

■ “Practice Makes Perfect”

Introduction

It is the branch of mathematics which deals with the collection, presentation, analysis and interpretation of numerical data.

In singular form, statistics is taken as a subject. And, in plural form, statistics means data.

Class interval: Each group into which the raw data is condensed, is called a class-interval.

Class limits: Each class is bounded by two figures, which are called **class limits**. The figure

on the left side of a class is called its **lower limit** and that on its right is called its **upper limit**.

Exclusive form (or continuous interval form): A frequency distribution in which the upper limit of each class is excluded and lower limit is included, is called an exclusive form.

Example

Suppose the marks obtained by some students in an examination are given. We may consider the classes 0 – 10, 10 – 20 etc. In class 0 – 10, we include 0 and exclude 10. In class 10–20, we include 10 and exclude 20.

Inclusive form (or discontinuous interval form):

A frequency distribution in which each upper limit as well as lower limit is included, is called an inclusive form. Thus, we have classes of the form 0 – 10, 11 – 20, 21 – 30 etc.

In 0 – 10, both 0 and 10 are included.

Important terms related to grouped data:

Class boundaries or true upper and true lower limits:

- (i) In the exclusive form, the upper and lower limits of a class are respectively known as the true upper limit and true lower limit.
- (ii) In the inclusive form, the number midway between the upper limit of a class and lower limit of the subsequent class gives the true upper limit of the class and the true lower limit of the subsequent class. Thus, in the above table of inclusive form, we have:

true upper limit of class 1 – 10 is $\left(\frac{10+11}{2}\right)$ = 10.5 and true lower limit of class 11 – 20 is 10.5.

Similarly, true upper limit of class 11 – 20 is $\left(\frac{20+21}{2}\right)$ = 20.5, and true lower limit of class 21 – 30 is 20.5.

Class size

The difference between the true upper limit and the true lower limit of a class is called its class size.

Class mark of a class

$$\text{Class mark} = \left(\frac{\text{True upper limit} + \text{True lower limit}}{2} \right)$$



Be Alert !

- The difference between any two successive class marks gives the class size.



**Do You
Remember ?**

- Average is the statistic which describes the center of a set of data, a set of numbers which are measurements or counts.

The most commonly used averages are the mean (arithmetic average), mode (most frequent number), median (middle number when numbers are listed smallest to largest).



The class marks of a frequency distribution are 7, 13, 19, 25, 31, 37, 43. Find the class-size and all the class-intervals.

Solution

Class size = Difference between two successive class-marks = $(13 - 7) = 6$.

Let the lower limit of the first-class interval be a . Then, its upper limit = $(a + 6)$.

$$\therefore \frac{a + (a + 6)}{2} = 7 \Rightarrow 2a + 6 = 14 \Rightarrow 2a = 8 \Rightarrow a = 4$$

So, the first class-interval is $4 - 10$.

Let the lower limit of last class-interval be b .

Then, its upper-class limit = $(b + 6)$.

$$\therefore \frac{b + (b + 6)}{2} = 43 \Rightarrow 2b + 6 = 86 \Rightarrow 2b = 80 \Rightarrow b = 40.$$

So, the last class-interval is $40 - 46$.

Hence, the required class-intervals are

$4 - 10, 10 - 16, 16 - 22, 22 - 28, 28 - 34, 34 - 40$ and $40 - 46$.

An average tends to lie centrally with the values of the variable arranged in ascending order of magnitude. So, we call an average a measure of central tendency of the data.

Three measures of central tendency are useful for analysing the data, namely

- (a) Mean (b) Median (c) Mode



**Do You
Remember ?**

□ We know that the mean of observations is the sum of the values of all the observations divided by the total number of observations i.e., if $x_1, x_2, x_3, \dots, x_n$ are n observations, then

$$\text{mean } \bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad \text{or} \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n},$$

where $\sum_{i=1}^n x_i$ denotes the sum $x_1 + x_2 + x_3 + \dots + x_n$.

Arithmetic mean

The arithmetic mean of grouped data may also be calculated by any one of the following methods:

- Direct method
- Assumed-mean method

Direct method

$x_1, x_2, x_3, \dots, x_n$ are observations with respective frequencies $f_1, f_2, f_3, \dots, f_n$ then mean, (\bar{x}) is defined by

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \dots + f_n x_n}{f_1 + f_2 + f_3 + \dots + f_n} \quad \text{or} \quad \bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i},$$

Where $\sum_{i=1}^n f_i = f_1 + f_2 + f_3 + \dots + f_n = N$

The following steps should be followed in finding the arithmetic mean of grouped data by direct method.

Step - 1: Find the class mark (x_i) of each class using,

$$x_i = \frac{\text{Lower limit} + \text{Upper limit}}{2}$$

Step - 2: Calculate $f_i x_i$ for each i

$$\frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i},$$

Step - 3: Use the formula : mean $\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}$

Sum of first ' n ' natural numbers is

$$\frac{n(n+1)}{2}$$

SPOT LIGHT



Numerical

2

Ability

Find the mean of the following data:

Class Interval	0 - 8	8 - 16	16 - 24	24 - 32	32 - 40
Frequency	6	7	10	8	9

Solution

We may prepare the table as given below:

Class interval	Frequency (f_i)	Class mark (x_i)	$f_i x_i$
0 - 8	6	4	24
8 - 16	7	12	84
16 - 24	10	20	200
24 - 32	8	28	224
32 - 40	9	36	324
	$\sum f_i = 40$		$\sum f_i x_i = 856$

$$\therefore \text{Mean, } \bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i} = \frac{856}{40} = 21.4$$



Numerical

3

Ability

The following distribution shows the daily pocket allowance of children of a locality. The mean pocket allowance is Rs. 18. Find the missing frequency f .

