

STATISTICS

8 CHAPTER

CONTENTS

- Class Mark
- Cumulative Frequency
- Histogram
- Frequency Polygon
- Mean
- Median
- Mode



IMPORTANT POINTS

- ◆ The word data means information. Statistical data are of two types :
 - (i) Primary data
 - (ii) Secondary data
- ◆ When an investigator collects data himself with a definite plan or design in his (her) mind, it is called **Primary data**.
- ◆ Data which are not originally collected rather obtained from published or unpublished sources are known as **Secondary data**.
- ◆ After collection of data, the investigator has to find ways to condense them in tabular form in order to study their salient features. Such an arrangement is called **Presentation of data**.
- ◆ Raw data (if the data is arranged in the manner as it was collected, then it is called as Raw data) when put in ascending or descending order of magnitude is called an array or arranged data.
- ◆ The number of times an observation occurs in the given data is called frequency of the observation
- ◆ Classes/class intervals are the groups in which all the observations are divided

- ◆ Suppose class-interval is 10-20, then 10 is called lower limit and 20 is called upper limit of the class

- ◆ Mid-value of class-interval is called **Class-mark**

$$\text{Class-mark} = \frac{\text{lower limit} + \text{upper limit}}{2}$$

$$\text{Class-mark} = \text{lower limit} + \frac{1}{2}$$

(difference between the upper and lower limits)

- ◆ If the frequency of first class interval is added to the frequency of second class and this sum is added to third class and so on then frequencies so obtained are known as **Cumulative Frequency (c.f.)**.
- ◆ There are two types of cumulative frequencies (a) less than, (b) greater than

◆ EXAMPLES ◆

- Ex.1** Given below are the ages of 25 students of class IX in a school. Prepare a discrete frequency distribution.

15, 16, 16, 14, 17, 17, 16, 15, 15, 16, 16, 17, 15, 16, 16, 14, 16, 15, 14, 15, 16, 16, 15, 14, 15.

- Sol.** Frequency distribution of ages of 25 students

Age	Tally marks	Frequency
14		4
15		8
16		10
17		3
Total		25

- Ex.2** Form a discrete frequency distribution from the following scores:-

15, 18, 16, 20, 25, 24, 25, 20, 16, 15, 18, 18, 16, 24, 15, 20, 28, 30, 27, 16, 24, 25, 20, 18, 28, 27, 25, 24, 24, 18, 18, 25, 20, 16, 15, 20, 27, 28, 29, 16.

Variate	Tally marks	Frequency
15		4
16		6
18		6
20		6
24		5
25		5
27		3
28		3
29		1
30		1
Total		40

Ex.3 The water tax bills (in rupees) of 30 houses in a locality are given below. Construct a grouped frequency distribution with class size of 10.

30, 32, 45, 54, 74, 78, 108, 112, 66, 76, 88, 40, 14, 20, 15, 35, 44, 66, 75, 84, 95, 96, 102, 110, 88, 74, 112, 14, 34, 44.

Sol. Here the maximum and minimum values of the variate are 112 and 14 respectively.

$$\therefore \text{Range} = 112 - 14 = 98.$$

It is given that the class size is 10, and

$$\frac{\text{Range}}{\text{Class size}} = \frac{98}{10} = 9.8$$

So, we should have 10 classes each of size 10. If we take the first class as 14-24 it includes the minimum value 14. If the last class is taken as 104-114, then it includes the maximum value 112. In the class 14-24, 14 is included but 24 is excluded. Similarly, in other classes, the lower limit is included and the upper limit is excluded. Now

Bill (in rupees)	Tally marks	Frequency
14-24		4
24-34		2
34-44		3
44-54		3
54-64		1
64-74		2
74-84		5
84-94		3
94-104		3
104-114		4
Total		30

Ex.4 The marks obtained by 40 students of class IX in an examination are given below :

18, 8, 12, 6, 8, 16, 12, 5, 23, 2, 16, 23, 2, 10, 20, 12, 9, 7, 6, 5, 3, 5, 13, 21, 13, 15, 20, 24, 1, 7, 21, 16, 13, 18, 23, 7, 3, 18, 17, 16.

Present the data in the form of a frequency distribution using the same class size, one such class being 15-20 (where 20 is not included)

Sol. Frequency Distribution of Marks

Marks	Tally marks	Frequency
0-5		6
5-10		10
10-15		8
15-20		8
20-25		8
	Total	40

Ex.5 The class marks of a distribution are :

47, 52, 57, 62, 67, 72, 77, 82, 87, 92, 97, 102
Determine the class size, the class limits and the true class limits.

Sol. Here the class marks are uniformly spaced. So, the class size is the difference between any two consecutive class marks

$$\therefore \text{Class size} = 52 - 47 = 5$$

We know that, if a is the class mark of a class interval and h is its class size, then the lower and upper limits of the class interval are

$$a - \frac{h}{2} \text{ and } a + \frac{h}{2} \text{ respectively.}$$

\therefore Lower limit of first class interval

$$= 47 - \frac{5}{2}$$

$$= 44.5$$

And, upper limit of first class interval

$$= 47 + \frac{5}{2}$$

$$= 49.5$$

So, first class interval is 44.5 – 49.5

Similarly, we obtain the other class limits as given under :

Class marks	Class limits
47	44.5 - 49.5
52	49.5 - 54.5
57	54.5 - 59.5
62	59.5 - 64.5
67	64.5 - 69.5
72	69.5 - 74.5
77	74.5 - 79.5
82	79.5 - 84.5
87	84.5 - 89.5
92	89.5 - 94.5
97	94.5 - 99.5
102	99.5 - 104.5

Since the classes are exclusive, so the true class limits are same as the class limits.

Ex.6 The class marks of a distribution are 26, 31, 36, 41, 46, 51, 56, 61, 66, 71. Find the true class limits.

Sol. Here the class marks are uniformly spaced. So, the class size is the difference between any two consecutive class marks.

$$\therefore \text{Class size} = 31 - 26 = 5.$$

If a is the class mark of a class interval of size h , then the lower and upper limits of the class interval are $a - \frac{h}{2}$ and $a + \frac{h}{2}$ respectively.

Here $h = 5$

\therefore Lower limit of first class interval

$$= 26 - \frac{5}{2} = 23.5$$

And, upper limit of first class interval

$$= 26 + \frac{5}{2} = 28.5$$

\therefore First class interval is 23.5 – 28.5.

Thus, the class intervals are:

23.5 – 28.5, 28.5 – 33.5, 33.5 – 38.5, 38.5 – 43.5, 43.5 – 48.5, 48.5 – 53.5

Since the class are formed by exclusive method. Therefore, these limits are true class limits.

CUMULATIVE FREQUENCY TABLE

A table which displays the manner in which cumulative frequencies are distributed over various classes is called a cumulative frequency distribution or cumulative frequency table.

Ex.7 The marks obtained by 35 students in a class are given below. Construct the cumulative frequency table :

Marks obtained	Number of students
0	1
1	2
2	4
3	4
4	3
5	5
6	4
7	6
8	3
9	2
10	1

Sol.

Marks	Frequency	Number of students
0	1	1
1	2	3 (=1 + 2)
2	4	7 (=1 + 2 + 4)
3	4	11 (=1 + 2 + 4 + 4)
4	3	14 (=1 + 2 + 4 + 4 + 3)
5	5	19 (=1 + 2 + 4 + 4 + 3 + 5)
6	4	23 (=1 + 2 + 4 + 4 + 3 + 5 + 4)
7	6	29 (=1 + 2 + 4 + 4 + 3 + 5 + 4 + 6)
8	3	32 (=1 + 2 + 4 + 4 + 3 + 5 + 4 + 6 + 3)
9	2	34 (=1 + 2 + 4 + 4 + 3 + 5 + 4 + 6 + 3 + 2)
10	1	35 (=1 + 2 + 4 + 4 + 3 + 5 + 4 + 6 + 3 + 2 + 1)
Total = 35		

Ex.8 The distribution of ages (in years) of 40 persons in a colony is given below.

