

Frequency Distribution Tables and Graphs

7.0 Introduction

Jagadeesh is watching sports news. A visual appeared on the T.V. screen giving details of the medals won by different countries in Olympics 2012.

Olympics 2012 - Medals Tally

Rank	Country	Gold	Silver	Bronze	Total
1	United States	46	29	29	104
2	China	38	27	23	88
3	Great Britain	29	17	19	65
4	Russia	24	26	32	82
5	Korea	13	8	7	28



The above table provides data about the top five countries that got the highest number of medals in the olympics 2012 as well as the number of medals they won.

Information, available in the numerical form or verbal form or graphical form that helps in taking decisions or drawing conclusions is called **Data**.

- Which country has got the highest number of medals ?
- Which country has got the highest number of bronze medals ?
- Write three more questions based on data provided in the table .



Try This

Give any three examples of data which are in situations or in numbers.

7.1 Basic measures of central tendency

Usually we collect data and draw certain conclusions based on the nature of a data. Understanding its nature, we do certain computations like mean, median and mode which are referred as measures of central tendency. Let us recall.

7.1.1 Arithmetic Mean

It is the most commonly used measure of central tendency. For a set of numbers, the mean is simply the average, i.e., sum of all observations divided by the number of observations.

Arithmetic mean of $x_1, x_2, x_3, x_4, \dots, x_n$ is

$$\text{Arithmetic mean} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{N}$$

$$\bar{x} = \frac{\sum x_i}{N} \text{ (short representation)}$$

$\sum x_i$ represents the sum of all x_i s where i takes the values from 1 to n

Example 1: Ashok got the following marks in different subjects in a unit test. 20, 11, 21, 25, 23 and 14. What is arithmetic mean of his marks?

Solution: Observations = 20, 11, 21, 25, 23 and 14

$$\text{Arithmetic mean } \bar{x} = \frac{\sum x_i}{N}$$

$$= \frac{20 + 11 + 21 + 25 + 23 + 14}{6} = \frac{114}{6}$$

$$\bar{x} = 19$$

Example 2: Arithmetic mean of 7 observations was found to be 32. If one more observation 48 was to be added to the data what would be the new mean of the data?

Solution: Mean of 7 observations $\bar{x} = 32$
 Sum of 7 observations is $\sum x_i = 32 \times 7 = 224$
 Added observation = 48
 Sum of 8 observations $\sum x_i = 224 + 48 = 272$
 \therefore Mean of 8 observations $\bar{x} = \frac{\sum x_i}{N} = \frac{272}{8} = 34$

Example 3: Mean age of 25 members of a club was 38 years. If 5 members with mean age of 42 years have left the club, what is the present mean age of the club members?

Solution: Mean age of 25 members of the club = 38 years

$$\begin{aligned}
 \text{Total age of all the 25 members} &= 38 \times 25 = 950 \\
 \text{Mean age of 5 members} &= 42 \text{ years} \\
 \text{Total age of 5 members} &= 42 \times 5 = 210 \\
 \text{Total age of remaining 20 members} &= 950 - 210 = 740 \\
 \therefore \text{Present mean age of club members } \bar{x} &= \frac{\sum x_i}{N} = \frac{740}{20} = 37 \text{ years}
 \end{aligned}$$

Example 4: Arithmetic mean of 9 observations was calculated as 45. In doing so an observation was wrongly taken as 42 for 24. What would then be the correct mean?

Solution:

$$\begin{aligned}
 \text{Mean of 9 observations} &= 45 \\
 \text{Sum of 9 observations} &= 45 \times 9 = 405 \\
 \text{When computing mean 42 was taken instead of 24} \\
 \therefore \text{Correct sum of 9 observations} &= 405 - 42 + 24 = 387 \\
 \text{Actual mean of 9 observations} &= \frac{\sum x_i}{N} = \frac{387}{9} = 43
 \end{aligned}$$

We observe,

- From the above examples we can see that Arithmetic Mean is a representative value of the entire data.
- Arithmetic mean depends on both number of observations and value of each observation in a data.
- It is unique value of the data.
- When all the observations of the data are increased or decreased by a certain number, the mean also increases or decreases by the same number.
- When all the observations of the data are multiplied or divided by a certain number, the mean also multiplied or divided by the same number.

7.1.2 Arithmetic Mean by Deviation Method

There are five observations in a data, 7, 10, 15, 21, 27. When the teacher asked to estimate the Arithmetic Mean of the data without actual calculation, three students Kamal, Neelima and Lekhya estimated as follows:

Kamal estimated that it lies exactly between minimum and maximum values: 17,

Neelima estimated that it is the middle value of the ordered (ascending or descending) data; 15,

Lekhya added all the observations and divided by their number; 16.

We call each of these estimations as 'estimated mean' or 'assumed mean' is represented with 'A'.

Let us verify which of the estimations coincides with the actual mean.

Case 1: Consider Kamal's estimated arithmetic mean $A = 17$

$$\text{Their actual arithmetic mean is } \bar{x} = \frac{\sum x_i}{N} = \frac{7+10+15+21+27}{5}$$

If each observation is written in terms of deviation from assumed mean A , we have

Score	A	in Terms of Deviations
7	17	$7 = 17 - 10$
10	17	$10 = 17 - 7$
15	17	$15 = 17 - 2$
21	17	$21 = 17 + 4$
27	17	$27 = 17 + 10$

$$\bar{x} = \frac{(17-10)+(17-7)+(17-2)+(17+4)+(17+10)}{5}$$

$$= \frac{5 \times 17}{5} + \frac{-10-7-2+4+10}{5}$$

$$= 17 + \frac{-5}{5} = 17 - 1 = 16$$

\therefore Arithmetic mean = Estimated mean + Average of deviations

Case 2: Consider Neelima's estimated arithmetic mean $A = 15$

$$\text{Their arithmetic mean is } \bar{x} = \frac{\sum x_i}{N} = \frac{7+10+15+21+27}{5}$$

$$\Rightarrow \bar{x} \text{ in terms of deviations} = \frac{(15-8)+(15-5)+(15-0)+(15+6)+(15+12)}{5}$$

$$= \frac{(5 \times 15)}{5} + \frac{(-8-5-0+6+12)}{5}$$

$$= 15 + \frac{5}{5} = 15 + 1 = 16$$

Case 3: Consider Lekhya's estimated arithmetic mean $A = 16$

$$\text{Their arithmetic mean is } \bar{x} = \frac{\sum x_i}{N} = \frac{7+10+15+21+27}{5}$$

$$\Rightarrow \bar{x} \text{ in terms of deviations} = \frac{(16-9)+(16-6)+(16-1)+(16+5)+(16+11)}{5}$$

$$= \frac{(5 \times 16)}{5} + \frac{(-9-6-1+5+11)}{5}$$

$$= 16 + \frac{0}{5} = 16$$



Try These

Prepare a table of estimated mean, deviations of the above cases. Observe the average of deviations with the difference of estimated mean and actual mean. What do you infer?

[Hint : Compare with average deviations]

It is clear that the estimated mean becomes the actual arithmetic mean if the sum (or average) of deviations of all observations from the estimated mean is 'zero'.

We may use this verification process as a means to find the Arithmetic Mean of the data.

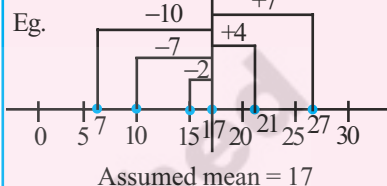
From the above cases, it is evident that the arithmetic mean may be found through the estimated mean and deviation of all observations from it.

Arithmetic mean = Estimated mean + Average of deviations

$$= \text{Estimated mean} + \frac{\text{Sum of deviations}}{\text{Number of observations}}$$

$$\bar{x} = A + \frac{\sum (x_i - A)}{N}$$

The difference between any score of data and assumed mean is called deviations.



Example 5: Find the arithmetic mean of 10 observations 14, 36, 25, 28, 35, 32, 56, 42, 50, 62 by assuming mean as 40. Also find mean by regular formula. Do you find any difference.

