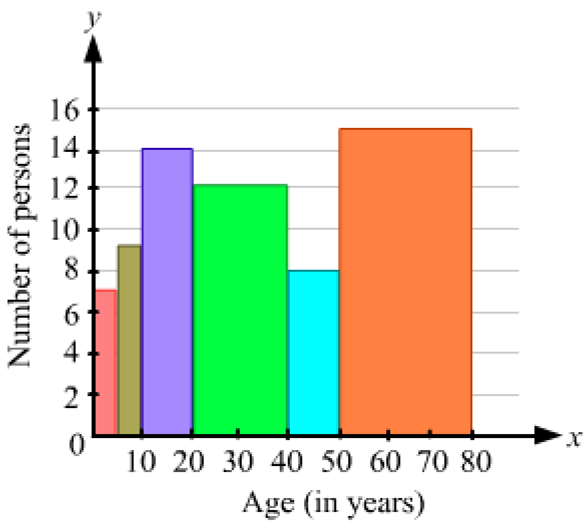
Construction of Histograms when Class Size is Different

Constructing Histogram for Data Having Different Class Intervals

The given **data** pertains to the ages of different persons in a particular colony.

Age (in years)	0–5	5–10	10–20	20–40	40–50	50–80
Number of persons	7	9	14	12	8	15

The **histogram** for the above data is shown in the given figure.



Do you think this graphical representation of the given data is correct?

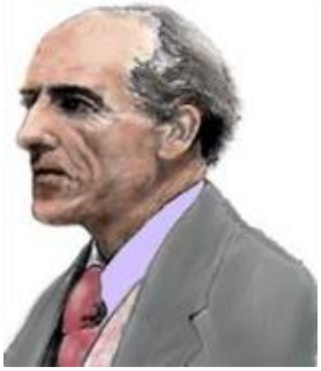
No, it is not. In a histogram, the area of each rectangle (and not its length) is directly proportional to the corresponding **frequency**. In case of a data with uniform or equal **class intervals**, the width of each rectangle is the same. The given graph, however, is misleading because the class intervals in the given data do not have a uniform **class size**. For this reason, we cannot use the conventional method of constructing a histogram.

In this lesson, we will learn how to construct a histogram when the class size is non-uniform.

Did You Know?

The word 'histogram' is regarded by many as the combination of the Greek words 'histos' and 'gramma'. 'Histos' refers to anything that is set upright and 'gramma' refers to any drawing, record or writing. The word was introduced by Karl Pearson in 1891. Some believe that Karl Pearson derived this word from 'historical diagram'.

Know Your Scientist



Karl Pearson (1857–1936) was a British mathematician who established the discipline of mathematical statistics. The term 'histogram' was introduced by him in 1891. In 1901, with Francis Galton and Raphael Weldon, he started the journal *Biometrika* for the development of statistical theory. He also headed the Department of Applied Statistics (now the Department of Statistical Science) at University College, London.

Know Your Scientist

Al-Kindi (c. 801–873 CE) was an Arab philosopher, mathematician, physician and musician. He wrote *Manuscript on Deciphering Cryptographic Messages*, which is one of the oldest books on statistics. In this book, Al-Kindi gave a detailed description of how to use statistics and frequency analysis to decipher encrypted messages. He combined mathematics and philosophy to disprove the eternity of the world through the demonstration of actual infinity. He said that eternity is just a mathematical and logical absurdity. In four volumes of *Ketab fi Isti'mal al-'Adad al-Hindi*, he wrote a lot about the use of Indian numerals. This book was instrumental in the diffusion of Indian numeral system in the Middle-Eastern and Western worlds.

Solved Examples

Example 1: **The given data pertains to the marks scored by the students of a particular class in a class test. Draw a histogram for this data.**

Marks	0–10	10–30	30–45	45–50	50–60
Number of students	8	32	18	10	18

Solution:

The class size in the given data is non-uniform, so we need to calculate the adjusted frequency for each class interval in order to draw the histogram.

Here, minimum class size = 5

We can find the adjusted frequencies using the following formula.

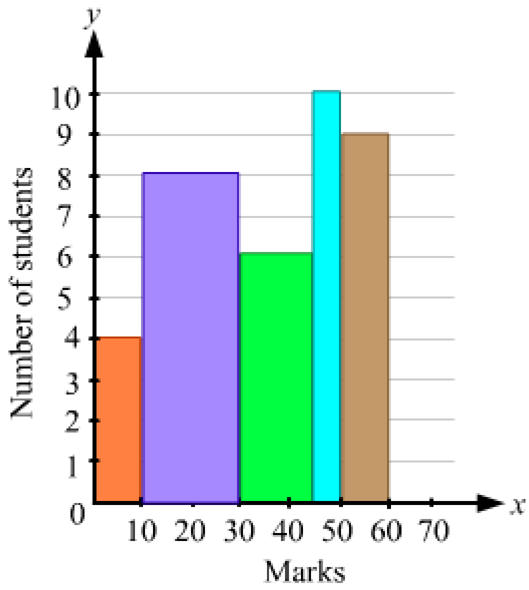
$$\text{Adjusted frequency of a class} = \frac{\text{Minimum class size}}{\text{Class size}} \times \text{Frequency}$$

The calculations of the adjusted frequencies are shown in the following table.

Marks	**Number of students (frequency)**	**Class size**	**Adjusted frequency**
0–10	8	10	$\frac{5}{10} \times 8 = 4$
10–30	32	20	$\frac{5}{20} \times 32 = 8$
30–45	18	15	$\frac{5}{15} \times 18 = 6$

45–50	10	5	$\frac{5}{5}\times 10=10$
50–60	18	10	$\frac{5}{10}\times 18=9$

The histogram of the given data can be drawn as follows:



Example 2: **The given data pertains to the prices of the different houses for sale in a particular society. Draw a histogram for this data.**

Price (in thousand rupees)	20–40	40–60	60–100	100–140	140–200
Number of houses	15	27	50	35	22

Solution:

The class size in the given data is non-uniform, so we need to calculate the adjusted frequency for each class interval in order to draw the histogram.

Here, minimum class size = 20

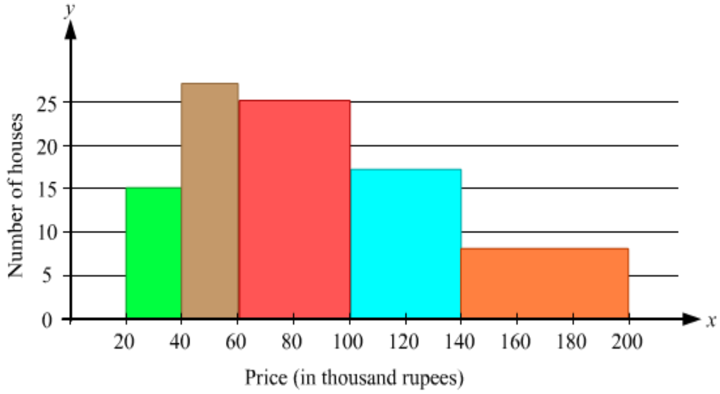
We can find the adjusted frequencies using the following formula.

$$\text{Adjusted frequency of a class} = \frac{\text{Minimum class size}}{\text{Class size}} \times \text{Frequency}$$

The calculations of the adjusted frequencies are shown in the following table.

Price (in thousand rupees)	**Number of houses (frequency)**	**Class size**	**Adjusted frequency**
20–40	15	20	$\frac{20}{20} \times 15 = 15$
40–60	27	20	$\frac{20}{20} \times 27 = 27$
60–100	50	40	$\frac{20}{40} \times 50 = 25$
100–140	35	40	$\frac{20}{40} \times 35 = 17.5$
140–200	22	60	$\frac{20}{60} \times 22 = 7.33$

The histogram of the given data can be drawn as follows:

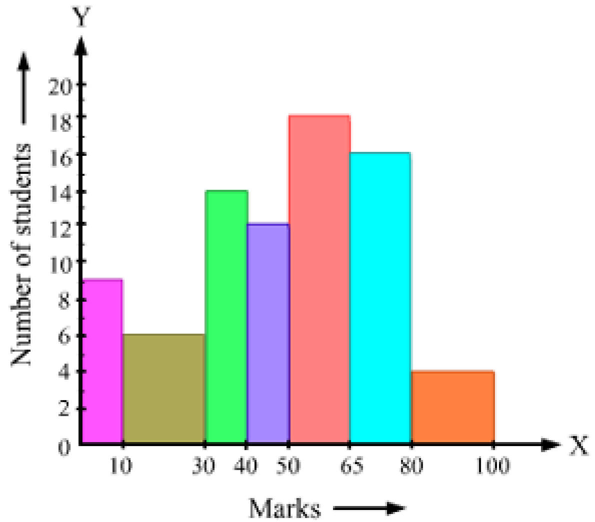


Interpretation of Histograms when Class Size is Different

Need for Interpreting Histograms

Histograms are one of the most commonly used tools for representing data in statistics. It is very important to be able to read histograms and interpret the data represented by them. The purpose of interpreting a histogram is to extract the data from the histogram and then use the same for specific purposes.