Age (in years)	Number of Persons
20-25	7
25-30	10
30-35	8
35-40	6
40-45	4
45-50	5

- Determine the class mark of each class
- What is the upper class limit of 4th class
- Determine the class size

Sol.(a) Class marks are

$$\frac{20+25}{2}, \frac{25+30}{2}, \frac{30+35}{2}, \frac{35+40}{2}, \frac{40+45}{2}, \frac{45+50}{2}.$$

$$= 22.5, 27.5, 32.5, 37.5, 42.5, 47.5$$

(b) The fourth class interval is 35–40. Its upper limit is 40

(c) The class size is $25 - 20 = 5$

Ex.9 Following is the distribution of marks of 40 students in a class. Construct a cumulative frequency distribution table.

Marks	Number of students
0-10	3
10-20	8
20-30	9
30-40	15
40-50	5

Sol.

Class interval	Frequency	Cumulative Frequency
0–10	3	3
10–20	8	11 (= 3 + 8)
20–30	9	20 (= 3 + 8 + 9)
30–40	15	35 (= 3 + 8 + 9 + 15)
40–50	5	40 (= 3 + 8 + 9 + 15 + 5)
Total = 40		

Ex.10 The class marks of a distribution are 25, 35, 45, 55, 65 and 75.

Determine the class size and class limit.

Sol. Class size = The difference between the class marks of two adjacent classes.

$$= 35 - 25$$

$$= 10$$

We need classes of size 10 with class marks as 25, 35, 45, 55, 65, 75

The class limits for the first class are

$$25 - \frac{10}{2} \text{ and } 25 + \frac{10}{2}$$

i.e. 20 and 30

First class is, therefore, 20–30

Similarly, the other classes are 30 – 40, 40 – 50, 50 – 60, 60 – 70, 70 – 80

Ex.11 Given below is the cumulative frequency distribution table showing the marks secured by 40 students.

Marks	Number of students
Below 20	5
Below 40	10
Below 60	25
Below 80	32
Below 100	40

Show in the class and their frequency form.

Sol.

Marks	Cumulative frequency	Frequency
0-20	5	5
20-40	10	5 (= 10 – 5)
40-60	25	15 (= 25 – 10)
60-80	32	7 (= 32 – 25)
80-100	40	8 (= 40 – 32)

◆ Graphical representation of data

The graphical representations such as **bar graphs, histograms, frequency polygons, etc.**

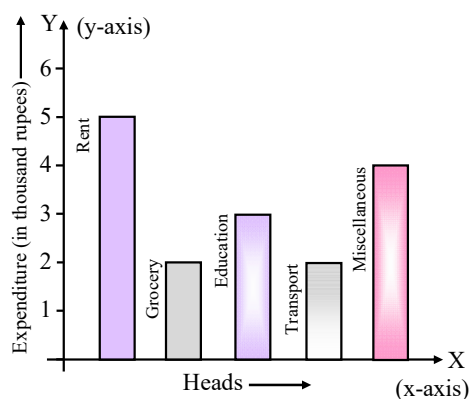
◆ Bar graph (diagram) of a data

A bar graph (diagram) is a pictorial representation of the data by a series of bars or rectangles of uniform width standing on the same horizontal (or vertical) base line with equal spacing between the bars. Each rectangle or bar represents only one numerical value of the data. The height (or length in case the base is on a vertical line) of each bar is proportional to the numerical values of the data.

For example, we are given a data about the household expenditure of a family as below :

Heads of expenditure	Expenditure (in thousand rupees)
Rent	5
Grocery	2
Education	3
Transport	2
Miscellaneous	4

We draw the bar graph for the above data as below:



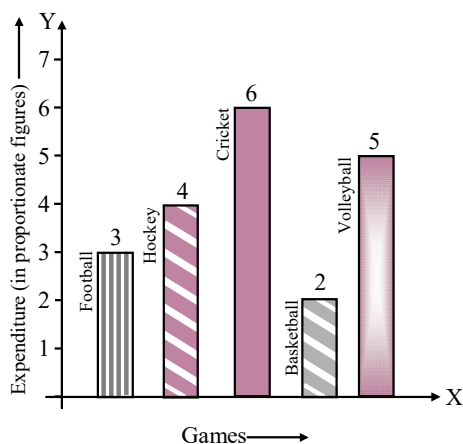
The horizontal axis is generally called x-axis and the vertical axis as the y-axis.

1. Each bar of a bar diagram has same width.
2. Space between two consecutive bars is same throughout.

❖ EXAMPLES ❖

Ex.12 In figure, the bar diagram presents the expenditure (in proportionate figures) on five different sports. If the total expenditure incurred on all the sports in a particular year be ₹ 2,00,000, then find the amount spent

- (i) on hockey
- (ii) on cricket



Sol. Proportionate amount spent on Football, Hockey, Cricket, Basketball, and volleyball are in the ratio 3 : 4 : 6 : 2 : 5 respectively. Total amount spent = ₹ 2,00,000.

Now, $3 + 4 + 6 + 2 + 5 = 20$.

Amount spent on hockey

$$= ₹ 2,00,000 \times \frac{4}{20}$$

$$= ₹ 40,000$$

Amount spent on cricket

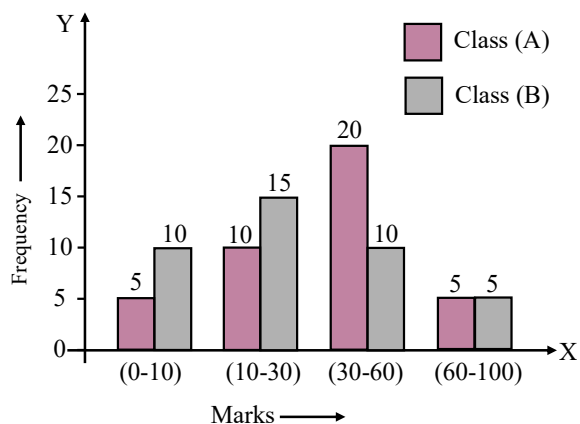
$$= ₹ 2,00,000 \times \frac{6}{20}$$

$$= ₹ 60,000$$

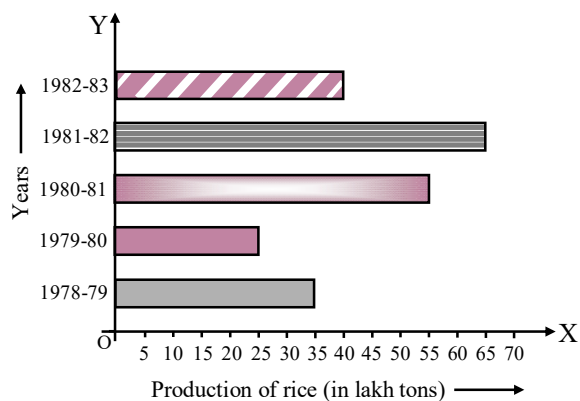
Ex.13 Compare the academic standard of two classes A and B each of 40 students on the basis of the following data by making bar diagram.

Marks :	0-10	10-30	30-60	60-100
No. of students in A :	5	10	20	5
No. of students in B :	10	15	10	5

Sol.



Ex.14 Read the bar graph in figure and answer the following :



- (i) What information is given by the bar graph ?

- (ii) What was the quantity of rice production in the year 1980-81.
- (iii) What is the difference between the maximum and minimum production of rice in the time span of 1978-1983.

Sol. (i) The bar graph, represents production of rice in the period 1978-1983 (year-wise).

- (ii) The bar for 1980-81 has length = 55.

Therefore, the production of rice in the year 1980-81 is 55 lakh tons.

- (iii) In the year 1981-82, the production of rice is maximum and is equal to 65 lakh tons. In the year 1979-80, the production of rice is minimum and is equal to 25 lakh tons.

