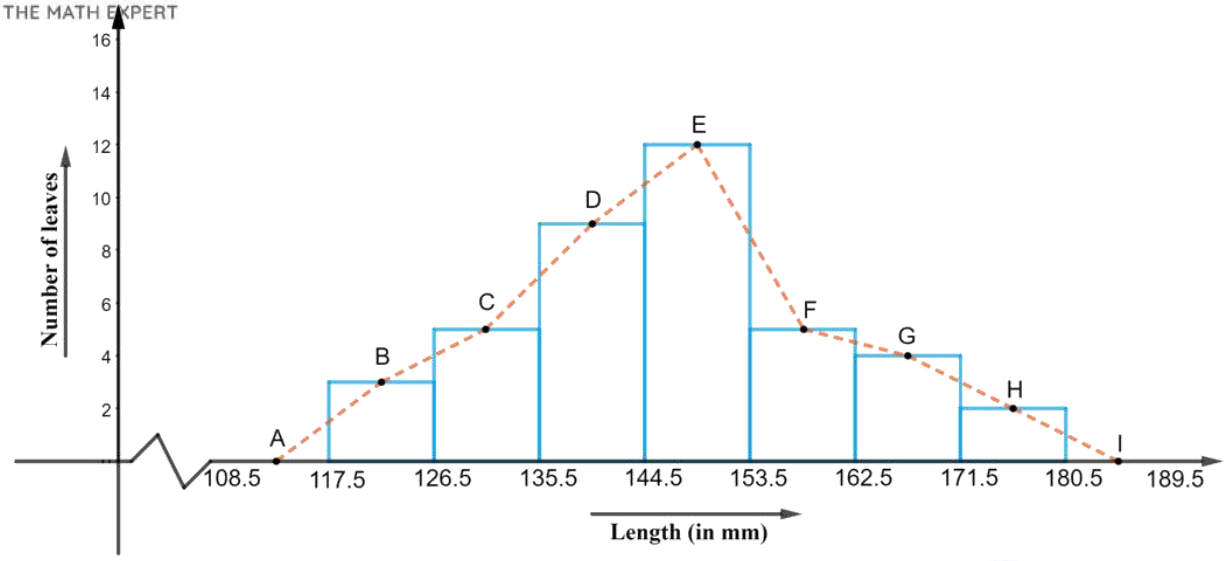
Solution:
Find the difference between upper limit of a class and the lower limit of its succeeding class. We then add half of this difference to each of the upper limits and subtract the same from each of the lower limit since the difference is 1 (127 – 126 = 1) half of 1 is 0.5

Now the frequency distribution table will look like as below.

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

The above data can be represented graphically through a histogram as below:

(i) Represent "Length of leaves (in mm) on *x*-axis and 'number of leaves' in *y*-axis.
(ii) And with a scale of '1 unit = 2 leaves', since the lower-class value is 2 and the highest is 12.
(iii) Also, since the first-class interval is starting from 117.5 and not zero, we show it on the graph by making a kink on a break on the *x*-axis.
(iv) We will make now draw rectangular bars of width equal in sizes and lengths according to the frequencies of class intervals. For eg, the rectangular bar for the class interval 117.5 – 126.5 will be of width 1cm and length 1.5 cm.



It can be observed from the above graph that the other suitable graphical representation of the same data would be a frequency polygon as below:

Join the mid points of the upper sides of all the rectangular bars through line segments (as shown in red dotted lines)

Let them B, C, D, E, F, G, H. To complete it look like a polygon figure, assume that there is a class interval with '0' frequency, before '117.5 – 126.5', and one after '171.5 – 180.5'. Now mark their imaginary mid points as 'a' and 'I' respectively.

Now, "ABCDEFGHI" is the frequency polygon that can be constructed for the given data. (Refer the red dotted line segment)

The maximum number of leaves have their length lie between 144.5 mm and 153.5 mm. Hence, we cannot say all leaves have their lengths as 153 mm.

Q5. The following table gives the lifetimes of 400 neon lamps:

Lifetime (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?

Difficulty Level:
Easy

What is known/given?

Lifetimes of 400 neon lamps.

What is unknown?