Take a look at the histogram shown below.



The given histogram represents data pertaining to the marks obtained by students in a test. It can be seen that the class intervals are not uniform in this histogram. Such histograms are quite common, so we need to learn to extract information from them.

In this lesson, we will learn how to interpret histograms having non-uniform class size.

Know More

A histogram generally shows:

1. The centre of the data.

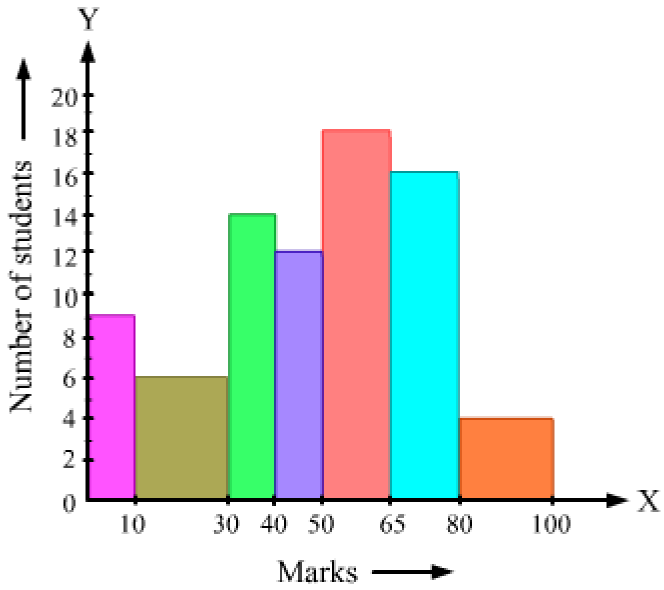
2. The spread of the data.

3. The presence of multiple peaks within the data

4. The presence of outliers (values outside the given range)

Interpretation of Histograms

We know that histograms are used to represent information. Now, suppose we are given a histogram. How do we go about interpreting the data represented by this graph? And what if the class size in the same histogram is non-uniform? Let us try to interpret one such histogram.



In the given histogram, the horizontal axis represents the marks scored by students in a test. The class intervals on the horizontal axis are 0−10, 10−30, 30−40, 40−50, 50−65, 65−80 and 80−100. The number of students in each class interval can be ascertained by looking at the vertical axis. Thus, we can say that the given histogram represents the numbers of students in the different groups of marks.

Here are a few questions relating to this histogram.

i) How many students are there in the highest group of marks?

The highest group of marks in the given histogram is the class interval 80−100. There are 4 students in this category as indicated by the height of the bar for this category. So, we can say that 4 students have secured 80 marks or more in the test.

ii) Which class interval contains the maximum number of students?

The tallest bar in the given histogram corresponds to the class interval 50−65. Thus, the maximum number of students belongs to this class interval. In other words, the maximum number of students in the class secured between 50 and 65 marks.

iii) If the students securing less than 50 marks have failed, then what number of students failed the test?

The students who secured less than 50 marks are those falling in the class intervals 0−10, 10−30, 30−40 and 40−50. We can find the total number of such students by adding the heights of the bars corresponding to these class intervals.

∴ Number of students who failed the test = 9 + 6 + 14 + 12 = 41

iv) What number of students took the test?

The number of students who took the test is given by the sum of the heights of the bars corresponding to all the class intervals in the histogram.

∴ Number of students who took the test = 9 + 6 + 14 + 12 + 18 + 16 + 4 = 79

v) What number of students secured 65 marks or more?

The number of students who scored 65 marks or more is given by the sum of the heights of the bars corresponding to the class intervals 65–80 and 80–100.

∴ Number of students who secured 65 marks or more = 16 + 4 = 20

Did You Know?

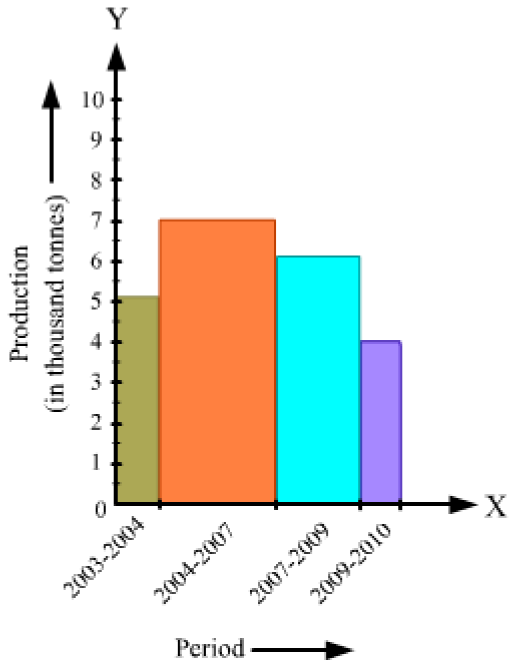
Florence Nightingale,an English social reformer and statistician, compared civilian mortality rates with those of the military by using circular histograms. She observed that as military patients died more frequently than civilian patients, it was important to make some improvements in the hygiene of the former.

Know Your Scientist

Florence Nightingale (1820–1910)was a British statistician. She is considered the pioneer of visual representation of information and of using statistical graphics. She developed a different form of pie chart known as the **polar area diagram** or the **Nightingale rose diagram**. It is equivalent to the modern circular histogram. She studied sanitation in Indian rural life and collected statistical data which lead to improvements in medical care and public health services in India.

Solved Examples

Example 1: **The following histogram shows the production of food grains (in thousand tonnes) over a period of time.**



i) What is the total production of food grains from 2004 to 2009?

ii) In which periods were the production of food grains the highest and the lowest?

Solution:

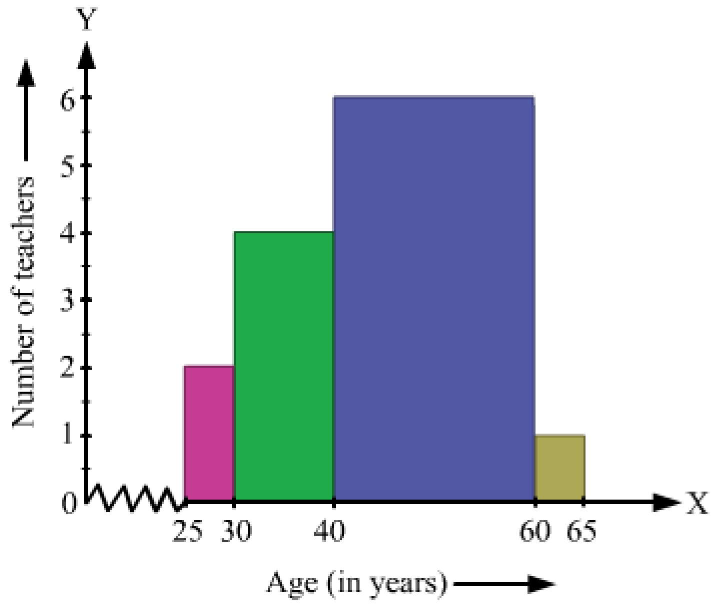
i) The total production of food grains from 2004 to 2009 can be ascertained by adding the heights of the class intervals 2004–2007 and 2007–2009.

∴ Total production of food grains from 2004 to 2009 = 7000 tonnes + 6000 tonnes = 13000 tonnes

ii) It is clear from the histogram that the bar corresponding to the class interval 2004–2007 is the tallest, and that corresponding to the class interval 2009–2010 is the shortest. So, the production of food grains was the highest in the period 2004–2007 and the lowest in the period 2009–2010 (i.e., 7000 tonnes and 4000 tonnes respectively).

Medium

Example 1: **The following histogram shows the ages of teachers in a school.**



Observe the histogram and answer the following questions.

i) How many teachers are aged 30 years or more, but less than 60 years?

ii) How many teachers are aged less than 40 years?

iii) Which age group contains the least number of teachers?

iv) To which age group do most teachers belong?

Solution:

i) The teachers aged 30 years or more, but less than 60 years are those falling in the class intervals 30–40 and 40–60. We can find the total number of such teachers by adding the heights of the bars corresponding to these class intervals.

∴ Number of teachers aged 30 or above, but below 60 = 4 + 6 = 10

ii) The teachers aged less than 40 years fall in the class intervals 25–30 and 30–40. We can find the total number of such teachers by adding the heights of the bars corresponding to these class intervals.