The difference between the maximum and the minimum production

$$= 65 \text{ lakh tons} - 25 \text{ lakh tons}$$

$$= 40 \text{ lakh tons}$$

➤ HISTOGRAM

A histogram is a graphical representation of frequency distribution in the form of rectangle with class interval as the bases and heights proportional to corresponding frequency. There is no gap between rectangles

◆ Construction of a histogram of a continuous grouped frequency distribution

- (i) Take a graph paper and draw two perpendicular lines, one horizontal and one vertical, intersecting at O (say). Mark them as OX and OY.
- (ii) Take horizontal line OX as X-axis and vertical line OY as Y-axis
- (iii) Choose a suitable scale and along X-axis represent class limits
- (iv) Choose a suitable scale for Y-axis and mark frequencies along Y-axis.
- (v) Construct rectangles with class intervals as bases and respective frequencies as heights.

◆ Construction of a histogram of a continuous grouped frequency distribution with unequal class intervals.

- (i) Take a graph paper and draw two perpendicular lines, one horizontal and one vertical, intersecting at O (say). Mark them as OX and OY.

- (ii) Take horizontal line OX as X-axis and vertical line OY as Y-axis.
- (iii) Choose a suitable scale along X-axis and represent class-limits on it.

- (iv) Determine a class-interval which has the minimum class size. Let the minimum class size be h.

- (v) Compute the adjusted frequencies of each class by using the following formula:

Adjusted Frequency of a class

$$= \frac{h}{\text{Class size}} \times \text{Frequency of the class}$$

These adjusted frequencies are the heights of each rectangle of histogram but widths will be according to class limits.

- (vi) Choose a suitable scale for Y-axis and mark adjusted frequencies along Y-axis.
- (vii) Construct rectangles with class intervals as bases and respective adjusted frequencies as heights.

◆ Construction of a histogram when mid-points of class-intervals are given

- (i) Compute the difference between second and first mid-point. Let it be h.
- (ii) Divide the difference h by 2.
- (iii) Subtract $\frac{h}{2}$ from first mid-point to get the lower limit of first class and add $\frac{h}{2}$ to first mid-point to get the upper limit of first class.
- (iv) Repeat first three steps for all other mid-points.

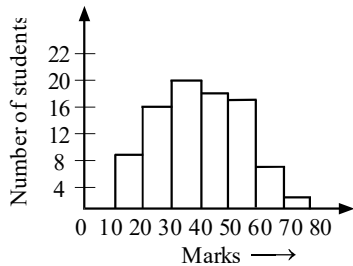
◆ EXAMPLES ◆

Ex.15 The following table gives the marks scored by 100 students in an entrance examination.

Marks	No. of students (Frequency)
0 - 10	4
10 - 20	10
20 - 30	16
30 - 40	22
40 - 50	20
50 - 60	18
60 - 70	8
70 - 80	2

Represent this data in the form of a histogram.

Sol. We represent the class limits along X-axis on a suitable scale and the frequencies along Y-axis on a suitable scale.



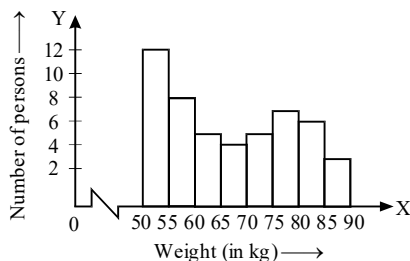
Ex.16 The following is the distribution of weights (in kg) of 50 persons :

weight (in kg)	No. of persons
50-55	12
55-60	8
60-65	5
65-70	4
70-75	5
75-80	7
80-85	6
85-90	3

Draw a histogram for the above data.

Sol. We represent the class limits along X-axis on a suitable scale and the frequencies along Y-axis on a suitable scale.

Since the scale on X-axis starts at 50, a kink (break) is indicated near the origin to signify that the graph is drawn to scale beginning at 50, and not at the origin.



Ex.17 Represent the following data by means of histogram.

Weekly wages (in Rs):	No. of workers (Frequency):
10-15	7
15-20	9
20-25	8
25-30	5
30-40	12
40-60	12
60-80	8

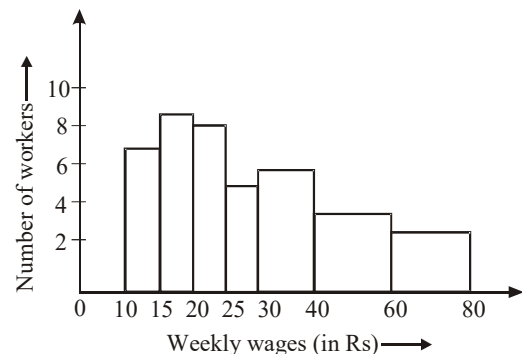
Sol. Here the class intervals are of unequal width. So, we shall first compute adjusted frequencies of each class. The minimum class size is $15 - 10 = 5$. The adjusted frequencies are computed by using the following formula :

Adjusted frequency of a class

$$= \frac{\text{Minimum class-size}}{\text{Class-size}} \times \text{Frequency of the class}$$

The adjusted frequencies are computed in the following table :

Weekly wages (in Rs):	Frequency	Adjusted frequency
10 - 15	7	$\frac{5}{5} \times 7 = 7$
15 - 20	9	$\frac{5}{5} \times 9 = 9$
20 - 25	8	$\frac{5}{5} \times 8 = 8$
25 - 30	5	$\frac{5}{5} \times 5 = 5$
30 - 45	12	$\frac{5}{10} \times 12 = 6$
40 - 60	12	$\frac{5}{20} \times 12 = 3$
60 - 80	8	$\frac{5}{20} \times 8 = 2$



Ex.18 Draw a histogram for the marks of students given below:

Marks :	No. of students :
0-10	8
10-30	32
30-45	18
45-50	10
50-60	6

Sol. Here, minimum class-size = 5.

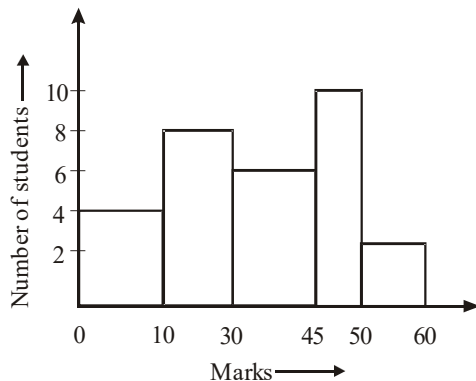
So, we adjust the frequencies by using the following formula:

Adjusted frequency of a class

$$= \frac{\text{Minimum class-size}}{\text{Class-size}} \times \text{Frequency of the class}$$

The adjusted frequencies are computed as follows :

Class intervals (Marks)	No. of students (Frequency)	Adjusted Frequency
0 - 10	8	$\frac{5}{10} \times 8 = 4$
10 - 30	32	$\frac{5}{20} \times 32 = 8$
30 - 45	18	$\frac{5}{15} \times 18 = 6$
45 - 50	10	$\frac{5}{5} \times 10 = 10$
50 - 60	6	$\frac{5}{10} \times 6 = 3$



Ex.19 Construct a histogram from the following distribution of total marks obtained by 65 students of IX class in the final examination.

Marks (mid - points) :	No. of students :
150	8
160	10
170	25
180	12
190	7
200	3

Sol. Ascertainment of lower and upper class limits
Since the difference between the second and first mid-point is $160 - 150 = 10$.

$$\therefore h = 10 \Rightarrow \frac{h}{2} = 5.$$

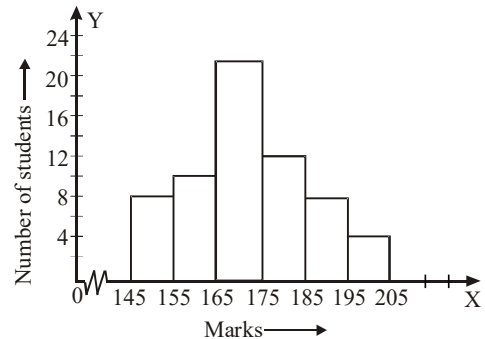
So, lower and upper limits of the first class are $150 - 5$ and $150 + 5$ i.e. 145 and 155 respectively.

