DATA HANDLING

1. The collection, organization, presentation, interpretation and analysis of Data is called Statistics.

2. The data collected can be presented in the form of tables, diagrams and graphs.

1.1 PICTOGRAMS:

1. **In a pictogram, a symbol / icon is used to represent a quantity of items.**

2. **A key is included to explain what each symbol / icon represents.**

EXAMPLE:

A group of students was asked to choose their favourite pet . The information is illustrated on the pictogram below .

HAMSTER	
RABBIT	
FISH	
TORTOISE	

Each  represents 25 students

a) Find the number of students who chose the rabbit .

b) Which is the most popular pet ?

c) Find the total number of students in the group .

d) What fraction of the students chose the fish as their favourite pet?

e) Calculate the percentage of students who chose the tortoise as their favourite pet.

SOLUTION:

a) Number of students who chose the rabbit

$= 5 \text{ X } 25$

$= 125$

b) The most popular pet is the hamster .

c) Total number of students

$= 22 \text{ X } 25$

$= 550$

d) Number of students who chose the fish

$= 6 \text{ X } 25$

$= 150$

Required percentage

$= \frac{150}{550}$

$= 0.27$

e) Number of students who chose the tortoise

$= 3 \text{ X } 25$

$= 75$

Required percentage

$= \frac{75}{550} \times 100\ \%$

$= 13.63$

TIPS FOR STUDENTS:

A pictogram is not a very accurate method of representing exact data . It only gives a quick comparison of the relative number of each type of pet .

1.2 BAR GRAPH

1. **A bar graph represents data by using rectangular bars of equal width.**

2. **The bars can be drawn vertically or horizontally.**

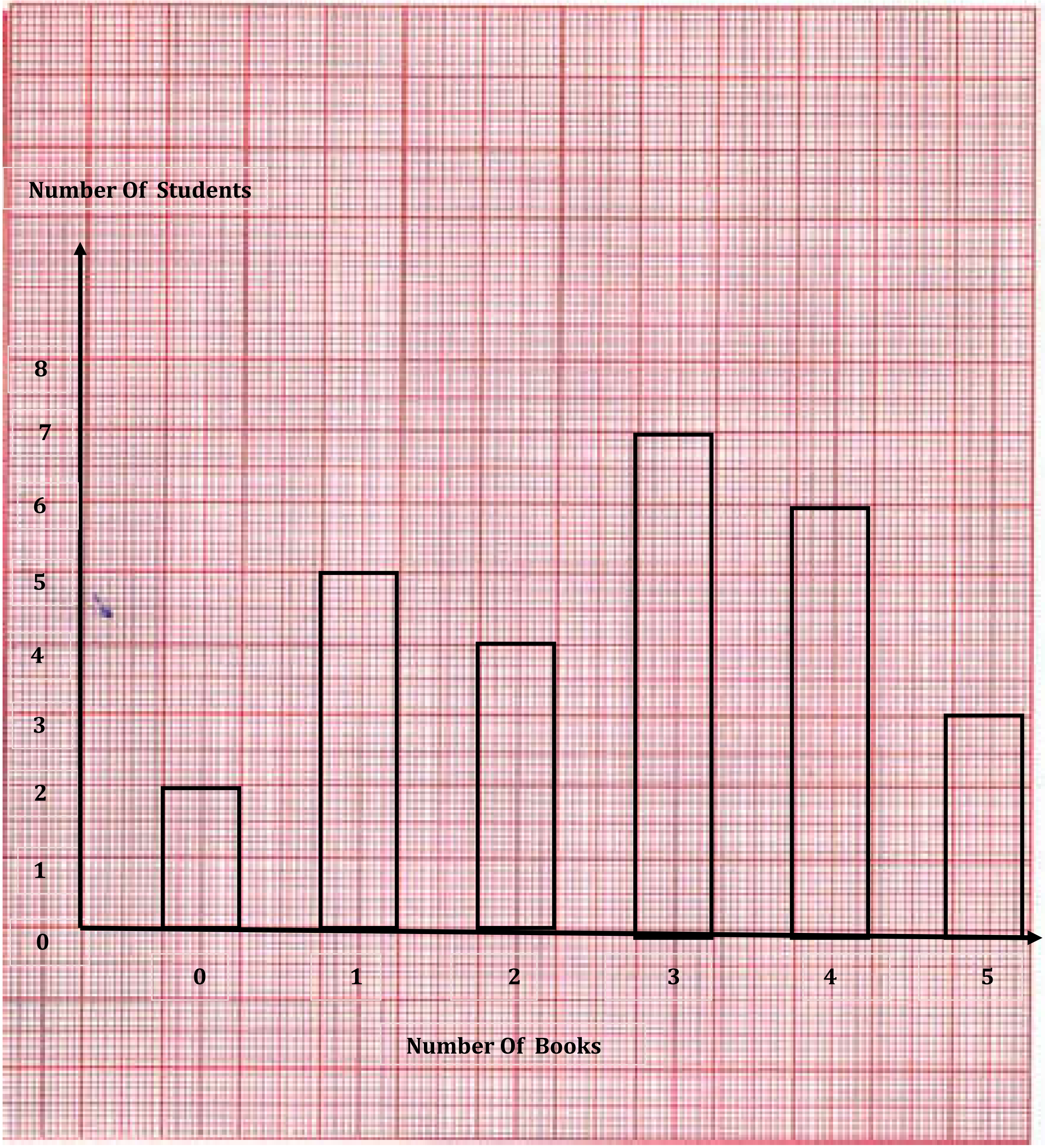
3. **The height / length of each bar is proportional to the data represented.**

4. **The spaces between the bars are uniform.**

EXAMPLE:

The graph illustrates the results of a survey to find the number of books borrowed from the school library by a group of students.

GRAPH DIAGRAM 1


Number Of Students
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
Number Of Books

FIND

a) The number of students in the group.

b) The total number of books borrowed.

c) The percentage of students who borrowed more than 2 books.

SOLUTION:

a) Number of Students = 2 + 5 + 4 + 7 + 6 + 3

= 27

b) Total Number of books borrowed

= (0 X 2) + (1 X 5) + (2 X 4) + (3 X 7) +(4 X 6) + (5 X 3)

= 0 + 5 + 8 + 21 + 24 + 15

= 73

c) Number of Students who borrowed more than 2 books

= 7 + 6 + 3

= 16

Required percentage

= $\frac{16}{27}$ X 100 %

= 59.3 % (correct to 3 sig. fig.)

