

BLENDED LEARNING MODULE



EDUCATION

(Honours)

Semester:-III

Paper:- Core-VII Unit:- II

Department of Education
BasudevGodabari Degree College
Kesaibahal, Sambalpur

Plan Unit – II

Unit. 2

Measures of Central Tendency and Variability

- i) **Mean, Median and Mode---**
Concept, Computational Process, Uses and Limitation.
- ii) **Range, Average Deviation, Quartile Deviation and Standard Deviation**
Concept, Computational Process, Uses and Limitation.

Learning Objective:

Compute and use various statistical measure of Average, Variation and bi-variable distribution to in analysis and interpretation of educational data.

Measures of Central Tendency

Central tendency is a descriptive summary of a dataset through a single value that reflects the center of the data distribution. Along with the variability (dispersion) of a dataset, central tendency is a branch of descriptive statistics.

The central tendency is one of the most quintessential concepts in statistics. Although it does not provide information regarding the individual values in the dataset, it delivers a comprehensive summary of the whole dataset.

Measures of Central Tendency

Generally, the central tendency of a dataset can be described using the following measures:

- **Mean (Average):** Represents the sum of all values in a dataset divided by the total number of the values.
- **Median:** The middle value in a dataset that is arranged in ascending order (from the smallest value to the largest value). If a dataset contains an even number of values, the median of the dataset is the mean of the two middle values.

- **Mode:** Defines the most frequently occurring value in a dataset. In some cases, a dataset may contain multiple modes while some datasets may not have any mode at all.

Arithmetic Mean (AP)

It is the ratio of all observations to the total number of observations. The arithmetic mean can also inform or model concepts outside of statistics. In a physical sense, the arithmetic mean can be thought of as a centre of gravity. From the mean of a data set, we can think of the average distance the data points are from the mean as standard deviation. The square of standard

Arithmetic Mean Definition

Arithmetic mean represents a number that is obtained by dividing the sum of the elements of a set by the number of values in the set. So you can use the layman term Average, or be a little bit fancier and use the word “Arithmetic mean” your call, take your pick -they both mean the same. The arithmetic mean may be either

- Simple Arithmetic Mean
- Weighted Arithmetic Mean

Arithmetic Mean Formula

If any data set consisting of the values $b_1, b_2, b_3, \dots, b_n$ then the arithmetic mean B is defined as:

$B = (\text{Sum of all observations}) / (\text{Total number of observation})$

$$= 1/n \sum_{i=1}^n b_i = (b_1 + b_2 + b_3 + \dots + b_n) / n$$

If these n observations have corresponding frequencies, the arithmetic mean is computed using the formula

$$X = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n}{N}$$

using Sigma notation $= \frac{\sum_{i=1}^n x_i f_i}{N}$

where $N = f_1 + f_2 + \dots + f_n$.

The above formula can also be used to find the weighted arithmetic mean by taking f_1, f_2, \dots, f_n as the weights of x_1, x_2, \dots, x_n .

When the frequencies divided by N are replaced by probabilities p_1, p_2, \dots, p_n we get the formula for the expected value of a discrete random variable.

$$X = x_1 p_1 + x_2 p_2 + \dots + x_n p_n \text{ or}$$

$$\text{using Sigma notation} = \sum_{i=1}^n x_i p_i$$

Properties of Arithmetic Mean

Some important properties of the arithmetic mean are as follows:

- The sum of deviations of the items from their arithmetic mean is always zero, i.e. $\sum (x - X) = 0$.
- The sum of the squared deviations of the items from Arithmetic Mean (A.M) is minimum, which is less than the sum of the squared deviations of the items from any other values.
- If each item in the arithmetic series is substituted by the mean, then the sum of these replacements will be equal to the sum of the specific items.

Representative Values of Data

We see the use of representative value quite regularly in our daily life. When you ask about the mileage of the car, you are asking for the representative value of the amount of distance travelled to the amount of fuel consumed. This doesn't mean that the temperature in Shimla is constantly the representative value but that overall, it amounts to the average value. Average here represents a number that expresses a central or typical value in a set of data, calculated by the sum of values divided by the number of values.

Arithmetic Mean Example

The Arithmetic means utilizes two basic mathematical operations, addition and division to find a central value for a set of values. If you wanted to find the arithmetic means of the runs scored by Virat Kohli in the last few innings, all you would have to do is sum up his runs to obtain sum total and then divide it by the number of innings. For example;

Innings	1	2	3	4	5	6	7	8	9	10
Runs	50	59	90	8	106	117	59	91	7	74

The arithmetic mean of Virat Kohli's batting scores also called his Batting Average is;

Sum of runs scored/Number of innings = $661/10$

The arithmetic mean of his scores in the last 10 innings is 66.1. If we add another score to this sum, say his 11th innings, the arithmetic mean will proportionally change. If the runs scored in 11th innings are 70, the new average becomes;

$$\frac{661 + 70}{10 + 1} = \frac{731}{11} = 66.45$$

Direct Method

Individual Series

Let X is the variable which takes values $x_1, x_2, x_3, \dots, x_n$ over 'n' times, then arithmetic mean, simply the mean of X, denoted by bar over the variable X is given by,

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Discrete Series

$$\bar{X} = \frac{\sum fx}{N} = \frac{\sum f}{N}$$

$\sum fx$ = Sum of the product of the values and their corresponding frequencies
N = Sum of the frequencies i.e., $\sum f$ or total number of observations.

Continuous Series

$$\bar{X} = \frac{\sum fm}{N} = \frac{\sum f}{\sum f}$$

$\sum fm$ = Sum of the product of mid values and their corresponding frequencies

N = Sum of the frequencies i.e., $\sum f$ or total number of observations.

Example 01

Calculate the mean of following data. Marks obtained by 6 students given: 20, 15, 23, 22, 25, 20.

Solution:

Mean marks,

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$= \frac{20 + 15 + 23 + 22 + 25 + 20}{6}$$

$$= \frac{125}{6} = 20.83$$

Six month income of a departmental store is given below. Find mean income of store.

