UNIT I: MEASURES OF CENTRAL TENDENCY

An average or measure of central tendency gives a single representative value for a set of usually unequal values. This value is the point around which all the values cluster. So, the measure of central tendency is also called a measure of central location.

Definition

An average is a value which is a representative of a set of data

Various important measures of central tendency are

A. Arithmetic mean
B. Geometric mean
C. Harmonic mean

Mathematical Averages

D. Median
E. Quartiles
F. Deciles
G. Percentiles
H. Mode

Positional Averages

Objectives or Functions of an average

i. Averages provide a quick understanding of complex data.
ii. Averages enable comparison
iii. Average facilitate sampling techniques.
iv. Averages pave the way for further statistical analysis.
v. Averages establish the relationship between variables.

A. ARITHMETIC MEAN

Definition

Arithmetic mean is the total (sum) of all values divided by the number of observations.

Calculation of Arithmetic mean for Raw data

When the observed values are given individually such as $x_1, x_2, x_3 \ldots. x_n$ the arithmetic mean is given by

$$\text{Arithmetic mean } \bar{X} = \frac{\text{Total of all values}}{\text{Number of the observations}}$$

$$= \frac{x_1 + x_2 + x_3 + \ldots. + x_n}{n} = \frac{\Sigma x_i}{n}$$

Example: Given $\bar{X}$ =1600 and n=5 find the total.

$\Sigma x_i = n * \bar{X}$ = 5 *1600 =8000.

Example: Calculate the arithmetic mean for the following

1600, 1590, 1560, 1610, 1640, 10.

Solution:

$$\text{Arithmetic mean, } \bar{X} = \frac{1600 + 1590 + 1560 + 1610 + 1640+10}{6}$$

$$= \frac{8010}{6} = 1335$$

Example: Calculate Arithmetic mean

S.No.	1	2	3	4	5	6	7	8	9	10
Sales in 1000's(x)	34	55	45	62	48	57	28	57	62	78

$$\text{Arithmetic mean, } \bar{X} = \frac{34+ 55 + 45 + 62 + 48 +57+28+57+62+78}{10}$$

$$= \frac{526}{10} = 52.6 \text{ (average sales)}$$

Discrete data

Let x_i, x_2, $x3 \ldots. x_n$ be the n values of the variable x with corresponding frequencies f_i, f_2, $f_3 \ldots. f_n$. then the arithmetic mean $\bar{X} = \frac{x_1.f_1 + x_2.\ f_2 + x_3.f_3 + \ldots. + x_n .\ f_n}{f_1 + f_2 + f_3 + \ldots. + f_n}$

$$= \frac{\Sigma\, x_i\, f_i}{\Sigma\, f_i}$$

Example: Calculate the arithmetic mean

x	f	xf
2	4	2x4= 8
4	6	24
6	10	60
8	12	96
10	8	80
12	7	84
14	3	42
	$\Sigma f_i =$ 50	$\Sigma x_i f_i$ = 394

$$\text{Arithmetic mean, } \bar{X} = \frac{\Sigma\, x_i\, f_i}{\Sigma\, f_i} = \frac{394}{50} = 78.8$$

Continuous data

Let $m_i, m_2, m_3 \dots m_n$ be the mid values of the class interval of the variable x with corresponding frequency $f_i, f_2, f_3 \dots f_n$. then

$$\text{the arithmetic mean } \bar{X} = \frac{m_1.f_1 + m_2.\,f_2 + m_3.f_3 + \dots + m_n\,.\,f_n}{f_1 + f_2 + f_3 + \dots + f_n}$$

$$= \frac{\Sigma\, m_i\, f_i}{\Sigma\, f_i}$$

Example: Calculate the arithmetic mean

Class interval(x)	m	f	mf
20-40	(20+40)/2 = 30	4	120
40-60	50	6	300
60-80	70	10	700
80-100	90	12	1080
100-120	110	8	880
		$\Sigma f_i =$ 40	$\Sigma m_i f_i$ = 3080

$$\text{Arithmetic mean, } \bar{X} = \frac{\Sigma\, m_i\, f_i}{\Sigma\, f_i} = \frac{3080}{40} = 77.0$$

Merits and Demerits of Arithmetic Mean:

Merits:

i) It is rigidly defined.
ii) It is easy to understand and easy to calculate.
iii) If the number of items is sufficiently large, it is more accurate and more reliable.
iv) It is a calculated value and is not based on its position in the series.
v) It is possible to calculate even if some of the details of the data are lacking.
vi) Of all averages, it is affected least by fluctuations of sampling.
vii) It provides a good basis for comparison.

Demerits:

i) It cannot be obtained by inspection nor located through a frequency graph.
ii) It cannot be in the study of qualitative phenomena not capable of numerical measurement i.e. Intelligence, beauty, honesty etc.,
iii) It can ignore any single item only at the risk of losing its accuracy.
iv) It is affected very much by extreme values.
v) It cannot be calculated for open-end classes.
vi) It may lead to fallacious conclusions, if the details of the data from which it is computed are not given.

B. MEDIAN

- ✓ It is the value which divides the data into two equal parts.
- ✓ 50% of the observations will be less than median value and 50% of the values will be more than the median value.

Calculation for Raw data

Median = value of (n+1)/2 th observation after the values are arranged in ascending order of magnitude.

For example, the median of 20,30,35,64,23,46,78,34,20

Arranging the data in ascending order; 20,20,23,30,34,35,46,64,78

Md = value of [(9+1)/2 = 5^{th} observation] = 34

Suppose the given number of observations is even then median will be the average of two central values

For example, if the data is the median of 20,30,35,64,23,46,78,34,20,56

Arranging the data in ascending order : 20,20,23,30,34,35,46,56,64,78

Md = value of (10+1)/2 =5.5^{th} observation

= (value of 5^{th} observation + value of 6^{th} observation)/2

= (34+35)/2 =34.5

Discrete data

Md = value of x corresponding to the cumulative frequency just greater than or equal to N/2

i) Arrange the data is in ascending order
ii) Find the c.f.; Calculate N/2
iii) In c.f. column see the value just $\geq$ N/2
iv) Md= value of x corresponding to this c.f.

Example: Find the median

x	f	
2	4	
4	6	
6	10	
8	12	
10	8	
12	7	
14	3	
	$\Sigma f_i = 50$	

Solution:

x	f	c.f.
2	4	4
4	6	10
6	10	20
8	12	32
10	8	40
12	7	47
14	3	50
	$\Sigma f_i = 50$	

N/2= 50/2= 25.5 , Md=8

Continuous data

Md = L + {(N/2 – c.f) x c/f}

where

L= lower limit of the median class

c =class interval of the median class

f =frequency of the median class

c.f.= cumulative frequency of the class preceding the median class

N= Σf_i

Md class is the class corresponding to the c.f. just ≥ N/2.

Example: Find the median

Class interval(x)	f
20-40	4
40-60	6
60-80	10
80-100	12
100-120	8
	$\Sigma f_i =$ 40

Solution:

Class interval(x)	f	c.f
20-40	4	4
40-60	6	10
60-80	10	20
80-100	12	32
100-120	8	40
	$\Sigma f_i = 40$	

- N/2= 40/2 =20
- Median class is 60-80
- L=60, c=80-60=20, f= 10, c.f= 10
- Md = L + {(N/2 – c.f) x c/f} = 60 + {(20 -10)x (20/10)}

  = 60 + {10x2} = 60+20 = 80.

Example: Find the median

Marks	No. of students	c.f.
10-25	6	6
25-40	20	26
40-55	44	70
55-70	26	96
70-85	3	99
85-100	1	100
	$\Sigma f_i = 100$	

- N/2= 100/2=50
- Median class is 40-55
- L= 40, f = 44, c = 55 – 40 =15, c.f. = 26
- Md = L + {(N/2 – c.f) x c/f}
  = 40 + {[50 – 26] x 15/44}
  =40 + {(24 x15)/44}
  =40 + [360/44]
  = 40+8.18 = 48.18

QUARTILES

- ✓ It is the value which divides the data into FOUR equal parts.
- ✓ There are three quartiles.
- ✓ Q_1, the first quartile or the lower quartile divides the data in such a way that 25 percent of the observations will be less than Q_1 value and 75% of the values will be more than the Q_1 value.
- ✓ Q_3, the Third quartile or upper quartile divides the data in such a way that 75 percent of the observations will be less than Q_3 value and 25% of the values will be more than the Q_3 value
- ✓ The second quartile is nothing but the median.
- ✓ 50% of the observations will be less than median value and 50% of the values will be more than the median value.

Calculation for Raw data

- ✓ Median = value of $(n+1)/2$ th observation in ascending order data.
- ✓ Q_1 = value of $(n+1)/4$ th observation in ascending order data.
- ✓ Q_3 = value of $3(n+1)/4$ th observation in ascending order data.

For example, the median of 20,30,35,64,23,46,78,34,20

Arranging the data in ascending order : 20,20,23,30,34,35,46,64,78

Md = value of $(9+1)/2 = 5^{th}$ observation

= 34

Example: Find Q_1 and Q_3, 20,30,35,64,23,46,78,34,20

Arranging the data in ascending order: 20,20,23,30,34,35,46,64,78

Q_1 = value of $(9+1)/4 = 2.5^{th}$ observation

= value of 2^{nd} observation + $0.5(3^{rd}$ value-2^{nd} value)

= 20+0.5(23 -20) =20 + 0.5 x 3 = 20+ 1.5 = 21.5

Q_3 = value of $3(9+1)/4$ =7.5th observation

= 7^{th} observation + 0.5 (8^{th}value – 7^{th} value)

= 46 + 0.5 (64-46) =46 +(0.5 x 18) = 46 + 9 = 55

Suppose the given number of observations is even then median will be the average of two central values.

