Measures of Dispersion

Dispersion and Types of Dispersion

Objective

After going through this lesson, you shall be able to understand the following concepts.

- Dispersion- Meaning and Need
- Types of Measures of Dispersion
    - Absolute Measure of Dispersion
    - Relative Measure of Dispersion

Introduction

In the previous few lessons, we studied about the central tendency and its measures which indicate a central value or a representative value for a series or a set of data. In other words, a single value is assumed to represent all the values of the series. However, these measures of central tendency do not reflect the distribution of various values in the series.

The values of different items in the series differ from the central value. While some of the values are close to the average value on the other hand, some others can show high divergence. For example, suppose a study of average marks is conducted in two classes. It is found that both the classes have the same mean marks.

However, while in one of the classes, the marks of different students did not vary much. On the other hand, in the other class, the marks of students showed higher variability. Thus, to judge how well the central value represents the values in the data, a study of the variability is important. The variability in the data is studied with the help of the measures of dispersion.

Dispersion

Dispersion measures the extent to which the different items tend to disperse away from the central tendency. In other words, while central tendency indicates a representative value, the measures of dispersion indicate the divergence of the values of different item from the central value. Thus, measures of dispersion provide a complete and comprehensive picture of the statistical series.

Need to Study Dispersion

The following points highlight the needs and importance of the study of dispersion.

1. Reliability of average: Dispersion measures the extent to which the values of different items in the series differ from the central value. By measuring the extent of variability it can be judged how well a central tendency is able to represent the values in the series. A high value of dispersion indicates that the central value does not appropriately represent the values in the series. On the other hand, a low value of dispersion indicates that the central value is a good representative of the data. Thus, a measure of dispersion helps in judging the reliability of the central value.

2. Comparison of different series: Measures of dispersion facilitates comparison of different series with regard to uniformity and homogeneity. A low dispersion indicates high uniformity and vice-versa. Measures of dispersion are used to assess variables such as profits of a firm, value of shares, etc. For example, a low dispersion in the data for profits over different years indicates consistency in the profit earnings of the firm over the years.

3. To control variability: Measures of dispersion helps in identifying the variability in the data. Accordingly, appropriate steps can be taken to control the variability. For example, an identification of the variability in the quality of the production can act as a guide for the firm to take appropriate measures.

4. Helpful in using other statistical measures: Measures of dispersion is not only helpful in measuring the variability in a given data but infact is useful in other areas also. In other words, it acts as a base in the calculation of higher order statistical measures such as correlation, regression, etc. and in this way it becomes important to study dispersion.

Types of Measures of Dispersion

Measures of dispersion can be broadly classified into the following two categories.

i. Absolute Measure of Dispersion

ii. Relative Measure of Dispersion

(i) Absolute measure of dispersion

Absolute measures of dispersion refers to those measures that are expressed in terms of original unit of series. For example, if the dispersion in the series for income is expressed in rupees then, it refers to the absolute dispersion. Such measures facilitate the comparison of variability in two or more series that are expressed in same units. Some of the absolute measures of dispersion are range, quartile deviation, mean deviation and standard deviation.

(ii) Relative measure of dispersion

Relative measures of dispersion refers to those measures that expresses the variability of data in relative value or percentage. With the help of relative measures a comparison of the dispersion can be made for two or more series that are expressed in different units.

For example, in order to compare the variability in the income of workers in India (expressed in rupees) and the variability in the income of the workers in USA (expressed in dollars), relative ,measures of dispersion are used. Some of the relative measures of dispersion are coefficient of range, coefficient of quartile deviation, coefficient of mean deviation and coefficient of variation.

The difference between the absolute measures of dispersion and relative measures of dispersion can be summarised in the following way:

Absolute Measures	**Relative Measures**
A dispersion expressed in terms of original unit of series is called absolute dispersion.	A dispersion that expresses the variability of data in relative value or percentage is called relative dispersion.
For example: range, quartile deviation, mean deviation, standard deviation.	For example: coefficient of range, coefficient of quartile deviation, coefficient of mean deviation, and coefficient of variation

Absolute Measures of Dispersion-Range, Quartile Deviation and Lorenz Curve

Objective

After going through this lesson, you shall be able to understand the following concepts.

