Chapter - 14

Statistics

Mean of Grouped Data - Direct Method

Introduction

Statistics is the study of collection, organization, interpretation, and presentation of data.

Data is a collection of numbers, measurements, observations or even descriptions of things, gathered to give some information. In our day to day life, we come across several kinds of tables consisting of numbers, figures like average cricket score of a team, election results. All these data give us some kind of information.

Application of Statistics

Statistics is widely used in many sectors such as,

- Weather forecasting
- Science and Medicine
- Psychology
- Geology

Mean of Grouped Data

The data that has been organized into several groups is called a grouped data while ungrouped data is just a list of numbers.

The mean or average of observations is the sum of the values of all the observations divided by the total number of observations.

1. Direct Method

Let x_1, x_2, x_3 ... ... , xn be the observations with respective frequencies f_1, f_2, f_3, ... f_n.

This means observation x_1 occurs f_1 times, x_2 occurs f_2 times and so on.

Sum of the values of all the observations = $f_1x_1 + f_2x_2 + f_3x_3 \ldots \ldots \ldots f_nx_n$

Number of observations = $f_1 + f_2 + f_3 \ldots \ldots \ldots . f_n$

Mean, $\overline{x} = \dfrac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots\ldots\ldots\ldots\ldots f_nx_n}{f_1 + f_2 + f_3 + \ldots\ldots fn}$

Mean, $\dfrac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}$

$\sum$ is the Greek letter which means the summation.
This method of finding the mean is known as the Direct Method.

Example: The following table shows the marks obtained by 22 students of Class X in a Mathematics paper consisting of 100 marks.

Marks obtained (x_i)	10	20	12	40	56	60	70	72
Number of students (f_i)	1	1	2	4	3	2	4	4

• Prepare the frequency table in such a way that the first column consists of the value of observations and the second column the corresponding frequencies.

• In the third column multiply the value of observations with the corresponding frequency.

• Calculate the sum of all entries in column III and column IV.

• Finally use the formula, Mean, $\bar{x} = \dfrac{\sum fixi}{\sum fi}$

Marks Obtained (x_i)	Number of students(f_i)	f_ix_i
10	1	10
20	1	20

12	2	24
40	5	200
56	3	168
60	2	120
70	4	280
72	4	288
	$\sum f_i = 22$	$\sum f_i x_i = 1110$

Mean, $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{1110}{22} = 50.45$

Example: The following table shows the weight of 12 students.

Weight in kg (x_i)	68	70	72	74	76
Number of students (f_i)	5	3	2	1	1

Find the mean weight of the students.

Weight in kg (x_i)	Number of students(f_i)	$f_i x_i$
68	5	340
70	3	210
72	2	144
74	1	74
76	1	76
	$\sum f_i = 12$	$\sum f_i x_i = 844$

Mean, $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{844}{12} = 70.33$

Method to convert Ungrouped Data into Grouped Data

In most real-life cases, the data is usually very large and to make a meaningful study it needs to be condensed as grouped data.

• We can convert an ungrouped data into grouped data by forming class intervals of equal width.

• While allocating frequencies to each class interval, the observation falling in the upper-class limit would be considered in the next class interval.

• The frequency of each class interval is centred around its midpoint.

• We can calculate the mid-point of a class (classmark) by finding the average of its upper and lower limits.

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

• Class interval or class represents the range in which the data are grouped. For example, groups 0 – 20, 20 – 40, 40 – 60, etc represent class interval.

• Lower Class Limit – the lowest number which occurs in a particular class interval is known as its lower-class limit. For

example, in the class interval 20 – 40, 20 is the lower-class limit of that interval.

• Upper-Class Limit – the highest number occurring in a particular class interval is known as its upper limit. For

example, in the class interval 20 – 40, 40 is the upper-class limit of that interval.

Example: Find the arithmetic mean of the following data.

x_i	3	5	7	9	11	13	15	17
f_i	6	8	15	9	8	4	2	3

Now we will first convert the ungrouped data into grouped data by forming the class interval of width 5.

Class Interval	$0-5$	$5-10$	$10-15$	15 - 20
f_i	6	(8 + 15 + 9) = 32	(8 + 4) = 12	(2 + 3) = 5

Class Interval	Class Marks(x_i)	f_i	$f_i x_i$
$0-5$	$\frac{0+5}{2} = 2.5$	6	15
$5-10$	$\frac{5+10}{2} = 7.5$	32	240
$10-15$	$\frac{10+15}{2} = 12.5$	12	150
$15-20$	$\frac{15+20}{2} = 17.5$	5	87.5
		$\sum f_i = 55$	$\sum f_i x_i = 492.5$

Mean, $\bar{x} = \frac{\sum fixi}{\sum fi} = \frac{492.5}{8.95} = 8.95$

Example: If the mean of the following distribution is 53.6, then find the value of p.