Month	Jan	Feb	Mar	Apr	May	June
Income(\$)	25000	30000	45000	20000	25000	20000

Solution:

$n = \text{Total number of items (observations)} = 6$

$\text{Total income} = \sum x_i = \$ (25000 + 30000 + 45000 + 20000 + 25000 + 20000)$
 $= \$165000$

Mean Income,

$= \frac{\sum x_i}{n} = \frac{\$165000}{6} = \$27500$

Example 03

Calculate the arithmetic mean by direct method from the following data

Wage	10	20	30	40	50
No. of Workers	4	5	3	2	5

Solution:

Let us denote the wages by x and number of workers by f .

Wages (x)	No. of Workers (f)	fx
10	4	40
20	5	100
30	3	90

40	2	80
50	5	250
	$\Sigma f = 19$	$\Sigma fx = 560$

Hence, mean wage of the workers,

$$\bar{X} = \frac{\Sigma fx}{\Sigma f} = \frac{560}{19} = 29.47$$

Example 04

Calculate the missing value when its mean is 115.86.

Wage	110	112	113	117	–	125	128	130
No. of Workers	25	17	13	15	14	8	6	2

Solution:

Let us denote the wages by x, number of workers by f and missing item by 'a'.

Wages (x)	No. of Workers (f)	fx
110	25	2750

112	17	1904
113	13	1469
117	15	1755
a	14	14a
125	8	1000
128	6	768
130	2	260
	$\sum f = 100$	$\sum fx = 9906 + 14a$

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

$$\Rightarrow 115.86 = \frac{9906 + 14a}{100}$$

$$\Rightarrow 115.86 \times 100 = 9906 + 14a$$

$$\Rightarrow 14a = 11586 - 9906$$

$$\Rightarrow 14a = 1680$$

$$\Rightarrow a = \frac{1680}{14}$$

$$\therefore a = 120$$

Example 05

Find the mean for the following distribution by using direct method.

Class Interval	84-90	90-96	96-102	102-108	108-114
frequency	8	12	15	10	5

This is the case of continuous series. Let 'm' be the mid-value and 'f' be the frequency.

Class interval	Mid-value (m)	Frequency (f)	fx
84-90	97	8	776
90-96	93	12	1116
96-102	99	15	1485
102-108	105	10	1050
108-114	111	5	555

		$\Sigma f = 50$	$\Sigma fx = 4982$
--	--	-----------------	--------------------

Hence, the mean,

$$\bar{X} = \frac{\Sigma fx}{\Sigma f} = \frac{4982}{50} = 99.64$$

Short Cut Method

Individual Series

The arithmetic mean can also be calculated by taking deviations from any arbitrary points. Steps of this method are given below.

Step 1: Assume any one value as a mean which is called arbitrary average (A).

Step 2: Find the difference (deviations) of each value from arbitrary average.

$$d = x_i - A$$

Step 3: Add all differences to get Σd .

Step 4: Use following equation and compute the mean value.

$$\bar{X} = A + \frac{\Sigma d}{n}$$

Discrete Series

$$\bar{X} = A + \frac{\Sigma fd}{\Sigma f}$$

Continuous Series

$$\bar{X} = A + \frac{\Sigma fd}{\Sigma f}$$

Where,

n = Total number of observations.

Σd = Total deviation value

d = Deviation of item from the assumed mean.

A = Assumed mean.

Σfd = Sum of products of deviations and their corresponding frequencies.

Example 01

Determine the average salary of a staff from the following data relating to the monthly salaries of the teaching staff of a college by using short cut method.

Salary (\$)	2200	2500	3000	3700	4500
No. of Staff	5	10	15	7	3

Solution:

Let assumed mean $A = 3000$

x	f	$D = (x - A)$	fd
2200	5	-800	-4000
2500	10	-500	-5000
3000	15	000	000
3700	7	700	4900
4500	3	1500	4500

	$\Sigma f = 40$		$\Sigma fd = 400$
--	-----------------	--	-------------------

We have,

$$X = A + \frac{\Sigma fd}{\Sigma f} = 3000 + \frac{400}{40} = 3000 + 10 = 3010$$

Example 02

Calculate the average marks obtained by BCA students in Mathematics paper by short cut method

Class of Marks	0-10	10-20	20-30	30-40	40-50
No. of students	5	3	7	25	20

Solution:

Let A = 25

Marks of Students	No. of Students(f)	Mid-values(m)	d = m – 25	fd
0-10	5	5	-20	-100
10-20	3	15	-10	-30

20-30	7	25	0	00
30-40	25	35	10	250
40-50	20	45	20	400
	$\sum f = 60$			$\sum fd = 520$

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

$$= 25 + \frac{520}{60} = 25 + 8.67 = 33.67$$

Step Deviation Method

This method is extension of short cut method. This method is used when the figures of deviations appear to be big and divisible by a common factor.

Individual Series

$$\bar{X} = A + \frac{\sum d'N}{N}$$

Discrete Series

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

Continuous Series

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

Where,

c = common factor by which each of the deviation is divided.

d' = the deviation from the assumed average divided by the common factor.

Example 01

Calculate the arithmetic mean by means of step deviation method

Marks	0-10	10-20	20-30	30-40	40-50
No. of students	20	24	40	36	20

Solution:

Let Assumed average $A = 25$

Marks	Mid-values(m)	f	$A = 25,$ $d = m - 25$	$d' = (d/10)$	fd'
0-10	5	20	-20	-2	-40
10-20	15	24	-10	-1	-24
20-30	$25 = A$	40	0	0	0
30-40	35	36	10	1	36

40-50	45	20	20	2	40
		$\sum f = 140$			$\sum fd' = 12$

$$\begin{aligned} X - &= A + \frac{\sum fd'}{\sum f} \times c \\ &= 25 + \frac{12}{140} \times 10 = 25.85 \end{aligned}$$

Merits of Arithmetic Mean

- The arithmetic mean is simple to understand and easy to calculate.
- It is influenced by the value of every item in the series.
- A.M is rigidly defined.
- It has the capability of further algebraic treatment.
- It is a measured value and not based on the position in the series.

