# STATISTICS

## Key Points

In this chapter, you have studied the following points:

1.  The mean for grouped data is calculated by:

    (i)     The direct method: $\bar{x} = \dfrac{\sum f_i x_i}{\sum f_i}$

    (ii)    The assumed mean method: $\bar{x} = a + \dfrac{\sum f_i d_i}{\sum f_i}$ $or$ $\bar{x} = a + \bar{d}$

    (iii)   The step deviation method: $\bar{x} = a + \left( \dfrac{\sum f_i u_i}{\sum f_i} \right) \times h$ $or$ $\bar{x} = a + h\,\bar{x}$

2.  The mode for grouped data can be found by using the formula:

    $$\text{Mode} = l + \left( \frac{f_1 - f_o}{2f_1 - f_o - f_2} \right) \times h$$

    Where, symbols have their usual meaning.

3.  The median for grouped data can be found formed by using the formula:

    $\text{Median} = l + \left( \dfrac{\dfrac{n}{2} - cf}{f} \right) \times h$ Where symbols have their usual meanings.

4.  In order to find median, class intervals should be continuous.

5.  Representing a cumulative frequency distribution graphically as a cumulative frequency curve, or an ogive of the less than type and of the more than type.

6.  While drawing ogives, class boundaries are taken on X- axis and corresponding cumulative frequencies are taken on Y- axis.

7.   Scale on both the axes may not be equal.

8.  The median of grouped data can be obtained graphically as the $x-$coordinate of the point of intersection of the two ogives of the data.

## Mean of Grouped Data

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

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} \ or, \ \overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

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

In other words, we can say that the arithmetic mean of a set of observations is equal to their sum divided by the total number of observations.

2. In a discrete frequency distribution the arithmetic mean may be computed by any one of the following methods:

(i) Direct method,

(ii) Short – cut method,

(iii) Step – Deviation method.

Let us now learn about these methods one by one.

## Direct Method

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

$$\overline{X} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_n x_n}{f_1 + f_2 + \ldots + f_n}$$

Or $$\overline{X} = \frac{\sum_{i=1}^{n} f_i x_i}{N}, \ where \ N = \sum_{i=1}^{n} f_1 = f_1 + f_2 \ldots + f_n$$

The following algorithm may be used to compute arithmetic mean by direct method.

## Algorithm

| | |
|---|---|
| <u>Step I</u> | Prepare the frequency table in such a way that its first column consists of the values of the variable and the second column the corresponding frequencies. |
| <u>Step II</u> | Multiply the frequency of each row with the corresponding values of variable to obtain third column containing $f_i x_i$ |
| <u>Step III</u> | Find the sum of all entries in column III to obtain $\sum f_i x_i$ |
| <u>Step IV</u> | Find the sum of all the frequencies in column II to obtain $\sum f_i = N$ |
| <u>Step V</u> | Use the formula: $\overline{X} = \dfrac{\sum f_i x_i}{N}$ |

**Example 1**   **Find the mean of the following distribution**:

$x$ :   10    30    50    70    89

$f$ :   7    8    10    15    10

**Solution**                         **Calculation of Mean**

| $x_i$ | $f_i$ | $f_i x_i$ |
|:---:|:---:|:---:|
| 10 | 7 | 70 |
| 30 | 8 | 240 |
| 50 | 10 | 500 |
| 70 | 15 | 1050 |
| 89 | 10 | 890 |
| | $N = \sum f_i = 50$ | $\sum f_i x_i = 2750$ |

$$\therefore \quad Mean = \frac{\sum f_i x_i}{N} = \frac{2750}{50} = 55$$

**Example : 2** **Find the value of p, if the mean of the following distribution is 7.5.**

$x$ :   3     5     7     9     11     13

$f$ :   6     8     15     p     8     4

**Solution**                                         **Calculation of Mean**

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

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

$\therefore \; Mean = \dfrac{\sum f_i x_i}{\sum f_i}$

$\Rightarrow \quad 7.5 = \dfrac{303 + 9p}{41 + p}$

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

$\Rightarrow \quad 307.5 + 7.5p = 303 + 9p \quad \Rightarrow 9p - 7.5p = 307.5 - 303 \Rightarrow 1.5p = 4.5 \Rightarrow p = 3$

**Short –Cut Method**

Let  $x_1, x_2 ..., x_n$  be values of a variable X with corresponding frequencies  $f_1, f_2, f_3, ...., f_n$  respectively. Taking deviations about an arbitrary point 'A', we have

$d_1 = x_i - A, i = 1, 2, 3, ..., n$

$\Rightarrow \quad \overline{X} = A + \dfrac{1}{N} \sum_{i=1}^{n} f_i d_i$

Finding AM by using the above formula is known as the short – cut method.

**Note :** The number 'A' is generally known as the assumed mean and is generally chosen in such a way that the deviations are small.

**Algorithm**

Step I          Prepare the frequency table in such a way that its first column consists of the values of the variable and the second column consists of the corresponding frequencies.

Step II         Choose a number 'A' (preferable among the values in first column) and take deviations $d_i = x_i - A$ of the value $x_i$ of variable X about A. Write these deviations against the corresponding frequencies in the third column.

Step III        Multiply the frequencies in column II with the corresponding deviations $d_i$ in column III to prepare column IV consisting of $f_i d_i$.

Step IV         Find the sum of all entries in column III to obtain $\sum\limits_{i=1}^{n} f_i d_i$ and the sum of all frequencies in column Ii to obtain $\sum\limits_{i=1}^{n} f_i = N$.

Step V          Use the formula: $\overline{X} = A + \dfrac{1}{N}\left(\sum\limits_{i=1}^{n} f_i d_i\right)$.