Example: Find the median of 20,30,35,64,23,46,78,34,20,56.

Arranging the data in ascending order:
20,20,23,30,34,35,46,56,64,78

Md = value of $(10+1)/2 = 5.5^{th}$ observation

= (value of 5^{th} observation + value of 6^{th} observation)/2

= (34+35)/2 =34.5

Example: Find Q_1 and Q_3 20,30,35,64,23,46,78,34,20,56

Solution: Arranging the data in ascending order: 20,20,23,30,34,35,46,56,64,78

Q_1 = value of $(10+1)/4 = 2.75^{th}$ observation

= 2^{nd} value + 0.75 (3^{rd} value – 2^{nd} value)

= 20 + 0.75 (23-20)

= 20 + (0.75 x3) = 20 +2.25 =22.25

Q_3= value of 3 (n+1\4)th observation

= (3 x2.75 $=8.25^{th}$) observation

= 8^{th} value + 0.25 (9^{th} value - 8^{th} value)

= 56 + 0.25 (64-56)

=56 + 0.25 (8) = 56 + 2 = 58

Discrete data

- Md = value of x corresponding to the cumulative frequency just $\geq N/2$
- Q_1= value of x corresponding to the cumulative frequency just $\geq N/4$
- Q_3= value of x corresponding to the cumulative frequency just $\geq 3N/4$

- Arrange the data is in ascending order
- Find the c.f.
- Calculate N/2
- In c.f. column see the value greater than or equal to N/2
- Md= value of x corresponding to this c.f.

Example: Find the median and the quartiles

x	f
2	4
4	6
6	10
8	12
10	8
12	7
14	3
	$\Sigma f_i =$ 50

Solution:

x	f	c.f.
2	4	4
4	6	10
6	10	20
8	12	32
10	8	40
12	7	47
14	3	50
	$\Sigma f_i =$ 50	

- N/2= 50/2= 25; Therefore Md=8
- N/4 =50/4 = 12.5

Q_1= value of x corresponding to the cumulative frequency just greater than or equal to N/4 = 20.

$Q_1 = 6$

- 3 N/4 = 37.5

Q_3= value of x corresponding to the cumulative frequency just greater than or equal to 3N/4; Q_3= value of x corresponding to the cumulative frequency just greater than 37.5 i.e.,40

$Q_3 = 10$

Continuous data

Md = L + {(N/2 – c.f) x c/f}

where L= lower limit of the median class

= class interval of the median class

f = frequency of the median class

c.f. = c.f. of the class preceding the median class

N= Σf_i

Md class is the class corresponding to the c.f. just ≥ N/2.

$Q_1 = L_1 + \{(N/4 – c.f_1) \times c_1/f_1\}$

L_1 lower limit of the Q_1 class

c_1 class interval of the Q_1 class

f_1 frequency of the Q_1 class

c.f.1 cumulative frequency of the class preceding the Q_1 class

N= Σf_i

Q_3 class is the class corresponding to the c.f. just greater than or equal to N/4.

$Q_3 = L_3 + \{(3N/4 – c.f_3) \times c_3/f_3\}$

L_3 lower limit of the Q_3 class

c_3 class interval of the Q_3 class

f_3 frequency of the Q_3 class

$c.f._3$ cumulative frequency of the class preceding the Q_3 class

N= Σf_i Q_3 class is the class corresponding to the c.f. just greater than or equal to 3N/4.

Example: Find Median:

Class interval(x)	f
20-40	4
40-60	6
60-80	10
80-100	12
100-120	8
	$\Sigma f_i = 40$

Solution:

Class interval(x)	f	c.f
20-40	4	4
40-60	6	10
60-80	10	20
80-100	12	32
100-120	8	40
	$\Sigma f_i = 40$	

- N/2= 40/2 =20
- Median class is 60-80
- L=60, c=80-60=20, f= 10, c.f= 10
- Md = L + {(N/2 – c.f) x c/f}
  = 60 + {(20 -10) x (20/10)}
  = 60 + {10x2}
  = 60 + 20 = 80

marks	No. of students	C.f.
10-25	6	6
25-40	20	26
40-55	44	70
55-70	26	96
70-85	3	99
85-100	1	100
	$\Sigma f_i = 100$	

- N/2= 100/2=50
  Median class is 40-55
  L= 40, f = 44, c = 55 – 40 =15, c.f. = 26
  Md = L + {(N/2 – c.f) x c/f}
  = 40 + {[50 – 26] x 15/44}
  =40 + {(24 x15)/44}
  =40 + [360/44]
  = 40+8.18 = 48.18

- $Q_1 = L_1 + \{(N/4 - c.f_1) \times c_1/f_1\}$
  Q_1 class 25-40, $L_1 = 25$, $c_1 = 40 - 25 = 15$, $f_1 = 20$, $c.f_1 = 6$
  $Q_1 = 25 + \{(25 - 6)(15/20)\}$
  = 25 + {19 x 15 /20}
  = 25 + 19x0.75
  = 25 + 14.25 = 39.25

- $Q_3 = L_3 + \{(3N/4 – c.f_3) \times c_3/f_3\}$

  $3N/4 = 3x25 = 75$

  Q_3 class is 55-70

  $L_3 = 55$, $c_3 = 70 - 55 = 15$, $f_3 = 26$, $c.f_3 = 70$

  $Q_3 = L_3 + \{(3N/4 – c.f_3) \times c_3/f_3\}$

  $= 55 + \{(75 – 70) \times 15 / 26\}$

  $= 55 + \{ 5 \times 0.57\}$

  $= 55 + 2.88 = 57.88.$

Merits of Median:

i) Median is not influenced by extreme values because it is a positional average.
ii) Median can be calculated in case of distribution with open end intervals.
iii) Median can be located even if the data are incomplete.
iv) Median can be located even for qualitative factors such as ability, honesty etc.

Demerits of Median:

i) A slight change in the series may bring drastic change in median value.
ii) In case of even number of items or continuous series, median is an estimated value other than any value in the series.
iii) It is not suitable for further mathematical treatment except its use in mean deviation.
iv) It does not consider all the observations.

C. MODE

Mode is the value of x which is repeated more often or more frequently.

Raw data

Mode is found by observation. The number of times each value occurs ids noted and the value which is repeated maximum number of times is the mode.

Example: Find mode 20,30,35,64,23,46,78,34,20,56

Mode is 20 as it is repeated twice while other values are repeated only once.

<u>Case i) Unimodal – only one mode</u>

In the series 40, 30, 20, 17, 18, 32, 29, 23, 17, 17, 24, 24, 12 mode is 17.

<u>Case ii) Bimodal – two modes</u>

In the series 40, 30, 20, 17, 18, 32, 29, 23, 17, 17, 24, 24, 12, 24, 23

Mode-1 =17, mode-2 is 24,

<u>Case iii) No model:</u>

In the series 40, 34, 45, 45, 34, 40 there is no mode or mode is ill-defined.

Discrete data

Mode = value of x corresponding to the highest frequency

Case i) Unimodal – only one mode

x	f
2	4
4	6
6	10
8	12
10	8
12	7
14	3
	$\Sigma f_i = 50$

Mode = value of x corresponding to the highest frequency 12

Mode =8.

Continuous data

Mode = $l + [\{(f_1-f_0)/(2f_1- f_0-f_2)\}xc]$

where f_1 is the frequency of the modal class

f_0 is the frequency of the class preceding the modal class

f_2 is the frequency of the class succeeding the modal class

c is the class interval of the modal class

l is the lower limit of the modal class

Modal class the class corresponding to the highest frequency.

marks	No. of students
10-25	6
25-40	20 f_0
40-55	44 f_1
55-70	26 f_2
70-85	3
85-100	1
	$\Sigma f_i = 100$

Modal class is 40 -55

$L= 40$, $f_1= 44$, $f_0= 20$, $f_2= 26$, $c =55-40 =15$.

$$\text{Mode} = l + [\{(f_1-f_0)/ (2f_1- f_0-f_2)\} \text{ x} c]$$
$$= 40 + [\{(44 – 20)/ (2x44 – 20 – 26)\}x\ 15]$$
$$= 40 + \{[24/ (88-46)]x15\}]$$
$$= 40 + [(24/42)\ x15]$$
$$= 40 + [0.5714\ x15]$$
$$= 40 + 8.57$$
$$= 48.57.$$

Relationship between mean, median and mode : **Mode = 3median - 2 mean**
Called the EMPERICAL FORMULA

Measures of Dispersion & Skewness

CONTENTS

Measures of Dispersion
2.1 Range
2.2 Quartile Deviation (Q.D)
2.3 Standard Deviation
2.3.1 Combined Standard Deviation:
2.4 Mean Deviation OR Average Deviation
2.4.1 Difference between the Mean Deviation and standard Deviation
2.5 Coefficient of Variation
2.6 Variance
2.7 Skewness

Measures of Dispersion

Definition:

It is hardly fully representative of a mass, unless we know the manner in which the individual items scatter around it. A further description of the series is necessary if we are to gauge how representative the average is.

.