1.3 PIE CHART

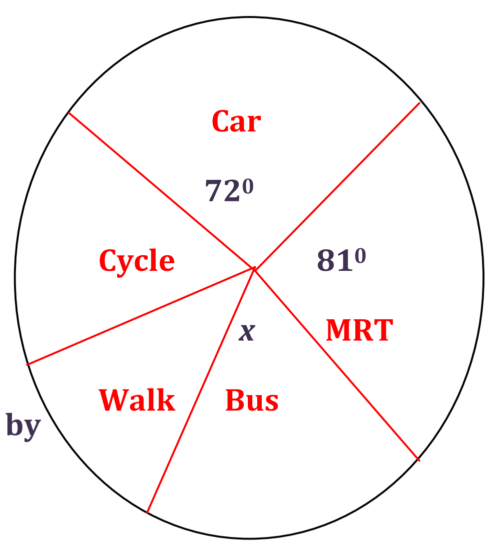
1. A pie chart displays the given data using sectors of a circle.

2. The angle in each sector is proportional to the number of items represented.

EXAMPLE:

120 students were asked how they travelled to school .

The pie chart shows the result of this survey .



a) CALCULATE

i) The value of *x* if 36 students travelled to school by bus,

ii) The number of students who travelled to school by MRT ,

iii) The percentage of students cycle than walk to school, find the number of students who walk to school .

SOLUTION:

a) i) $x = \frac{36}{120} \text{ X } 360^0$

$= 108^0$

ii) Number of students who travelled to school by MRT

$= \frac{81^0}{360^0} \text{ X } 120$

$= 27$

iii) Percentage of students who travelled to school by car

$= \frac{72^0}{360^0} \text{ X } 100\ \%$

$= 20\%$

b) Angle of the sector representing students who walk or cycle to school

$= 360^0 - 72^0 - 81^0 - 108^0$ ($\angle$s at a point)

Number of students who walk or cycle to school

$= \frac{99}{360^0} \text{ X } 120$

$= 33$

Number of students who walk to school

$= (33 - 3) \div 2$

$= 15$

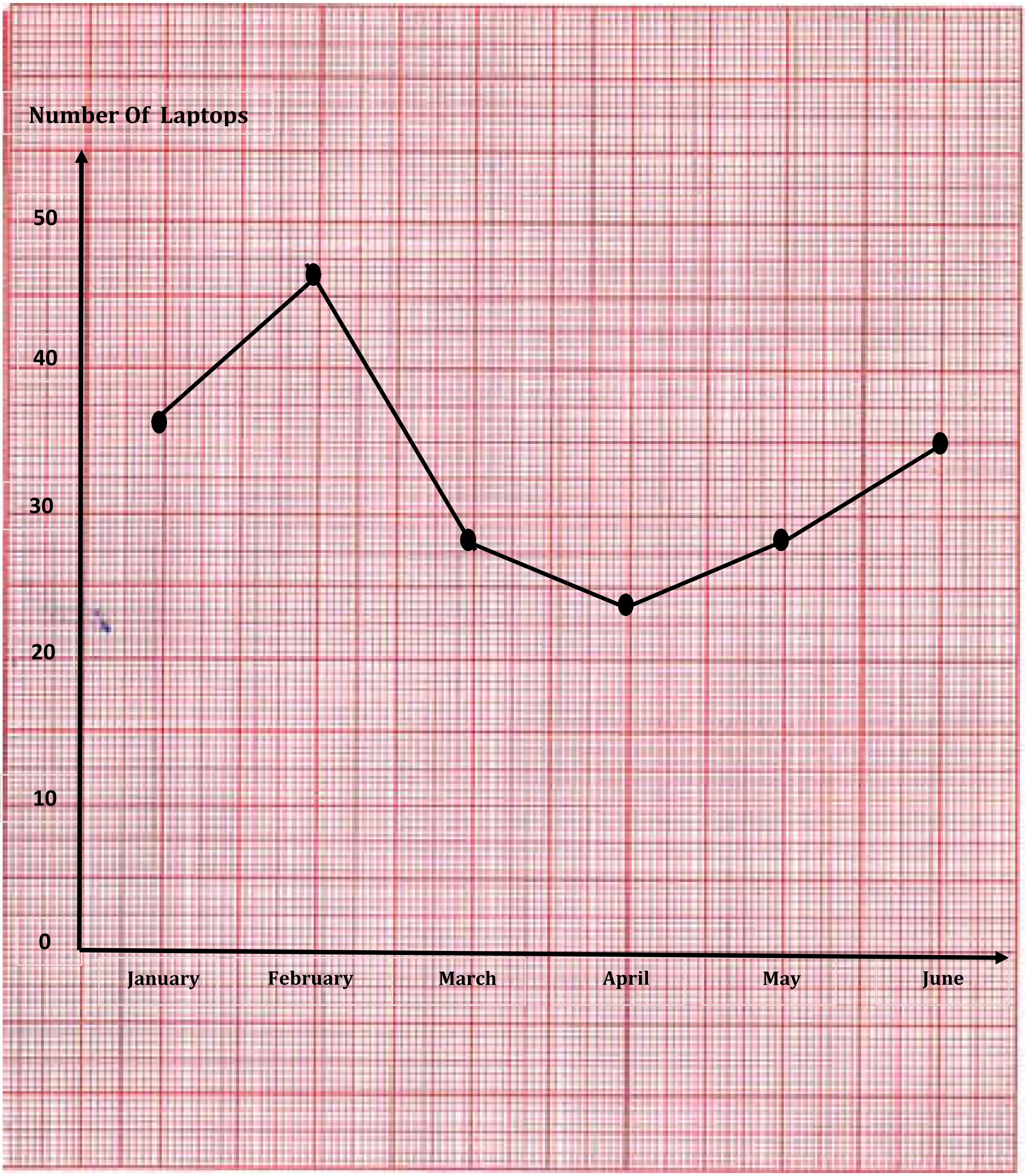
1.4 LINE GRAPHS

A line graph is drawn by plotting the points corresponding to the data and then joining the points with line segments .

EXAMPLE:

The line graph shows the number of laptops sold by a company over a period of 6 months.

GRAPH DIAGRAM 2


Number Of Laptops
50
40
30
20
10
0
January
February
March
April
May
June

a) In which month was the sale of laptops the greatest ?

b) Between which months was there the greatest decrease in the sale of laptops?

SOLUTION:

a) The scale of laptops was the greatest in February.

b) The greatest decrease in the sale of laptops occurred between February and March.

1.5 FREQUENCY TABLE AND HISTOGRAMS

1. A frequency table shows how often a value occurs.

2. The number of times a value occurs is called its frequency.

EXAMPLE :

The number of siblings each one of a group of 30 students has is shown below.

1	2	0	3	0	2	0	2	1	0

3	**4**	**1**	**2**	**3**	**1**	**3**	**4**	**2**	**3**
0	**1**	**0**	**0**	**1**	**0**	**2**	**1**	**0**	**1**

Construct a frequency table to represent the distribution above .