Class	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
Frequency	8	p	10	9	13

Class	Class marks (x_i)	Frequency(f_i)	$f_i x_i$
0 - 20	$\frac{0+20}{2} = 10$	8	80
$20-40$	$\frac{20+40}{2} = 30$	p	$30p$
$40-60$	$\frac{40+60}{2} = 50$	10	500
$60-80$	$\frac{60+80}{2} = 70$	9	630
$80-100$	$\frac{80+100}{2} = 90$	13	1170
		$\sum f_i = 40 + p$	$\sum f_i x_i = 2380 + 30p$

Mean, $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{2380 + 30p}{40 + p}$

$53.6 = \frac{2380 + 30p}{40 + p} \Rightarrow 53.6\,(40 + p) = 2380 + 30p$

$\Rightarrow 2144 + 53.6\,p = 2380 + 30p$

$\Rightarrow 53.6p - 30p = 2380 - 2144$

$23.6p = 236 \Rightarrow p = 236 \div 23.6 = 10$

Mean of Grouped Data - Assumed Mean Method

2. Assumed Mean Method

Sometimes the numerical values of x_i and fi are large and finding the product of x_i and f_i is very tedious and time-consuming.

• In this method, the first step is to choose an assumed mean among the x_i 's, and denote it by 'a'. We may take 'a' to be that x_i which lies in the center of x_1, x_2, x_3 … … . . x_n.

• The next step is to find the difference d_i between 'a' and each of the x_i 's i.e. $d_i = x_i - a$

• The third step is to find the product of di with the corresponding fi and take the sum of all the $f_i d_i$'s.

• Mean of the deviation, di = $\frac{\sum f_i d_i}{\sum f_i}$

The value of the mean obtained does not depend on 'a'.

Relation between $\overline{d}$ and $\overline{x}$

Mean of deviation $\overline{d} = \frac{\sum f_i d_i}{\sum f_i}$

so, $\overline{d} = \frac{\sum f_{i(x_i - a)}}{\sum f_i}$ $\quad (\because d_i = x_i - A)$

$\overline{d} = \frac{\sum f_i x_i}{\sum f_i} - \frac{\sum f_i a}{\sum f_i}$

$\overline{d} = \overline{x} - \frac{\sum f_i a}{\sum f_i}$ $\quad (\because \bar{x} = \frac{\sum f_i x_i}{\sum f_i})$

$\overline{d} = \overline{x} \frac{-a \sum f_i}{\sum f_i}$

$\overline{d} = \overline{x} - a \Rightarrow \overline{x} = a + \overline{d}$

$\overline{x} = a + \frac{\sum f_i d_i}{\sum f_i} (\because \overline{d} = \frac{\sum f_i d_i}{\sum f_i})$

Example: Consider the following distribution of daily wages of 50 workers of a factory:

Daily wages (in Rs.)	100 - 120	120 - 140	140 - 160	160 - 180	180 - 200
Number of workers	12	14	8	6	10

(REFERENCE: NCERT)

Find the mean daily wages of the workers of the factory.

- Prepare the frequency table.
- Choose the assumed mean 'a' from the values in the third column.
- Take deviations $d_i = x_i - a$ of the values x_i. Write these deviations against the corresponding frequency in the fourth column.
- Multiply the frequencies in column II with the corresponding deviations d_i in column IV and write the product $f_i d_i$ in column V.

• Use the formula, $\bar{x} = a + \frac{\sum f_i d_i}{\sum f_i}$ to calculate the mean.

Let the assumed mean be 150. (We have chosen the middle value as the assumed mean)

Example: Consider the following distribution of daily wages of 50 workers of a factory:

sClass	Frequency(f_i)	Classmarks(x_i)	$d_i = x_i - 150$	$f_i d_i$
100 - 120	12	$\frac{100+120}{2} = 110$	110 – 150 = -40	-480
120 – 140	14	$\frac{120+140}{2} = 130$	130 – 150 = -20	- 280
140 – 160	8	$\frac{140+160}{2} = 150 = a$	150 – 150 = 0	0
160 – 180	6	$\frac{160+180}{2} = 170$	170 – 150 = 20	120
180 – 200	10	$\frac{180+200}{2} = 190$	190 – 150 = 40	400
	$\sum f_i = 50$			$\sum f_i d_i = -240$

Mean, $\overline{x} = a + \frac{\sum f_i d_i}{\sum f_i}$

$$\overline{x} = 150 + (\frac{-240}{50}) \Rightarrow \overline{x} = 150 + (4.8) \Rightarrow \overline{x} = 145.20$$

Therefore, the mean daily wage = Rs 145.20

Mean of Grouped Data - Step Deviation Method

3. Step deviation Method

Step deviation is used in the cases, where the deviations from the assumed mean are multiples of a common number.