**Example: 3**   **Find the mean wages from the data given below:**

Wage (in Rs.):      800      820      860      900      920      980      1000

No. of workers:     7        14       19       25       20       10       5

**Solution**      Let the assumed mean be A =900.

<div align="center">

**Calculation of Mean**

</div>

| Wage (in Rs. ) $x_i$ | No. of workers $f_i$ | $d_i = x_i - A = x_i - 900$ | $f_i d_i$ |
|---|---|---|---|
| 800 | 7 | -100 | -700 |
| 820 | 14 | -80 | -1120 |
| 860 | 19 | -40 | -760 |
| 900 | 25 | 0 | 0 |

| 920 | 20 | 20 | 400 |
| 980 | 10 | 80 | 800 |
| 1000 | 5 | 100 | 500 |

$$N = \sum f_i = 100 \qquad\qquad \sum f_i d_i = -880$$

We have,

$$N = 100, \sum f_i d_i = -880 \ and \ A = 900$$

$\therefore \quad$ Mean $\overline{X} = A + \dfrac{\sum f_i d_i}{N}$

$\Rightarrow \quad$ Mean $^{-} = 900 + \dfrac{-880}{100} = 900 - 8.8 = 891.2$

Hence, mean wage = Rs. 891.2.

## Step – Deviation Method

Sometimes, during the application of the short – cut method for finding AM, the deviations $d_i$ are divisible by a common number h (say). In such a case the arithmetic is reduced to a great extent by taking

$$u_i = \frac{x_i - A}{h} ; i = 1,2,3,....,n$$

$\Rightarrow \quad x_i = A + h u_i, i = 1,2,3,....n$

$\Rightarrow \quad \overline{x} = A + h\left\{ \dfrac{1}{N} \sum_{i=1}^{n} f_i u_i \right\}$

Finding AM by using this formula is known as the step – deviation method.

Following algorithm may be used to find the arithmetic mean by step deviation method :

## Algorithm

Step I        Obtain the frequency distribution and prepare the frequency table in such a way that its first column consists of the values of the variable and the second column corresponding frequencies.

| Step II | Chose a number 'A' (generally known as the assumed mean) and take deviations $d_i = x_i - A$ about A. write these deviations against the corresponding frequencies in the third column. |
|---|---|
| Step III | Choose a number h, generally common factor of all $d_i's$ in III column, divide deviations $d_i$ by h to get $u_i$. Write these $u_i's$ against the corresponding $d_i's$ in the IV column. |
| Step IV | Multiply the frequencies in II column with the corresponding $u_i's$ in IV column to prepare V column of $f_i u_i$. |
| Step V | Find the sum of all entries in V column to obtain $\sum_{i=1}^{n} f_i u_i$ and the sum of all frequencies in II column to obtain $N = \left( \sum_{i=1}^{n} f_i \right)$ |
| Step VI | Use the formula: $\overline{X} = A + h \left\{ \dfrac{1}{N} \sum_{i=1}^{n} f_i u_i \right\}$ |

**Example 4:** **The weights in kilograms of 60 workers in a factory are given in the following frequency table. Find the mean weight of a worker.**

Weight (in kg) $x$ :  60    61    62    63    64    65

No. of workers $f$ :  5    8    14    16    10    7

**Solution:**    Let the assumed mean be A=63

### Calculation of Mean

| Weight (in kg ) $x$ | No. of workers $f$ | $d_i = x_i - 63$ | $f_i d_i$ |
|---|---|---|---|
| 60 | 5 | -3 | -15 |
| 61 | 8 | -2 | -16 |
| 62 | 14 | -1 | -14 |
| 63 | 16 | 0 | 0 |
| 64 | 10 | 1 | 10 |

| 65 | 7 | 2 | 14 |
|---|---|---|---|
| | $N = \sum f_i = 60$ | | $\sum f_i d_i = -21$ |

We have, $N = 60, A = 63$ and $\sum f_i d_i = -21$

$$\therefore \quad Mean = A + \left\{ \frac{1}{N} \sum f_i d_i \right\}$$

$$\Rightarrow \quad Mean = 63 + \left( \frac{-21}{60} \right) = 63 - \frac{7}{20} = 63 - 0.35 = 62.65$$

Hence, mean weight of a worker = 62.65 kg.

## Arithmetic Mean of a Continuous Frequency Distribution

**Example 5**    **Find the mean of the following frequency distribution**:

Classes:     0-20    20-40   40-60   60-80   80-100

Frequency:   15       18       21       29       17

**Solution**                            **Calculation of Mean**

| Classes | Mid- values $(x_i)$ | Frequency $(f_i)$ | $d_i = x_i - 50$ | $u_i = \dfrac{x_i - 50}{20}$ | $f_i u_i$ |
|---|---|---|---|---|---|
| 0-20 | 10 | 15 | -40 | -2 | -30 |
| 20-40 | 30 | 18 | -20 | -1 | -18 |
| 40-60 | 50 | 21 | 0 | 0 | 0 |
| 60-80 | 70 | 29 | 20 | 1 | 29 |
| 80-100 | 90 | 17 | 40 | 2 | 34 |
| | | $\sum f_i = 100$ | | | $\sum f_i u_i = 15$ |

We have,

$$A = 50, h = 20, N = 100 \text{ and } \sum f_i u_i = 15$$

$$\therefore \quad Mean = A + h\left\{\frac{1}{N}\sum f_i u_i\right\}$$

$$\Rightarrow \quad Mean = 50 + 20 \times \frac{15}{100}$$

$$\Rightarrow \quad Mean = 50 + 3 = 53$$

**Example: 6** the mean of the following frequency table 50. But the frequencies $f_1$ and $f_2$ in class 20-40 and 60-80 are missing. Find the missing frequencies.

| Class: | 0-20 | 20-40 | 40-60 | 60-80 | 80-100 | Total |
|--------|------|-------|-------|-------|--------|-------|
| Frequency: | 17 | $f_1$ | 32 | $f_2$ | 19 | 120 |

**Solution:** Let the assumed mean be A =50 and h = 20.

**Calculation of Mean**

| Class $f_1$ | Frequency $x_i$ | Mid- values | $u_i = \dfrac{x_i - A}{h}$ | $f_i u_i$ |
|-------------|-----------------|-------------|----------------------------|-----------|
| 0-20 | 17 | 10 | -2 | -34 |
| 20-40 | $f_1$ | 30 | -1 | $-f_1$ |
| 40-60 | 32 | 50 | 0 | 0 |
| 60-80 | $f_2$ | 70 | 1 | $f_2$ |
| 80-100 | 19 | 90 | 2 | 38 |
| $N = \sum f_i = 68 + f_i + f_2$ | | | | $\sum f_i u_i = 4 - f_1 + f_2$ |

