8

$$l+\frac{\left(\frac{N}{2}-m\right)}{f}\times c \qquad \bar{X}=\frac{\sum fx}{\sum f}$$

STATISTICS

"Lack of statistics is to hide inconvenient facts."

- Albert Bertilsson

Sir Ronald Aylmer Fisher (AD (CE) 1890 - 1962)

Sir Ronald Aylmer Fisher was a British Statistician and Biologist. Also he was known as the Father of Modern Statistics and Experimental Design. Fisher did experimental agricultural research, which saved millions from starvation. He was awarded the Linnean Society of London's prestigious Darwin-Wallace Medal in 1958.

BEC9CY

Learning Outcomes

- To recall different types of averages known already.
- To recall the methods of computing the Mean, Median and Mode for ungrouped data.
- To compute the Mean, Median and Mode for the grouped data.

8.1 Introduction

Statistics is the science of collecting, organising, analysing and interpreting data in order to make decisions. In everyday life, we come across a wide range of quantitative and qualitative information. These have profound impact on our lives.

Data means the facts, mostly numerical, that are gathered; statistics implies collection of data. We analyse the data to make decisions. The methods of statistics are tools to help us in this.

Cricket News

Team U19	Matches	Won	Lost	NR/Tied
India	71	52	18	0/1
Australia	67	50	15	0/2
Pakistan	69	50	19	0/0
Bangladesh	64	45	17	1/1
West Indies	71	44	27	0/0
South Africa	61	43	17	0/1
England	69	40	28	0/1
Sri Lanka	68	36	31	0/1
New Zealand	66	30	35	0/1
Zimbabwe	62	28	34	0/0
Ireland	49	16	32	1/0
Afghanistan	24	11	13	0/0
Namibia	47	9	37	1/0
Kenya	17	5	12	0/0
Canada	29	4	23	1/1
PNG	41	3	38	0/0

Customer Satisfaction Survey

Hotel Tamilnadu

Tell us how you were satisfied with our service

Very Satisfied

Satisfied

Neutral

Unsatisfied

Annoyed

India's 2018 GDP Forecast

UN	7.2%
IMF	7.4%
World Bank	7.3%
Morgan Stanley	7.5%
Moody's	7.6%
HSBC	7%
Bank of America	7.2%
Merill Lynch	7.5%
Goldman Sachs	8%

8.2 Collection of Data

Primary data are first-hand original data that we collect ourselves. Primary data collection can be done in a variety of ways such as by conducting personal interviews (by phone, mail or face-to-face), by conducting experiments, etc.

Secondary data are the data taken from figures collected by someone else. For example, government-published statistics, available research reports etc.

Progress Check

Identify the primary data

(i) Customer surveys
(ii) Medical researches
(iii) Economic predictions
(iv) School results
(v) Political polls
(vi) Marketing details
(vii) Sales forecasts
(viii) Price index details

8.2.1 Getting the Facts Sorted Out

When data are initially collected and before it is edited and not processed for use, they are known as Raw data. It will not be of much use because it would be too much for the human eye to analyse.

For example, study the marks obtained by 50 students in mathematics in an examination, given below:

Activity - 1

Prepare an album of pictures, tables, numeric details etc that exhibit data. Discuss how they are related to daily life situations.

61	60	44	49	31	60	79	62	39	51	67	65	43	54	51	42
52	43	46	40	60	63	72	46	34	55	76	55	30	67	44	57
62	50	65	58	25	35	54	59	43	46	58	58	56	59	59	45
42	44														

In this data, if you want to locate the five highest marks, is it going to be easy? You have to search for them; in case you want the third rank among them, it is further complicated. If you need how many scored less than, say 56, the task will be quite time consuming.

Hence **arrangement of an array of marks** will make the job simpler.

With some difficulty you may note in the list that 79 is the highest mark and 25 is the least. Using these you can subdivide the data into convenient classes and place each mark into the appropriate class. Observe how one can do it.

Class Interval	Marks
25-30	30, 25
31-35	31, 34, 35
36-40	39, 40
41-45	44, 43, 42, 43, 44, 43, 45, 42, 44
46-50	49, 46, 46, 50, 46
51-55	51, 54, 51, 52, 55, 55, 54
56-60	60, 60, 60, 57, 58, 59, 58, 58, 56, 59, 59
61-65	61, 62, 65, 63, 62, 65
66-70	67, 67
71-75	72
76-80	79, 76

From this table can you answer the questions raised above? To answer the question, "how many scored below 56", you do not need the actual marks. You just want "how many" were there. To answer such cases, which often occur in a study, we can modify the table slightly and just note down how many items are there in each class. We then may have a slightly simpler and more useful arrangement, as given in the table.

Class Interval	Number of items
25-30	2
31-35	3
36-40	2
41-45	9
46-50	5
51-55	7
56-60	11
61-65	6
66-70	2
71-75	1
76-80	2

This table gives us the number of items in each class; each such number tells you how many times the required item occurs in the class and is called the frequency in that class.

The table itself is called a frequency table.

We use what are known as tally marks to compute the frequencies. (Under the column 'number of items', we do not write the actual marks but just tally marks). For example, against the class 31-35, instead of writing the actual marks 31, 34, 35 we simply put |||. You may wonder if for the class 56-60 in the example one has to write |||||||||||, making it difficult to count. To avoid confusion, every fifth tally mark is put across the four preceding it, like this 𝍸 . For example, 11 can be written as 𝍸 𝍸 | . The frequency table for the above illustration will be seen as follows:

Class Interval	Tally Marks	Frequency
25-30	||	2
31-35	|||	3
36-40	||	2
41-45	~~||||~~ ||||	9
46-50	~~||||~~	5
51-55	~~||||~~ ||	7
56-60	~~||||~~ ~~||||~~ |	11
61-65	~~||||~~ |	6
66-70	||	2
71-75	|	1
76-80	||	2
	TOTAL	50

Note

Consider any class, say 56- 60; then 56 is called the lower limit and 60 is called the upper limit of the class.