Daily pocket allowance (in Rs.)	Number of children
11-13	7
13-15	6
15-17	9
17-19	13
19-21	f
21-23	5
23-25	4

Solution

We may prepare the table as given below:

Daily Pocket	Number of	Class mark (x_i)	$f_i x_i$
--------------	-----------	----------------------	-----------

allowance	Children (f_i)		
11 - 13	7	12	84
13 - 15	6	14	84
15 - 17	9	16	144
17 - 19	13	18	234
19 - 21	f	20	20f
21 - 23	5	22	110
23 - 25	4	24	96
	$\sum f_i = 44 + f$		$\sum f_i x_i = 752 + 20f$

$$\therefore \text{Mean } \bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i} = \frac{752 + 20f}{44 + f}$$

Given, mean = 18

$$\therefore 18 = \frac{752 + 20f}{44 + f}$$

$$\Rightarrow 792 + 18f = 752 + 20f$$

$$\Rightarrow 2f = 40$$

$$\Rightarrow f = 20$$



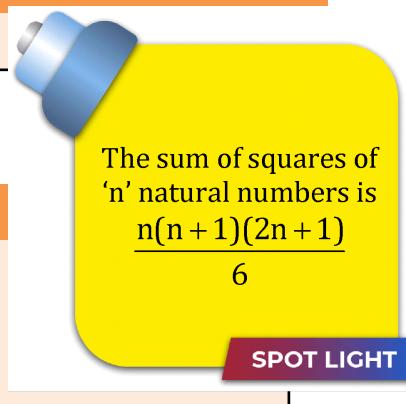
Find the missing frequencies f_1 and f_2 in the table given below, it is being given that the mean of the given frequency distribution is 50.

Class	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100	Total
Frequency	17	f_1	32	f_2		

Solution

We may prepare the table as given below:

Class	Frequency (f_i)	Class mark (x_i)
0 - 20	17	10
20 - 40	f_1	30
40 - 60	32	50



60 - 80	f_2	70	$70f_2$
80 - 100	19	90	1710
$\Sigma f_i = 68 + f_1 + f_2$		$\Sigma f_i x_i = 3480 + 30f_1 + 70f_2$	

$$\therefore \text{Mean, } \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{3480 + 30f_1 + 70f_2}{68 + f_1 + f_2}$$

Given, mean = 50

$$\therefore 50 = \frac{3480 + 30f_1 + 70f_2}{68 + f_1 + f_2}$$

$$\Rightarrow 3400 + 50f_1 + 50f_2 = 3480 + 30f_1 + 70f_2$$

$$\Rightarrow 20f_1 - 20f_2 = 80$$

$$\Rightarrow f_1 - f_2 = 4 \quad \dots(i)$$

And $\Sigma f_i = 68 + f_1 + f_2$

$$\therefore 120 = 68 + f_1 + f_2 \quad [\because \Sigma f_i = 120]$$

$$\Rightarrow f_1 + f_2 = 52 \quad \dots(ii)$$

Adding (i) and (ii), we get $2f_1 = 56$

$$\Rightarrow f_1 = 28$$

$$\Rightarrow f_2 = 24$$

Hence, the missing frequencies f_1 and f_2 are 28 and 24 respectively.



Find the mean of the following distribution.

Class	50 - 70	70 - 90	90 - 110	110 - 130	130 - 150	150 - 170
Frequency	18	12	13	27	8	22

Solution

Class	Frequency (f_i)	Class mark (x_i)	$f_i x_i$
50 - 70	18	60	1080
70 - 90	12	80	960
90 - 110	13	100	1300
110 - 130	27	120	3240
130 - 150	8	140	1120
150 - 170	22	160	3520
	$N = 100$		$\Sigma f_i x_i = 11220$

∴ Mean,

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{11220}{100} = 112.2$$



Find the mean marks from the following data:

Marks	No. of Students
Below 10	5
Below 20	9
Below 30	18
Below 40	29
Below 50	45
Below 60	60
Below 70	70
Below 80	78
Below 90	83
Below 100	85

Solution

We may prepare the table as given below:

Marks	No. of Students	Class Interval	f_i	Class mark (x_i)	$f_i x_i$
Below 10	5	0 - 10	5	5	25
Below 20	9	10 - 20	4	15	60
Below 30	18	20 - 30	9	25	225

Below 40	29	30 – 40	11	35	385
Below 50	45	40 – 50	16	45	720
Below 60	60	50 – 60	15	55	825
Below 70	70	60 – 70	10	65	650
Below 80	78	70 – 80	8	75	600
Below 90	83	80 – 90	5	85	425
Below 100	85	90 – 100	2	95	190
			$\sum f_i =$ 85		$\sum f_i x_i = 4105$

∴ Mean,

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{4105}{85} = 48.29$$



Quick Tips

If less than type or more than type frequency distribution table is given then convert table is usual form to find the mean.



Numerical Ability

7

Find the mean marks of students from the adjoining frequency distribution table.