SOLUTION:

NUMBER OF SIBLINGS	TALLY	FREQUENCY
0	~~IIII~~ IIII	9
1	~~IIII~~ III	8
2	~~IIII~~ I	6
3	~~IIII~~	5
4	II	2
		30

HISTOGRAMS:

1. A histogram is a vertical bar graph with no gaps in between the bars .

2. The area of each bar is proportional to the frequency it represents .

3. We can use a histogram to display the information given in a frequency table .

EXAMPLE:

A group of 40 students was asked to write down the number of hours they spent surfing the internet on a certain day . The data collected are as shown below .

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

a) Construct a frequency table for the number of hour the students spent surfing the internet

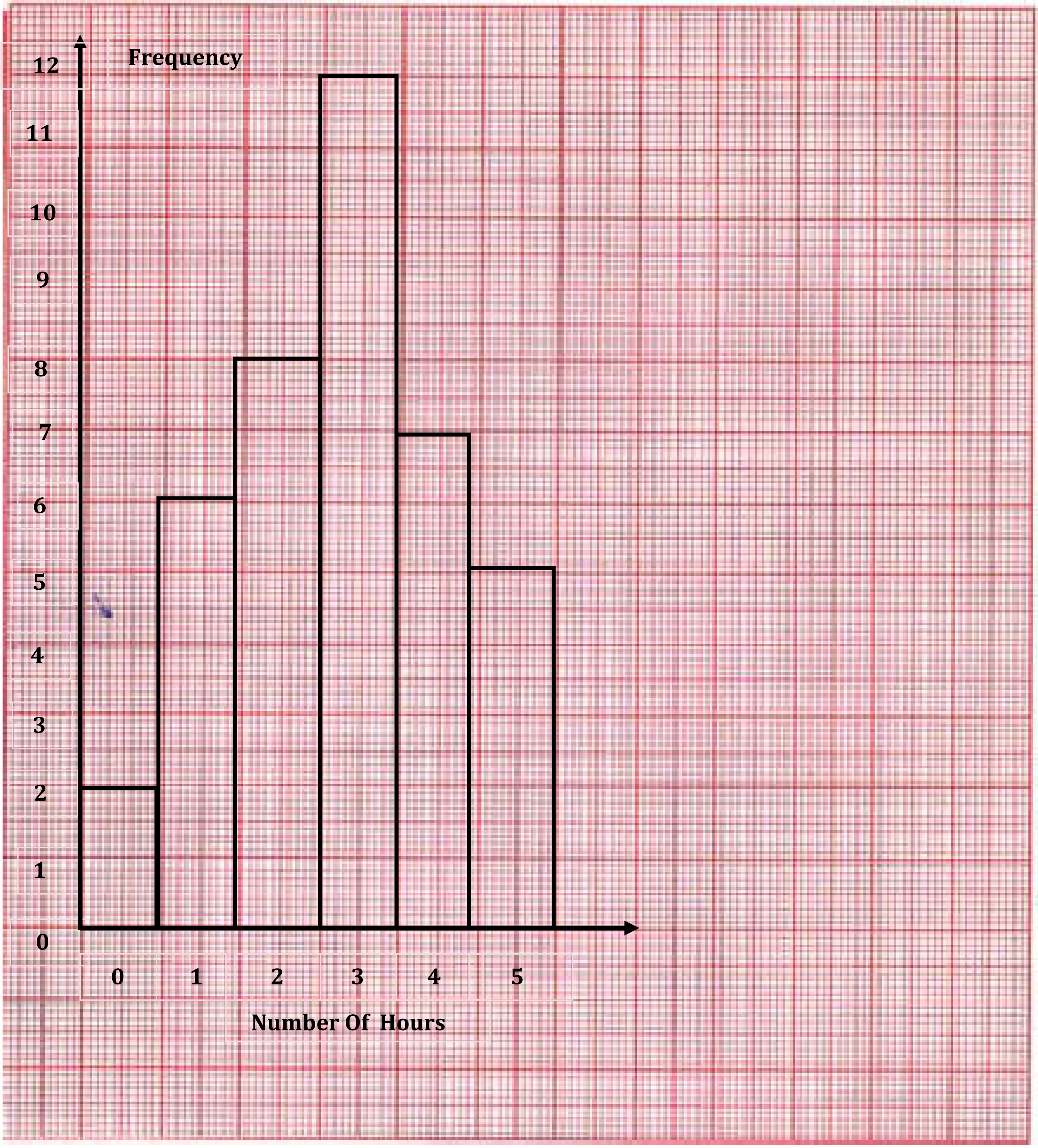
b) Draw a histogram to illustrate this distribution .

SOLUTION:

a)

NUMBER OF HOURS	TALLY	FREQUENCY
0	II	2
1	~~IIII~~ I	6
2	~~IIII~~ III	8
3	~~IIII~~ ~~IIII~~ II	12
4	~~IIII~~ II	7
5	~~IIII~~	5
		40

FIND b) GRAPH DIAGRAM 3



Frequency
12
11
10
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
Number Of Hours

GROUP FREQUENCY TABLE

In a group frequency table, the data are grouped into class intervals of equal sizes .

EXAMPLE :

The masses, in kilograms, of 40 students are shown below .

45	64	40	48	51	42	54	57	60	49
58	68	49	55	65	63	45	50	41	64
50	67	52	41	55	43	48	61	54	40
58	44	51	46	45	50	47	52	46	58

a)

Construct a grouped frequency table to represent the information .