Solution: Observations of the data = 14, 25, 28, 32, 35, 36, 42, 50, 56, 62

Let the assumed mean is $A = 40$

$$\therefore \text{Arithmetic mean} = A + \frac{\sum (x_i - A)}{N}$$

$$\begin{aligned} \bar{x} &= 40 + \frac{(14 - 40) + (25 - 40) + (28 - 40) + (32 - 40) + (35 - 40) + (36 - 40) + (42 - 40) + (50 - 40) + (56 - 40) + (62 - 40)}{10} \\ &= 40 + \frac{(-26) + (-15) + (-12) + (-8) + (-5) + (-4) + (2) + (10) + (16) + (22)}{10} \\ &= 40 + \frac{(-70 + 50)}{10} \\ &= 40 - \frac{20}{10} \\ &= 40 - 2 = 38 \end{aligned}$$

$$\begin{aligned} \text{By usual formula } \bar{x} &= \frac{\sum x_i}{N} = \frac{14 + 25 + 28 + 32 + 35 + 36 + 42 + 50 + 56 + 62}{10} \\ &= \frac{380}{10} = 38 \end{aligned}$$

In both the methods we got the same mean.

This way of computing arithmetic mean by deviation method is conveniently used for data with large numbers and decimal numbers.

Consider the following example.

Example 6: Market value (in rupees) of a share through a week is changing as 3672, 3657, 3673, 3665, 3668. Find the arithmetic mean of the market value of the share.

Solution: Observations of the data = 3657, 3665, 3668, 3672, 3673

Estimated mean = 3668

$$\begin{aligned} \text{Arithmetic mean } \bar{x} &= A + \frac{\sum(x_i - A)}{N} \\ &= 3668 + \frac{(3657 - 3668) + (3665 - 3668) + (3668 - 3668) + (3672 - 3668) + (3673 - 3668)}{5} \\ &= 3668 + \frac{(-11 - 3 - 0 + 4 + 5)}{5} = 3668 + \frac{(-5)}{5} = 3668 - 1 = ₹ 3667. \end{aligned}$$



Try These

1. Estimate the arithmetic mean of the following data
 - (i) 17, 25, 28, 35, 40
 - (ii) 5, 6, 7, 8, 8, 10, 10, 10, 12, 12, 13, 19, 19, 19, 20

Verify your answers by actual calculations.

Project work

1. Collect marks of 10 of your classmates in different subjects in the recent examinations. Estimate the arithmetic mean of marks in each subject and verify them by actual calculations. How many of your estimations represent exact mean?
2. Measure the heights of students of your class and estimate the mean height. Verify their mean from records of your physical education teacher. Do you notice any difference?

7.1.3 Median

Median is another frequently used measure of central tendency. The median is simply the middle term of the distribution when it is arranged in either ascending or descending order, i.e. there are as many observations above it as below it.

If **n** number of observations in the data arranged in ascending or descending order

- When **n** is odd, $\left(\frac{n+1}{2}\right)^{\text{th}}$ observation is the median.

- When n is even, arithmetic mean of two middle observations $\left(\frac{n}{2}\right)^{\text{th}}$ and $\left(\frac{n}{2}+1\right)^{\text{th}}$ is the median of the data.

Example 7: Find the median of 9 observations 14, 36, 25, 28, 35, 32, 56, 42, 50.

Solution: Ascending order of the data = 14, 25, 28, 32, 35, 36, 42, 50, 56

No of observations $n = 9$ (odd number)

$$\begin{aligned}\text{Median of the data} &= \left(\frac{n+1}{2}\right)^{\text{th}} \text{ observation} \\ &= 5^{\text{th}} \text{ observation} = 35\end{aligned}$$

$$\therefore \text{Median} = 35$$

Example 8: If another observation 61 is also included to the above data what would be the median?

Solution: Ascending order of the data = 14, 25, 28, 32, 35, 36, 42, 50, 56, 61

No of observations $n = 10$ (even number)

Then there would be two numbers at the middle of the data.

$$\begin{aligned}\text{Median of the data} &= \text{arithmetic mean of } \left(\frac{n}{2}\right)^{\text{th}} \text{ and } \left(\frac{n}{2}+1\right)^{\text{th}} \text{ observations} \\ &= \text{arithmetic mean of } 5^{\text{th}} \text{ and } 6^{\text{th}} \text{ observations} \\ &= \frac{35+36}{2} = 35.5\end{aligned}$$

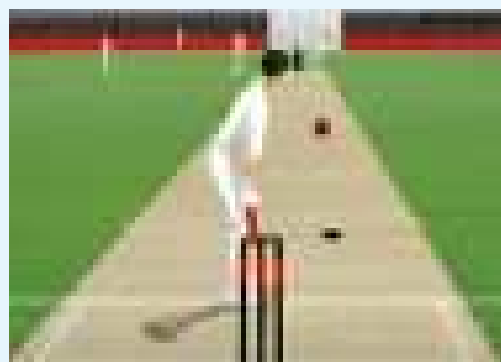


Do This

Here are the heights of some of Indian cricketers. Find the median height of the team.

S.No.	Players Name	Heights
1.	VVS Laxman	5'11"
2.	Parthiv Patel	5'3"
3.	Harbhajan Singh	6'0"
4.	Sachin Tendulkar	5'5"
5.	Gautam Gambhir	5'7"
6.	Yuvraj Singh	6'1"
7.	Robin Uthappa	5'9"
8.	Virender Sehwag	5'8"
9.	Zaheer Khan	6'0"
10.	MS Dhoni	5'11"

5' 10" means 5 feet 10 inches



Note :

- Median is the middle most value in ordered data.
- It depends on number of observations and middle observations of the ordered data. It is not effected by any change in extreme values.

**Try These**

1. Find the median of the data 24,65,85,12,45,35,15.
2. If the median of $x, 2x, 4x$ is 12, then find mean of the data.
3. If the median of the data 24, 29, 34, 38, x is 29 then the value of ' x ' is
(i) $x > 38$ (ii) $x < 29$ (iii) x lies in between 29 and 34 (iv) none

7.1.4 Mode

When we need to know what is the favourite uniform colour in a class or most selling size of the shirt in shop, we use mode. The mode is simply the most frequently occurring value. Consider the following examples.

Example 9: In a shoe mart different sizes (in inches) of shoes sold in a week are; 7, 9, 10, 8, 7, 9, 7, 9, 6, 3, 5, 5, 7, 10, 7, 8, 7, 9, 6, 7, 7, 7, 10, 5, 4, 3, 5, 7, 8, 7, 9, 7. Which size of the shoes must be kept more in number for next week to sell? Give the reasons.

Solution: If we write the observations in the data in order we have
3, 3, 4, 5, 5, 5, 5, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 9, 10, 10, 10.

From the data it is clear that 7 inch size shoes are sold more in number. Thus the mode of the given data is 7. So 7 inch size shoes must be kept more in number for sale.

Example 10: The blood group of 50 donors, participated in blood donation camp are A, AB, B, A, O, AB, O, O, A, AB, B, A, O, AB, O, O, A, B, A, O, AB, O, O, A, AB, B, O, AB, O, B, A, O, AB, O, O, A, AB, B, A, O, AB, O, A, AB, B, A, O, AB, O, O. Find the mode of the above verbal data.

Solution: By observing the data we can find that A group is repeated for 12, B group is repeated for 7, AB group is repeated for 12, O group is repeated for 19 times.
 \therefore Mode of the data is = 'O' group.

Think, Discuss and Write

Is their any change in mode, if one or two more observations, equal to mode are included in the data?

Note :

- Mode is the most frequent observation of the given data.
- It depends neither on number of observations nor values of all observations.
- It is used to analyse both numerical and verbal data.
- There may be 2 or 3 or many modes for the same data.

