BASIC STATISTICS

INTRODUCTION

Statistics is an important branch of mathematics that is widely used in a variety of traditional disciplines like economics, commerce, research, surveys, etc. In this present digital age, emerging technologies like data science and machine learning have boomed up. These technologies are also centered around statistics. After all, statistics is all about the data collection, data interpretation, and presentation of data. Basically, statistics provide insights into the data.

In statistics, the central tendency is the descriptive summary of a data set. Through the single value from the dataset, it reflects the centre of the data distribution. Moreover, it does not provide information regarding individual data from the dataset, where it gives a summary of the dataset. Generally, the central tendency of a dataset can be defined using some of the measures in statistics.

The idea of central tendency is that there may be one single value that can possibly describe the data to the best extent. Measures of central tendency or averages reduce the large number of observations to one figure. Actually the measures of central tendency describe the tendency of items of group around the middle in a frequency distribution of numerical values.

Definition

The **central tendency (average)** is stated as the statistical measure that represents the single value of the entire distribution or a dataset. The purpose for computing an average value for a set of observations is to obtain a single value which is representative of all the items and which the mind can grasp simply and quickly. The single value is the point of location around which the individual items cluster.

Characteristics of a good average

(1) It should be rigidly defined.
(2) It should be based on all the observations.
(3) It should be capable of further algebraic treatment.
(4) It should not be affected by fluctuations of sampling.
(5) It should be easy to compute.

MEASURES OF CENTRAL TENDENCY

The following are the five measures of average or central tendency that are in common use :

(i) Arithmetic average or arithmetic mean or simple mean
(ii) Median
(iii) Mode
(iv) Geometric mean
(v) Harmonic mean

Arithmetic mean, Geometric mean and Harmonic means are usually called Mathematical averages while Mode and Median are called Positional averages.

ARITHMETIC MEAN

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.

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. $\sum(xi - \bar{x}) = 0$.
- 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.

  i.e. $\sum(xi - \bar{x})^2 < \sum(xi - A)^2$
- 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.

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.
- It 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.

CALCULATION OF ARITHMETIC MEAN

Following methods used to compute the Arithmetic Mean for three types of series:

I. Individual Data Series
II. Discrete Data Series
III. Continuous Data Series

I. Individual Data Series

If any data set consisting of the values $x_1, x_2, x_3, \ldots, x_n$ then the arithmetic mean $\bar{x}$ is defined as:

$\bar{x}$ = (Sum of all observations)/ (Total number of observation)

$$= \frac{x_1+x_2+x_3+\cdots+x_n}{n}$$

$$= \frac{1}{n}\sum_{i=1}^{n} xi$$

Examples

Q. 1: The marks obtained by 6 students in a class test are 20, 22, 24, 26, 28, and 30. Find the mean.

Solution:

$\bar{x}$ = (20+22+24+26+28+30)/6

= 25

Therefore, mean = 25

Q. 2: If the arithmetic mean of 14 observations 26, 12, 14, 15, x, 17, 9, 11, 18, 16, 28, 20, 22, 8 is 17. Find the missing observation.

Solution:

Given 14 observations are: 26, 12, 14, 15, x, 17, 9, 11, 18, 16, 28, 20, 22, 8

Arithmetic mean = 17

We know that Arithmetic mean = Sum of observations/Total number of observations

17 = (216 + x)/14

17 * 14 = 216 + x

216 + x = 238

x = 238 – 216

x = 22

Therefore, the missing observation is 22.

Q 3: Find the arithmetic means of the runs scored by Virat Kohli in the last few innings.

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

Solution:

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.

Q 4: Calculate Arithmetic Mean for the following individual data:

Items 14 36 45 70 105

II. Discrete Data Series

Discrete series means where frequencies of a variable are given but the variable is without class intervals.

Here the mean can be found by Two Methods.

(i) Direct Method

(ii) Short Cut Method

(i) Direct Method

If any data set consisting of the values $x_1, x_2, x_3, \ldots, x_n$ with corresponding frequencies $f_1, f_2, f3, \ldots, f_n$, then the arithmetic mean $\bar{x}$ is defined as:

$$\bar{x} = \frac{x_1f_1 + x_2f_2 + \cdots \ldots + x_nf_n}{N}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_if_i}{N}$$

where $N = f_1 + f_2 + \ldots\ldots\ldots + f_n = \sum_{i=1}^{n} f_i$

(ii) Short Cut Method

This method is used to make the calculations simpler. Consider any data set consisting of the values $x_1, x_2, x_3, \ldots, x_n$ with corresponding frequencies $f_1, f_2, f3, \ldots, f_n$. Let A be any assumed mean (or any assumed number), then the arithmetic mean $\bar{x}$ is defined as:

$$\bar{x} = A + \frac{\sum_{i=1}^{n} f_id_i}{N}$$

where di the deviation of the arithmetic mean and can be calculated by $d_i=(x_i-A)$; and

$N = f_1 + f_2 + \ldots\ldots\ldots\ldots + f_n$.

<u>Examples</u>

Q1. Calculate the arithmetic mean from the following data using direct method.

X	25	35	45	55	65
Frequency	8	26	30	20	16

Solution

xi	Frequency (fi)	xifi
25	8	200

35	26	910
45	30	1350
55	20	1100
65	16	1040
	N=100	Σxifi = 4600

Arithmetic Mean, $\bar{x}$ = Σxifi / N
= 4600 / 100
= 46

Q2. Calculate the arithmetic mean from the following data using short cut method.

X	25	35	45	55	65
Frequency	8	26	30	20	16

Solution

Xi	Frequency (fi)	di=xi-A (di=xi-45)	fi*di
25	8	-20	-160
35	26	-10	-260
45	30	0	0
55	20	10	200
65	16	20	320
	N=100		Σfidi=100

Arithmetic Mean, $\bar{x}$ =A + Σfidi/N
= 45 + (100 / 100)
= 45 + 1 = 46

Q3. Calculate the arithmetic mean from the following data.

III. Continuous Data Series

Continuous series is a series where the variables are in the class interval, and each has corresponding frequency. The arithmetic mean is calculated in continuous series in the same way as it is calculated in discrete series. The only difference is that in continuous series, the midpoints (x) or mid-values of all classes are determined in the beginning. The mid value x can be calculated by

x=(upper limit + lower limit)/2

Upon taking the mid-values, the continuous series takes the form of discrete series.
Here the mean can be found by Three Methods.

(i) Direct Method
(ii) Short Cut Method
(iii) Step Deviation Method

(i) Direct Method

Consider the mid values $x_1, x_2, x_3, \ldots, x_n$ of a continuous series with corresponding frequencies $f_1, f_2, f3, \ldots, f_n$, then the arithmetic mean $\bar{x}$ is defined as:

$$\bar{x} = \frac{x_1f_1+x_2f_2+\cdots\ldots+x_nf_n}{N}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} xifi}{N}$$

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

(ii) Short Cut Method

This method is used to make the calculations simpler. Consider the mid values x_1, x_2, x_3, …., x_n of a continuous series with corresponding frequencies f_1, f_2, f3, …., f_n. Let A be any assumed mean (or any assumed number), then the arithmetic mean $\bar{x}$ is defined as:

$$\bar{x} = A + \frac{\sum_{i=1}^{n} fidi}{N}$$

where di the deviation of the arithmetic mean and can be calculated by di=(xi-A)

(iii) Step Deviation Method

Consider the mid values x_1, x_2, x_3, …., x_n of a continuous series with corresponding frequencies f_1, f_2, f3, …., f_n. Let A be any assumed mean (or any assumed number) and c is the common factor (usually class interval), then the arithmetic mean $\bar{x}$ is defined as:

$$\bar{x} = A + \frac{\sum_{i=1}^{n} fiui}{N} \times \underline{c}$$

where ui can be calculated by ui=(xi-A)/c

<u>Examples</u>

Q1. Calculate the arithmetic mean from the following data using direct method.