> *"Dispersion is the measure of the variation of the items."*
>
> *—A.L. Bowley*
>
> *"Dispersion is a measure of the extent to which the individual items vary."* *—L.R. Connor*
>
> *"Dispersion or spread is the degree of the scatter or variation of the variables about a central value."*
>
> *—B.C. Brooks and W.F.L. Dicks*
>
> *"The degree to which numerical data tend to spread about an average value is called the variation or dispersion of the data."* *—Spiegal*

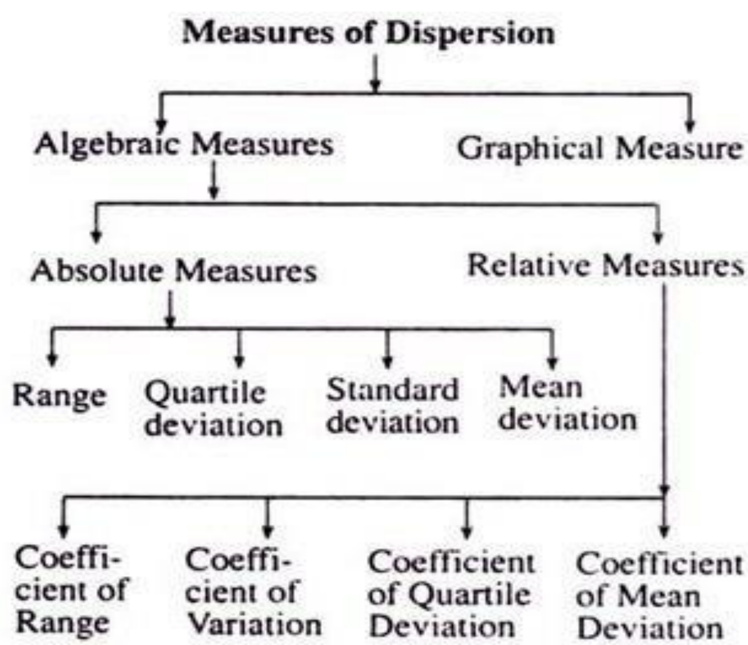
Importance or Significance of Measures of Dispersion:

The reliability of a measure of central tendency is known.

1. Measures of Dispersion provide a basis for the control of variability.
2. They help to compare two or more sets of data with regard to their variability.
3. They enhance the utility and scope of statistical techniques.

The Usual Measures of Dispersion:

The usual measures of dispersion, very often suggested by the statisticians, are exhibited with the aid of the following chart:



Difference between the Absolute measure and Relative Measure:

Absolute measure	Relative measure
Range	Co-efficient of Range
Quartile deviation	Co-efficient of Quartile Deviation
Mean Deviation (about Mean) Median Deviation (about Median) Mode Deviation (about Mode	Co-efficient of Mean Deviation (about Mean) Co-efficient of Median Deviation (about Median) Co-efficient of Mode Deviation (about Mode)
Standard deviation and Variance	Co-efficient of Variation

Absolute Measures of Dispersion:

Range

Definition: Range is the difference between the greatest (largest) and the smallest of the values.

In Symbols, Range = L-S.

L- Largest Value

S- Smallest Value

In individual observations and discrete series, L and S are easily identified. In Continuous series, the following two method are followed.

Method-I

L- Upper boundary of the highest class

S - Lower boundary of the lowest class

Method-2

L - Mid value of the highest class

S - Mid value of the lowest class

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

Uses of Range:

1. Range is used in finding the control limits of mean chart and Range chart in S.Q.C.

2. While Quoting the prices of shares, bonds, gold, etc. on daily basis or yearly basis, the minimum and the maximum prices are mentioned.

3. The minimum and the maximum temperature likely to prevail on each day are forecasted.

Merits:

1. It is simple to understand and easy to calculate.
2. It can be calculated in no time.

Demerits:

1. Its definition does not seem to suit continuous series.
2. It is based on the two extreme items. It does not consider the other items.
3. It is usually affected by the extreme items.
4. It cannot be manipulated algebraically. The Range of combined set cannot be found from the range of the individual sets.
5. It does not have sampling stability.
6. It cannot be calculated from open-end class intervals. It is a very rarely used measure. Its Scope is limited.

Quartile Deviation (Q.D)

Definition : Quartile Deviation is half of the difference between the first and the third quartiles. Hence it is called Semi Inter Quartile Range.

In Symbols, Q.D = $\frac{Q_3 - Q_1}{2}$.

Coefficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

Merits:

1. It is simple to understand and easy to calculate.
2. It is not affected by extreme items.
3. It can be calculated for data with open and classes also.

Demerits:

1. It is not based on all items. It is based on two positional values Q_1 and Q_3 and ignores the extreme 50% of the items.
2. It cannot be manipulated algebraically.
3. It is affected by sampling fluctuations.
4. Like range, it does not measure the deviation about any measure of central tendency.

Mean Deviation or Average Deviation:

Definition: Mean deviation is the arithmetic mean of the absolute deviations of the values about their arithmetic mean or median or mode.

M.D. is the abbreviation for Mean Deviation. There are three kinds of mean deviations, Viz.,

1. mean deviation or mean deviation about mean
2. mean deviation about median

3. mean deviation about mode.

Mean deviation about median is the least. it could be easily verified in individual observations and discrete series where the actual values are considered.

The relative measures are the following:

Coefficient of Mean deviation (about Mean)

$$= \frac{Mean\ Deviation\ about\ Mean}{Mean}$$

Coefficient of Mean deviation about Median

$$= \frac{Mean\ Deviation\ about\ Median}{Median}$$

Coefficient of Mean deviation about Mode

$$= \frac{Mean\ Deviation\ about\ Mode}{Mode}$$

Individual Observations:

Mean Deviation (about Mean) = $\frac{\sum |X - \overline{X}|}{N}$

Mean Deviation about Median = $\frac{\sum |X - M|}{N}$

Mean Deviation about Mode = $\frac{\sum |X - Z|}{N}$

Discrete Series:

Mean Deviation (about Mean) = $\frac{\sum f|X - \overline{X}|}{N}$

Mean Deviation about Median = $\frac{\sum f|X - M|}{N}$

Mean Deviation about Mode = $\frac{\sum f|X - Z|}{N}$

Continuous Series:

Mean Deviation (about Mean) = $\frac{\sum f|m - \overline{X}|}{N}$

Mean Deviation about Median = $\dfrac{\sum f|m-M|}{N}$

Mean Deviation about Mode = $\dfrac{\sum f|m-Z|}{N}$

Coefficient of M.D about Mean = $\dfrac{M.D.about\ Mean}{Mean}$

Coefficient of M.D about Median = $\dfrac{M.D.about\ Median}{Median}$

Coefficient of M.D about Mode = $\dfrac{M.D.about\ Mode}{Mode}$

Uses of Mean Deviation:

Mean deviation provides an opportunity to calculate deviation, absolute deviation, total deviation and average of the deviations. Standard deviation is the most important absolute measure of dispersion. knowledge of the principle of mean deviation facilities understanding the concept of standard deviation. Standard deviation is a part of almost all the theories of Statistics, Viz., skewness, kurtosis, correlation, regression, sampling, estimation, inference, S.Q.C., etc. Mean deviation is preferred when a particular discussion is not carried to other spheres. It is found to be much useful in forecasting business cycles and a few other statistical activities connected with business, economic and sociology.

Merits:

1. Mean deviations are rigidly defined.
2. They are based on all the items.
3. They are affected less by extreme items than standard deviation. Among the three mean deviations, mean deviation about median is the least.
4. They are simple to understand and not difficult to calculate.
5. They do not vary much from sample to sample.
6. They provide choice. among the three mean deviations, the one that is suitable to a particular situation can be used.
7. Formation of different distributions can be compared on the basis of a mean deviation.

Demerits:

1. Omission of negative sign of deviations makes them non-algebraic. It is pointed out as a great drawback.
2. They could not be manipulated. Combined mean deviation could not be found.
3. it is not widely used in business or economics.

Standard Deviation:

Definition: Standard Deviation is the root mean square deviation of the values from their arithmetic mean. S.D denoted by **σ** (read , sigma). **Variance** is denoted by $\boldsymbol{\sigma^2}$. S.D. is the positive square root of variance. Karl Pearson introduced the concept of standard deviation in 1893. S.D is also called **root mean square deviation**. The corresponding relative measure is **Coefficient of Variation**.

$$\text{Coefficient of variation} = \frac{Standard\ Deviation}{Arithmetic\ Mean} \times 100$$

Individual Observation:

Method-1

$$\text{Standard deviation } \sigma = \sqrt{\frac{\sum x^2}{N}}$$

Method-2

$$\text{Standard deviation } \sigma = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2}$$

Discrete Series:

Method-1

$$\text{Standard deviation } \sigma = \sqrt{\frac{\sum fx^2}{N}} \quad \text{where } x = X - \overline{X} \text{ and } N = \sum f$$

Method – 2

$$\text{Standard deviation } \sigma = \sqrt{\frac{\sum fX^2}{N} - \left(\frac{\sum fX}{N}\right)^2}$$

Continuous Series:

Method – 1

$$\text{Standard deviation } \sigma = \sqrt{\frac{\sum f\left(m - \overline{X}\right)^2}{N}}$$

Method – 2

$$\text{Standard deviation } \sigma = \sqrt{\frac{\sum fm^2}{N} - \left(\frac{\sum fm}{N}\right)^2}$$

Combined Standard Deviation:

When two or three groups merge, the mean and standard deviation of the combined group are calculated as follows.

