Statistics

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	𝍸	5
154	||	2
155	||	2
158	𝍸 ||	7
160	𝍸 |	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 𝍸 (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.

To understand the concept with the help of an example, look at the following video.

The table obtained in the video can be represented using tally marks as follows:

Group	Tally Mark	Frequency
0 – 10	$\|\|$	2
10 – 20	$\cancel{\|\|\|\|}\ \cancel{\|\|\|\|}\ \|\|\|\|$	14
20 – 30	$\cancel{\|\|\|\|}\ \cancel{\|\|\|\|}\ \|\|\|\|$	14
30 – 40	$\cancel{\|\|\|\|}\ \cancel{\|\|\|\|}$	10
40 – 50	$\cancel{\|\|\|\|}\ \|\|\|$	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:

Class interval (Runs scored)	**Tally marks**	**Frequency**
0 – 20	卌	5
20 – 40	卌 |	6
40 – 60	||||	4
60 – 80	||||	4
80 – 100	|||	3
100 – 120	||	2
120 – 140	|	1

(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. $\text{Class mark} = \dfrac{\text{Lower class limit} + \text{Upper class limit}}{2}$

$$\Rightarrow \text{Class mark} = \frac{48 + 60}{2}$$
$$\Rightarrow \text{Class mark} = \frac{108}{2}$$
$$\Rightarrow \text{Class mark} = 54$$

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	卌	5
10 – 20	卌 |	6
20 – 30	卌 |||	8
30 – 40	卌 |	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		7

525 – 550	𝍸 |||	8
550 – 575	𝍸		6
575 – 600	𝍸	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

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

Calculating Cumulative Frequency from Grouped Frequency Distribution Table

When the data is very large, we take the help of the frequency distribution table to represent the same. Frequencies of different classes are given in this table, which helps us to extract the desired information.

Apart from this, there is more to learn about frequency. For the same, let us consider the data collected from a random survey conducted about the number of persons of different ages in a locality. The data is represented in the following table.

Age (in years)	5 − 15	15 − 25	25 − 35	35 − 45	45 − 55	55 − 65
Number of persons	7	19	17	12	8	2

(Table 1)

Now, if we want to find out the number of people of age more than 35 or that of less than 65 then how would we proceed?

This can be done by simply adding and subtracting the frequencies of the required classes. We have a method of calculating the same by finding the cumulative frequency.

Cumulative frequency for a class is the total of frequencies above or below it. Cumulative frequency (*c.f.*) can be of the following two types:

(1) Less than type

(2) More than type

Let us understand them one by one.

•**Less than type cumulative frequency**

Less than type cumulative frequency is calculated to collect information about the frequency less than the upper limit of each class.

Let us consider the Table 1, less than type cumulative frequency for this can be calculated in the following manner:

Age(in years) (Class)	Number of people (f)	Cumulative frequency ($c.f.$) (less than type)
5 − 15	7	7
15 − 25	19	7 + 19 = 26
25 − 35	17	26 + 17 = 43
35 − 45	12	43 + 12 = 55
45 − 55	8	55 + 8 = 63
55 − 65	2	63 + 2 = 65
Total	**65**	

This data can be represented in a modified table in the following manner:

Age (in years) (Class)	$c.f.$	Cumulative frequency less than the upper limit of class
5 − 15	7	7 people are of age less than 15

5 − 25	26	26 people are of age less than 25
5 − 35	43	43 people are of age less than 35
5 − 45	55	55 people are of age less than 45
5 − 55	63	63 people are of age less than 55
5 − 65	65	65 people are of age less than 65

From this table, we can easily find the number of people of age less than the upper limit of each class.

• **More than type cumulative frequency**

More than type cumulative frequency is calculated to collect information about the frequency more than the lower limit of each class.

To calculate more than type cumulative frequency, we start from last class to first class. The table given below shows the more than type cumulative frequency for Table 1.

Age (in years) (Class)	**Number of people (*f*)**	**Cumulative frequency (*c.f.*) (more than type)**
5 − 15	7	58 + 7 = 65
15 − 25	19	39 + 19 = 58

25 − 35	17	22 + 17 = 39
35 − 45	12	10 + 12 = 22
45 − 55	8	2 + 8 = 10
55 − 65	2	2
Total	**65**	

This data can be represented in a modified table in the following manner:

Age (in years) (Class)	***c.f.***	**Cumulative frequency more than the lower limit of class**
5 − 65	65	65 persons are of age more than 5
15 − 65	58	58 persons are of age more than 15
25 − 65	39	39 persons are of age more than 25
35 − 65	22	22 persons are of age more than 35
45 − 65	10	10 persons are of age more than 45

55 − 65	2	2 people are of age more than 55

From this table, we can easily find the number of people of age more than the lower limit of each class.