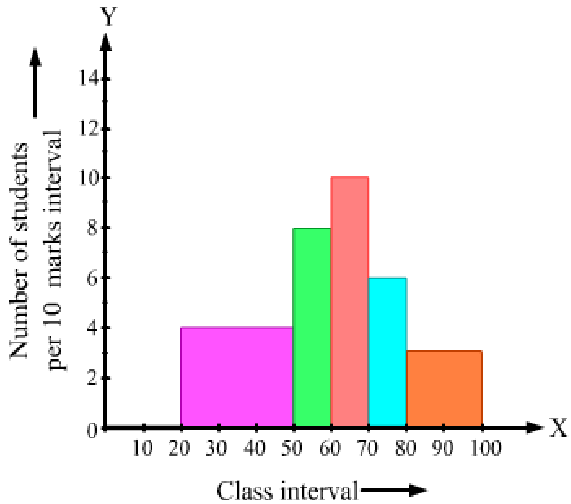
∴ Number of teachers aged below 40 = 2 + 4 = 6

iii) It is clear from the histogram that the bar corresponding to the class interval 60–65 is the shortest. Thus, the least number of teachers belongs to the age group 60–65.

iv) It is clear from the histogram that the bar corresponding to the class interval 40−60 is the tallest. Thus, most teachers belong to the age group 40−60.

Hard

Example 1: The given histogram shows the marks obtained (out of 100) by 42 students in a class.



Which of the following statements regarding the given histogram is correct?

A. The number of students whose marks are less than 50 is equal to the number of students who got 70 marks or above.

B. The number of students having marks in the ranges 50−60 and 70−80 is less than the number of students whose marks are less than 50.

C. The number of students whose marks are 80 or above is 3.

D. The number of students having marks in the range 70−80 is more than the number of students who got 80 marks or above.

Solution:

The correct answer is A.

The number of students for the corresponding class intervals can be found as follows:

Class interval	**Class size**	**Number of students (frequency)**

20−50	30	$\frac{4\times30}{10}=4\times3=12$
50−60	10	$\frac{8\times10}{10}=8$
60−70	10	$\frac{10\times10}{10}=10$
70−80	10	$\frac{6\times10}{10}=6$
80−100	20	$\frac{3\times20}{10}=6$

It is clear that the number of students having marks less than 50 is 12.

The number of students with 70 marks or above includes students in the class intervals 70−80 and 80−100, i.e., 6 + 6 = 12.

Thus, statement A is correct.

The number of students having marks in the ranges 50−60 and 70−80 is given by the frequencies of the corresponding class intervals, i.e., 8 + 6 = 14. This number is more than the number of students who scored less than 50 marks.

Thus statement B is incorrect.

The students having 80 marks or above are placed in the class interval 80−100. The number of such students is 6.

Thus statement C is incorrect.

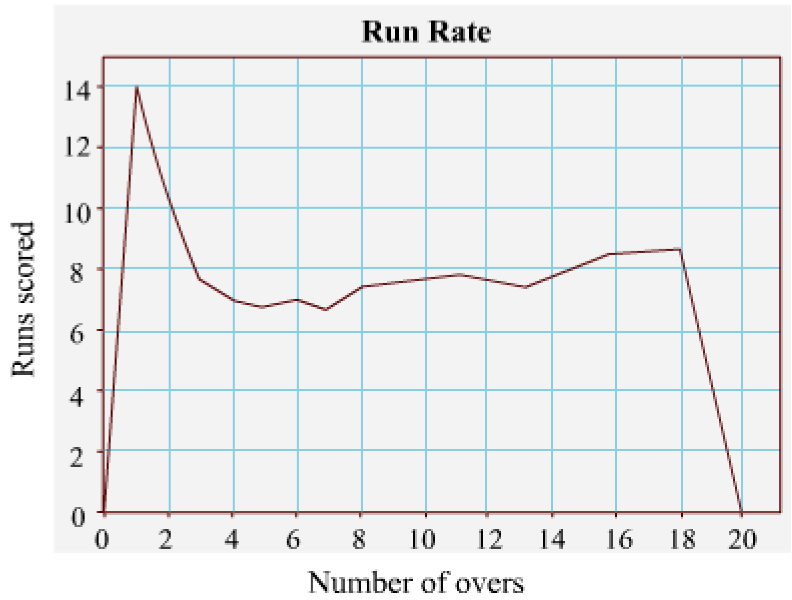
The number of students having marks in the range 70−80 is given by the frequency of the corresponding class interval, i.e., 6. This number is the same as the number of students who scored 80 marks or above.

Thus, statement D is incorrect.

Construction of Frequency Polygons

Frequency Polygons: An Introduction

A lot of people love watching and talking about cricket. Perhaps you too are one of them. You must be familiar with the run-rate graphs shown during the telecast of a cricket match. Here is one such run-rate graph. It shows the runs scored by a team in a T20 match.



Do you know what this type of graph is called? It is a **frequency polygon**. Such graphs are used in a variety of contexts. They are used while preparing business reports, **census** reports, weather reports, etc. These graphs are really useful in real life.

In this lesson, we will learn to draw such frequency polygons.

Frequency Polygons: In Depth

A frequency polygon is a continuous curve obtained by plotting and joining the ordered pairs of **class marks** and their corresponding frequencies. Another way of drawing a frequency polygon is by joining the midpoints of the tops of the bars of a histogram and the midpoints of the classes preceding and succeeding the lowest and highest class intervals respectively of the histogram.

A frequency polygon can be drawn with or without using a histogram. In the first method, we need to first draw the histogram before drawing the frequency polygon. In the second method, we draw the frequency polygon directly using the given data.

Did You Know?

Sociologists prefer to use frequency polygons instead of frequency distributions. This is because frequency polygons provide more information without much observation about the distribution of data. Another reason for this preference is that multiple frequency polygons can be easily analyzed simultaneously. This enables sociologists to analyze data more efficiently.

Solved Examples

Example 1: **Here are the weights (in kg) of the babies born in a hospital during a particular week.**

2.3, 2.0, 2.5, 2.7, 3.0, 3.2, 3.1, 2.2, 3.0, 2.5, 2.4, 3.0, 2.3, 2.4, 2.8

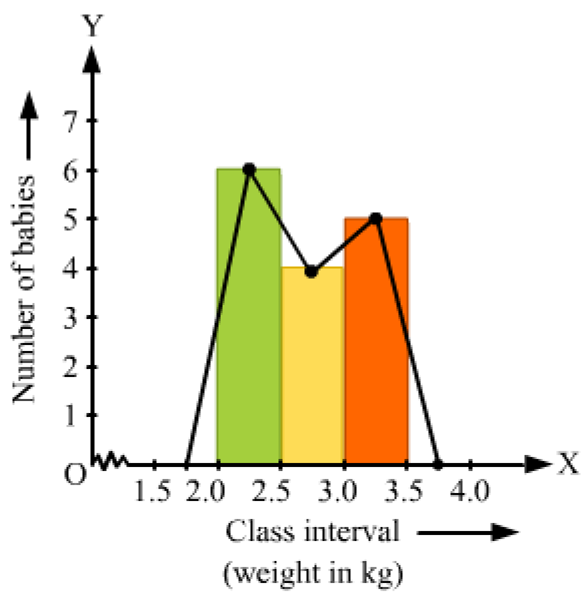
Draw a histogram for the data and then draw a frequency polygon using it.

Solution:

The frequency distribution table of the given data is as follows:

Class interval	Frequency
2.0−2.5	6
2.5−3.0	4
3.0−3.5	5

The histogram and frequency polygon for the given data are drawn as is shown.



Solved Examples

Example 1: **Draw a frequency polygon for the following data without using a histogram.**

Daily earnings (in rupees)	300–350	350–400	400–450	450–500	500–550
Number of stores	5	10	17	20	3

Solution:

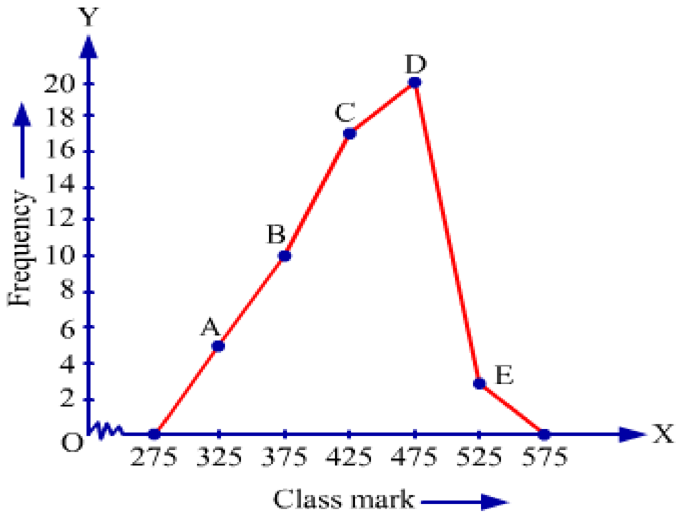
We will first calculate the class marks and write the data as follows:

10. Gravitation	10. Gravitation	10. Gravitation	10. Gravitation
10. Gravitation	10. Gravitation	10. Gravitation	10. Gravitation

The class interval preceding the lowest class is 250–300 and the class interval succeeding the highest class is 550–600. We assume the frequency of each of these two classes as zero.

The class marks of the intervals 250–300 and 550–600 are 275 and 575 respectively.

Now, by plotting and joining the points (275, 0), A, B, C, D, E and (575, 0), we obtain the required frequency polygon as is shown.