- Range and Coefficient of Range
- Measurement of Range in
    - Individual Series
    - Discrete Series
    - Continuous Series
- Merits and Demerits of Range
- Quartile Deviation
- Measurement of Quartile Deviation in

    - Individual Series
    - Discrete Series
    - Continuous Series
- Merits and Demerits of Quartile Deviation
- Lorenz Curve

Range

Range refers to the difference between the highest value and the lowest value in a series. In other words, range measures the spread of the data in the series.

Algebraically,

$R = H – L$
Here,
R represents range
H represents the highest value
L represents the lowest value

Let us understand the measurement of range.

Measurement of Range in Individual Series

Example: Calculate range for the following data.

Weight (kg)	30	40	58	60	45	37	49	27

Solution: Here,

Highest Value (H) = 60

Lowest Value (L) = 27

$R = H – L$

R = 60 – 27 = 33

Thus, range is 33 kg

Measurement of Range in Discrete Series

The measurement of range in case of discrete series remains the same as that in individual series. In other words, it is calculated as the difference of the highest value and the lowest value without taking account of the respective frequencies.

Example- For the following data, calculate the range.

Marks	10	20	30	40	50	60
Number of Students	2	5	12	15	4	5

Solution

Here, for the calculation of range, the number of students corresponding to marks (frequency) will not be taken into account.

Highest value (H) = 60
Lowest value (L) = 10

Range = Highest Value – Lowest Value
= 60 – 10
= 50

Measurement of Range in Continuous Series

In the case of continuous series, range can be calculated by using either of the following two methods.

Method 1: Range is the difference of the upper limit of the highest class interval and the lower limit of the lowest class interval.

Range= Upper Limit of the Highest Class Interval - Lower Limit of the Lowest Class Interval

Method 2: Range is the difference of the mid value of the highest class and that of the lowest class.

Range= Mid-Value of the Highest Class Interval - Mid-Value of the Lowest Class Interval

Example: For the following data calculate the range.

X	f
0-10	10
10-20	15
20-30	11
30-40	17

40-50	13

Solution:

Method 1: Range = Upper Limit of the Highest Class Interval - Lower Limit of the Lowest Class Interval

= 50 – 0 = 50

Alternatively,

Method 2: Range = Mid-Value of the Highest Class Interval - Mid-Value of the Lowest Class Interval

$$\text{Mid-value of the Highest class interval} = \frac{50+40}{2} = 45$$

$$\text{Mid-value of the Lowest class interval} = \frac{10+0}{2} = 5$$

Range = 45 – 5= 40

Note: In case the data is given in inclusive series, the series is first converted into exclusive series.

Example- The following data presents the marks of 66 students in a class. Calculate range.

Marks	Number of Students
1-5	5
6-15	8
16-20	9
21-25	10
26-30	17
31-35	15
36-40	2

Solution

Here the data is given in inclusive series. To calculate range, we will have to first convert the data in exclusive series as follows.

Marks	Number of Students
0.5-5.5	5
5.5-15.5	8

15.5-20.5	9
20.5-25.5	10
25.5-30.5	17
30.5-35.5	15
35.5-40.5	2

Range = Upper limit of highest class interval – Lower limit of lowest class interval
= 40.5 – 0.5 = 40

Merits of Range

The following are the merits of range as a measure of dispersion.

i. ***Rigidly defined***: The value and method for calculations of range is very well and rigidly defined.

ii. ***Easy to calculate***: Range is very easy to calculate. That is, just by subtracting the lowest value from the highest value.

iii. ***Simple to understand***: The range is simple and easy to understand.

iv. ***Unaffected by units***: It gives the measures of dispersion in the same units that of the variable.

v. ***Unaffected by frequencies***: In calculation of range in case of discrete and continuous series, frequencies are not taken into account.