Class	20-30	30-40	40-50	50-60	60-70
Frequency	8	26	30	20	16

Solution

Class	Frequency (fi)	Mid value (xi)	xi*fi
20-30	8	25	200
30-40	26	35	910
40-50	30	45	1350
50-60	20	55	1100
60-70	16	65	1040
	N= 100		Σxifi= 4600

Arithmetic Mean, $\bar{x}$ = Σxifi/N

= 4600 / 100

= 46

Q2. Calculate the arithmetic mean from the following data using shortcut method.

Class	20-30	30-40	40-50	50-60	60-70
Frequency	8	26	30	20	16

Solution

Class	Frequency (fi)	mid value (xi)	di=xi-A (di=xi-45)	fi*di
20-30	8	25	-20	-160
30-40	26	35	-10	-260
40-50	30	45	0	0

50-60	20	55	10	200
60-70	16	65	20	320
	N=100			Σxidi= 100

Arithmetic Mean, $\bar{x}$ = A+ $\frac{\sum_{i=1}^{n} \text{fidi}}{N}$

Therefore, $\bar{x}$ = 45+100/100

= 46

Q3. Calculate the arithmetic mean from the following data using step deviation method.

Class	20-30	30-40	40-50	50-60	60-70
Frequency	8	26	30	20	16

Solution

Class	Frequency (fi)	mid value (xi)	ui=(xi-A)/c [ui=(xi-45)/10]	fi*ui
20-30	8	25	-2	-16
30-40	26	35	-1	-26
40-50	30	45	0	0
50-60	20	55	1	20
60-70	16	65	2	32
	N=100			Σxiui= 10

Arithmetic Mean, $\bar{x}$ = A+ $\frac{\sum_{i=1}^{n} \text{fiui}}{N}$ x c

Here A=45, c=10

Therefore, $\bar{x}$ = 45+(10/100)x10

= 46

Q4. Find the arithmetic mean for income from the following continuous series.

Income(in thousands) :	10-20	20-30	30-40	40-50	50-60	60-70
No of person :	4	7	16	20	15	8

Q5. Calculate mean from following data using short-cut method.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No of Students	3	6	7	5	22	17	20	16	13	5

MEDIAN

Median is one of the measures of central tendency. It is a positional average. Median may be defined as 'the middlemost or central value of the series when the values are arranged in ascending or descending order of magnitude'.

It can be also defined as "The median is the value of the variable which divides the group into two equal parts one part comprising all values greater the median and the other all values less than the median".

For example, the marks obtained, by seven students in a paper of Statistics are

15, 20, 23, 32, 34, 39, 48

the maximum marks being 50, then the median is 32 since it is the value of the 4th term, which is situated such that the marks of 1st, 2nd and 3rd students are less than this value and those of 5th, 6th and 7th students are greater than this value.

If there is an *odd number of an item*, then the median is found out by taking the middle most items of the series only after arranging the data in order of magnitude.

If there is an *even number of items*, then the median is half-way between the two middle ones and it is found by taking the average of these two items.

Properties of Median

- The median can be calculated graphically, while the mean cannot.
- The sum of the absolute deviations taken from the median is less than the sum of absolute deviations from any other observations in the data.
- The median is not much affected by extreme values and would therefore be a better representative as a centre or average.

COMPUTATION OF MEDIAN

Following methods used to compute the Median for three types of series:

I. Individual Data Series
II. Discrete Data Series
III. Continuous Data Series

I. Median in individual series.

Let n be the number of values in the dataset. First of all we write the values of the in ascending or descending order of magnitudes

Here two cases arise:

Case 1. If n is odd then value of $(n+1)/2^{th}$ term gives the median.

Case 2. If n is even then there are two central terms i.e., $n/2^{th}$ and $(n/2)+1^{th}$, then the mean of these two values gives the median.

Examples

Q1. According to the census of 2011, following are the population figure, in thousands, of 10 cities :

1400, 1250, 1670, 1800, 700, 650, 570, 488, 2100, 1700.

Find the median.

Solution

Arranging the terms in ascending order.

488, 570, 650, 700, 1250, 1400, 1670, 1700, 1800, 2100.

Here n=10, therefore the median is the mean of the measure of the 5th and 6th terms.

Here 5th term is 1250 and 6th term is 1400.

Median = (1250+1400)/2 Thousands

= 1325 Thousands

Q2. The number of runs scored by 11 players of a cricket team of a college are

5, 19, 42, 11, 50, 30, 21, 0, 52, 36, 27

Find the median?

Solution

Arranging the terms in ascending order.

0,5,11,19,21,27,30,36,42,50,52

Here, N=11

Median = (N+1)/2 th item

=6th item

= 27

Q3. The weights (in Kg) of 15 students are as follows:

31, 35, 27, 29, 32, 43, 37, 41, 34, 28, 36, 44, 45, 42, 30

Find the median. If the weight 44 Kg is replaced by 46 Kg and 27 Kg by 25 kg, find the median.

Solution:

Cumulative Frequency: (cf)

Cumulative frequency of each class is the sum of the frequency of the class and the frequencies of the pervious classes, ie adding the frequencies successively, so that the last cumulative frequency gives the total number of items.

II. Median in Discrete Series

Step1: Find cumulative frequencies.

Step2: Find $(\frac{N+1}{2})$

Step3: See in the cumulative frequencies the value just greater than $(\frac{N+1}{2})$

Step4: Then the corresponding value of x is median.

Examples

Q1. The following data pertains to the number of members in a family. Find median size of the family.

Number of members x	1	2	3	4	5	6	7	8	9	10	11	12
Frequency, f	1	3	5	6	10	13	9	5	3	2	2	1

Solution

x	f	cf
1	1	1
2	3	4
3	5	9
4	6	15
5	10	25

6	13	38
7	9	47
8	5	52
9	3	55
10	2	57
11	2	59
12	1	60
	N=60	

Size of (N+1/2) th item

Size of (61/2) th item

Size of 30.5 th item

The cumulative frequency just greater than 30.5 is 38, and the value of x corresponding to 38 is 6. Hence the median size is 6 members per family.

Q2. Calculate the median for the following data

No. of Students	6	4	16	7	8	2
Marks	20	9	25	50	40	80

III. Median in Continuous Series

The steps given below are followed for the calculation of median in continuous series.

Step1: Find cumulative frequencies.

Step2: Find $\left(\frac{N}{2}\right)$

Step3: See in the cumulative frequency the value greater than $\left(\frac{N}{2}\right)$, then the corresponding class interval is called the Median class. Then apply the formula

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

where L = Lower limit of the median class

m = cumulative frequency preceding the median class

c = width of the median class

f = frequency in the median class.

N=Total frequency.