Case 1. Merger of Two Groups

	Size	Mean	S.D.
Group I	N_1	$\bar{X}_1$	σ_1
Group II	N_2	$\bar{X}_2$	σ_2

That is,

N_1 – Number of items in the first group.

N_2 – Number of items in the second group.

$\bar{X}_1$ - Mean of items in the first group

$\bar{X}_2$ - Mean of items in the second group.

σ_1 – Standard Deviation of items in the first group.

σ_2 - Standard Deviation of items in the second group.

The Mean of the combined group,

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

The Standard Deviation of the combined group,

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

where $d_1 = \bar{X}_1 - \bar{X}_{12}$ and $d_2 = \bar{X}_2 - \bar{X}_{12}$

Case 2. Merger of Three Groups

	Size	Mean	S.D.
Group I	N_1	$\bar{X}_1$	σ_1
Group II	N_2	$\bar{X}_2$	σ_2

Group III N_3 $\overline{X}_3$ $\sigma 3$

That is

N_1 – Number of items in the first group.

N_2 – Number of items in the second group.

$\overline{X}_1$ - Mean of items in the first group

$\overline{X}_2$ - Mean of items in the second group.

σ_1 – Standard Deviation of items in the first group.

σ_2 - Standard Deviation of items in the second group.

The Mean of the combined group,

$$\overline{X}_{123} = \frac{N_1 \overline{X}_1 + N_2 \overline{X}_2 + N_3 \overline{X}_3}{N_1 + N_2 + N_3}$$

The Standard Deviation of the combined group,

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1 d_1^2 + N_2 d_2^2 + N_3 d_3^2}{N_1 + N_2 + N_3}}$$

where $d_1 = \overline{X}_1 - \overline{X}_{123}$, $d_2 = \overline{X}_2 - \overline{X}_{123}$ and $d_3 = \overline{X}_3 - \overline{X}_{123}$

Difference between the Mean Deviation and standard Deviation:

Mean Deviations	**Standard Deviations**
Deviations are calculated from Mean or Median or Mode.	Deviations are always calculated from Mean.
While finding the deviations, negative sign is omitted.	Deviations are squared. Finally square root is taken.
M.D.'s are simple to understand and not difficult to calculate.	It is not simple to under standard and it is not easy to calculate.
Omission of negative sign is considered to be deficient mathematically.	It has many desirable mathematically properties.

Uses

Standard deviation is the best absolute measure of dispersion. It is a part of many statistical concepts such as Skewness, Kurtosis, Correlation, Regression, Estimation, sampling, tests of Significance and Statistical Quality Control. Not only in statistics but also in Biology, education, Psychology and other disciplines standard deviation is of immense use.

Merits:

1. Standard deviation is rigidly defined.
2. It is calculated on the basis of the magnitudes of all the items.
3. It could be manipulated further. The combined S.D. can be calculated.
4. Mistakes in its calculation can be corrected. The entire calculation need not be redone.
5. Coefficient of variation is based on S.D. It is the best and most widely used relative measure of dispersion.
6. It is free from sampling fluctuations. This property of sampling stability has brought it an indispensable place in tests of significance.
7. It reduces the complexity in the approach of normal distribution by providing standard normal variable.
8. It is the most important absolute measure of dispersion. It is used in all the areas of statistics. It is widely used in other disciplines such as Psychology, Education and Biology as well.
9. Scientific calculators show the standard deviation of any series.
10. Different forms of the formula are available.

Demerits:

1. Compared with other absolute measures of dispersion, it is difficult to calculate.
2. It is not simple to understand.
3. It gives more weight age to the items away from the mean than those near the mean as the deviations are squared.

Coefficient of Variation

Definition: Coefficient of variation is the most widely used relative measure of dispersion. It is based on the best absolute measure of dispersion and the best measure of central tendency. It is a percentage. While comparing two or more groups, the group which has less coefficient of variation is less variable or more consistent or more stable or more uniform or more homogeneous. Coefficient of Variation is denoted by the C.V.

$$\text{Coefficient of Variation} = \frac{Standard\ Deviation}{Arithmetic\ Mean} \times 100$$

$$\text{C.V.} = \frac{S.D.}{A.M.} \times 100 \qquad \textbf{OR} \qquad \text{C.V.} = \frac{\sigma}{\overline{\overline{X}}} \times 100$$

Variance

Variance is the mean square deviation of the values from their arithmetic mean. It is denoted by σ2. Standard deviation is the positive square root of variance and is denoted by σ. The term of variance was introduced by R.A. fisher in the year 1913. It is used much in sampling, analysis of variance, etc., In analysis of variance, total variation is split into a few components. Each component is ascribable to one factor of variation. The significance of the variation is then tested.

Individual Observation:

$$\text{Variance, } \sigma 2 = \frac{\sum \left(X - \overline{X}\right)^2}{N}$$

Discrete Series:

$$\text{Variance, } \sigma 2 = \frac{\sum fX^2}{N} - \left(\frac{\sum fX}{N}\right)^2$$

Continuous Series:

$$\text{Variance, } \sigma 2 = C^2 \left[\frac{\sum d'^2}{N} - \left(\frac{\sum fd'^2}{N}\right)^2\right]$$

Combined Variance:

Based on the notations used in combined mean and combined variance,

Combined variance of two groups,

$$\sigma_{12}^2 = \frac{N_1 \sigma_1^2 + N_2 \sigma_2^2 + +N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}$$

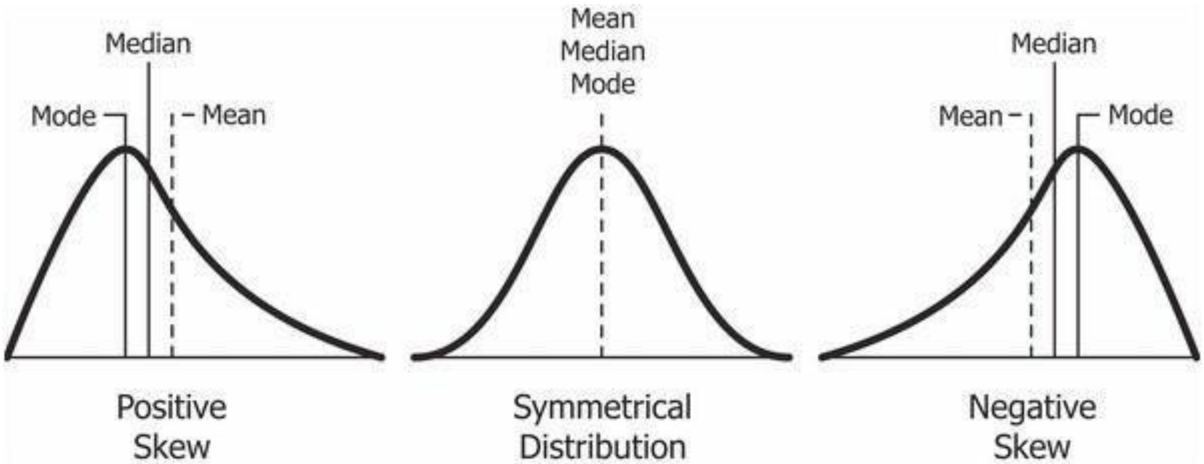
Combined variance of three groups,

$$\sigma_{123}^2 = \frac{N_1 \sigma_1^2 + N_2 \sigma_2^2 + N_3 \sigma_3^2 + N_1 d_1^2 + N_2 d_2^2 + N_3 d_3^2}{N_1 + N_2 + N_3}$$

Skewness:

Definition: Skewness is the degree of asymmetry, or departure from symmetry, of a distribution.

There are two types of Skewness: Positive and Negative



Positive Skewness means when the tail on the right side of the distribution is longer or fatter. The mean and median will be greater than the mode.

Negative Skewness is when the tail of the left side of the distribution is longer or fatter than the tail on the right side. The mean and median will be less than the mode.
Karl – Pearson (1867- 1936) was great British Biometric and Statistician. He introduced the formula given below.

Illustration 1. The following are the prices of shares of *AB* Co. Ltd. from Monday to Saturday :

Day	Price (Rs.)	Day	Price (Rs.)
Monday	200	Thursday	160
Tuesday	210	Friday	220
Wednesday	208	Saturday	250

Calculate range and its coefficient.

Solution. Range = $L - S$

Here $L = 250$ and $S = 160$

Range = $250 - 160$ = Rs. 90

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{250-160}{250+160} = \frac{90}{410} = 0.22.$$

Continuous Series. There are two methods of determining the range from data grouped into a frequency distribution. The first method is to find the difference between the upper limit of the highest wage class and the lower limit of the lowest wage class. The other method is to subtract the midpoint of the lowest wage class from the mid-point of the highest wage class. In practice, both the methods are used.

Illustration 2. Calculate Coefficient of Range from the following data :

Marks	No. of students	Marks	No. of students
10—20	8	40—50	8
20—30	10	50—60	4
30—40	12		

Solution. $\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{60-10}{60+10} = \frac{50}{70} = 0.714$

Merits and Limitations The merits and limitations of Range can be enumerated here.

Merits.

- Amongst all the methods of studying dispersion range is the simplest to understand and the easiest to compute.
- It takes minimum time to calculate the value of range. Hence, if one is interested in getting a quick rather than a very accurate picture of variability one may compute range.

Limitations.