Let $u_i = \frac{x_i - a}{h}$ where a is the assumed mean and h is the class size.

$$\bar{u} = \frac{\sum f_i u_i}{\sum f_i}$$

Relation between ū and x

ū = $\frac{\sum f_i u_i}{\sum f_i}$

$$Now, \overline{u} = \frac{\sum f_{i(\frac{x_i - a}{h})}}{\sum f_i} (\because u_i = \frac{x_i - a}{h})$$

$$\overline{u} = \frac{1}{h}[\frac{\sum f_i x_i}{\sum f_i} - \frac{\sum f_i a}{\sum f_i}]$$

$$\overline{u} = \frac{1}{h}[\frac{\sum f_i x_i}{\sum f_i} - \frac{a \sum f_i}{\sum f_i}]$$

$$\overline{u} = \frac{1}{h}[\overline{x} - a] \qquad (\because \overline{x} = \frac{\sum f_i x_i}{\sum f_i})$$

$$h\overline{u} = \overline{x} - a$$

$$\overline{x} = a + h\overline{u}$$

$$\overline{x} = a + h(\frac{\sum f_i u_i}{\sum f_i})(\because \overline{u} = \frac{\sum f_i u_i}{\sum f_i})$$

Example: In a health check-up, the number of heartbeats of 40 women were recorded in the following table:

Number of heartbeats/minute	65 - 69	70 - 74	75 - 79	80 - 84
Number of women	2	18	10	4

Here, class intervals are not continuous. But mid-value xi of each class, the interval would be the same even if the class interval is not continuous.

Let the assumed mean be 72 and h = 5 (Class size)

Class	Frequency (f_i)	Class marks (x_i)	$d_i = x_i - 72$	$u_i = \frac{x_i - a}{h} = \frac{d_i}{h}$	$f_i u_i$
65 - 69	2	$\frac{65+69}{2} = 67$	$67 - 72 = -5$	$\frac{-5}{5} = -1$	-2
70 – 74	18	$\frac{70+74}{2} = 72 = a$	$72 - 72 = 0$	0	0
75 – 79	10	$\frac{75+79}{2} = 77$	$77 - 72 = 5$	$\frac{5}{5} = 1$	10
80 – 84	4	$\frac{80+84}{2} = 82$	$82 - 72 = 10$	$\frac{10}{5} = 2$	8
	$\sum f_i = 40$				$\sum f_i u_i = 16$

Mean, $\overline{x} = a + h(\frac{\sum f_i u_i}{\sum f_i})$

$$\overline{x} = 72 + 5\frac{16}{40} \Rightarrow \overline{x} = 74$$

Mean = 74

- The step deviation method is convenient to apply if all the d_i's have a common factor.
- The mean obtained by all the three methods are same.
- The assumed mean method and step deviation method are just simplified forms of the direct method.
- If the data is discontinuous, there is no need to convert it into continuous data, while finding the mean. Because the mid value x_i of each class interval is same even if the class interval is discontinuous.

Mean

• It is the most commonly used central tendency because it takes into account all the observation, and lies between the extremes.

• It is useful in comparison of two or more distributions.

• If we compare the mean results of the students of a class then we can easily tell which student is the best performer.

• However, extremes values in the data affect the mean.

• The mean of classes having frequencies more or less the same will be a good representative of data. But if the frequency of one of the classes varies largely from the other frequencies then mean is not a good representative of the data.

Mode of Grouped Data

Mode of Grouped Data

Mode is that value among the observations which occurs most often or it is the value of the observation having the maximum frequency.

Mode of the numbers 2, 3, 4, 4, 4, 5, 5, 6, 6, 8, 8, 8, 8 and 9 is 8 because it is repeated the maximum number of times.

Sometimes, in data, more than one value may have the same maximum frequency. Such data is called multimodal.

In a grouped frequency distribution, it is not possible to determine the mode by looking at the frequencies. So, here we first locate a class with the maximum frequency. This class is called the modal class.

Class Interval	0 - 10	10 - 20	20 - 30	30 - 40	40 – 50	50 – 60	60 - 70
Number of students	2	9	14	20	22	28	30

In the above example, the modal class is 50 – 60 as the highest frequency, 28 is of class 50 – 60.

The mode is a value inside the modal class and is given by the formula:

$\text{Mode} = l + (f1 - f02f1 - f0 - f2) \times h$

lower limit of the modal class $(l) = 50$

size of the class interval$(h) = 10$

frequency of the modal class$(f_1) = 28$

frequency of the class preceding the modal class(f_0) = 22

frequency of the class succeeding the modal class(f_2) = 30

Mode = l + ($\frac{f_1 + f_0}{2f_1 - f_0 - f_2}$) × h

Mode = 50 + ($\frac{28 - 22}{2X28 - 22 - 30}$) × 10 = 50 + ($\frac{6}{56 - 52}$) = 50 + $\frac{6}{4}$

= $\frac{200 + 6}{4}$ = $\frac{206}{4}$ = 51. 5

Example: An NGO working for the welfare of TB patients, maintained its record as follows:

Age of patient (in years)	0 – 20	20 - 40	40 - 60	60 - 80
Number of patients	30	300	125	50

We see that the maximum frequency is 315 and the class corresponding to this frequency is 20 – 40.