Progress Check

Form a frequency table for the following data:

23	44	12	11	45	55	79	20
52	37	77	97	82	56	28	71
62	58	69	24	12	99	55	78
21	39	80	65	54	44	59	65
17	28	65	35	55	68	84	97
80	46	30	49	50	61	59	33
11	57						

8.3 Measures of Central Tendency

It often becomes necessary in everyday life to express a quantity that is typical for a given data. Suppose a researcher says that on an average, people watch TV serials for 3 hours per day, it does not mean that everybody does so; some may watch more and some less. The average is an acceptable indicator of the data regarding programmes watched on TV.

Averages summarise a large amount of data into a single value and indicate that there is some variability around this single value within the original data.

A mathematician's view of an average is slightly different from that of the commoner. There are three different definitions of average known as the Mean, Median and Mode. Each of them is found using different methods and when they are applied to the same set of original data they often result in different average values. It is important to figure out what each of these measures of average tells you about the original data and consider which one is the most appropriate to calculate.

8.4 Arithmetic Mean

8.4.1 Arithmetic Mean-Raw Data

The Arithmetic Mean of a data is the most commonly used of all averages and is found by adding together all the values and dividing by the number of items.

For example, a cricketer, played eight (T20) matches and scored the following scores 25, 32, 36, 38, 45, 41, 35, 36.

Then, the mean of his scores (that is the arithmetic average of the scores) is obtained by

$$\overline{X} = \frac{\sum x}{n} = \frac{25 + 32 + 36 + 38 + 45 + 41 + 35 + 36}{8} = \frac{288}{8} = 36 .$$

In general, if we have n number of observations $x_1, x_2, x_3, \ldots, x_n$ then their arithmetic mean denoted by $\overline{X}$ (read as X bar) is given by

$$\overline{X} = \frac{\text{sum of all the observations}}{\text{number of observations}} \qquad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

We express this as a formula: $\overline{X} = \frac{\sum x}{n}$

Assumed Mean method: Sometimes we can make calculations easy by working from an entry that we guess to be the right answer. This guessed number is called the assumed mean.

In the example above on cricket scores, let us assume that 38 is the assumed mean. We now list the differences between the assumed mean and each score entered:

Note

It does not matter which number is chosen as the assumed mean; we need a number that would make our calculations simpler. Perhaps a choice of number that is closer to most of the entries would help; it need not even be in the list given.

25–38 = –13, 32–38 = –6, 36–38 = –2, 38–38 = 0,

45–38 = 7, 41–38 = 3, 35–38 = –3, 36–38 = –2

The average of these differences is $\frac{-13-6-2+0+7+3-3-2}{8} = \frac{-16}{8} = -2$

We add this ‘mean difference’ to the assumed mean to get the correct mean.

Thus the correct mean = Assumed Mean +Mean difference = 38 – 2 = 36 .

This method will be very helpful when large numbers are involved.

8.4.2 Arithmetic Mean-Ungrouped Frequency Distribution

Consider the following list of heights (in cm) of 12 students who are going to take part in an event in the school sports.

140, 142, 150, 150, 140, 148, 140, 147, 145, 140, 147, 145.

How will you find the Mean height?

There are several options.

(i) You can add all the items and divide by the number of items.

$$\frac{140+142+150+150+140+148+140+147+145+140+147+145}{12} = \frac{1734}{12} = 144.5$$

(ii) You can use Assumed mean method. Assume, 141 as the assumed mean.

Then the mean will be given by

$$= 141 + \frac{(-1)+(1)+(9)+(9)+(-1)+(7)+(-1)+(6)+(4)+(-1)+(6)+(4)}{12}$$

$$= 141 + \frac{-4+46}{12} = 141 + \frac{42}{12} = 141 + 3.5 = 144.5$$

(iii) A third method is to deal with an ungrouped frequency distribution. You find that 140 has occurred 4 times, (implying 4 is the frequency of 140), 142 has occurred only once (indicating that 1 is the frequency of 142) and so on. This enables us to get the following frequency distribution.

Height(cm)	140	142	150	148	145	147
No. of students	4	1	2	1	2	2

You find that there are four 140s; their total will be $140 \times 4 = 560$

There is only one 142; so the total in this case is $142 \times 1 = 142$

There are two 150s; their total will be $150 \times 2 = 300$ etc.

These details can be neatly tabulated as follows:

Height (x)	Frequency (f)	fx
140	4	560
142	1	142
150	2	300
148	1	148
145	2	290
147	2	294
	12	1734

$$\text{Mean} = \frac{\text{Sum of all } fx}{\text{No.of items}}$$

$$= \frac{1734}{12} = 144.5 \text{ cm}$$

Looking at the procedure in general terms, you can obtain a formula for ready use. If $x_1, x_2, x_3, \ldots x_n$ are n observations whose corresponding frequencies are $f_1, f_2, f_3, \ldots f_n$ then the mean is given by

Note

Study each step and understand the meaning of each symbol.

$$\bar{X} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_n x_n}{f_1 + f_2 + \ldots f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum fx}{\sum f}$$

Can you adopt the above method combining with the assumed mean method? Here is an attempt in that direction:

(iv) Let the assumed mean be 145. Then we can prepare the following table:

Height(x)	d = deviation from the assumed mean	Frequency (f)	fd
140	140 – 145 = – 5	4	–20
142	142– 145 = – 3	1	– 3
150	150 – 145 = + 5	2	+10
148	148 – 145 = + 3	1	+ 3
145 (Assumed)	145 – 145 = 0	2	0
147	147 – 145 = + 2	2	+ 4
Total		$\sum f = 12$	$\sum fd = -23 + 17 = -6$

Arithmetic mean = Assumed mean + Average of the sum of deviations

$$= A + \frac{\sum fd}{\sum f} = 145 + \left(\frac{-6}{12}\right) = 145.0 - 0.5 = 144.5$$

When large numbers are involved, this method could be useful.

