

STATISTICS AND PROBABILITY

(A) Main Concepts and Results

Statistics

Meaning of 'statistics', Primary and secondary data, Raw/ungrouped data, Range of data, Grouped data-class intervals, Class marks, Presentation of data - frequency distribution table, Discrete frequency distribution and continuous frequency distribution.

- Graphical representation of data :
 - (i) Bar graphs
 - (ii) Histograms of uniform width and of varying widths
 - (iii) Frequency polygons
- Measures of Central tendency
 - (a) **Mean**
 - (i) **Mean of raw data**

$$\text{Mean} = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

where x_1, x_2, \dots, x_n are n observations.

(ii) **Mean of ungrouped data**

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

where f_i 's are frequencies of x_i 's.

(b) Median

A median is the value of the observation which divides the data into two equal parts, when the data is arranged in ascending (or descending) order.

Calculation of Median

When the ungrouped data is arranged in ascending (or descending) order, the median of data is calculated as follows :

- (i) When the number of observations (n) is odd, the median is the value of the

$\left(\frac{n+1}{2}\right)^{th}$ observation.

- (ii) When the number of observations (n) is even, the median is the average or

mean of the $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ observations.

(c) Mode

The observation that occurs most frequently, i.e., the observation with maximum frequency is called **mode**. Mode of ungrouped data can be determined by observation/inspection.

Probability

- Random experiment or simply an experiment
- Outcomes of an experiment
- Meaning of a trial of an experiment
- The experimental (or empirical) probability of an event E (denoted by P(E)) is given by

$$P(E) = \frac{\text{Number of trials in which the event has happened}}{\text{Total number of trials}}$$

- The probability of an event E can be any number from 0 to 1. It can also be 0 or 1 in some special cases.

(B) Multiple Choice Questions

Write the correct answer in each of the following :

Sample Question 1: The marks obtained by 17 students in a mathematics test (out of 100) are given below :

91, 82, 100, 100, 96, 65, 82, 76, 79, 90, 46, 64, 72, 68, 66, 48, 49.

The range of the data is :

- (A) 46 (B) 54 (C) 90 (D) 100

Solution : Answer (B)

Sample Question 2: The class-mark of the class 130-150 is :

- (A) 130 (B) 135 (C) 140 (D) 145

Solution : Answer (C)

Sample Question 3 : A die is thrown 1000 times and the outcomes were recorded as follows :

Outcome	1	2	3	4	5	6
Frequency	180	150	160	170	150	190

If the die is thrown once more, then the probability that it shows 5 is :

- (A) $\frac{9}{50}$ (B) $\frac{3}{20}$ (C) $\frac{4}{25}$ (D) $\frac{7}{25}$

Solution : Answer (B)

EXERCISE 14.1

Write the correct answer in each of the following :

1. The class mark of the class 90-120 is :

- (A) 90 (B) 105 (C) 115 (D) 120

2. The range of the data :

25, 18, 20, 22, 16, 6, 17, 15, 12, 30, 32, 10, 19, 8, 11, 20 is

- (A) 10 (B) 15 (C) 18 (D) 26

3. In a frequency distribution, the mid value of a class is 10 and the width of the class is 6. The lower limit of the class is :

- (A) 6 (B) 7 (C) 8 (D) 12

the adjusted frequency for the class 25-45 is :

- (A) 6 (B) 5 (C) 3 (D) 2

11. The mean of five numbers is 30. If one number is excluded, their mean becomes 28. The excluded number is :

- (A) 28 (B) 30 (C) 35 (D) 38

12. If the mean of the observations :

$$x, x + 3, x + 5, x + 7, x + 10$$

is 9, the mean of the last three observations is

- (A) $10\frac{1}{3}$ (B) $10\frac{2}{3}$ (C) $11\frac{1}{3}$ (D) $11\frac{2}{3}$

13. If \bar{x} represents the mean of n observations x_1, x_2, \dots, x_n , then value of $\sum_{i=1}^n (x_i - \bar{x})$ is:

- (A) -1 (B) 0 (C) 1 (D) $n - 1$

14. If each observation of the data is increased by 5, then their mean

- (A) remains the same (B) becomes 5 times the original mean
(C) is decreased by 5 (D) is increased by 5

15. Let \bar{x} be the mean of x_1, x_2, \dots, x_n and \bar{y} the mean of y_1, y_2, \dots, y_n . If \bar{z} is the mean of $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n$, then \bar{z} is equal to

- (A) $\bar{x} + \bar{y}$ (B) $\frac{\bar{x} + \bar{y}}{2}$ (C) $\frac{\bar{x} + \bar{y}}{n}$ (D) $\frac{\bar{x} + \bar{y}}{2n}$

16. If \bar{x} is the mean of x_1, x_2, \dots, x_n , then for $a \neq 0$, the mean of $ax_1, ax_2, \dots, ax_n, \frac{x_1}{a},$

$$\frac{x_2}{a}, \dots, \frac{x_n}{a}$$
 is

- (A) $\left(a + \frac{1}{a}\right)\bar{x}$ (B) $\left(a + \frac{1}{a}\right)\frac{\bar{x}}{2}$ (C) $\left(a + \frac{1}{a}\right)\frac{\bar{x}}{n}$ (D) $\frac{\left(a + \frac{1}{a}\right)\bar{x}}{2n}$

17. If $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_n$ are the means of n groups with n_1, n_2, \dots, n_n number of observations respectively, then the mean \bar{x} of all the groups taken together is given by :

$$(A) \sum_{i=1}^n n_i \bar{x}_i \quad (B) \frac{\sum_{i=1}^n n_i \bar{x}_i}{n^2} \quad (C) \frac{\sum_{i=1}^n n_i \bar{x}_i}{\sum_{i=1}^n n_i} \quad (D) \frac{\sum_{i=1}^n n_i \bar{x}_i}{2n}$$

18. The mean of 100 observations is 50. If one of the observations which was 50 is replaced by 150, the resulting mean will be :

- (A) 50.5 (B) 51 (C) 51.5 (D) 52

19. There are 50 numbers. Each number is subtracted from 53 and the mean of the numbers so obtained is found to be -3.5 . The mean of the given numbers is :

- (A) 46.5 (B) 49.5 (C) 53.5 (D) 56.5

20. The mean of 25 observations is 36. Out of these observations if the mean of first 13 observations is 32 and that of the last 13 observations is 40, the 13th observation is :

- (A) 23 (B) 36 (C) 38 (D) 40

21. The median of the data

78, 56, 22, 34, 45, 54, 39, 68, 54, 84 is

- (A) 45 (B) 49.5 (C) 54 (D) 56

22. For drawing a frequency polygon of a continuous frequency distribution, we plot the points whose ordinates are the frequencies of the respective classes and abscissae are respectively :

- (A) upper limits of the classes (B) lower limits of the classes
(C) class marks of the classes (D) upper limits of preceding classes

23. Median of the following numbers :

4, 4, 5, 7, 6, 7, 7, 12, 3 is

- (A) 4 (B) 5 (C) 6 (D) 7

24. Mode of the data

15, 14, 19, 20, 14, 15, 16, 14, 15, 18, 14, 19, 15, 17, 15 is

- (A) 14 (B) 15 (C) 16 (D) 17

25. In a sample study of 642 people, it was found that 514 people have a high school certificate. If a person is selected at random, the probability that the person has a high school certificate is :

- (A) 0.5 (B) 0.6 (C) 0.7 (D) 0.8

26. In a survey of 364 children aged 19-36 months, it was found that 91 liked to eat potato chips. If a child is selected at random, the probability that he/she does not like to eat potato chips is :

- (A) 0.25 (B) 0.50 (C) 0.75 (D) 0.80

27. In a medical examination of students of a class, the following blood groups are recorded:

Blood group	A	AB	B	O
Number of students	10	13	12	5

A student is selected at random from the class. The probability that he/she has blood group B, is :