Therefore, the modal class is 20 – 40.

Here, lower limit(l) of the modal class = 20,

Class size(h) = 20,

Frequency(f_1) of the modal class = 300,

Frequency(f_0) of class preceding the modal class= 30,

Frequency (f_2) of class succeeding the modal class = 125

Mode = l + ($\frac{f_1 + f_0}{2f_1 - f_0 - f_2}$) × h

Mode = 20 + ($\frac{300 - 30}{2X300 - 35 - 125}$) × 20 ⇒ 20 + ($\frac{270}{600 - 160}$) × 20

Mode $= 20 + (\frac{270}{440}) \times 20 = 20 + 12.27 = 32.27$

Therefore, the average age of the maximum number of students = 32.27

Example: Find the mode of the following data.

Class	1 - 3	3 - 5	5 - 7	7 - 9	9 – 11
Frequency	14	16	4	3	3

Here, the maximum frequency is 16 and the class corresponding to this frequency is 3 – 5. So, 3 – 5 is the modal class.

Here, lower limit (l) of the modal class = 3,

Class size (h) = 2,

Frequency (f1) of the modal class = 16,

Frequency (f0) of class preceding the modal class= 14,

Frequency (f2) of class succeeding the modal class = 4

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

Mode $= 3 + (\frac{16 - 14}{2X16 - 14 - 4}) \times 2 \Rightarrow 3 + (\frac{2}{32 - 18}) \times 2$

Mode $= 3 + (\frac{2}{14}) \times 2 = 3 + 0.2857 = 3.2857$

Therefore, mode = 3.2857

Example: Monthly consumption of electricity of some consumers is given below as a distribution. Find the missing frequency (x), if the mode of distribution is given to be

200 units.

Monthly consumption (in units)	90 - 120	120 - 150	150 - 180	180 - 210	210 - 240
Number of consumers	30	25	x	65	40

The given mode is 200 units, which lies between 180 – 210.

So, 180 – 210 is the modal class.

Here, lower limit (l) of the modal class = 180,

Class size (h) = 30,

Frequency (f1) of the modal class = 65,

Frequency (f0) of class preceding the modal class= x,

Frequency (f2) of class succeeding the modal class = 40

Mode = l + $(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}) X h$

$200 = 180 + (\frac{65 - x}{2X65 - x - 40}) \times 30$

$\Rightarrow 200 - 180 = (\frac{65 - x}{40 - x}) \times 30$

$20 = \frac{65 - x}{40 - x} \times 30$

$20(90 - x) = 30(65 - x) \Rightarrow 1800 - 20x = 1950 - 30x$

$30x - 20x = 1950 - 1800$

$10x = 150$

$x = 15$

Therefore, the missing frequency (x) = 15

Mode

• In cases where we have to find the most frequent value or the most popular item then the mode is the best choice.

• For example, the most popular game among students of a class, the most popular song of the week.

Median of Grouped Data

Median of Grouped Data

The median is a measure of central tendency which gives the value of the middlemost observation in the data.

A measure of central tendency is a single value that describes the center of the data. It is also called an average or the center of the distribution.

To find the median of the ungrouped data, we first arrange the data values of the observation in ascending order.

Let n be the total number of observations.

• Case I – If n is odd

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

• Case II – If n is even

Median is the average of the n2 th and the$\left(\frac{n}{2} + 1\right)$ th observations

Cumulative Frequency

The frequency of observation in data is the number of times that observation occurs in the data.

Cumulative frequency of a class is defined as the sum of all frequencies up to the given class.

Cumulative Frequency Distribution

Cumulative frequency distribution is of two types:

• Less than type

• More than type

We will understand the formation of these two distributions with the help of examples.

Example: The following distribution gives the daily income of 50 workers of a factory:

Daily income (in Rs)	100 - 120	120 - 140	140 - 160	160 - 180	180 - 200
Number of workers	12	14	8	6	10

Write the above distribution as 'less than type' cumulative frequency.

(REFERENCE: NCERT)

The given distribution is less than type.

The number of workers earning less than Rs $120 = 12$

The number of workers earning less than Rs 140 also includes the number of workers earning between Rs 100 to Rs 120 as well as those earning between Rs 120 to Rs 140.