\therefore First class interval is 145 – 155.

Using the same procedure, we get the classes of other mid-points as under :

Marks :	No. of students :
145 - 155	8
155 - 165	10
165 - 175	25
175 - 185	12
185 - 195	7
195 - 205	3

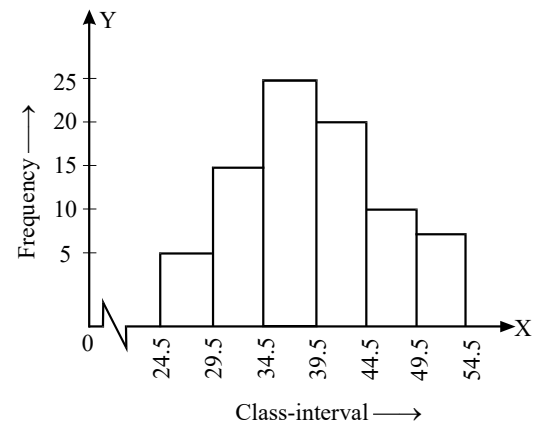
The histogram of the above frequency distribution is given in Fig.



Ex.20 Draw a histogram for the following data:

Class :	Frequency :
25 - 29	5
30 - 34	15
35 - 39	23
40 - 44	20
45 - 49	10
50 - 54	7

The histogram of the above frequency distribution is given in Fig.

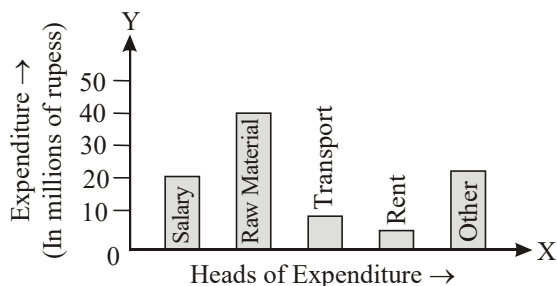


Ex.21 The expenditure of a company under different heads (in millions of rupees) is given below.

Head	Expenditure
salary	20
Raw material	40
Transport	10
Rent	5
Others	15

Draw a bar chart to depict the above data.

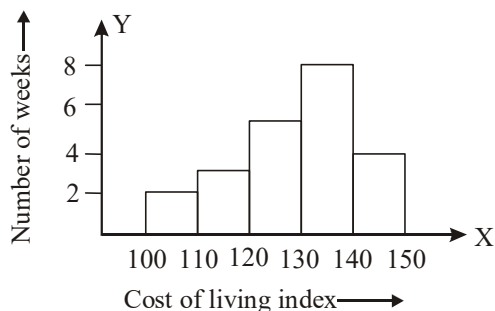
Sol. The required bar chart is given below :



Ex.22 Draw a histogram for the data given below.

Cost of living index	number of weeks
100-110	2
110-120	3
120-130	5
130-140	8
140-150	4

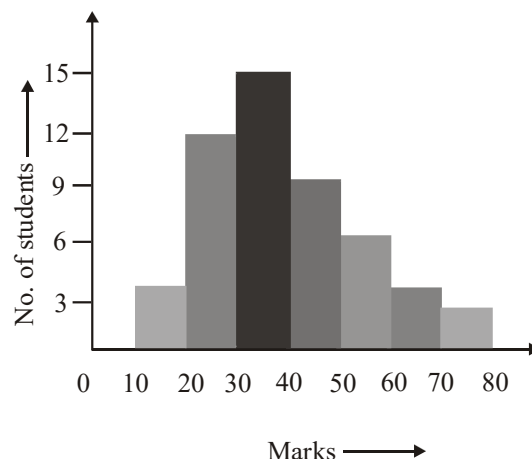
Sol. The required histogram is given below



Ex.23 Draw a histogram to represent the following data of marks obtained by 50 students of a class.

Marks	No. of students
10-20	4
20-30	12
30-40	15
40-50	8
50-60	5
60-70	4
70-80	2

Sol. The class intervals showing the marks are taken along the X-axis and the number of students along the Y-axis. The histogram is shown below.



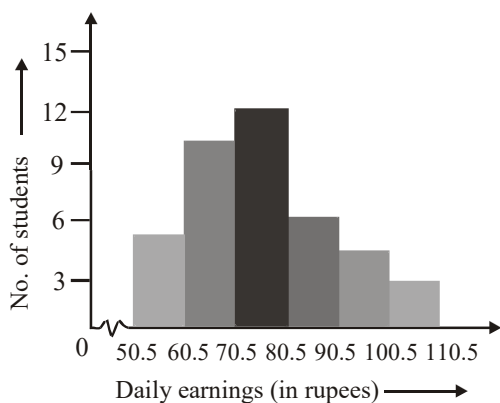
Ex.24 The daily earnings of 40 workers of a factory are given below :

Daily earnings (in rupees)	No. of workers
51-60	5
61-70	10
71-80	12
81-90	6
91-100	4
101-110	3

Sol. The given distribution is inclusive (discontinuous). So we are to change it to a continuous distribution. As the gap between the upper limit of the first class and the lower limit of the second class is 1, the true class limits of the first class are 50.5 and 60.5. The class boundaries (true class limits) of the other classes are similarly determined. The continuous distribution table is then shown below.

Daily earnings (in rupees)	No. of workers
50.5 – 60.5	5
60.5 – 70.5	10
70.5 – 80.5	12
80.5 – 90.5	6
90.5 – 100.5	4
100.5 – 110.5	3

The histogram is shown below.



The earnings are taken along the horizontal axis and the number of workers along the vertical axis. As we start from 50.5 (not zero according to scale) a kink (break) is indicated near the origin to show that the graph is drawn to scale, starting from 50.5.

➤ FREQUENCY POLYGON

A frequency polygon is the polygon obtained by joining the mid-points of upper horizontal sides of all the rectangles in a histogram.

◆ Construction of a frequency polygon with Histogram.

- Obtain the frequency distribution and draw a histogram representing it.
- Obtain the mid-points of the upper horizontal side of each rectangle.
- Join these mid-points of the adjacent rectangles of the histogram by dotted line segments.
- Obtain the mid-points of two class-intervals of zero frequency i.e. on X-axis, one adjacent to the first, on its left and one adjacent to the last, on its right.

These class-intervals are known as imagined class intervals.

- Complete the polygon by joining the mid-points of first and last class intervals to the mid-points of imagined class-intervals adjacent to them.

◆ Construction of a frequency polygon without using a histogram

- Obtain the frequency distribution.
- Compute the mid-points of class intervals i.e. class marks.
- Represent class marks on X-axis on a suitable scale.

- Represent frequencies on Y-axis on a suitable scale.
- Plot the points (x_i, f_i) where x_i denotes class mark and f_i corresponding frequency.
- Join the points plotted in step V by line segments.
- Take two class intervals of zero frequency, one at the beginning and the other at the end. Obtain their mid-points.

These classes are known as imagined classes.

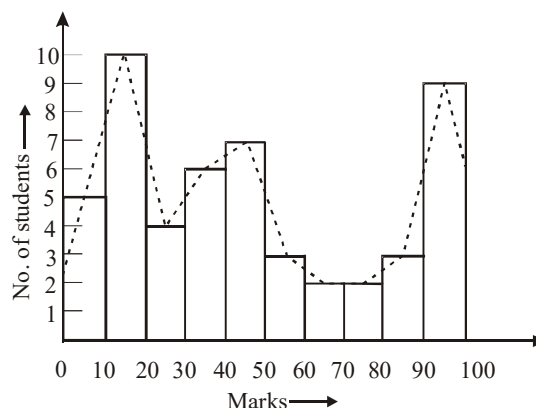
- Complete the frequency polygon by joining the mid-points of first and last class intervals to the mid-points of the imagined classes adjacent to them.