We have, $N = \sum f_i = 120$

$$\Rightarrow \qquad 68 \ + \ f_i \ + \ f_2 \ = \ 120$$

$$\Rightarrow \qquad f_1 \ + \ f_2 \ = \ 52$$

*Now* ,

$$Mean \ = \ 50$$

$$\Rightarrow \qquad A \ + \ h \left\{ \frac{1}{N} \sum f_i u_i \right\} \ = \ 50$$

$$\Rightarrow \quad 50 \ + \ 20 \ \times \ \left\{ \frac{4 \ - \ f_1 \ + \ f_2}{120} \right\} \ = \ 50$$

$$\Rightarrow \quad 50 \ + \ \frac{4 \ - \ f_1 \ + \ f_2}{6} \ = \ 50$$

$$4 + f_2 = f_1 \quad \frac{\wedge \ - \ f_1 \ + \ f_2}{6} \ = \ 0$$

$$\Rightarrow \quad 4 \ - \ f_1 \ = \ f_2 \ = \ 0$$

$$\Rightarrow \quad f_1 \ - \ f_2 \ = \ 4$$

Solving equations (i) and (ii), we get $f_1 = 28$ *and* $f_2 = 24$.

## Median of Grouped Data

Median is a measure of central tendency which gives the value of the middle – most observation in the data. Recall that for finding the median of ungrouped data. We first arrange the data values or the observations in ascending order.

Then, if n is odd, the median is the $\left( \frac{n+1}{2} \right)^{th}$ observation and

If n is even, then the median will be the average of the $\left( \frac{n}{2} \right)^{th}$ *and* $\left( \frac{n}{2}+1 \right)^{th}$ observations.

Suppose, we have to find the median of the following data, which is about the marks, out of 50 obtained by 100 students in attest:

| Marks obtained | 20 | 29 | 28 | 33 | 42 | 38 | 43 | 25 |
|---|---|---|---|---|---|---|---|---|
| Number of students | 6 | 28 | 24 | 15 | 2 | 4 | 1 | 20 |

First, we arrange the marks in ascending order and prepare a frequency table 14.9 as follows:

| Marks Obtained | Number of Students (Frequency ) |
|---|---|
| 20 | 6 |
| 25 | 20 |
| 28 | 24 |
| 29 | 28 |
| 33 | 15 |
| 38 | 4 |
| 42 | 2 |
| 43 | 1 |
| Total | 100 |

Here n = 100, which is even. The median will be the average of the $\left(\frac{n}{2}\right)^{th}$ and the $\left(\frac{n}{2}+1\right)^{th}$ observations, i.e., the 50th and 51st observations. To find the position of these middle values, we construct cumulative frequency.

| Marks obtained | Number of students | Cumulative frequency |
|---|---|---|
| 20 | 6 | 6 |
| Upto 25 | 6+20 =26 | 26 |
| Upto 28 | 26+24=50 | 50 |
| Upto 29 | 50+28=78 | 78 |
| Upto 33 | 78+15=93 | 93 |
| Upto 38 | 93+4=97 | 97 |
| Upto 42 | 97+2=99 | 99 |
| Upto 43 | 99+1=100 | 100 |

Now we add another column depicting this information to the frequency table above and name it as cumulative frequency column.

From the table above, we see that:

50th observation is 28 (why?)

51st observation is 29

Median $= \dfrac{28+29}{2} = 28.5$

## Median

### Illustration 1

(i) **The following are the marks of 9 students in a class. Find the median 34, 32, 48, 38, 24, 30, 27, 21, 35**

(ii)     Find the median of the daily wages of ten workers from the following data :

Rs. 20, 25, 17, 18, 8, 15, 22, 11, 9, 14.

**Solution:**   (i) Arranging the data in ascending order of magnitude, we have

21, 24, 27, 30, 32, 34, 35, 38, 48

Since there are 9 i.e., an odd number of items. Therefore, median is the value of $\left(\dfrac{9+1}{2}\right)^{th}$

observation i.e., 32

(ii) Arranging the wages in ascending order of magnitude, we have

8, 9, 11, 14, 15, 17, 18, 20, 22, 25

Since there are 10 observations therefore, median is the arithmetic mean of

$\left(\dfrac{10}{2}\right)^{th} and \left(\dfrac{10}{2}+1\right)^{th}$  observations.

Hence, Median $= \dfrac{15+17}{2} = 16$

**Example - 7** A survey regarding the heights (in cm) of 51 girls of Class X of a school was conducted and data was obtained as shown in table find their median.

| Height (in cm) | Number of girls |
|---|---|
| Less than 140 | 4 |
| Less than 145 | 11 |
| Less than 150 | 29 |
| Less than 155 | 40 |
| Less than 160 | 46 |
| Less than 165 | 51 |

Number of observation, n = 51

$\frac{n}{2} = \frac{51}{2} = 25.5^{th}$ observation, which lies in the class 145-150.

$\therefore$ $145-150$ is median class

The, $l$ 9the lower boundary) = 145.

$cf$ (the cumulative frequency of the class preceding 145-150) =11,

$f$ (the frequency of the median class 145-150) = 18,

h (the class size) = 5.