Therefore, the total numbers of workers earning less than Rs 140

$=12 + 14 = 26$

Similarly, the number of workers earning less than Rs 160

$= 12 + 14 + 8 = 34$

The number of workers earning less than Rs 180

$= 12 + 14 + 8 + 6 = 40$

Finally, the number of workers earning less than Rs 200

$= 12 + 14 + 8 + 6 + 10 = 50$

Daily Income (in Rs)	Number of students (Cumulative Frequency)
Less than 120	12
Less than 140	26
Less than 160	34
Less than 180	40
Less than 200	50

Example: The following distribution gives cumulative frequencies of 'more than type':

Marks obtained (more than or equal to)	5	10	15	20
Number of students (cumulative frequency)	30	24	9	2

Change the above data into a continuous grouped frequency distribution.

The given distribution is more than type distribution.

We see that 30 students have scored more than or equal to 5.

Now, 23 students have scored more than or equal to 10.

Therefore, the number of students in class 5 – 10 is 30 – 24 $= 6$

Similarly, the number of students in class 10 – 15 is 24 – 9 $= 15$ and the number of students in class 15 – 20 is 9 – 2 $= 7$

Number of students scoring more than or equal to 20 is 2

Class	Frequency(f_i)
5 – 10	6
10 – 15	15
15 – 20	7
More than equal to 20	2

Median of Grouped Data

In grouped data, it is difficult to find the middle observation (median) just by looking at the cumulative frequencies as the middle observation will be some value in the class interval.

Therefore, we find the value inside a class that divides the whole distribution into two halves.

- We first find the cumulative frequencies of all the classes and $\frac{n}{2}$, where n is the number of observations

- Then we use the formula of the median to calculate the median.

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

Where l = lower limit of the median class

n = number of observations

cf = cumulative frequency of the class preceding the median class

f = frequency of the median class

h = class size

The class whose cumulative frequency is greater than (and nearest to $\frac{n}{2}$) is the median class. (nis the number of observations)

Example: The following table gives the literacy rate (in %) of 25 cities.

Literacy Rate (in percent)	50 - 60	60 - 70	70 - 80	80 - 90
Number of cities	9	6	7	3

Find the median class and modal class.

The cumulative frequency table is

Literacy Rate	Frequency(f_i)	Cumulative frequency (cf)
50 - 60	9	9
60 – 70	6	9 + 6 = 15
70 – 80	8	9 + 6 + 7= 22
80 – 90	2	9 + 6 + 7 + 3 = 25

Now n = 25

So, $\frac{n}{2} = \frac{25}{2} = 12.5$

The cumulative frequency just greater than 12.5 is 15 and the corresponding class is 60 – 70.

Therefore, the median class = 60 – 70

The maximum frequency is 9 and the class corresponding to this frequency is 50 – 60.

Therefore, the modal class is 50 – 60.

Example: Find the median for the following data.

Height (in cm) [less than]	120	140	160	180	200
Number of students	12	26	34	39	50

We will first convert the above data into a continuous grouped frequency distribution.

The given distribution is of less than type.

We see that the number of students having a height of less than 120 cm = 12

Now, 26 students have got height less than or equal to 140.

Therefore, the number of students having a height in class 120 – 140 is 26 – 12 = 14

Similarly, the number of students having a height in class 140 – 160 is 34 – 26 = 8

The number of students having a height in class 160 – 180 is 39 – 34 = 5

Finally, the number of students having a height in class 180 – 200 is 50 – 39 = 11

Height (in cm)	Frequency(f_i)	Cumulative frequency (cf)
Below 120	12	12
120 - 140	14	12 + 14 = 26
140 – 160	8	12+ 14 + 8 = 34
160 – 180	5	12 + 14 + 8 + 5 = 39
180 – 200	11	12 + 14 + 8 + 6 + 11 = 50

Here, $n = 50$

So, $\frac{n}{2} = \frac{50}{2} = 25$

The cumulative frequency just greater than 25 is 26and the corresponding class is 120 – 140.

Therefore, the median class = 120 – 140

l, the lower limit = 120

cf (cumulative frequency of the class preceding 120 – 140) = 12

f(frequency of the median class 120 - 140) = 14

h (class size) = 20

$$Mediab = l + (\frac{\frac{n}{2} - cf}{f})$$

$$120 + (\frac{\frac{50}{2} + 12}{14})X20 = 120 + (\frac{25 - 12}{14})X20$$

$$= 120 + (\frac{13}{7})X10 = \frac{840 + 130}{7} = \frac{970}{7}$$

=138.57 CM

Example: If the median of the following frequency distribution is 24, then find the missing frequency x

Age (in years)	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
Number of persons	5	25	x	17	8

Age (in years)	Frequency(f_i)	Cumulative frequency (cf)
0 – 10	5	5
10 – 20	25	5 + 25 = 30
20 – 30	x	5 + 25 + x = 30 + x
30 - 40	18	5 + 25 + x + 17 = 47 + x
40 - 50	7	5 + 25 + x +18 + 8 = 55 + x

It is given that the median is 24 which lies in class 20 – 30.