Examples

Q1. For the frequency distribution of weights of Samsung ear-heads given in table below, calculate the median.

Weights of ear heads (in g)	No of ear heads (f)
60-80	22
80-100	38
100-120	45
120-140	35
140-160	20

Solution

Weights of ear heads (in g)	No of ear heads (f)	Cumulative frequency (m)
60-80	22	22
80-100	38	60
100-120	45	105
120-140	35	140
140-160	20	160
	N=160	

Here

N=160

N/2=160/2=80

So the median class is 100-120

L=100

f=45

c=20

m=60

$$\text{Median} = L + \frac{(\frac{N}{2}) - m}{f} \text{ x } c$$

= 100+[(80-60)/45]x20

= 108.89

Q2. Find the median for the following distribution:

Wages in Rs.	0-10	10-20	20-30	30-40	40-50
No. of workers	22	38	46	35	20

Solution

Wages in Rs.	No. of Workers f	Cumulative Frequencies (c.f.)
0-10	22	22
10-20	38	60
20-30	46	106

30-40	35	141
40-50	20	161
	N=161	

Here

N=161

N/2=161/2=80.5

Clearly 81st term is situated in the class 20-30. Thus 20-30 is the median class.

L=20

f=46

c=10

m=60

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

= 20+((80.5-60)/46)*10

= 24.46

Q3. Find the median of the following frequency distribution:

Marks	No. of students
Less than 10	15
Less than 20	35
Less than 30	60
Less than 40	84
Less than 50	106
Less than 60	120
Less than 70	125

Solution

Class (Marks)	Frequency f	Cumulative Frequency (C. F.)
0-10	15	15
10-20	20	35
20-30	25	60
30-40	24	84
40-50	22	106
50-60	14	120
60-70	5	125
	N=125	

Here N=125

N/2=125/2=62.5

Therefor Median class is 30-40

L=30

f=24

c=10

m=60

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

= 30+((62.5-60)/24)*10

= 31.04

Example:

Following data relates to daily wages of persons working in a factory. Compute the median daily wage.

Daily wages (in Rs):	55–60	50–55	45–50	40–45	35–40	30–35	25–30	20–25
Number of workers:	7	13	15	20	30	33	28	14

Calculate median from following data.

Class group	110-120	120-130	130-140	140-150	150-160	160-170	170-180	180-190	190-200
Frequency	6	25	48	72	116	60	38	22	3

MODE

Mode is one of the measures of central tendency. In statistics, mode or modal value is that observation which occurs at the maximum time or has the highest Frequency in the given set of data. A given set of data may have one or more than one mode. A set of numbers with one mode is unimodal, a set of numbers having two modes is bimodal, a set of numbers having three modes is trimodal, and any set of numbers having four or more than four modes is known as multimodal.

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.

Characteristics

- Mode is one of the measures of central tendency.
- Mode is that value in which f(x) is maximum.
- Mode is not affected by extreme values.
- The mode is not calculated on all observations in a data set.
- The value of the mode can be computed graphically whereas the value of the mean cannot be calculated graphically.
- We have the following relationship between the mean, median and the mode:

  Mode = 3*Median - 2*Mean
- The mode can be conveniently found even if the frequency distribution has class intervals of unequal magnitude.
- The mode can be used for qualitative data.

CALCULATION OF MODE

We're going to discuss methods to compute the Arithmetic Mode for three types of series:

- Individual Data Series
- Discrete Data Series
- Continuous Data Series

I. Calculation of Mode - Individual Data Series

The terms are arranged in any order. Ascending or Descending. If each term of the series is occurring once, then there is no mode, otherwise the value that occurs Maximum Times is known as Mode.

Example 1

Calculate Arithmetic Mode for the following individual data:

Items 14 36 45 36 105 36

Solution

Arranging the above data in ascending order

14, 36, 36, 36, 45, 105

The Arithmetic Mode of the given numbers is 36 as it is repeated maximum number of times, 3.

Example 2

Following are the monthly incomes of 10 employees of a company. Find out the mode monthly income:

Monthly Income (in Rs) 2000 2220 1800 2000 1600 2100 2000 2400 2000 1900

Solution:

In the above example, there are 4 employees of 2000 monthly income, this is the highest number of employees, so the Rs 2000 monthly income will be mode.

Example 3

The following series has the sizes of shoes that are worn by the customers. Find out the mode shoe size :

4, 3, 4, 5, 6, 7, 6, 4, 5, 7, 3, 7, 8, 7, 9, 6

The maximum frequency is 4, whose value is 7, so the mode size of the shoes = 7.

Example 4

Find the Mode of the following model size number of shoes.

Model size no. of shoes: 3, 4, 2, 1, 7, 6, 6, 7, 5, 6, 8, 9, 5.

II. Calculation of Mode - Discrete Data Series

When the values (or measures) of all the terms (or items) are given. In this case the mode is the value (or size) of the term (or item) which occurs most frequently.

Example

Find the mode from the following size of shoes

Size of shoes	1	2	3	4	5	6	7	8	9
Frequency	1	1	1	1	2	3	2	1	1

Here maximum frequency is 3 whose term value is 6. Hence the mode is modal size number 6.

Example

Find out the mode from the following term series :

Age(in years) :	9	10	11	12	13	14	15
No. of Boys :	2	3	12	9	7	4	2

It is clear from inspection that the frequency of age 11 has the highest number of boys (12). So the mode age = 11 years.

Example

Calculate Arithmetic Mode for the following discrete data:

Items	14	36	45	70	105	145
Frequency	2	5	1	3	12	0

Solution:

The Arithmetic Mode of the given numbers is 105 as the highest frequency, 12 is associated with 105.

III. Calculation of Mode - Continuous Data Series

In continuous frequency distribution the computation of mode is done by the following formula

$$\boxed{\text{Mode} = L + \frac{f1 - f0}{2f1 - f0 - f2} \times c}$$

L = lower limit of class,

f1 = frequency of modal class,

f0 =frequency of the class just preceding to the modal class,

f2 =frequency of the class just following of the modal class,

c =class interval

Example

Calculate the Arithmetic Mode from the following data:

Wages (in Rs.)	No. of workers
0-5	3
5-10	7
10-15	15
15-20	30
20-25	20
25-30	10
30-35	5

Solution

Here maximum frequency 30 lies in the class-interval 15-20. Therefore 15-20 is the modal class.

$$\text{Mode} = L + \frac{f1 - f0}{2f1 - f0 - f2} \times c$$

L = 15

f1 = 30

f0 = 15

f2 = 20

c = 5

$$\text{Mode} = 15 + \frac{30 - 15}{2x30 - 15 - 20} \times 5$$

= 15 + 3

= 18

Example 2: Compute the mode of the following distribution:

Class:	0-7	7-14	14-21	21-28	28-35	35-42	42-49
Frequency:	19	25	36	72	51	43	28

Solution.

Here maximum frequency 72 lies in the class-interval 21-28. Therefore 21-28 is the modal class.

$$\text{Mode} = L + \frac{f1 - f0}{2f1 - f0 - f2} \times c$$

L = 21

f1 = 72

f0 = 36

f2 = 51

c = 7

$$\text{Mode} = 21 + \frac{72 - 36}{2x72 - 36 - 51} \times 7$$

=21+(36/(144-36-51))*7

= 25.421

Example: Let us take the following frequency distribution:

Class intervals	Frequency
30-40	4
40-50	6
50-60	8

60-70	12
70-80	9
80-90	7
90-100	4

Calculate the mode in respect of this series.