Marks	No. of Students
Above 0	80
Above 10	77
Above 20	72
Above 30	65
Above 40	55
Above 50	43
Above 60	23
Above 70	16
Above 80	10
Above 90	8
Above 100	0

Solution

We may prepare the table as given below:

Marks	No. of	Class Interval	f_i	Class	$f_i x_i$

Students			mark (x_i)		
Above 0	80	0 – 10	3	5	15
Above10	77	10 – 20	5	15	75
Above20	72	20 – 30	7	25	175
Above30	65	30 – 40	10	35	350
Above40	55	40 – 50	12	45	540
Above50	43	50 – 60	20	55	1100
Above60	23	60 – 70	7	65	455
Above 70	16	70 – 80	6	75	450
Above 80	10	80 – 90	2	85	170
Above 90	8	90 – 100	8	95	760
Above 100	0	100 – 110	0	105	0
			$\sum f_i =$ 80		$\sum f_i x_i = 4090$

∴ Mean,

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{4090}{80} = 51.125 = 51.1 \text{ (approx)}$$

Assumed mean method

In this case, to calculate the mean, we follow the following steps:

$$\frac{\text{Lower limit} + \text{Upper limit}}{2}$$

Step - 1: Find the class mark x_i of each class using $x_i =$

Step - 2 : Choose a suitable value of x_i in the middle as the assumed mean and denote it by 'a'.

Step - 3 : Find $d_i = x_i - a$ for each i

Step - 4 : Find $f_i \times d_i$ for each i

Step - 5 : Find $N = \sum f_i$

Step - 6 : Calculate the mean, (\bar{x}) by using the formula $\bar{x} = a + \frac{\sum f_i d_i}{N}$



The following table gives the marks scored by 100 students in a class test:

Mark	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
No. of Students	12	18	27	20	17	60

Find the mean marks scored by a student in class test.

Solution

We may prepare the table with assumed mean, $a = 35$ as given below:

Marks	No. of Students	Class mark (x_i)	$d_i = x_i - a = x_i - 35$	$f_i d_i$
0 - 10	12	5	-30	-360
10 - 20	18	15	-20	-360
20 - 30	27	25	-10	-270
30 - 40	20	35 = a	0	0
40 - 50	17	45	10	170
50 - 60	6	55	20	120
	N = 100			$\sum f_i d_i = -700$

$$\therefore \text{Mean, } \bar{x} = a + \frac{\sum f_i d_i}{N}$$

$$= 35 + \frac{(-700)}{100}$$

$$= 35 - 7 = 28$$



Thirty women were examined in a hospital by a doctor and the number of heart beats per minute were recorded and summarised as follows. Find the mean heart beats per minute for these women, by using assumed-mean method.

No. of heart

beats per minute 65 – 68 68 – 71 71 – 74 74 – 77 77 – 80 80 – 83 83 – 86

Frequency	2	4	3	8	7	4	2
------------------	---	---	---	---	---	---	---

Solution

We may prepare the table with assumed mean, $a = 75.5$ as given below:

No. of heart beats per minute	No. of women (f_i)	Class mark (x_i)	$d_i = x_i - a = x_i - 75.5$	$f_i d_i$
65 – 68	2	66.5	-9	-18
68 – 71	4	69.5	-6	-24
71 – 74	3	72.5	-3	-9
74 – 77	8	75.5 = a	0	0
77 – 80	7	78.5	3	21
80 – 83	4	81.5	6	24
83 – 86	2	84.5	9	18
	N = 30			$\sum f_i d_i = 12$

$$\therefore \text{Mean, } \bar{x} = a + \frac{\sum f_i d_i}{N} = 75.5 + \frac{12}{30} = 75.5 + \frac{2}{5} = 75.9$$



Find the arithmetic mean of the following frequency distribution.

Class	25 – 29	30 – 34	35 – 39	40 – 44	45 – 49	50 – 54	55 – 59
Frequency	14	22	16	6	5	3	4

Solution

The given series is in inclusive form. We will prepare the table in exclusive form with assumed mean $a = 42$ as given below:

Class	Frequency (f_i)	Class mark (x_i)	$d_i = x_i - a = x_i - 42$	$f_i d_i$
24.5 – 29.5	14	27	-15	-210
29.5 – 34.5	22	32	-10	-220
34.5 – 39.5	16	37	-5	-80
39.5 – 44.5	6	42 = a	0	0
44.5 – 49.5	5	47	5	25
49.5 – 54.5	3	52	10	30
54.5 – 59.5	4	57	15	60

	$N = 70$			$\sum f_i d_i = -395$
--	----------	--	--	-----------------------

$$\therefore \text{Mean, } \bar{x} = a + \frac{\sum f_i d_i}{N} = 42 + \frac{(-395)}{70}$$

$$= \frac{2940 - 395}{70}$$

$$= \frac{2545}{70} = 36.36 \text{ (approx.)}$$


**Quick
Tips**

- It is not necessary to convert the data in continuous interval form for finding mean.

Step-deviation method or short-cut method

Sometimes, the values of x and f are so large that the calculation of mean by assumed mean method becomes quite inconvenient. In this case, we follow the following steps:

Step-1 : Find the class mark x_i of each class by using $x_i = \frac{\text{lower limit} + \text{Upper limit}}{2}$

Step-2 : Choose a suitable values of x_i in the middle as the assumed mean and denote it by 'a'.