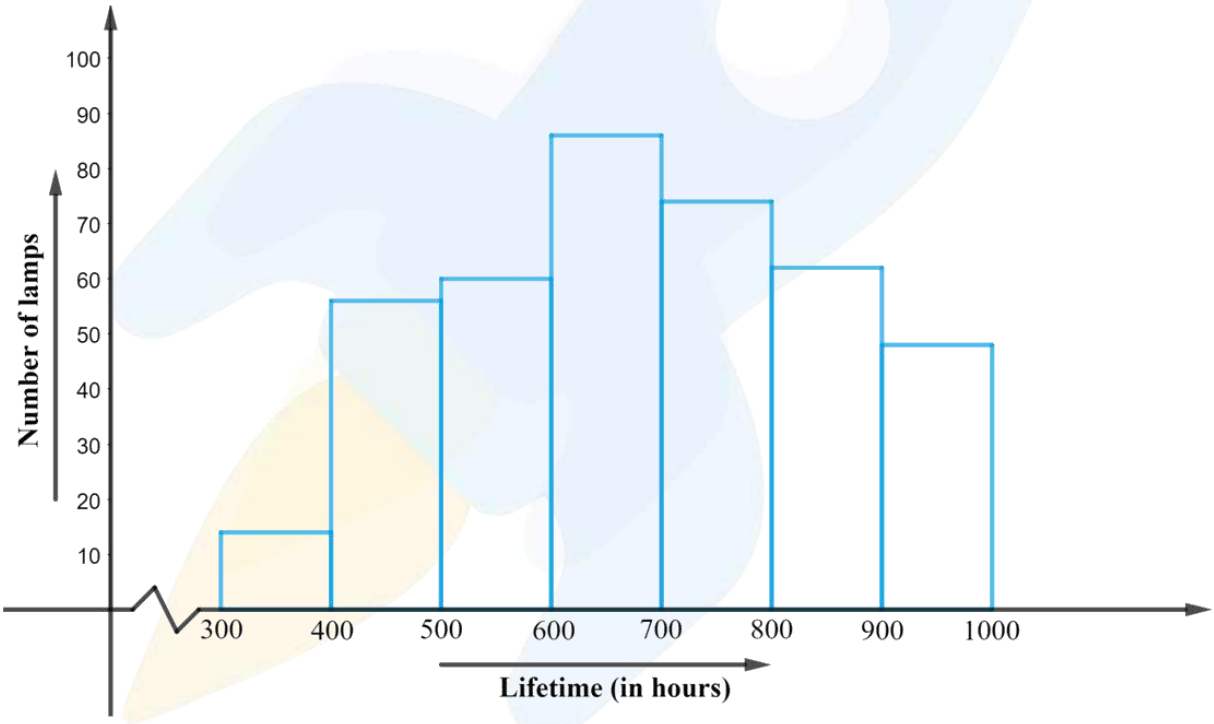
(i) A histogram representation for the given data.
(ii) Number of lamps that have a lifetime of more than 700 hours.

Reasoning:

The given data can be represented with the help of a histogram as above:

(i) Represent the 'lifetime (in hours) in *x*-axis.
(ii) Represent the 'number of lamps' in *y*-axis.
(iii) Class intervals are continuous.
(iv) Take "1 unit = 10 lamps" on *y*–axis as the lowest value of frequency is 14 and highest 86.
(v) Also, since the first interval is starting from 300 and not '0', we show it by marking a 'kink' or a break on the *x*-axis.

Answer:



From the above graph, it can be concluded that:

(i) The number of neon lamps having their lifetime more than 700 is under class intervals "700 – 800, 800 – 900, 900 – 1000"
Hence, their corresponding frequencies when added up will be (72 + 62 + 48) 184 lamps.

Q6. The following table gives the distribution of students of two sections according to the marks 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.

Difficulty Level:
Medium

What is known/given?
The distribution of students of two sections according to the marks obtained by them.

What is unknown?
(i) A frequency polygon to represent the marks of the students of both the sections.
(ii) Comparisons of performance of those two sections.