- (A) $\frac{1}{4}$ (B) $\frac{13}{40}$ (C) $\frac{3}{10}$ (D) $\frac{1}{8}$

28. Two coins are tossed 1000 times and the outcomes are recorded as below :

Number of heads	2	1	0
Frequency	200	550	250

Based on this information, the probability for at most one head is

- (A) $\frac{1}{5}$ (B) $\frac{1}{4}$ (C) $\frac{4}{5}$ (D) $\frac{3}{4}$

29. 80 bulbs are selected at random from a lot and their life time (in hrs) is recorded in the form of a frequency table given below :

Life time (in hours)	300	500	700	900	1100
Frequency	10	12	23	25	10

One bulb is selected at random from the lot. The probability that its life is 1150 hours, is

- (A) $\frac{1}{80}$ (B) $\frac{7}{16}$ (C) 0 (D) 1

30. Refer to Q.29 above :

The probability that bulbs selected randomly from the lot has life less than 900 hours is :

- (A) $\frac{11}{40}$ (B) $\frac{5}{16}$ (C) $\frac{7}{16}$ (D) $\frac{9}{16}$

(C) Short Answer Questions with Reasoning

Sample Question 1 : The mean of the data :

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

is given to be 5. Based on this information, is it correct to say that the mean of the data:

10, 12, 10, 2, 18, 8, 12, 6, 12, 10, 8, 10, 12, 16, 4

is 10? Give reason.

Solution : It is correct. Since the 2nd data is obtained by multiplying each observation of 1st data by 2, therefore, the mean will be 2 times the mean of the 1st data.

Sample Question 2 : In a histogram, the areas of the rectangles are proportional to the frequencies. Can we say that the lengths of the rectangles are also proportional to the frequencies?

Solution: No. It is true only when the class sizes are the same.

Sample Question 3 : Consider the data : 2, 3, 9, 16, 9, 3, 9. Since 16 is the highest value in the observations, is it correct to say that it is the mode of the data? Give reason.

Solution : 16 is not the mode of the data. The mode of a given data is the observation with highest frequency and not the observation with highest value.

EXERCISE 14.2

1. The frequency distribution :

Marks	0-20	20-40	40-60	60-100
Number of Students	10	15	20	25

has been represented graphically as follows :

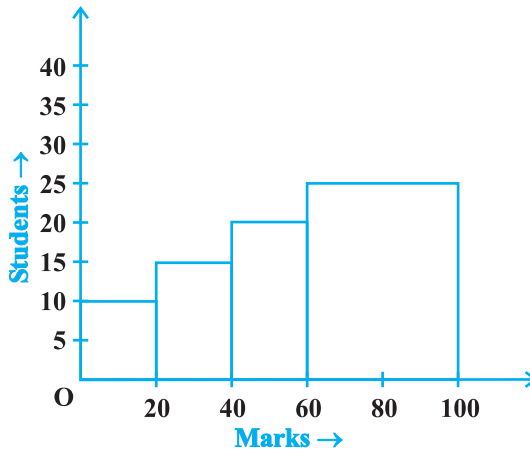


Fig. 14.1

Do you think this representation is correct? Why?

- In a diagnostic test in mathematics given to students, the following marks (out of 100) are recorded :

46, 52, 48, 11, 41, 62, 54, 53, 96, 40, 98, 44

Which ‘average’ will be a good representative of the above data and why?

- A child says that the median of 3, 14, 18, 20, 5 is 18. What doesn’t the child understand about finding the median?

- A football player scored the following number of goals in the 10 matches :

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

Since the number of matches is 10 (an even number), therefore, the median

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

$$= \frac{8+6}{2} = 7$$

Is it the correct answer and why?

- Is it correct to say that in a histogram, the area of each rectangle is proportional to the class size of the corresponding class interval? If not, correct the statement.

- The class marks of a continuous distribution are :

1.04, 1.14, 1.24, 1.34, 1.44, 1.54 and 1.64

Is it correct to say that the last interval will be 1.55 - 1.73? Justify your answer.