Demerits of Arithmetic Mean

- It is changed by extreme items such as very small and very large items.
- It can rarely be identified by inspection.
- In some cases, A.M. does not represent the original item. For example, average patients admitted to a hospital are 10.7 per day.
- The arithmetic mean is not suitable in extremely asymmetrical distributions.

Median

Median, in statistics, is the middle value of the given list of data, when arranged in an order. The arrangement of data or observations can be done either in ascending order or descending order.

Example: The median of 2,3,4 is 3.

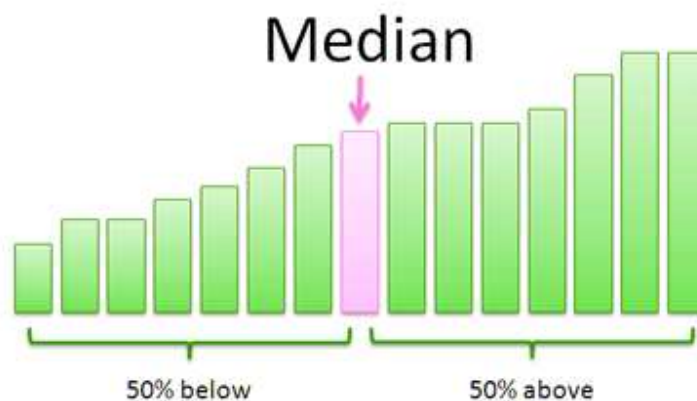
In Maths, the median is also a type of average, which is used to find the center value. Therefore, it is also called measure of central tendency.

Apart from the median, the other two central tendencies are mean and mode. Mean is the ratio of sum of all observations and total number of observations. Mode is the value in the given data-set, repeated most of the time.

is also defined as the center point of a polygon. For example, the median of a triangle is the line segment joining the vertex of triangle to the center of the opposite sides. Therefore, a median bisects the sides of triangle.

The median of a set of data is the middlemost number or center value in the set. The median is also the number that is halfway into the set.

To find the median, the data should be arranged, first, in order of least to greatest or greatest to the least value. A median is a number that is separated by the higher half of a data sample, a population or a probability distribution, from the lower half. The median is different for different types of distribution.



For example, the median of 3, 3, 5, 9, 11 is 5. If there is an even number of observations, then there is no single middle value; the median is then usually defined to be the mean of the two middle values: so the median of 3, 5, 7, 9 is $(5+7)/2 = 6$.

Median Formula

The formula to calculate the median of the finite number of data set is given here. Median formula is different for even and odd numbers of observations. Therefore, it is necessary to recognise first if we have odd number of values or even number of values in a given data set.

The formula to calculate the median of the data set is given as follow.

Odd Number of Observations

If the total number of observation given is odd, then the formula to calculate the median is:

$$\text{Median} = \{(n+1)/2\}^{\text{th}} \text{term}$$

where n is the number of observations

Even Number of Observations

If the total number of observation is even, then the median formula is:

$$\text{Median} = [(n/2)^{\text{th}} \text{ term} + \{(n/2)+1\}^{\text{th}}]/2$$

where n is the number of observations

How to Calculate the Median?

To find the median, place all the numbers in the ascending order and find the middle.

Example 1:

Find the Median of 14, 63 and 55

solution:

Put them in ascending order: 14, 55, 63

The middle number is 55, so the median is 55.

Example 2:

Find the median of the following:

4, 17, 77, 25, 22, 23, 92, 82, 40, 24, 14, 12, 67, 23, 29

Solution:

When we put those numbers in the order we have:

4, 12, 14, 17, 22, 23, 23, 24, 25, 29, 40, 67, 77, 82, 92,

There are fifteen numbers. Our middle is the eighth number:

The median value of this set of numbers is 24.

Example 3:

Rahul's family drove through 7 states on summer vacation. The prices of Gasoline differ from state to state. Calculate the median of gasoline cost.

1.79, 1.61, 2.09, 1.84, 1.96, 2.11, 1.75

Solution:

By organizing the data from smallest to greatest, we get:

1.61, 1.75, 1.79, **1.84** , 1.96, 2.09, 2.11

Hence, the median of the gasoline cost is 1.84. There are three states with greater gasoline costs and 3 with smaller prices.

How to calculate Median step by step?

You can use the following steps to calculate the median.

For ungrouped data:

Step 1. Arrange the given values in the ascending order.

Step 2. Find the number of observations in the given set of data. It is denoted by n .

Step 3. If n is odd, the median equals the $[(n+1)/2]^{\text{th}}$ observation.

Step 4. If n is even, then the median is given by the mean of $(n/2)^{\text{th}}$ observation and $[(n/2)+1]^{\text{th}}$ observation.

For grouped data:

Step 1. Make a table with 3 columns. First column for the class interval, second column for frequency, f , and the third column for cumulative frequency, cf .

Step 2. Write the class intervals and the corresponding frequency in the respective columns.

Step 3. Write the **cumulative frequency** in the column cf . It is done by adding the frequency in each step.

Step 4. Find the sum of frequencies, $\sum f$. It will be the same as the last number in the cumulative frequency column.

Step 5. Find $n/2$. Then find the class whose cumulative frequency is greater than and nearest to $n/2$. This is the median class.

Step 6. Now we use the formula
$$\text{Median} = l + \left(\frac{\frac{n}{2} - cf}{f} \right) \times h$$

l = lower limit of median class

n = no. of observations

cf denotes cumulative frequency of the class preceding the median class

f = frequency of median class

h = class size (assuming classes are of equal size)

Formula

For ungrouped data:

Median = $[(n+1)/2]^{\text{th}}$ observation, if n is odd.

Median = mean of $(n/2)^{\text{th}}$ observation and $[(n/2)+1]^{\text{th}}$ observation, if n is even.

