UNIT SEVEN

Statistics

- Tabulation of Data
- Diagrammatic Representation of Data

Let's Recap

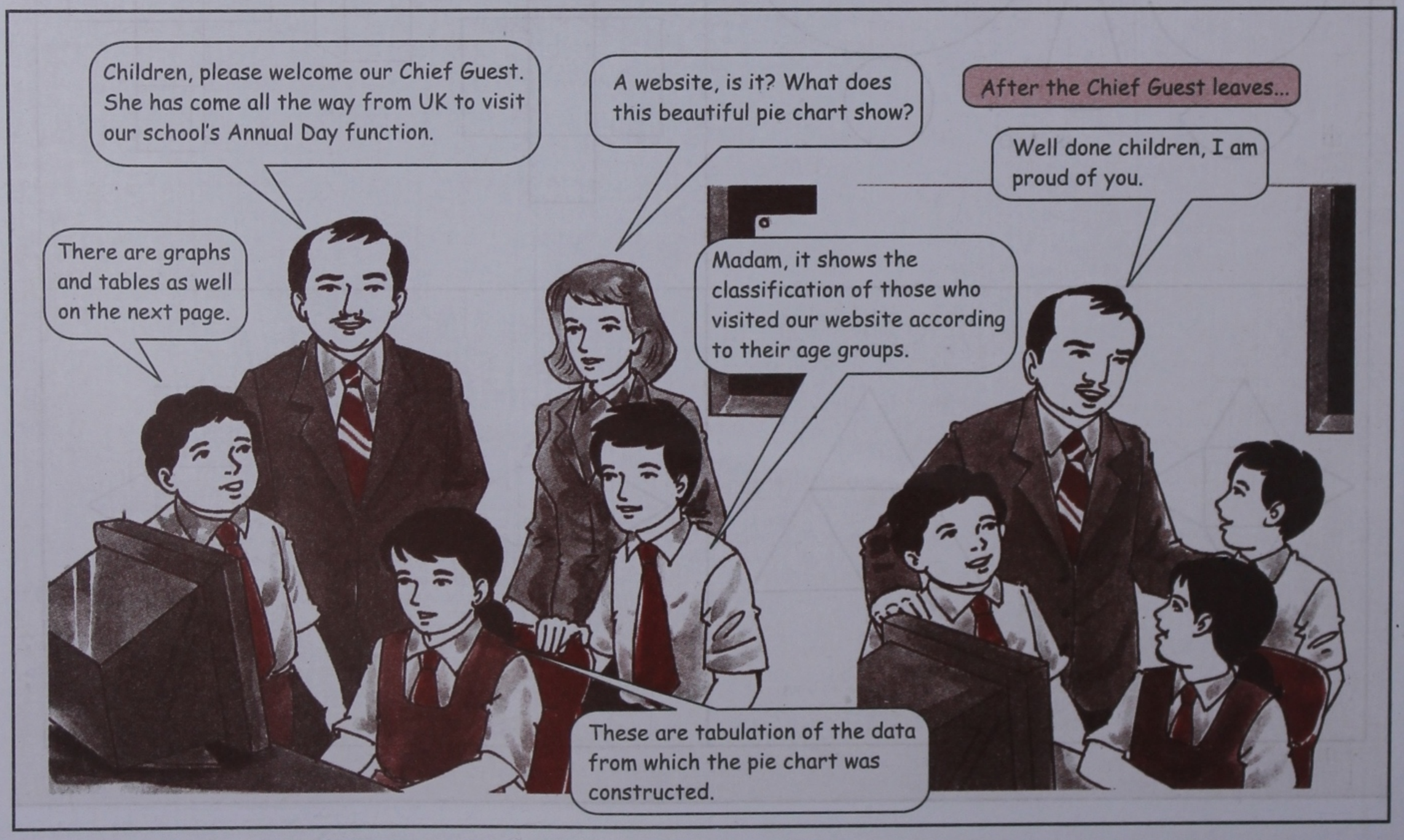
1. Construct a column graph to display the sales figures recorded by a cycle manufacturer in the year 2009.

Month	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
Units sold	1050	850	3620	3850	2170	1190	780	350	580	2900	3410	1730

2. An advertising agency incurs the following expenditure in the month of January 2010. Construct a pie chart showing the comparison of expense heads.