Solution

Here maximum frequency 12 lies in the class-interval 60-70. Therefore 60-70 is the modal class.

Let Mode = L + $\frac{f1 - f0}{2f1 - f0 - f2}$ x c

L = 60

f1 = 12

f0 = 8

f2 = 9

c = 10

Mode = 60 + $\frac{12 - 8}{2x12 - 8 - 9}$ x 10

= 60+(4/(24-8-9))*10

= 65.714

Example

Calculate the value of modal worker family's monthly income from the following data:

Income per month (in 1000 Rs)	Cumulative Frequency
Less than 50	97
Less than 45	95
Less than 40	90
Less than 35	80
Less than 30	60
Less than 25	30
Less than 20	12
Less than 15	4

Solution

Income Group (in 1000 Rs)	Frequency
45–50	97 - 95 = 2
40–45	95 -90 = 5
35–40	90 -80 = 10
30–35	80 -60 = 20
25–30	60 -30 = 30
20–25	30 -12 = 18
15–20	12 -4 = 8
10–15	4

Here maximum frequency 30 lies in the class-interval 25–30. Therefore 25–30 is the modal class.

Let Mode = L + $\frac{f1 - f0}{2f1 - f0 - f2}$ x c

L = 25

f1 = 30

$f0 = 18$

$f2 = 20$

$c = 5$

$$\text{Mode} = 25 + \frac{30 - 18}{2x30 - 18 - 20} \text{ x } 5$$

$= 25+ (12/(60-18-20))*5$

Mode $= 27.727$

Example: Calculate mode from following data

marks :	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
students :	2	18	30	45	35	20	6	4

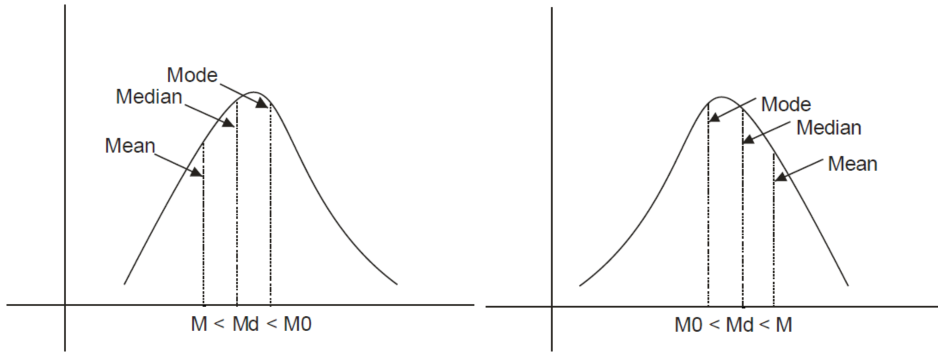
Empirical Relation among Mean, Median and Mode

For moderately asymmetrical distribution (or for asymmetrical curve), the relation

Mean – Mode = 3 (Mean - Median)

approximately holds. In such a case, first evaluate mean and median and then mode is determined by

Mode = 3 Median – 2 Mean



If in the asymmetrical curve the area on the left of mode is greater than area on the right then

Mean < Median < Mode, i. e., (M < Md < Mo)

If in the asymmetrical curve, the area on the left of mode is less than the area on the right then in this case

Mode < Median< Mean, i.e. (Mo < Md < M).

MEASURES OF DISPERSION

Dispersion is the state of getting dispersed or spread. Statistical dispersion means the extent to which a numerical data is likely to vary about an average value. In statistics, the measures of dispersion help to interpret the variability of data i.e. to know how much homogenous or heterogeneous the data is.

Definition

The degree to which numerical data tend to spread about an average value is called variation or dispersion of the data

Properties of measure of dispersion

- It should be easy to understand simple to calculate.
- It should be based on all observations.
- It should be capable of further algebraic treatment.
- It should be least affected by sampling fluctuations.
- It should not be unduly affected by extreme observations.

TYPES OF MEASURES OF DISPERSION

There are two main types of dispersion methods in statistics which are:

- Absolute Measure of Dispersion
- Relative Measure of Dispersion

Absolute Measure of Dispersion

These measures give us an idea about the amount of dispersion in a set of observations. They give the answers in the same units as the units of the original observations. When the observations are in kilograms, the absolute measure is also in kilograms. The absolute measures which are commonly used are:

- The Range
- The Quartile Deviation
- The Mean Deviation
- The Standard Deviation and Variance

Relative Measure of Dispersion

The relative measures of dispersion are used to compare the distribution of two or more data sets. These measures are free of the units in which the original data is measured. If the original data is in dollars or kilometers, we do not use these units with relative measures of dispersion. These measures are a sort of ratio and are called coefficients. Each absolute measure of dispersion can be converted into its relative measure. The relative measures of dispersion are:

- Coefficient of Range
- Coefficient of Quartile Deviation
- Coefficient of Mean Deviation

- Coefficient of Standard Deviation
- Coefficient of Variation (a special case of Coefficient of Standard Deviation)

RANGE

It is the simplest possible measure of dispersion. The range of a set of numbers (data) is the difference between the largest and the smallest numbers in the set i.e. values of the variable. If this difference is small then the series of numbers is supposed to be regular and if this difference is large then the series is supposed to be irregular.

Range is defined and given by the following function:

$$\boxed{\text{Range} = L - S}$$

where

L = Largest item

S = Smallest item

This is an absolute measure. The relative measure called as coefficient of range is given by

$$\boxed{\text{Coefficient of Range} = \frac{L-S}{L+S}}$$

<u>Example 1</u>

Rani took 7 math tests in one year. What is the range of her test scores and coeff. of range?

89 73 84 91 87 77 94

Solution:

Ordering the test scores from least to greatest, we get:

73 77 84 87 89 91 94

Range = Largest − Smallest

= 94−73

= 21

Largest + Smallest = 94+73

=167

Coefficient of Range $= \frac{L-S}{L+S}$

=21/167

= 0.1257

The Range of these test scores is 21 points and coeff. of range is 0.1257.

<u>Example 2</u>

Compute the range for the following observation

15 20 25 25 30 35

Solution

Range = Largest – Smallest

= 35-15

= 20

<u>Example 3</u>

Find out absolute and relative dispersion of range from the following observations:

Marks: 63, 68, 71.5, 83, 50, 27, 64, 38, 40

Solution

Absolute measure of Range = Largest – Smallest

= 83 – 27

= 56 marks

Largest + Smallest = 83 + 27

=110

Relative Measure $= \frac{L-S}{L+S}$

= 56/110

= 0.509

Example 4

The yearly income of a person (in thousands) for the last ten years in given below. Find the range and its coefficient.

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Income	40	30	80	100	80	90	120	110	130	150

Solution

From the data, L = 150, S = 30.