**Exercise - 7.1**

1. Find the arithmetic mean of the sales per day in a fair price shop in a week.
₹10000, ₹.10250, ₹.10790, ₹.9865, ₹.15350, ₹.10110
2. Find the mean of the data; 10.25, 9, 4.75, 8, 2.65, 12, 2.35
3. Mean of eight observations is 25. If one observation 11 is excluded, find the mean of the remaining.
4. Arithmetic mean of nine observations is calculated as 38. But in doing so, an observation 27 is mistaken for 72. Find the actual mean of the data.
5. Five years ago mean age of a family was 25 years. What is the present mean age of the family?
6. Two years ago the mean age of 40 people was 11 years. Now a person left the group and the mean age is changed to 12 years. Find the age of the person who left the group.
7. Find the sum of deviations of all observations of the data 5, 8, 10, 15, 22 from their mean.
8. If sum of the 20 deviations from the mean is 100, then find the mean deviation.
9. Marks of 12 students in a unit test are given as 4, 21, 13, 17, 5, 9, 10, 20, 19, 12, 20, 14. Assume a mean and calculate the arithmetic mean of the data. Assume another number as mean and calculate the arithmetic mean again. Do you get the same result? Comment.
10. Arithmetic mean of marks (out of 25) scored by 10 students was 15. One of the student, named Karishma enquired the other 9 students and find the deviations from her marks are noted as $-8, -6, -3, -1, 0, 2, 3, 4, 6$. Find Karishma's marks.
11. The sum of deviations of ' n ' observations from 25 is 25 and sum of deviations of the same ' n ' observations from 35 is -25 . Find the mean of the observations.
12. Find the median of the data; 3.3, 3.5, 3.1, 3.7, 3.2, 3.8
13. The median of the following observations, arranged in ascending order is 15.
10, 12, 14, $x - 3$, x , $x + 2$, 25. Then find x .
14. Find the mode of 10, 12, 11, 10, 15, 20, 19, 21, 11, 9, 10.

15. Mode of certain scores is x . If each score is decreased by 3, then find the mode of the new series.
16. Find the mode of all digits used in writing the natural numbers from 1 to 100.
17. Observations of a raw data are 5, 28, 15, 10, 15, 8, 24. Add four more numbers so that mean and median of the data remain the same, but mode increases by 1.
18. If the mean of a set of observations $x_1, x_2, \dots, \dots, x_{10}$ is 20. Find the mean of $x_1 + 4, x_2 + 8, x_3 + 12, \dots, \dots, x_{10} + 40$.
19. Six numbers from a list of nine integers are 7, 8, 3, 5, 9 and 5. Find the largest possible value of the median of all nine numbers in this list.
20. The median of a set of 9 distinct observations is 20. If each of the largest 4 observations of the set is increased by 2, find the median of the resulting set.

7.2 Organisation of Grouped Data

We have learnt to organize smaller data by using tally marks in previous class. But what happens if the data is large? We organize the data by dividing it into convenient groups. It is called grouped data. Let us observe the following example.

A construction company planned to construct various types of houses for the employees based on their income levels. So they collected the data about monthly net income of the 100 employees, who wish to have a house. They are (in rupees) 15000, 15750, 16000, 16000, 16050, 16400, 16600, 16800, 17000, 17250, 17250..... 75000.

This is a large data of 100 observations, ranging from ₹ 15000 to ₹ 75000. Even if we make frequency table for each observation the table becomes large. Instead the data can be classified into small income groups like 10001 – 20000, 20001 – 30000, . . . , 70001 – 80000.

These small groups are called ‘class intervals’. The intervals 10001 – 20000 has all the observations between 10001 and 20000 including both 10001 and 20000. This form of class interval is called ‘inclusive form’, where 10001 is the ‘lower limit’, 20000 is the ‘upper limit’.

7.2.1 Interpretation of Grouped frequency distribution:

Example 11: Marks of 30 students in mathematics test are given in the adjacent grouped frequency distribution.

- (i) Into how many groups the data is classified?

Sl. No	Marks	No of Students
1	0 – 5	5
2	5 – 10	7
3	10 – 15	10
4	15 – 20	6
5	20 – 25	2

- (ii) How many students are there in the third group?
- (iii) If a student gets 10 marks, should he be included in 2nd or 3rd class?
- (iv) What are the marks of 6 students who are in 4th class interval?
- (v) What are the individual marks of 2 students in the fifth group?

Answers

- (i) The data is classified into 5 groups or 5 classes.
- (ii) There are 10 students in the third group.
- (iii) Here 10 is the upper limit of 2nd class and lower limit of 3rd class. In such case upper limit is not included in the class. So 10 is included in the 3rd class interval.
- (iv) Marks of 6 students in 4th class interval varies from 15 and below 20.
- (v) Individual marks of students can't be identified from this frequency distribution, they may be from 20 and below 25.



Do This

Ages of 90 people in an apartment are given in the adjacent grouped frequency distribution

- (i) How many Class Intervals are there in the table?
- (ii) How many people are there in the Class Interval 21-30?

(iii) Which age group people are more in that apartment?

(iv) Can we say that both people the last age group (61-70) are of 61, 70 or any other age?

Ages	No of People
1 – 10	15
11 – 20	14
21 – 30	17
31 – 40	20
41 – 50	18
51 – 60	4
61 – 70	2

7.2.2 Limits and Boundaries

Suppose we have to organize a data of marks in a test. We make class intervals like 1-10, 11-20, If a student gets 10.5 marks, where does it fall? In class 1-10 or 11-20 ? In this situation we make use of real limits or boundaries.

Consider the class intervals shown in the adjacent table.

Class Intervals	
Limits	Boundaries
1 – 10	0.5 - 10.5
11 – 20	10.5 – 20.5
21 – 30	20.5 – 30.5
31 – 40	30.5 – 40.5

- Average of Upper Limit (UL) of first class and Lower Limit (LL) of second class becomes the Upper Boundary (UB) of the first class and Lower Boundary (LB) of the second class.
i.e., Average of 10, 11; $\frac{10+11}{2} = 10.5$ is the boundary.
- Now all the observations below 10.5 fall into group 1-10 and the observations from 10.5 to below 20.5 will fall into next class i.e 11-20 having boundaries 10.5 to 20.5. Thus 10.5 falls into class interval of 11-20.
- Imagine the UL of the previous class interval (usually zero) and calculate the LB of the first class interval. Average of 0, 1 is $\frac{0+1}{2} = 0.5$ is the LB.
- Similarly imagine the LL of the class after the last class interval and calculate the UB of the last class interval. Average of 40, 41 is $\frac{40+41}{2} = 40.5$ is the UB.

These boundaries are also called “true class limits”.

Observe limits and boundaries for the following class intervals.

Class interval	Limits		Boundaries	
	Lower limit	Upper limit	Lower boundary	Upper boundary
1-10	1	10	0.5	10.5
11-20	11	20	10.5	20.5
21-30	21	30	20.5	30.5

Class interval	Limits		Boundaries	
	Lower limit	Upper limit	Lower boundary	Upper boundary
0-10	0	10	0	10
10-20	10	20	10	20
20-30	20	30	20	30

There in the above illustration we can observe that in case of discrete series (Inclusive class intervals) limit and boundaries are different. But in case of continuous series (exclusive class

intervals) limits and boundaries are the same. Difference between upper and lower boundaries of a class is called ‘class length’, represented by ‘C’.



Do These

1. Long jump made by 30 students of a class are tabulated as

Distance (cm)	101 – 200	201 – 300	301 – 400	401 – 500	501 – 600
No of students	4	7	15	3	1

- I. Are the given class intervals inclusive or exclusive?
 - II. How many students are in second class interval?
 - III. How many students jumped a distance of 3.01m or more ?
 - IV. To which class interval does the student who jumped a distance of 4.005 m belongs?
2. Calculate the boundaries of the class intervals in the above table.
 3. What is the length of each class interval in the above table?

7.2.3 Construction of grouped frequency Distribution

Consider the marks of 50 students in Mathematics secured in Quarterly examination 31, 14, 0, 12, 20, 23, 26, 36, 33, 41, 37, 25, 22, 14, 3, 25, 27, 34, 38, 43, 32, 22, 28, 18, 7, 21, 20, 35, 36, 45, 9, 19, 29, 25, 33, 47, 35, 38, 25, 34, 38, 24, 39, 1, 10, 24, 27, 25, 18, 8.

After seeing the data, you might be thinking, into how many intervals the data could be classified? How frequency distribution table could be constructed?

The following steps help in construction of grouped frequency distribution.

Step1: Find the range of the data.

$$\begin{aligned} \text{Range} &= \text{Maximum value} - \text{Minimum value} \\ &= 47 - 0 = 47 \end{aligned}$$

Step2: Decide the number of class intervals. (Generally number of class intervals are 5 to 8)

If no of class intervals = 6

$$\Rightarrow \text{Length of the class interval} = \frac{47}{6} \approx 8 \text{ (approximately)}$$

Class Intervals (Marks)	Tally Marks	Frequency (No of students)
0 – 7		4
08 – 15		6
16 – 23		9
24 – 31		13
32 – 39		14
40 – 47		4

Step 3: Write inclusive class intervals starting from minimum value of observations.

i.e 0-7, 8-15 and so on...