- Range is not based on each and every item of the distribution.
- It is subject to fluctuations of considerable magnitude from sample to sample.
- Range cannot tell us anything about the character of the distribution within the two extreme observations. For example, observe the following three series :

Series A	46	6	46	46	46	46	46	46
Series B	6	10	6	6	46	46	46	46
Series C	6	6	15	25	30	32	40	46

In all the three series range is the same, *i.e.*, $(46 - 6) = 40$. But it does not mean that the distributions are alike. The range takes no account of the form interquartile range. In other words, interquartile range represents the difference between the third quartile and the first quartile. Symbolically,

$$\text{Interquartile range} = Q_3 - Q_1.$$

Very often the interquartile range is reduced to the form of the Semi-interquartile range or quartile deviation by dividing it by 2. Symbolically,

$$\text{Quartile Deviation or } Q.D. = \frac{Q_3 - Q_1}{2}.$$

Quartile deviation gives the average amount by which the two quartiles differ from the median. In asymmetrical distribution the two quartiles (Q_1 and Q_3) are equidistant from the median, *i.e.*, Med $- Q_1 = Q_3 -$ Med. and as such the difference can be taken as a measure of dispersion. The Median $\pm$ $Q.D.$ covers exactly 50 per cent of the observations.

In reality, however, one seldom finds a series in business and economic data that is perfectly symmetrial. Nearly all distributions of social series are asymmetrical. In an asymmetrical distribution, Q_1 and Q_3 are not equidistant from the median. As a result an asymmetrical distribution includes only approximately 50 per cent of observations.

When quartile deviation is very small, it describes high uniformity or small variation of the central 50% items and a high quartile deviation means that the variation among the central items is large.

Quartile deviation is an absolute measure of dispersion. The relative measure corresponding to this measure, called the coefficient of quartile deviation, is calculated as follows.

$$\text{Coefficient of } Q.D. = \frac{(Q_3 - Q_1)/2}{(Q_3 + Q_1)/2} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Coefficient of quartile deviation can be used to compare the degree of variation in different distributions.

Computation of Quartile Deviation The process of computing quartile deviation is very simple, we have just to compute the values of the upper and lower quartiles. The following illustrations would clarify calculations.

Individual Observations

Illustration 3. Find out the value of quartile deviation and its coefficient from the following data:

Roll No.	1	2	3	4	5	6	7
Marks	20	28	40	12	30	15	50

Solution. CALCULATION OF QUARTILE DEVIATION

Marks arranged in ascending order : 12 15 20 28 30 40 50

$$Q_1 = \text{Size of } \frac{N+1}{4} \text{ th item} = \text{Size of } \frac{7+1}{4} = \text{2nd item}$$

Size of 2nd item is 15. Thus $Q_1 = 15$

$$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right) \text{th item} = \text{Size of } \left(\frac{3 \times 8}{4}\right) \text{th item} = \text{6th item}$$

Size of 6th item is 40. Thus $Q_3 = 40$

$$Q.D. = \frac{Q_3 - Q_1}{2} = \frac{40 - 15}{2} = 12.5.$$

$$\text{Coefficient of } Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = = \frac{40 - 15}{40 + 15} = \frac{25}{55} = 0.455.$$

Discrete Series

Illustration 4. Compute coefficient of quartile deviation from the following data :

Marks	10	20	30	40	50	60
No. of Students	4	7	15	8	7	2

(B.Com., Madras Univ., 1998)

Solution. CALCULATION OF COEFFICIENT OF QUARTILE DEVIATION

Marks	*Frequency*	*c.f.*	*Marks*	*Frequency*	*c.f.*
10	4	4	40	8	34
20	7	11	50	7	41
30	15	26	60	2	43

$$Q_1 = \text{Size of } \frac{N+1}{4} \text{ th item} = \frac{43+1}{4} = 11\text{th item.}$$

Size of 11th item is 20. Thus, $Q_1 = 20$

$$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)\text{th item} = \frac{3 \times 44}{4} = 33\text{rd item}$$

Size of 33rd item is 40. Thus, $Q_3 = 40$

$$Q.D. = \frac{Q_3 - Q_1}{2} = \frac{40 - 20}{2} = 10$$

$$\text{Coefficient of } Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{40 - 20}{40 + 20} = 0.333$$

Illustration 5. Calculate quartile deviation and the coefficient of quartile deviation from the following data :

Wages in Rupees per week	less then 35	35–37	38–40	41–43	over 43
Number of wage earners	14	62	99	18	7

(B. Com., Madras Univ., 1998)

Solution. CALCULATION OF Q.D. AND ITS COEFFICIENT

Wages (Rs. per week)	*f*	*c.f.*
Less than 35	14	14
35–37	62	76
38–40	99	175
41–43	18	193
over 43	7	200

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

$$Q_1 = \text{Size of } \frac{N}{4} \text{th item} = \frac{200}{4} = 50 \text{ th item}$$

Q_1 lies in the class 35–37.

$$Q_1 = L + \frac{N/4 - c.f.}{f} \times i$$

$$L = 35,\ N/4 = 50,\ c.f. = 14,\ f = 62,\ i = 2$$

$$Q_1 = 35 + \frac{50 - 14}{62} \times 2 = 35 + 1.16 = 36.16$$

$$Q_3 = \text{Size of } \frac{3N}{4} \text{th item} = \frac{3 \times 200}{4} = 150 \text{ th item}$$

Q_3 lies in the class 38–40.

$$Q_3 = L + \frac{3N/4 - c.f.}{f} \times i$$

$$L = 38, \quad 3N/4 = 150, \quad c.f. = 76, \quad f = 99, \quad i = 2$$

$$Q_3 = 38 + \frac{150 - 76}{99} \times 2 = 38 + 1.49 = 39.49$$

$$Q.D. = \frac{39.49 - 36.16}{2} = 1.67$$

$$\text{Coefficient of } Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$= \frac{39.49 - 36.16}{39.49 + 36.16} = \frac{3.33}{75.65} = 0.044$$

Merits and Limitations

Merits. In certain respects it is superior to range as a measure of dispersion.

- It has a special utility in measuring variation in case of open end distributions or one in which the data may be ranked but measured quantitatively.
- It is also useful in erratic or badly skewed distributions, where the other measures of dispersion would be warped by extreme values. The quartile deviation is not affected by the presence of **extreme values.***

Limitations. Quartile deviation ignores 50% items, *i.e.*, the first 25% and the last 25%. As the value of quartile deviation does not depend upon every item of the series, it cannot be regarded as a good method of measuring dispersion.

- It is not capable of mathematical manipulation.
- Its value is very much affected by sampling fluctuations.
- It is in fact not a measure of dispersion as it really does not show the scatter around an average but rather a distance on a scale, *i.e.*, quartile deviation is not itself measured from an average, but it is a positional average. Consequently, some statisticians speak of quartile deviation as a measure of *partition* rather than a measure of dispersion. If we really desire to measure variation in the sense of showing the scatter round an average, we must include the deviation of each and every item from an average in the measurement.

Because of the above limitations quartile deviation is not often useful for statistical inference.

Percentile Range. Like semi-interquartile range, the percentile range is also used as a measure of dispersion. Percentile range of a set of data is defined as :

$$\text{Percentile Range} = P_{90} - P_{10}$$

* The Range and Q.D. are positional measures of dispersion as they are based on the position of certain items.

where P_{90} and P_{10} are the 90th and 10th percentiles respectively. The semi-percentile range, *i.e.*, $\left(\frac{P_{90} - P_{10}}{2}\right)$ can also be used, but is not commonly employed.

The Mean Deviation

The two methods of dispersion discussed above, namely, range and quartile deviation, are not measures of dispersion in the strict sense of the term because they do not show the scatterness around an average. However, to study the formation of a distribution we should take the deviations from an average. The two other measures namely, the average deviation and the standard deviation, help us in achieving this goal.

The mean deviation is also known as the average deviation. It is the average difference between the items in a distribution and the median or mean of that series. Theoretically there is an advantage in taking the deviations from median because *the sum of deviations of items from median is minimum when signs are ignored.* However, in practice, the arithmetic mean is more frequently used in calculating the value of average deviation and this is the reason why it is more commonly called mean deviation. In any case, the average used must be clearly stated in a given problem so that any possible confusion in meaning is avoided.

Computation of Mean Deviation—Individual Observations.* If X_2, X_1, X_3, X_N are N given observations then the deviation about an average A is given by

$$\text{M.D.} = \frac{1}{n} \Sigma \mid X - A \mid$$

$$\frac{1}{N} \Sigma \mid D \mid \quad \text{or} \quad \frac{\Sigma \mid D \mid}{N}$$

where $\mid D \mid = X - A \mid$. Read as mod $(X - A)$ is the modulus value or absolute value of the deviation ignoring plus and minus signs.

Steps.

- Compute the median of the series.
- The deviations of items from median ignoring ± signs and denote these deviations by $\mid D \mid$.
- Obtain the total of these deviations, *i.e.*, $\Sigma \mid D \mid$.
- Divide the total obtained in step (*iii*) by the number of observations.

If a distribution is normal, the mean ± mean deviation is the range that will include 57.7 per cent of the items in the series. If it is moderately skewed, then we may expect approximately 57.5 per cent of the items to fall within this range. Hence, if average deviation is small, the distribution is highly compact or uniform, since more than half of the cases are concentrated within a small range around the mean.

* If the mean deviation is computed from mean then in that case $\mid D \mid$ shall denote deviations of the items from mean, ignoring signs.

The relative measure corresponding to the mean deviation, called the coefficient of mean deviation, is obtained by dividing mean deviation by the particular average used in computing mean deviation. Thus, if mean deviation has been computed from median, the coefficient of mean deviation shall be obtained by dividing mean deviation by median.