Demerits of Range

i. ***Not based on all observations***: Range takes into account only the highest value and the lowest value in the series. So, a change in the maximum and the minimum values can affect the value of Range to a large extent.

ii. ***Independent of measures of central tendency***: The value of the range is independent of the measures of central tendency such as mean, median and mode.

iii. ***Affected by sampling fluctuations***: While calculating range, the sampling fluctuations affect its estimation and calculation.

iv. ***Not used in open-ended series***: Range cannot be used in case of open-ended series.

Uses of Range

i. ***Forecast weather conditions***: Range is used by meteriological department to forecast weather conditions such as temperature, rainfall, etc.

ii. ***Check quality variations***: Range is used to study and check the quality variations in the manufactured products. If the quality of the product is within the prescribed range then, the product quality is under control.

iii. ***Study fluctuations in economic variables***: Range is used to study fluctuations in various economic variables such as prices, exchange rates etc.

iv. ***Measure variability***: Range is used to measure variability in of wide number of variables such as wages, sales etc. that are used in our daily life.

Quartile Deviation

Quartile deviation is the half of the difference between upper quartile and lower quartile. Algebraically,

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

Here,

Q.D. represents Quartile Deviation

Q_3 represents third quartile

Q_1 represents first quartile

Quartile Deviation is an absolute measure of dispersion and does not take into consideration all the items of the series and is thus known as the methods of limits.

The value of the quartile deviation is unaffected by the extreme values in the series.

However, the relative measure corresponding to quartile deviation is called the 'Coefficient of quartile deviation',

$$\text{Coefficient of Quartile deviation} = \frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

The coefficient of quartile deviation is used for comparing the degree of dispersion among different series.

Calculation of Quartile Deviation in Individual Series

Example: The following data presents the food consumption (in 100 calories) for 9 workers. Calculate the Quartile Deviation.

Weight (kg)	15	17	20	23	24	26	27	29	30

Solution:

$$Q_1 = \text{Size of} \left(\frac{n+1}{4}\right)^{\text{th}} \text{item} = \text{Size of} \left(\frac{9+1}{4}\right)^{\text{th}} \text{item}$$

$$\text{so, } Q_1 = \text{Size of } 2.5^{\text{th}} \text{ item}$$

$$\text{or, } Q_1 = \text{Size of } 2^{\text{nd}} \text{ item} + 0.5 \left(\text{Size of } 3^{\text{rd}} \text{ item} - \text{Size of } 2^{\text{nd}} \text{ item}\right)$$

$$\text{or, } Q_1 = 17 + 0.5\,(20 - 17)$$

$$\Rightarrow Q_1 = 18.5$$

$$Q_3 = \text{Size of } 3\left(\frac{n+1}{4}\right)^{\text{th}} \text{item} = \text{Size of } 3\left(\frac{9+1}{4}\right)^{\text{th}} \text{item}$$

$$\text{so, } Q_3 = \text{Size of } 7.5^{\text{th}} \text{ item}$$

$$\text{or, } Q_3 = \text{Size of } 7^{\text{th}} \text{ item} + 0.5 \left(\text{Size of } 8^{\text{th}} \text{ item} - \text{Size of } 7^{\text{th}} \text{ item}\right)$$

$$\text{or, } Q_3 = 27 + 0.5\,(29 - 27)$$

$$\Rightarrow Q_3 = 28$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{28 - 18.5}{2} = 4.75$$

$$\text{Coefficient of Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{28 - 18.5}{28 + 18.5} = 0.204$$

Calculation of Quartile Deviation in Discrete Series

The calculation of quartile deviation in discrete series is same as that in the individual series.

Example: The following data presents the wages of 50 workers in a factory. Calculate the Quartile Deviation.