Step-3 : Find $h = (\text{upper limit} - \text{lower limit})$ for each class.

Step-4 : Find $u_i = \frac{x_i - a}{h}$ for each class.

Step-5 : Find $f_i u_i$ for each i .

Step-6 : Calculate, the mean by using the formula $= a + \left\{ \frac{\sum f_i \times u_i}{N} \right\} \times h$, where $N = \sum f_i$.


**Numerical
Ability**
11

Find the mean of the following distribution by step-deviation method:

Class	50 - 70	70 - 90	90 - 110	110 - 130	130 - 150	150 - 170
Frequency	18	12	13	27	8	22

Solution

We may prepare the table with assumed mean $a = 120$ and $h = 20$ as given below:

Class	Frequency (f_i)	Class mark (x_i)	$u_i = \frac{x_i - a}{h} = \frac{x_i - 120}{20}$	$f_i u_i$
50 - 70	18	60	-3	-54
70 - 90	12	80	-2	-24
90 - 110	13	100	-1	-13
110 - 130	27	120 = a	0	0
130 - 150	8	140	1	8
150 - 170	22	160	2	44
	$N = 100$			$\sum f_i u_i = -39$

$$\therefore \text{Mean, } \bar{x} = a + \frac{\sum f_i u_i h}{N}$$

$$= 120 + \frac{(-39) \times 20}{100} = 120 - \frac{39}{5} = \frac{561}{5} = 112.2$$



Quick Tips

- If class size is different then h is taken as HCF of the class marks.



Calculate the mean for the following frequency distribution (By step deviation method).

Class interval	0 - 80	80 - 160	160 - 240	240 - 320	320 - 400
Frequency	22	35	44	25	24

Solution

Class Interval	Mid - value (x_i)	f_i	$u_i = (x_i - a)/h$	$f_i u_i$
0 - 80	40	22	-2	-44
80 - 160	120	35	-1	-35
160 - 240	200(a)	44	0	0
240 - 320	280	25	1	25

320 - 400	360	24	2	48
Total		$\sum f_i = 150$		$\sum f_i u_i = -6$

Mean $(\bar{x}) = a + \frac{\sum f_i u_i}{\sum f_i} \times h = 200 + \left(\frac{-6}{150} \right) \times 80$

$$= 200 - \left(\frac{2 \times 8}{5} \right) = 200 - \frac{16}{5} = \frac{1000 - 16}{5} = \frac{984}{5} = 196.8$$

Median

Median: It is a measure of central tendency which gives the value of the middle most observation in the data. In a grouped data, it is not possible to find the middle observation by looking at the cumulative frequencies as the middle observation will be some value in a class interval. It is, therefore, necessary to find the value inside a class that divides the whole distribution into two halves.

$\frac{N}{2}$

Median Class: The class whose cumulative frequency is greater than $\frac{N}{2}$ is called the median class.

To calculate the median of a grouped data, we follow the following steps.

Step - 1: Prepare the cumulative frequency table corresponding to the given frequency distribution and obtain $N = \sum f_i$

Step - 2: Find $\frac{N}{2}$

Step - 3: Look at the cumulative frequency just greater than $\frac{N}{2}$ and find the corresponding class (Median class).

$$\ell + \left\{ \frac{\frac{N}{2} - cf}{f} \right\} \times h$$

Step - 4: Use the formula Median =

Where, ℓ = Lower limit of median class.

f = Frequency of the median class.

cf = Cumulative frequency of the class preceding the median class.

h = Size of the median class.

$$N = \sum f_i$$



- Data must be in continuous interval form to find median and mode of grouped data.



Find the median of the following frequency distribution:

Mark	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	Total
No. of Students	8	20	36	24	12	100

Solution

At first, we prepare a cumulative frequency distribution table as given below:

Marks	No. of students (f_i)	Cumulative frequency
0 - 10	8	8
10 - 20	20	28
20 - 30	36	64
30 - 40	24	88
40 - 50	12	100
	N = 100	

Here, $N = 100$

$$\therefore \frac{N}{2} = 50$$

The cumulative frequency just greater than 50 is 64 and the corresponding class is 20 – 30.

So, the median class is 20 – 30.

$$\therefore \ell = 20, N = 100,$$

$$cf = 28,$$

$$f = 36 \text{ and } h = 10$$

$$\ell + \left\{ \frac{\frac{N}{2} - cf}{f} \right\} \times h$$

Therefore, median =

$$= 20 + \left(\frac{50 - 28}{36} \right) \times 10$$

$$= 20 + \frac{22 \times 10}{36}$$

$$= 20 + \frac{55}{9}$$

$$= \frac{180 + 55}{9}$$

$$= \frac{235}{9}$$

$$= 26.1$$



The median is the middle of a distribution: half the scores are above the median and half are below the median.

SPOT LIGHT



Numerical Ability

14

A health insurance agent found the following data for distribution of ages of 100 policy holders. Calculate the median age, if policies are given only to persons having age 18 years onwards but less than 60 years.