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\text{Median}}.$$

If mean has been used while calculating the value of mean deviation, in such a case coefficient of mean deviation shall be obtained by dividing mean deviation by the mean.

Illustration 6. Calculate the mean deviation and its coefficient of the two income groups of five and seven members given below :

I (Rs.) :	4,000	4,200	4,400	4,600	4,800		
II (Rs.) :	3,000	4,000	4,200	4,400	4,600	4,800	5,800

Solution. CALCULATION OF MEAN DEVIATION

Group I		*Group II*	
	Deviation from median 4400 $\mid D \mid$		*Deviation from median 4400* $\mid D \mid$
4,000	400	3,000	1,400
4,200	200	4,000	400
4,400	0	4,200	200
4,600	200	4,400	0
4,800	400	4,600	200
		4,800	400
		5,800	1,400
$N = 5$	$\Sigma \mid D \mid = 1200$	$N = 7$	$\Sigma \mid D \mid = 4,000$

Mean Deviation : Group I : $\text{M.D.} = \frac{\Sigma \mid D \mid}{N}$

$\mid D \mid$ = Deviation from median ignoring signs,

$$\text{Median} = \text{Size of } \frac{N+1}{2} \text{ th item} = \frac{5+1}{2} = \text{3rd item}$$

Size of 3rd item is 4,400 $\quad \text{M.D.} = \frac{1,200}{5} = 240$

This means that the average deviation of the individual incomes from the median income is Rs. 240.

Mean Deviation : Group II

$$\text{Mean} = \text{Size of } \frac{N+1}{2} \text{ th item} = \frac{7+1}{2} = \text{4th item}$$

Size of 4th item is 4,400

$$\Sigma \mid D \mid = 4,000, \quad N = 7.$$

$$\therefore \quad \text{M.D.} = \frac{4,000}{7} = 571.43.$$

Note. If we were to compute coefficient of mean deviation we shall divide mean deviation by median. Thus for the first group :

$$\text{Coefficient of M.D.} = \frac{240}{4{,}400} = 0.054$$

and for the second group

$$\text{Coefficient of M.D.} = \frac{571.43}{4{,}400} = 0.130.$$

Calculation of Mean Deviation

Discrete Series In discrete series the formula for calculating mean deviation is

$$\text{M.D.} = \frac{\Sigma f \mid D \mid}{N} \text{ (by the same logic as given before)}$$

$\mid D \mid$ denotes deviation from median ignoring signs.

Steps.

- Calculate the median of the series.
- Take the deviations of the items from median ignoring signs and denote them by $\mid D \mid$.
- Multiply these deviations by the respective frequencies and obtain the total $\Sigma f \mid D \mid$.
- Divide the total obtained in Step (*ii*) by the number of observations. This gives us the value of mean deviation.

Illustration 7. (*a*) Calculate mean deviation from the following series :

X	10	11	12	13	14
f	3	12	18	12	3

Solution. CALCULATION OF MEAN DEVIATION

X	*f*	$\mid D \mid$	$f \mid D \mid$	*c.f.*
10	3	2	6	3
11	12	1	12	15
12	18	0	0	33
13	12	1	12	45
14	3	2	6	48
	$N = 48$		$\Sigma f \mid D \mid = 36$	

$$\text{M.D.} = \frac{\Sigma f \mid D \mid}{N}$$

$$\text{Median} = \text{Size of } \frac{N+1}{2} \text{ th item} = \frac{48+1}{2} = 24.5\text{th item}$$

Size of 24.5th item is 12, hence Median = 12

$$\text{M.D.} = \frac{36}{48} = 0.75.$$

(*b*) Calculate the mean deviation from the mean for the following data :

Size :	2	4	6	8	10	12	14	16
Frequency :	2	2	4	5	3	2	1	1

Solution. CALCULATION OF MEAN DEVIATION FROM MEAN

X	f	fX	$\lvert X-8 \rvert$ $\lvert D \rvert$	$f\lvert D \rvert$
2	2	4	6	12
4	2	8	4	8
6	4	24	2	8
8	5	40	0	0
10	3	30	2	6
12	2	24	4	8
14	1	14	6	6
16	1	16	8	8
	$N = 20$	$\Sigma fX = 160$		$\Sigma f\lvert D \rvert = 56$

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{160}{20} = 8$$

$$\text{M.D.} = \frac{\Sigma f\lvert D \rvert}{N} = \frac{56}{20} = 2.8.$$

Calculation of Mean Deviation—Continuous Series

For calculating mean deviation in continuous series the procedure remains the same as discussed above. The only difference is that here we have to obtain the mid-point of the various classes and take deviations of these points from median. The formula is same, *i.e.*,

$$\text{M.D.} = \frac{\Sigma f \mid D \mid}{N}.$$

Illustration 8. (*a*) Find the median and mean deviation of the following data :

Size	*Frequency*	*Size*	*Frequency*
0—10	7	40—50	16
10—20	12	50—60	14
20—30	18	60—70	8
30—40	25		

(*B. Com. Mysore Univ.* 1998)

Solution. CALCULATION OF MEDIAN AND MEAN DEVIATION

Size	f	*c.f.*	*m.p.* m	$\lvert m-35.2 \rvert$ $\lvert D \rvert$	$f\lvert D \rvert$
0—10	7	7	5	30.2	211.4
10—20	12	19	15	20.2	242.4
20—30	18	37	25	10.2	183.6
30—40	25	62	35	0.2	5.0
40—50	16	78	45	9.8	156.8
50—60	14	92	55	19.8	277.2
60—70	8	100	65	29.8	238.4
	$N = 100$				$\Sigma f\lvert D \rvert = 1314.8$

$$\text{Med.} = \text{Size of } \frac{N}{2}\text{th item} = \frac{100}{2} = 50\text{th item}$$

Median lies in the class 30–40

$$\text{Med.} = L + \frac{N/2 - c.f.}{f} \times i$$

$L = 30$, $N/2 = 50$, $c.f. = 37$, $f = 25$, $i = 10$

$$\text{Med.} = 30 + \frac{50 - 37}{25} \times 10 = 30 + 5.2 = 35.2$$

$$\text{M.D.} = \frac{\Sigma f|D|}{N} = \frac{1314.8}{100} = 13.148$$

(b) Calculate the mean deviation and its coefficient from the following data :

Class	*Frequency*	*Class*	*Frequency*
0—10	5	40—50	20
10—20	8	50—60	14
20—30	12	60—70	12
30—40	15	70—80	6

(B.Com., Andhra Univ., 1996)

Solution. Since nothing is special we will calculate mean deviation from median :

CALCULATION OF MEAN DEVIATION

Class	*Frequency* f	*c.f.*	*m.p.* m	$\|m - 43\|$ $\|D\|$	$f\|D\|$
0–10	5	5	5	38	190
10–20	8	13	15	28	224
20–30	12	25	25	18	216
30–40	15	40	35	8	120
40–50	20	60	45	2	40
50–60	14	74	55	12	168
60–70	12	86	65	22	264
70–80	6	92	75	32	192
	$N = 92$				$\Sigma f\|D\| = 1414$

$$\text{Med.} = \text{Size of } \frac{N}{2}\text{th item} = \frac{92}{2} = 46\text{th item}$$

Median lies in the class 40—50.

$$\text{Med.} = L + \frac{N/2 - c.f.}{f} \times i$$

$L = 40$, $N/2 = 46$, $c.f. = 40$, $f = 20$, $i = 10$

$$\therefore \quad \text{Med} = 40 + \frac{46 - 40}{20} \times 10 = 40 + 3 = 43$$

$$\text{M.D.} = \frac{\Sigma f|D|}{N} = \frac{1414}{92} = 15.37$$

$$\text{Coeff. of M.D.} = \frac{\text{M.D.}}{\text{Median}} = \frac{15.37}{43} = 0.357.$$

Merits and Limitations

Merits. The outstanding advantage of the average deviation is its relative simplicity. It is simple to understand and easy to compute. Any one familiar with the concept of the average can readily appreciate the

* A distinction is often made between population standard deviation and sample standard deviation. Population standard deviation is denoted by σ whereas sample standard deviation by S.

The standard deviation measures the absolute dispersion (or variability of distribution; the greater the amount of dispersion or variability), the greater the standard deviation, for the greater will be the magnitude of the deviations of the values from their mean. A small standard deviation means a high degree of uniformity of the observation 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.

Difference between Mean Deviation and Standard Deviation

Both these measures of dispersion are based on each and every item of the distribution. But they differ in the following respects :

- Algebraic signs are ignored while calculating mean deviation whereas in the calculation of standard deviation signs are taken into account.
- Mean deviation can be computed either from median or mean. The standard deviation, on the other hand, is always computed from the arithmetic mean because the sum of the squares of the deviation of items from arithmetic mean is the least.

Calculation of Standard Deviation

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

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

Deviations taken from Actual Mean. When deviations are taken from actual mean the following formula is applied :

$$\sigma* = \sqrt{\frac{\Sigma x^2}{N}}$$

where, $x = (X - \overline{X})$.

Steps :

- Calculate the actual mean of the series, *i.e.*, $\overline{X}$.
- Take the deviations of the items from the mean, *i.e.*, find $(X - \overline{X})$. Denote these deviations by x.
- Square these deviations and obtain the total Σx^2.
- Divide Σx^2 by the total number of observations, *i.e.*, N and extract the square-root. This gives us the value of standard deviation.