For grouped data:

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

Solved Example

Example: The median of a set of 9 distinct observations is 20.5. If each of the largest 4 observations of the set is increased by 2, then the median of the new set

- (A) Is increased by 2
- (B) Is decreased by 2
- (C) Is two times the original median
- (D) Remains the same as that of the original set

Solution:

Given $n = 9$

Median = 20.5

Median term = $[(n+1)/2]^{\text{th}}$ term

= $[(9+1)/2]^{\text{th}}$ term

= $10/2^{\text{th}}$ term

= 5^{th} term

Frequently Asked Questions – FAQs

- What is the Median? Give Example.
- A median is the center value of a given list of observations when arranged in an order.

For example, a list of observations is 33, 55, 77, 22, 11.

Arranging in ascending order, we get:

11, 22, 33, 55, 77

Hence, the median is 33.

- What is the median of two numbers?
- If the number of given set of observations is 2, then we have to apply the formula of median for even number of observations, i.e.
Median = $[(n/2)\text{th term} + \{(n/2)+1\}\text{th term}]/2$
Example: Median of 15 and 20 is: $[(15)+(20)]/2 = 35/2 = 17.5$
- What is the median of 10 number of observations?
- The median of 10 numbers of observations is: $(5\text{th term} + 6\text{th term})/2$
- What is the median of odd numbers of observations?
- The formula to find median of odd number of observations is:
Median = $[(n+1)\text{th term}]/2$
Where n is the number of observations.

What is the difference between mean and median?

Median is defined as the center value of ordered list of values.

Mean is the ratio of sum of list of values and number of values, order of values does not matter.

Merits and Demerits of Median

Advantages or Merits of Median

Simple to Understand and Easy to Calculate

The biggest advantage of the median is that very simple to understand as you have to arrange any data first in ascending form and then find out the middle value from that arrangement and hence any individual with basic mathematics knowledge can learn this method and find out the median value of any data. Besides calculation part is also very easy unlike mean where one has to add all numbers and then divide it median requires no addition or division as one has to simply arrange the numbers in ascending order and find the middle number. Not Affected by Extreme Values

Another advantage of the median is that it does not get affected by an extreme value which is the case with mean. Hence for example, if in the above data if one

replaces \$10000 with \$100000 the also median will come to \$6000 whereas mean would have increased from \$6000 to \$24000 due to one extreme value.

Represent Data Graphically

Median can be used to represent the data graphically which enables the user to understand the data easily as 1 visual speaks thousands of words.

Disadvantages or Demerits of Median

Large Number of Items

This method is not practical when there are thousands of items in the data because one has to arrange that data first in ascending order and then find the median value which naturally will take lot of time and effort of an individual and hence when the data is huge than median method is not feasible as it will be time-consuming and impractical.

Problem with Even Observations

In case of median if there are odd numbers of observations that one gets the exact median which is the case with above example but if there were even numbers of observations than one had to average the 2 numbers which are in middle and one would have got an approximate figure of the median.

Does not take into Account all Items

In case of median the middle value is given the most significance and therefore other values are not taken into account or given significance which again is a major limitation of some data set or observations.

Percentage or Fraction

If data is in percentage or fraction than one will find it difficult to calculate median of those observations and hence when it comes to fraction or percentage median may not be the ideal method for calculation.

As one can see from the above that median has pros, as well as cons and any individual or company using it, should be aware of pros as well as cons before reaching to any conclusion about the results obtained using median.

Mode

In statistics, the **mode** is the value which is repeatedly occurring in a given set. We can also say that the value or number in a data set, which has a high

frequency or appears more frequently is called mode or **modal value**. It is one of the three measures of central tendency, apart from mean and median.

For example, mode of the set {3, 7, 8, 8, 9}, is 8. Therefore, for a finite number of observations, we can easily find the mode. A set of values may have one mode or more than one mode or no mode at all.

Table of Contents:

- Definition
- Mode Math
- Formula
- Finding Mode
- Example
- Grouped Data
- Mean Median Mode
- FAQs

Definition

A mode is defined as the value that has a higher frequency in a given set of values. It is the value that appears the most number of times.

Example: In the given set of data: 2, 4, 5, 5, 6, 7, the mode of the data set is 5 since it has appeared in the set twice.

Statistics deals with the presentation, collection and analysis of data and information for a particular purpose. To represent this data, we use tables, graphs, pie-charts, **bar graphs**, pictorial representation and so on. After the proper organization of the data, it must be further analyzed to infer some useful information from it.

For this purpose, frequently in statistics, we tend to represent a set of data by a representative value which would roughly define the entire collection of data. This representative value is known as the measure of central tendency. By the name itself, it suggests that it is a value around which the data is centred. These measures of central tendency allow us to create a statistical summary of the vast organized data. One such measure of central tendency is the mode of data.

Example: In the given set of data: 2, 4, 5, 5, 6, 7, the mode of the data set is 5 since it has appeared in the set twice.

Bimodal, Trimodal & Multimodal (More than one mode)

- When there are two modes in a data set, then the set is called **bimodal**
For example, The mode of Set A = {2,2,2,3,4,4,5,5,5} is 2 and 5, because both 2 and 5 is repeated three times in the given set.
- When there are three modes in a data set, then the set is called **trimodal**
For example, the mode of set A = {2,2,2,3,4,4,5,5,5,7,8,8,8} is 2, 5 and 8
- When there are four or more modes in a data set, then the set is called **multimodal**

Mode in Maths

The value occurring most frequently in a set of observations is its mode. In other words, the mode of data is the observation having the highest frequency in a set of data. There is a possibility that there exists more than one observation having the same frequency, i.e. a data set could have more than one mode. In such a case, the set of data is said to be multimodal.

Let us look into an example to get a better insight.

Example: The following table represents the number of wickets taken by a bowler in 10 matches. Find the mode of the given set of data.

Match No.	1	2	3	4	5	6	7	8	9	10
No. of Wickets	2	1	1	3	2	3	2	2	4	1

It can be seen that 2 wickets were taken by the bowler frequently in different matches. Hence, the mode of the given data is 2.