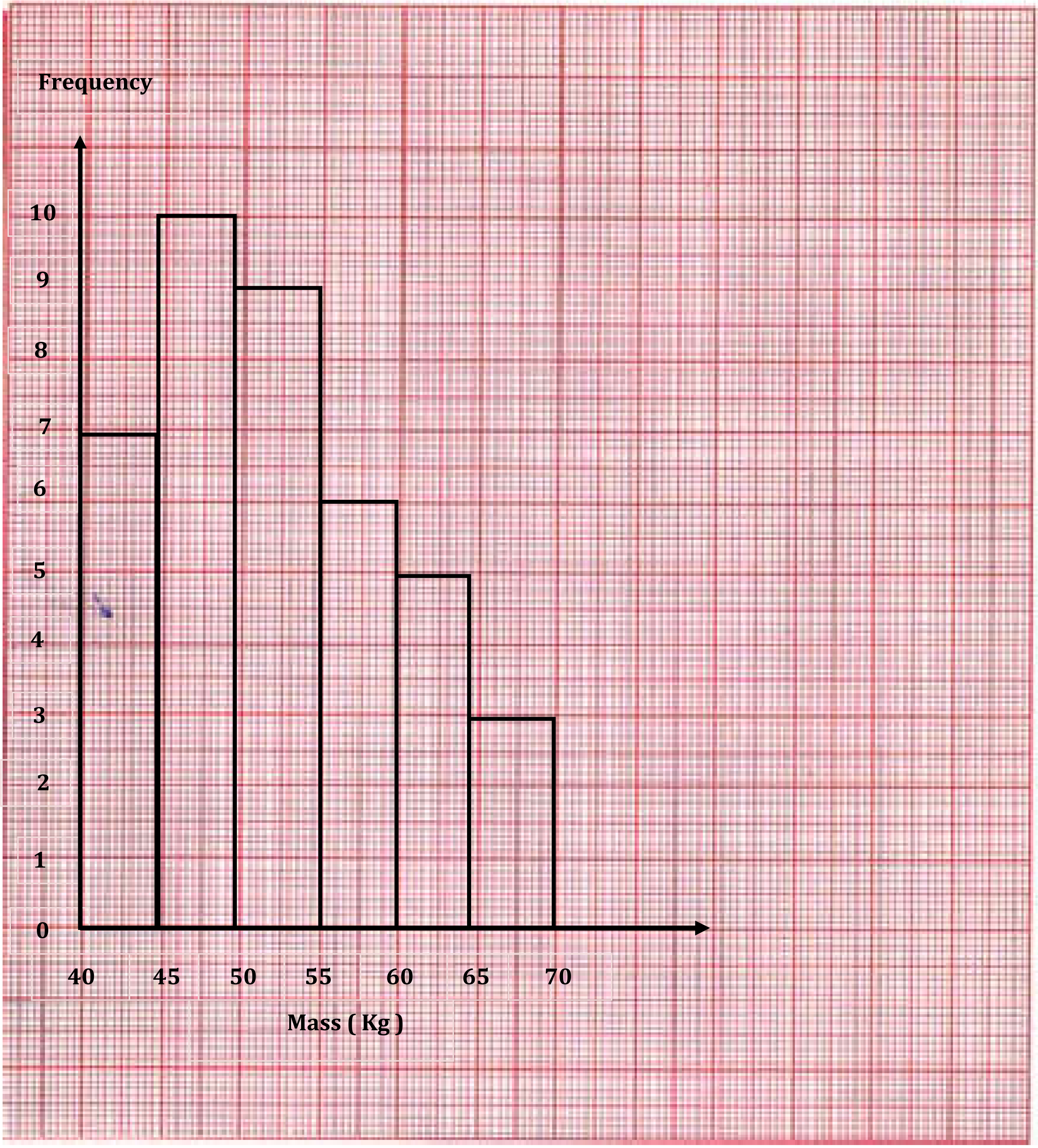
b) Draw a histogram to represent the information .

SOLUTION :

a)

MASS (*x* kg)	**TALLY**	**FREQUENCY**
$40 \leq x < 45$	~~llll~~ ll	7
$45 \leq x < 50$	~~llll~~ ~~llll~~	10
$50 \leq x < 55$	~~llll~~ llll	9
$55 \leq x < 60$	~~llll~~ l	6
$60 \leq x < 65$	~~llll~~	5
$65 \leq x < 70$	Lll	3
		40

b) GRAPH DIAGRAM 4


Frequency
10
9
8
7
6
5
4
3
2
1
0
40
45
50
55
60
65
70
Mass (Kg)

1.6 DOT DIAGRAMS

1. A dot diagram is drawn by placing dots that represent the values of a set of data above a horizontal number line .

2. The number of dots above each value indicates how many times the value occurred .

EXAMPLE1:

The heights, in centimeters, of 30 seedlings are shown below.

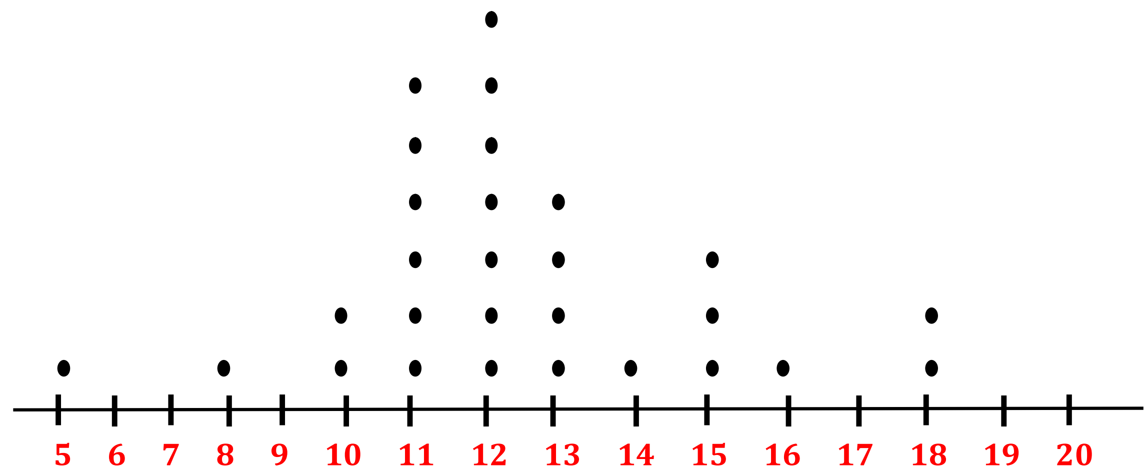
12	16	14	11	18	12	15	18	12	10
15	11	12	20	05	13	11	12	13	15
11	08	13	19	11	10	12	13	11	12

a) Represent the data in a dot diagram.

b) Find the most common height of a seeding.

SOLUTION:

a)

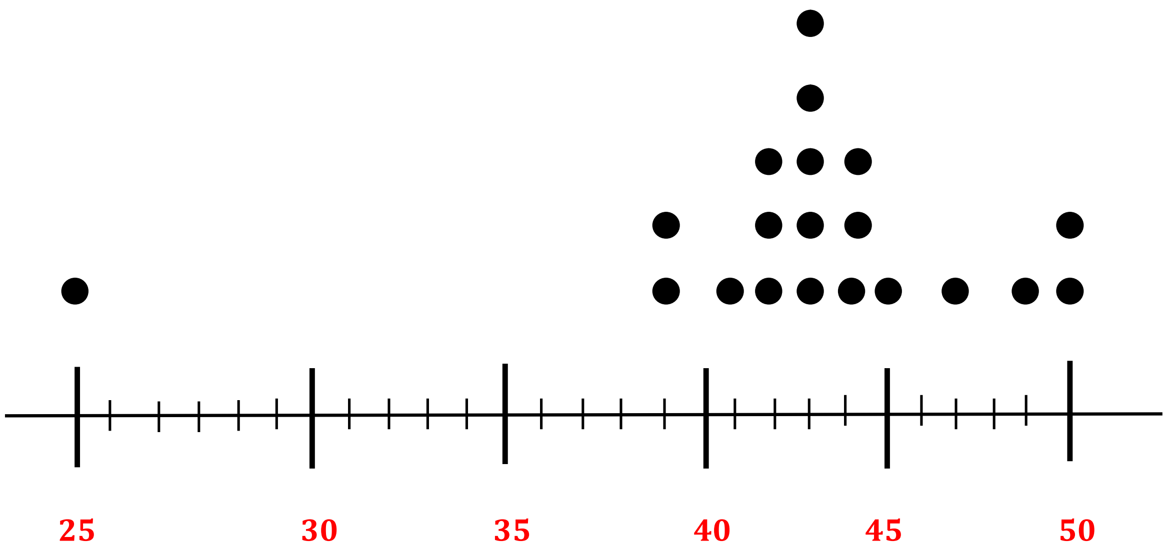


b) **The most common height of a seeding is 12 cm .**

The height that occurs the most often.