Using the formula, Median $= l + \dfrac{\left(\dfrac{n}{2} - cf\right)}{f} \times h$

$$= 145 + \frac{(25.5 - 11)}{18} \times 5$$
$$= 145 + \frac{72.5}{18} = 149.03$$

| Class intervals | Frequency | Cumulative Frequency |
|---|---|---|
| Below 140 | 4 | 4 |
| 140-145 | 7 | 11 cf |
| 145-150 | 18 f | 29 |
| 150-155 | 11 | 40 |
| 155-160 | 6 | 46 |
| 160-165 | 5 | 51 |

**Example – 8**  **The median of the following data is 525. Find the values, of $x$ and $y$, if the total frequency is 100. Here, CI stands for class interval and Fr for frequency.**

| CI | 0-100 | 100-200 | 200-300 | 300-400 | 400-500 | 500-600 | 600-700 | 700-800 | 800-900 | 900-1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| Fr | 2 | 5 | $x$ | 12 | 17 | 20 | $y$ | 9 | 7 | 4 |

**Solution:**

It is given that $n = 100$

So, $76 + x + y = 100$, i.e., $x + y = 24(1)$

The median is 525, which lies in the class 500-600

So, $l = 500, f = 20, cf = 36 + x, h = 100$

Using the formula

$$\text{Median} = l + \frac{\left(\dfrac{n}{2} - cf\right)}{f} \times h$$

$$525 = 500 + \frac{50 - 36 - x}{20} \times 100$$

i.e., $525 - 500 = (14 - x) \times 5$

i.e., $25 = 70 - 5x$

i.e., $5x = 70-25 = 45$

so, $x = 9$

Therefore, from (1), we get $9 + y = 24$

i.e., $y = 15$

| Class Intervals | Frequency | Cumulative Frequency |
|---|---|---|
| 0-100 | 2 | 2 |
| 100-200 | 5 | 7 |
| 200-300 | $x$ | $7 + x$ |
| 300-400 | 12 | $19 + x$ |
| 400-500 | 17 | $36 + x$ |
| 500-600 | 20 | $56 + x$ |
| 600-700 | $y$ | $56 + x + y$ |
| 700-800 | 9 | $65 + x + y$ |
| 800-900 | 7 | $72 + x + y$ |
| 900-1000 | 4 | $76 + x + y$ |

**Note:**

The median of grouped data with unequal class sizes can also be calculated.

**Obtaining Median from Ogive Curve:**

Is it possible to obtain the median from these two cumulative frequency curves? Let us see.

One obvious way is to locate on $\dfrac{n}{2} = \dfrac{53}{2} = 26.5$ on the $y$ - axis. From this point, draw a line parallel to the $x$ - axis cutting the curve at a point. From this point, draw a perpendicular to the $x$ -axis. Foot of this perpendicular determines the median of the data.

**Another way of obtaining the median:**

Draw both ogives (i.e., of the less than type and of the more than type) on the same axis. The two ogives will intersect each other at a point. From this point, if we draw a perpendicular on the $x$ - axis, the point at which it cuts the $x$ - axis gives us the median.
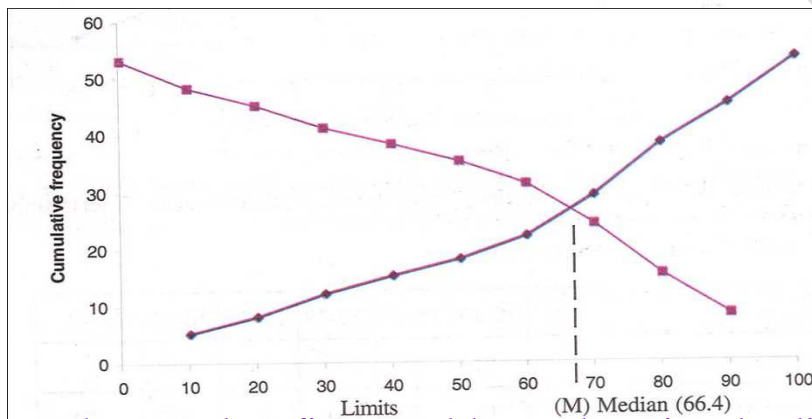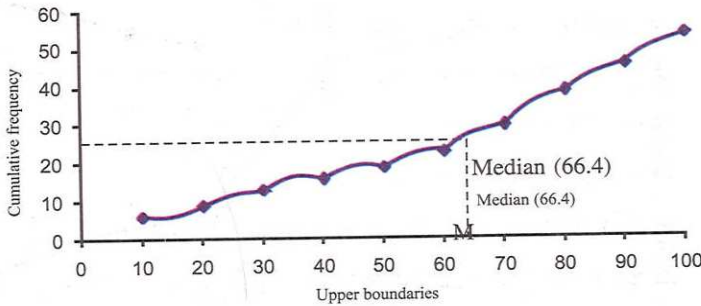




**Example – 9** **The annual profits earned by 30 shops in a locality give rise to the following distribution :**

| Profit (in lakhs) | Number of Shops (frequency) |
|---|---|
| More than or equal to 5 | 30 |
| More than or equal to 10 | 28 |
| More than or equal to 15 | 16 |
| More than or equal to 20 | 14 |
| More than or equal to 25 | 10 |
| More than or equal to 30 | 7 |
| More than or equal to 35 | 3 |

Draw both ogives for the data above. Hence obtain the median profit.

**Solution:** We first draw the coordinate axes, with lower boundaries of the profit along the horizontal axis, and the cumulative frequency along the vertical axes. Then, we plot the points (5,30), (10, 28), (15, 16), (20, 14), (25, 10), (30, 7) and (35, 3). We join these points with a smooth curve to get the more than ogive, as shown in the figure below –
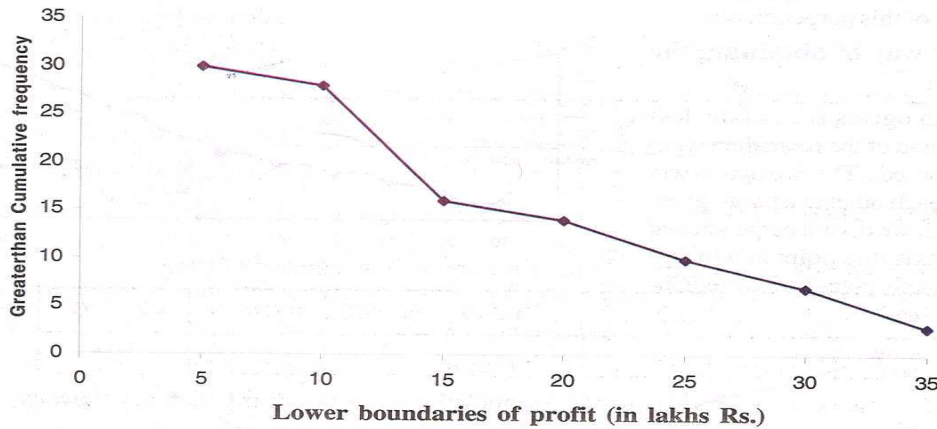


Now, let us obtain the classes, their frequencies and the cumulative frequency from the table above.

| Classes | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 |
|---|---|---|---|---|---|---|---|
| Number of Shops | 2 | 12 | 2 | 4 | 3 | 4 | 3 |
| Cumulative frequency | 2 | 14 | 16 | 20 | 23 | 27 | 30 |

Using these values, we plot the points (10,2), (15, 14), (20, 16), (25, 20), (30, 23), (35, 27), (40, 30) on the same axes as in last figure to get the less than ogive, as shown in figure below.