Client servicing	Creative	Production	Administration	Overheads	Other Office expenditure
24000	32000	40000	8000	8000	16000

Construct a pie chart showing the comparison of expense heads.



34 TABULATION OF DATA

- Raw Data
- Mean, Median, and Mode
- Simple Frequency Distribution
- Grouped Data
- Class Limits and Class Interval
- Grouped Frequency Distribution
- Range of Data

Introduction

Statistics involves the collection, classification, presentation, and analysis of numerical facts or data.

Data needs to be collected from 'sources', by conducting a 'survey' or a search. If data on the income of households in a particular colony is required, a door-to-door survey will have to be conducted. Geographical data may be obtained by searching through an atlas. Data on a larger scale, like the population of cities, can be collected from Government sources. Even for simple data like the shoe size of all your classmates you will have to do a small survey in your class. Consider the data given here.

Ankit 6, 5, Anu, Saibal, Vivek, 4, 5, 5, 6, Atula, Rohini

The above collection does not make sense unless it is presented properly.

Name of the Student	Shoe Size
Ankit	6
Anu	5
Saibal	5
Vivek	6
Atula	4
Rohini	5

The tabulation of data makes it easier to understand the information. In algebra the preparation of 'truth tables' helped us make graphs of equations. As the volume of data increases, new methods of tabulating the data will need to be used.

Raw Data

Consider the following data collected on the number of siblings (brothers and sisters) each student in Class VIII of a particular school has.

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

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

Each digit above represents an **observation**, which is the number of siblings a particular student has. This raw data does not help us get much information on the subject.

We know that the **arithmetic mean**, or average

$$\text{of n terms} = \frac{\text{Sum of n terms}}{\text{n}}$$

The given raw data on siblings per student has 32 observations, the sum of which is 38.

$$\text{Thus Arithmetic Mean of given data} = \frac{38}{32}$$
$$= 1.1875$$

Now, no student can have 1.1875 brother or sister. The mean merely indicates that most students have 'about' 2 siblings each.

Example 1: If the weights of 5 students are given as 38 kg, 41 kg, 36 kg, 39, kg, and 41 kg, find the mean of the given data.

The Arithmetic Mean of the given data

$$= \frac{38+41+36+39+41}{5} = \frac{195}{5} = 39 \text{ kg}$$

Thus, the mean weight of the 5 students is 39 kg.

Array

If the data on the number of siblings is arranged in ascending or descending order, it forms an **array**.

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 3 3 3 3

From this array we can deduce that the number of siblings each classmate has varies from 0—3.

In geometry we have seen how a median divides a triangle in two equal halves.

In statistics the middle term of all observations is known as the **median**.

Once the raw data has been written in an array, the $\left(\frac{n+1}{2}\right)^{th}$ term is its median.

Example 2: Find the median in the following array:

7, 9, 11, 13, 15, 17, 19

There are 7 observations in the array.

The median is the $\left(\frac{7+1}{2}\right)^{th}$ or the 4^{th} observation.

Thus, median of the given array = 13

The given array on siblings per student has 32 observations.

The median is given as:

$$\frac{32+1}{2} = 16.5.$$

The median of this array will be the average of the 16^{th} and the 17^{th} observations.

Thus, the median of the given array $= \frac{1+1}{2} = 1.$

Frequency

Now, many students have 0 sibling while some have 3 siblings. *The number of times an observation occurs in a set of data is known as its frequency of occurrence* or simply *frequency.*

The value of the greatest frequency is known as the **mode** of the given data.

Example 3: The ages of 10 students are given as 13.5, 14.5, 14, 14, 14, 14.5, 14, 13.5, 14, and 13.5 years. Find the mode.

We observe that 3 students are 13.5 years old
5 students are 14 years old
2 students are 14.5 years old

As the maximum number of students are 14 years old, mode of the given data = 14 years.

In the given array on siblings per student, 1 is the observation that occurs the maximum number of times.

Thus, the mode of the given array = 1.

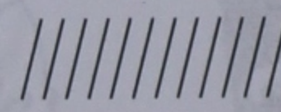
Try this!