Age(in years)	No. of policy holders
Below 20	4
Below 25	8
Below 30	24
Below 35	45
Below 40	78
Below 45	89
Below 50	90
Below 55	96
Below 60	100

Solution

From the given table we can find the frequency and cumulative frequencies as given below :

Age in(years)	No. of holders (f_i)	Cumulative frequency
15 - 20	4	4
20 - 25	4	8
25 - 30	16	24
30 - 35	21	45
35 - 40	33	78
40 - 45	11	89
45 - 50	1	90
50 - 55	6	96
55 - 60	4	100
	N = 100	

Here, $N = 100$

$$\therefore \frac{N}{2} = 50$$

The cumulative frequency just greater than 50 is 78 and the corresponding class is 35 – 40.

So, the median class is 35 – 40.

$$\therefore \ell = 35,$$

$$N = 100,$$

$$cf = 45,$$

$$f = 33 \text{ and } h = 5$$

$$\ell + \left\{ \frac{\frac{N - cf}{2}}{f} \right\} \times h$$

Therefore, median =

$$= 35 + \left(\frac{50 - 45}{33} \right) \times 5$$

$$= 35 + \frac{5 \times 5}{33}$$

$$= \frac{1155 + 25}{33} = \frac{1180}{33} = 35.76$$

Hence, the median age is 35.76 years.



Numerical Ability

15

From the following frequency distribution, calculate the median.

Marks	No. of students
5 – 10	7
10 – 15	15
15 – 20	24
20 – 25	31
25 – 30	42
30 – 35	30
35 – 40	26
40 – 45	15
45 – 50	10

Solution

$$\text{Here } \frac{N}{2} = \frac{200}{2} = 100$$

The median lies in the class 25 – 30.

$$\ell = 25, c.f. = 77, f = 42 \text{ and } h = 5$$

Calculation of Median

Marks	f	c.f.
5 – 10	7	7
10 – 15	15	22
15 – 20	24	46
20 – 25	31	77
25 – 30	42	119
30 – 35	30	149
35 – 40	26	175
40 – 45	15	190
45 – 50	10	200

$$\ell + \frac{\frac{N}{2} - c.f.}{f} \times h$$

Applying the formula, Median =

$$25 + \frac{100 - 77}{42} \times 5$$

We get, Median =

$$= 25 + \frac{115}{42} = 25 + 2.74 = 27.74$$

So, above half the student have scored marks less than 27.74 and the other half scored marks more than 27.74.



Calculate the missing frequency 'a' from the following distribution, it is being given that the median of the distribution is 24.

Age (in years)	0-10	10-20	20-30	30-40	40-50
No. of persons	5	25	a	18	7

Solution

At first we prepare a cumulative frequency distribution table as given below :

Age (in years)	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	Total
No. of persons (f_i)	5	25	a	18	7	$55 + a$
Cumulative frequency	5	30	$30+a$	$48+a$	$55+a$	

Since the median is 24, therefore, the median class will be 20 - 30.

Hence, $\ell = 20$, $N = 55 + a$, $cf = 30$, $f = a$ and $h = 10$

$$\ell + \left\{ \frac{\frac{N}{2} - cf}{f} \right\} \times h$$

Therefore, median =

$$\Rightarrow 24 = 20 + \left(\frac{\frac{55+a}{2} - 30}{a} \right) \times 10$$

$$\Rightarrow 24 = 20 + \frac{(a-5)}{2a} \times 10$$

$$\Rightarrow 4 = \frac{(a-5)}{a} \times 5$$

$$\Rightarrow 4a = 5a - 25$$

$$\Rightarrow a = 25$$

Hence, the value of missing frequency a is 25.



Numerical Ability

17

The median of the following data is 525. Find the values of x and y, if the total frequency is 100.

Class Interval	Frequency(f_i)
0 - 100	2
100 - 200	5
200 - 300	x
300 - 400	12
400 - 500	17
500 - 600	20
600 - 700	y
700 - 800	9
800 - 900	7
900 - 1000	4
	N = 100

Solution

At first we prepare a cumulative frequency distribution table as given below :

Class interval	Frequency (f_i)	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
	N = 100	

We have $N = 100$

$$\therefore 76 + x + y = 100$$

$$\Rightarrow x + y = 24 \quad \dots(i)$$

Since the median is 525, so, the median class is 500–600

$$\therefore \ell = 500, N = 100, cf = 36 + x, f = 20 \text{ and } h = 100$$

$$\text{Therefore, median} = \ell + \left\{ \frac{\frac{N}{2} - cf}{f} \right\} \times h$$

$$\Rightarrow 525 = 500 + \left(\frac{50 - 36 - x}{20} \right) \times 100$$

$$\Rightarrow 25 = (14 - x) \times 5$$

$$\Rightarrow 5 = 14 - x$$

$$\Rightarrow x = 9$$

Also, putting $x = 9$ in (i),

we get $9 + y = 24$

$$\Rightarrow y = 15$$

Hence, the values of x and y are 9 and 15 respectively.

Mode

Mode : Mode is that value among the observations which occurs most often i.e. the value of the observation having the maximum frequency.

In a grouped frequency distribution, it is not possible to determine the mode by looking at the frequencies.

Modal Class : The class of a frequency distribution having maximum frequency is called modal class of a frequency distribution.

The mode is a value inside the modal class and is calculated by using the formula.

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

Where ℓ = Lower limit of the modal class.

h = Size of class interval

f_1 = Frequency of modal class

f_0 = Frequency of the class preceding the modal class

f_2 = Frequency of the class succeeding the modal class.



Be Alert !

- A disadvantage of the mode is that many distributions have more than one mode.