Let us now look at an example to understand this concept better.

Example:

For the given distribution, find the cumulative frequency of less than and more than type.

Class	20 − 30	30 − 40	40 − 50	50 − 60	60 − 70	70 − 80
Frequency	21	14	18	12	9	16

Solution:

Class	**Frequency**	***c.f.*(Less than type)**	***c.f.*(More than type)**
20 − 30	21	21	69 + 21 = 90
30 − 40	14	21 + 14 = 35	55 + 14 = 69
40 − 50	18	35 + 18 = 53	37 + 18 = 55
50 − 60	12	53 + 12 = 65	25 + 12 = 37

60 − 70	9	65 + 9 = 74	16 + 9 = 25
70 − 80	16	74 + 16 = 90	16
Total	**90**		

Construction of Histograms when Class Size is Same

The frequency distribution table of the marks of 26 students in a particular subject is as follows.

Class interval (marks of students)	Frequency (number of students)
0 – 10	4
10 – 20	2
20 – 30	10
30 – 40	8
40 – 50	2

Can we represent this data graphically?

This data can be represented in the form of a histogram.

To understand the concept of histograms and the method used to draw a histogram, let's look at this video.

Let us now look at an example to understand this concept better.

Example 1:

The given tally table represents the total runs scored by 39 batsmen in 10 different test matches.

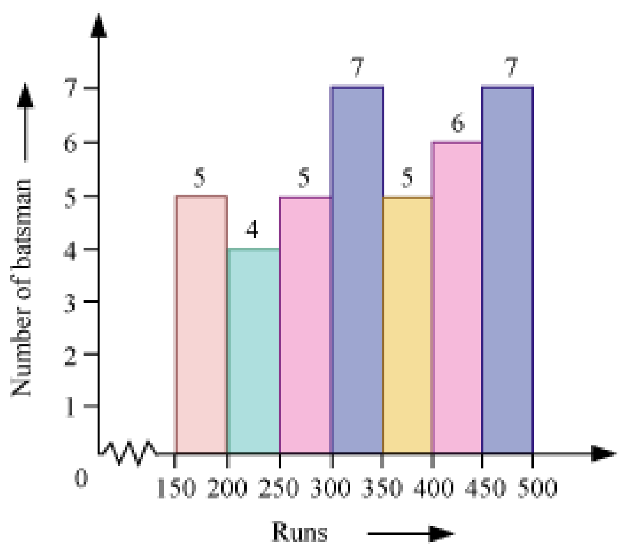
Runs	Tally marks	Frequency (Number of batsmen)
150 – 200	$\cancel{\vert\vert\vert\vert}$	5
200 – 250	$\vert\vert\vert\vert$	4

250 – 300	~~||||~~	5
300 – 350	~~||||~~ ||	7
350 – 400	~~||||~~	5
400 – 450	~~||||~~ |	6
450 – 500	~~||||~~ ||	7

Draw a histogram for the above given distribution table.

Solution:

In order to draw the histogram of the given frequency distribution table, we represent the runs on the horizontal axis and the number of batsmen on vertical axis. The height of each bar represents the frequency. The width of all the bars is same.



Here, we will use a broken line ( ) to indicate that the values between 0 – 150 are not represented.

Example 2:

The given table represents the data related to intelligent quotient (IQ) of the students of a class.

IQ	Number of students

61 – 70	4
71 – 80	3
81 – 90	5
91 – 100	8
101 – 110	15
111 – 120	12
121 – 130	13
Total	60

Draw a histogram for the above given distribution table.