Mean of Data Sets

Application of Mean in Real Life

The runs scored by the two opening batsmen of a team in ten successive matches of a cricket series are listed in the table.

Player A	24	50	34	24	20	96	105	50	13	27
Player B	26	22	30	10	42	98	40	54	10	122

Using this data, we can compare the performances of the players for each individual game. For example, player B performed better than player A in the first match, player A then performed better than player B in the second match, etc.

This method, however, is not useful in trying to determine the overall performances of the two players and comparing them. For this we need to calculate the average or mean score of each player. The player having the better average or mean score has the better overall performance.

In this lesson, we will learn how to find the mean of a data set.

Did You Know?

1. Arithmetic mean (AM), mean or average are all the same.

2. Mean is used in calculating average temperature, average mark, average score, average age, etc. It is also used by the government to find the average individual expense and income.

3. Mean cannot be determined graphically.

4. Mean is supposed to be the best measure of central tendency of a given data.

5. Mean can be determined for almost every kind of data.

Properties of Mean

1. Sum of the deviations taken from the arithmetic mean is zero.
If the mean of *n* observations
$x_1, x_2, x_3 \dots . x_n$ is $\bar{x}$ then $(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \dots + (x_n - \bar{x}) = 0$.

2. If each observation is increased by *p* then the mean of the new observations is also increased by *p*. If the mean of *n* observations
$x_1, x_2, x_3 \dots . x_n$ is $\bar{x}$ then the mean of $(x_1 + p), \ (x_2 + p), \ (x_3 + p), \ \dots, \ (x_n + p)$ is $(\bar{x} + p)$.

3. If each observation is decreased by *p* then the mean of the new observations is also decreased by *p*. If the mean of *n* observations
$x_1, x_2, x_3 \dots . x_n$ is $\bar{x}$ then the mean of $(x_1 - p), \ (x_2 - p), \ (x_3 - p), \ \dots, \ (x_n - p)$ is $(\bar{x} - p)$.

4. If each observation is multipled by *p*(where p ≠ 0) then the mean of the new observations is also multiplied by *p*. If the mean of *n* observations

$x_1, x_2, x_3 \dots . x_n$ is $\bar{x}$ then the mean of $px_1, \ px_2, \ px_3, \ \dots, \ px_n$ is $p\bar{x}$.

5. If each observation is divided by p(where p ≠ 0) then the mean of the new observations is also divided by p. If the mean of n observations

$x_1, x_2, x_3 \ldots . x_n$ is $\bar{x}$ then the mean of $\frac{x_1}{p}, \frac{x_2}{p}, \frac{x_3}{p}, \ldots, \frac{x_n}{p}$ is $\frac{\bar{x}}{p}$.

Solved Examples

Example 1: The amounts of money spent by Sajan during a particular week are listed in the table.

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Money spent (in rupees)	270	255	195	230	285	225	115

Find the average amount of money spent by him per day.

Solution:

$$\text{Average amount of money spent by Sajan per day} = \frac{\text{Total money spent}}{\text{Total number of days}}$$
$$= \text{Rs}\frac{270+255+195+230+285+225+115}{7}$$
$$= \text{Rs}\frac{1575}{7}$$
$$= \text{Rs}\,225$$

Example 2: The average weight of the students in a class is 42 kg. If the total weight of the students is 1554 kg, then find the total number of students in the class.

Solution:

Let the total number of students in the class be x.

$$\text{Average weight of the students} = \frac{\text{Total weight of the students}}{\text{Total number of students}}$$
$$\Rightarrow \text{Total number of students} = \frac{\text{Total weight of the students}}{\text{Average weight of the students}}$$
$$\Rightarrow \therefore x = \frac{1554}{42}$$
$$= 37$$

Thus, there are 37 students in the class.

Medium

Example 1: **For what value of *x* is the mean of the data 28, 32, 41, *x*, *x*, 5, 40 equal to 31?**

Solution:

$$\text{Mean of the given data set} = \frac{\text{Sum of all observations}}{\text{Number of observations}}$$
$$\Rightarrow 31 = \frac{28+32+41+x+x+5+40}{7}$$
$$\Rightarrow 217 = 2x+146$$
$$\Rightarrow 2x = 71$$
$$\Rightarrow \therefore x = 35.5$$

Thus, for $x = 35.5$, the mean of the data 28, 32, 41, x, x, 5, 40 is 31.

Example 2: **The numbers of children in five families are 0, 2, 1, 3 and 4. Find the average number of children. If two families having 6 and 5 children are included in this data set, then what is the new mean or average?**

Solution:

$$\text{Mean of the given data set} = \frac{\text{Sum of all observations}}{\text{Number of observations}}$$
$$\therefore \text{Mean of the initial data set} = \frac{0+2+1+3+4}{5} = \frac{10}{5} = 2$$

Thus, the average number of children for the five families in the initial data set is 2.

Two families are added to the initial set of families.

$$\therefore \text{Mean of the new data set} = \frac{0+2+1+3+4+6+5}{7} = \frac{21}{7} = 3$$

Thus, the average number of children for the seven families in the new data set is 3.

Example 3: **The mean of fifteen numbers is 7. If 3 is added to every number, then what will be the new mean?**

Solution:

Let $x_1, x_2, x_3, \ldots, x_{15}$ be the fifteen numbers having the mean as 7 and x¯x¯ be the mean.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$
$$\Rightarrow 7 = \frac{x_1 + x_2 + x_3 + \ldots + x_{15}}{15}$$
$$\Rightarrow x_1 + x_2 + x_3 + \ldots + x_{15} = 15 \times 7$$
$$\Rightarrow x_1 + x_2 + x_3 + \ldots + x_{15} = 105 \qquad \ldots(1)$$

The new numbers are $x_1 + 3,\ x_2 + 3,\ x_3 + 3,\ \ldots,\ x_{15} + 3$.

Let $\overline{X}$ be the mean of the new numbers.

$$\overline{X} = \frac{(x_1 + 3) + (x_2 + 3) + \ldots + (x_{15} + 3)}{15}$$
$$\Rightarrow \overline{X} = \frac{(x_1 + x_2 + \ldots + x_{15}) + 3 \times 15}{15}$$
$$\Rightarrow \overline{X} = \frac{105 + 45}{15} \qquad (\text{By equation 1})$$
$$\Rightarrow \overline{X} = \frac{150}{15}$$
$$\Rightarrow \therefore \overline{X} = 10$$

Thus, the mean of the new numbers is 10.

Hard

Example 1: **The average salary of five workers in a company is Rs 2500. When a new worker joins the company, the average salary is increased by Rs 100. What is the salary of the new worker?**

Solution:

Let the salary of the new worker be Rs *x*.

Before the joining of the new worker, we have:

$$\text{Mean salary of the five workers} = \frac{\text{Sum of the salaries of the five workers}}{5}$$

$$\Rightarrow 2500 = \frac{\text{Sum of the salaries of the five workers}}{5}$$

$$\Rightarrow \therefore \text{Sum of the salaries of the five workers} = 2500 \times 5 = 12500 \quad \ldots(1)$$

After the joining of the new worker, we have:

Number of workers = 5 + 1 = 6

Average salary = Rs (2500 + 100) = Rs 2600

$$\text{Mean salary of the six workers} = \frac{\text{Sum of the salaries of the six workers}}{6}$$

$$\Rightarrow 2600 = \frac{\text{Sum of the salaries of the five workers} + \text{Salary of the new worker}}{6}$$

$$\Rightarrow 2600 = \frac{12500 + x}{6} \qquad (\text{By equation 1})$$

$$\Rightarrow 15600 = 12500 + x$$

$$\Rightarrow \therefore x = 15600 - 12500 = 3100$$

Thus, the salary of the new worker is Rs 3100.

Example 2:

Find two numbers that lie between $\frac{2}{5}$ and $\frac{1}{2}$.

Solution:

The given numbers are $\frac{2}{5}$ and $\frac{1}{2}$.

Mean of the two numbers = $\frac{\frac{2}{5}+\frac{1}{2}}{2} = \frac{\frac{4+5}{10}}{2} = \frac{9}{10 \times 2} = \frac{9}{20}$

Now, we know that the mean of any two numbers lies between the numbers.

Hence, $\frac{2}{5} < \frac{9}{20} < \frac{1}{2}$

Mean of $\frac{9}{20}$ and $\frac{1}{2}$ = $\frac{\frac{9}{20}+\frac{1}{2}}{2}=\frac{\frac{9+10}{20}}{2}=\frac{19}{20\times 2}=\frac{19}{40}$

Hence, $\frac{9}{20}<\frac{19}{40}<\frac{1}{2}$

And $\frac{2}{5}<\frac{9}{20}<\frac{19}{40}<\frac{1}{2}$

So, two numbers that lie between $\frac{2}{5}$ and $\frac{1}{2}$ are $\frac{9}{20}$ and $\frac{19}{40}$.