❖ EXAMPLES ❖

Ex.25 For the following data, draw a histogram and a frequency polygon :

Marks :	No. of students :
0-10	5
10-20	10
20-30	4
30-40	6
40-50	7
50-60	3
60-70	2
70-80	2
80-90	3
90-100	9

Sol. In Fig. a histogram and a frequency polygon (in dotted lines) are drawn on the same scale.

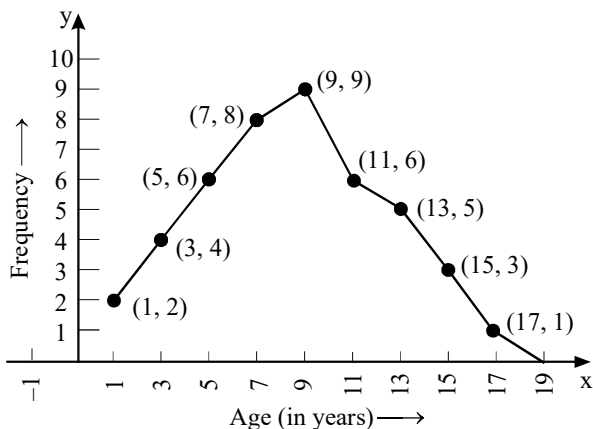


Ex.26 Construct a frequency polygon for the following data :

Age(in years) :	Frequency :
0 - 2	2
2 - 4	4
4 - 6	6
6 - 8	8
8 - 10	9
10 - 12	6
12 - 14	5
14 - 16	3
16 - 18	1

Sol. First we obtain the class marks as given in the following table.

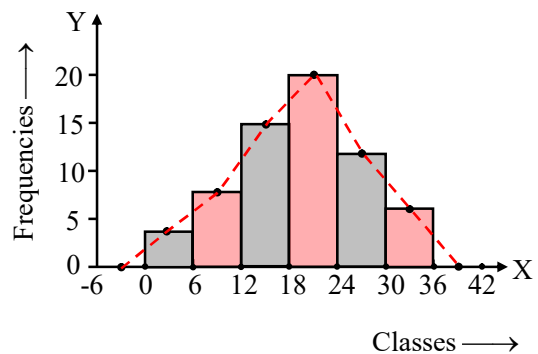
Age(in years) :	Class marks	Frequency
0 - 2	1	2
2 - 4	3	4
4 - 6	5	6
6 - 8	7	8
8 - 10	9	9
10 - 12	11	6
12 - 14	13	5
14 - 16	15	3
16 - 18	17	1



Ex.27 Prepare frequency polygon with the help of histogram from the following data :

Classes	0-6	6-12	12-18	18-24	24-30	30-36
Frequency	4	8	15	20	12	6

Sol. We plot the classes (-6-0), (0-6), (6-12), (12-18), (18-24), (24-30), (30-36) and (36-42) along the x-axis. The frequencies of these classes are respectively, 0, 4, 8, 15, 20, 12, 6 and 0. These frequencies are plotted along the y-axis.



After drawing the histogram for the data, we join the mid-points of the top sides of the rectangles of the histogram. The frequency polygon is made with dotted line segments.

Ex.28 Draw frequency polygon for the data given below, without drawing histogram :

Classes	140-150	150-160	160-170	170-180	180-190	190-200
Frequencies	5	10	20	9	6	2

Sol. We have the classes

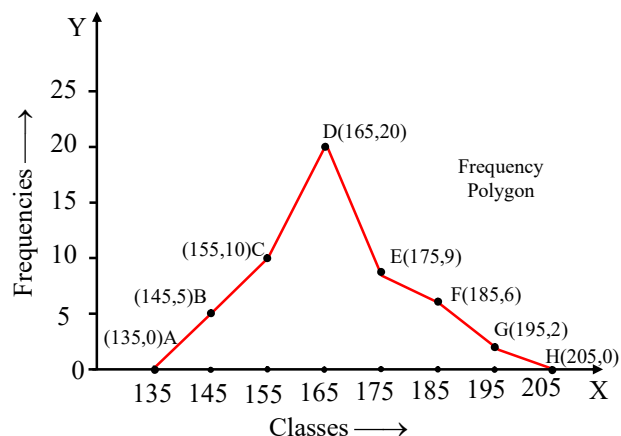
130-140, 140-150, 150-160, 160-170, 170-180, 180-190, 190-200, 200-210 having frequencies 0, 5, 10, 20, 9, 6, 2, 0 respectively.

The class marks of the classes are

$$\frac{130+140}{2} = 135, \frac{140+150}{2} = 145, \dots, \frac{200+210}{2} = 205 \text{ respectively.}$$

We take class marks along x-axis and the frequencies along y-axis.

We plot the points (135, 0), (145, 5), (155, 10), (165, 20), (175, 9), (185, 6), (195, 2) and (205, 0). Now, we join these points and get the required frequency polygon ABCDEFGH as in figure.



Ex.29 Ages (in years) of the members of two sports clubs were recorded and the data collected is as under.

Age (in years)	Number of members of Club A	Number of members of Club B
20-29	5	10
30-39	10	15
40-49	15	10
50-59	10	5
60-69	20	10
70-79	5	15
Total	65	65

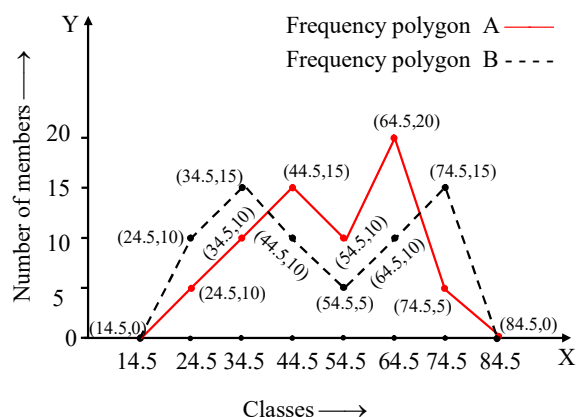
Represent the ages of the members of the sports clubs A and B on the same graph by two frequency polygons. Compare the pattern of the two clubs according to ages of the members.

Sol. The class marks of the classes

20-29, 30-39, 40-49, 50-59, 60-69, 70-79 are 24.5, 34.5, 44.5, 54.5, 64.5, 74.5 respectively. The corresponding frequencies for clubs A and B are 5, 10, 15, 10, 20, 5 ; 10, 15, 10, 5, 10, 15 respectively.

For the club A, the frequency polygon is drawn by joining the points (14.5, 0), (24.5, 5), (34.5, 10), (44.5, 15), (54.5, 10), (64.5, 20), (74.5, 5), (84.5, 0).

For the club B, the frequency polygon is drawn by joining the points (14.5, 0), (24.5, 10), (34.5, 15), (44.5, 10), (54.5, 5), (64.5, 10), (74.5, 15), (84.5, 0).



We observe that the club B is more youthful in comparison to club A.

Ex.30 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

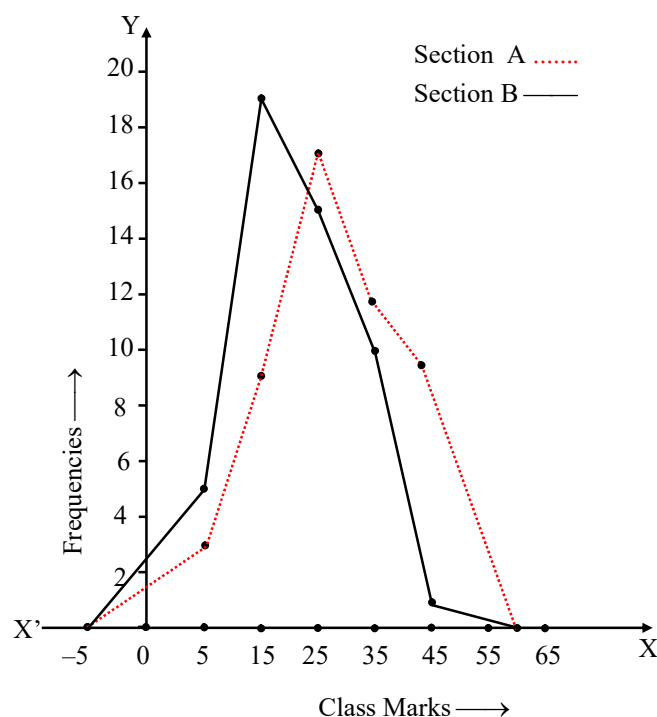
Represent the marks of the students of both the sections on the same graph by two frequency polygons. From the two polygons compare the performance of the two section.