Find the mean, median, and mode of the following array:

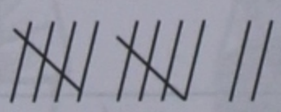
2, 4, 6, 8, 10, 12, 14, 16

Simple Frequency Distribution

The tabular presentation of the frequency of all the observations is known as a frequency distribution. To construct a frequency distribution the observations are first listed in a column. Each occurrence of an observation is marked with a tally mark like '/'. After 4 tally marks against an observation, the fifth time an observation is repeated, it is marked with a long reverse tally mark that crosses out the first four marks. This makes it easier to read off the frequency of each observation.



is not easy to read off as 12.



is easy to read off as 12 as each crossed out bunch of marks represents a frequency of 5.

A **tally chart** can thus be used to construct a simple frequency distribution.

TALLY CHART		
Observations	Tally Marks	Total Tally Marks
0	𝍸 ////	9
1	𝍸 𝍸 //	12
2	𝍸 //	7
3	////	4

SIMPLE FREQUENCY DISTRIBUTION	
Number of Siblings per Student	Number of Students
0	9
1	12
2	7
3	4

The above frequency distribution allows us to deduce the following:

1. Most of the students in the class have just 1 sibling.
2. Least number of students in the class have 3 siblings.
3. $12 + 7 + 4 = 23$ students in the class have at least one sibling.
4. 9 students in the class do not have a brother or a sister.

Example 4: A survey was conducted covering 40 households collecting data on the number of family members each had. Construct a simple frequency distribution with the given data.

3	6	4	4	3	5	4	6	5	4
5	2	5	3	4	4	5	3	4	6
4	3	6	5	4	3	4	2	5	2
6	2	4	4	3	3	5	4	4	6

Arranging the given data in an array, we have

2	2	2	2	3	3	3	3	3	3
3	3	4	4	4	4	4	4	4	4
4	4	4	4	4	4	5	5	5	5
5	5	5	5	6	6	6	6	6	6

TALLY CHART		
Observation	Tally Marks	Total Tally Marks
2	////	4
3	𝍸 ///	8
4	𝍸 𝍸 ////	14
5	𝍸 ///	8
6	𝍸 /	6

SIMPLE FREQUENCY DISTRIBUTION	
Number of Family Members	Number of Households
2	4
3	8
4	14
5	8
6	6

Try this!

Construct a frequency table for the following data :

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

Grouped Data

While constructing the above simple frequency distribution in Example 4, the data was such that each observation could be individually listed in the table.

Now consider the following data on the percentage of marks obtained by 32 students in an examination.

78%	84%	53%	62%	71%	86%	43%	66%
36%	77%	48%	59%	76%	81%	92%	58%
68%	74%	79%	85%	65%	49%	81%	75%
57%	78%	84%	65%	73%	42%	87%	74%

Arranging the above data in an array, we have

36%	42%	43%	48%	49%	53%	57%	58%
59%	62%	65%	65%	66%	68%	71%	73%
74%	74%	75%	76%	77%	78%	78%	79%
81%	81%	84%	84%	85%	86%	87%	92%

We find that there are only 5 observations (65%, 74%, 78%, 81%, and 84%) with a frequency of 2 and a total of 27 (32 – 5) observations. Listing all the observations in a table would not only be tedious, but will also not help us to make any meaningful deductions.

In such cases the data is grouped into **classes**.

As 36% is the lowest observation and 92% is the highest observation, the entire data can be grouped in only 7 classes, viz. 30–40, 40–50, 50–60, 60–70, 70–80, 80–90, and 90 –100.

Class Limits and Class Interval

In the class 40–50, 40 is the **lower class limit** while 50 is the **upper class limit**. The difference between the two class limits is known as the **class interval** (50 – 40 = 10). The mid-point of a class interval is known as the **class mark**. The class mark of the class interval in 40–50 is $\frac{40 + 50}{2} = 45$, or the class mark is the average of the lower and upper class limits.

Grouped Frequency Distribution

A tally chart can now be used to count the frequency of observations in the different classes, leading to the construction of a grouped frequency distribution.