Range = L - S

=150 – 30

= 120 (in thousands)

Coefficient of Range $= \frac{L-S}{L+S}$

$= \frac{150-30}{150+30}$

= 120/180

= 0.667

Example 5

Find out absolute and relative measure of range in the following distribution:

Scores	Frequency
2	6
3	7
4	3
5	11
6	4
7	2
8	5
9	8
10	3

Solution

Absolute Measure of Range = L – S

= 10 – 2

= 8 score

Relative Measure $= \frac{L-S}{L+S}$

= 8/12

= 0.667

Example 6

Calculate Coefficient of range from the following distribution:

Marks	No. of Students
25—30	6
30—35	3
35—40	12
40—45	8

45—50	22
50—55	9
55—60	5

Solution: The above question can be calculated by:

(1) Limits Method: L = 60, S = 25

Coefficient of Range = L – S / L + S

= 60 – 25 / 60 + 25

= 0.411

(2) Mid-Value Method: L = 57.5, S = 27.5

Coefficient of Range = $\frac{L-S}{L+S}$

= (57.5 – 27.5) / (57.5 + 27.5)

= 0.352

<u>Problems to be solved</u>

1. The price of a share for a six-day week fluctuated as follows (Figures are in Rupees):

122, 200, 88, 145, 99, 164.

Calculate the range and coefficient of range.

2. From the following data calculate Range and coefficient of Range

Weight (Kg.)	5	8	10	12	25	30	38
No. of Students	2	3	8	10	9	3	2

MEAN DEVIATION

The mean deviation is the mean of the absolute deviations of the observations or values from a suitable average. This suitable average may be the mean, median or mode.

CALCULATION OF MEAN DEVIATION

The mean deviation of the data values can be easily calculated using the below procedure.

Step 1: Find the average value (mean/median/mode) for the given data values

Step 2: Now, subtract the average value form each of the data value given (Note: Ignore the minus symbol)

Step 3: Now, find the mean of those values obtained in step 2.

I. Formula of Mean Deviation from Mean

Formula to calculate mean deviation from mean for individual series, discrete series, and continuous series are as follows.

1) Individual Series: The formula to find the mean deviation from mean for an individual series is:

M.D about mean = $\frac{\sum |x_i - \bar{x}|}{n}$

x_i = Observations / Values

$\bar{x}$ = Mean

n = Number of observations

2) Discrete Series: The formula of mean deviation from mean for a discrete series is:

M.D about mean = $\frac{\sum f_i |x_i - \bar{x}|}{N}$

x_i = Observations / Values

$\bar{x}$ = Mean

f_i = frequency of observations

$N = f_1 + f_2 + \cdots + f_n = \sum f_i$

3) Continuous Series: The formula to find the mean deviation from mean for a continuous series is:

M.D about mean = $\frac{\sum f_i |x_i - \bar{x}|}{N}$

x_i = mid value of the class

$\bar{x}$ = Mean

f_i = frequency of observations

$N = f_1 + f_2 + \cdots + f_n = \sum f_i$

Relative Measure for Mean Deviation about Mean

The relative measure corresponding to the mean deviation about mean is called the coefficient of mean deviation about mean and it is obtained as follows

Coefficient of MD about mean= $\frac{\text{MD about mean}}{\text{Mean}}$

Example 1

Calculate the mean deviation about the mean of the following values; 2, 4, 7, 8 and 9. Also find coefficient of MD about mean.

Mean of the given data ($\bar{x}$) = Sum of all the terms/total number of terms

= (2+4+7+8+9)/5

= 6

x_i	$\lvert x_i - \bar{x} \rvert$
2	4
4	2
7	1
8	2
9	3
n=5	$\sum \lvert x_i - \bar{x} \rvert = 12$

Mean deviation about mean = $\frac{\sum |x_i - \bar{x}|}{n}$

= 12 / 5

= 2.4

Coefficient of MD about mean= $\frac{\text{MD about mean}}{\text{Mean}}$

= 2.4/6

= 0.4

Example 2

Calculate the mean deviation about the mean using the following data

6, 7, 10, 12, 13, 4, 8, 12

Solution

First we have to find the mean of the data that we are provided with

Mean of the given data ($\bar{x}$) = Sum of all the terms/total number of terms

= (6+7+10+12+13+4+8+12)/8

= 9

Next, we have to find the mean deviation

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x}\rvert$
6	6 - 9 = -3	|-3| = 3
7	7 - 9 = -2	|-2| = 2
10	10 - 9 = 1	|1|= 1
12	12 - 9 = 3	|3|= 3
13	13 - 9 = 4	|4|= 4
4	4 - 9 = -5	|-5|= 5
8	8 - 9 = -1	|-1|= 1
12	12 - 9 = 3	|3|= 3
n= 8		$\sum \lvert xi - \bar{x}\rvert = 22$

Mean deviation about mean $= \frac{\sum \lvert xi - \bar{x}\rvert}{n}$

= 22 / 8

= 2.75

Example 3

Find the mean deviation from the arithmetic mean of the following distribution:

Profits of companies (Rs in lakh) Class intervals	Number of Companies
10–20	5
20–30	8
30–50	16
50–70	8
70–80	3

Solution

Class	f_i	Mid value x_i	$x_i f_i$	x_i - x'	d_i= |x_i - x'|	$f_i d_i$
10–20	5	15	75	-25.5	25.5	127.5
20–30	8	25	200	-15.5	15.5	124
30–50	16	40	640	-0.5	0.5	8
50–70	8	60	480	19.5	19.5	156
70–80	3	75	225	34.5	34.5	103.5
	N= 40	Σxi=215	Σxifi= 1620		Σdi=95.5	Σxidi =519

$\bar{x}$ $= \Sigma x_i f_i / N$

= 1620/40

= 40.5

M.D about mean $= \frac{\sum fi\lvert xi - \bar{x}\rvert}{N}$

= 519/ 40

= 12.975

II. Mean Deviation from Median

1) Individual Series: The formula to find the mean deviation from median for an individual series is:

$$\text{M.D about median} = \frac{\sum |\mathbf{xi - Md}|}{\mathbf{n}}$$

xi = Observations / Values
M_d = Median
n = Number of observations

2) Discrete Series: The formula to find the mean deviation from median for a discrete series is:

$$\text{M.D about median} = \frac{\sum \mathbf{fi|xi - Md|}}{\mathbf{N}}$$

xi = Observations / Values
M_d = Median
fi = frequency of observations
N = $f1 + f2 + \cdots + fn = \sum fi$

3) Continuous Series: The formula to find the mean deviation from median for a continuous series is:

$$\text{M.D about median} = \frac{\sum \mathbf{fi|xi - Md|}}{\mathbf{N}}$$

xi = mid value of the class
M_d = Median
fi = frequency of observations
N = $f1 + f2 + \cdots + fn = \sum fi$

Relative Measure for Mean Deviation about Median

The relative measure corresponding to the mean deviation about median is called the coefficient of mean deviation about median and it is obtained as follows

$$\text{Coefficient of MD about median} = \frac{\text{MD about Median}}{\text{Median}}$$

Example 1

Calculate the mean deviation about the median of the following values; 2, 4, 7, 8 and 9.

Median of the given data (M_d) = 7

x_i	$\lvert x_i - M_d \rvert$
2	5
4	3
7	0
8	1
9	2
n=5	$\sum \lvert x - Md \rvert = 11$

$$\text{Mean deviation about median} = \frac{\sum |xi - Md|}{n}$$
$$= 11 / 5$$
$$= 2.2$$

Example 2

Calculate the mean deviation about the median and the coefficient of mean deviation about the median using the data given below:

Test Marks of 9 students are as follows: 86, 25, 87, 65, 58, 45, 12, 71, 35 respectively.