Sol. Class marks for section A are : 5, 15, 25, 35, 45 and corresponding frequencies as 3, 9, 17, 12, 9 respectively.

Its frequency polygon is the join of the points (by line segments) (-5, 0), (5, 3), (15, 9), (25, 17), (35, 12), (45, 9) and (60, 0).

Similarly for the section B, the frequency polygon is the join of the points (-5, 0), (5, 5), (15, 19), (25, 15), (35, 10), (45, 1) and (60, 0).

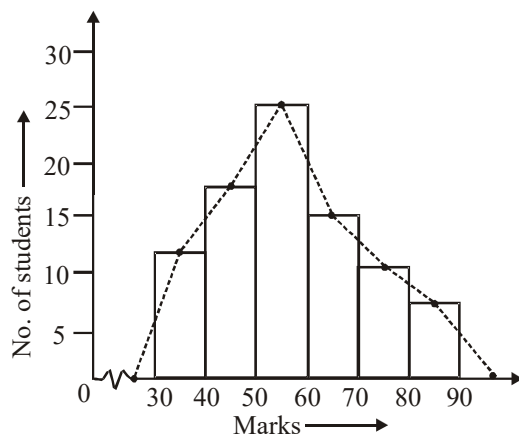
We draw the two frequency polygon :



Ex.31 Make a histogram and a frequency polygon from the given data :

Marks	No. of students
30-40	12
40-50	18
50-60	26
60-70	16
70-80	10
80-90	6

Sol. The frequency polygon has been shown by dotted lines



➤ MEAN

If $x_1, x_2, x_3, \dots, x_n$ are n values of a variable X , then the arithmetic mean or simply the mean of these values is denoted by \bar{X} and is defined as

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{1}{n} \left(\sum_{i=1}^n x_i \right)$$

Here the symbol $\sum_{i=1}^n x_i$ denotes the sum $x_1 + x_2 + x_3 + \dots + x_n$.

- ◆ If \bar{X} is the mean of n observations x_1, x_2, \dots, x_n , then the mean of the observations $x_1 + a, x_2 + a, \dots, x_n + a$ is $\bar{X} + a$ i.e. if each observation is increased by a , then the mean is also increased by a
- ◆ If \bar{X} is the mean of x_1, x_2, \dots, x_n then the mean of ax_1, ax_2, \dots, ax_n is $a\bar{X}$, where a is any number different from zero i.e. if each observation is multiplied by a non-zero number a , then the mean is also multiplied by a .

◆ If \bar{X} is the mean of n observations $x_1, x_2, x_3, \dots, x_n$, then the mean of $\frac{x_1}{a}, \frac{x_2}{a}, \frac{x_3}{a}, \dots, \frac{x_n}{a}$ is $\frac{\bar{X}}{a}$, where a is any non-zero number

◆ If \bar{X} is the mean of n observations x_1, x_2, \dots, x_n , then the mean of $x_1 - a, x_2 - a, \dots, x_n - a$ is $\bar{X} - a$, where a is any real number

◆ Advantages

- (i) Arithmetic mean is simple to understand and easy to calculate.
- (ii) It is rigidly defined.
- (iii) It is suitable for further algebraic treatment.
- (iv) It is least affected fluctuation of sampling.
- (v) It takes into account all the values in the series.

◆ Disadvantages

- (i) It is highly affected by the presence of a few abnormally high or abnormally low scores.
- (ii) In absence of a single item, its value becomes inaccurate.
- (iii) It can not be determined by inspection

❖ EXAMPLES ❖

Ex.32 If the mean of n observations $ax_1, ax_2, ax_3, \dots, ax_n$ is $a\bar{X}$, show that

$$(ax_1 - a\bar{X}) + (ax_2 - a\bar{X}) + \dots + (ax_n - a\bar{X}) = 0$$

Sol. We have

$$a\bar{X} = \frac{ax_1 + ax_2 + \dots + ax_n}{n}$$

$$\Rightarrow ax_1 + ax_2 + \dots + ax_n = n(a\bar{X}) \quad \dots(i)$$

$$\begin{aligned} \text{Now, } (ax_1 - a\bar{X}) + (ax_2 - a\bar{X}) + \dots + (ax_n - a\bar{X}) \\ = (ax_1 + ax_2 + \dots + ax_n) - (\underbrace{a\bar{X} + a\bar{X} + \dots + a\bar{X}}_{n \text{ terms}}) \\ = n(a\bar{X}) - n(a\bar{X}) = 0. \end{aligned}$$

Ex.33 The mean of 40 observations was 160. It was detected on rechecking that the value of 165 was wrongly copied as 125 for computation of mean. Find the correct mean.

Sol. Θ Here, $n = 40$, $\bar{X} = 160$

$$\text{So, } \bar{X} = \frac{1}{n}(\sum x_i) \Rightarrow 160 = \frac{1}{40}(\sum x_i)$$

$$\Rightarrow \sum x_i = 160 \times 40 = 6400$$

$$\Rightarrow \text{Incorrect value of } \sum x_i = 6400$$

Now,

$$\text{Correct value of } \sum x_i$$

$$= \text{Incorrect value of } \sum x_i - \text{Incorrect item} \\ + \text{Correct item}$$

$$\Rightarrow \text{Correct value of } \sum x_i = 6400 - 125 \\ + 165 = 6440$$

\therefore Correct mean

$$= \frac{\text{Correct value of } \sum x_i}{n} = \frac{6440}{40} = 161.$$

Ex.34 The mean of 10 numbers is 20. If 5 is subtracted from every number, what will be the new mean?

Sol. Let x_1, x_2, \dots, x_{10} be 10 numbers with their mean equal to 20. Then,

$$\bar{X} = \frac{1}{n}(\sum x_i)$$

$$\Rightarrow 20 = \frac{x_1 + x_2 + \dots + x_{10}}{10}$$

$$\Rightarrow x_1 + x_2 + \dots + x_{10} = 200 \quad \dots(i)$$

New numbers are $x_1 - 5, x_2 - 5, \dots, x_{10} - 5$.

Let \bar{X}' be the mean of new numbers.

Then,

$$\bar{X}' = \frac{(x_1 - 5) + (x_2 - 5) + \dots + (x_{10} - 5)}{10}$$

$$\bar{X}' = \frac{(x_1 + x_2 + \dots + x_{10}) - 5 \times 10}{10} = \frac{200 - 50}{10}$$

[Using (i)]

$$\bar{X}' = 15.$$

❖ EXAMPLES ❖

Ex.35 Neeta and her four friends secured 65, 78, 82, 94 and 71 marks in a test of mathematics.

Find the average (arithmetic mean) of their marks.