Therefore, the median class = 20 - 30

l, the lower limit = 20

cf (cumulative frequency of the class preceding 20 – 30) = 30

f (frequence of the median class20-30)=x

h(class size)=10

$$Median = l + (\frac{\frac{n}{2} - cf}{f})Xh$$

$$24 = 20 + \left(\frac{\frac{55+x}{2} - 30}{x}\right) X10 \Rightarrow 24 - 20$$

$$= \left(\frac{55 + x - 60}{2Xx}\right) X10$$

$$4 = \left(\frac{x - 5}{2x}\right)_{\text{x10}}$$

$8x = (x - 5) \times 10 \Rightarrow 8x = 10x - 50$

$50 = 10x - 8x \Rightarrow 2x = 50$

$x = 25$

Therefore, the missing frequency is 25.

Median

• In cases where individual observations are not important, and we wish to find out a 'typical' observation then the median is a better measure of central tendency.

• For example, if we have to find the average wage in a country then the median is more appropriate.

Cumulative Frequency Distribution

Cumulative Frequency

The frequency of observation in data is the number of times that observation occurs in the data.

Cumulative frequency of a class is defined as the sum of all frequencies up to the given class.

Cumulative Frequency Distribution

The cumulative frequency distribution is of two types:

• Less than type

• More than type

We will understand the formation of these two distributions with the help of examples.

Example: The following distribution gives the daily income of 50 workers of a factory:

Daily income (in Rs)	100 - 120	120 - 140	140 - 160	160 - 180	180 - 200
Number of workers	12	14	8	6	10

Write the above distribution as 'less than type' cumulative frequency.

(REFERENCE: NCERT)

The given distribution is less than type.

The number of workers earning less than Rs $120 = 12$

The number of workers earning less than Rs 140 also includes the number of workers earning between Rs 100 to Rs 120 as well as those earning between Rs 120 to Rs 140.

Therefore, the total numbers of workers earning less than Rs $140 =$

$12 + 14 = 26$

Similarly, the number of workers earning less than Rs 160

$= 12 + 14 + 8 = 34$

The number of workers earning less than Rs 180

$= 12 + 14 + 8 + 6 = 40$

Finally, the number of workers earning less than Rs 200

$= 12 + 14 + 8 + 6 + 10 = 50$

Daily Income (in Rs)	Number of students (Cumulative Frequency)
Less than 120	12
Less than 140	26
Less than 160	34
Less than 180	40
Less than 200	50

Example: The following distribution gives cumulative frequencies of 'more than type':

Marks obtained (more than or equal to)	5	10	15	20
Number of students (cumulative frequency)	30	24	8	2

Change the above data into a continuous grouped frequency distribution.

The given distribution is more than type distribution.

We see that 30 students have scored more than or equal to 5.

Now, 23 students have scored more than or equal to 10.

Therefore, the number of students in class 5 – 10 is 30 – 24 = 6

Similarly, the number of students in class 10 – 15 is 24 – 8 = 16 and

the number of students in class 15 – 20 is 8 – 2 = 6

Number of students scoring more than or equal to 20 is 2

Class	Frequency(f_i)
5 - 10	6
10 – 15	16
15 – 20	6
More than equal to 20	2

Graphical Representation of Cumulative Frequency Distribution

A graphical representation helps us to understand the given data at a glance.

Cumulative Frequency Curve or an Ogive Curve

Cumulative Frequency curve is a curve that represents the cumulative frequency distribution of grouped data on a graph. This curve is also called an Ogive Curve.

The term 'ogive' is derived from the word ogee. An ogee is a shape consisting of a concave arc flowing into a convex arc, forming an S shaped curve with vertical ends.

The two types of Cumulative Frequency Curve (or Ogive) are

1) More than type Cumulative Frequency Curve

2) Less than type Cumulative frequency Curve

More than type Cumulative Frequency Curve

Steps to plot the ogive of more than type:

1)Mark the lower limits of the class intervals on the horizontal axis (x-axis) and the corresponding cumulative frequency on the vertical axis(y axis)

2) Now plot the points using lower limits and the corresponding cumulative frequencies on a graph paper.

3) Join the points by a free hand smooth curve to get the ogive.

Less than type Cumulative Frequency Curve

Steps to plot the ogive of less than type:

1)Mark the upper limits of the class intervals on the horizontal axis (x-axis) and corresponding cumulative frequency on the vertical axis (y-axis)

2) Now plot the points using upper limits and the corresponding cumulative frequencies on a graph paper.

3) Join the points by a free hand smooth curve to get the ogive.