Mode Formula

In the case of grouped frequency distribution, calculation of mode just by looking into the frequency is not possible. To determine the mode of data in such cases we calculate the modal class. Mode lies inside the modal class. The mode of data is given by the formula:

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

Where,

l = lower limit of the modal class

h = size of the class interval

f_1 = frequency of the modal class

f_0 = frequency of the class preceding the modal class

f_2 = frequency of the class succeeding the modal class

Let us take an example to understand this clearly.

Finding the Mode

Let us learn here how to find the mode of a given data with the help of examples.

Example 1: Find the mode of the given data set: 3, 3, 6, 9, 15, 15, 15, 27, 27, 37, 48.

Solution: In the following list of numbers,

3, 3, 6, 9, 15, 15, 15, 27, 27, 37, 48

15 is the mode since it is appearing more number of times in the set compared to other numbers.

Example 2: Find the mode of 4, 4, 4, 9, 15, 15, 15, 27, 37, 48 data set.

Solution: Given: 4, 4, 4, 9, 15, 15, 15, 27, 37, 48 is the data set.

As we know, a data set or set of values can have more than one mode if more than one value occurs with equal frequency and number of time compared to the other values in the set.

Hence, here both the number 4 and 15 are modes of the set.

Example 3: Find the mode of 3, 6, 9, 16, 27, 37, 48.

Solution: If no value or number in a data set appears more than once, then the set has no mode.

Hence, for set 3, 6, 9, 16, 27, 37, 48, there is no mode available.

Mode of Grouped Data

Example 4: In a class of 30 students marks obtained by students in mathematics out of 50 is tabulated as below. Calculate the mode of data given.

Marks Obtained	Number of student
10-20	5
20-30	12
30-40	8
40-50	5

Solution:

The maximum class frequency is 12 and the class interval corresponding to this frequency is 20 – 30. Thus, the modal class is 20 – 30.

Lower limit of the modal class (l) = 20

Size of the class interval (h) = 10

Frequency of the modal class (f_1) = 12

Frequency of the class preceding the modal class (f_0) = 5

Frequency of the class succeeding the modal class (f_2) = 8

Substituting these values in the formula we get;

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) \times h = 20 + \left(\frac{12 - 5}{2 \times 12 - 5 - 8} \right) \times 10 = 26.364$$

Mean Median Mode Comparison

Mean	Median	Mode
<p>Mean is the average value that is equal to the ratio of sum of values in a data set and total number of values.</p> <p>Mean = Sum of observations/Number of observations</p>	<p>Median is the central value of given set of values when arranged in an order.</p>	<p>Mode is the most repetitive value of a given set of values.</p>
<p>For example, if we have set of values = 2,2,3,4,5, then;</p>		
<p>Mean = $(2+2+3+4+5)/5 = 3.2$</p>	<p>Median = 3</p>	<p>Mode = 2</p>

Merits of Mode:

- i. It is simple and easily understood.
- ii. Mode is not affected by the values of extreme items provided they follow to the natural law relating to extremes.
- iii. For determination of mode all values in the series are not considered.

Demerits of Mode:

- i. As mode is not based on all observations of a series, therefore, it is not rigidly defined.
- ii. Mode is not capable of further mathematical treatment.
- iii. Mode may be unrepresentative in many cases and it may be impossible to set a definite value of mode as in a set of observations 2 or 3 or more modal values may occur.

Frequently Asked Questions – FAQs

What is mode in statistics?

A mode, in statistics, is defined as the value that has higher frequency in a given set of values. It is the value that appears the most number of times.

How to find mode for given set of values?

If we have a set of values equal to 33,44,55,55,66. Then the most repeated value in the given set is 55. Therefore, mode of the given set is 55.

Can there be two modes in a given set of data?

Yes, there can be two modes in a given set of data. Such values are called bimodal.

What is trimodal and multimodal mode?

If there are three modes in a data set, then it is called trimodal and if there are four or more than four modes then it is called multimodal mode.

What is no mode condition?

If the given set of observations do not have any value that is repeated in the set, more than once, then it is said to be no mode.

[Sources :<https://www.youtube.com/watch?v=EtXR1wMCzgY>]

Range

In statistics, range is defined simply as the difference between the maximum and minimum observations. It is intuitively obvious why we define range in statistics this way - range should suggest how diversely spread out the values are, and by computing the difference between the maximum and minimum values, we can get an estimate of the spread of the data.

For example, suppose an experiment involves finding out the weight of lab rats and the values in grams are 320, 367, 423, 471 and 480. In this case, the range is simply computed as $480 - 320 = 160$ grams.

Limitations of Range

Range is quite a useful indication of how spread out the data is, but it has some serious limitations. This is because sometimes data can have outliers that are widely off the other data points. In these cases, the range might not give a true indication of the spread of data.

Average Deviation

The average deviation is a part of several indices of variability that is used by statisticians to characterize the dispersion among the measures in a given population. The average deviation of a set of scores is calculated by computing the mean and then the specific distance between each score and that mean without regard to whether the score is above or below the mean. It is also called

an average absolute deviation. Below mentioned is the formula to calculate the average deviation.

$$\text{Average Deviation} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

Where,

x represents the observation.

\bar{x} represents the mean.

n represents the number of observations.

Solved Examples

Question 1: Calculate the average deviation for the given data: 4, 6, 8, 10, 12, 14.

Solution:

Given:

$$n = 6$$

First let's find the mean by using the formula,

$$\bar{x} = \frac{4+6+8+10+12+14}{6}$$

$$\bar{x} = 9$$

$$\text{Average Deviation} = \frac{|4-9| + |6-9| + |8-9| + |10-9| + |12-9| + |14-9|}{6}$$

$$\text{Average Deviation} = \frac{5+3+1+1+3+5}{6}$$

$$\text{Average Deviation} = 3$$

Question 2: Find the average deviation for the following set of observations.