Step 4: Using the tally marks distribute the observations of the data into respective class intervals.

Step 5: Count the tally marks and write the frequencies in the table.

Now construct grouped frequency distribution table for exclusive classes.

Think, Discuss and Write



1. Make a frequency distribution of the following series. 1, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7.
2. Construct a frequency distribution for the following series of numbers.
2, 3, 4, 6, 7, 8, 9, 9, 11, 12, 12, 13, 13, 13, 14, 14, 14, 15, 16, 17, 18, 18, 19, 20, 20, 21, 22, 24, 24, 25. (Hint: Use inclusive classes)
3. What are the differences between the above two frequency distribution tables?
4. From which of the frequency distributions we can write the raw data again?

7.2.4 Characteristics of Grouped Frequency Distribution

1. It divides the data into convenient and small groups called 'class intervals'.
2. In a class interval 5-10, 5 is called lower limit and 10 is called upper limit.
3. Class intervals like 1-10, 11-20, 21-30 ... are called inclusive class intervals, because both lower and upper limits of a particular class belong to that particular class interval.
4. Class intervals like 0-10, 10-20, 20-30 ... are called exclusive class intervals, because only lower limit of a particular class belongs to that class, but not its upper limit.
5. Average of upper limit of a class and lower limit of the next class is called upper bound of the first class and lower bound of the next class.
6. In exclusive class intervals, both limits and boundaries are equal but in case of inclusive class intervals limits and boundaries are not equal.
7. Difference between upper and lower boundaries of a class is called 'length of the class'.
8. Individual values of all observations can't be identified from this table, but value of each observation of a particular class is assumed to be the average of upper and lower boundaries of that class. This value is called 'class mark' or 'mid value' (\bar{x}).

Example 12: The following marks achieved by 30 candidates in mathematics of SSC examination held in the year 2010.

45, 56, 75, 68, 35, 69, 98, 78, 89, 90, 70, 56, 59, 35, 46, 47, 13, 29, 32, 39, 93, 84, 76, 79, 40, 54, 68, 69, 60, 59. Construct the frequency distribution table with the class intervals ; failed (0 – 34), third class (35 – 49), second class (50 – 60), first class (60 – 74) and distinction (75 – 100).

Solution: Class intervals are already given. So proceed from step 3

Step 3: Write class intervals as given.

Step 4: These are inclusive class intervals. Recall that upper limits also belong to the class. Using the tally marks, distribute the observations of the data into different class intervals.

Step 5: Count the tally marks and write the frequencies in the table.

Class Intervals (Marks)	Tally Marks	Frequency (No of students)
0 – 34		3
35 – 49		7
50 – 59		5
60 – 74		6
75 – 100		9

(Note : The lengths of class intervals are not same in this case)

Example 13: A grouped frequency distribution table is given below with class mark (mid values of class intervals) and frequencies. Find the class intervals.

Class marks	7	15	23	31	39	47
Frequency	5	11	19	21	12	6

Solution: We know that class marks are the mid values of class intervals. That implies class boundaries lie between every two successive class marks.

Step 1: Find the difference between two successive class marks; $h = 15 - 7 = 8$.

(Find whether difference between every two successive classes is same)

Step 2: Calculate lower and upper boundaries of every class with class mark 'x', as $x - h/2$ and $x + h/2$

For example boundaries of first class are $7 - \frac{8}{2} = 3$ or $7 + \frac{8}{2} = 11$

Class Marks	Class intervals	Frequency
7	$(7 - 4) - (7 + 4) = 03 - 11$	5
15	$(15 - 4) - (15 + 4) = 11 - 19$	11
23	$(23 - 4) - (23 + 4) = 19 - 27$	19
31	$(31 - 4) - (31 + 4) = 27 - 35$	21
39	$(39 - 4) - (39 + 4) = 35 - 43$	12
47	$(47 - 4) - (47 + 4) = 43 - 51$	6

7.3 Cumulative Frequency

In a competitive examination 1000 candidates appeared for a written test. Their marks are announced in the form of grouped frequency distribution as shown in the adjacent table.

Two candidates Sarath, Sankar are looking at the table and discussing like ...

Sarath: How many candidates have appeared for the test?

Sankar: It seems 1000 candidates appeared for the test.

Sarath: See, 360 candidates achieved 50-60 marks.

Sankar: If 60 is the cut off mark, how many candidates are eligible to get call letter?

Sarath: Do you mean how many got 60 and above marks in altogether?

Sankar: It is $50 + 25 + 10 + 5$, that is 90 candidates will be eligible.

Sarath: But there are only 105 jobs. How many candidates are eligible, if cut off mark as 50.

Sankar: In that case, $360 + 50 + 25 + 10 + 5$, that is totally 450 candidates are eligible to get call letter for interview.

Similarly we can make some more conclusions.

Number of candidates, who got equal or more than 90 (Lower boundary) = 5

Number of candidates, who got equal or more than LB of ninth CI = $10 + 5 = 15$

Number of candidates, who got equal or more than LB of eighth CI = $25 + 15 = 40$

Number of candidates, who got equal or more than LB of seventh CI = $50 + 40 = 90$

Class Interval (Marks)	No of Candidates
0 - 10	25
10 - 20	45
20 - 30	60
30 - 40	120
40 - 50	300
50 - 60	360
60 - 70	50
70 - 80	25
80 - 90	10
90 - 100	5

We are getting these values by taking progressive total of frequencies from either the first or last class to the particular class. These are called cumulative frequencies. The progressive sum of frequencies from the last class of the to the lower boundary of particular class is called 'Greater than Cumulative Frequency' (G.C.F.).

Watch out how we can write these greater than cumulative frequencies in the table.

1. Frequency in last class interval itself is greater than cumulative frequency of that class.
2. Add the frequency of the ninth class interval to the greater than cumulative frequency of the tenth class interval to give the greater than cumulative frequency of the ninth class interval
3. Successively follow the same procedure to get the remaining greater than cumulative frequencies.

Class Interval (Marks)	LB	Frequency (No of Candidates)	Greater than cumulative frequency
0 – 10	0	25	$25+975 = 1000$
10 – 20	10	45	$45+930 = 975$
20 – 30	20	60	$60+870 = 930$
30 – 40	30	120	$120+750 = 870$
40 – 50	40	300	$300+450 = 750$
50 – 60	50	360	$360+ 90 = 450$
60 – 70	60	50	$50 + 40 = 90$
70 – 80	70	25	$25 \rightarrow 25 + 15 = 40$
80 – 90	80	10	$10 \rightarrow 10 + 5 = 15$
90 – 100	90	5	$5 \rightarrow 5$

The distribution that represent lower boundaries of the classes and their respective Greater than cumulative frequencies is called Greater than Cumulative Frequency Distribution .

Similarly in some cases we need to calculate less than cumulative frequencies.

For example if a teacher wants to give some extra support for those students, who got less marks than a particular level, we need to calculate the less than cumulative frequencies.

Thus the progressive total of frequencies from first class to the upper boundary of a particular class is called Less than Cumulative Frequency (L.C.F.).

Class Interval (Marks)	UB	No of Candidates	Less than cumulative frequency
0 – 5	5	7	$7 \rightarrow 7$
5 – 10	10	10	$10 \rightarrow 10+7 = 17$
10 – 15	15	15	$15+17 = 32$
15 – 20	20	8	$8+32 = 40$
20 – 25	25	3	$3+40 = 43$

Consider the grouped frequency distribution expressing the marks of 43 students in a unit test.

1. Frequency in first class interval is directly written into less than cumulative frequency.
2. Add the frequency of the second class interval to the less than cumulative frequency of the first class interval to give the less than cumulative frequency of the second class interval

3. Successively follow the same procedure to get remaining less than cumulative frequencies.

The distribution that represents upper boundaries of the classes and their respective less than cumulative frequencies is called Less than Cumulative Frequency Distribution.