Example: The following table gives production yield per hectare of wheat of 100 farms of a village:

Production yield (in kg/ha)	Number of farms
50 – 55	2
55 – 60	8
60 – 65	12
65 – 70	24
70 – 75	38
75 – 80	16

Change the distribution to a more than type distribution, and draw its ogive.

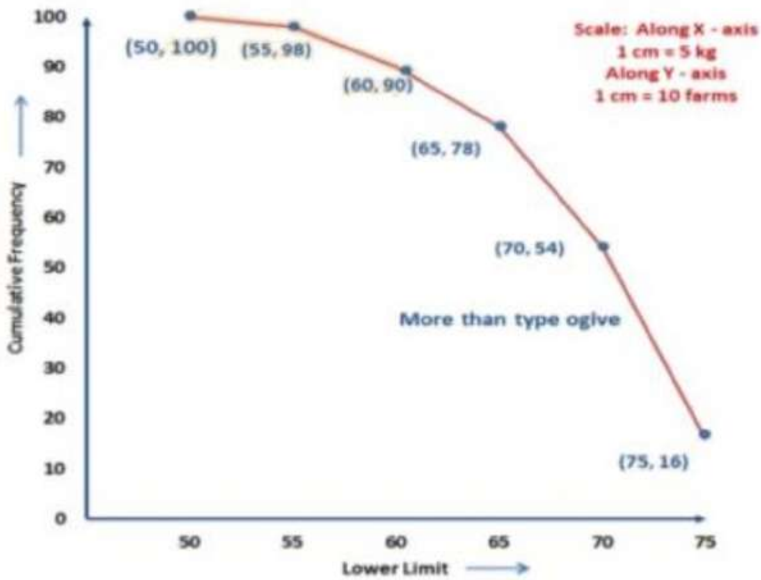
(REFERENCE: NCERT)

We first prepare the more than type cumulative frequency distribution

Production yield (in kg/ha)	Number of farms (f)	Cumulative Frequency(cf)
More than or equal to 50	2	100
More than or equal to 55	8	100 - 2 = 98
More than or equal to 60	12	98 – 8 = 90
More than or equal to 65	24	90 – 12 = 78
More than or equal to 70	38	78 – 24 = 54
More than or equal to 75	16	54 – 38 = 16
	$\sum f = 100$	

Then mark the lower limit on the x-axis and the corresponding cumulative frequency on the y-axis by choosing a convenient scale.

Plot the points (50, 100), (55, 98), (60, 90), (65, 78), (70, 54) and (75, 16) and join the points by a freehand smooth curve.



Median from Cumulative Frequency Curve or Ogive

We can use ogives to find the median of a frequency distribution.

1st method

• Locate $\frac{n}{2}$ 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.

• The point of intersection of this perpendicular with the x −axis gives the median of the data.

Example: Draw 'more than ogive' for the frequency distribution and hence obtain the median.

Class Interval	5 - 10	10 – 15	15 - 20	20 - 25	25 - 30	30 - 35	35 - 40
Frequency	3	11	2	4	3	4	3

We will first prepare the more than type cumulative frequency distribution.

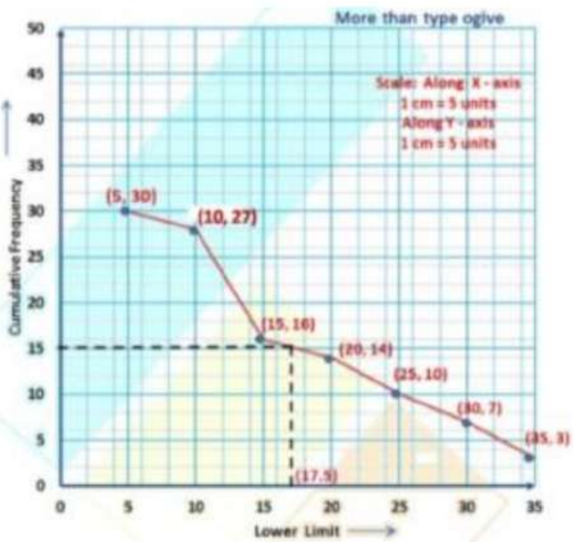
Class Interval	Frequency	Class Interval	Cumulative Frequency(cf)
$5-10$	3	More than or equal to 5	30
$10-15$	11	More than or equal to 10	30 - 3 = 27
$15-20$	2	More than or equal to 15	$27-11=16$
$20-25$	4	More than or equal to 20	$16-2=14$
$25-30$	3	More than or equal to 25	$14-4=10$
$30-35$	4	More than or equal to 30	$10-3=7$
$35-40$	3	More than or equal to 35	$7-4=3$
	$\sum f=30$		

Then mark the lower limit on the x-axis and the corresponding cumulative frequency on the y-axis by choosing a convenient scale.