8.4.3 Arithmetic Mean-Grouped Frequency Distribution

When data are grouped in class intervals and presented in the form of a frequency table, we get a frequency distribution like this one:

Age (in years)	10-20	20-30	30 – 40	40 – 50	50 - 60
Number of customers	80	120	50	22	8

The above table shows the number of customers in the various age groups. For example, there are 120 customers in the age group 20 – 30, but does not say anything about the age of any individual. (When we form a grouped frequency table the identity of the individual observations is lost). Hence we need a value that represents the particular class interval. Such a value is called mid value (mid-point or class mark) The mid-point or class mark can be found using the formula given below.

Mid Value = $\frac{UCL + LCL}{2}$, *UCL – Upper Class Limit, LCL – Lower Class Limit*

In grouped frequency distribution, arithmetic mean may be computed by applying any one of the following methods.

(i) Direct Method (ii) Assumed Mean Method (iii) Step Deviation Method

Direct Method

When direct method is used, the formula for finding the arithmetic mean is

$$\overline{X} = \frac{\sum fx}{\sum f}$$

Where x is the mid-point of the class interval and f is the corresponding frequency

Steps

(i) Obtain the mid-point of each class and denote it by x

(ii) Multiply those mid-points by the respective frequency of each class and obtain the sum of fx

(iii) Divide Σfx by Σf to obtain mean

Example 8.1 The following data gives the number of residents in an area based on their age. Find the average age of the residents.

Age	0-10	10-20	20-30	30-40	40-50	50-60
Number of Residents	2	6	9	7	4	2

Solution

Age	Number of Residents(f)	Midvalue(x)	fx
0-10	2	5	10
10-20	6	15	90
20-30	9	25	225
30-40	7	35	245
40-50	4	45	180
50-60	2	55	110
	$\sum f = 30$		$\sum fx = 860$

Mean= $\overline{X} = \frac{\sum fx}{\sum f} = \frac{860}{30} = 28.67$

Hence the average age $= 28.67$.

Assumed Mean Method

We have seen how to find the arithmetic mean of a grouped data quickly using the direct method formula. However, if the observations are large, finding the products of the observations and their corresponding frequencies, and then adding them is not only difficult and time consuming but also has chances of errors. In such cases, we can use the Assumed Mean Method to find the arithmetic mean of grouped data.

Steps

1. Assume any value of the observations as the Mean (A). Preferably, choose the middle value.
2. Calculate the deviation $d = x - A$ for each class
3. Multiply each of the corresponding frequency 'f' with 'd' and obtain Σfd
4. Apply the formula $\overline{X} = A + \dfrac{\sum fd}{\sum f}$

Example 8.2 Find the mean for the following frequency table:

Class Interval	100-120	120-140	140-160	160-180	180-200	200-220	220-240
Frequency	10	8	4	4	3	1	2

Solution

Let Assumed mean $A = 170$

Class Interval	Frequency f	Mid value x	$d = x - A$ $d = x - 170$	fd
100-120	10	110	–60	–600
120-140	8	130	–40	–320
140-160	4	150	–20	–80
160-180	4	170	0	0
180-200	3	190	20	60
200-220	1	210	40	40
220-240	2	230	60	120
	$\sum f = 32$			$\sum fd = -780$

Mean $\overline{X} = A + \dfrac{\sum fd}{\sum f}$

$= 170 + \left(\dfrac{-780}{32}\right)$

Therefore, $\overline{X} = 170 - 24.375$

$= 145.625$

Step Deviation Method

In order to simplify the calculation, we divide the deviation by the width of class intervals (i.e. calculate $\dfrac{x-A}{c}$) and then multiply by c in the formula for getting the mean of the data. The formula to calculate the Arithmetic Mean is

$$\bar{X} = A + \left[\frac{\sum fd}{\sum f} \times c\right], \text{ where } d = \frac{x - A}{c}$$

Example 8.3

Find the mean of the following distribution using Step Deviation Method.

Class Interval	0-8	8-16	16-24	24-32	32-40	40-48
Frequency (f)	10	20	14	16	18	22

Solution

Let Assumed mean $A = 28$, class width c =8

Class Interval	Mid Value x	Frequency f	$d = \frac{x-A}{c}$	fd
0-8	4	10	–3	–30
8-16	12	20	–2	–40
16-24	20	14	–1	–14
24-32	28	16	0	0
32-40	36	18	1	18
40-48	44	22	2	44
		$\sum f = 100$		$\sum fd = -22$

Mean

$$\bar{X} = A + \frac{\sum fd}{\sum f} \times c$$

$$= 28 + \left(\frac{-22}{100}\right) \times 8$$

$$= 28 - 1.76 = 26.24$$

Note

- When x_i and f_i are small, then Direct Method is the appropriate choice.
- When x_i and f_i are numerically large numbers, then Assumed Mean Method or Step Deviation Method can be used.
- When class sizes are unequal and d numerically large,we can still use Step Deviation Method.