11, 6, 6, 12, 12, 7, 7, 9

Solution:

Given set of observations:

11, 6, 6, 12, 12, 7, 7, 9

Sum of observations = $11 + 6 + 6 + 12 + 12 + 7 + 7 + 9 = 70$

Number of observations = $n = 8$

Mean = Sum of observations/Total number of observations

$$\bar{x} = 70/8 = 8.75$$

$$\begin{aligned}\text{Average deviation} &= |11-8.75| + |6-8.75| + |6-8.75| + |12-8.75| + |12-8.75| + |7-8.75| + |7-8.75| + |9-8.75| \\ &= 2.25 + 2.75 + 2.75 + 3.25 + 3.25 + 1.75 + 1.75 + 0.25 \\ &= 18 \\ &= 18/8 \\ &= 2.25\end{aligned}$$

Merits and Limitation of Average Deviation

Merits:

- (i) The outstanding advantage of the average deviation is its relative simplicity. It is simply to understand and easy to compute. Any one familiar with the concept of the average can readily appreciate the meaning of the average deviation. If a situation requires a measure of dispersion that will be presented to the general public or any group not thoroughly grounded in statistics, the average deviation is very useful.
- (ii) It is based on each and every item of the data. Consequently, change in the value of any item would change the value of mean deviation.
- (iii) Mean deviation is less affected by the values of extreme items than the standard deviation.
- (iv) Since deviations are taken from a central value, comparison about formation of different distributions can easily be made.

Limitations:

- (i) The greatest drawback of this method is that algebraic signs are ignored while taking the deviations of the items. For example, if from twenty, fifty is deducted we write 30 and not -30. This is mathematically wrong and makes the method non – algebraic. If the signs of the deviations are not ignored the net sum of the deviations will be zero if the reference point is the mean or approximately zero if the references point is median.
- (ii) This method may not give us very accurate results. The reason is that mean deviation gives us best results when deviations are taken from median. But median is not a satisfactory measure when the degree of variability in a series is very high. And if we compute mean deviation from mean that is also not desirable because the sum of the deviations from mean (ignoring signs) is greater than the sum of the deviations from median (ignoring signs). If mean deviation is computed from mode that is also not scientific because the value of mode cannot always be determined.
- (iii) It is not capable of further algebraic treatment.
- (iv) It is rarely used in sociological studies.

Quartile Deviation

Quartile deviation is one of the measures of dispersion. Before getting into a deeper understanding, let's recall quartiles and how we can define them. Quartiles are the values that divide a list of numerical data into three-quarters, such as Q_1 , Q_2 and Q_3 . The middle part of the three quarters measures the central point of distribution and shows the data values near the midpoint (or the central value; this is referred to as the median). The lower part of the quarters indicates just half the information set, which comes under the median, and the upper part shows the remaining half, which falls above the median. Thus, the quartiles represent the distribution or dispersion of the given data set.

Quartile Deviation in Statistics can be defined as the statistic that measures the dispersion. Here, the Dispersion is the state of getting dispersed or spread. Statistical dispersion means the extent to which numerical data is likely to vary about an average value. In other words, dispersion helps to understand the distribution of the data.

Quartile Deviation Definition

The Quartile Deviation can be defined mathematically as half of the difference between the upper and lower quartile. Here, quartile deviation can be represented as QD; Q_3 denotes the upper quartile and Q_1 indicates the lower quartile.

Quartile Deviation is also known as the Semi Interquartile range.

Quartile Deviation Formula

Suppose Q_1 is the lower quartile, Q_2 is the median, and Q_3 is the upper quartile for the given data set, then its quartile deviation can be calculated using the following formula.

$$QD = (Q_3 - Q_1)/2$$

In the next section, you will learn how to calculate these quartiles for both ungrouped and grouped data separately.

Quartile Deviation for Ungrouped Data

For an ungrouped data, quartiles can be obtained using the following formulas,

$$Q_1 = [(n+1)/4]\text{th item}$$

$$Q_2 = [(n+1)/2]\text{th item}$$

$$Q_3 = [3(n+1)/4]\text{th item}$$

Where n represents the total number of observations in the given data set.

Also, Q_2 is the median of the given data set, Q_1 is the median of the lower half of the data set and Q_3 is the median of the upper half of the data set.

Before, estimating the quartiles, we have to arrange the given data values in ascending order. If the value of n is even, we can follow the similar procedure of finding the median.

Click here to get the [quartiles calculator](#) for ungrouped data.

Quartile Deviation for Grouped Data

For a grouped data, we can find the quartiles using the formula,

$$Q_r = l_1 + \frac{r\left(\frac{N}{4}\right) - c}{f} (l_2 - l_1)$$

Here,

Q_r = the r th quartile

l_1 = the lower limit of the quartile class

l_2 = the upper limit of the quartile class

f = the frequency of the quartile class

c = the cumulative frequency of the class preceding the quartile class

N = Number of observations in the given data set

Quartile Deviation Example

Let's understand the quartile deviation of ungrouped and grouped data with the help of examples given below.

Example 1:

Find the quartiles and quartile deviation of the following data:

17, 2, 7, 27, 15, 5, 14, 8, 10, 24, 48, 10, 8, 7, 18, 28

Solution:

Given data:

17, 2, 7, 27, 15, 5, 14, 8, 10, 24, 48, 10, 8, 7, 18, 28

Ascending order of the given data is:

2, 5, 7, 7, 8, 8, 10, 10, 14, 15, 17, 18, 24, 27, 28, 48

Number of data values = $n = 16$

Q_2 = Median of the given data set

n is even, median = $(1/2) [(n/2)\text{th observation and } (n/2 + 1)\text{th observation}]$

= $(1/2)[8\text{th observation} + 9\text{th observation}]$

= $(10 + 14)/2$

= $24/2$

= 12

$Q_2 = 12$

Now, lower half of the data is:

2, 5, 7, 7, 8, 8, 10, 10 (even number of observations)

Q_1 = Median of lower half of the data

= $(1/2)[4\text{th observation} + 5\text{th observation}]$

= $(7 + 8)/2$

= $15/2$

$$= 7.5$$

Also, the upper half of the data is:

14, 15, 17, 18, 24, 27, 28, 48 (even number of observations)

Q_3 = Median of upper half of the data

$$= (1/2)[4\text{th observation} + 5\text{th observation}]$$

$$= (18 + 24)/2$$

$$= 42/2$$

$$= 21$$

$$\text{Quartile deviation} = (Q_3 - Q_1)/2$$

$$= (21 - 7.5)/2$$

$$= 13.5/2$$

$$= 6.75$$

Therefore, the quartile deviation for the given data set is 6.75.