Wage (Rs in'000)	**Number of workers (*f*)**
5	7
16	8
7	15
18	10

9	4
10	6

Solution:

Wage (Rs in'000)	Number of workers (*f*)	Cumulative frequency (*c.f*)	
5	7	7	
7	15	22	– Q_1
9	4	26	
10	6	32	
16	8	40	– Q_3
18	10	50	

$$Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{\text{th}} \text{item}$$

$$\text{or, } Q_1 = \text{Size of } \left(\frac{50+1}{4}\right)^{\text{th}} \text{item}$$

$$\text{or, } Q_1 = \text{Size of } 12.75^{\text{th}} \text{ item}$$

12.75^{th} item corresponds to 22 in the cumulative frequency.

$$\therefore Q_1 = 7$$

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

$$\text{or, } Q_3 = \text{Size of } 3\left(\frac{50+1}{4}\right)^{\text{th}} \text{item}$$

$$\text{or, } Q_3 = \text{Size of } 38.25^{\text{th}} \text{ item}$$

38.25^{th} item corresponds to 40 in the cumulative frequency.

$$\therefore Q_3 = 16$$

$$\text{Quartile Deviation or Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{16 - 7}{2} = 4.5$$

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{16 - 7}{16 + 7} = 0.39$$

Calculation of Quartile Deviation in Continuous Series

Let us look at the calculation of Quartile Deviation with the help of the following example.

Example: For the following data calculate the Quartile Deviation.

X	***f***
0-10	10
10-20	17
20-30	13
30-40	11
40-50	9
50-60	10

Solution

X	***f***	***c.f***	
0-10	10	10	
10-20	17	27	– Q_1
20-30	13	40	
30-40	11	51	
40-50	9	60	– Q_3
50-60	10	70	

$$Q_1 = \text{size of } \left(\frac{N}{4}\right)^{th} \text{item} = \text{size of } \left(\frac{70}{4}\right)^{th} \text{item}$$

$$= 17.5^{th}\text{item}$$

Thus, the class Interval for Q_1 is $10-20$

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f}{f} \times i$$

$$= 10 + \frac{17.5-10}{17} \times 10$$

$$= 14.41$$

$$Q_3 = \text{size of } 3\left(\frac{N}{4}\right)^{th} \text{item} = 3\left(\frac{70}{4}\right)^{th} \text{item}$$

$$= 52.5^{th}\text{item}$$

Thus, Q_3 lies in class interval for $40-50$

$$Q_3 = l_1 + \frac{3\left(\frac{N}{4}\right) - c.f}{f} \times i$$

$$= 10 + \frac{52.5 - 51}{9} \times 10$$

$$= 41.66$$

$$\text{Quartile Diviation} = \frac{Q_3 - Q_1}{2}$$

$$= \frac{41.66 - 14.41}{2}$$

$$= \frac{27.25}{2} = 13.62$$

The following table summarises the formulas to calculate Quartile Deviation

Formulas to Calculate Quartile Deviation		
For Individual Series		
$Q.D. = \frac{Q_3 - Q_1}{2}$	$\text{Size of } \left(\frac{n+1}{4}\right)^{th} \text{ item}$ *n* = Number of observations	$\text{Size of } \left(\frac{3(n+1)}{4}\right)^{th} \text{ item}$ *n* = Number of observations
For Discrete Series		
$\frac{10-0}{2} = 5$	Locate the $Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{th} \text{ item}$ In the *CF* column and corresponding *x* value is Q_1 *N* = Sum of frequencies	Locate the $Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)^{th} \text{ item}$ in the *CF* column and corresponding *x* value is Q_3 *N* = Sum of frequencies
For Continuous Series		
$Q.D. = \frac{Q_3 - Q_1}{2}$	Locate the size of (N/4)th item in *CF* column and the value of Q_1 will lie in the corresponding class interval. $Q_1 = l_1 + \frac{\frac{N}{4} - CF}{f} \times i$ where, l_1= Lower limit of class interval *N* = Sum of frequencies *CF* = Cumulative frequency of the class preceding the Q_1 class	Locate the size of 3(N/4)th item in *CF* column and the value of Q_3 will lie in the corresponding class interval. $Q_3 = l_1 + \frac{3\left(\frac{N}{4}\right) - CF}{f} \times i$ where, l_1= Lower limit of class interval *N* = Sum of frequencies *CF* = Cumulative frequency of the class preceding the Q_3 class *i* = Class interval

	i = Class interval	

Merits of Quartile Deviation

i. ***Easy to calculate***: Quartile Deviation is very easy to calculate and simple to understand.

ii. ***Not much affected by extreme values***: The values of Quartile deviation is not much affected by the extreme values in the sample.

iii. ***Depend on only few values***: Quartile Deviation does not depend on all the values of the sample.

iv. ***Measured in same units:*** Quartile Deviation gives the measures of dispersion in same units that of the variable.

v. ***Calculated in open-ended series***: Quartile Deviation can also be calculated in open-ended series.

vi. ***Reliability***: Quartile Deviation is more superior and reliable as compared to range.