7. 30 children were asked about the number of hours they watched TV programmes last week. The results are recorded as under :

Number of hours	0-5	5-10	10-15	15-20
Frequency	8	16	4	2

Can we say that the number of children who watched TV for 10 or more hours a week is 22? Justify your answer.

8. Can the experimental probability of an event be a negative number? If not, why?
 9. Can the experimental probability of an event be greater than 1? Justify your answer.
 10. As the number of tosses of a coin increases, the ratio of the number of heads to the

total number of tosses will be $\frac{1}{2}$. Is it correct? If not, write the correct one.

(D) Short Answer Questions

Sample Question 1 : Heights (in cm) of 30 girls of Class IX are given below:

140, 140, 160, 139, 153, 153, 146, 150, 148, 150, 152,
 146, 154, 150, 160, 148, 150, 148, 140, 148, 153, 138,
 152, 150, 148, 138, 152, 140, 146, 148.

Prepare a frequency distribution table for this data.

Solution : Frequency distribution of heights of 30 girls

Height (in cm)	Tally Marks	Frequency
138		2
139		1
140		4
146		3
148	 	6
150	 	5
152		3
153		3
154		1
160		2
Total		30

Sample Question 2 : The following observations are arranged in ascending order :
 26, 29, 42, 53, x , $x + 2$, 70, 75, 82, 93

If the median is 65, find the value of x .

Solution : Number of observations (n) = 10, which is even. Therefore, median is the mean of $\left(\frac{n}{2}\right)^{\text{th}}$ and $\left(\frac{n}{2} + 1\right)^{\text{th}}$ observation, i.e., 5th and 6th observation.

Here, 5^{th} observation = x
 6^{th} observation = $x + 2$

$$\text{Median} = \frac{x + (x + 2)}{2} = x + 1$$

Now, $x + 1 = 65$ (Given)

Therefore, $x = 64$

Thus, the value of x is 64.

Sample Question 3 : Here is an extract from a mortality table.

Age (in years)	Number of persons surviving out of a sample of one million
60	16090
61	11490
62	8012
63	5448
64	3607
65	2320

- (i) Based on this information, what is the probability of a person ‘aged 60’ of dying within a year?
- (ii) What is the probability that a person ‘aged 61’ will live for 4 years?

Solution :

- (i) We see that 16090 persons aged 60, (16090-11490), i.e., 4600 died before reaching their 61st birthday.

$$\text{Therefore, } P(\text{a person aged 60 die within a year}) = \frac{4600}{16090} = \frac{460}{1609}$$

- (ii) Number of persons aged 61 years = 11490
 Number of persons surviving for 4 years = 2320

$$P(\text{a person aged 61 will live for 4 years}) = \frac{2320}{11490} = \frac{232}{1149}$$

EXERCISE 14.3

- The blood groups of 30 students are recorded as follows:
 A, B, O, A, AB, O, A, O, B, A, O, B, A, AB, B, A, AB, B,
 A, A, O, A, AB, B, A, O, B, A, B, A
 Prepare a frequency distribution table for the data.
- The value of π upto 35 decimal places is given below:
 3. 14159265358979323846264338327950288
 Make a frequency distribution of the digits 0 to 9 after the decimal point.
- The scores (out of 100) obtained by 33 students in a mathematics test are as follows:
 69, 48, 84, 58, 48, 73, 83, 48, 66, 58, 84
 66, 64, 71, 64, 66, 69, 66, 83, 66, 69, 71
 81, 71, 73, 69, 66, 66, 64, 58, 64, 69, 69
 Represent this data in the form of a frequency distribution.
- Prepare a continuous grouped frequency distribution from the following data:

Mid-point	Frequency
5	4
15	8
25	13
35	12
45	6

Also find the size of class intervals.

- Convert the given frequency distribution into a continuous grouped frequency distribution:

Class interval	Frequency
150-153	7
154-157	7
158-161	15
162-165	10
166-169	5
170-173	6

In which intervals would 153.5 and 157.5 be included?