Example 3:

If $\bar{x}$ is the mean of the *n* observations $x_1, x_2, x_3, \ldots, x_n$, then prove that

$$\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)}{n}=0$$

Solution:

It is given that $\bar{x}$ is the mean of the *n* observations $x_1, x_2, x_3, \ldots, x_n$.

Thus,

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

$$\Rightarrow x_1+x_2+x_3+\ldots+x_n=n\bar{x} \qquad \ldots(1)$$

Now,

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

$$=\frac{\left(x_1+x_2+x_3+\ldots+x_n\right)-\left(\bar{x}+\bar{x}+\bar{x}+\ldots+\bar{x}\right)}{n}$$

$$=\frac{n\bar{x}-n\bar{x}}{n} \qquad \text{(By equation 1)}$$

$$=\frac{0}{n}$$

$$=0$$

Thus, the given result is proved.

Mean of Frequency Distribution Using Summation Formula

Calculating Mean from a Frequency Distribution Table

During a cricket match, two batsmen score the following runs off twelve balls of two consecutive overs.

4, 2, 0, 1, 2, 0, 6, 1, 1, 1, 0, 2

We can find the mean number of runs scored per ball by applying the following formula.

$$\text{Mean}\left(\bar{x}\right)=\frac{x_1+x_2+x_3+\ldots+x_n}{n}$$

Sometimes data can be given in the form of a **frequency distribution table**, as is shown.

Runs scored	0	1	2	4	6
Frequency	3	4	3	1	1

Can you find the average or mean number of runs scored per ball from this table?

Since the data given in the table is ungrouped, we cannot apply the conventional formula to find the mean.

In this lesson, we will learn how to calculate the mean of an ungrouped data presented as a frequency distribution table.

Did You Know?

1. Mean is least affected by fluctuation of data as it depends on the whole data, and not on any particular value in the data.

2. Mean is capable of further mathematical treatment. For example, if we have the mean of two or more groups, then we can calculate the combined mean of the group formed by the combination of these groups.

Mean has one and only one interpretation. So, if different persons compute its measure using the same observations, then they will get the same result as the mean.

Solved Examples

Easy

Example 1: **The table lists the average heights of students of five sections of class 10 in a school.**

Section	A	B	C	D	E
Number of students	30	35	25	28	32
Average height (in cm)	150	156	152	149	157

Find the average height of each student of class 10 in the school.

Solution: To find the mean of a given data, we first need to find $\sum f_i$ and $\sum f_i x_i$. This is shown in the following table.

Average height (xi) (in cm)	**Number of students (fi)**	**fixi (in cm)**
150	30	4500
156	35	5460
152	25	3800
149	28	4172
157	32	5024

Total	150	22956

Total number of students $= \sum f_i = 150$

$\sum f_i x_i = 22956$

We know that

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i} = \left(\frac{22956}{150}\right) \text{cm} = 153.04 \text{ cm}$$

So, the average height of each student of class 10 in the school is 153.04 cm.

Medium

Example 1: **Students of four cities take part in a quiz. The data obtained from the quiz is listed in the table.**

City	I	II	III	IV
Number of students	40	48	?	60
Average score	50	80	55	75

If the mean score of the students of all the four cities is 66, then find the number of students that took part in the quiz from city III.

Solution:

Let the number of students that took part in the quiz from city III be *x*.

To find the mean of a given data, we first need to find $\sum f_i$ and $\sum f_i x_i$. This is shown in the following table.

Average score (xi)	**Number of students (fi)**	**fixi**
50	40	2000
80	48	3840

55	x	55x
75	60	4500
Total	148 + x	10340 + 55x

Total number of students $= \sum f_i = 148 + x$

$\sum f_i x_i = 10340 + 55x$

Mean = 66

We know that

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i}$$
$$\Rightarrow 66 = \frac{10340 + 55x}{148 + x}$$
$$\Rightarrow 66(148 + x) = 10340 + 55x$$
$$\Rightarrow 9768 + 66x = 10340 + 55x$$
$$\Rightarrow 11x = 572$$
$$\Rightarrow \therefore x = \frac{572}{11} = 52$$

So, the number of students that took part in the quiz from city III is 52.

Example 2: **Find the value of *x* if the mean of the given data is 13.2.**

Observation	3	7	11	x	19	23
Frequency	8	6	11	10	13	2

Solution:

To find the mean of a given data, we first need to find $\sum f_i$ and $\sum f_i x_i$. This is shown in the following table.

Observation (xi)	**Frequency (fi)**	**fixi**

3	8	24
7	6	42
11	11	121
x	10	10x
19	13	247
23	2	46
Total	50	480 + 10x

Total frequency $= \sum f_i = 50$

$\sum f_i x_i = 480 + 10x$

Mean = 13.2

We know that

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i}$$
$$\Rightarrow 13.2 = \frac{480 + 10x}{50}$$
$$\Rightarrow 480 + 10x = 50 \times 13.2$$
$$\Rightarrow 10x = 660 - 480$$
$$\Rightarrow 10x = 180$$
$$\Rightarrow \therefore x = \frac{180}{10} = 18$$

Hard

Example 1: **In the given frequency table, the frequencies of observations 30 and 40 are missing.**

Observation	10	20	30	40	50
Frequency	6	3	-	-	4

If the frequency of 40 is two more than that of 30 and the mean of this data is 30, then find the total frequency of the given data?

Solution: Let the frequency of 30 be *x*.

So, frequency of 40 = *x* + 2

To find the mean of a given data, we first need to find $\sum f_i$ and $\sum f_i x_i$. This is shown in the following table.

Observation (xi)	**Frequency (fi)**	**fixi**
10	6	60
20	3	60
30	x	30x
40	x + 2	40x + 80
50	4	200
Total	15 + 2x	400 + 70x

Total frequency = $\sum f_i = 15 + 2x$

$\sum f_i x_i = 400 + 70x$

Mean = 30

We know that

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i}$$
$$\Rightarrow 30 = \frac{400 + 70x}{15 + 2x}$$
$$\Rightarrow 30(15 + 2x) = 400 + 70x$$
$$\Rightarrow 450 + 60x = 400 + 70x$$
$$\Rightarrow 10x = 50$$
$$\Rightarrow \therefore x = \frac{50}{10} = 5$$

So, total frequency of the given data = 15 + 2 × 5 = 25

Medians of Data Sets Having Odd or Even Number of Terms

Median as the Measure of Central Tendency

Let us consider the group of half dozen individuals shown in the picture. There are five children standing with a very tall man.

You can see that the distribution of height in this group is unbalanced or asymmetrical because one individual is much taller than the others. When we calculate the mean height of the five children, the value so obtained will be close to the actual height of each child. However, when we calculate the mean height of the six persons (including the really tall man), the value so obtained will give the impression that each individual in the group is quite tall. So, in this case, the mean will not be an appropriate measure of central tendency. In situations such as this, we use median as the measure of central tendency.

In this lesson, we will study about median and the method to calculate the same for any given data.

Method to Find Median

Median can be defined as follows:

Median is the value of the middlemost observation when the data is arranged in increasing or decreasing order.

The method to find median can be summarized as follows:

Step 1: Arrange the data in increasing or decreasing order.

Step 2: Let *n* be the number of observations. Here, two cases arise.

Case 1: When *n* is even, the median of the observations is given by the formula

Median = Mean of the $\left(\frac{n}{2}\right)^{\text{th}}$ **and** $\left(\frac{n}{2}+1\right)^{\text{th}}$ **observations**

Case 2: When *n* is odd, the median of the observations is given by the formula

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

Did You Know?

1. Median is used to measure the distribution of earnings, to calculate the poverty line, etc.

2. Median is independent of the range of the series as it is the middle value of a data. So, it is not affected by extreme values or end values.

3. The median of a data is incapable of further algebraic or mathematical treatment. For example, if we have the median of two or more groups, then we cannot find the median of the bigger group formed by combining the given groups.

4. Median is affected by the fluctuation in data as it depends only on one item, i.e., the middle term.

Know Your Scientist

Antoine Augustin Cournot

Antoine Augustin Cournot (1801–1877) was a French economist, philosopher and mathematician. The term median was introduced by him in 1843. He used this term for the value that divides a probability distribution into two equal parts or halves.

In the field of economic analysis, he developed the concept of functions and probability. He introduced the demand curve to show the relationship between price and demand for any given item. He is best remembered for his theory of strategic behaviour of competitors in a market having only two players, i.e., in a duopoly.

Know More

Advantages of median

1. Median is better suited for non-symmetrical distributions as it is not much affected by very low and high values. Non-symmetrical distribution means the data is distributed in such a way that the values toward one end are much higher or lower than the values toward the other end. For example, 1, 2, 3, 4, 25, 30, 50, 60 is a non-symmetrical distribution.

2. Knowing the median test score is important to people who want to know whether they belong to the 'better half of the population' or not.

Mode can be defined as:

"The observation which occurs the maximum number of times is called **mode**".
Or "the observation with maximum frequency is called **mode**".