Example 2:

Calculate the quartile deviation for the following distribution.

Class	0- 10	10- 20	20- 30	30- 40	40- 50	50- 60	60- 70	70- 80	80- 90	90- 100
Frequency	5	3	4	3	3	4	7	9	7	8

Solution:

Let us calculate the cumulative frequency for the given distribution of data.

Class	Frequency	Cumulative Frequency
0 – 10	5	5
10 – 20	3	5 + 3 = 8
20 – 30	4	8 + 4 = 12
30 – 40	3	12 + 3 = 15
40 – 50	3	15 + 3 = 18
50 – 60	4	18 + 4 = 22
60 – 70	7	22 + 7 = 29
70 – 80	9	29 + 9 = 38
80 – 90	7	38 + 7 = 45
90 – 100	8	45 + 8 = 53

Here, N = 53

We know that,

$$Q_r = l_1 + \frac{r\left(\frac{N}{4}\right) - c}{f} (l_2 - l_1)$$

Finding Q_1 :

$$r = 1$$

$$N/4 = 53/4 = 13.25$$

Thus, Q_1 lies in the interval 30 – 40.

In this case, quartile class = 30 – 40

l_1 = the lower limit of the quartile class = 30

l_2 = the upper limit of the quartile class = 40

f = the frequency of the quartile class = 3

c = the cumulative frequency of the class preceding the quartile class = 12

Now, by substituting these values in the formula we get:

$$Q_1 = 30 + [(13.25 - 12)/3] \times (40 - 30)$$

$$= 30 + (1.25/3) \times 10$$

$$= 30 + (12.5/3)$$

$$= 30 + 4.167$$

$$= 34.167$$

Finding Q_3 :

$$r = 3$$

$$3N/4 = 3 \times 13.25 = 39.75$$

Thus, Q_3 lies in the interval 80 – 90.

In this case, quartile class = 80 – 90

l_1 = the lower limit of the quartile class = 80

l_2 = the upper limit of the quartile class = 90

f = the frequency of the quartile class = 7

c = the cumulative frequency of the class preceding the quartile class = 38

Now, by substituting these values in the formula we get:

$$Q_3 = 80 + [(39.75 - 38)/7] \times (90 - 80)$$

$$= 80 + (1.75/7) \times 10$$

$$= 80 + (17.5/7)$$

$$= 80 + 2.5$$

$$= 82.5$$

Finally, the quartile deviation = $(Q_3 - Q_1)/2$

$$QD = (82.5 - 34.167)/2$$

$$= 48.333/2$$

$$= 24.1665$$

Hence, the quartile deviation of the given distribution is 24.167 (approximately).

Merits and Demerits of Quartile Deviation

Merits

- It can be easily calculated and simply understood.
- It does not involve much mathematical difficulties.
- As it takes middle 50% terms hence it is a measure better than Range and percentile Range.
- It is not affected by extreme terms as 25% of upper and 25% of lower terms are left out.
- Quartile Deviation also provides a short cut method to calculate Standard Deviation using the formula $6 \text{ Q.D.} = 5 \text{ M.D.} = 4 \text{ S.D.}$
- In case we are to deal with the centre half of a series this is the best measure to use.

Demerits or Limitations

- As Q1 and Q3 are both positional measures hence are not capable of further algebraic treatment.
- Calculation are much more, but the result obtained is not of much importance.
- It is too much affected by fluctuations of samples.
- 50% terms play no role; first and last 25% items ignored may not give reliable result.
- If the values are irregular, then result is affected badly.
- We can't call it a measure of dispersion as it does not show the scatter-ness around any average.
- The value of quartile may be same for two or more series or Q.D. is not affected by the distribution of terms between Q1 and Q3 or outside these positions.

Standard Deviation

Standard Deviation is a measure which shows how much variation (such as spread, dispersion, spread,) from the mean exists. The standard deviation indicates a “typical” deviation from the mean. It is a popular measure of variability because it returns to the original units of measure of the data set. Like the variance, if the data points are close to the mean, there is a small variation whereas the data points are highly spread out from the mean, then it has a high variance. Standard deviation calculates the extent to which the values differ from the average. Standard Deviation, the most widely used measure of dispersion, is based on all values. Therefore a change in even one value affects the value of standard deviation. It is independent of origin but not of scale. It is also useful in certain advanced statistical problems.

Standard Deviation Formula

The population standard deviation formula is given as:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2}$$

Here,

σ = Population standard deviation

N = Number of observations in population

X_i = ith observation in the population

μ = Population mean

Similarly, the sample standard deviation formula is:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Here,

s = Sample standard deviation

n = Number of observations in sample

x_i = ith observation in the sample

\bar{x} = Sample mean

Variance Formula:

The population variance formula is given by:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2$$

The sample variance formula is given by:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

How is Standard Deviation calculated?

The formula for standard deviation makes use of three variables. The first variable is the value of each point within a data set, with a sum-number indicating each additional variable (x, x_1, x_2, x_3 , etc). The mean is applied to the values of the variable M and the number of data that is assigned to the variable n . Variance is the average of the values of squared differences from the [arithmetic mean](#).

To calculate the mean value, the values of the data elements have to be added together and the total is divided by the number of data entities that were involved.

Standard deviation, denoted by the symbol σ , describes the square root of the mean of the squares of all the values of a series derived from the arithmetic mean which is also called the root-mean-square deviation. 0 is the smallest value of standard deviation since it cannot be negative. When the elements in a series are more isolated from the mean, then the standard deviation is also large.