6. The expenditure of a family on different heads in a month is given below:

Head	Food	Education	Clothing	House Rent	Others	Savings
Expenditure (in Rs)	4000	2500	1000	3500	2500	1500

Draw a bar graph to represent the data above.

7. Expenditure on Education of a country during a five year period (2002-2006), in crores of rupees, is given below:

Elementary education	240
Secondary Education	120
University Education	190
Teacher’s Training	20
Social Education	10
Other Educational Programmes	115
Cultural programmes	25
Technical Education	125

Represent the information above by a bar graph.

8. The following table gives the frequencies of most commonly used letters *a, e, i, o, r, t, u* from a page of a book :

Letters	<i>a</i>	<i>e</i>	<i>i</i>	<i>o</i>	<i>r</i>	<i>t</i>	<i>u</i>
Frequency	75	125	80	70	80	95	75

Represent the information above by a bar graph.

9. If the mean of the following data is 20.2, find the value of p :

x	10	15	20	25	30
f	6	8	p	10	6

10. Obtain the mean of the following distribution:

Frequency	Variable
4	4
8	6
14	8
11	10
3	12

11. A class consists of 50 students out of which 30 are girls. The mean of marks scored by girls in a test is 73 (out of 100) and that of boys is 71. Determine the mean score of the whole class.
12. Mean of 50 observations was found to be 80.4. But later on, it was discovered that 96 was misread as 69 at one place. Find the correct mean.
13. Ten observations 6, 14, 15, 17, $x + 1$, $2x - 13$, 30, 32, 34, 43 are written in an ascending order. The median of the data is 24. Find the value of x .
14. The points scored by a basket ball team in a series of matches are as follows:
17, 2, 7, 27, 25, 5, 14, 18, 10, 24, 48, 10, 8, 7, 10, 28
Find the median and mode for the data.
15. In Fig. 14.2, there is a histogram depicting daily wages of workers in a factory. Construct the frequency distribution table.

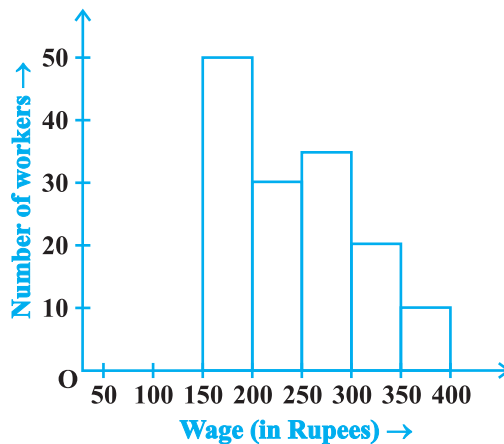


Fig. 14.2

- 16.** A company selected 4000 households at random and surveyed them to find out a relationship between income level and the number of television sets in a home. The information so obtained is listed in the following table:

Monthly income (in Rs)	Number of Televisions/household			
	0	1	2	Above 2
< 10000	20	80	10	0
10000 - 14999	10	240	60	0
15000 - 19999	0	380	120	30
20000 - 24999	0	520	370	80
25000 and above	0	1100	760	220

Find the probability:

- (i) of a household earning Rs 10000 – Rs 14999 per year and having exactly one television.
 - (ii) of a household earning Rs 25000 and more per year and owning 2 televisions.
 - (iii) of a household not having any television.
- 17.** Two dice are thrown simultaneously 500 times. Each time the sum of two numbers appearing on their tops is noted and recorded as given in the following table:

Sum	Frequency
2	14
3	30
4	42
5	55
6	72
7	75
8	70
9	53
10	46
11	28
12	15

If the dice are thrown once more, what is the probability of getting a sum

- (i) 3? (ii) more than 10?
 (iii) less than or equal to 5? (iv) between 8 and 12?