EXAMPLE2:

The dot diagram shows the marks obtained by a group of students in a mathematics test .



a) **Find the fraction of students who scored less than 42 marks in the mathematics test.**

b) **If a distinction is awarded to students who scored at least 45 marks, find the percentage of students who scored a distinction .**

c) **Comment, briefly, on the data.**

SOLUTION:

a) **Total number of students = 20** ← **Find the total no . of dots .**

Number of students who scored < 42 marks = 4

$$\text{Required fraction} = \frac{4}{20}$$

$$= \frac{1}{5}$$

b) **Total number of students who scored ≥ 45 marks = 5**

$$= \frac{5}{20} \times 100\ \%$$

$$= 25\ \%$$

c) **The data vary between 25 and 50. The lowest marks obtained was 25 marks and the highest marks obtained was 50 marks . The most common marks is 43 and the data cluster around 43 .**

1.7 STEM AND LEAF DIAGRAMS

In a stem and leaf diagram, each value is split into two parts, the stem and the leaf by a vertical line .

EXAMPLE : 1

The masses, in grams, of 20 apples are given below .

70	81	93	74	72	89	90	81	69	89
85	69	77	85	68	90	72	91	83	72

Represent the data in a stem and leaf diagram .

SOLUTION:

The ten digit of a number .

The ones digit of a number.

STEM	LEAF						
6	8	9	9				
7	0	2	2	2	4	7	
8	1	1	3	5	5	9	9
9	0	0	1	3			

This now shows the data 70, 72, 72, 72, 74 and 77.

KEY : 6 | 8 represents 68 g.

EXAMPLE2 :

The stem and leaf diagram below shows the times, in minutes, taken by a group of workers in a factory to assemble a toy .

STEM	LEAF							
1	0	2	4					
1	5	5	8	8	9			
2	1	1	3	3	4			
2	5	6	6	6	7	7	8	9
3	0	0	0	1	1	3		
3	7							

KEY : 1 | 0 represents 10 minutes .

a) Find the time taken by the fastest and the slowest worker to assemble a toy .

b) Find the percentage of workers who took longer than 25 minutes to assemble a toy .

c) Write down the most common time taken to assemble a toy .

d) 25 % of the workers took less than x minutes to assemble a toy . Find the value of x .

SOLUTION :

a) Time taken by the fastest worker = 10 Min

Time taken by the slowest worker = 37 Min

b) Total number of workers = 28

Number of workers who took 25 > min = 14

Required percentage = $\frac{14}{28}$ X 100 %

c) Most common time taken = 26 Min and 30 Min

The time that occurs the most often .

d) 25 % of 28

$= \frac{25}{100} \times 28$

$= 7$

From the stem and leaf diagram, 7 workers took less than 19 minutes to assemble a toy .

$\therefore$ x = 19 min

EXAMPLE : 3

The stem and leaf diagram shows the ages of the employees of two companies, ***A*** and ***B*** .

COMPANY *A*	STEM	COMPANY *B*
8 7 4 3 3 2 0 0	2	8 8 9
9 9 8 8 8 1 0 0 0	3	2 3 3 5 5 5 5 8 9
2 1 1	4	0 0 0 1 1 3 5 7
0	5	

a) Which company has the oldest employee ?

b) Find the age of the youngest employee .

c) Find the most common age of the employees .

d) Compare the distribution of the ages of the employees of these two companies .

SOLUTION:

a) Company *B* has the oldest employee. (57 years old)

b) The youngest employee is 20 years old.

c) The most common age is 38 years .

There are 3 employees in a company *A* and 2 employees in Company *B* are 38 years old.

d) The ages of the employees in Company *A* cluster around 20 to 30 years old . The ages of the employees in Company *B* cluster around 40 to 50 years old . Thus, the average age of the employees in Company *B* is more than those in Company *A* .

1.8 MODE, MEDIAN AND MEAN

The mode, median and mean are measures of central tendency. A measure of central tendency is a single value that describes where the data are centred, i.e. its average value.

MODE:

1. The mode of a set of data is the value that occurs most frequency.

2. In some distributions, no value appears more than once. So there is no mode. In other distributions, there may be more than one mode .

EXAMPLE:

Find the mode (s) of the following sets of data.

a) 1, 1, 2, 6, 7, 7, 7, 9, 10

b) 1, 2, 2, 2, 3, 5, 5, 8, 8, 8

c) 1, 3, 4, 7, 10

SOLUTION:

a) 1, 1, 2, 6, 7, 7, 7, 9, 10

Mode = 7 ← 7 Appears the most often, i.e. 3 times .

b) 1, 2, 2, 2, 3, 5, 5, 8, 8, 8

Mode = 2 and 8 ← This set of numbers is BIMODAL as it has two modes, 2 and 8.

c) 1, 3, 4, 7, 10 ← None of the numbers appear more than once .

There is no mode.

MEDIAN

1. The value exactly in the middle of a set of ordered number (ascending or descending) is the median.

2. To find the median of a set of *n* data:

1. **Arrange the number in ascending order, i.e. from the smallest to the greatest .**

2. **If *n* is odd, the median is the middle value.**

   If *n* is even, the median is the mean of the two middle values.

3. To find the middle position of a set of *n* data:

MIDDLE POSITION $= \frac{n+1}{2}$

EXAMPLE1:

Find the median of the following sets of data.

a) 3, 8, 2, 6, 10

b) 1, 6, 9, 5, 4, 8

SOLUTION:

a) **Middle position**

2, 3, 6, 8, 10 ← **Arrange the numbers in ascending order first .**

Median = 6

b) **Middle position**

1, 6, 9, 5, 4, 8

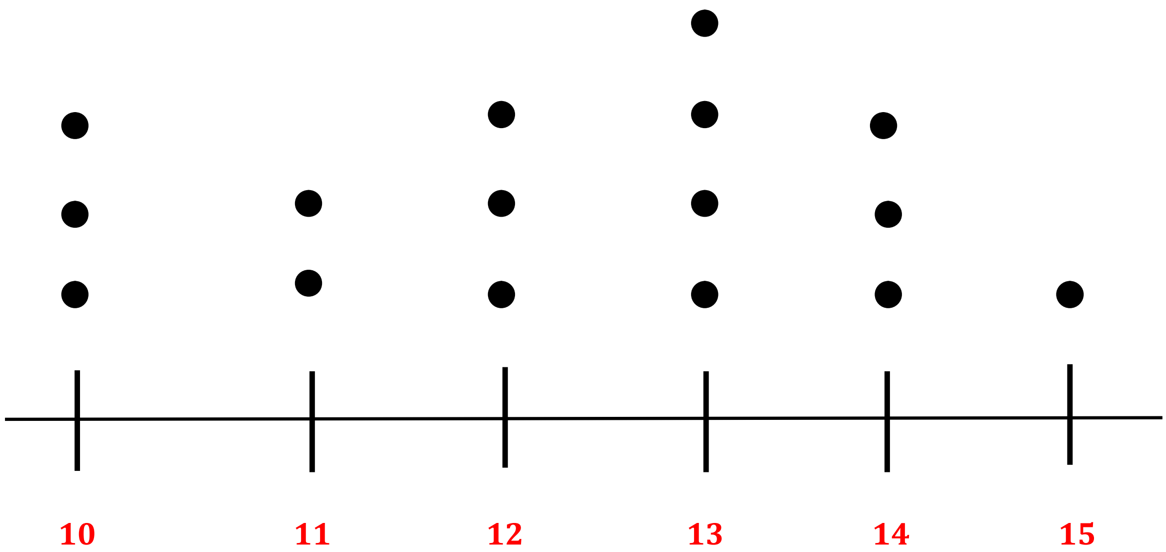
Median $= \frac{5 + 6}{2}$

$= 5.5$

EXAMPLE : 2

Find the mode and median of each of the following .

a)



b)

STEM	LEAF						
4	0	0	1	2	8		
5	1	3	3	3	5	6	7
6	0	0	0	2	8	9	
7	6						

KEY : 4 | 0 represents 40 cm.

c)

TIME (MIN)	20	21	22	23	24	25
NUMBER OF BOYS	3	7	9	5	3	2

SOLUTION:

a) **Mode** = 13 g ← The data with the highest frequency .

Middle Position = $\frac{16 + 1}{2}$

= 58.5^{th} position

$\therefore$ **Median** = Mean of 8^{th} and 9^{th} values

= $\frac{12 + 13}{2}$

= 12.5 g ← The data with the middle position .

b) **Mode** = 53 cm and 60 cm

Middle Position = $\frac{19 + 1}{2}$

= 10^{th} position

$\therefore$ **Median** = 55 cm

c) **Mode** = **22 min**

Total number of boys = **3 + 7 + 9 + 5 + 3 + 2**

= **29**

Middle Position = $\frac{29 + 1}{2}$

= 15^{th} **position**

$\therefore$ **Median** = **22 min**

TIPS FOR STUDENTS

$$\overbrace{20, 20, 20}^{3}, \overbrace{21, \ldots\ldots, 21}^{7}, \overbrace{22, 22, 22, 22, 22, 22, 22, 22, 22}^{9}$$

15^{th} **position**

MEAN

1. The mean of a set of data is obtained by dividing the sum, of all the data by the total number of data .

$$\boxed{\text{MEAN} = \frac{\text{Sum Of Data}}{\text{Number Of Data}}}$$

2. The Mean of a set of data, x_1, x_2, x_3, . . . , x_n, denoted by $\overline{x}$ (red x bar) is given by :

$$\boxed{\overline{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}}$$

3. Given a set of data, x_1, x_2, x_3, . . . , x_n, occurring with corresponding frequencies, f_1, f_2, f_3, . . . , f_n, it mean, $\overline{x}$ is given by :

$$\overline{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \ldots + x_n f_n}{f_1 + f_2 + f_3 + \ldots + f_n}$$

$$= \frac{\sum fx}{\sum f}$$

x	x_1	x_2	x_3		x_n
f	f_1	f_2	f_3		f_n

EXAMPLE : 1

a) Find the mean of the following set of numbers .

8, 3, 1, 2, 6, 7, 5, 5, 4, 9

b) The mean of six number is 18 . Four of the numbers are 10, 12, 23 and 31 . If each of the other two numbers is equal to x , find the value of x .

c) The mean of five numbers is 82 . If another number is added, the mean of the six numbers is 84 . Find the number added .

SOLUTION :

a) Mean $= \dfrac{8+3+1+2+6+7+5+5+4+9}{10}$

$= \dfrac{50}{10}$

$= 5$

b) Sum of the 6 numbers $= 6 \times 18$

$= 108$

$$10 + 12 + 23 + 31 + x + x = 108$$

$$2x + 76 = 108$$

$$2x = 32$$

$$x = 16$$

c) Sum of the 5 numbers $= 5 \times 82 = 410$

Sum of the 6 numbers $= 6 \times 84 = 504$

$\therefore$ Number added $= 504 - 410 = 94$

EXAMPLE : 2

The below shows the performances in 25 matches of a hockey team .

Number Of Goals (x)	0	1	2	3	4	5
Number Of Matches (f)	3	5	6	8	2	1

FIND

a) **The Mean ,**

b) **The Median ,**

c) **The Mode .**

SOLUTION :

a) **Mean** $= \frac{\sum fx}{\sum f}$

$= \frac{(0 \text{ X } 3) + (1 \text{ X } 5) + (2 \text{ X } 6) + (3 \text{ X } 8) + (4 \text{ X } 2) + (5 \text{ X } 1)}{25}$

$= \frac{54}{25}$

$=$ **2.16 goals**

b) **Middle Position** $= \frac{25 + 1}{2}$

$=$ **13^{th} position**

$\therefore$ **Median = 2 goals** ← **The data with the middle position .**

c) **Mode = 3 goals** ← **The data with the highest frequency .**

EXAMPLE : 3

A survey was conducted to find the number of occupants in each unit of an apartment . The results are shown in the table below .

Number Of Occupants	**2**	**3**	**4**	**5**	**6**	**7**
Number Of Units	**3**	**6**	x	**4**	**10**	**5**

a) If the mode is 4, write down the smallest possible value of x **.**

b) If the mode is 6, write down the smallest possible value of x **.**

c) If the median is 5, find

i) The greatest possible value of x **.**

ii) The smallest possible value of x **.**

d) If the mean number of occupants is 2.75 find the value of x **.**

SOLUTION :

a) Smallest position value of x (mode = 4)

$= 11 \leftarrow \boxed{x > 10, \ \therefore x = 11}$

b) Greatest position value of x (mode = 6)

$= 9 \leftarrow \boxed{x > 10, \ \therefore x = 9}$

The Greatest position value of x occurs when the median is here .