Try These

1. Less than cumulative frequency is related to _____
2. Greater than cumulative frequency is related to _____
3. Write the Less than and Greater than cumulative frequencies for the following data

Class Interval	1 -10	11 - 20	21 - 30	31 - 40	41 - 50
Frequency	4	7	12	5	2

4. What is total frequency and less than cumulative frequency of the last class above problem? What do you infer?

Example 14: Given below are the marks of students in a less than cumulative frequency distribution table.. Write the frequencies of the respective classes. Also write the Greater than cumulative frequencies. How many students' marks are given in the table?

Class Interval (Marks)	1 -10	11 - 20	21 - 30	31 - 40	41 - 50
L.C.F. (No of students)	12	27	54	67	75

Solution:

Class Interval (Marks)	L.C.F.	Frequency (No of students)	G.C.F.
1 – 10	12	12	$12 + 63 = 75$
11 – 20	27	$27 - 12 = 15$	$15 + 48 = 63$
21 – 30	54	$54 - 27 = 27$	$27 + 21 = 48$
31 – 40	67	$67 - 54 = 13$	$13 + 8 = 21$
41 – 50	75	$75 - 67 = 8$	8

Total number of students mentioned in the table is nothing but total of frequencies or less than cumulative frequency of the last class or greater than cumulative frequency of the first class interval, i.e. 75.



Exercise - 7.2

1. Given below are the ages of 45 people in a colony.

33 8 7 25 31 26 5 50 25 48 56
 33 28 22 15 62 59 16 14 19 24 35
 26 9 12 46 15 42 63 32 5 22 11
 42 23 52 48 62 10 24 43 51 37 48
 36

Construct grouped frequency distribution for the given data with 6 class intervals.

2. Number of students in 30 class rooms in a school are given below. Construct a frequency distribution table for the data with a exclusive class interval of 4 (students).

25 30 24 18 21 24 32 34 22 20 22
 32 40 28 30 22 26 31 34 15 38 28
 20 16 15 20 24 30 25 18

3. Class intervals in a grouped frequency distribution are given as 4 – 11, 12 – 19, 20 – 27, 28 – 35, 36 – 43. Write the next two class intervals. (i) What is the length of each class interval? (ii) Write the class boundaries of all classes, (iii) What are the class marks of each class?

4. In the following grouped frequency distribution table class marks are given.

Class Marks	10	22	34	46	58	70
Frequency	6	14	20	21	9	5

- (i) Construct class intervals of the data. (Exclusive class intervals)
 (ii) Construct less than cumulative frequencies and
 (iii) Construct greater than cumulative frequencies.
5. The marks obtained by 35 students in a test in statistics (out of 50) are as below.

35 1 15 35 45 23 31 40 21 13 15
 20 47 48 42 34 43 45 33 37 11 13
 27 18 12 37 39 38 16 13 18 5 41
 47 43

Construct a frequency distribution table with equal class intervals, one of them being 10-20 (20 is not included).

6. Construct the class boundaries of the following frequency distribution table. Also construct less than cumulative and greater than cumulative frequency tables.

Ages	1 - 3	4 - 6	7 - 9	10 - 12	13 - 15
No of children	10	12	15	13	9

7. Cumulative frequency table is given below. Which type of cumulative frequency is given. Try to build the frequencies of respective class intervals.

Runs	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
No of cricketers	3	8	19	25	30

8. Number of readers in a library are given below. Write the frequency of respective classes. Also write the less than cumulative frequency table.

Number of books	1-10	11-20	21-30	31-40	41-50
Greater than Cumulative frequency	42	36	23	14	6

7.4 Graphical Representation of Data:

Frequency distribution is an organised data with observations or class intervals with frequencies. We have already studied how to represent of discrete series in the form of pictographs, bar graphs, double bar graph and pie charts.

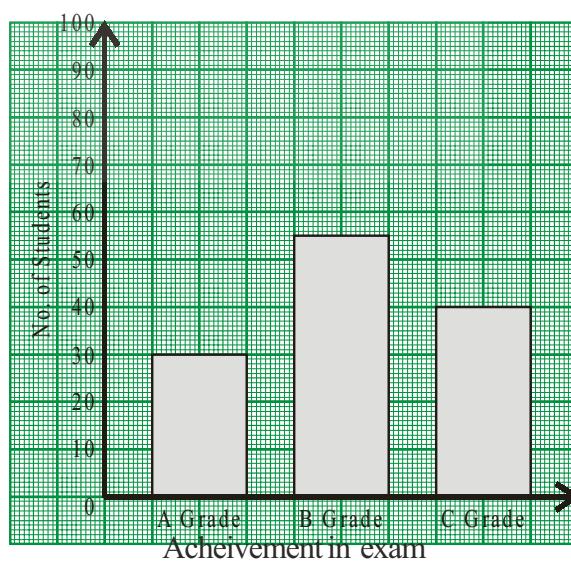
Let us recall bar graph first.

7.4.1 Bar Graph

A display of information using vertical or horizontal bars of uniform width and different lengths being proportional to the respective values is called a bar graph.

Let us see what a bar graph can represent. Study the following vertical bar graph.

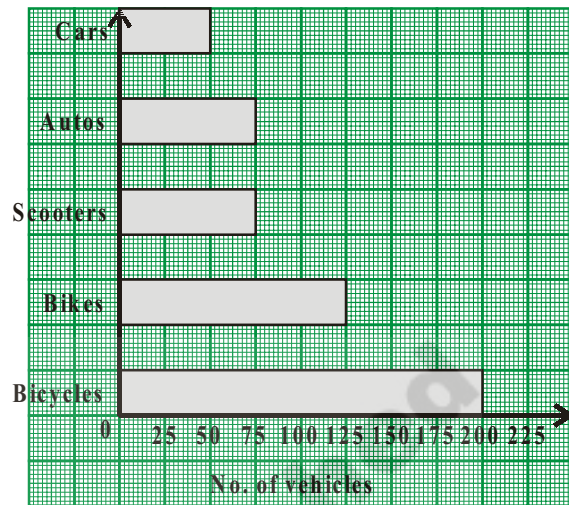
- What does this bar graph represent?
- How many students secured A, B or C grades?



- (iii) Which grade is secured by more number of the students?
- (iv) How many students are there in the class?

It is easy to answer the questions from the graph.

Similarly in some graphs bars may be drawn horizontally. For example observe the second bar graph. It gives the data about number of vehicles in a village Sangam in Nellore district.



Think, Discuss and Write



- All the bars (or rectangles) in a bar graph have
 - same length
 - same width
 - same area
 - equal value
- Does the length of each bar depend on the lengths of other bars in the graphs?
- Does the variation in the value of a bar affect the values of other bars in the same graph?
- Where do we use vertical bar graphs and horizontal bar graphs.

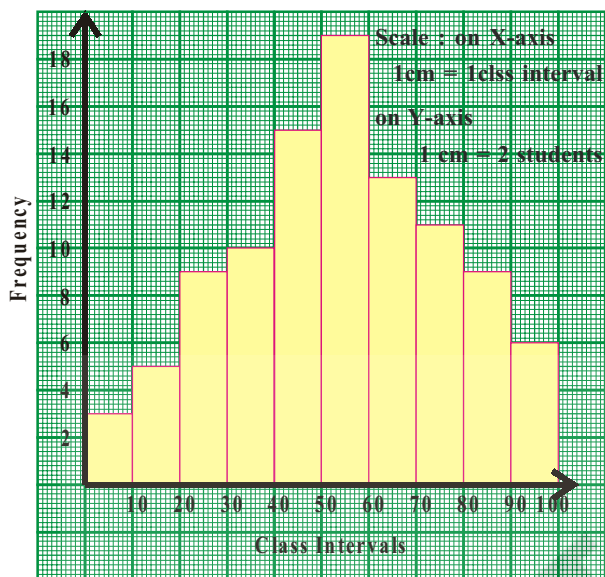
7.5 Graphical Representation of Grouped Frequency Distribution

Let us learn the graphical representation of grouped frequency distributions of continuous series i.e. with exclusive class intervals. First one of its kind is histogram.

7.5.1 Histogram

7.5.1.1 Interpretation of Histogram:

Observe the following histogram for the given grouped frequency distribution.