These distribution are called “multi modal” and is therefore not recommended to be used as the only measure of central tendency.



Numerical Ability

18

The following data gives the information on the observed lifetimes (in hours) of 225 electrical components:

Lifetimes (in hours)	0-20	20-40	40-60	60-80	80-100	100-120
Frequency	10	35	52	61	38	29

Determine the modal lifetime of the components.

Solution

Here the class 60 – 80 has maximum frequency, so it is the modal class.

$$\therefore \ell = 60, h = 20, f_1 = 61, f_0 = 52 \text{ and } f_2 = 38$$

$$\text{Therefore, mode} = \ell + \left\{ \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right\} \times h$$

$$= 60 + \left(\frac{61 - 52}{2 \times 61 - 52 - 38} \right) \times 20$$

$$= 60 + \frac{9}{32} \times 20$$

$$= 60 + 5.625$$

$$= 65.625$$

Hence, the modal lifetime of the components is 65.625 hours.



Given below is the frequency distribution of the heights of players in a school.

Heights (in cm)	160 – 162	163 – 165	166 – 168	169 – 171	172 – 174
No. of students	15	118	142	127	18

Find the modal class.

Solution

The given series is in inclusive form. We prepare the table in exclusive form, as given below :

Heights (in cm)	159.5 – 162.5	162.5 – 165.5	165.5 – 168.5	168.5 – 171.5	171.5 – 174.5
No. of students	15	118	142	127	18

Here, the class 165.5-168.5 has maximum frequency, so it is the modal class.



The mode of the following series is 36. Find the missing frequency f in it.

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
Frequency	8	10	f	16	12	6	7

Solution

Since the mode is 36, so the modal class will be 30 – 40

$$\therefore \ell = 30, h = 10, f_1 = 16, f_0 = f \text{ and } f_2 = 12$$

$$\text{Therefore, mode} = \ell + \left\{ \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right\} \times h$$

$$\Rightarrow 36 = 30 + \left(\frac{16 - f}{2 \times 16 - f - 12} \right) \times 10$$

$$\Rightarrow 6 = \frac{(16-f)}{(20-f)} \times 10$$

$$\Rightarrow 120 - 6f = 160 - 10f$$

$$\Rightarrow 4f = 40$$

$$\Rightarrow f = 10$$

Hence, the value of the missing frequency f is 10.



Do You Remember ?

- The empirical relationship between mean, mode and median is

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$$



The mean and mode of a frequency distribution are 28 and 19 respectively. Then find the median.

Solution

$$\text{Mean} = 28 \text{ and Mode} = 19$$

$$\text{Median} = ?$$

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$$

$$19 = 3 \text{ Median} - 2(28)$$

$$3 \text{ Median} = 19 + 56$$

$$\text{Median} = \frac{75}{3}$$

$$\text{Median} = 25$$

Graphical representation of cumulative frequency distribution

Cumulative frequency polygon curve (Ogive)

Cumulative frequency is of two types and corresponding to these, the ogive is also of two types.

- **Less than ogive**
- **More than ogive**

Less than ogive : To construct a cumulative frequency polygon and an ogive, we follow

these steps :

Step-1 : Mark the upper class limit along x-axis and the corresponding cumulative frequencies along y-axis.

Step-2 : Plot these points successively by line segments. We get a polygon, called cumulative frequency polygon.

Step-3 : Plot these points successively by smooth curves, we get a curve called cumulative frequency curve or an ogive.

More than ogive To construct a cumulative frequency polygon and an ogive, we follow these steps:

Step-1 : Mark the lower class limits along x-axis and the corresponding cumulative frequencies along y-axis.

Step-2 : Plot these points successively by line segments, we get a polygon, called cumulative frequency polygon.

Step-3 : Plot these points successively by smooth curves, we get a curve, called cumulative frequency curve or an ogive.

Application of an ogive

Ogive can be used to find the median of a frequency distribution. To find the median, we follow these steps.

Method-I

Step-1 : Draw anyone of the two types of frequency curves on the graph paper.

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

Step-3 : Draw a line parallel to x-axis from the point marked in step 2, cutting the cumulative frequency curve at a point P.

Step-4 : Draw perpendicular PM from P on the x-axis. The x-coordinate of point M gives the median.

Method-II

Step-1 : Draw less than type and more than type cumulative frequency curves on the graph paper.

Step-2 : Mark the point of intersecting (P) of the two curves drawn in step 1.

Step-3 : Draw perpendicular PM from P on the x-axis. The x- coordinate of point M gives the median.


Numerical Ability
22

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
No. of workers	12	14	8	6	10

Convert the distribution above to a less than type cumulative frequency distribution and draw its ogive.

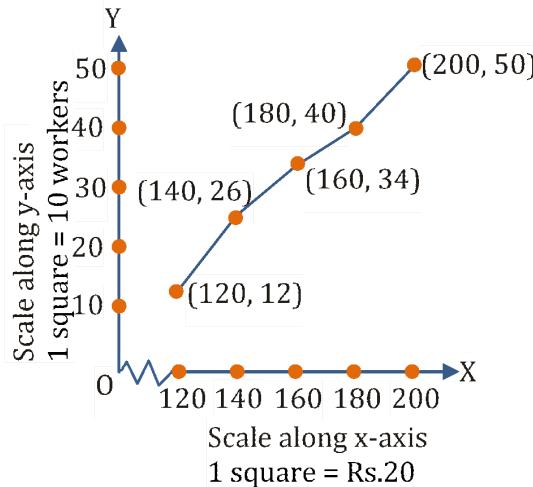
Solution

From the given table, we prepare a less than type cumulative frequency distribution table, as given below:

Income less than (in Rs.)	120	140	160	180	200
Cumulative frequency	12	26	34	40	50

Now, plot the points (120, 12), (140, 26), (160, 34), (180, 40) and (200, 50).