8.4.4 A special property of the Arithmetic Mean

1. *The sum of the deviations of the entries from the arithmetic mean is always zero.*

If $x_1, x_2, x_3, \ldots, x_n$ are n observations taken from the arithmetic mean $\bar{X}$

then $(x_1 - \bar{X}) + (x_2 - \bar{X}) + (x_3 - \bar{X}) + \ldots + (x_n - \bar{X}) = 0$ .Hence $\sum_{i=1}^{n}(x_i - \bar{X}) = 0$

2. *If each observation is increased or decreased by k (constant) then the arithmetic mean is also increased or decreased by k respectively.*

3. *If each observation is multiplied or divided by k, $k \neq 0$, then the arithmetic mean is also multiplied or divided by the same quantity k respectively.*

Example 8.4 Find the sum of the deviations from the arithmetic mean for the following observations:

21, 30, 22, 16, 24, 28, 18, 17

Solution

$$\overline{X} = \frac{\sum_{i=1}^{8} x_i}{n} = \frac{21+30+22+16+24+28+18+17}{8} = \frac{176}{8} = 22$$

Deviation of an entry x_i from the arithmetic mean $\overline{X}$ is $x_i - \overline{X}$, $i = 1, 2, ...8$.

Sum of the deviations

$= (21-22)+(30-22)+(22-22)+(16-22)+(24-22)+(28-22)+(18-22)+(17-22)$

$= 16-16 = 0.$ or equivalently, $\sum_{i=1}^{8}(x_i - \overline{X}) = 0$

Hence, we conclude that sum of the deviations from the Arithmetic Mean is zero.

Example 8.5 The arithmetic mean of 6 values is 45 and if each value is increased by 4, then find the arithmetic mean of new set of values.

Solution

Let $x_1, x_2, x_3, x_4, x_5, x_6$ be the given set of values then $\frac{\sum_{i=1}^{6} x_i}{6} = 45$.

If each value is increased by 4, then the mean of new set of values is

$$\text{New A.M. } \overline{X} = \frac{\sum_{i=1}^{6}(x_i + 4)}{6}$$

$$= \frac{(x_1+4)+(x_2+4)+(x_3+4)+(x_4+4)+(x_5+4)+(x_6+4)}{6}$$

$$= \frac{\sum_{i=1}^{6} x_i + 24}{6} = \frac{\sum_{i=1}^{6} x_i}{6} + 4$$

$$\overline{X} = 45 + 4 = 49.$$

> **Progress Check**
>
> Mean of 10 observations is 48 and 7 is subtracted to each observation, then mean of new observation is __________

Example 8.6 If the arithmetic mean of 7 values is 30 and if each value is divided by 3, then find the arithmetic mean of new set of values

Solution

Let X represent the set of seven values $x_1, x_2, x_3, x_4, x_5, x_6, x_7$.

Then $$\overline{X} = \frac{\sum_{i=1}^{7} x_i}{7} = 30 \text{ or } \sum_{i=1}^{7} x_i = 210$$

If each value is divided by 3, then the mean of new set of values is

> **Progress Check**
>
> 1. The Mean of 12 numbers is 20. If each number is multiplied by 6, then the new mean is __
> 2. The Mean of 30 numbers is 16. If each number is divided by 4, then the new mean is___

$$\frac{\sum_{i=1}^{7}\frac{x_i}{3}}{7} = \frac{\left(\frac{x_1}{3}+\frac{x_2}{3}+\frac{x_3}{3}+\frac{x_4}{3}+\frac{x_5}{3}+\frac{x_6}{3}+\frac{x_7}{3}\right)}{7}$$

$$= \frac{\sum_{i=1}^{7}x_i}{21} = \frac{210}{21} = 10$$

Aliter

If Y is the set of values obtained by dividing each value of X by 3.

Then, $\bar{Y} = \frac{\bar{X}}{3} = \frac{30}{3} = 10.$

Example 8.7 The average mark of 25 students was found to be 78.4. Later on, it was found that score of 96 was misread as 69. Find the correct mean of the marks.

Solution

Given that the total number of students $n = 25, \bar{X} = 78.4$

So, Incorrect $\sum x = \bar{X} \times n \quad = 78.4 \times 25 = 1960$

Correct $\Sigma x = incorrect \sum x - wrong\ entry + correct\ entry$

$= 1960 - 69 + 96 = 1987$

Correct $\bar{X} = \frac{correct \sum x}{n} = \frac{1987}{25} = 79.48$

Progress Check

There are four numbers. If we leave out any one number, the average of the remaining three numbers will be 45, 60, 65 or 70. What is the average of all four numbers?

Exercise 8.1

1. In a week, temperature of a certain place is measured during winter are as follows 26°C, 24°C, 28°C, 31°C, 30°C, 26°C, 24°C. Find the mean temperature of the week.

2. The mean weight of 4 members of a family is 60kg.Three of them have the weight 56kg, 68kg and 72kg respectively. Find the weight of the fourth member.

3. In a class test in mathematics, 10 students scored 75 marks, 12 students scored 60 marks, 8 students scored 40 marks and 3 students scored 30 marks. Find the mean of their score.

4. In a research laboratory scientists treated 6 mice with lung cancer using natural medicine. Ten days later, they measured the volume of the tumor in each mouse and given the results in the table.

Mouse marking	1	2	3	4	5	6
Tumor Volume(mm^3)	145	148	142	141	139	140

Find the mean.