Reasoning:
Frequency polygons can also be drawn independently without drawing histograms. For this requires the midpoints of the class-intervals used in the data. The mid-points are called class-marks.

$$\text{Class Mark} = \frac{\text{Upper Limit} + \text{Lower Limit}}{2}$$

Solution:

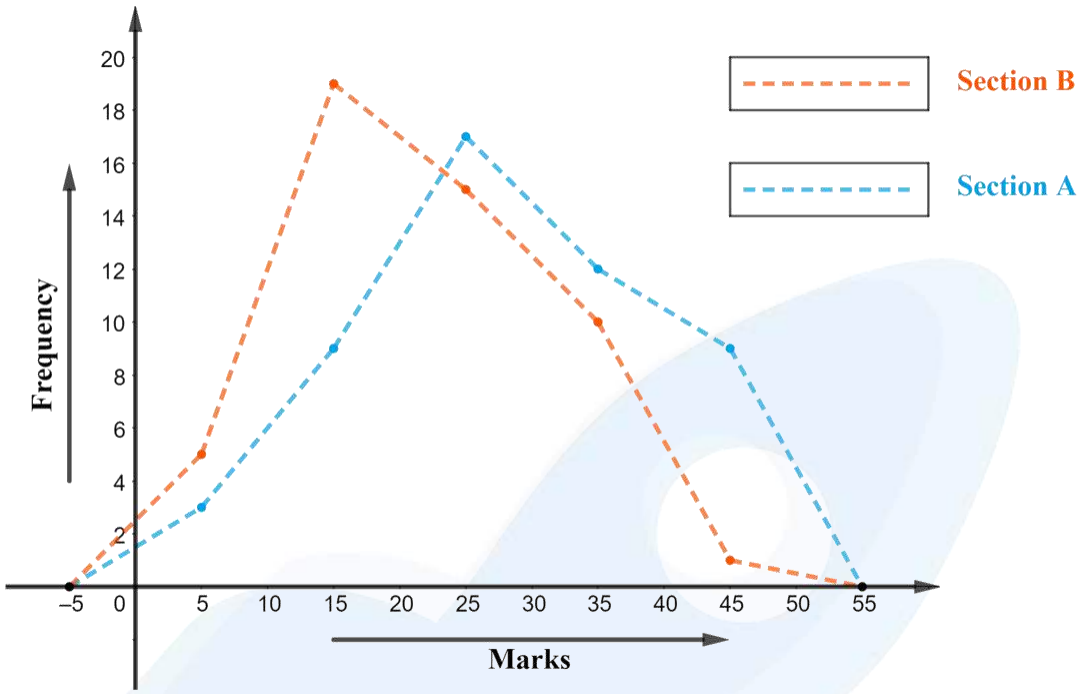
Now, the data table with inclusion of class marks:

	Section A		Section B		
Marks	**Class Mark**	**Frequency**	**Marks**	**Class Mark**	**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

The frequency polygon for the above data can be constructed by:

(i) Taking class marks on *x*-axis.

(ii) Taking 'frequency' on *y*-axis with an appropriate scale of 1 unit = 2 students as the lowest frequency value observed is 1 and the highest frequency value is 19.



It can be observed that the performance of students of Section 'A' is better than the students of Section B as section 'A' shows more students securing marks between '40 – 50' class and '30 – 40' class.

Q7. 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.]

Difficulty Level:
Medium

What is known/given?
The runs scored by two teams A and B on the first 60 balls in a cricket match are given.

What is unknown?
A frequency polygon to represent the data of both the teams.

Reasoning:

(i) It can be observed from the given data that the class intervals of the given data are not continuous. There is a gap of 'I' unit between them. So, to make the class intervals continuous, 0.5 has to be added to every upper-class limit and 0.5 has to be subtracted from the lower-class limit:

(ii) Class mark should also be found as below:

$$\text{Class Mark} = \frac{\text{Upper Limit} + \text{Lower Limit}}{2}$$

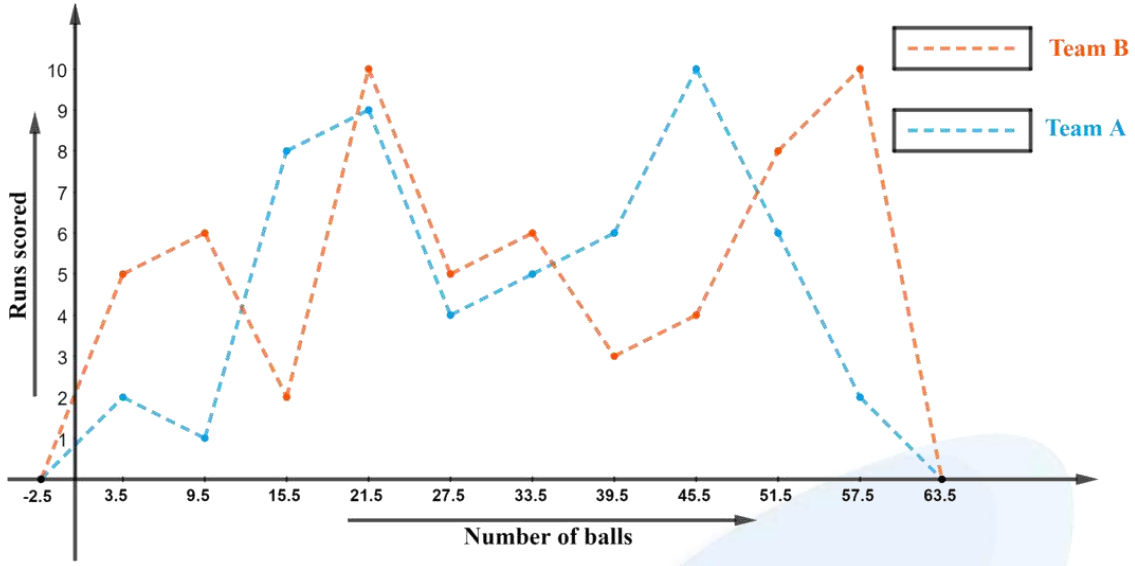
Solution:
The data table with continuous interval and with class mark is as below:

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

The frequency polygon for the above data can be constructed by

(i) Number of balls on x-axis

(ii) Runs scored on y-axis with an approximate scale of "1 unit = 1 run" as the lowest run was at 1 and the highest was at 10.



Q8. 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.

Difficulty Level:
Medium

What is known/given?
A random survey of the number of children of various age groups playing in a park.

What is unknown?
A Histogram to represent the data.

Reasoning:

(i) From the given data, we can observe that the class intervals have varying width. This will make the rectangular bars to have varying widths and will give us a misleading picture of the data.
(ii) The areas of the rectangles should be proportional to the frequencies in a histogram.

(iii) So, we need to make certain modification in the lengths so that area's again proportional to the frequencies. For that,

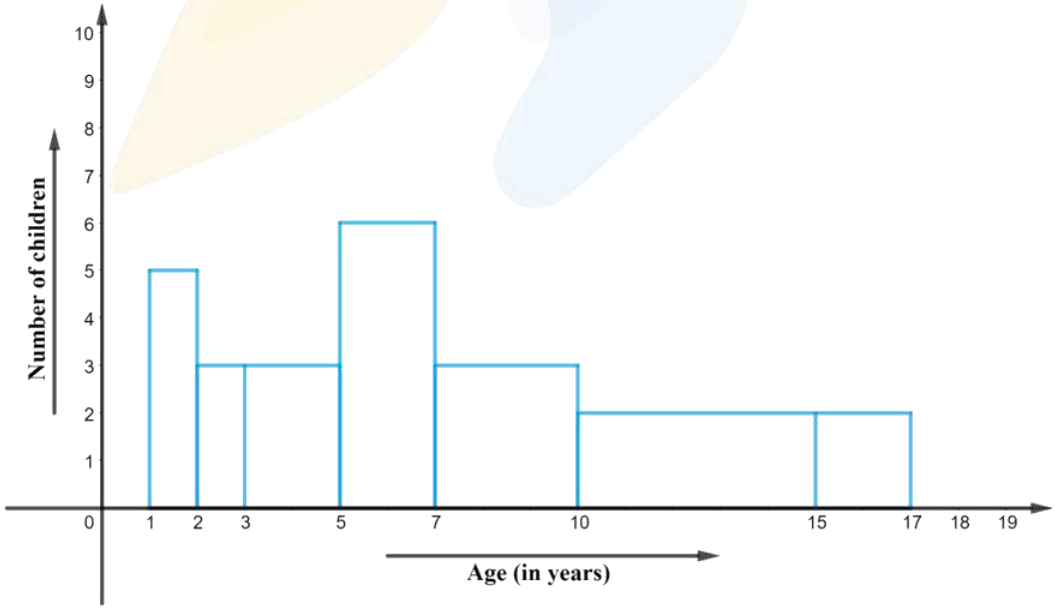
(a) Select a class interval in the minimum. Class size
(b) The lengths of the rectangles are then modified to be proportionate to the class size 1.
(c) For instance, when the class size is 5, the length of the rectangle is 10. So when the class size is 1, the length of the rectangle will be $\frac{10}{5}\times 1 = 2$