18. Bulbs are packed in cartons each containing 40 bulbs. Seven hundred cartons were examined for defective bulbs and the results are given in the following table:

Number of defective bulbs	0	1	2	3	4	5	6	more than 6
Frequency	400	180	48	41	18	8	3	2

One carton was selected at random. What is the probability that it has

- (i) no defective bulb?
 (ii) defective bulbs from 2 to 6?
 (iii) defective bulbs less than 4?
19. Over the past 200 working days, the number of defective parts produced by a machine is given in the following table:

Number of defective parts	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Days	50	32	22	18	12	12	10	10	10	8	6	6	2	2

Determine the probability that tomorrow's output will have

- (i) no defective part
 (ii) atleast one defective part
 (iii) not more than 5 defective parts
 (iv) more than 13 defective parts
20. A recent survey found that the ages of workers in a factory is distributed as follows:

Age (in years)	20 - 29	30 - 39	40 - 49	50 - 59	60 and above
Number of workers	38	27	86	46	3

If a person is selected at random, find the probability that the person is:

- (i) 40 years or more
 (ii) under 40 years

- (iii) having age from 30 to 39 years
- (iv) under 60 but over 39 years

(E) Long Answer Questions

Sample Question 1: Following is the frequency distribution of total marks obtained by the students of different sections of Class VIII.

Marks	100 - 150	150 - 200	200 - 300	300 - 500	500 - 800
Number of students	60	100	100	80	180

Draw a histogram for the distribution above.

Solution: In the given frequency distribution, the class intervals are not of equal width.

Therefore, we would make modifications in the lengths of the rectangles in the histogram so that the areas of rectangles are proportional to the frequencies. Thus, we have:

Marks	Frequency	Width of the class	Length of the rectangle
100 - 150	60	50	$\frac{50}{50} \times 60 = 60$
150 - 200	100	50	$\frac{50}{50} \times 100 = 100$
200 - 300	100	100	$\frac{50}{100} \times 100 = 50$
300 - 500	80	200	$\frac{50}{200} \times 80 = 20$
500 - 800	180	300	$\frac{50}{300} \times 180 = 30$

Now, we draw rectangles with lengths as given in the last column. The histogram of the data is given below :

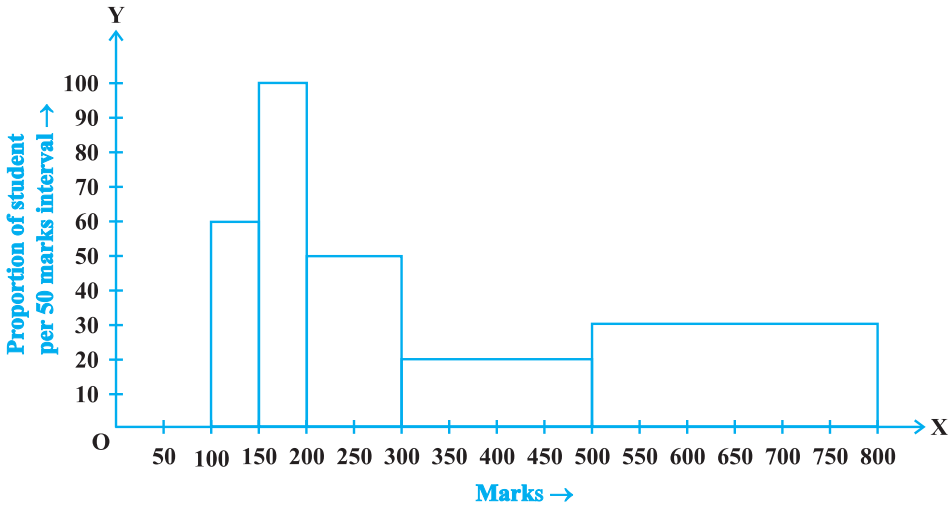


Fig. 14.3

Sample Question 2 : Two sections of Class IX having 30 students each appeared for mathematics olympiad. The marks obtained by them are shown below:

46 31 74 68 42 54 14 61 83 48 37 26 8 64 57
 93 72 53 59 38 16 88 75 56 46 66 45 61 54 27
 27 44 63 58 43 81 64 67 36 49 50 76 38 47 55
 77 62 53 40 71 60 58 45 42 34 46 40 59 42 29

Construct a group frequency distribution of the data above using the classes 0-9, 10-19 etc., and hence find the number of students who secured more than 49 marks.