Deviations taken from Assumed Mean. When the actual mean is in fractions, say, it is 123.674 it would be too cumbersome to take deviations from it and then obtain squares of these deviations. In such a case either the mean may be approximated or else the deviations be taken from an assumed mean and the necessary adjustment made in the value of the standard deviation. The former method of approximation is less accurate and, therefore, invariably in such a case deviations are taken from assumed mean.

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

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$

Steps.

- Take the deviations of the items from an assumed mean, *i.e.*, obtain $(X - A)$. Denote these deviations by d. Take the total of these deviations, *i.e.*, obtain Σd.
- Square these deviations and obtain the total Σd^2.
- Substitute the values of Σd^2, Σd and N in the above formula.

Illustration 9. Blood serum cholesterol levels of 10 persons are as under:

240, 260, 290, 245, 255, 288, 272, 263, 277, 251.

Calculate standard deviation with the help of assumed mean.

Solution. CALCULATION OF STANDARD DEVIATION BY THE ASSUMED MEAN METHOD

X	$(X - 264)$* d	d^2
240	−24	576
260	− 4	16
290	+26	676
245	−19	361
255	− 9	81
288	+24	576
272	+ 8	64
263	− 1	1
277	+13	169
251	−13	169
$\Sigma X = 2641$	$\Sigma d = +1$	$\Sigma d^2 = 2689$

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$

$$\Sigma d^2 = 2689, \ \Sigma d = +1, \ N = 10$$

$$\therefore \quad \sigma = \sqrt{\frac{2689}{10} - \left(\frac{1}{10}\right)^2}$$

$$= \sqrt{268.9 - 0.01} = 16.398.$$

Illustration 10. Calculate the standard deviation from the following observations :

240.12	240.13	240.15	240.12	240.17
240.15	240.17	240.16	240.22	240.21

(B.Sc., Madras Univ.)

* The assumed mean should be as nearer to the actual mean possible to minimise calculations. In this case the actual mean is 264.1 and we have taken 264 as assumed mean.

Solution.

CALCULATION OF STANDARD DEVIATION

X	$(X-240)$ d	d^2
240.12	+0.12	.0144
240.13	+0.13	.0169
240.15	+0.15	.0225
240.12	+0.12	.0144
240.17	+0.17	.0289
240.15	+0.15	.0225
240.17	+0.17	.0289
240.16	+0.16	.0256
240.22	+0.22	.0484
240.21	+0.21	.0441
$N = 10$	$\Sigma d = +1.60$	$\Sigma d^2 = 0.2666$

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2} = \sqrt{\frac{0{\cdot}2666}{10} - \left(\frac{1{\cdot}6}{10}\right)^2} = \sqrt{0{\cdot}02666 - 0{\cdot}0256} = 0{\cdot}033$$

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.
3. Step deviation method.

(a) *Actual Mean Method.* When this method is applied, deviations are taken from the actual mean, *i.e.*, we find $(X - \overline{X})$ and denote these deviations by x. These deviations are then squared and multiplied by the respective frequencies. The following formula is applied :

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N}}, \text{ where } x = (X - \overline{X})$$

However, in practice this method is rarely used because if the actual mean is in fractions the calculations take a lot of time.

(b) *Assumed Mean Method.* When this method is used, the following formula is applied :

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}, \text{ where } d = (X - A).$$

Steps.

- Take the deviations of the items from an assumed mean and denote these deviations by d.
- Multiply these deviations by the respective frequencies and obtain the total, Σfd.
- Obtain the squares of the deviations, *i.e.*, calculate d^2.
- Multiply the squared deviations by the respective frequencies, and obtain the total, Σfd^2.

- Substitute the values in the above formula.

Illustration 11. Calculate the standard deviation from the data given below :

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

(B. Com., Calicut Univ., 1996)

Solution. CALCULATION OF STANDARD DEVIATION

X *Size of item*	f	$(X-6.5)$ d	fd	fd^2
3.5	3	−3	−9	27
4.5	7	−2	−14	28
5.5	22	−1	−22	22
6.5	60	0	0	0
7.5	85	+1	+ 85	85
8.5	32	+2	+ 64	128
9.5	8	+3	+ 24	72
	$N = 217$		$\Sigma fd = +128$	$\Sigma fd^2 = 362$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

$$\Sigma fd^2 = 362, \quad \Sigma fd = 128, \quad N = 217$$

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

$$= \sqrt{1.668 - .348} = 1{\cdot}149$$

(c) *Step Deviation Method.* When this method is used we take deviations of midpoints from an assumed mean and divide these deviations by the width of class interval, *i.e.*, 'i'. In case class intervals are unequal, we divide the deviations of midpoints by the lowest common factor and use 'c' instead of 'i' in the formula for calculating standard deviation. The formula for calculating standard deviation is :

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i$$

where, $d = \frac{(X - A)}{i}$ and i = class interval.

The use of the above formula simplifies calculations.

Illustration 12. The annual salaries of a group of employees are given in the following table:

Salaries (in Rs. 000)	45	50	55	60	65	70	75	80
Number of persons	3	5	8	7	9	7	4	7

Calculate the standard deviation of the salaries. *(B. Com., Kerela Univ., 1993)*

Solution.

CALCULATION OF STANDARD DEVIATION

Salaries X	No. of persons f	$(X-60)/5$ d	fd	fd^2
45	3	−3	−9	27
50	5	−2	−10	20
55	8	−1	−8	8
60	7	0	0	0
65	9	+1	+9	9
70	7	+2	+14	28
75	4	+3	+12	36
80	7	+4	+28	112
	$N = 50$		$\Sigma fd = 36$	$\Sigma fd^2 = 240$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i \qquad \text{Here, } \Sigma fd^2 = 240,\ N = 50,\ \Sigma fd = 36,\ i = 5$$

$$\sigma = \sqrt{\frac{240}{50} - \left(\frac{36}{50}\right)^2} \times 5 = \sqrt{4.8 - .5184} \times 5 = 10.35$$

Calculation of Standard Deviation—Continuous Series. In continuous series any of the methods discussed above for discrete frequency distribution can be used. However, in practice it is the step deviation method that is most used. The formula is

$$= \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i$$

$$\text{where } d = \frac{(m-A)}{i},\ i = \text{class interval}$$

Steps.

- Find the mid-points of various classes.
- Take the deviations of these mid-points from an assumed mean and denote these deviations by d.
- Wherever possible take a common factor and denote this column by d.
- Multiply the frequencies of each class with these deviations and obtain Σfd.
- Square the deviations and multiply them with the respective frequencies of each class and obtain Σfd^2.

Thus the only difference in procedure in case of continuous series is to find mid-points of the various classes.

Illustration 13. Calculate mean and standard deviation of following frequency distribution of marks :

Marks	No. of Students	Marks	No. of Students
0—10	5	40—50	50
10—20	12	50—60	37
20—30	30	60—70	21
30—40	45		

(B.A. & B.Sc., Guwahati Univ., B.Com., Calicut Univ., 1994)

Solution. CALCULATION OF MEAN AND STANDARD DEVIATION

Marks	*m.p.* *m*	*f*	*(m − 35)/10* *d*	*fd*	*fd*²
0—10	5	5	−3	−15	45
10—20	15	12	−2	−24	48
20—30	25	30	−1	−30	30
30—40	**35**	45	0	0	0
40—50	45	50	+1	+50	50
50—60	55	37	+2	+74	148
60—70	65	21	+3	+63	189
		$N = 200$		$\Sigma fd = 118$	$\Sigma fd^2 = 510$

$$\overline{X} = A + \frac{\Sigma fd}{N} \times i = 35 + \frac{118}{200} \times 10 = 35 + 5.9 = 40.9$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times 10 = \sqrt{\frac{510}{200} - \left(\frac{118}{200}\right)^2} \times 10$$

$$= \sqrt{2.55 - .3481} \times 10 = 1.4839 \times 10 = 14.839.$$

Illustration 14. Find the standard deviation from the following data :

Age under :	10	20	30	40	50	60	70	80
No. of persons dying :	15	30	53	75	100	110	115	125

(B. Com., Kerala Univ., B. Com.; Nagarjuna Univ.; B.Com., Bangalore Univ., 1996)

Solution. CALCULATION OF STANDARD DEVIATION

Age	*f*	*m.p.* *m*	*(m − 35)/10* *d*	*fd*	*fd*²
0—10	15	5	−3	−45	135
10—20	15	15	−2	−30	60
20—30	23	25	−1	−23	23
30—40	22	**35**	0	0	0
40—50	25	45	+1	+25	25
50—60	10	55	+2	+20	40
60—70	5	65	+3	+15	45
70—80	10	75	+4	+40	160
	$N = 125$			$\Sigma fd = 2$	$\Sigma fd^2 = 488$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{488}{125} - \left(\frac{2}{125}\right)^2} \times 10$$

$$= \sqrt{3.904 - .0003} \times 10 = 1.976 \times 10 = 19.76.$$

Illustration 15. Find the standard deviation of the following distribution :

Age :	20—25	25—30	30—35	35—40	40—45	45—50
No. of persons :	170	110	80	45	40	35

Take assumed average = 32.5*.

[B.A. (H) Eco. Delhi Univ.]

* When we are asked in the question to take a specified value as assumed mean we should take deviations only from that value.

Solution.