Percentage of Marks	Tally Marks	Number of Students
30–40	/	1
40–50	////	4
50–60	////	4
60–70	/////	5
70–80	///// /////	10
80–90	///// //	7
90–100	/	1
Total number of students		**32**

The grouped frequency table now allows us to deduce the following:

1. Most students in the class scored between 70%–80%.
2. Only 1 student did extremely well, scoring over 90% and only 1 student did very poorly, scoring less than 40%.
3. 5 + 10 + 7 + 1 = 23 students scored more than 60%, earning a 1st division.
4. 1 out of 32 students who appeared for the examination failed, or the failure percentage was only 3.125%.
5. 31 out of 32 students passed the examination or the passing percentage was 96.875%.

Inclusive and Exclusive Forms of Class Intervals

In the previous frequency distribution, what if a student had scored 80% marks? Would the observation belong to the (70–80)% class or the (80–90)% class?

The class intervals like 10–20, 20–30 are the **exclusive** forms of class intervals. In this form, the upper limit is excluded. Thus an observation of 20 would belong to the 20–30 class and not the 10–20 class.

Class intervals like 11–20, 21–30 are the **inclusive** forms of class intervals. In this form, both class limits are included, and clearly an observation of 20 would belong to the 11–20 class.

An inclusive form of class intervals can be converted into an exclusive form by following the steps given below.

Step 1: Find the difference between the lower limit of a class and the upper limit of the previous class.

Step 2: Subtract half to this difference from the lower class limit to obtain a new lower class limit.

Step 3: Add half the difference (from *Step* 1) to the upper class limit to obtain a new upper class limit.

Example 5: Convert discontinuous class intervals 11–20, 21–30, 31–40, and 41–50 to an exclusive form.

Let us consider the class 21–30.

Difference between the lower class limit and the upper class limit of the previous class = 21 – 20 = 1

Half of 1 = 0.5.

Thus new lower class limit = 21 – 0.5 = 20.5 and new upper class limit = 30 + 0.5 = 30.5.

Similarly, the other classes will be 11–20.5, 20.5–30.5, 30.5–40.5, and 40.5–50.

The lower and upper class limits of this exclusive form of class intervals are called the true class limits.

Try this!

Convert the following discontinuous class intervals to an exclusive form 31-40, 41-50, 51-60, and 61-70.

Range of Data

The range of data is the difference between the highest observation and the lowest observation. It gives an indication as to how the data should be grouped into classes and what the class interval should be.

As the data for the grouped frequency distribution we constructed ranged from 36% to 92%,

1. there was no need to list classes 0–10, 10–20, and 20–30;
2. a class interval of 10 was selected such that the information could be presented in only 7 rows;
3. an exclusive form of class intervals was selected for convenience and its visual appeal.

Example 6: The maximum relative humidity recorded in Kolkata in the month of June 2004 is given as follows (all figures are in percentage.)

82 81 86 80 76 83 88 92 96 94 98 97 99 100 98
96 100 94 92 94 93 90 92 93 96 94 89 90 88 88

Distribute the data in appropriate classes and construct a grouped frequency distribution.

Arranging the given data in an array, we have

76 80 81 82 83 86 88 88 88 89 90 90 92 92 92
93 93 94 94 94 94 96 96 96 97 98 98 99 100 100

The observations range from 76 to 100. Thus, the data from 75 to 100 can be distributed in 5 classes with a class interval of 5 each, as 75–80, 80–85, 85–90, 90–95, and 95–100.

As the upper limit is excluded in each class, an observation like 80 would belong to the 80–85 class, but which class would the observation 100 belong to? Thus, we choose the inclusive form of class intervals as 76–80, 81–85, 86–90, 91–95, and 96–100.

Maximum Relative Humidity (Percentage)	Tally Marks	Number of Days
76–80	//	2
81–85	///	3
86–90	𝍸 //	7
91–95	𝍸 ////	9
96–100	𝍸 ////	9
Total number of days		**30**

Try this!

The runs scored by OUP XI are as follows : 81, 47, 52, 63, 44, 72, 20, 27, 73, 99. Find the range.

Exercise 34.1

1. The given data represents the number of graduates in 40 apartments in a housing complex. Construct an array by arranging the given data in ascending order, and find the arithmetic mean and median.