5. If the mean of the following data is 20.2, then find the value of p

Marks	10	15	20	25	30
No. of students	6	8	p	10	6

6. In the class, weight of students is measured for the class records. Calculate mean weight of the class students using Direct method.

Weight in kg	15-25	25-35	35-45	45-55	55-65	65-75
No. of students	4	11	19	14	0	2

7. Calculate the mean of the following distribution using Assumed Mean Method:

Class Interval	0-10	10-20	20-30	30-40	40-50
Frequency	5	7	15	28	8

8. Find the Arithmetic Mean of the following data using Step Deviation Method:

Age	15-19	20-24	25-29	30-34	35-39	40-44
No. of persons	4	20	38	24	10	9

8.5 Median

The arithmetic mean is typical of the data because it 'balances' the numbers; it is the number in the 'middle', pulled up by large values and pulled down by smaller values.

Suppose four people of an office have incomes of ₹5000, ₹6000, ₹7000 and ₹8000. Their mean income can be calculated as $\frac{5000+6000+7000+8000}{4}$ which gives ₹6500. If a fifth person with an income of ₹ 29000 is added to this group, then the arithmetic mean of all the five would be $\frac{5000+6000+7000+8000+29000}{5} = \frac{55000}{5}$ = ₹11000. Can one say that the average income of ₹11000 truly represents the income status of the individuals in the office? Is it not, misleading? The problem here is that an extreme score affects the Mean and can move the mean away from what would generally be considered the central area.

In such situations, we need a different type of average to provide reasonable answers.

Median is the value which occupies the middle position when all the observations are arranged in an ascending or descending order. It is a positional average.

For example, the height of nine students in a class are 122 cm, 124 cm, 125 cm, 135 cm, 138 cm, 140cm, 141cm, 147 cm, and 161 cm.

(i) Usual calculation gives Arithmetic Mean to be 137 cm.

(ii) If the heights are neatly arranged in, say, ascending order, as follows 122 cm, 124 cm, 125 cm, 135 cm, 138 cm, 140cm, 141cm, 147 cm, 161 cm, one can observe the value 138 cm is such that equal number of items lie on either side of it. Such a value is called the Median of given readings.

(iii) Suppose a data set has 11 items arranged in order. Then the median is the 6th item because it will be the middlemost one. If it has 101 items, then 51st item will be the Median.

If we have an odd number of items, one can find the middle one easily. In general, if a data set has n items and n is odd, then the median will be the $\left(\frac{n+1}{2}\right)^{th}$ item.

(iv) If there are 6 observations in the data, how will you find the Median? It will be the average of the middle two terms. (Shall we denote it as 3.5^{th} term?) If there are 100 terms in the data, the Median will be 50.5^{th} term!

In general, if a data set has n items and n is even, then the Median will be the average of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{\text{th}}$ items.

Example 8.8 The following are scores obtained by 11 players in a cricket match 7, 21, 45, 12, 56, 35, 25, 0, 58, 66, 29. Find the median score.

Solution

Let us arrange the values in ascending order.

0,7,12,21,25,29,35,45,56,58,66

The number of values = 11 which is odd

$$\text{Median} = \left(\frac{11+1}{2}\right)^{\text{th}} \text{ value}$$

$$= \left(\frac{12}{2}\right)^{\text{th}} \text{ value} = 6^{\text{th}} \text{ value} = 29$$

Example 8.9 For the following ungrouped data 10, 17, 16, 21, 13, 18, 12, 10, 19, 22. Find the median.

Solution

Arrange the values in ascending order.

10, 10, 12, 13, 16, 17, 18, 19, 21, 22.

The number of values = 10

$$\text{Median} = \text{Average of } \left(\frac{10}{2}\right)^{\text{th}} \text{ and } \left(\frac{10}{2}+1\right)^{th} \text{ values}$$

$$= \text{Average of } 5^{\text{th}} \text{ and } 6^{\text{th}} \text{ values}$$

$$= \frac{16+17}{2} = \frac{33}{2} = 16.5$$

Example 8.10 The following table represents the marks obtained by a group of 12 students in a class test in Mathematics and Science.

Marks (Mathematics)	52	55	32	30	60	44	28	25	50	75	33	62
Marks (Science)	54	42	48	49	27	25	24	19	28	58	42	69

Indicate in which subject, the level of achievement is higher?

Solution

Let us arrange the marks in the two subjects in ascending order.

Marks (Mathematics)	25	28	30	32	33	44	50	52	55	60	62	75
Marks (Science)	19	24	25	27	28	42	42	48	49	54	58	69

Since the number of students is 12, the marks of the middle-most student would be the mean mark of 6^{th} and 7^{th} students.

Therefore , Median mark in Mathematics $= \dfrac{44+50}{2} = 47$

Median mark in Science $= \dfrac{42+42}{2} = 42$

Here the median mark in Mathematics is greater than the median mark in Science. Therefore, the level of achievement of the students is higher in Mathematics than Science.

8.5.1 Median-Ungrouped Frequency Distribution

(i) Arrange the data in ascending (or) decending order of magnitude.

(ii) Construct the cumulative frequency distribution. Let N be the total frequency.

(iii) If N is odd, median $= \left(\frac{N+1}{2}\right)^{th}$ observation.