Solution

First we have to arrange the marks into ascending order, i.e., 12, 25, 35, 45, 58, 65, 71, 86, 87.

Then we have to find out the median so,

Median (M_d) = Value of the (n+1)/2 -th term

= Value of the 5th term

= 58

Now we have to calculate the mean deviation

x_i	$\lvert x_i - M_d \rvert$
12	46
25	33
35	23
45	13
58	0
65	7
71	13
86	28
87	29
n = 9	$\sum \lvert xi - Md \rvert = 192$

M.D about median $= \frac{\sum \lvert xi - Md \rvert}{n}$

= 192/9

= 21.333

Lastly, we have to find the coefficient of mean deviation from median so,

Coefficient of the mean deviation from median = M.D about median/ median

= 21.33/58

= 0.368

<u>Example 3</u>

Find the mean deviation from the median of the following distribution:

Class intervals	Frequencies
20–30	5
30–40	10
40–60	20
60–80	9
80–90	6

Solution

Example 4

A batch of 10 students obtained the following marks out of 100. Calculate the mean deviation and its coefficient.

Marks: 58, 39, 22, 11, 44, 28, 49, 55, 41 and 42.

Solution:

The series in ascending order:

11, 22, 28, 39, 41, 42, 44, 49, 55, 58

The median value for the series is:

n= 10 (N is even)

(n/2) = 10/2= 5

(n/2)+1= 5+1= 6

$$\text{Median(Md)} = \frac{(n/2)\text{th term} + (n/2)\text{th term}}{2}$$

$$= \frac{5\text{th term} + 6\text{th term}}{2}$$

$$= \frac{41 + 42}{2}$$

= 41.5

Items (xi)	|xi - Md| (|xi- 41.5|)
11	30.5
22	19.5
28	13.5
39	2.5
41	0.5
42	0.5
44	2.5
49	7.5
55	13.5
58	16.5
n=10	107

$$\text{M.D about median} = \frac{\sum |xi - Md|}{n}$$

$$= \frac{107}{10}$$

= 10.7 marks

$$\text{Coefficient of MD about median} = \frac{\text{MD about median}}{\text{Median}}$$

= 10.7/ 41.5

= 0.26

Example 5

Calculate Mean Deviation and Coefficient of Mean Deviation for the following discrete data using median:

Items	14	36	45	50	70
Frequency	2	5	1	1	3

Solution:

III. Mean Deviation from Mode

1) Individual Series: The formula to find the mean deviation from mode for an individual series is:

$$\text{M.D about mode} = \frac{\sum |xi - Mo|}{n}$$

xi = Observations / Values

M_o = Mode

n = Number of observations

2) Discrete Series: The formula to find the mean deviation from mode for a discrete series is:

$$\text{M.D about mode} = \frac{\sum fi|xi - Mo|}{N}$$

xi = Observations / Values

M_o = Mode

fi = frequency of observations

N $= f1 + f2 + \cdots + fn = \sum fi$

3) Continuous Series: The formula to find the mean deviation from mode for a continuous series is:

M.D about mode = $\frac{\sum \mathbf{fi|xi-Mo|}}{\mathbf{N}}$

xi = mid value of the class
M_o = Mode
fi = frequency of observations
N $= f1 + f2 + \cdots + fn = \sum fi$

Relative Measure for Mean Deviation about Mode

The relative measure corresponding to the mean deviation about mode is called the coefficient of mean deviation about mode and it is obtained as follows

Coefficient of MD about mode $= \frac{\text{MD about mode}}{\text{Mode}}$

Example 1

The following series has the sizes of shoes that are worn by the customers. Find out the mean deviation about the mode of shoe size.

4, 3, 4, 7, 6, 7, 8, 7

Solution

The maximum frequency is 3, whose value is 7, so the mode size of the shoes = 7.

xi	xi- Mo	|xi-Mo|
4	-3	3
3	-4	4
4	-3	3
7	0	0
6	-1	1
7	0	0
8	1	1
7	0	0
n= 8		$\sum \lvert xi - Mo \rvert = 12$

M.D about mode $= \frac{\sum \mathbf{|xi-Mo|}}{\mathbf{n}}$

= 12/8

= 1.5

STANDARD DEVIATION

The standard deviation measures the absolute dispersion or variability of a distribution; the greater the amount of dispersion or variability, the greater the standard deviation, the greater will be the magnitude of the deviations of the values from their mean. It is denoted as ‘σ’.

A small standard deviation means a high degree of uniformity of the observations as well as homogeneity of a series; a large standard deviation means just the opposite. Thus, if we have two or more comparable series with identical or nearly identical means, it is the distribution with the smallest standard deviation that has the most representative mean. Hence standard deviation is extremely useful in judging the representativeness of the mean.

Variance

The variance is a measure of how far a set of data are dispersed out from their mean or average value. It is denoted as '$\sigma 2$'. The variance is a measure of variability. It is calculated by taking the average of squared deviations from the mean. Variance tells you the degree of spread in the data set.

Variance = (Standard deviation)^2= σ^2

CALCULATION OF STANDARD DEVIATION

I. Individual Observations

In case of individual observations, standard deviation may be computed by applying any of the following two methods:

1. By taking deviations of the items from the actual mean.
2. By taking deviations of the items from an assumed mean.

1. Deviations taken from Actual Mean:

When deviations are taken from actual mean the following formula is applied:

$$\sigma = \sqrt{\frac{\Sigma\, di^2}{n}}$$

where di = $(x_i - \bar{x})$ and n = number of observations.

Steps:

(i) Calculate the actual mean of the series, i.e., $\bar{x}$

(ii) Take the deviations of the items from the mean, i.e., find $(x_i - \bar{x})$. Denote this deviation by di.

(iii) Square these deviations and obtain the total Σd_i^2.

(iv) Divide Σd_i^2 by the total number of observations, i.e., n, and extract the square root. This gives us the value of standard deviation.

2. Deviations taken from Assumed Mean

When deviations are taken from assumed mean the following formula is applied:

$$\sigma = \sqrt{\frac{\Sigma\, di^2}{n} - \left(\frac{\Sigma\, di}{n}\right)^2}$$

Steps:

(i) Take the deviations of the items from an assumed mean i.e., obtain $(x_i - A)$. Denote these deviations by d_i. Take the total of these deviations, i.e., obtain Σd_i.

(ii) Square these deviations and obtain the total Σd_i^2 .

(iii) Substitute the value of Σd_i^2, Σd_i and *n* in the formula.

<u>Example 1</u>

Calculate standard deviation (using actual mean) from the following observations of marks of 5 students of a tutorial group:

Marks (out of 25): 8 12 13 15 22

Solution:

xi	di=$(x_i-\bar{x})$	d_i^2
8	-6	36
12	-2	4
13	-1	1
15	1	1

22	8	64
$\Sigma x_i=70$		$\Sigma d_i^2=106$

Here

n = 5

Mean, $\bar{x}$ $= \Sigma xi/n$

$= 70/5$

$=14$

Therefore standard deviation σ $= \sqrt{\frac{\Sigma\, di^2}{n}}$

$= \sqrt{\frac{106}{5}}$

$= 4.604$

<u>Example 2</u>

Use actual mean method to find standard deviation for following values.