Class Interval (Marks)	Frequency (No of Students)
0 – 10	3
10 – 20	5
20 – 30	9
30 – 40	10
40 – 50	15
50 – 60	19
60 – 70	13
70 – 80	11
80 – 90	9
90 -100	6

- (i) How many bars are there in the graph?
- (ii) In what proportion the height of the bars are drawn?
- (iii) Width of all bars is same. What may be the reason?
- (iv) Shall we interchange any two bars of the graph?

From the graph you might have understood that

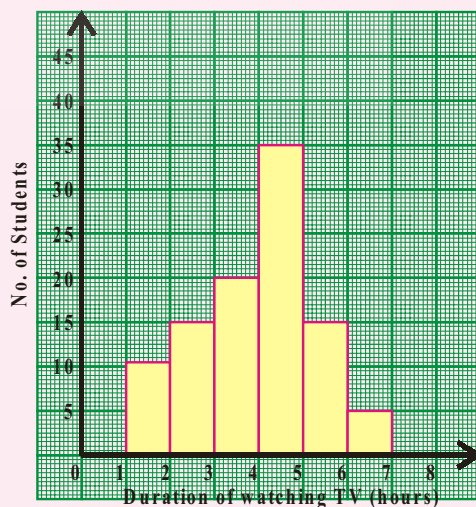
- (i) There are 10 bars representing frequencies of 10 class intervals.
- (ii) Heights of the bars are proportional to the frequencies,
- (iii) Width of bars is same because width represents the class interval. Particularly in this example length of all class intervals is same.
- (iv) As it is representing a continuous series, (with exclusive class intervals), we can't interchange any two bars.



Try These

Observe the adjacent histogram and answer the following questions-

- (i) What information is being represented in the histogram?
- (ii) Which group contains maximum number of students?
- (iii) How many students watch TV for 5 hours or more?
- (iv) How many students are surveyed in total?



7.5.1.2 Construction of a Histogram

A TV channel wants to find which age group of people are watching their channel. They made a survey in an apartment. Represent the data in the form of a histogram.

Step 1 : If the class intervals given are inclusive (limits) convert them into the exclusive form (boundaries) since the histogram has to be drawn for a continuous series.

Class Interval (Age group)	Frequency (No of viewers)	Class Intervals
11 – 20	10	10.5 – 20.5
21 – 30	15	20.5 – 30.5
31 – 40	25	30.5 – 40.5
41 – 50	30	40.5 – 50.5
51 – 60	20	50.5 – 60.5
61 – 70	5	60.5 – 70.5
Limits		Boundaries

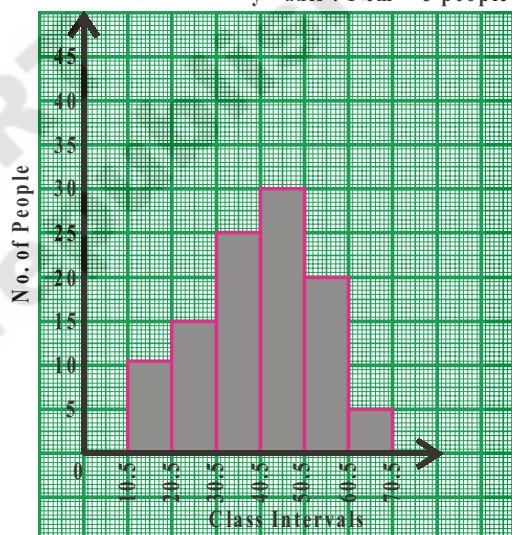
Step 2 : Choose a suitable scale on the X-axis and mark the class intervals on it.

Step 3 : Choose a suitable scale on the Y-axis and mark the frequencies on it. (The scales on both the axes may not be same)

Scale : X-axis 1 cm = one class interval
Y-axis 1 cm = 5 people

Step 4 : Draw rectangles with class intervals as bases and the corresponding frequencies as the corresponding heights.

Scale :
x - axis : 1 cm = 1 class interval
y - axis : 1 cm = 5 people



7.5.1.3 Histogram with Varying Base Widths

Consider the following frequency distribution table.

Category	Class Intervals (Marks)	Percentage of Students
Failed	0-35	28
Third Class	35-50	12
Second Class	50-60	16
First Class	60-100	44

You have noticed that for different categories of children performance the range of marks for each category is not uniform.

If we observe the table, the students who secured first class is 44 % which spreads over the class length 40 (60 to 100). Whereas the student who have secured second class is 16% of the

students spread over the class length 10 (50 to 60) only. Therefore to represent the above distribution table into histogram we have take the widths of class intervals also into account.

In such cases frequency per unit class length (frequency density) has to be calculated and histogram has to be constructed with respective heights. Any class interval may be taken as unit class interval for calculating frequency density. For convenience least class length is taken as unit class length.

∴ Modified length of any rectangle is proportional to the corresponding frequency

$$\text{Density} = \frac{\text{Frequency of class}}{\text{Length of that class}} \times \text{Least class length}$$

Class intervals (Marks)	Percentage of students	Class length	Length of the rectangle
0 – 35	28	35	$\frac{28}{35} \times 10 = 8$
35 – 50	12	15	$\frac{12}{15} \times 10 = 8$
50 – 60	16	10	$\frac{16}{10} \times 10 = 16$
60 – 100	44	40	$\frac{44}{40} \times 10 = 11$

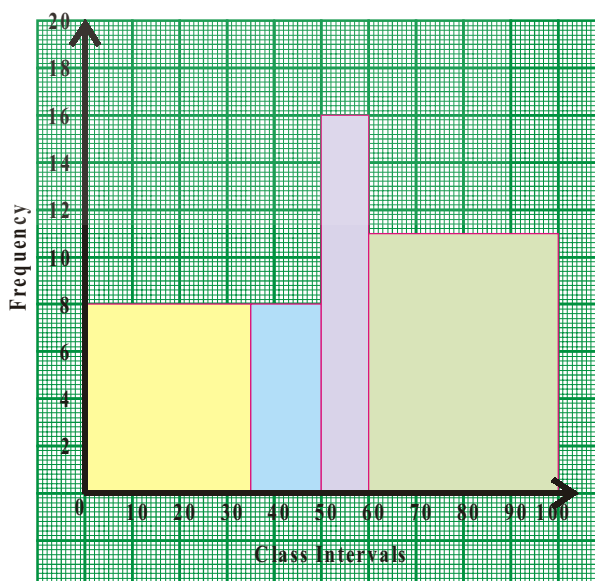
With the modified lengths histogram has to be constructed as in the previous example.

Step 1: Choose a suitable scale on the X-axis and mark the class intervals on it.

Step 2: Choose a suitable scale on the Y-axis and mark the frequencies on it. (The scales on both the axes may not be same)

Scale: X-axis 1 cm = 1 Min. class interval
Y-axis 1 cm = 2 %

Step 3: Draw rectangles with class intervals as bases and the corresponding frequencies as the heights.



7.5.1.4 Histogram for grouped frequency distribution with class marks

Example 15: Construct a histogram from the following distribution of total marks obtained by 65 students of class VIII.

Marks (Mid points)	150	160	170	180	190	200
No of students	8	10	25	12	7	3

Solution: As class marks (mid points) are given, class intervals are to be calculated from the class marks.

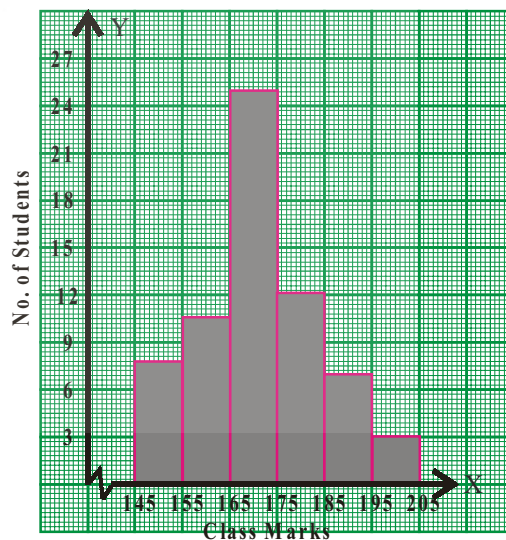
Step 1: Find the difference between two successive classes. $h = 160 - 150 = 10$.
(Find whether difference between every two successive classes is same)

Step 2: Calculate lower and upper boundaries of every class with class mark ' x ', as $x - \frac{h}{2}$ and $x + \frac{h}{2}$.