(iv) If N is even, median $= \left(\frac{\left(\frac{N}{2}\right)^{\text{th}} \text{observation} + \left(\frac{N}{2}+1\right)^{\text{th}} \text{observation}}{2}\right)$

Example 8.11 Calculate the median for the following data:

Height (cm)	160	150	152	161	156	154	155
No. of Students	12	8	4	4	3	3	7

Solution

Let us arrange the marks in ascending order and prepare the following data:

Height (cm)	Number of students (f)	Cumulative frequency (cf)
150	8	8
152	4	12
154	3	15
155	7	22
156	3	25
160	12	37
161	4	41

Here N = 41

Median = size of $\left(\frac{N+1}{2}\right)^{th}$ value = size of $\left(\frac{41+1}{2}\right)^{th}$ value = size of 21^{st} value.

If the 41 students were arranged in order (of height), the 21^{st} student would be the middle most one, since there are 20 students on either side of him/her. We therefore need to find the height against the 21^{st} student. 15 students (see cumulative frequency) have height less than or equal to 154 cm. 22 students have height less than or equal to 155 cm. This means that the 21^{st} student has a height 155 cm.

Therefore, Median = 155 cm

8.5.2 Median - Grouped Frequency Distribution

In a grouped frequency distribution, computation of median involves the following

Steps

(i) Construct the cumulative frequency distribution.

(ii) Find $\left(\frac{N}{2}\right)^{th}$ term.

(iii) The class that contains the cumulative frequency $\frac{N}{2}$ is called the median class.

(iv) Find the median by using the formula:

$$\text{Median} = l + \frac{\left(\frac{N}{2} - m\right)}{f} \times c$$

Where *l = Lower limit of the median class,* *f = Frequency of the median class*

c = Width of the median class, *N = The total frequency $\left(\sum f\right)$*

m = cumulative frequency of the class preceeding the median class

Example 8.12 The following table gives the weekly expenditure of 200 families. Find the median of the weekly expenditure.

Weekly expenditure (₹)	0-1000	1000-2000	2000-3000	3000-4000	4000-5000
Number of families	28	46	54	42	30

Solution

Weekly Expenditure	Number of families (*f*)	Cumulative frequency (*cf*)
0-1000	28	28
1000-2000	46	74
2000-3000	54	128
3000-4000	42	170
4000-5000	30	200
	N=200	

Median class $= \left(\frac{N}{2}\right)^{th}$ value $= \left(\frac{200}{2}\right)^{th}$ value

$= 100^{th}$ value

Median class $=$ 2000 – 3000

$\frac{N}{2} = 100 \quad l = 2000$

$m = 74, \quad c = 1000, \quad f = 54$

$$\text{Median} = l + \frac{\left(\frac{N}{2} - m\right)}{f} \times c$$

Progress Check

1. The median of the first four whole numbers ______.
2. If 4 is also included to the collection of first four whole numbers then median value is _________.
3. The difference between two median is ______.

$= 2000 + \left(\frac{100-74}{54}\right) \times 1000$

$= 2000 + \left(\frac{26}{54}\right) \times 1000 = 2000 + 481.5$

$= 2481.5$

Example 8.13 The Median of the following data is 24. Find the value of x.

Class Interval (*CI*)	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
Frequency (*f*)	6	24	x	16	9

Solution

Class Interval (*CI*)	Frequency (*f*)	Cumulative frequency (*cf*)
0-10	6	6
10-20	24	30
20-30	x	$30 + x$
30-40	16	$46 + x$
40-50	9	$55 + x$
	$N = 55 + x$	

Since the median is 24 and median class is 20 – 30

$l = 20 \quad N = 55 + x, \quad m = 30, \quad c = 10, \quad f = x$

$$\text{Median} = l + \frac{\left(\frac{N}{2} - m\right)}{f} \times c$$

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

$$4 = \frac{5x - 25}{x} \quad \text{(after simplification)}$$

$$4x = 5x - 25$$

$$5x - 4x = 25$$

$$x = 25$$

Note

The median is a good measure of the average value when the data include extremely high or low values, because these have little influence on the outcome.

BEL5EL

Exercise 8.2

1. Find the median of the given values : 47, 53, 62, 71, 83, 21, 43, 47, 41.
2. Find the Median of the given data: 36, 44, 86, 31, 37, 44, 86, 35, 60, 51

3. The median of observation 11, 12, 14, 18, x+2, x+4, 30, 32, 35, 41 arranged in ascending order is 24. Find the values of x.

4. A researcher studying the behavior of mice has recorded the time (in seconds) taken by each mouse to locate its food by considering 13 different mice as 31, 33, 63, 33, 28, 29, 33, 27, 27, 34, 35, 28, 32. Find the median time that mice spent in searching its food.

5. The following are the marks scored by the students in the Summative Assessment exam

Class	0-10	10-20	20-30	30-40	40-50	50-60
No. of Students	2	7	15	10	11	5

Calculate the median.

6. The mean of five positive integers is twice their median. If four of the integers are 3, 4, 6, 9 and median is 6, then find the fifth integer.

8.6 Mode

(i) The votes obtained by three candidates in an election are as follows:

Name of the Candidate	Votes Polled
Mr. X	4, 12, 006
Mr. Y	9, 87, 991
Mr. Z	7, 11, 973
Total	21, 11, 970

Who will be declared as the winner? Mr. Y will be the winner, because the number of votes secured by him is the highest among all the three candidates. Of course, the votes of Mr. Y do not represent the majority population (because there are more votes against him). However, he is declared winner because the mode of selection here depends on the highest among the candidates.

BEV1G9

(ii) An Organisation wants to donate sports shoes of same size to maximum number of students of class IX in a School. The distribution of students with different shoe sizes is given below.

Shoe Size	5	6	7	8	9	10
No. of Students	10	12	27	31	19	1

If it places order, shoes of only one size with the manufacturer, which size of the shoes will the organization prefer?