The abscissa of their point of intersection is nearly 17.5, which is the median. This can also be verified by using the formula. Hence, the median profit (in lakhs) is ₹17.5.

Limits
Median (17.5)
Profit (in lakhs Rs.)
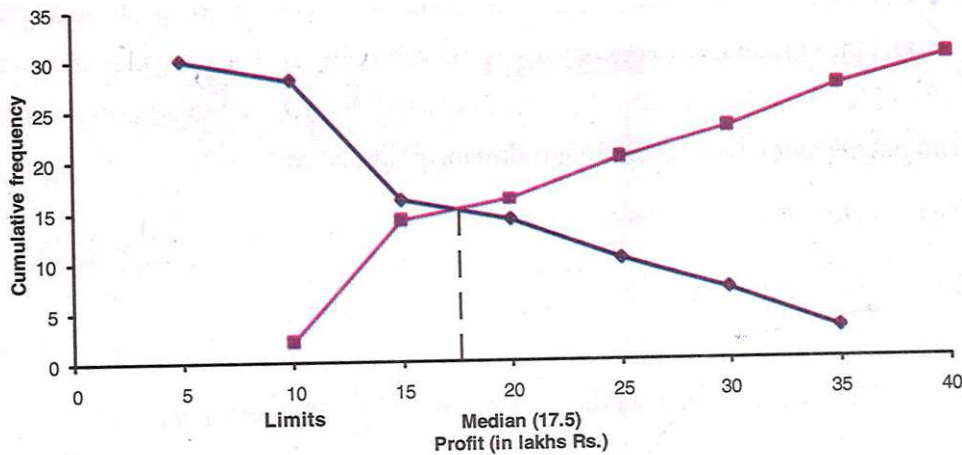
## Median of Discrete Frequency Distribution

In case of a discrete frequency distribution $x_i / f_i; i = 1, 2, \ldots, n$ we calculate the median by using the following algorithm.

## ALGORITHM

**Step I** : Find the cumulative frequencies (c.f.)

**Step II:** Find $\dfrac{N}{2}$, *where* $N = \sum\limits_{i=1}^{n} f_i$

**Step III** : See the cumulative frequency (c.f.) just greater than $\dfrac{N}{2}$ and determine the corresponding value of the variable.

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

**Illustration 2 Obtain the median for the following frequency distribution:**

$x$ :   1    2    3    4    5    6    7    8    9

$f$ :   8    10    11    16    20    25    15    9    6

**Solution:**                **Calculation of Median**

| $x$ | $f$ | $cf$ |
|---|---|---|
| 1 | 8 | 8 |
| 2 | 10 | 18 |
| 3 | 11 | 29 |

| | | |
|---|---|---|
| 4 | 16 | 45 |
| 5 | 20 | 65 |
| 6 | 25 | 90 |
| 7 | 15 | 105 |
| 8 | 9 | 114 |
| 9 | 6 | 120 |
| | $N = 120$ | |

Here, $N = 120 \Rightarrow \dfrac{N}{2} = 60$

We find that the cumulative frequency just greater than $\dfrac{N}{2}$ i.e, 60 is 65 and the value of $x$ corresponding to 65 is 5. Therefore, Median = 5.

## Median of a Grouped or Continuous Frequency Distribution

In order to calculate the median of a grouped or continuous frequency distribution, we use the following algorithm.

## Algorithm

<u>Step I</u>      Obtain the frequency distribution.

<u>Step II</u>      Prepare the cumulative frequency column and obtain $N = \sum f_i$.

<u>Step III</u>      Find N/2.

<u>Step IV</u>      See the cumulative frequency just greater than N/2 and determine the corresponding class. This class is known as the median class.

<u>Step V</u>      Use the following formula:

$$\text{Median} = l + \left\{ \dfrac{\dfrac{N}{2} - F}{f} \right\} \times h$$

Where,        $l$ = lower limit of the median class

$f$ = frequency of the median class

$h$ = with (size) of the median class

F= cumulative frequency of the class preceding the median class $N = \sum f_i$.

## Illustration 3   Calculate the median from the following distribution:

Class        :        5-10    10-15   15-20   20-25   25-30   30-35   35-40   40-45

Frequency   :        5        6        15      10       5        4        2        2

**Solution:**   First we prepare the following cumulative table to compute the median

| Class | Frequency | Cumulative Frequency |
|-------|-----------|----------------------|
| 5-10 | 5 | 5 |
| 10-15 | 6 | 11 (F) |
| 15-20 | 15 f | 26 |
| 20-25 | 10 | 36 |
| 25-30 | 5 | 41 |
| 30-35 | 4 | 45 |
| 35-40 | 2 | 47 |
| 40-45 | 2 | 49 |
| | $N = 49$ | |

We have,   N=49

$\therefore \dfrac{N}{2} = \dfrac{49}{2} = 24.5$

The cumulative frequency just greater than N/2 is 26 and the corresponding class is 15-20. Thus, 15-20 is the median class such that

$l = 15, f = 15. \ F = 11 \ and \ h = 5$

$$\therefore \ Median = l + \frac{\dfrac{N}{2} - F}{f} \times h = 15 + \frac{24.5 - 11}{15} \times 5 = 15 + \frac{13.5}{3} = 19.5$$

**Example 10 :** If the median of the distribution given below is 28.5, find the value of $x$ and $y$ .

Class interval:       0-10   10-20  20-30  30-40  40-50  50-60

No. of students:     5       $x$     20     15     $y$     5

**Solution:**               Computation of Median

| Class intervals | Frequency (f) | Cumulative Frequency (cf) |
|---|---|---|
| 0-10 | 5 | 5 |
| 10-20 | $x$ | $5 + x$ (F) |
| 20-30 | 20 f | $25 + x$ |
| 30-40 | 15 | $40 + x$ |
| 40-50 | $y$ | $40 + x + y$ |
| 50-60 | 5 | $45 + x + y$ |
| | $\sum f_1 = 60$ | |

We have,

    Median = 28.5

Clearly, it lies in the class interval 20-30., So, 20-30 is the median class.