5, 10, 25, 30, 50

<u>Example 3</u>

Following figures give the income of 10 persons in rupees. Find the standard deviation using actual mean.

227, 235, 255, 269, 292, 299, 312, 321, 333, 348

(Hint: Standard deviation σ = 39.24)

Example 4

Calculate standard deviation (using assumed mean) from the following observations of marks of 5 students of a tutorial group:

Marks (out of 25): 8 12 13 15 22

Solution

x_i	$d_i=(xi-A)$	d_i^2
8	-5	25
12	-1	1
13	0	0
15	2	4
22	9	81
	$\Sigma d_i = 5$	$\Sigma d_i^2 = 111$

Here

n= 5

Therefore standard deviation, σ $= \sqrt{\frac{\Sigma di^2}{n} - \left(\frac{\Sigma di}{n}\right)^2}$

$= \sqrt{\frac{111}{5} - \left(\frac{5}{5}\right)^2}$

= 4.604

Example 5

Use assumed mean method to find standard deviation for following data.

5 10 25 30 50

II. Calculation of Standard Deviation—Discrete Series

For calculating standard deviation in discrete series any of the following methods may be applied:

1. Actual mean method.
2. Assumed mean method.

1. Actual Mean Method:

When this method is used, the following formula is applied.

Standard Deviation, $\boldsymbol{\sigma} = \sqrt{\frac{\Sigma \mathbf{fidi^2}}{\mathbf{N}}}$

where $di = (x_i - \bar{x})$ and $N = f_1+f_2+\ldots+f_n = \Sigma f_i$

Steps

(i) Calculate, $di = xi-\bar{x}$ where $\bar{x}$ is the actual mean.

(ii) Obtain the squares of the deviations, i.e., calculate d_i^2.

(iii) Multiply the squared deviations by respective frequencies and obtain the total, $\Sigma f_i d_i^2$. Substitute the values in the above formula.

2. Assumed Mean Method

When this method is used, the following formula is applied.

$$\text{Standard Deviation, } \boldsymbol{\sigma} = \sqrt{\frac{\sum \mathbf{fidi^2}}{\mathbf{N}} - \left(\frac{\sum \mathbf{fidi}}{\mathbf{N}}\right)^{\mathbf{2}}}$$

where d = $(x_i – A)$ and N = $f_1+f_2+\ldots+f_n= \Sigma f_i$

Steps:

(i) Take the deviations of the items from an assumed mean and denote these deviations by d_i.
(ii) Multiply the deviations by the respective frequencies and obtain the total, $\Sigma f_i d_i$.
(iii) Obtain the squares of the deviations, i.e., calculate d_i^2 .
(iv) Multiply the squared deviations by respective frequencies and obtain the total, Σf_d^2. Substitute the values in the above formula.

Example 1

Calculate the standard deviation (using actual mean) from the data given below:

Size of item	Frequency
3.5	3
4.5	7
5.5	22
6.5	60
7.5	85
8.5	32
9.5	8

Solution:

xi	fi	xifi	di= xi-$\bar{x}$	di^2	fidi^2
3.5	3	10.5	-3.59	12.89	38.66
4.5	7	31.5	-2.59	6.708	46.96
5.5	22	121	-1.59	2.528	55.62
6.5	60	390	-0.59	0.348	20.89
7.5	85	637.5	0.41	0.168	14.29
8.5	32	272	1.41	1.988	63.62
9.5	8	76	2.41	5.808	46.46
	N= 217	$\Sigma x_i f_i$ =1539			$\Sigma f_i d_i^2$ =286.5

Here N=217

$\bar{x} = \Sigma x_i f_i / N$

$= 1539/217$

$= 7.09$

Standard Deviation, $\sigma = \sqrt{\frac{\sum fidi^2}{N}}$

$= \sqrt{\frac{286.5}{217}}$

$= 1.149$

Example 2

Calculate the standard deviation (using assumed mean) from the data given below:

Size of item	Frequency

3.5	3
4.5	7
5.5	22
6.5	60
7.5	85
8.5	32
9.5	8

Solution

x_i	f_i	$d_i = x_i - A$	$f_i d_i$	d_i^2	$f_i d_i^2$
3.5	3	-3	-9	9	27
4.5	7	-2	-14	4	28
5.5	22	-1	-22	1	22
6.5	60	0	0	0	0
7.5	85	1	85	1	85
8.5	32	2	64	4	128
9.5	8	3	24	9	72
	N= 217		$\Sigma f_i d_i = 128$		$\Sigma f_i d_i^2 = 362$

Let A= 65

N= 217

$$\text{Standard Deviation, } \sigma = \sqrt{\frac{\sum \text{fidi}^2}{\text{N}} - \left(\frac{\sum \text{fidi}}{\text{N}}\right)^2}$$

$$= \sqrt{\frac{362}{217} - \left(\frac{128}{217}\right)^2}$$

$$= 1.149$$

<u>Example 3</u>

Calculate standard deviation for following data.

Item	frequency
10.5	5
20.5	8
30.5	16
50.5	8
70.5	3

III. Calculation of Standard Deviation—Continuous Series

In continuous series any of the methods discussed above for discrete frequency distribution can be used. For calculating standard deviation in discrete series any of the following methods may be applied:

1. Actual mean method.
2. Assumed mean method.
3. Step deviation method.

However, in practice it is the step deviation method that is mostly used.

1. Actual Mean Method:

When this method is used, the following formula is applied.

$$\text{Standard Deviation, } \boldsymbol{\sigma} = \sqrt{\frac{\sum \mathbf{fidi^2}}{\mathbf{N}}}$$

where di = ($x_i - \bar{x}$), N = $f_1+f_2+\ldots+f_n = \Sigma f_i$ and x_i is the mid-values of corresponding class.

Steps:

(i) Find the mid-points(x_i) of various classes.

(ii) Calculate, di =xi-$\bar{x}$ where $\bar{x}$ is the actual mean.

(iii) Obtain the squares of the deviations, i.e., calculate d_i^2.

(iv) Multiply the squared deviations by respective frequencies and obtain the total, $\Sigma f_i d_i^2$. Substitute the values in the above formula.

2. Assumed Mean Method

When this method is used, the following formula is applied.

$$\text{Standard Deviation, } \boldsymbol{\sigma} = \sqrt{\frac{\sum \mathbf{fidi^2}}{\mathbf{N}} - \left(\frac{\sum \mathbf{fidi}}{\mathbf{N}}\right)^2}$$

where d = (x_i – A), N = $f_1+f_2+\ldots+f_n = \Sigma f_i$ and x_i is the mid-values of corresponding class.

Steps:

(i) Find the mid-points (x_i) of various classes.

(ii) Calculate, $d_i = x_i$ - A, where A is the assumed mean.

(iii) Multiply the deviations by the respective frequencies and obtain the total, $\Sigma f_i d_i$.

(iv) Obtain the squares of the deviations, i.e., calculate d_i^2 .

(v) Multiply the squared deviations by respective frequencies and obtain the total, $\Sigma f_i d_i^2$.

Substitute the values in the above formula.