The Smallest position value of x occurs when the median is here .

c) $\underbrace{2, 2, 2}_{3}, \underbrace{3, \ldots, 3}_{6}, \underbrace{4, \ldots, 4}_{x}, \underbrace{5, \ldots, 5}_{4}, \underbrace{6, \ldots, 6}_{10}, \underbrace{7, \ldots, 7}_{5}$

i) $3 + 6 + x = 3 + 10 + 5$

$9x + x = 18$

$x = 9$

$\therefore$ The Greatest possible value of x is 9 .

ii) $3 + 6 + x = 3 + 10 + 5$

$12 + x = 15$

$x = 3$

$\therefore$ The **Smallest** possible value of x is 3 .

d) **Mean** = 4.75

$$= \frac{(2 \times 3) + (3 \times 6) + (4 \times 1) + (5 \times 4) + (6 \times 10) + (7 \times 5)}{3 + 6 + x + 4 + 10 + 5} = 4.75$$

$$\frac{6 + 18 + 4x + 20 + 60 + 35}{28 + x} = 4.75$$

$$\frac{139 + 4x}{28 + x} = 4.75$$

Multiply both sides by $28 + x$ → $139 + 4x = 4.75\,(28 + x)$

$$139 + 4x = 133 + 4.75x$$

$$6 = 0.75x$$

$$x = \frac{6}{0.75}$$

$$= 8$$

EXAMPLE : 4

The number of story books each of a group of 30 students read in a certain month is shown in the table below .

Number Of Books	0	1	2	3	4	5	6
Number Of Students	2	5	x	4	2	y	1

a) Show that $x + y = 16$.

b) Given that the mean number of story books read is 2.8, show that $2x + 5y = 53$.

c) Find the value of x and y .

d) Hence, state the modal number of story books read .

SOLUTION :

a) Total number of students $= 30$

$$2 + 5 + x + 4 + 2 + y + 1 = 30$$

$$14 + x + y = 30$$

$$x + y = 16 \text{ (Shown)}$$

b) Mean $= 2.8$

$$\frac{(0 \times 2) + (1 \times 5) + (2 \times x) + (3 \times 4) + (4 \times 12) + (5 \times y) + (6 \times 1)}{3 + 6 + x + 4 + 10 + 5} = 2.8$$

Multiply both sides by 30 .

$$5 + 2x + 12 + 8 + 5y + 6 = 2.8 (30)$$

$$31 + 2x + 5y = 84$$

$$2x + 5y = 53 \text{ (Shown)}$$

c)

$$x + y = 16 \quad \text{—— } (1)$$

$$2x + 5y = 53 \quad \text{—— } (2)$$

$$(1) \times (2): \quad 2x + 2y = 32 \quad \text{—— } (3)$$

$$(2) - (3): \quad 3y = 21$$

$$y = 7$$

Solve the two equations simultaneously to find the value of x and y .

SUBSTITUTE

$y = 7$ into (1) :

$$x + 7 = 16$$

$$x = 9$$

$\therefore x = 9$ and $y = 7$

d)

Number Of Books	0	1	2	3	4	5	6
Number Of Students	2	5	9	4	2	7	1

Modal number of story books read = 2

Substitute the values of x and y into the table to find the modal no . of story books read .

1.9 MEAN FOR GROUPED DATA

To calculate the estimated mean when data are grouped into intervals use :

MEAN $\bar{x} = \frac{\sum fx}{\sum f}$

Where x = Mid – value of the class interval

and f = Frequency of the class interval .

EXAMPLE :

The speeds of 50 cars passing a speed camera on an expressway are shown in the table below .

a) Calculate an estimate of the mean speed of the cars .

b) Write down the modal class of this distribution .

c) Find the class interval where the median lies .

Speed (s km / h)	$50 < s \leq 60$	$60 < s \leq 70$	$70 < s \leq 80$	$80 < s \leq 90$	$90 < s \leq 100$
Number Of Cars	8	15	19	6	2

SOLUTION :

a)

SPEED (s km / h)	FREQUENCY (f)	MID – VALUE (x)	fx
$50 < s \leq 60$	8	55	440
$60 < s \leq 70$	15	65	975
$70 < s \leq 80$	19	75	1425
$80 < s \leq 90$	6	85	510
$90 < s \leq 100$	2	95	190
	$\sum f = 50$		$\sum fx = 3540$

Estimated mean speed $= \frac{\sum fx}{\sum f}$

$= \frac{3540}{50}$

$=$ 70.8 km / h

TIPS FOR STUDENTS

The mid – value of the first class interval = $\frac{50 + 60}{2}$ = 55

b) **Modal Class** = $70 < s \leq 80$

The class interval with the highest frequency is called the modal class .

c) **Middle Position** = $\frac{50 + 1}{2}$

= 25.5^{th} Position

Median = Mean of the 25^{th} and 26^{th} values

The 25^{th} and 26^{th} values fall in the class interval $70 < s \leq 80$.

$\therefore$ The median lies in the class interval $70 < s \leq 80$

1.10 STANDARD DEVIATION

The standard deviation , S measures the spread of a set of data from mean.

STANDARD DEVIATION FOR UNGROUPED DATA

1. **To find the standard deviation of a data set $\{ x_1,\ x_2,\ x_3, \ldots, x_n \}$ use :**

$$\boxed{S = \sqrt{\frac{\Sigma(x - \overline{x})^2}{n}}}$$

Where $\overline{x}$ = Mean = $\frac{\Sigma x}{n}$

And n = Total Number Of Data In The Set .

2. **Another formula to calculate the standard deviation is shown below :**

$$\boxed{S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}}$$

Where $\frac{\Sigma x^2}{n} = \frac{1}{n}\,(x_1^2,\ x_2^2,\ x_3^2, \ldots, x_n^2)$

And $\left(\frac{\sum x}{n}\right)^2$ = Square of the mean .

EXAMPLE : 1

The lengths of 5 worms are given below:

2 cm, 3 cm, 5 cm, 8 cm, and 10 cm.

Find the means and the standard deviation of these lengths.

SOLUTION :

METHOD 1 : Using the formula $S = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$

Mean $\bar{x} = \frac{\sum x}{n}$

$= \frac{2 + 3 + 5 + 8 + 10}{5}$

$= \frac{28}{5}$

$=$ 5.6 cm

x	$x - \bar{x}$	$(x - \bar{x})^2$
2	- 3.6	12.96
3	- 2.6	6.76
5	- 0.6	0.36
8	- 2.4	5.76
10	- 4.4	19.36
		$\sum(x-\bar{x})^2$ = **45.2**

Standard Deviation, S = $\sqrt{\frac{\sum(x - \bar{x})^2}{n}}$

= $\sqrt{\frac{45.2}{5}}$

= **3.01 cm (Correct to 3 sig . fig)**

METHOD 2 : Using the formula $S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

Mean $\bar{x} = \frac{\sum x}{n}$

$= \frac{2 + 3 + 5 + 8 + 10}{5}$

$= \frac{28}{5}$

$= 5.6$ cm

$\sum x^2 \;=\; 2^2 + 3^2 + 5^2 + 8^2 + 10^2$

Standard Deviation, S $= \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

$= \sqrt{\frac{202}{5} - 5.6^2}$

$= 3.01$ cm (Correct to 3 sig . fig)

EXAMPLE : 2

a) **Calculate the mean and the standard deviation of the marks of two groups of students, *A* and *B* .**