$\therefore \quad l = 20, h = 10, f = 20, \ F = 5 + x \ and \ N = 60 , \dfrac{N}{2} = 30$

Now,

$$Median = l + \frac{\dfrac{N}{2} - F}{f} \times h$$

$$\Rightarrow \ 28.5 = 20 + \frac{30 - (5 + x)}{20} \times 10$$ **www.sakshieducation.com**

$\Rightarrow \quad 28.5 = 20 + \dfrac{3-(5+x)}{20} \times 10$

$\Rightarrow \quad 28.5 = 20 + \dfrac{25-x}{2}$

$\Rightarrow \quad 8.5 = \dfrac{25-x}{2} \Rightarrow 25 - x = 17 \Rightarrow x = 8$

We have,

$N = 60$

$\therefore \quad 45 + x + y = 60 \Rightarrow x + y = 15$

*Putting $x = 8$ in $x + y = 15$, we get $y = 7$*

*Hence, $x = 8$ and $y = 7$.*

## Mode

A mode is that value among the observations which occurs most frequently.

Before learning about calculating the mode of grouped data let us first recall how we found the mode for ungrouped data through the following example.

**Example: 11    The wickets taken by a bowler in 10 cricket matches are as follows: 2, 6, 4, 5, 0, 2, 1, 3, 2, 3. Find the mode of the data.**

**Solution:** Let us arrange the observations in order i.e., 0, 1, 2, 2, 2, 3, 3, 4, 5, 6

Clearly, 2 is the number of wickets taken by the bowler in the maximum number of matches (i.e., 3 times) So, the mode of this data is 2.

In a grouped frequency distribution, it is not possible to determine the mode by looking at the frequencies. Here, we can only locate a class with the maximum frequency, called the modal class. The mode is a value inside the modal class, and is given by the formula.

$$\text{Mode} = l + \left( \dfrac{f_i - f_o}{2f_i - f_o - f_2} \right) \times h$$

Where,        $l$ = lower boundary of the modal class,

h = size of the modal class interval,

$f_i$ = frequency of the modal class,

$f_o$ = frequency of the class preceding the modal class,

$f_2$ = frequency of the class succeeding the modal class.

Let us consider the following examples to illustrate the use of this formula.

**Example: 12** **A survey conducted on 20 households in a locality by a group of students resulted in the following frequency table for the number of family members in a household.**

| Family size | 1-3 | 3-5 | 5-7 | 7-9 | 9-11 |
|---|---|---|---|---|---|
| Number of families | 7 | 8 | 2 | 2 | 1 |

Find the mode of this data.

**Solution:** Here the maximum class frequency is 8, and the class corresponding to this frequency is 3-5. So the modal class is 3-5.

Now,

Modal class = 3-5, boundary limit $(l)$ of modal class = 3, class size (h) = 2

Frequency of the modal class $(f_1) = 8$,

Frequency of class preceding the modal class $(f_0) = 7$,

Frequency of class succeeding the modal class $(f_2) = 2$.

Now, let us substitute these values in the formula –

$$\text{Mode} = l + \left( \frac{f_1 - f_o}{2f_1 - f_0 - f_2} \right) \times h$$

$$= 3 + \left( \frac{8-7}{2 \times 8 - 7 - 2} \right) \times 2 = 3 + \frac{2}{7} = 3.286$$

Therefore, the mode of the data above is 3.286.

**Example: 13**   **The marks distribution of 30 students in a mathematics examination are given in the adjacent table. Find the mode of this data. Also compare and interpret the mode and the mean.**

| Class interval | Number of students ($f_i$) | Class Marks ($x_i$) | $f_i x_i$ |
|---|---|---|---|
| 10-25 | 2 | 17.5 | 35.0 |
| 25-40 | 3 | 32.5 | 97.5 |
| 40-55 | 7 | 47.5 | 332.5 |
| 55-70 | 6 | 62.5 | 375.0 |
| 70-85 | 6 | 77.5 | 465.0 |
| 85-100 | 6 | 92.5 | 555.0 |
| Total | $\sum f_i = 30$ | | $\sum f_i x_i = 1860.0$ |

**Solution:**   Since the maximum number of students (i.e., 7) have got marks in the interval, 40-65 the modal class is 40-55.

The tower boundary ($l$) of the modal class = 40.

The class size (h) = 15,

The frequency of modal class ($f_i$) = 7,

The frequency of the class preceding the modal class ($f_0$) = 3,

The frequency of the class succeeding the modal class ($f_2$) = 6.

Now using the formula :

$$\text{Mode} = l + \left( \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) \times h$$

$$= 40 + \left( \frac{7 - 3}{2 \times 7 - 6 - 3} \right) \times 15 = 40 + 12 = 52$$

**Example: 14** **Find the mode of the following data:**

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

**Solution:** Let us first form the frequency table for the given data as given below:

Value $x$ : 110 120 130 140

Frequency $f_i$ : 2 4 2 2

We observe that the value 120 has the maximum frequency.

Hence, the mode or modal value is 120.