Sol. Arithmetic mean or average

$$= \frac{65 + 78 + 82 + 94 + 71}{5}$$

$$= \frac{390}{5} = 78$$

Hence, arithmetic mean = 78

Ex.36 The mean of 5, 7, p, 11, 15, 17, and 20 is 12, find p.

$$\text{Sol. Mean} = \frac{5 + 7 + p + 11 + 15 + 17 + 20}{7}$$

$$\Rightarrow 12 = \frac{75 + p}{7}$$

$$\Rightarrow p + 75 = 12 \times 7$$

$$\Rightarrow p = 9$$

Ex.37 If the mean of 5 observations is 15 and that of another 10 observations is 20, find the mean of all 15 observations

Sol. Let first five observations be x_1, \dots, x_5

$$\Rightarrow \text{Mean} = \frac{x_1 + x_2 + \dots + x_5}{5}$$

$$\Rightarrow 15 = \frac{x_1 + x_2 + \dots + x_5}{5}$$

$$\Rightarrow x_1 + \dots + x_5 = 75 \quad \dots(i)$$

Let next ten observations be $y_1 + \dots + y_{10}$.

$$\Rightarrow \text{Mean} = \frac{y_1 + \dots + y_{10}}{10}$$

$$\Rightarrow 20 = \frac{y_1 + \dots + y_{10}}{10}$$

$$y_1 + \dots + y_{10} = 200 \quad \dots(ii)$$

The mean of all 15 observations will be

$$\frac{(x_1 + \dots + x_5) + (y_1 + \dots + y_{10})}{15}$$

$$= \frac{75 + 200}{15} \quad (\text{from (i) and (ii)})$$

$$= 18.33$$

If a variate X takes values x_1, x_2, \dots, x_n with corresponding frequencies $f_1, f_2, f_3, \dots, f_n$ respectively, then arithmetic mean of these values is

$$\bar{X} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_n x_n}{f_1 + f_2 + \dots + f_n}$$

$$\text{or } \bar{X} = \frac{\sum_{i=1}^n f_i x_i}{N}, \text{ where } N = \sum_{i=1}^n f_i = f_1 + f_2 + \dots + f_n$$

Ex.38 Find the mean of the following distribution :

$x :$ 4 6 9 10 15

$f :$ 5 10 10 7 8

Sol. Calculation of Arithmetic Mean

x_i	f_i	$f_i x_i$
4	5	20
6	10	60
9	10	90
10	7	70
15	8	120
	$N = \Sigma f_i = 40$	$\Sigma f_i x_i = 360$

$$\therefore \text{Mean} = \bar{X} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{360}{40} = 9.$$

Ex.39 Find the mean of the following distribution :

x	10	30	50	70	89
f	7	8	10	15	10

Sol. Calculation of Mean

x_i	f_i	$f_i x_i$
10	7	70
30	8	240
50	10	500
70	15	1050
89	10	890
	$\Sigma f_i = N = 50$	$\Sigma f_i x_i = 2750$

$$\therefore \text{Mean} = \frac{\Sigma f_i x_i}{N} = \frac{2750}{50} = 55.$$

Ex.40 Find the value of p , if the mean of following distribution is 7.5.

$x :$	3	5	7	9	11	13
$y :$	6	8	15	P	8	4

Sol.

x_i	f_i	$f_i x_i$
3	6	18
5	8	40
7	15	105
9	p	$9p$
11	8	88
13	4	52
	$N = \Sigma f_i = 41 + p$	$\Sigma f_i x_i = 303 + 9p$

We have, $\Sigma f_i = 41 + p$, $\Sigma f_i x_i = 303 + 9p$

$$\therefore \text{Mean} = \frac{\Sigma f_i x_i}{\Sigma f_i} \Rightarrow 7.5 = \frac{303 + 9p}{41 + p}$$

$$\Rightarrow 7.5 \times (41 + p) = 303 + 9p$$

$$\Rightarrow 307.5 + 7.5p = 303 + 9p$$

$$\Rightarrow 9p - 7.5p = 307.5 - 303$$

$$\Rightarrow 1.5p = 4.5 \Rightarrow p = 3$$

Ex.41 Find the missing frequencies in the following frequency distribution if it is known that the mean of the distribution is 1.46.

Number of accidents (x) :	0	1	2	3	4	5	Total
Frequency (f) :	46	?	?	25	10	5	200

Sol. Let the missing frequencies be f_1 and f_2

x_i	f_i	$f_i x_i$
0	46	0
1	f_1	f_1
2	f_2	$2f_2$
3	25	75
4	10	40
5	5	25
	$N = 86 + f_1 + f_2$	$\Sigma f_i x_i = 140 + f_1 + 2f_2$

We have : $N = 200$

$$\therefore 200 = 86 + f_1 + f_2 \Rightarrow f_1 + f_2 = 114 \dots(i)$$

Also,

$$\text{Mean} = 1.46$$

$$\Rightarrow 1.46 = \frac{\sum f_i x_i}{N} \Rightarrow 1.46 = \frac{140 + f_2 + 2f_2}{200}$$

$$\Rightarrow 292 = 140 + f_1 + 2f_2$$

$$\Rightarrow f_1 + 2f_2 = 152 \quad \dots(ii)$$

Solving (i) and (ii) we get $f_1 = 76$ and $f_2 = 38$.

Ex.42 If the mean of the following data be 9.2, find the value of p.

x	4	6	7	p+4	12	12
f	5	6	4	10	8	7

Sol.

x	f	f.x
4	5	20
6	6	36
7	4	28
p+4	10	10p + 40
12	8	96
14	7	98
Total	40	318 + 10p

Now,

$$\text{Mean } \bar{x} = \frac{\sum f.x}{\sum f} = \frac{318 + 10p}{40}$$

$$\therefore 9.2 = \frac{318 + 10p}{40}$$

$$\Rightarrow 318 + 10p = 368 \Rightarrow 10p = 50 \Rightarrow p = 5$$

Ex.43 The marks of 30 students are given below, find the mean marks.

Marks	Number of Students
10	4
11	3
12	8
13	6
14	7
15	2

Sol.

x	f	fx
10	4	40
11	3	33
12	8	96
13	6	78
14	7	98
15	2	30
	$\Sigma f = 30$	$\Sigma fx = 375$

$$\begin{aligned} \text{Mean} &= \frac{\sum fx}{\sum f} = \frac{375}{30} \\ &= 12.5 \end{aligned}$$



GROUPED FREQUENCY DISTRIBUTION

There are 3 methods for calculation of mean :

1. Direct Method
2. Assumed mean deviation method
3. Step deviation method.



Direct Method for Calculation of Mean

mid-value	frequency	$f_i x_i$
x_1	f_1	$f_1 x_1$
x_2	f_2	$f_2 x_2$
\vdots	\vdots	\vdots
x_k	f_k	$f_k x_k$
Total	N	$\sum_{i=1}^k f_i x_i$

According to direct method

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_k f_k}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{i=1}^k x_i f_i}{\sum f_i} = \frac{1}{N} \sum_{i=1}^k f_i x_i$$

$$[N = f_1 + f_2 + \dots + f_k]$$

❖ EXAMPLES ❖

Ex.44

Mid-values	2	3	4	5	6
Frequencies	49	43	57	38	13

Find the mean by direct method.

Sol.

Mid-values	Frequencies (f_i)	$f_i x_i$
2	49	98
3	43	129
4	57	228
5	38	190
6	13	78
Total	$N = \sum f_i = 200$	$\sum f_i x_i = 723$

$$\begin{aligned} \text{Mean} &= \frac{\sum f_i x_i}{\sum f_i} = \frac{723}{200} \\ &= 3.615 \end{aligned}$$

Ex.45 Find the mean of the following frequency distribution :

Class Interval	Frequency
10 – 30	90
30 – 50	20
50 – 70	30
70 – 90	20
90 – 110	40

Sol.

Class interval	f	Mid value (x)	f x
10 – 30	90	20	1800
30 – 50	20	40	800
50 – 70	30	60	1800
70 – 90	20	80	1600
90 – 110	40	100	4000
	$\Sigma f = 200$		$\Sigma f x = 10000$

$$\text{Mean } (\bar{x}) = \frac{\Sigma f x}{\Sigma f} = \frac{10000}{200} = 50$$

Ex.46 A survey was conducted by a group of students as a part of their environment awareness programme, in which they collected the following data regarding the number of plants in 20 houses in a locality. Find the mean number of plants per house.

Number of plants	0 – 2	2 – 4	4 – 6	6 – 8	8 – 10	10 – 12	12 – 14
No. of houses	1	2	1	5	6	2	3

Which method did you use for finding the mean and why ?