Join these points by a freehand curve to get an ogive of 'less than' type.


Numerical Ability
23

The following table gives the weight of 120 articles :

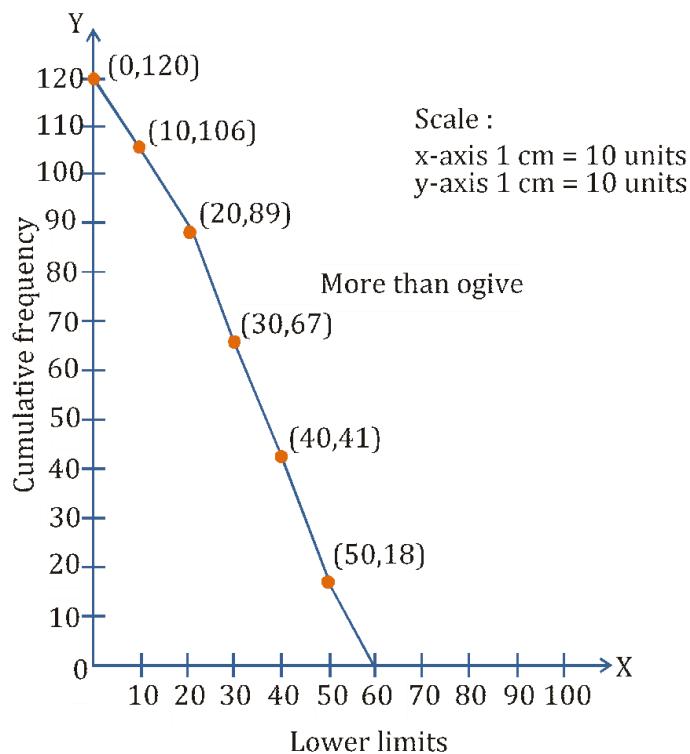
Weight (in kg)	0-10	10-20	20-30	30-40	40-50	50-60
Number of Students	14	17	22	26	23	18

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

Solution

Weight (in Kg)	Cumulative Frequency
More than or equal to 0	120
More than or equal to 10	106
More than or equal to 20	89
More than or equal to 30	67
More than or equal to 40	41
More than or equal to 50	18

Plotting the points :

**Numerical****24**

The annual profits earned by 30 shops of a shopping complex in a locality gives rise to the following distribution:

Profit (in lakhs Rs.)	No. 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

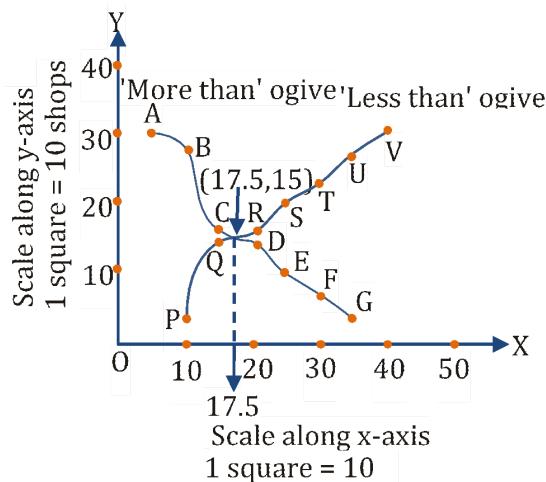
We have a more than type cumulative frequency distribution table. We may also prepare a less than type cumulative frequency distribution table from the given data, as given below:

More than type	
Profit more than (Rs. in lakhs)	No. of shops
5	30
10	28
15	16
20	14
25	10
30	7
35	3

Less than type	
Profit less than (Rs. in lakhs)	No. of shops
10	2
15	14
20	16
25	20
30	23
35	27
40	30

Now, plot the points A (5,30), B (10,28), C (15,16), D (20,14), E (25,10), F (30,7) and G (35,3) for the more than type cumulative frequency and the points P (10,2), Q (15,14), R (20,16), S (25,20), T (30,23), U (35,27) and V (40,30) for the less than type cumulative frequency distribution table.

Join these points by a freehand to get ogives for 'more than' type and 'less than' type.



The two ogives intersect each other at point (17.5, 15).

Hence, the median profit is Rs. 17.5 lakhs.



The following data gives the information on marks of 70 students in a periodical test:

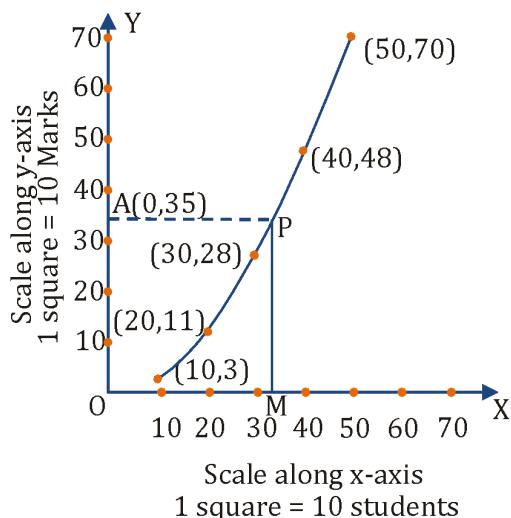
Draw a cumulative frequency curve for the given data and find the median.