3. Step Deviation Method

When this method is used we take a common factor from the given data. The formula for computing standard deviation is:

$$\text{Standard Deviation, } \boldsymbol{\sigma} = \sqrt{\frac{\sum \mathbf{fiui^2}}{\mathbf{N}} - \left(\frac{\sum \mathbf{fiui}}{\mathbf{N}}\right)^2} \ \mathbf{x\ c}$$

where ui = $\frac{(x_i - A)}{c}$ and c = common factor(class interval)

Steps:

(i) Find the mid-points (x_i) of various classes.

(ii) Calculate, $u_i = \frac{(xi - A)}{c}$, where A is the assumed mean and c is the class interval.

(iii) Multiply the frequency of each class with these deviations and obtain $\Sigma f_i u_i$.

(iv) Square the deviations and multiply them with the respective frequencies of each class and obtain $\Sigma f_i u_i^2$.

Example 1

The following table gives the distribution of income of 100 families in a village. Calculate standard deviation using actual mean:

Income (Rs.)	No. of families
0–1000	18
1000–2000	26
2000–3000	30
3000–4000	12
4000–5000	10
5000–6000	4

Solution

Income	f_i	x_i	$x_i f_i$	$d_i = x_i - \bar{x}$	$f_i d_i^2$
0–1000	18	500	9000	-1820	59623200
1000–2000	26	1500	39000	-820	17482400
2000–3000	30	2500	75000	180	972000
3000–4000	12	3500	42000	1180	16708800
4000–5000	10	4500	45000	2180	47524000
5000–6000	4	5500	22000	3180	40449600
	N= 100		232000		182760000

Here

Actual Mean, $\bar{x} = \frac{\sum \mathbf{fixi}}{\mathbf{N}}$

$= 232000/100$

$= 2320$

Standard Deviation, $\sigma = \sqrt{\frac{\sum fidi^2}{N}}$

$= \sqrt{\frac{182760000}{100}}$

$= 1351.888$

Example 2

The following table gives the distribution of income of 100 families in a village. Calculate standard deviation using assumed mean:

Income (Rs.)	No. of families
0–1000	18
1000–2000	26

2000–3000	30
3000–4000	12
4000–5000	10
5000–6000	4

Solution

Income (Rs.)	No. of families f_i	Mid-value x_i	$d_i = x_i - A$	$f_i d_i$	$f_i d_i^2$
0–1000	18	500	-2000	-36000	72000000
1000–2000	26	1500	-1000	-26000	26000000
2000–3000	30	2500	0	0	0
3000–4000	12	3500	1000	12000	12000000
4000–5000	10	4500	2000	20000	40000000
5000–6000	4	5500	3000	12000	36000000
	N= 100			-18000	186000000

Let A= 2500

$$\text{Standard Deviation, } \sigma = \sqrt{\frac{\sum fidi^2}{N} - \left(\frac{\sum fidi}{N}\right)^2}$$

$$= \sqrt{\frac{186000000}{100} - \left(\frac{-18000}{100}\right)^2}$$

$$= 1351.888$$

<u>Example 3</u>

The following table gives the distribution of income of 100 families in a village. Calculate standard deviation using step deviation method:

Income (Rs.)	No. of families
0–1000	18
1000–2000	26
2000–3000	30
3000–4000	12
4000–5000	10
5000–6000	4

Solution

Income (Rs.)	No. of families f_i	x_i	$u_i=(x_i-A)/c$	$f_i u_i$	$f_i u_i^2$
0–1000	18	500	-2	-36	72
1000–2000	26	1500	-1	-26	26
2000–3000	30	2500	0	0	0
3000–4000	12	3500	1	12	12
4000–5000	10	4500	2	20	40
5000–6000	4	5500	3	12	36
	N=100			-18	186

Let

A= 2500

C= 1000

Standard Deviation, $\sigma = \sqrt{\frac{\sum fiui^2}{N} - \left(\frac{\sum fiui}{N}\right)^2} \times c$

$= \sqrt{\frac{186}{100} - \left(\frac{-18}{100}\right)^2} \times 1000$

$= 1351.888$

<u>Example 4</u>

Calculate standard deviation from following data using step deviation method.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No of Students	2	5	6	4	21	16	19	15	12	4

<u>Example 5</u>

For the following distribution of marks scored by a class of 40 students, calculate standard deviation.

Class interval	0-10	10-20	20-40	40-60	60-90
No of Students	4	9	15	9	5

COEFFICIENT OF VARIATION

The Standard deviation discussed above is an absolute measure of dispersion. The corresponding relative measure is known as the coefficient of variation. It is used in such problems where we want to compare the variability of two or more than two series.

That series (or group) for which the coefficient of variation is greater is said to be more variable or conversely less consistent, less uniform, less stable or less homogeneous.

On the other hand, the series for which coefficient of variation is less is said to be less variable or more consistent, more uniform more stable or more homogeneous.

Coefficient of variation is denoted by the symbol C.V. and is obtained as follows:

Coefficient of variation or C.V. $= \dfrac{\sigma}{\bar{x}} \times 100$

Example 1

The scores of two batsmen A and B in ten innings during a certain season are:

A	32	28	47	63	71	39	10	60	96	14
B	19	31	48	53	67	90	10	62	40	80

Find (using coefficient of variation) which of the batsmen A, B is more consistent in scoring.

Solution:

Coefficient of variation of batsman A

x_1	$d_1 = x1-46$	d_1^2
32	-14	196
28	-18	324
47	1	1
63	17	289
71	25	625
39	-7	49
10	-36	1296
60	14	196
96	50	2500
14	-32	1024
$\sum x1=460$		$\sum d1^2 = 6500$

Here n1=10

Mean, $\overline{x1} = \dfrac{\sum x1}{n1} = 460/10 = 46$

Standard Deviation of A, $\sigma_1 = \sqrt{\dfrac{\sum d1^2}{n1}}$

$= \sqrt{\dfrac{6500}{10}}$

$= 25.495$

Coefficient of variation of A, $CV_1 = \dfrac{\sigma 1}{\overline{x1}} \times 100$

$= (25.495/46) \times 100$

$= 55.424$

Coefficient of variation of batsman B

x_2	$d_2 = x2-50$	d_2^2
19	-31	961
31	-19	361
48	-2	4
53	3	9
67	17	289
90	40	1600
10	-40	1600
62	12	144
40	-10	100
80	30	900
$\sum x2=500$		5968

Here $n_2=10$

Mean, $\overline{x2} = \frac{\sum x2}{n2} = 500/10 = 50$

Standard Deviation of B, $\sigma_2 = \sqrt{\frac{\sum d2^2}{n2}}$

$= \sqrt{\frac{5968}{10}}$

$= 24.429$

Coefficient of variation of B, $CV_2 = \frac{\sigma 2}{\overline{x2}} \times 100$

$= (24.429/50) \times 100$

$= 48.858$

Since coefficient of variation is less for batsman B, hence batsman B is more consistent.

Example

From the marks given below obtained by two students taking the same course, find out who is the more consistent student.

A	: 58	59	60	65	66	52	75	31	46	48
B	: 56	87	89	46	93	65	44	54	78	68

<u>Example</u>

The scores of two batsmen A and B in ten innings during a certain season are:

A	:	35	23	49	68	73	36	12	62	90	12
B	:	18	32	44	59	67	91	11	61	41	82

Find which of the two batsmen A, B is more consistent in scoring.

Example

Calculate coefficient of variation from following data.

Size of item	6	7	8	9	10	11	12
Frequency	3	6	9	13	8	5	4