GROUP A : 12, 15, 15, 17, 18, 20, 21, 22

GROUP B : 3, 5, 6, 16, 17, 24, 28, 40

b) **Compare and comment briefly on the two group of marks .**

SOLUTION :

a) GROUP A :

$$\text{Mean} = \frac{12 + 15 + 15 + 17 + 18 + 20 + 21 + 22}{8}$$

$$= \frac{140}{8}$$

$$= 17.5 \text{ Marks}$$

$$\sum x^2 = 12^2 + 15^2 + 15^2 + 17^2 + 18^2 + 20^2 + 21^2 + 22^2$$

$$= 2532$$

$$\text{Standard Deviation, } S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$= \sqrt{\frac{2532}{8} - 17.5^2}$$

$$= 3.20 \text{ Marks (Correct to 3 sig . fig)}$$

b) GROUP B :

$$\text{Mean} = \frac{3 + 5 + 6 + 16 + 17 + 24 + 28 + 40}{8}$$

$$= \frac{139}{8}$$

$$= 17.375 \text{ Marks}$$

$\sum x^2 = 3^2 + 5^2 + 6^2 + 16^2 + 17^2 + 24^2 + 28^2 + 40^2$

$= 2532$

Standard Deviation, $S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

$= \sqrt{\frac{3575}{8} - 17.375^2}$

$=$ **12.0 Marks (Correct to 3 sig . fig)**

b) **The two groups have about the same values of mean but very different values of standard deviation has a narrower spread of marks around the mean . Most students in Group *A* scored about 17.5 marks . Since the standard deviation of Group *B* is higher, the marks are more widely spreed from the mean . The students in Group *B* have more extreme performances .**

TIPS FOR STUDENTS

- ✓ **A smaller spread has a smaller standard deviation .**
- ✓ **A wider spread has a higher standard deviation .**
- ✓ **Standard deviation is always greater or equal to zero .**

STANDARD DEVIATION FOR GROUPED DATA

1. To find the standard deviation of a set of grouped data in the form of a frequency table, use :

$$\boxed{S = \sqrt{\frac{\sum f(x - \bar{x})^2}{\sum f}}}$$

Where $\bar{x}$ = Mean = $\frac{\sum f}{n}$

And $\sum f$ = Total Frequency .

2. Another formula to calculate the standard deviation for grouped data is shown below :

$$\boxed{S = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}}$$

Where $\left(\frac{\sum fx}{\sum f}\right)^2$ = Square of the mean

and $\sum f$ = Total Frequency .

EXAMPLE : 1

The table below shows the time taken by 40 students for their 2.4 km run .

TIME (MIN)	10	11	12	13	14
FREQUENCY	1	5	8	21	5

Find the standard deviation of the time taken .

SOLUTION :

METHOD 1 : Using the formula $S = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}}$

a)

x	FREQUENCY (f)	fx	$x - \bar{x}$	$(x - \bar{x})^2$	$f(x - \bar{x})^2$
10	1	10	- 2.6	6.76	6.76
11	5	55	- 1.6	2.56	12.8
12	8	96	- 0.6	0.36	2.88
13	21	273	0.4	0.16	3.36
14	5	70	1.4	1.96	9.8
	$\sum f = 40$	$\sum fx = 540$			$\sum(x - \bar{x})^2 = 35.6$

Mean $\overline{x} = \frac{\sum fx}{\sum f}$

$= \frac{504}{40}$

$=$ **12.6 Min**

Standard Deviation, $S = \frac{\sum f(x - \overline{x})^2}{\sum f}$

$= \sqrt{\frac{35.6}{40}}$

$=$ **0.943 Min (Correct to 3 sig . fig)**

METHOD 2 : Using the formula $S = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$

x	FREQUENCY (f)	fx	$f(x - \overline{x})^2$
10	1	10	100
11	5	55	605
12	8	96	1152
13	21	273	3549
14	5	70	980
	$\sum f = 40$	$\sum fx = 540$	$\sum (x - \overline{x})^2 = 6386$

Standard Deviation, S = $\sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$

= $\sqrt{\frac{6386}{40} - \left(\frac{504}{40}\right)^2}$

= **0.943 Min (Correct to 3 sig . fig)**

EXAMPLE : 2

The table shows the ages of 50 participants in a cooking competition .

AGE (A YEARS)	$20 < a \leq 30$	$30 < a \leq 40$	$40 < a \leq 50$	$50 < a \leq 60$	$60 < a \leq 70$
FREQUENCY	8	11	14	12	5

Calculate the mean and the standard deviation of the ages .

SOLUTION :

$$\frac{20 + 30}{2}$$

AGE (A YEARS)	FREQUENCY (f)	MID – VALUE (x)	fx	fx^2
$20 < a \leq 30$	8	25	200	5000
$30 < a \leq 40$	11	35	385	13 475
$40 < a \leq 50$	14	45	630	28 350
$50 < a \leq 60$	12	55	660	36 300
$60 < a \leq 70$	5	65	325	21 125
	$\sum f = 50$		$\sum fx = 2200$	$\sum fx^2 = 104\,250$

Mean Age $= \dfrac{\sum fx}{\sum f}$

$= \dfrac{2200}{50}$

$=$ 44 Years

Standard Deviation, $S = \dfrac{\sum f(x - \bar{x})^2}{\sum f}$

$= \sqrt{\dfrac{104\,250}{50} - 44^2}$

$=$ 12.2 Years (Correct to 3 sig . fig)

EXAMPLE : 3

A group of 100 students each from School *A* and School B were asked the amount of time they use their handphones each week . The results are given in the tables below .

SCHOOL A

TIME (HOURS)	0 - 4	5 - 9	10 - 14	15 - 19	20 - 24
NUMBER OF STUDENTS	6	9	28	32	25

SCHOOL B

MEAN = 15.05 Hours
STANDARD DEVIATION = 7.82 Hours

a) For school *A*, calculate the mean and the standard deviation of the number of hours the group of students use their handphones each week .

b) Compare, briefly the results of these two schools .

SOLUTION :

a) **SCHOOL A**

$\dfrac{0 + 4}{2}$

TIME (HOURS)	FREQUENCY (f)	MID – VALUE (x)	fx	fx^2
0 – 4	6	2	12	24
5 – 9	9	7	63	441
10 – 14	28	12	336	4032
15 – 19	32	17	544	9248
20 – 24	25	22	550	12 100
	$\sum f = 100$		$\sum fx = 1505$	$\sum fx^2 = 25\,845$

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

$$= \frac{1505}{100}$$

$$= 15.05 \text{ h}$$

Standard Deviation, S = $\sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$

$= \sqrt{\frac{25845}{100} - 15.05^2}$

= **5.65 h (Correct to 3 sig . fig)**

b) Both school have the same mean . School *B* has a higher standard deviation . That means that the time spent using the hanhphones in school *B* are more widely spread . Some of its students spent a long time on their handphones while some other spent a very short time .