Example 2: **Find the mode of the following marks obtained by 15 students.**
2, 5, 1, 0, 8, 11, 8, 12, 19, 18, 13, 10, 9, 8, 1
Solution:
We arrange the given data as follows:
0, 1, 1, 2, 5, 8, 8, 8, 9, 10, 11, 12, 13, 18, 19
We observe that 8 occurs most often. So, the mode is 8.

Solved Examples

Easy

Example 1: **Find the median of these observations: 324, 250, 234, 324, 250, 196, 189, 250, 313, 227.**

Solution:

On writing the observations in ascending order, we have the following sequence.

189, 196, 227, 234, 250, 250, 250, 313, 324, 324

Here, the number of observations (n) is 10, which is an even number.

∴ Median = Mean of the $\left(\frac{10}{2}\right)^{\text{th}}$ and $\left(\frac{10}{2}+1\right)^{\text{th}}$ observations

⇒ Median = Mean of the 5^{th} and 6^{th} observations

Here, the 5^{th} and 6^{th} observations are the same value, i.e., 250.

∴ Median $\frac{250+250}{2}$ = 250

Medium

Example 1: **The minimum temperatures (in °C) for fifteen days in a city are recorded as follows:**

4.5, 4.7, 3.9, 5.2, 5.0, 4.2, 4.6, 4.2, 4.2, 4.5, 5.7, 2.3, 6.0, 3.5, 4.0

Find the median of the minimum temperatures.

Solution: On arranging the data in ascending order, we obtain the following sequence.

2.3, 3.5, 3.9, 4.0, 4.2, 4.2, 4.2, 4.5, 4.5, 4.6, 4.7, 5.0, 5.2, 5.7, 6.0

Here, the number of observations (n) is 15, which is an odd number.

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

⇒ Median = $\left(\frac{15+1}{2}\right)^{\text{th}}$ observation

Here, the 8^{th} observation is 4.5.

Thus, the median of the minimum temperatures for the fifteen days was 4.5°C.

Example 2: **The marks obtained (out of 50) by fifteen students are 27, 31, 29, 35, 30, 42, 45, 41, 37, 32, 28, 36, 44, 34 and 43. Find the median. If the marks 27 and 44 are replaced by 25 and 46, then what will be the new median?**

Solution: The given marks can be arranged in ascending order as follows:

27, 28, 29, 30, 31, 32, 34, 35, 36, 37, 41, 42, 43, 44, 45

Here, the number of observations (n) is 15, which is an odd number.

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

= $\left(\frac{15+1}{2}\right)^{\text{th}}$ observation

= 8^{th} observation

= 35

After replacing 27 and 44 by 25 and 46, the marks are arranged in ascending order as follows:

25, 28, 29, 30, 31, 32, 34, 35, 36, 37, 41, 42, 43, 45, 46

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

= $\left(\frac{15+1}{2}\right)^{\text{th}}$ observation

= 8^{th} observation

= 35

Hard

Example 1: **The following observations are arranged in ascending order.**

11, 17, 20, 25, 39, 2y, 3y + 1, 69, 95, 112, 135, 1204

If the median of the data is 53, then find the value of y.

Solution: The observations in ascending order are given as follows:

11, 17, 20, 25, 39, $2y$, $3y + 1$, 69, 95, 112, 135, 1204

Here, the number of observations (n) is 12, which is even.

$\therefore$ Median = Mean of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ observations

It is given that the median of the observations is 53.

$\therefore$ 53 = Mean of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ observations.

$\Rightarrow$ 53 = Mean of $\left(\frac{12}{2}\right)^{th}$ and $\left(\frac{12}{2}+1\right)^{th}$ observations

$\Rightarrow$ 53 = Mean of $(6)^{th}$ and $(7)^{th}$ observations

The 6^{th} and 7^{th} observations are $2y$ and $3y + 1$ respectively.

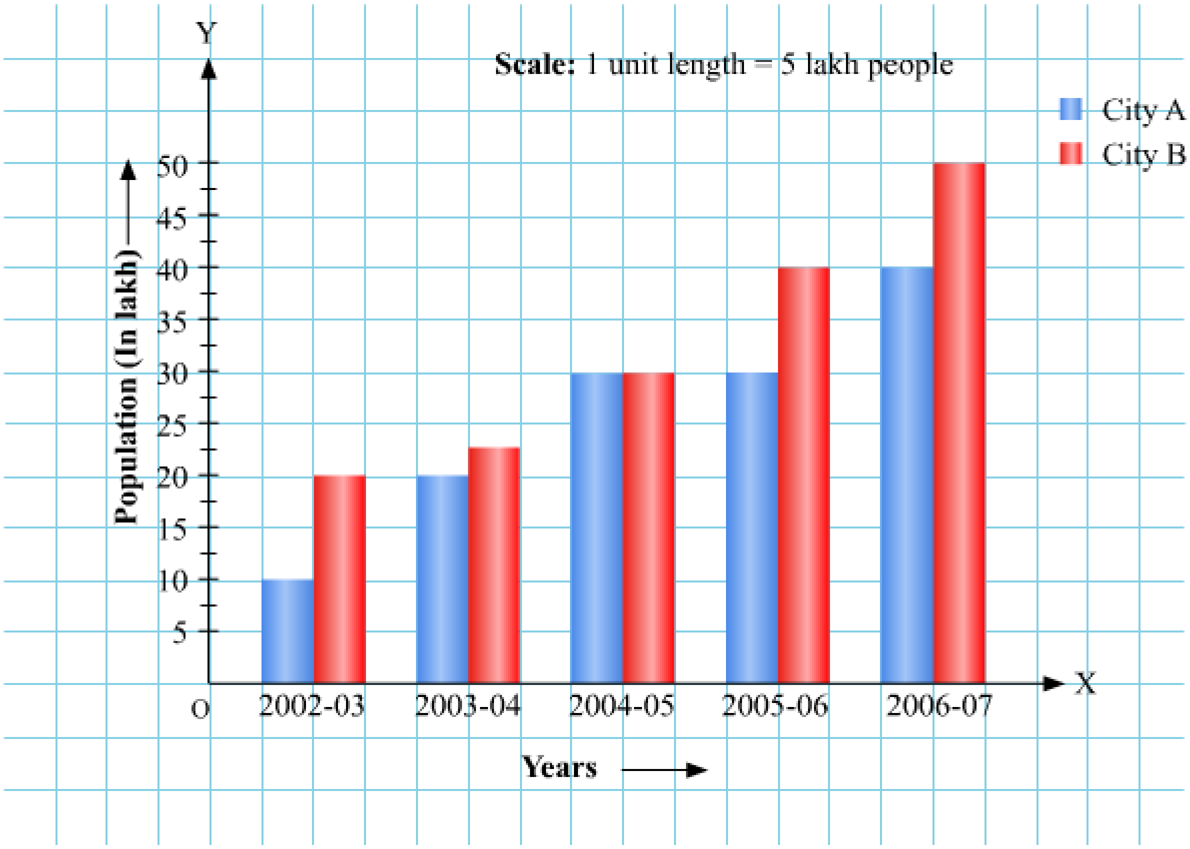
So,

$$53 = \frac{(2y)+(3y+1)}{2}$$
$$\Rightarrow 106 = 5y+1$$
$$\Rightarrow 5y = 105$$
$$\Rightarrow \therefore y = \frac{105}{5} = 21$$

Interpretation of Double Bar Graphs

We know how to construct a double bar graph but do you know that we can get much information from a double bar graph.

Example 1: Look at the double bar graph given below and answer the following questions.