Solution:

We need to proceed in similar manner, to get the following table:

Age (in years)	Number of children	Width of the Class	Length of the 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	$\frac{6\times 1}{2}\times 1=3$
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	$\frac{4\times 1}{2}=2$

Will take the age of children on x-axis and proportion of children per 1 year interval per year on y-axis, the histogram can be



Q9. 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.

Difficulty Level:
Medium

What is known/given?
Distribution of the number of letters in the English alphabet in the surnames.

What is unknown?
(i) A Histogram to depict the given data
(ii) The class interval in which the maximum number of surnames lies.

Reasoning:
(i) It can be observed from the given data that it has class intervals of varying width.
(ii) The proportion of number of surnames per 2 letter interval (class interval of min class size for reference) can be made.

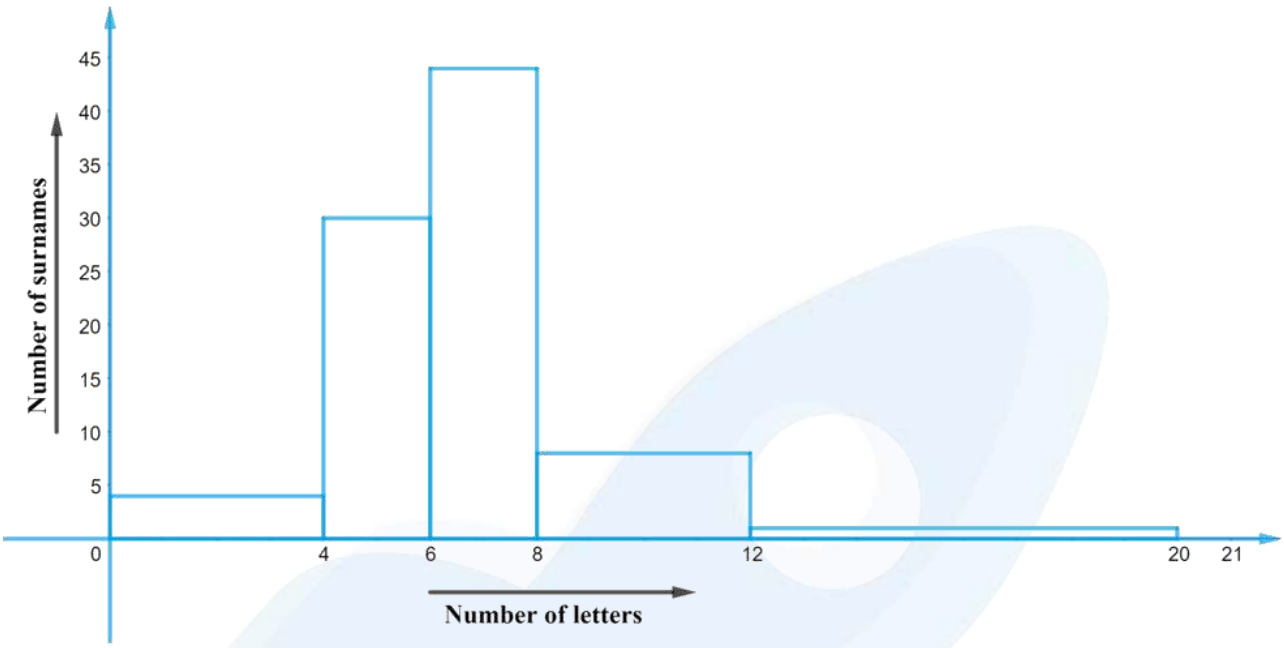
Solution:
The proportion of number of surnames per 2 letter interval (class interval of min class size for reference) can be made as follows

Number of letters	Number of surnames	Width of the 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	$\frac{4\times 2}{8}=1$

We will take the number of letters on x-axis and proportion of the number of surnames per 2 letter interval on *y*-axis.

And choose an appropriate scale of 1 unit = 4 surnames for y-axis.