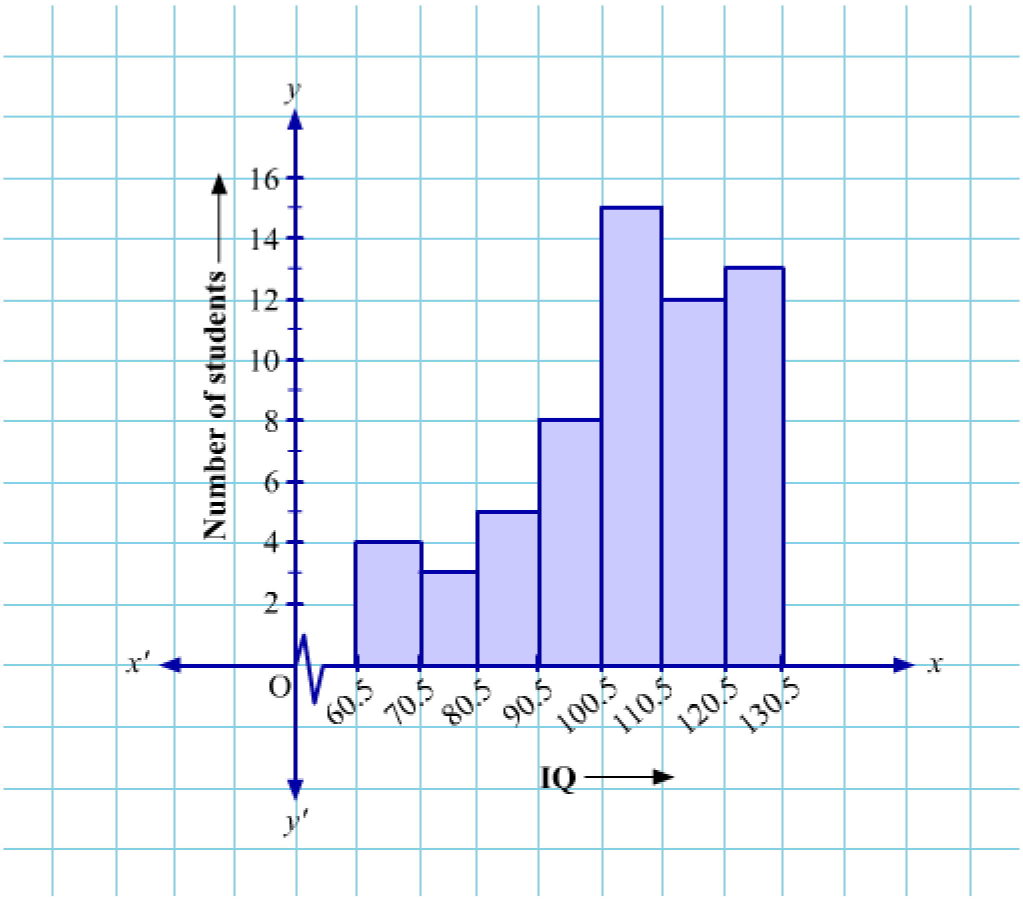
Solution:

In the given table the class intervals are of inclusive type, so we need to make them of exclusive type. Here, the difference between the upper limit of a class and lower limit of next class is 1. So, we need to subtract half of this i.e., 0.5 from lower limit and add 0.5 to upper limit of each class. Thus, we will get extended classes according to which, we can draw the histogram.

The modified table consisting extended classes is as follows:

Original Class	**Extended Class**	**Number of students (Frequency)**
61 – 70	60.5 – 70.5	4
71 – 80	70.5 – 80.5	3
81 – 90	80.5 – 90.5	5
91 – 100	90.5 – 100.5	8
101 – 110	100.5 – 110.5	15
111 – 120	110.5 – 120.5	12
121 – 130	120.5 – 130.5	13
Total		60

In order to draw the histogram of this frequency distribution table, we represent the IQ on the horizontal axis and the number of students on vertical axis. The height of each bar represents the frequency. The width of all the bars is same.

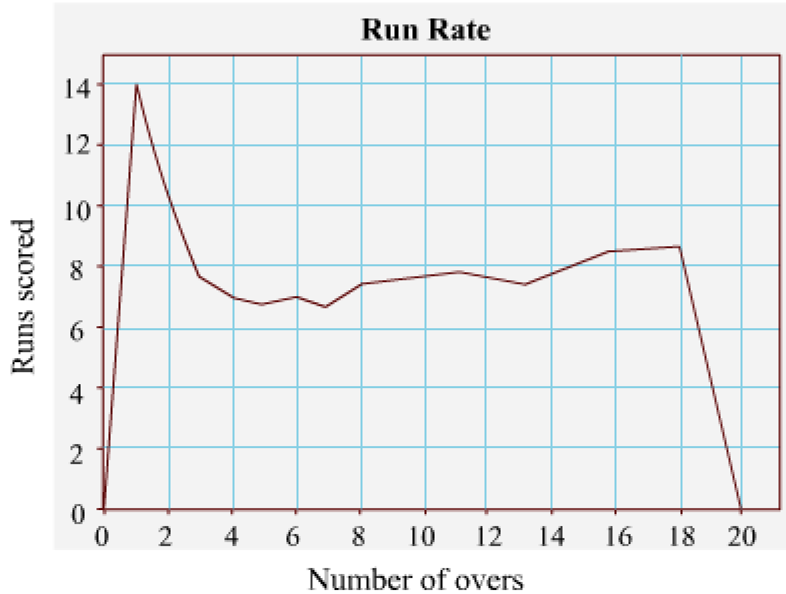


Here, we will use a broken line ( ) to indicate that the values between 0 – 60.5 are not represented.

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.