Step 3: Choose a suitable scale. X-axis 1 cm = one class interval
Y-axis 1 cm = 4 students

Step 4: Draw rectangles with class intervals as bases and the corresponding frequencies as the heights.

Class Marks (x)	Class Intervals	Frequency (No of students)
150	145 – 155	8
160	155 – 165	10
170	165 – 175	25
180	175 – 185	12
190	185 – 195	7
200	195 – 205	3



Think, Discuss and Write



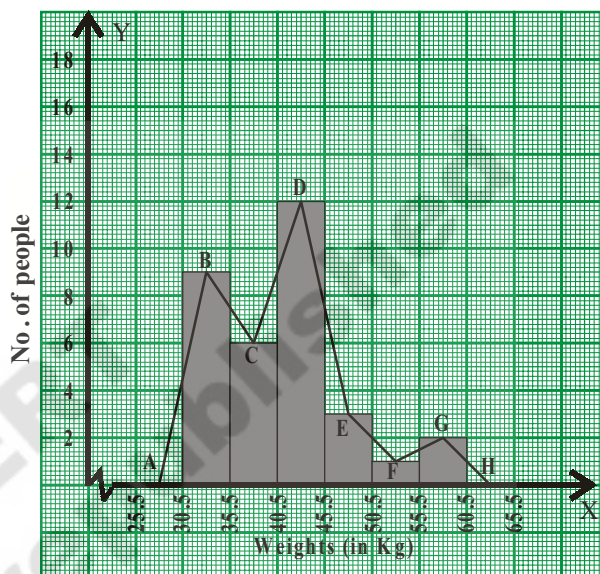
1. Class boundaries are taken on the ' X '-axis. Why not class limits?
2. Which value decides the width of each rectangle in the histogram?
3. What does the sum of heights of all rectangles represent?

7.5.2 Frequency Polygon

7.5.2.1 Interpretation of Frequency Polygon

Frequency polygon is another way of representing a quantitative data and its frequencies. Let us see the advantages of this graph.

Consider the adjacent histogram representing weights of 33 people in a company. Let us join the mid-points of the upper sides of the adjacent rectangles of this histogram by means of line segments. Let us call these mid-points B, C, D, E, F and G. When joined by line segments, we obtain the figure BCDEFG. To complete the polygon, we assume that there is a class interval with frequency zero before 30.5-35.5 and one after 55.5 - 60.5, and their mid-points are A and H, respectively. ABCDEFGH is the frequency polygon.



Although, there exists no class preceding the lowest class and no class succeeding the highest class, addition of the two class intervals with zero frequency enables us to make the area of the frequency polygon the same as the area of the histogram. Why is this so?

Think, Discuss and Write

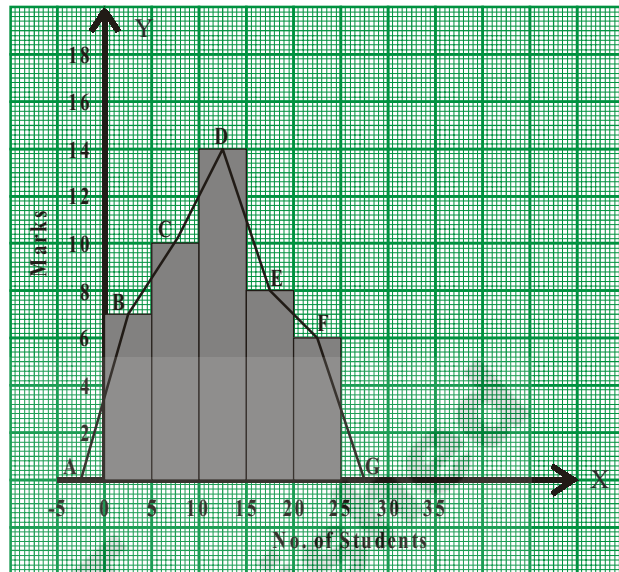


1. How do we complete the polygon when there is no class preceding the first class?
2. The area of histogram of a data and its frequency polygon are same. Reason how.
3. Is it necessary to draw histogram for drawing a frequency polygon?
4. Shall we draw a frequency polygon for frequency distribution of discrete series?

7.5.2.2 Construction of a Frequency Polygon

Consider the marks, (out of 25), obtained by 45 students of a class in a test. Draw a frequency polygon corresponding to this frequency distribution table.

Class Interval (Marks)	Frequency (No. of students)	Mid Values
0-5	7	2.5
5-10	10	7.5
10-15	14	12.5
15-20	8	17.5
20-25	6	22.5
Total	45	



Steps of construction

- Step 1: Calculate the mid points of every class interval given in the data.
- Step 2: Draw a histogram for this data and mark the mid-points of the tops of the rectangles (here in this example B, C, D, E, F respectively).
- Step 3: Join the mid points successively.
- Step 4: Assume a class interval before the first class and another after the last class. Also calculate their mid values (A and H) and mark on the axis. (Here, the first class is 0 – 5. So, to find the class preceding 0 - 5, we extend the horizontal axis in the negative direction and find the mid-point of the imaginary class-interval $-5 - 0$)
- Step 5: Join the first end point B to A and last end point F to G which completes the frequency polygon.

Frequency polygon can also be drawn independently without drawing histogram. For this, we require the midpoints of the class interval of the data.



Do These

1. Construct the frequency polygons of the following frequency distributions.
- (i) Runs scored by students of a class in a cricket friendly match.

Runs scored	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No of students	3	5	8	4	2

- (ii) Sale of tickets for drama in an auditorium.

Rate of ticket	10	15	20	25	30
No of tickets sold	50	30	60	30	20

7.5.2.3 Characteristics of a Frequency Polygon:

1. Frequency polygon is a graphical representation of a frequency distribution (discrete / continuous)
2. Class marks or Mid values of the successive classes are taken on X-axis and the corresponding frequencies on the Y-axis.
3. Area of frequency polygon and histogram drawn for the same data are equal.

Think, Discuss and Write



1. Histogram represents frequency over a class interval. Can it represent the frequency at a particular point value?
2. Can a frequency polygon give an idea of frequency of observations at a particular point?

7.5.2.4 Construction of a Frequency Polygon for a grouped frequency distribution without using histogram:

In a study of diabetic patients, the following data were obtained.

Ages	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No of patients	5	9	16	11	3

Let us construct frequency polygon for it without using the histogram.

Step 1: Find the class marks of different classes.

Step 2: Select the scale :

X-axis 1 cm = 1 class interval

Y-axis 1 cm = 2 marks

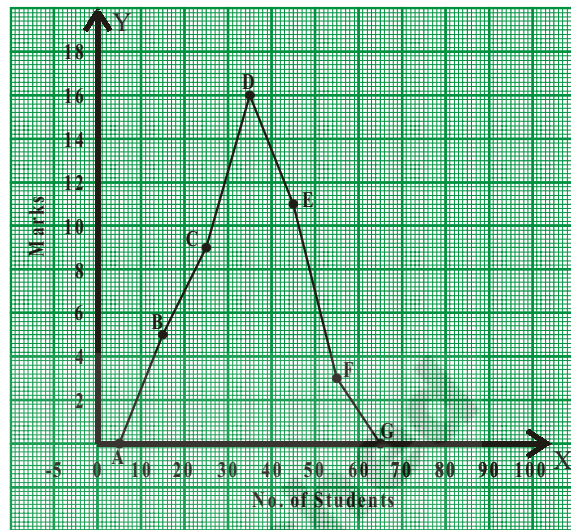
Step 3: If 'x' denotes the class mark and f denotes the corresponding frequency of a particular class, then plot ('x', f) on the graph.

Step 4: Join the consecutive points in order by line segments.

Step 5: Imagine two more classes, one before the first class and the other after the last class each having zero frequency. Mark their mid values on the graph.

Step 6: Complete the polygon.

Class Interval (Ages)	No of Patients	Class Mark	Points
0 – 10	0	5	(5, 0)
10 – 20	5	15	(15, 5)
20 – 30	9	25	(25, 9)
30 – 40	16	35	(35, 16)
40 – 50	11	45	(45, 11)
50 – 60	3	55	(55, 3)
60 – 70	0	65	(65, 0)



7.5.3 Frequency Curve for a grouped frequency distribution

It is another way of representation of the data by a free hand curve.