CALCULATION OF STANDARD DEVIATION

Age	m.p. m	No. of persons f	$(m - 32.5)/5$ d	fd	fd^2
20–25	22.5	170	–2	–340	680
25–30	27.5	110	–1	–110	110
30–35	32.5	80	0	0	0
35–40	37.5	45	+1	+45	45
40–45	42.5	40	+2	+80	160
45–50	47.5	35	+3	+105	315
		$N = 480$		$\Sigma fd = -220$	$\Sigma fd^2 = 1310$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{1310}{480} - \left(\frac{-220}{480}\right)^2} \times 5$$

$$= \sqrt{2.729 - .21} \times 5 = \sqrt{2.519} \times 5 = 1.587 \times 5 = 7.936.$$

Mathematical Properties of Standard Deviation Standard deviation has some very important mathematical properties which considerably enhance its utility in statistical work.

1. **Combined Standard Deviation** Just as it is possible to compute combined mean of two or more than two groups, similarly we can also compute combined standard deviation of two or more groups. Combined standard deviation is denoted by σ_{12} and is computed as follows :

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

σ_{12} = combined standard deviation;

σ_1 = standard deviation of first group;

σ_2 = standard deviation of second group;

$d_1 = |\overline{X}_1 - \overline{X}_{12}|$; $d_2 = |\overline{X}_2 - \overline{X}_{12}|$;

The above formula can be extended to find out the standard deviation of three or more groups. For example, combined standard deviation of three groups would be :

$$\sigma_{123} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1d_1^2 + N_2d_2^2 + N_3d_3^2}{N_1 + N_2 + N_3}}$$

where, $d_1 = |X_1 - X_{123}|$; $d_2 = |X_2 - X_{123}|$: $d_3 = |X_3 - X$

Illustration 16. The following are some of the particulars of the distribution of weight of boys and girls in a class

	Boys	Girls
Number	100	50
Mean weight	60 kg	45 kg
Variance	9	4

(a) Find the standard deviation of the combined data

(b) Which of the two distributions is more variable ? *(B.Com., MD. Univ. 1997)*

Solution.

(a) Combined S.D. $\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$

For finding combined standard deviation, we have to calculate combined mean.

$$\bar{X}_{12} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2}{N_1 + N_2}$$

$$= \frac{100\,(60) + 50\,(45)}{100 + 50} = \frac{6000 + 2250}{150} = 55$$

$N_1 = 100$, $\sigma_1^2 = 9$, $N_2 = 50$, $\sigma_2^2 = 4$, $d_1 = |\bar{X}_1 - \bar{X}_{12}| = 60 - 55 = 5$

$d_2 = |\bar{X}_2 - \bar{X}_{12}| = |45 - 55| = 10.$

Substituting the values $\sigma_{12} = \sqrt{\frac{100\,(9) + 50\,(4) + 100\,(5)^2 + 50\,(10)^2}{150}}$

$$= \sqrt{\frac{900 + 200 + 2500 + 5000}{150}} = \sqrt{\frac{8600}{150}} = 7.57$$

(*b*) For finding which distribution is more variable compare the coefficient of variation of two distributions :

$$\text{C.V. (Boys)} = \frac{\sigma}{\bar{X}} \times 100 = \frac{3}{60} \times 100 = 5.00$$

$$\text{C.V. (Girls)} = \frac{\sigma}{\bar{X}} \times 100 = \frac{2}{45} \times 100 = 4.44$$

Since coefficient of variation is more for distribution of weight of boys hence this distribution shows greater variability.

Illustration 17. The number of workers employed, the mean wages (in Rs.) per month and standard deviation (in Rs.) in each section of a factory are given below. Calculate the mean wages and standard deviation of all the workers taken together.

Section	*No. of workers employed*	*Mean wages in Rs.*	*Standard deviation in Rs.*
A	50	1113	60
B	60	1120	70
C	90	1115	80

Solution.

$$\bar{X}_{123} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2 + N_3 \bar{X}_3}{N_1 + N_2 + N_3}$$

$$= \frac{(50 \times 1113) + (60 \times 1120) + (90 \times 1115)}{50 + 60 + 90}$$

$$= \frac{55{,}650 + 67{,}200 + 1{,}00{,}350}{200}$$

$$= \frac{2{,}23{,}200}{200} = \text{Rs. } 1116.$$

Combined standard deviation of three series :

$$\sigma_{123} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1 d_1^2 + N_2 d_2^2 + N_3 d_3^2}{N_1 + N_2 + N_3}}$$

$d_1 = |\bar{X}_1 - \bar{X}_{123}|$ or $|1113 - 1116| = 3$

$d_2 = |\bar{X}_2 - \bar{X}_{123}|$ or $|1120 - 1116| = 4$

$d_3 = |\bar{X}_3 - \bar{X}_{123}|$ or $|1115 - 1116| = 1$

$$\sigma_{123} = \sqrt{\frac{50\,(60)^2 + 60\,(70)^2 + 90\,(80)^2 + 50\,(3)^2 + 60\,(4)^2 + 90\,(1)^2}{50 + 60 + 90}}$$

$$= \sqrt{\frac{1{,}80{,}000 + 2{,}94{,}000 + 5{,}76{,}000 + 450 + 960 + 90}{200}}$$

$$= \sqrt{\frac{10{,}51{,}500}{200}} = \sqrt{5{,}257.5} = 72.51$$

Mean : $\bar{X} = A + \frac{\Sigma fd}{N} \times i = 60 - \frac{26}{80} \times 5 = 60 - 1.625 = 58.375$

S.D. : $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{352}{80} - \left(\frac{-26}{80}\right)^2} \times 5$

$= \sqrt{4.4 - .106} \times 5 = 2.072 \times 5 = 10.36$

$$\text{C.V.} = \frac{\sigma}{\bar{X}} \times 100 = \frac{10.36}{58.375} \times 100 = 17.75\%.$$

Illustration 20. From the prices of shares of X and Y below find out which is more stable in value :

X	35	54	52	53	56	58	52	50	51	49
Y	108	107	105	105	106	107	104	103	104	101

(B. Com., Bangalore Univ., 1998)

Solution. In order to find out which shares are more stable, we have to compare coefficient of variations.

CALCULATION OF COEFFICIENT OF VARIATION

X	$(X-\bar{X})$ x	x^2	Y	$(Y-\bar{Y})$ y	y^2
35	−16	256	108	+3	9
54	+3	9	107	+2	4
52	+1	1	105	0	0
53	+2	4	105	0	0
56	+5	25	106	+1	1
58	+7	49	107	+2	4
52	+1	1	104	−1	1
50	−1	1	103	−2	4
51	0	0	104	−1	1
49	−2	4	101	−4	16
$\Sigma X = 510$	$\Sigma x = 0$	$\Sigma x^2 = 350$	$\Sigma Y = 1050$	$\Sigma y = 0$	$\Sigma y^2 = 40$

Coefficient of Variation X :

$$\text{C.V.} = \frac{\sigma}{\bar{X}} \times 100$$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{510}{10} = 51$$

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{350}{10}} = 5.916$$

$$\text{C.V.} = \frac{5.916}{51} \times 100 = 11.6$$

Coefficient of Variation Y :

$$\text{C.V.} = \frac{\sigma}{\bar{Y}} \times 100$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{1050}{10} = 105$$

$$\sigma = \sqrt{\frac{\Sigma y^2}{N}} = \sqrt{\frac{40}{10}} = 2$$

$$\text{C.V.} = \frac{2}{105} \times 100 = 1.905$$

Since coefficient of variation is much less in case of shares Y, hence they are more stable in value.

Illustration 21. Goals scored by two teams in a Football session were as follows :

No. of Goals Scored in a Football Match	No. of Football Matches Played	
	Team 'A'	*Team 'B'*
0	15	20
1	10	10
2	07	05
3	05	04
4	03	02
5	02	01
Total	42	42

Calculate coefficient of variation and state which team is more consistent.

(B. Com., M.D. Univ., 1996)

Solution. In order to find out which team is more consistent we shall have to compare the coefficient of variation.

X	$(X-7)$ x	x^2	Y	$(Y-7)$ y	y^2
15	+8	64	20	+13	169
10	+3	9	10	+3	9
7	0	0	5	−2	4
5	−2	4	4	−3	9
3	−4	16	2	−5	25
2	−5	25	1	−6	36
$\Sigma X = 42$	$\Sigma x = 0$	$\Sigma x^2 = 118$	$\Sigma Y = 42$	$\Sigma y = 0$	$\Sigma y^2 = 252$

Team A

$$\text{C.V.} = \frac{\sigma}{\overline{X}} \times 100$$

$$\overline{X} = \frac{\Sigma X}{N} = \frac{42}{6} = 7$$

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{118}{6}} = 4.43$$

$$\text{C.V.} = \frac{4.43}{7} \times 100 = 63.29$$

Team B

$$\text{C.V.} = \frac{\sigma}{\overline{Y}} \times 100$$

$$\overline{Y} = \frac{\Sigma Y}{N} = \frac{42}{6} = 7$$

$$\sigma = \sqrt{\frac{\Sigma y^2}{N}} = \sqrt{\frac{252}{6}} = 6.48$$

$$\text{C.V.} = \frac{6.48}{7} \times 100 = 92.57$$

Illustration 22. Two brands of tyres are tested with the following results:

Life (in'000 miles)	*No. of tyres brand*	
	X	Y
20—25	1	0
25—30	22	24
30—35	64	76
35—40	10	0
40—45	3	0

(a) Which brand of tyres have greater average life ?

(b) Compare the variability and state which brand of tyres would you use on your fleet of trucks ?

(B.Com., Bangalore Univ., 1997)