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

2. 40 respondents chose the following options in a multiple choice questionnaire in a survey. Construct an array by arranging the given data in an alphabetical order, and find the mode.

a	d	d	a	b	d	a	a	d	c
d	a	b	a	d	d	d	a	d	a
a	c	a	c	d	b	a	d	b	a
b	c	a	d	d	b	a	a	b	d

3. The given data represents the outcome when a dice is cast 40 times. Construct a simple frequency distribution and deduce if the dice is fair or not.

1	6	2	5	6	6	3	6	6	6
5	4	4	6	2	4	5	1	5	2
6	5	3	2	6	6	3	6	6	6
1	5	6	6	5	1	6	2	1	4

4. The ages of 40 students, rounded off to the nearest quarter, in class VIII of a particular school, are given below. Construct a simple frequency distribution and deduce how many students are at least $14\frac{1}{2}$ years old.

$14\frac{3}{4}$ $14\frac{1}{2}$ $14\frac{3}{4}$ $14\frac{1}{4}$ $14\frac{1}{2}$ $14\frac{3}{4}$ $14\frac{1}{4}$ $14\frac{1}{2}$ $15\frac{1}{4}$ $14\frac{1}{2}$

$15\frac{1}{4}$ $14\frac{1}{4}$ $14\frac{1}{2}$ $14\frac{1}{4}$ 15 $14\frac{1}{4}$ 14 $14\frac{1}{4}$ $14\frac{3}{4}$ $14\frac{1}{4}$

$14\frac{1}{4}$ 14 $14\frac{1}{4}$ $14\frac{3}{4}$ $14\frac{1}{2}$ $15\frac{1}{2}$ $14\frac{1}{2}$ $14\frac{3}{4}$ $13\frac{3}{4}$ $14\frac{1}{2}$

$14\frac{1}{4}$ $14\frac{3}{4}$ $14\frac{1}{2}$ $14\frac{3}{4}$ $14\frac{1}{2}$ $14\frac{1}{2}$ 15 $14\frac{1}{2}$ $14\frac{1}{2}$ $14\frac{3}{4}$

5. Write the lower class limit, upper class limit, and class marks of the following class intervals:
   (i) 60–70 (ii) 0–5 (iii) 45–50
   (iv) 31–40 (v) 40–49

6. The heights of 40 students of a class, rounded off to the nearest centimetre, are given below. Construct a grouped frequency distribution with the given data with 6 classes, the first class interval being 145–150, and deduce how many children are 160 cm or more in height.

169 161 150 160 151 156 147 161 163 159

154 159 166 155 170 161 153 162 158 152

171 164 171 155 163 168 166 154 160 157

167 158 152 164 168 156 163 162 167 153

(Remember, an observation of 160 cm will belong to the 160–165 class interval.)

7. Select appropriate class intervals, given the following data on the maximum temperatures recorded in a city over 30 days and construct a grouped frequency distribution. On how many days was the maximum temperature above 34°C?

35	38	34	30	37	33	36	36	40	33
26	32	29	36	31	33	37	29	35	28
36	28	35	38	36	32	37	34	31	37

Revision Exercise

1. The given data represents the heights (in cm) of 11 girls of a class. Construct an array by arranging the given data in ascending order, and find the arithmetic mean and median.
   145, 144, 146, 134, 126, 141, 136, 156, 152, 144, 149
2. The number of members in 20 families is given below. Construct a simple frequency distribution of data.
   3, 7, 4, 6, 3, 7, 3, 4, 4, 5, 4, 4, 5, 3, 5, 5, 8, 3, 5, 4, 8
3. The given data represents the outcome when a die is cast 30 times. Construct a simple frequency distribution and deduce if the die is fair or not.

2	4	1	2	6	3	2	1	4	5
3	1	1	4	4	1	1	3	5	6
4	3	2	2	6	4	6	5	5	2

4. The weights (in kg) of 30 persons are given below. Construct a grouped frequency distribution with the given data with 5 classes, the first class interval being 40-45, and deduce how many children are 55kg, or more in weight.

43	51	47	62	50	38	48	64	55	58
42	46	54	55	52	40	54	53	50	44
47	52	50	52	47	44	55	60	63	60

5. Select appropriate class-intervals, given the following data on the weekly wages (in rupees) of 30 workers of a factory and construct a grouped frequency distribution.

680	700	670	612	640	660	666	622	721	718
720	642	672	654	692	706	718	702	704	678
613	642	682	714	707	613	636	656	654	670