Marks	Less than 10	Less than 20	Less than 30	Less than 40	Less than 50
No. of students	3	11	28	48	70

Solution

We have a less than cumulative frequency table. We mark the upper class limits along the x-axis and the corresponding cumulative frequencies (no. of students) along the y-axis.

Now, plot the points (10, 3), (20, 11), (30, 28), (40, 48) and (50, 70). Join these points by a freehand curve to get an ogive of 'less than' type.



A distribution is said to be symmetric if
Mean = Median = Mode

SPOT LIGHT

Here, $N = 70$

$$\therefore \frac{N}{2} = 35$$

Take a point A(0, 35) on the y-axis and draw AP \parallel x-axis, meeting the curve at P.

Draw PM \perp x -axis, intersecting the x-axis, at M.

Then, OM = 33.

Hence, the median marks is 33.

Memory map

Class Interval

Each group into which the raw data is condensed, is called a class-interval.

Class Limits

Each class is bounded by two figures, which are called class limits. The figure on the left side of a class is called its lower limit and that on its right is called its upper limit.

Class size

The difference between the true upper limit and the true lower limit of a class is called its class size.

Class mark of a class



$$\text{Class mark} = \left(\frac{\text{True upper limit} + \text{True lower limit}}{2} \right)$$

Three measures of central tendency are useful for analysing the data, namely,

- (a) Mean (b) Median (c) Mode

Arithmetic mean



$$\text{Mean } (\bar{x}) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \text{ or } \bar{x} = \frac{\sum_{i=1}^n x_i}{n},$$

where $\sum_{i=1}^n x_i$ denotes the sum $x_1 + x_2 + x_3 + \dots + x_n$.

The arithmetic mean of grouped data may also be calculated by any one of the following methods.

1 Direct method

$x_1, x_2, x_3, \dots, x_n$ are observations with respective frequencies

$f_1, f_2, f_3, \dots, f_n$ then mean, (\bar{x}) is defined by

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \dots + f_n x_n}{f_1 + f_2 + f_3 + \dots + f_n} \text{ or } \bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i},$$

Where $\sum_{i=1}^n f_i = f_1 + f_2 + f_3 + \dots + f_n = N$ and $x_i = \frac{\text{Lowerlimit} + \text{Upperlimit}}{2}$

2 Assumed Mean method

In this case, to calculate the mean, we follow the following steps:

Calculate the mean, (\bar{x}) by using the formula.

$$\bar{x} = a + \frac{\sum f_i d_i}{N}$$

where, $x_i = \frac{\text{Lower limit} + \text{Upper limit}}{2}$,

a = assumed mean, $d_i = x_i - a$ and $N = \sum f_i$

3 Step-deviation method or short-cut method

$$\text{Mean } (\bar{x}_i) = a + \left\{ \frac{\sum f_i \times u_i}{N} \right\} \times h, \text{ where } N = \sum f_i$$

$$x_i = \frac{\text{lower limit} + \text{Upper limit}}{2}$$

a = assumed mean, h = (upper limit - lower limit)

$$\text{and } u_i = \frac{x_i - a}{h}$$

Median

It is a measure of central tendency which gives the value of the middle most observation in the data.

Median Class

The class whose cumulative frequency is greater than $\frac{N}{2}$ is called the median class.

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

Where,

l = Lower limit of median class.

f = Frequency of the median class.

cf = Cumulative frequency of the class preceding the median class.

h = Size of the median class.

$$N = \sum f$$



Mode

Mode is that value among the observations which occurs most often i.e., the value of the observation having the maximum frequency.

Modal Class

The class of a frequency distribution having maximum frequency is called modal class of a frequency distribution.

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

Where,

l = Lower limit of the modal class.

h = Size of class interval

f_1 = Frequency of modal class

f_0 = Frequency of the class preceding the modal class

f_2 = Frequency of the class succeeding the modal class.



Graphical representation of cumulative frequency distribution



Cumulative frequency polygon curve (Ogive)

Cumulative frequency is of two types and corresponding to these, the ogive is also of two types.

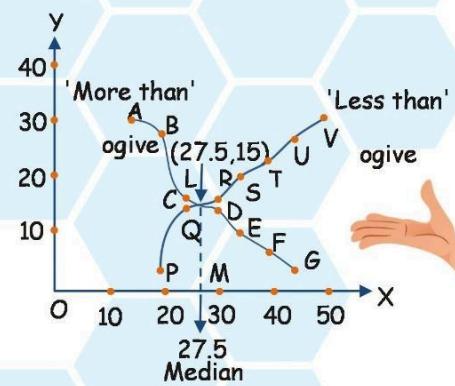
- ✓ Less than ogive
- ✓ More than ogive

Application of an ogive

Step-1 : Draw less than type and more than type cumulative frequency curves on the graph paper.

Step-2 : Mark the point of intersecting (L) of the two curves drawn in step 1.

Step-3 : Draw perpendicular PM from P on the x-axis. The x- coordinate of point M gives the median.



Statistics