Let us construct frequency curve for the above data without using the histogram.

Step 1: Find the class marks of different classes.

Step 2: Select the scale :

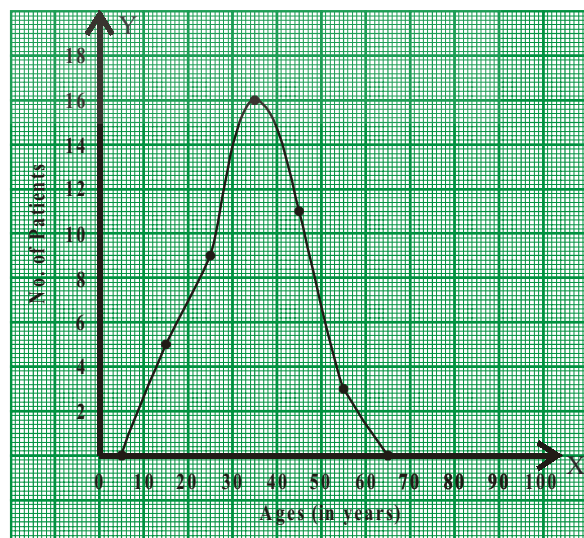
X-axis 1 cm = 1 class interval

Y-axis 1 cm = 2 marks

Step 3: If 'x' denotes the class mark and f denotes the corresponding frequency of a particular class, then plot (x, f) on the graph.

Step 4: Join the consecutive points successively by a free hand curve.

Class Interval (Ages)	No of Patients	Class Mark	Points
0 – 10	0	5	(5, 0)
10 – 20	5	15	(15, 5)
20 – 30	9	25	(25, 9)
30 – 40	16	35	(35, 16)
40 – 50	11	45	(45, 11)
50 – 60	3	55	(55, 3)
60 – 70	0	65	(65, 0)



7.5.4 Graph of a Cumulative Frequency Distribution

A graph representing the cumulative frequencies of a grouped frequency distribution against the corresponding lower / upper boundaries of respective class intervals is called Cumulative Frequency Curve or Ogive Curve.

These curves are useful in understanding the accumulation or outstanding number of observations at every particular level of continuous series.

7.5.4.1 Less than Cumulative frequency curve

Consider the grouped frequency distribution of number of tenders received by a department from the contractors for a civil work in a course of time.

CI (days)	0 – 4	4 - 8	8 - 12	12 – 16	16 – 20
No of tenders	2	5	12	10	3

Step 1: If the given frequency distribution is in inclusive form, then convert it into an exclusive form.

Step 2: Construct the less than cumulative frequency table.

Step 3: Mark the upper boundaries of the class intervals along X-axis and their corresponding cumulative frequencies along Y-axis

Select the scale :

X-axis 1 cm = 1 class interval

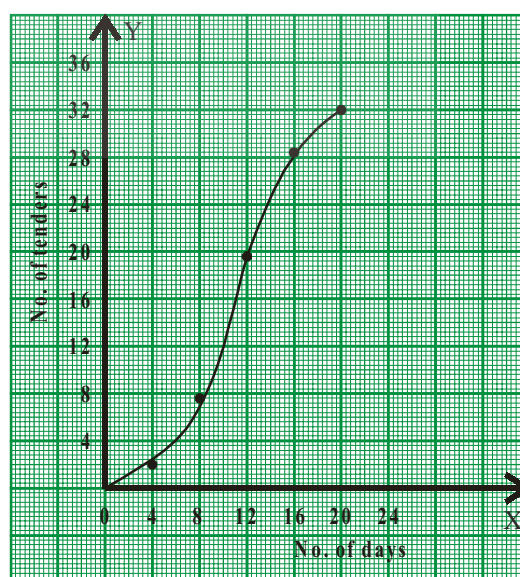
Y-axis 1 cm = 4 tenders

Step 4: Also, plot the lower boundary of the first class (upper boundary of the class previous to first class) interval with cumulative frequency 0.

Step 5: Join these points by a free hand curve to obtain the required ogive.

Similarly we can construct ‘Greater than cumulative frequency curve’ by taking greater than cumulative on Y-axis and corresponding ‘Lower Boundaries’ on the X-axis.

Class Interval (Days)	No of Tenders	UB	L.Cu.Fr
0 – 4	2	4	2
4 – 8	5	8	7
8 – 12	12	12	19
12 – 16	10	16	29
16 – 20	3	20	32





Exercise - 7.3

1. The following table gives the distribution of 45 students across the different levels of Intelligent Quotient. Draw the histogram for the data.

IQ	60-70	70-80	80-90	90-100	100-110	110-120	120-130
No of students	2	5	6	10	9	8	5

2. Construct a histogram for the marks obtained by 600 students in the VII class annual examinations.

Marks	360	400	440	480	520	560
No of students	100	125	140	95	80	60

3. Weekly wages of 250 workers in a factory are given in the following table. Construct the histogram and frequency polygon on the same graph for the data given.

Weekly wage	500-550	550-600	600-650	650-700	700-750	750-800
No of workers	30	42	50	55	45	28

4. Ages of 60 teachers in primary schools of a Mandal are given in the following frequency distribution table. Construct the Frequency polygon and frequency curve for the data without using the histogram. (Use separate graph sheets)

Ages	24 – 28	28 – 32	32 – 36	36 – 40	40 – 44	44 – 48
No of teachers	12	10	15	9	8	6

5. Construct class intervals and frequencies for the following distribution table. Also draw the ogive curves for the same.

Marks obtained	Less than 5	Less than 10	Less than 15	Less than 20	Less than 25
No of students	2	8	18	27	35



What we have discussed

- Arithmetic mean of the ungrouped data = $\frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$ or $\bar{x} = \frac{\sum x_i}{N}$ (short representation) where $\sum x_i$ represents the sum of all x_i s where 'i' takes the values from 1 to n
- Arithmetic mean = Estimated mean + Average of deviations
Or $\bar{x} = A + \frac{\sum (x_i - A)}{N}$
- Mean is used in the analysis of numerical data represented by unique value.
- Median represents the middle value of the distribution arranged in order.
- The median is used to analyse the numerical data, particularly useful when there are a few observations that are unlike mean, it is not affected by extreme values.
- Mode is used to analyse both numerical and verbal data.
- Mode is the most frequent observation of the given data. There may be more than one mode for the given data.
- Representation of classified distinct observations of the data with frequencies is called 'Frequency Distribution' or 'Distribution Table'.
- Difference between upper and lower boundaries of a class is called length of the class denoted by 'C'.
- In a class the initial value and end value of each class is called the lower limit and upper limit respectively of that class.
- The average of upper limit of a class and lower limit of successive class is called upper boundary of that class.
- The average of the lower limit of a class and upper limit of preceding class is called the lower boundary of the class.
- The progressive total of frequencies from the last class of the table to the lower boundary of particular class is called Greater than Cumulative Frequency (G.C.F).

- The progressive total of frequencies from first class to the upper boundary of particular class is called Less than Cumulative Frequency (L.C.F.).
- Histogram is a graphical representation of frequency distribution of exclusive class intervals.
- When the class intervals in a grouped frequency distribution are varying we need to construct rectangles in histogram on the basis of frequency density.

Frequency density =

$$\frac{\text{Frequency of class}}{\text{Length of that class}} \times \text{Least class length in the data}$$

- Frequency polygon is a graphical representation of a frequency distribution (discrete / continuous)
- In frequency polygon or frequency curve, class marks or mid values of the classes are taken on X-axis and the corresponding frequencies on the Y-axis.
- Area of frequency polygon and histogram drawn for the same data are equal.
- A graph representing the cumulative frequencies of a grouped frequency distribution against the corresponding lower / upper boundaries of respective class intervals is called Cumulative Frequency Curve or “Ogive Curve”.

Thinking Critically

The ability of some graphs and charts to distort data depends on perception of individuals to figures. Consider these diagrams and answer each question both before and after checking.

- Which is longer, the vertical or horizontal line?
- Are lines l and m straight and parallel?
- Which line segment is longer: \overline{AB} or \overline{BC}
- How many sides does the polygon have? Is it a square?
- Stare at the diagram below. Can you see four large posts rising up out of the paper? State some and see four small posts.