In the above two cases, we observe that mean or median does not fit into the situation. We need another type of average, namely the Mode.

The mode is the number that occurs most frequently in the data.

When you search for some good video about Averages on You Tube, you look to watch the one with maximum views. Here you use the idea of a mode.

8.6.1 Mode - Raw Data

For an individual data mode is the value of the variable which occurs most frequently.

Example 8.14 In a rice mill, seven labours are receiving the daily wages of ₹500, ₹600, ₹600, ₹800, ₹800, ₹800 and ₹1000, find the modal wage.

Solution

In the given data ₹800 occurs thrice.Hence the mode is ₹ 800.

Example 8.15 Find the mode for the set of values 17, 18, 20, 20, 21, 21, 22, 22.

Solution

In this example, three values 20, 21, 22 occur two times each. There are three modes for the given data!

Note

- A distribution having only one mode is called unimodal.
- A distribution having two modes is called bimodal.
- A distribution having Three modes is called trimodal.
- A distribution having more than three modes is called multimodal.

8.6.2 Mode for Ungrouped Frequency Distribution

In a ungrouped frequency distribution, the value of the item having maximum frequency is taken as the mode.

Example 8.16 A set of numbers consists of five 4's, four 5's, nine 6's,and six 9's. What is the mode.

Solution

Size of item	4	5	6	9
Frequency	5	4	9	6

6 has the maximum frequency 9. Therefore 6 is the mode.

8.6.3 Mode – Grouped Frequency Distribution:

In case of a grouped frequency distribution, the exact values of the variables are not known and as such it is very difficult to locate mode accurately. In such cases, if the class intervals are of equal width, an appropriate value of the mode may be determined by

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

The class interval with maximum frequency is called the modal class.

Where l - lower limit of the modal class; f - frequency of the modal class

f_1 - frequency of the class just preceding the modal class

f_2 - frequency of the class succeeding the modal class

c - width of the class interval

Example 8.17 Find the mode for the following data.

Marks	1-5	6-10	11-15	16-20	21-25
No. of students	7	10	16	32	24

Solution

Marks	f
0.5-5.5	7
5.5-10.5	10
10.5-15.5	16
15.5-20.5	32
20.5-25.5	24

Note

To convert discontinuous class interval into continuous class interval, 0.5 is to be substracted at the lower limit and 0.5 is to be added at the upper limit for each class interval.

Modal class is 16 -20 since it has the maximum frequency.

$l = 15.5, f = 32, f_1 = 16, f_2 = 24, c = 20.5 - 15.5 = 5$

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

$$= 15.5 + \left(\frac{32 - 16}{64 - 16 - 24}\right) \times 5$$

$$= 15.5 + \left(\frac{16}{24}\right) \times 5 = 15.5 + 3.33 = 18.83.$$

8.6.4 An Empirical Relationship between Mean, Medan and Mode

We have seen that there is an approximate relation that holds among the three averages discussed earlier, when the frequencies are nearly symmetrically distributed.

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

Example 8.18 In a distribution, the mean and mode are 66 and 60 respectively. Calculate the median.

Solution

Given, Mean = 66 and Mode = 60.

Using, Mode $\approx$ 3Median – 2Mean

$$60 \approx 3\text{Median} - 2(66)$$

$$3\text{ Median} \approx 60 + 132$$

Therefore, $\text{Median} \approx \dfrac{192}{3} \approx 64$

Exercise 8.3

1. The monthly salary of 10 employees in a factory are given below :
₹5000, ₹7000, ₹5000, ₹7000, ₹8000, ₹7000, ₹7000, ₹8000, ₹7000, ₹5000
Find the mean, median and mode.

2. Find the mode of the given data : 3.1, 3.2, 3.3, 2.1, 1.3, 3.3, 3.1

3. For the data 11, 15, 17, x+1, 19, x–2, 3 if the mean is 14 , find the value of x. Also find the mode of the data.

4. The demand of track suit of different sizes as obtained by a survey is given below:

Size	38	39	40	41	42	43	44	45
No. of Persons	36	15	37	13	26	8	6	2

Which size is in greater demanded?

5. Find the mode of the following data:

Marks	0-10	10-20	20-30	30-40	40-50
Number of students	22	38	46	34	20

6. Find the mode of the following distribution:

Weight(in kgs)	25-34	35-44	45-54	55-64	65-74	75-84
Number of students	4	8	10	14	8	6

Exercise 8.4

Multiple choice questions

1. Let m be the mid point and b be the upper limit of a class in a continuous frequency distribution. The lower limit of the class is

   (1) $2m - b$ (2) $2m+b$ (3) $m-b$ (4) $m-2b$.

2. The mean of a set of seven numbers is 81. If one of the numbers is discarded , the mean of the remaining numbers is 78 . The value of discarded number is

   (1) 101 (2) 100 (3) 99 (4) 98.

3. A particular observation which occurs maximum number of times in a given data is called its

   (1) Frequency (2) range (3) mode (4) Median.

4. For which set of numbers do the mean, median and mode all have the same values?

   (1) 2,2,2,4 (2) 1,3,3,3,5 (3) 1,1,2,5,6 (4) 1,1,2,1,5.

5. The algebraic sum of the deviations of a set of n values from their mean is

   (1) 0 (2) n-1 (3) n (4) n+1.

6. The mean of a,b,c,d and e is 28 . If the mean of a, c and e is 24 , then mean of b and d is_

   (1) 24 (2) 36 (3) 26 (4) 34

7. If the mean of five observations x, $x+2$, $x+4$, $x+6$, $x+8$, is 11 , then the mean of first three observations is

   (1) 9 (2) 11 (3) 13 (4) 15.