The statistical tool of standard deviation is the measures of dispersion that computes the erraticism of the dispersion among the data. For instance, mean, median and mode are the [measures of central tendency](#). Therefore, these are considered to be the central first order averages. The measures of dispersion that are mentioned directly over are averages of deviations that result from the average values, therefore these are called second-order averages.

Standard Deviation Example

Let's calculate the standard deviation for the number of gold coins on a ship run by pirates.

There are a total of 100 pirates on the ship. Statistically, it means that the population is 100. We use the standard deviation equation for the entire population if we know a number of gold coins every pirate has.

Statistically, let's consider a sample of 5 and here you can use the standard deviation equation for this sample population.

This means we have a sample size of 5 and in this case, we use the standard deviation equation for the sample of a population.

Consider the number of gold coins 5 pirates have; 4, 2, 5, 8, 6.

Mean:

$$\begin{aligned} \bar{x} &= \frac{\sum x_n}{n} \\ &= \frac{x_1 + x_2 + x_3 + x_4 + \dots + x_n}{n} \\ &= \frac{(4 + 2 + 5 + 6 + 8)}{5} \end{aligned}$$

$$= 5$$

$x_n - \bar{x}$ for every value of the sample:

$$x_1 - \bar{x} = 4 - 5 = -1$$

$$x_2 - \bar{x} = 2 - 5 = -3$$

$$x_3 - \bar{x} = 5 - 5 = 0$$

$$x_4 - \bar{x} = 8 - 5 = 3$$

$$x_5 - \bar{x} = 6 - 5 = 1$$

$$\sum (x_n - \bar{x})^2$$

$$= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_5 - \bar{x})^2$$

$$= (-1)^2 + (-3)^2 + 0^2 + 3^2 + 1^2$$

$$= 20$$

Standard deviation:

$$S.D = \sqrt{\frac{\sum (x_n - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{20}{4}}$$

$$= \sqrt{5}$$

$$= 2.236$$

Standard deviation of Grouped Data

In case of grouped data or grouped frequency distribution, the standard deviation can be found by considering the frequency of data values. This can be understood with the help of an example.

Question: Calculate the mean, variance and standard deviation for the following data:

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	27	10	7	5	4	2

Solution:

Class Interval	Frequency (f)	Mid Value (x_i)	fx_i	fx_i^2
0 – 10	27	5	135	675
10 – 20	10	15	150	2250
20 – 30	7	25	175	4375
30 – 40	5	35	175	6125
40 – 50	4	45	180	8100
50 – 60	2	55	110	6050
	$\Sigma f = 55$		$\Sigma fx_i = 925$	$\Sigma fx_i^2 = 27575$

$$N = \Sigma f = 55$$

$$\text{Mean} = (\Sigma fx_i)/N = 925/55 = 16.818$$

$$\text{Variance} = 1/(N - 1) [\Sigma fx_i^2 - 1/N(\Sigma fx_i)^2]$$

$$= 1/(55 - 1) [27575 - (1/55) (925)^2]$$

$$= (1/54) [27575 - 15556.8182]$$

$$= 222.559$$

$$\text{Standard deviation} = \sqrt{\text{variance}} = \sqrt{222.559} = 14.918$$

Practice Problems on Standard Deviation

1. Calculate the standard deviation of the following values:

5, 10, 25, 30, 50

2. Find the mean and standard deviation for the following data.

x	60	61	62	63	64	65	66	67	68
f	2	1	12	29	25	12	10	4	5

3. The diameters of circles (in mm) drawn in a design are given below:

Diameters	33 – 36	37 – 40	41 – 44	45 – 48	49 – 52
No.of circles	15	17	21	22	25

Calculate the standard deviation and mean diameter of the circles.

[Hint: First make the data continuous by making the classes as 32.5-36.5, 36.5-40.5, 40.5-44.5, 44.5 – 48.5, 48.5 – 52.5 and then proceed.]

Check out more problems on variance and standard deviation of grouped data and Statistics, register with BYJU’S – The Learning App to learn with ease.

Frequently Asked Questions – FAQs

How do you calculate the standard deviation?

The procedure to calculate the standard deviation is given below:

Step 1: Compute the mean for the given data set.

Step 2: Subtract the mean from each observation and calculate the square in each instance.

Step 3: Find the mean of those squared deviations.

Step 4: Finally, take the square root obtained mean to get the standard deviation.

What does Standard Deviation tell you?

Standard deviation tells us how far is the mean from each observation in the given data set. In other words, it shows the typical deviation from the mean.

What is the standard deviation and variance?

Standard deviation indicates how the spread of observations of a data set is from the mean by studying at the variance's square root. The variance estimates the average degree to which each observation differs from the mean of all observations of the data.

Merits and Limitations of Standard Deviation

Merits:

- (i) The standard deviation is the best measure of variation because of its mathematical characteristics. It is based on every item of the distribution. Also it is amenable to algebraic treatment and is less affected by fluctuations of sampling than most other measures of dispersion.
- (ii) It is possible to calculate the combined standard deviation of two or more groups. This is not possible with any other measure.
- (iii) For comparing the variability of two or more distribution coefficient of variation is considered to be most appropriate and this is based on mean standard deviation.
- (iv) Standard deviation is most prominently used in further statistical work. For example, in computing skewness, correlation, etc., use is made of standard deviation. It is keynote in sampling and provides a unit of measurement for the normal distribution.

Limitations:

- (i) As compared to other measures it is difficult to compute. However, it does not reduce the importance of this measure because of high degree of accuracy of results it gives.
- (ii) It gives more weight to extreme items and less to those which are near the mean. It is because of the fact that the squares of the deviations which are big in size would be proportionately greater than the squares of those deviations which are comparatively small. The deviations 2 and 8 are in the ratio of 1:4 but their squares, i.e., 4 and 64 would be in the ratio of 1:16.

[Sources : <https://www.youtube.com/watch?v=eqDnC6CVhvw>]

[Sources : <https://www.youtube.com/watch?v=N0rKb3OFybU>]

[Sources: <https://www.youtube.com/watch?v=KA1pGreZU5I>]