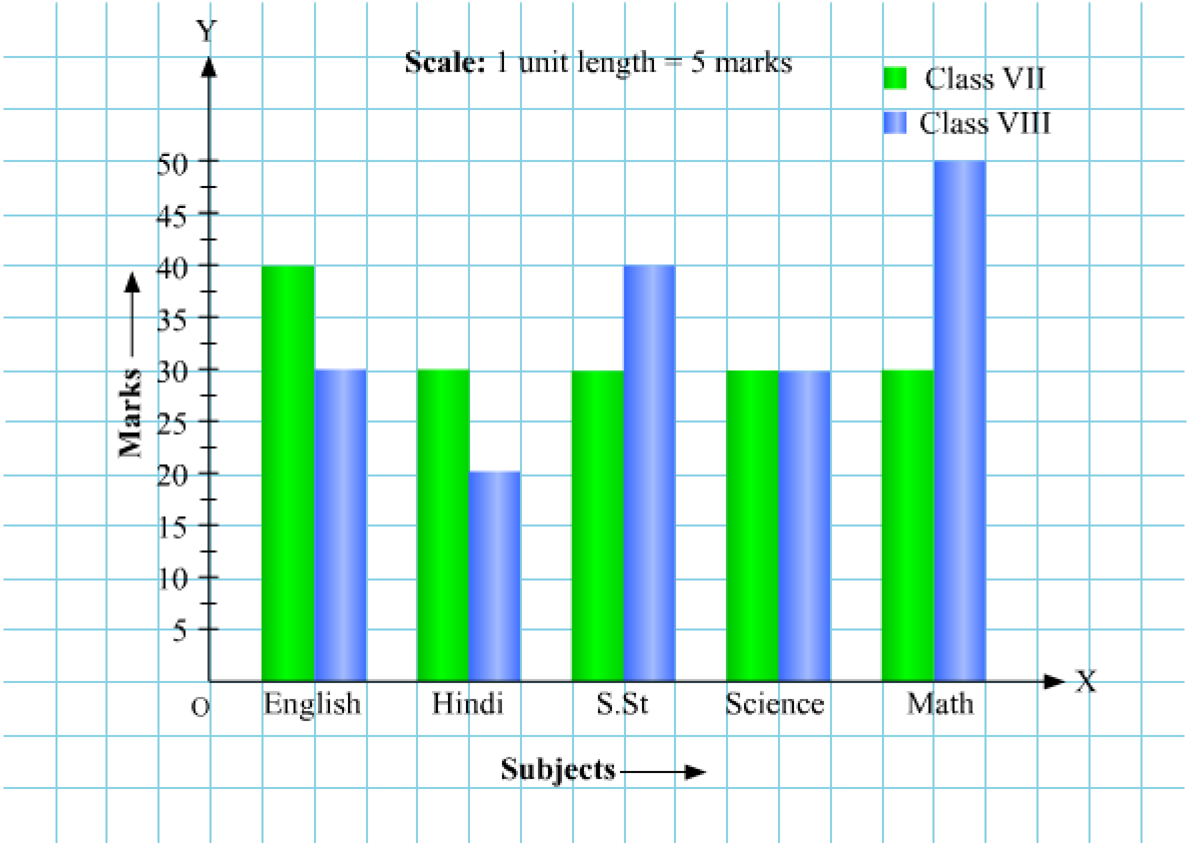


1. **What information is represented in the bar graph?**
2. **In which year was the population of city B the largest?**
3. **In which year was the population of city A same as that of city B?**
4. **Can you estimate the population of city A in the year 2003-04 and also in the year 2005-06?**

Solution:

1. The bar graph represents the population of two cities A and B for five consecutive years from 2002-03 to 2006-07.
2. Since the bar representing city B is biggest in the year 2006-07, the population of city B was the largest in the year 2006-07.
3. The population of the two cities was same in the year 2004-05, since the length of both the bars are the same in this year.
4. As seen from the graph in the year 2003-04, the population of city A was 20 lakhs and in the year 2005-06, the population of city A was 30 lakhs.

Example 2: The following graph shows the marks obtained by Ritesh in class VII and VIII in five subjects.



Observe the graph and answer the following questions.

1. **In which subject was the performance of Ritesh same in both classes?**
2. **In which subject did the performance deteriorate?**
3. **In which subject was there the maximum difference in marks?**
4. **How many marks did Ritesh score in English in both the classes?**

Solution:

1. The performance of Ritesh in Science was same in both the classes, since the length of the bar is same for both classes.
2. The performance deteriorated in Hindi and English both, since the blue bar is smaller than the green bar in these two subjects.
3. The maximum difference of marks can be seen in Maths as we can see that the difference between the heights of the blue and green bars is maximum for Maths.
4. Ritesh got 40 marks in English in class VII and 30 marks in class VIII.

Frequency Distribution Table and Terminology Related to It

Observe the information given in the following cases.

(1) The weights (in kg) of 15 students in the same class are as follows:

45, 50, 48, 47, 58, 52, 49, 54, 48, 51, 46, 57, 56, 50, 44

(2) Minimum temperature (in °C) of a city for each day of a week is given as follows:

1.5, 2, 0, 2.5, 3.5, 1, 1

(3) Runs scored by 8 players of a cricket team in a match are as follows:

Player	**Runs**
Harry	21
Venkat	16
Robin	74
Dinesh	09
Vikram	81
Laxmipati	42
Jairaj	36
Ysuf	27

It can be seen that in each case, we have some numeric information.

Numeric information collected for a particular purpose is known as raw data and each number involved in this raw data is known as score.

For example, in case (1), weight of each student is a score.

Similarly, in case (2), temperature of each day and in case (3), runs scored by each player are scores.

Raw data is found in unorganized form and in many real life situations, we have to deal with it.

Data can be of two types such as **primary data** and **secondary data**.

Primary data: When the data is collected by an investigator according to a plan for a particular objective then the collected data is called primary data.

For example, if a person collects information about the people using a particular mobile phone network in a particular locality, then the data collected by the person will be called primary data.

Secondary data: When the required data is taken from the data already collected by other private agency, government agency, an organization or any other party then the data is called secondary data.

For example, if an organization extracts the data from the records of census published by the government, then the data is called secondary data.

To draw meaningful inference, we organize the data into systematic pattern in the form of frequency distribution table.

Let us understand this with the help of an example.

Heights (in cm) of 30 students of a class are given as follows:

152, 160, 154, 151, 158, 165, 152, 160, 160, 152, 152, 161, 158, 160, 152, 165, 165, 155, 158, 158, 154, 158, 160, 161, 158, 161, 158, 155, 161, 160

Now, it can be seen that the lowest score is 151 and the highest score is 165.

The difference between the highest observation and lowest observation in a given data set is called the **range**. Range of the above data is 165 – 151 = 14.

It can be observed that few scores occur more than once in the data.

Number of times by which a score occurs in the data is called the frequency of that score.

Score 151 occurs just once, so its frequency is 1.

Similarly,

Score 152 occurs five times, so its frequency is 5.

Score 154 occurs two times, so its frequency is 2.

Score 155 occurs two times, so its frequency is 2.

Score 158 occurs seven times, so its frequency is 7.

Score 160 occurs six times, so its frequency is 6.

Score 161 occurs four times, so its frequency is 4.

Score 165 occurs three times, so its frequency is 3.

The sum of all frequencies or total frequency is 30 which gives us the total number of scores in the data. Total frequency is denoted by N.

Now, we can arrange these scores in a table according to their respective frequencies and such a table is known as frequency distribution table.

Frequency distribution table for the given data is as follows:

Height	**Tally Mark**	**Frequency**
151	|	1
152	~~||||~~	5
154	||	2
155	||	2
158	~~||||~~ ||	7
160	~~||||~~ |	6
161	||||	4
165	|||	3
	Total (N)	30

The bars in the second column are known as tally marks which are used to represent the numbers.

In tally marks representation, 1 is represented by one bar i.e., |, 2 is represented by the group of two bars i.e., || and 5 is represented by ~~||||~~ (four vertical bars are intersected by one bar diagonally). Similarly, each number is represented by putting that many of bars in a group.

We can make frequency distribution table by arranging the data in small groups or intervals also.

Group	**Tally Mark**	**Frequency**
0 – 10	||	2

10 – 20	卌 卌 IIII	14
20 – 30	卌 卌 IIII	14
30 – 40	卌 卌	10
40 – 50	卌 III	8

In the video, we have discussed about class limits, class size and class frequency. There is one more term that is used while talking about frequency table. The term is **class mark.**

Class mark is the arithmetic mean of the upper and lower limits of a class. It is also known as the mid value of the class interval.

Therefore,

$$\text{Class mark} = \frac{\text{Lower class limit} + \text{Upper class limit}}{2}$$

Let us now go through the given examples to understand this concept better.

Example 1: The marks obtained by 10 students out of 100 are given below:

55, 79, 68, 85, 96, 48, 39, 67, 80, 72

Find the range of marks.

Solution: From the given marks, we observe that the highest mark is 96 and the lowest mark is 39.

∴ Range of the marks = Highest mark – Lowest mark = 96 – 39 = 57

Example 2: The number of runs scored by a cricket player in 25 innings are given below:

64, 94, 26, 35, 46, 49, 107, 56, 3, 36, 41, 73, 8, 63, 128, 17, 33, 68, 5, 11, 23, 77, 28, 85, 117

Prepare a frequency distribution table, taking the size of the class interval as 20, and answer the following questions:

(i)What are the class intervals of highest and lowest frequency.

(ii)What does the frequency 2 corresponding to the class interval (100 – 120) indicates?

(iii)What is the class mark of the class interval (100 – 120).

(iv)What is the range of the runs scored by the player?

Solution:

The frequency distribution for the given data is as follows:

<table>
<tr><th>Class interval (Runs scored)</th><th>Tally marks</th><th>Frequency</th></tr>
<tr><td>0 – 20</td><td>卌</td><td>5</td></tr>
<tr><td>20 – 40</td><td>卌 |</td><td>6</td></tr>
<tr><td>40 – 60</td><td>||||</td><td>4</td></tr>
<tr><td>60 – 80</td><td>||||</td><td>4</td></tr>
<tr><td>80 – 100</td><td>|||</td><td>3</td></tr>
<tr><td>100 – 120</td><td>||</td><td>2</td></tr>
<tr><td>120 – 140</td><td>|</td><td>1</td></tr>
</table>

(i) Class interval with the highest frequency is 20 – 40 whereas the class interval with the lowest frequency is 120 – 140.