The histogram can be constructed as follows:



The class interval in which the maximum number of surnames lies 6 – 8 and has surnames in it.

Chapter 14 - Statistics

Exercise (14.4)

Q1. 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.

Difficulty level:
Easy

What is known/given?
Number of goals was scored by a team in a series of 10 matches

What is unknown?
Mean, median and mode.

Reasoning:
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.

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 can be calculated based on no. of observation are even or odd.

The mode is that value of the observation which occurs most frequently.

Solution:
The number of goals scored by the team is
2, 3, 4, 5, 0, 1, 3, 3, 4, 3

$$\text{Mean of data} = \frac{\text{Sum of all observation}}{\text{Total number of observation}}$$

$$\text{Mean score} = \frac{2+3+4+5+0+1+3+3+4+3}{10}$$

$$= \frac{28}{10} = 2.8$$

$$= 2.8 \text{ goals}$$

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, median score will be the mean of $\frac{10}{2}$ i.e., 5^{th} and $\frac{10}{2}+1$ i.e., 6th observation while arranged in ascending or descending order.

$$\text{Median score} = \frac{5^{th}\text{ observation} + 6^{th}\text{ observation}}{2}$$

$$= \frac{3+3}{2} = \frac{6}{2} = 3$$

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.

Q2. 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.

Difficulty level:
Easy

What is known/given?
Marks obtained by 15 students in mathematics test

What is unknown?
Mean, median and mode.

Reasoning:
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.

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 can be calculated based on no. of observation are even or odd.

The mode is that value of the observation which occurs most frequently.

Solution:
The marks of 15 students in mathematics test are

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

$$\text{Mean of data} = \frac{\text{sum of all observation}}{\text{Total number of observation}}$$

$$= \frac{41+39+48+52+46+62+54+40+96+52+98+40+42+52+60}{15}$$

$$= \frac{822}{15} = 54.8$$

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 $\frac{15+1}{2}$ 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.

Q3. 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

What is known/given?

Ungrouped data and median of data is 63.

What is unknown?

Value of x.

Reasoning:

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 can be calculated based on no. of observation are even or odd.

Solution:

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 $\frac{10}{2}$ i.e., 5^{th} and $\frac{10}{2}$ + 1 i.e., 6^{th} observation.

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

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

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

$$\Rightarrow 63 = x + 1$$

$$\Rightarrow x = 62$$

Q4. Find the mode of 14, 25, 14, 28, 18, 17, 18, 14, 23, 22, 14, 18.

Difficulty level:
Easy

What is known/given?
Ungrouped data.

What is unknown?
Mode of the given data.

Reasoning:
The mode is that value of the observation which occurs most frequently.

Solution:
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.

Q5. Find the mean salary of 60 workers of a factory from the following table:

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

Difficulty level:
Medium

What is known/given?
Salary of 60 workers in tabular form

What is unknown?

Mean of the given data

Solution:

We know that

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

The value of $\sum f_i x_i$ and $\sum f_i$ f can be calculated as follows.

Salary (in ₹) x_i	**Numbers of workers** f_i	$f_i x_i$
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	$10000 \times 1 = 10000$
	$\sum f_1 = 60$	$\sum f_i x_i = 305000$

$$\text{Mean salary} = \frac{305000}{60}$$

Therefore, mean salary of 60 workers is ₹ 5083.33.

Q6. 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.

Reasoning:

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.

Solution:

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.

cuemath
THE MATH EXPERT

When you learn math in an interesting way, you never forget.

25 Million
Math classes & counting

100K+
Students learning Math the right way

20+ Countries
Present across USA, UK, Singapore, India, UAE & more.

Why choose Cuemath?

"Cuemath is a valuable addition to our family. We love solving puzzle cards. My daughter is now visualizing maths and solving problems effectively!"

- Gary Schwartz

"Cuemath is great because my son has a one-on-one interaction with the teacher. The instructor has developed his confidence and I can see progress in his work. One-on-one interaction is perfect and a great bonus."

- Kirk Riley

"I appreciate the effort that miss Nitya puts in to help my daughter understand the best methods and to explain why she got a problem incorrect. She is extremely patient and generous with Miranda."

- Barbara Cabrera

Get the Cuemath advantage

Book a FREE trial class