Solution :

Class	Tally Marks	Frequency
0-9		1
10-19		2
20-29		4
30-39		6
40-49		15
50-59		12
60-69		10
70-79		6
80-89		3
90-99		1
Total		60

From the table above, we find that the number of students who secure more than 49 marks is $(12 + 10 + 6 + 3 + 1)$, i.e., 32.

EXERCISE 14.4

1. The following are the marks (out of 100) of 60 students in mathematics.

16, 13, 5, 80, 86, 7, 51, 48, 24, 56, 70, 19, 61, 17, 16, 36, 34, 42, 34, 35, 72, 55, 75, 31, 52, 28, 72, 97, 74, 45, 62, 68, 86, 35, 85, 36, 81, 75, 55, 26, 95, 31, 7, 78, 92, 62, 52, 56, 15, 63, 25, 36, 54, 44, 47, 27, 72, 17, 4, 30.

Construct a grouped frequency distribution table with width 10 of each class starting from 0 - 9.

2. Refer to Q1 above. Construct a grouped frequency distribution table with width 10 of each class, in such a way that one of the classes is 10 - 20 (20 not included).
3. Draw a histogram of the following distribution :

Heights (in cm)	Number of students
150 - 153	7
153 - 156	8
156 - 159	14
159 - 162	10
162 - 165	6
165 - 168	5

4. Draw a histogram to represent the following grouped frequency distribution :

Ages (in years)	Number of teachers
20 - 24	10
25 - 29	28
30 - 34	32
35 - 39	48
40 - 44	50
45 - 49	35
50 - 54	12

5. The lengths of 62 leaves of a plant are measured in millimetres and the data is represented in the following table :

Length (in mm)	Number of leaves
118 - 126	8
127 - 135	10
136 - 144	12
145 - 153	17
154 - 162	7
163 - 171	5
172 - 180	3

Draw a histogram to represent the data above.

6. The marks obtained (out of 100) by a class of 80 students are given below :

Marks	Number of students
10 - 20	6
20 - 30	17
30 - 50	15
50 - 70	16
70 - 100	26

Construct a histogram to represent the data above.

7. Following table shows a frequency distribution for the speed of cars passing through at a particular spot on a high way :

Class interval (km/h)	Frequency
30 - 40	3
40 - 50	6
50 - 60	25
60 - 70	65
70 - 80	50
80 - 90	28
90 - 100	14

Draw a histogram and frequency polygon representing the data above.

8. Refer to Q. 7 :

Draw the frequency polygon representing the above data without drawing the histogram.

9. Following table gives the distribution of students of sections A and B of a class according to the marks obtained by them.

Section A		Section B	
Marks	Frequency	Marks	Frequency
0 - 15	5	0 - 15	3
15 - 30	12	15 - 30	16
30 - 45	28	30 - 45	25
45 - 60	30	45 - 60	27
60 - 75	35	60 - 75	40
75 - 90	13	75 - 90	10

Represent the marks of the students of both the sections on the same graph by two frequency polygons. What do you observe?

10. The mean of the following distribution is 50.

x	f
10	17
30	$5a + 3$
50	32
70	$7a - 11$
90	19

Find the value of a and hence the frequencies of 30 and 70.

11. The mean marks (out of 100) of boys and girls in an examination are 70 and 73, respectively. If the mean marks of all the students in that examination is 71, find the ratio of the number of boys to the number of girls.

12. A total of 25 patients admitted to a hospital are tested for levels of blood sugar, (mg/dl) and the results obtained were as follows :

87	71	83	67	85
77	69	76	65	85
85	54	70	68	80
73	78	68	85	73
81	78	81	77	75

Find mean, median and mode (mg/dl) of the above data.