Plot the points (5, 30), (10, 27), (15, 16), (20, 14), (25, 10), (30, 7) and (35, 3). Join the points by a freehand smooth curve.

• Locate $\frac{n}{2}=\frac{30}{2}=15$ on the y-axis. From this point, draw a line parallel to the x-axis cutting the curve at a point.

• The point of intersection of this perpendicular with the x-axis gives the median of the data.



2nd Method

- Draw the less than type and more than type ogive on the same axis.
- The two ogives intersect each other at a point.
- From this point draw a perpendicular on the x-axis.
- The point at which it cuts the x-axis gives the median.

Example: During the medical check-up of 35 students of a class, their weights were recorded as follows:

Weight (in kg)	Number of students
Less than or equal to 38	0
Less than or equal to 40	3
Less than or equal to 42	5
Less than or equal to 44	9
Less than or equal to 46	14
Less than or equal to 48	28
Less than or equal to 50	32
Less than or equal to 52	35

Draw a 'less than type' ogive and 'more than type' ogive for the given data. Hence, obtain the median weight from the graph.

(REFERENCE: NCERT)

The given distribution is the cumulative frequency distribution of less than type.

Weight (in kg)	Number of students (cf)
Less than or equal to 38	0
Less than or equal to 40	3
Less than or equal to 42	5
Less than or equal to 44	9
Less than or equal to 46	14
Less than or equal to 48	28
Less than or equal to 50	32
Less than or equal to 52	35

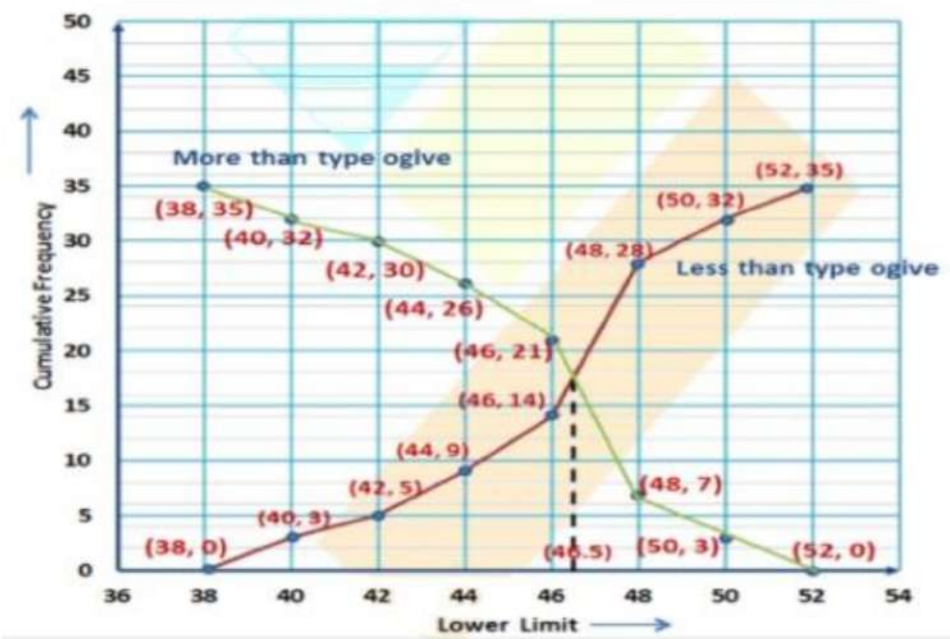
Then mark the upper limit on the x-axis and the corresponding cumulative frequency on the y-axis by choosing a convenient scale.

Plot the points (38, 0), (40, 3), (42, 5), (44, 9), (46, 14), (48, 28), (50, 32) and (52, 35). Join the points by a freehand smooth curve.

We will now prepare the more than type cumulative frequency distribution.

Weight (in kg)	Frequency	Number of students (cf)
More than or equal to 38	0	35
More than or equal to 40	$3 - 0 = 3$	$35 - 3 = 32$
More than or equal to 42	$5 - 3 = 2$	$32 - 2 = 30$
More than or equal to 44	$9 - 5 = 4$	$30 - 4 = 26$
More than or equal to 46	$14 - 9 = 5$	$26 - 5 = 21$
More than or equal to 48	$28 - 14 = 14$	$21 - 14 = 7$
More than or equal to 50	$32 - 28 = 4$	$7 - 4 = 3$
More than or equal to 52	$35 - 32 = 3$	$3 - 3 = 0$

Now we plot the points (38, 35), (40, 32), (42, 30), (44, 26), (46, 21), (48, 7), (50, 3) and (52, 0). Join the points by a freehand smooth curve to obtain an ogive of more than type.



The two ogives intersect each other at a point. From this point draw a perpendicular on the x-axis. The point at which it cuts the x-axis gives the median. Therefore, the required median is 46.5 kg.