(ii) The frequency 2 in the class interval 100 – 120 indicates that the player has scored runs in the range 100 to 120 twice in 25 innings.

(iii)Class mark of the interval 100 – 120 $= \frac{100+120}{2} = 110$

(iv) Range of the runs scored = Highest run – Lowest run = 128 – 3 = 125

Example 3: Observe the given frequency distribution table and answer the following questions:

Salary per month(Rupees in thousands)	**Number of employees(Frequency)**
15	20
20	35
25	30
30	25
35	20
40	20

45	18
50	12

I. How many employees are involved in the survey?

II. How many employees earn Rs 25,000 per month?

III. What is the difference between the number of employees getting the highest salary and the number of employees getting the lowest salary?

IV. How many employees earn more than Rs 35,000 per month?

V. What is the monthly salary that is being paid to the maximum number of employees?

Solution: I. Number of employees involved in the survey = Sum of all frequencies

∴ Number of employees involved in the survey = 20 + 35 + 30 + 25 + 20 + 20 + 18 + 12

⇒ Number of employees involved in the survey = 180

II. 30 employees earn Rs 25,000 per month.

III. Number of employees getting the highest salary = 12

Number of employees getting the lowest salary = 20

∴ Required difference = 20 – 12

⇒ Required difference = 8

IV. Number of employees earning more than Rs 35,000 per month = Sum of number of employees earning Rs 40,000, Rs 45,000 and Rs 50,000 per month

∴ Number of employees earning more than Rs 35,000 per month = 20 + 18 + 12

⇒ Number of employees earning more than Rs 35,000 per month = 50

V. Highest frequency in the table is 35 which represents the maximum number of employees in any salary group. Also, each employee in this group earns Rs 20,000 per month.

Thus, Rs 20,000 is the monthly salary that is being paid to the maximum number of employees.

Example 4: Observe the given frequency distribution table and then answer the following questions.

Class interval (height in cm)	**Frequency (number of students)**
0 – 12	2
12 – 24	3
24 – 36	5
36 – 48	10
48 – 60	3

1. What is the size of the class intervals?
2. Which class interval has the highest frequency?
3. Which two classes have the same frequency?
4. How many students have height less than 36 cm?
5. What is the lower limit of the class interval 24 – 36?
6. What is the class mark of the class interval 48 – 60?

Solution:

1. The difference between the upper and lower class limits for each class interval is 12. Therefore, the class size is 12.
2. The class 36 – 48 has the highest frequency. 10 students height belong to this category.
3. The classes 12 – 24 and 48 – 60 have the same frequency.
4. The number of students having height less than 36 cm is 2 + 3 + 5 = 10.
5. The lower limit of the class interval 24 – 36 is 24.
6. $\text{Class mark} = \frac{\text{Lower class limit} + \text{Upper class limit}}{2}$

$$\Rightarrow \text{Class mark} = \frac{48+60}{2}$$

$$\Rightarrow \text{Class mark} = \frac{108}{2}$$

$$\Rightarrow \text{Class mark} = 54$$

Organise Data in the Form of Grouped Frequency Distribution Table

The ages of some residents of a particular locality are given as follows.

7, 28, 30, 32, 18, 19, 37, 36, 14, 27, 12, 8, 17, 24, 22, 2, 21, 5, 21, 36, 38, 25, 10, 25, 9.

How will you represent this raw data in a systematic form?

We represent such type of data with the help of ***grouped frequency distribution table.***

Let us now see how to draw it.

There are two ways to group the data to make frequency distribution table. These are as follows:

Inclusive method (Discontinuous form):

In this method, we group the data into small classes of convenient size. Let us take class size as 10 to group the data in different classes. In the above data, minimum value is 2 and maximum value is 38. The classes can be defined in inclusive method as 1 – 10, 11 – 20, 21 – 30 and 31 – 40. Here, both limits are inclusive in each class. Now, a number of residents falling in each group is obtained. All the given observations get covered in these four classes.

Now, frequency distribution table can be drawn as follows:

Class intervals	Tally marks	Frequency
1 – 10	卌 |	6
11 – 20	卌	5
21 – 30	卌 ||||	9
31 – 40	卌	5

Exclusive method (Continuous form):

First of all, we will choose the class intervals. In exclusive method, we take the class intervals as 0 – 10, 10 – 20, 20 – 30, 30 – 40 and obtain the number of residents falling in each group.

Now, the observations which are more than 0 but less than 10 will come under the group 0 – 10; the numbers which are more than 10 but less than 20 will come under the group 10 – 20 and so on.

We must note one thing, 10 occurs in two classes, which are 0 – 10 and 10 – 20. But it is not possible that an observation can be included in both classes. To avoid this, we can make any of lower limit or upper limit inclusive. Here, we adopt the convention that the common observation will belong to the higher class, i.e. 10 will be included in the class interval 10 – 20 and similarly we follow this for the other observations also.

The grouped frequency distribution table will be as follows:

Class intervals	Tally marks	Frequency
1 – 10	~~IIII~~	5
10 – 20	~~IIII~~ I	6
20 – 30	~~IIII~~ III	8
30 – 40	~~IIII~~ I	6

The above frequency distribution tables help to draw many inferences.

We can also tell the frequency, class limits, class size, etc. from the above frequency distribution tables.

The most commonly used method to make frequency distribution table among the above discussed methods is exclusive method.

Class boundaries:
From the table given for inclusive or discontinuous method, it can be observed that there is a gap between the upper limit of a class and the lower limit of its next consecutive class. We can convert this table into a table having continuous classes without altering class size,

class-marks and frequency column. For doing this, we just need to take the average of the upper limit of a class and the lower limit of its next consecutive class. This average is used as the **true upper limit** of that class and **true lower limit** of its next consecutive class.

Therefore,

True upper limit of the class

$= \frac{\text{Upper limit of the class} + \text{Lower limit of the next consecutive class}}{2}$ = True lower limit of the next consecutive class

For example, let us take two consecutive classes such as 1 – 10 and 11 – 20 from the table given for inclusive or discontinuous method.

Now,

True upper limit of the class 1 – 10 = $\frac{10 + 11}{2}$ = 10.5 = True lower limit of the class 11 – 20

In this manner, we obtain the continuous classes as 0.5 – 10.5, 10.5 – 20.5, 20.5 – 30.5 and 30.5 – 40.5.

Note: In this method, true lower limit of first class is obtained by subtracting the value added to its upper limit. Also, true upper limit of last class is obtained by adding the value subtracted from its lower limit.

There is one more method of finding the true upper and lower limits which is explained as follows:

Step 1: Find the difference by subtracting the upper limit of a class from the lower limit of the next consecutive class.

Step 2: Divide the difference by 2.

Step 3: Subtract the difference from the lower limit of each class to find the true lower limit of each class.

Step 4: Add the difference to the upper limit of each class to find the true upper limit of each class.

It can be observed that in the table given for inclusive or discontinuous method, difference between the upper limit of each class and the lower limit of its consecutive class is 1. On dividing this difference by 2, we get 0.5. Now, the continuous classes will be 0.5 – 10.5, 10.5 – 20.5, 20.5 – 30.5 and 30.5 – 40.5.

These methods are very helpful at times.

Now, we know that the range of a data set is the span from lowest value to highest value in the data. We should choose class intervals for a particular range of data very carefully.

Few points to be remembered while choosing class intervals are as follows:

1. Classes should not be overlapping and all values or observations should be covered in these classes.
2. The class size for all classes should be equal.
3. The number of class intervals is normally between five and ten.
4. Class marks and class limits should be taken as integers or simple fractions.

Let us now look at some more examples to understand the concept better.

Example 1: Construct a frequency distribution table for the given data of weekly income of workers by using class intervals as 500 – 525, 525 – 550 and so on. The incomes for the 26 workers for a week are as follows.

540, 530, 545, 510, 520, 580, 570, 575, 555, 516, 527, 560, 550, 525, 535, 535, 565, 575, 585, 580, 560, 510, 515, 510, 520, 525

Solution:

The class intervals to be used are 500 – 525, 525 – 550 and so on. Therefore, the class size is 25. The frequency distribution table will be as follows.

Class intervals	**Tally marks**	**Frequency**
500 – 525	卌 ||	7
525 – 550	卌 |||	8
550 – 575	卌 |	6
575 – 600	卌	5

Example 2: **Observe the following distribution table.**

Class intervals	**Frequency**
0 – 5	2

5 – 10	5
10 – 15	10
15 – 20	2
20 – 25	20
25 – 30	10
30 – 35	50
35 – 40	30

Example 2: **Form a frequency distribution table by taking the class intervals as 0 – 10, 10 – 20 and so on.**

Solution:

Here, in the first class interval 0 – 10, we have to include both the classes (0 – 5 and 5 – 10) of the given table and to find the frequency of class interval (10 – 20), we include the classes 10 – 15 and 15 – 20. In the similar way, we can form the whole table. Thus, the new frequency distribution table will be as follows.

Class intervals	**Frequency**
0 – 10	7
10 – 20	12
20 – 30	30
30 – 40	80