Drawing a Frequency Polygon Using a Histogram

Solved Examples

Medium

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

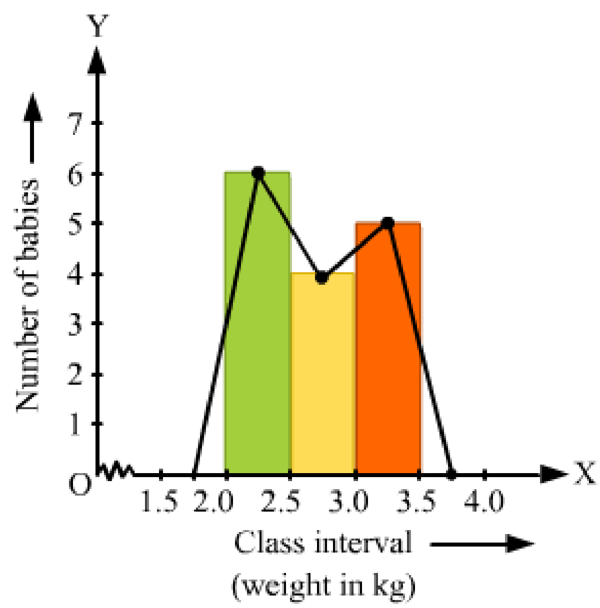
Draw a histogram for the data and then draw a frequency polygon using it.

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

Medium

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:

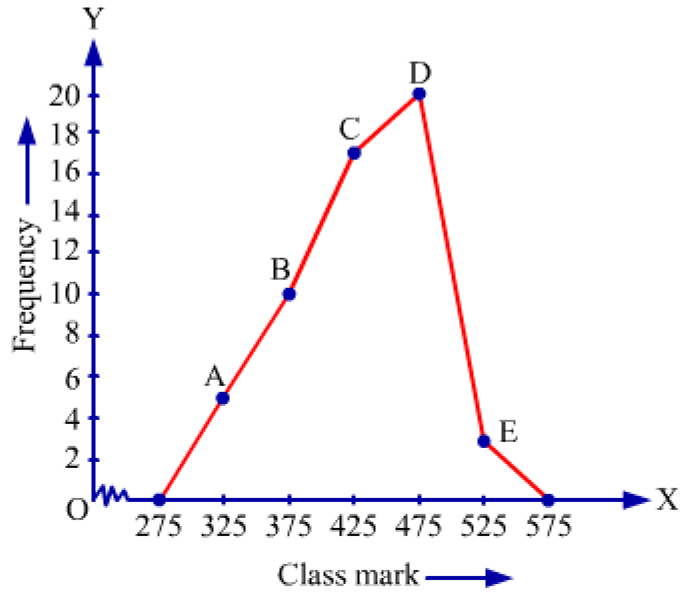
We will first calculate the class marks and write the data as follows:

Class interval	**Class mark**	**Frequency**	**Points**
300−350	325	5	A (325, 5)
350−400	375	10	B (375, 10)
400−450	425	17	C (425, 17)
450−500	475	20	D (475, 20)
500−550	525	3	E (525, 3)

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≠0p≠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≠0p≠0) then the mean of the new observations is also divided by *p*.

If the mean of *n* observations $x_1, x_2, x_3 \dots . x_n$ is $\bar{x}$
then the mean of $\frac{x_1}{p},\ \frac{x_2}{p},\ \frac{x_3}{p},\ \dots,\ \frac{x_n}{p}$ is $\frac{\bar{x}}{p}$.

Solved Examples

Easy

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¯ 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:

$$\text{Mean salary of the five workers} = \frac{\text{Sum of the salaries of the five workers}}{5}$$
$$\Rightarrow 2500 = \frac{\text{Sum of the salaries of the five workers}}{5}$$
$$\Rightarrow \therefore \text{Sum of the salaries of the five workers} = 2500 \times 5 = 12500 \qquad ...(1)$$

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 = $\dfrac{\frac{2}{5} + \frac{1}{2}}{2} = \dfrac{\frac{4+5}{10}}{2} = \dfrac{9}{10 \times 2} = \dfrac{9}{20}$

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,

$$
\begin{aligned}
\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)}{n} &= \frac{\left(x_1-\overline{x}\right)+\left(x_2-\overline{x}\right)+\left(x_3-\overline{x}\right)+\ldots+\left(x_n-\overline{x}\right)}{n} \\
&= \frac{\left(x_1+x_2+x_3+\ldots+x_n\right)-\left(\overline{x}+\overline{x}+\overline{x}+\ldots+\overline{x}\right)}{n} \\
&= \frac{n\overline{x}-n\overline{x}}{n} \qquad \text{(By equation 1)} \\
&= \frac{0}{n} \\
&= 0
\end{aligned}
$$

Thus, the given result is proved.

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, 2*y*, 3*y* + 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, 2*y*, 3*y* + 1, 69, 95, 112, 135, 1204

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

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

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

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

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

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

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

So,

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