**Example: 15** **For the following grouped frequency distribution find the mode :**

**Class:** 3-6 6-9 9-12 12-15 15-18 18-21 21-24

**Frequency:** 2 5 10 23 21 12 3

**Solution:** We observe that the class 12-15 has maximum frequency. Therefore, this is the modal class such that $l = 12$, $h = 3$, $f = 23$, $f_1 = 10$ and $f_2 = 21$.

$$\therefore \quad Mode = l + \frac{f - f_1}{2f - f_1 - f_2} \times h$$

$$\Rightarrow \quad Mode = 12 + \frac{23 - 10}{46 - 10 - 21} \times 3 = 12 + \frac{13}{15} \times 3 = 12 + \frac{13}{5} = 14.6$$

**Example: 16** **The following data gives the distribution of total household expenditure (in rupees) of manual workers in a city:**

| Expenditure (in ₹) | Frequency | Expenditure (in ₹) | Frequency |
|---|---|---|---|
| 1000-1500 | 24 | 3000-3500 | 30 |
| 1500-2000 | 40 | 3500-4000 | 22 |
| 2000-2500 | 33 | 4000-4500 | 16 |
| 2500-3000 | 28 | 4500-5000 | 7 |

Find the average expenditure which is being done by the maximum number of manual workers.

**Solution:** We know that the mode is the value of the variable which occurs maximum number of times in a frequency distribution. So, the average expenditure done by the maximum number of workers is the modal value. We observe that the class 1500-2000 has the maximum frequency 40. So, it is the modal class such that $l = 1500, h = 500, f = 40, f_1 = 24$ *and* $f_2 = 33$.

$$\therefore \quad Mode = l + \frac{f - f_1}{2f - f_1 - f_2} \times h$$

$$\Rightarrow \quad Mode = 1500 + \frac{40 - 24}{80 - 24 - 33} \times 500 = 1500 + \frac{16}{23} \times 500 = 1847.826.$$

### Relationship among Mean, Median and Mode

We have learnt about three measures of central value, namely; arithmetic mean, median and mode. These three measures are closely connected by the following relations.

Mode = 3 Median – 2 Mean

or, Median = Mode + $\frac{2}{3}$ (Mean – Mode)

or, Mean = Mode + $\frac{3}{2}$ (Median – Mode)

### Cumulative Frequency Polygon Curve (An Ogive)

There are two methods of constructing a frequency polygon and an ogive. Let us now discuss the two methods.

(i) Less than method                    (ii) More than method.

### Less than Method

To construct a cumulative frequency polygon and an ogive by less than method, we use the following algorithm

## Algorithm

Step I        Start with the upper limits of class intervals and add class frequencies to obtain the cumulative frequency distribution.

Step II       Mark upper class limits along X- axis on a suitable scale.

Step III        Mark cumulative frequencies along Y – axis on a suitable scale.

Step IV        Plot the points $(x_i, f_i)$ *where* $x_i$ is the upper limit of a class and $f_i$ is corresponding cumulative frequency.

Step V         Join the points obtained in step IV by a free hand smooth curve to get the ogive and to get the cumulative frequency polygon join the points obtained in step IV by line segments.

## More than Method

To construct a cumulative frequency polygon and an ogive by more than method, we use the following algorithm.

### Algorithm

Step I         Start with the lower limits of the class intervals and from the total frequency subtract the frequency of each class to obtain the cumulative frequency distribution.

Step II        Mark the lower class limits along X – axis on a suitable scale.

Step III       Mark the cumulative frequencies along Y – axis on a suitable scale.

Step IV        Plot the points $(x_i, f_i)$ *where* $x_i$ is the lower limit of a class and $f_i$ is the corresponding cumulative frequency.

Step V         Join the points obtained in step IV by a free hand smooth curve to get the ogive and to get the cumulative frequency polygon join these points by line segments.

**Example 17    Draw a cumulative frequency curve and cumulative frequency polygon for the following frequency distribution by less than method.**

| Age (in years): | 0-9 | 10-19 | 120-29 | 30-39 | 40-49 | 50-59 | 60-69 |
|---|---|---|---|---|---|---|---|
| No. of persons: | 5 | 15 | 20 | 23 | 17 | 11 | 9 |

**Solution:**   The give frequency distribution is not continuous. So, we first make it continuous and prepare the cumulative frequency distribution as under:

| Age (in years) | Frequency | Age less than | Cumulative Frequency |
|---|---|---|---|
| -0.5-9.5 | 5 | 9.5 | 5 |
| 9.5-19.5 | 15 | 19.5 | 20 |
| 19.5-29.5 | 20 | 29.5 | 40 |
| 29.5-39.5 | 23 | 39.5 | 63 |
| 39.5-49.5 | 17 | 49.5 | 80 |
| 49.5-59.5 | 11 | 59.5 | 91 |
| 59.5-69.5 | 9 | 69.5 | 100 |
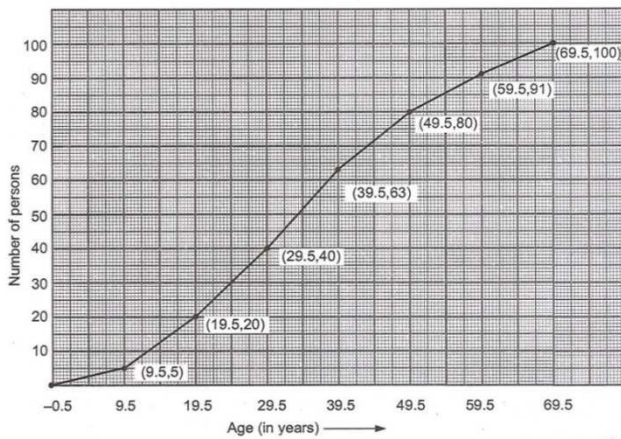