Sol.

Number of plants	Number of houses (f)	Mid value x	f x
0 – 2	1	1	1
2 – 4	2	3	6
4 – 6	1	5	5
6 – 8	5	7	35
8 – 10	6	9	54
10 – 12	2	11	22
12 – 14	3	13	39
	$\Sigma f = 20$		$\Sigma f x = 162$

$$\text{Mean} = \frac{\Sigma f x}{\Sigma f} = \frac{162}{20} = 8.1$$

➤ MEDIAN

Median of a distribution is the value of the variable which divides the distribution into two equal parts i.e. it is the value of the variable such that the number of observations above it is equal to the number of observations below it.

- ◆ If the values x_i in the raw data, are arranged in order of increasing or decreasing magnitude, then the middle, most value in the arrangement is called the median.

Algorithm :

Step I : Arrange the observations (values of the variate) in ascending or descending order of magnitude.

Step II : Determine the total number of observations, say, n .

Step III : If n is odd, then

$$\text{Median} = \text{value of } \left(\frac{n+1}{2} \right)^{\text{th}} \text{ observation}$$

If n is even, then
Median

$$= \frac{\text{Value of } \left(\frac{n}{2} \right)^{\text{th}} \text{ observation} + \text{Value of } \left(\frac{n}{2} + 1 \right)^{\text{th}} \text{ observation}}{2}$$

- ◆ The median can be calculated graphically.
- ◆ Median is not affected by extreme values.

❖ EXAMPLES ❖

Ex.47 Find the median of the following data :

25, 34, 31, 23, 22, 26, 35, 28, 20, 32

Sol. Arranging the data in ascending order, we get 20, 22, 23, 25, 26, 28, 31, 32, 34, 35

Here, the number of observations $n = 10$ (even).

∴ Median

$$= \frac{\text{Value of } \left(\frac{10}{2} \right)^{\text{th}} \text{ observation} + \text{Value of } \left(\frac{10}{2} + 1 \right)^{\text{th}} \text{ observation}}{2}$$

⇒ Median

$$\begin{aligned} & \text{Value of 5}^{\text{th}} \text{ observation} + \text{value} \\ & = \frac{\text{of 6}^{\text{th}} \text{ observation}}{2} \end{aligned}$$

$$\therefore \text{Median} = \frac{26+28}{2} = 27$$

Hence, median of the given data is 27.

Ex.48 Find the median of the following values :

37, 31, 42, 43, 46, 25, 39, 45, 32

Sol. Arranging the data in ascending order, we have 25, 31, 32, 37, 39, 42, 43, 45, 46

Here, the number of observations $n = 9$ (odd)

$$\begin{aligned} \therefore \text{Median} &= \text{Value of } \left(\frac{9+1}{2} \right)^{\text{th}} \text{ observation} \\ &= \text{Value of 5th observation} = 39. \end{aligned}$$

Ex.49 The median of the observations 11, 12, 14, 18, $x + 2$, $x + 4$, 30, 32, 35, 41 arranged in ascending order is 24. Find the value of x .

Sol. Here, the number of observations $n = 10$. Since n is even, therefore

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

$$\Rightarrow 24 = \frac{5^{\text{th}} \text{ observation} + 6^{\text{th}} \text{ observation}}{2}$$

$$\Rightarrow 24 = \frac{(x+2) + (x+4)}{2}$$

$$\Rightarrow 24 = \frac{2x+6}{2}$$

$$\Rightarrow 24 = x + 3 \Rightarrow x = 21.$$

Hence, $x = 21$.

Ex.50 Find the median of the following data : 19, 25, 59, 48, 35, 31, 30, 32, 51. If 25 is replaced by 52, what will be the new median.

Sol. Arranging the given data in ascending order, we have 19, 25, 30, 31, 32, 35, 48, 51, 59

Here, the number of observations $n = 9$ (odd)

Since the number of observations is odd. Therefore.

$$\text{Median} = \text{Value of } \left(\frac{9+1}{2} \right)^{\text{th}} \text{ the observations}$$

$$\Rightarrow \text{Median} = \text{value of 5}^{\text{th}} \text{ observation} = 32.$$

Hence, Median = 32

If 25 is replaced by 52, then the new observations arranged in ascending order are :

19, 30, 31, 32, 35, 48, 51, 52, 59

$$\therefore \text{New median} = \text{Value of 5}^{\text{th}} \text{ observation} = 35.$$

Ex.51 Calculate the median for the following distribution

Weight (in kg)	Number of students
46	3
47	2
48	4
49	6
50	5
51	2
52	1

Sol. The cumulative frequency table is constructed as shown below :

Weights x_i	Number of students f_i	Cumulative frequency
46	3	3
47	2	5
48	4	9
49	6	15
50	5	20
51	2	22
52	1	23

Here, $n = 23$, which is odd

$$\text{Median} = t_{\frac{23+1}{2}} = t_{12} = 49$$

(i.e. weight of the 12th student when the weights have been arranged in order)

Ex.52 The following data have been arranged in descending orders of magnitude 75, 70, 68, $x + 2$, $x - 2$, 50, 45, 40

If the median of the data is 60, find the value of x .

Sol. The number of observations are 8, the median will be the average of 4th and 5th number

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

$$\Rightarrow 60 = \frac{2x}{2} \Rightarrow x = 60$$



MODE

Mode is also known as norm.

- Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster density.

Algorithm

Step I : Obtain the set of observations.

Step II : Prepare the frequency distribution.

Step III : Obtain the value which has the maximum frequency.

Step IV : The value obtained in step III is the mode.

Ex.53 Find the mode from the following data :

110, 120, 130, 120, 110, 140, 130, 120, 140, 120.

Sol. Arranging the data in the form of a frequency table, we have

Value	Tally bars	Frequency
110		2
120		4
130		2
140		2

Since the value 120 occurs maximum number of times i.e. 4. Hence, the modal value is 120.

Ex.54 Compute mode for the following data

7, 7, 8, 8, 8, 9, 9, 10, 10, 10, 11, 11, 12, 13, 13

Sol. Here, both the scores 8 and 10 occurs thrice (maximum number of times). So, we apply the empirical formula.

Here,

$$\text{mean} = \frac{7 \times 2 + 8 \times 3 + 9 \times 2 + 10 \times 3 + 11 \times 2 + 12 + 13 \times 2}{2 + 3 + 2 + 3 + 2 + 1 + 2}$$

$$= \frac{14 + 24 + 18 + 30 + 22 + 12 + 26}{15} = \frac{146}{15} = 9.73$$

No. of scores = 15 (odd)

$$\therefore \text{Median} = t_{\frac{15+1}{2}} = t_8 = 10$$

$$\therefore \text{Mode} = 3 \text{ median} - 2 \text{ mean}$$

$$= 3 \times 10 - 2 \times 9.73 = 30 - 19.46 = 10.54$$

Ex.55 The following table gives the weights of 40 men. Calculate mode.

Weights (in kg)	Number of men
54	6
72	6
80	1
64	2
62	6
60	5
58	5
56	4
63	5

Sol. Here, each of the scores 54, 72 and 62 occurs maximum number of times (six times). So we apply the empirical formula.

We construct the following table :

Weights x	No. of men f	Cumulative frequency	Product f.x
54	6	6	324
56	4	10	224
58	5	15	290
60	5	20	300
62	6	26	372
63	5	31	315
64	2	33	128
72	6	39	432
80	1	40	80
Total	40		2465

$$\text{Mean} = \frac{\Sigma f.x}{\Sigma f} = \frac{2465}{40} = 61.625$$

Here, No. of scores = 40 (even)

$$\text{Median} = \frac{t_{20} + t_{21}}{2} = \frac{60 + 62}{2} = 61$$

$$\therefore \text{Mode} = 3 \text{ median} - 2 \text{ mean}$$

$$= 3 \times 61 - 2 \times 61.625$$

$$= 183 - 123.25 = 59.75$$

Thus, modal weight = 59.75 kg