8. The mean of 5, 9, x, 17, and 21 is 13, then find the value of x

   (1) 9 (2) 13 (3) 17 (4) 21

9. The mean of the square of first 11 natural numbers is

   (1) 26 (2) 46 (3) 48 (4) 52.

10. The mean of a set of numbers is $\bar{X}$. If each number is multiplied by z, the mean is

    (1) $\bar{X}+z$ (2) $\bar{X} - z$ (3) $z\ \bar{X}$ (4) $\bar{X}$

Project

1. Prepare a frequency table of the top speeds of 20 different land animals. Find mean, median and mode. Justify your answer.

2. From the record of students particulars of the class,

   (i) Find the mean age of the class (using class interval)

   (ii) Calculate the mean height of the class(using class intervals)

Points to Remember

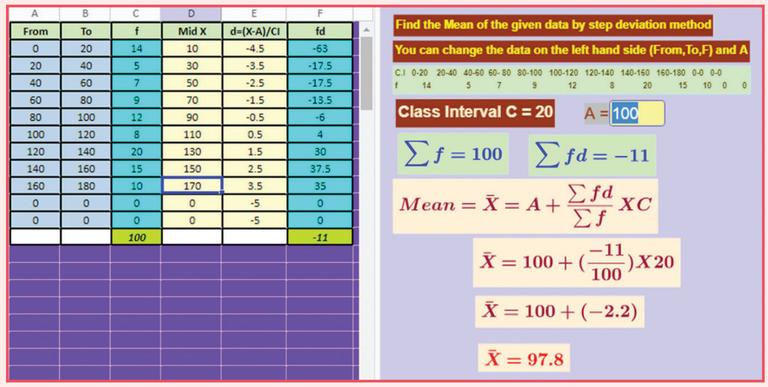
- The information collected for a definite purpose is called data.
- The data collected by the investigator are known as primary data. When the information is gathered from an external source,the data are called secondary data.
- Initial data obtained through unorganized form are called Raw data.
- Mid Value $= \frac{UCL + LCL}{2}$ (where UCL–Upper Class Limit, LCL–Lower Class Limit).
- Size of the class interval $= UCL - LCL$.
- The mean for grouped data:

Direct Method	Assumed Mean Method	Step-Deviation Method
$\bar{X} = \frac{\sum fx}{\sum f}$	$\bar{X} = A + \frac{\sum fd}{\sum f}$	$\bar{X} = A + \left[\frac{\sum fd}{\sum f} \times c\right]$

- The cumulative frequency of a class is the frequency obtained by adding the frequency of all up to the classes preceeding the given class.
- Formula to find the median for grouped data: Median $= l + \frac{\left(\frac{N}{2} - m\right)}{f} \times c$.
- Formula to find the mode for grouped data: Mode $= l + \left(\frac{f - f_1}{2f - f_1 - f_2}\right) \times c$.

ICT Corner

Expected Result is shown in this picture

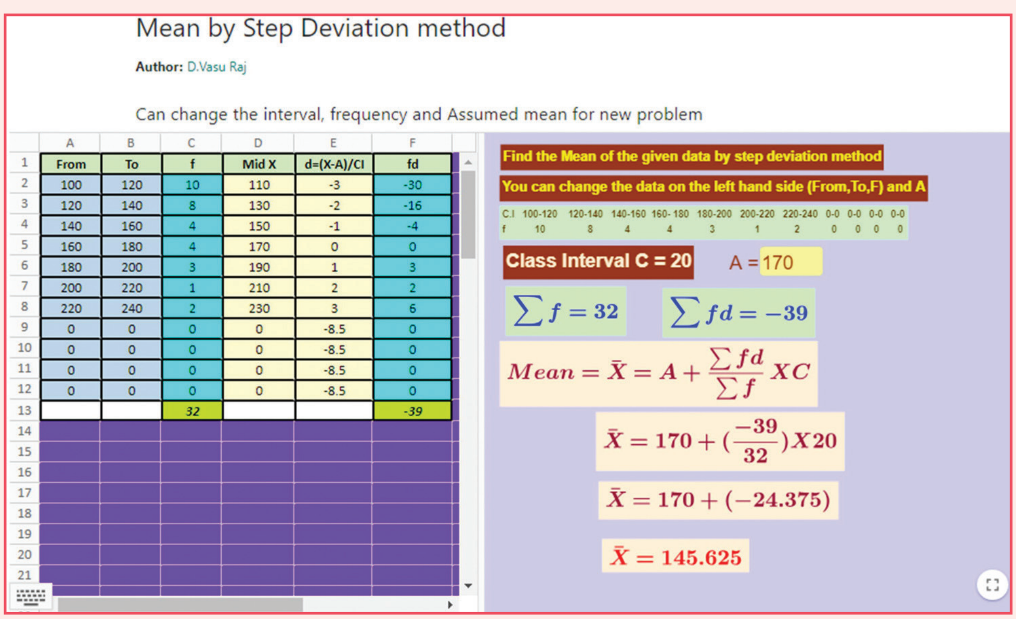


Step

Open the browser, type the URL Link given below (or) Scan the QR Code. GeoGebra work sheet named "Mean by step deviation method" will open.

In the work sheet Example 5.5 is given. Observe the steps. You can change the question by typing new data "From", "To" and "Frequency f" in the spread sheet on Left hand side. After that change, the Assumed mean on the right-hand side and check the calculation.

Step 1



Browse in the link

Mean by step deviation method:
https://ggbm.at/NWcKTRtA or Scan the QR Code

B566_9_MAT_EM_T3