**Cumulative frequency curve**

Now, we plot points (9.5, 5), (19.5, 20), (29.5, 40), (39.5, 63), (49.5, 80), (59.5, 91) and (69.5, 100) and join them by a free hand smooth curve to obtain the required ogive as shown in fig. The cumulative frequency polygon is obtained by joining these points by line segments as shown in Fig.
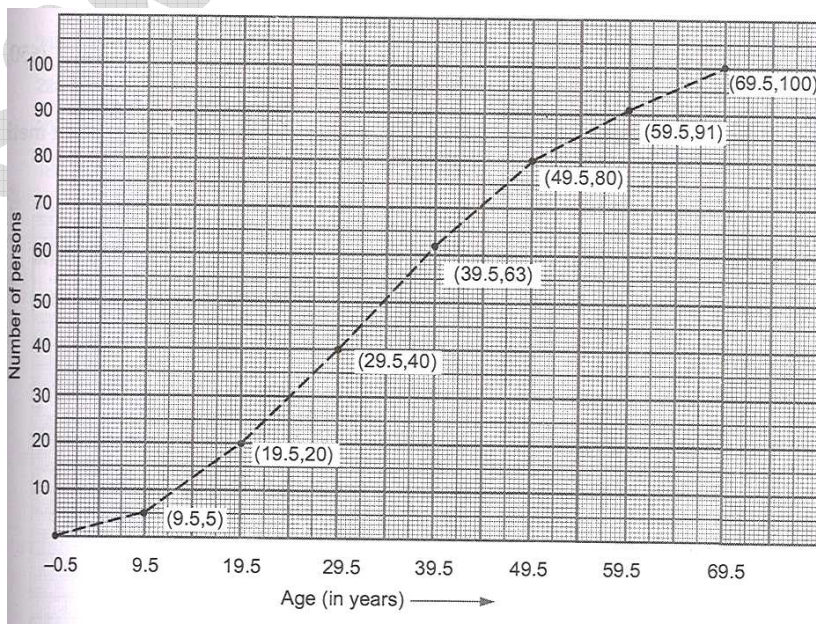


**Cumulative frequency polygon**

## Some Applications of Ogives

Ogives can be used to find the median of a frequency distribution. In order to determine the same, we may use the following algorithms:

## Algorithm 1

Step I Draw any one of the two types of frequency curves on the graph paper.

Step II Compute $\frac{N}{2}(N = \sum f_i)$ and mark the corresponding point on y- axis.

Step III Draw a line parallel to $x - axis$, from the point makred in step Ii, cutting the cumulative frequency curve at a point P (say).

Step IV Draw perpendicular PM from P on the X- axis. The x- coordinate of point M gives the median.

## Algorithm 2

Step I Draw less than type and greater than type cumulative frequency curves on the graph paper.

Step II Mark the point of intersection of the two curves drawn in step I. Let this point be P

Step III Draw perpendicular PM from P on the X – axis. The x – coordinate of point M gives the Median.

**Example 18:** **The table below gives the percentage distribution of female teachers in the primary schools of rural areas of various states and union territories (U.T) of India. Find the mean percentage of female teachers using all the three methods.**

| Percentage of female teachers | 15-25 | 25-35 | 35-45 | 45-55 | 55-65 | 65-75 | 75-85 |
|---|---|---|---|---|---|---|---|
| Number of States / U.T. | 6 | 11 | 7 | 4 | 4 | 2 | 1 |

**Solution:** Let us find the class marks $x_i$ of each class, and put them in a table

Here we take a = 50, h = 10,

Then $d_i = x_i - 50$ *and* $u_i = \dfrac{x_i - 50}{10}$

We now find $d_i$ *and* $u_i$ and put them in the table

| Percentage of female teachers | Number of States/ U.T. | $x_i$ | $d_i = x_i - 50$ | $u_i = \dfrac{x_i - 50}{10}$ | $f_i x_i$ | $f_i d_i$ | $f_i u_i$ |
|---|---|---|---|---|---|---|---|
| 15-25 | 6 | 20 | -30 | -3 | 120 | -180 | -18 |
| 25-35 | 11 | 30 | -20 | -2 | 330 | -220 | -22 |
| 35-45 | 7 | 40 | -10 | -1 | 280 | -70 | -7 |
| 45-55 | 4 | 50 | 0 | 0 | 200 | 0 | 0 |
| 55-65 | 4 | 60 | 10 | 1 | 240 | 40 | 4 |
| 65-75 | 2 | 70 | 20 | 2 | 140 | 40 | 4 |
| 75-85 | 1 | 80 | 30 | 3 | 80 | 30 | 3 |
| Total | 35 | | | | 1390 | -360 | -36 |

From the table above, we obtain $\sum f_i = 35, \sum f_i x_i = 1390, \sum f_i d_i = -360, \ \sum f_i u_i = -36.$

Using the direct method $\bar{x} = \dfrac{\sum f_i x_i}{\sum f_i} = \dfrac{1390}{35} = 39.71$

Using the assumed mean method $\bar{x} = a + \dfrac{\sum f_i d_i}{\sum f_i} = 50 + \dfrac{360}{35} = 50 - 10.29 = 39.71$

Using the step – deviation method $\bar{x} = a + \left( \dfrac{\sum f_i u_i}{\sum f_i} \right) \times h = 50 + \dfrac{-36}{35} \times 10 = 39.71$

Therefore, the mean percentage of female teachers in the primary schools of rural areas is 39.71.

**Example - 19** **The distribution below shows the number of wickets taken by bowlers in one-day cricket matches. Find the mean number of wickets by choosing a suitable method. What does the mean signify?**

| Number of wickets | 20-60 | 60-100 | 100-150 | 150-250 | 250-350 | 350-450 |
|---|---|---|---|---|---|---|
| Number of bowlers | 7 | 5 | 16 | 12 | 2 | 3 |

**Solution:** Here, the class size varies, and the $x_i's$ are large. Let us still apply the step deviation method with a = 200 and h = 20. Then, we obtain the data as given in the table.

| Number of wickets | Number of bowlers $f_i$ | $x_i$ | $d_i = x_i - a$ | $u_i = \dfrac{x_i - a}{h}$ (h = 20) | $f_i u_i$ |
|---|---|---|---|---|---|
| 20-60 | 7 | 40 | -160 | -8 | -56 |
| 60-100 | 5 | 80 | -120 | -6 | -30 |
| 100-150 | 16 | 125 | -75 | -3.75 | -60 |
| 150-250 | 12 | 200(a) | 0 | 0 | 0 |
| 250-350 | 2 | 300 | 100 | 5 | 10 |
| 350-450 | 3 | 400 | 200 | 10 | 30 |
| Total | 45 | | | | -106 |

So $\bar{x} = a + \left( \dfrac{\sum f_i u_i}{\sum f_i} \right) \times h = 200 + \dfrac{-106}{45} \times 20 = 200 - 47.11 = 152.89$

Thus, the average number of wickets taken by these 45 bowlers in one – day cricket is 152.89.