Demerits of Quartile Deviation

i. ***Cannot be used for further treatment***: Quartile deviation cannot be used for further mathematical treatment.

ii. ***Considers only few observations***: It considers only first 25% and the last 25% of the observations, so, it ignores roughly 50% of the observations.

iii. ***Inappropriate comparisons***: Quartile deviation does not provide enough result to conduct meaningful comparisons.

iv. ***Affected by sampling fluctuations***: The value of Quartile Deviation is affected by the change in the sampling fluctuations.

Lorenz Curve

Lorenz curve is a graphic method of measuring dispersion. This curve was first used by economic statistician Max O. Lorenz to measure the distribution of income. This curve measures the deviation from the actual distribution from the line of equal distribution. Greater the distance of the Lorenz curve from the line of equal distribution, greater is the inequality (dispersion).

Steps to prepare Lorenz Curve

***Step* 1**: Convert the given series into cumulative series.

***Step* 2**: After cumulating the frequencies and sizes of items, obtain the percentages for the respective cumulative data.

Step* 3**: Now, plot the percentages obtained in ***step 2 on a graph. Since we are taking percentages, start from 0 to 100 on both the axis.

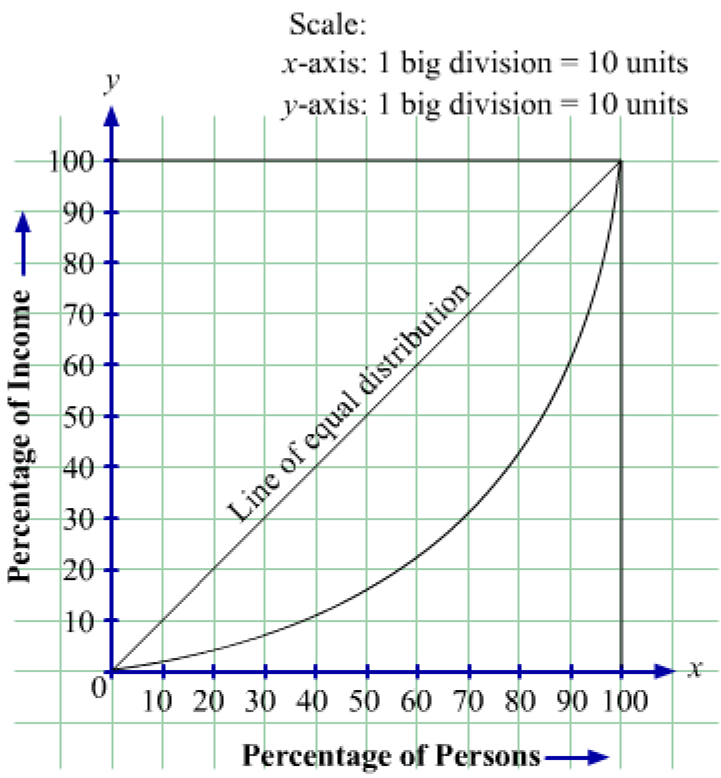
***Step* 4**: In the graph, draw a line of equal distribution is drawn at an angle of 45°.

***Step* 5**: Now, plot the percentages of the cumulated values of the given variable against the percentages of the corresponding cumulated frequencies and join the points to get a smooth curve. The curve obtained is the 'Lorenz Curve'.

Example: For the following data construct a Lorenz Curve.

Income (Rs in '000)	**No. of Persons**	***c.f* of Income**	***c.f* of Persons**	**Cumulative % of Income**	**Cumulative % of Persons**
10	7	10	7	5	14
15	15	25	22	12.5	44
30	8	55	30	27.5	60
35	5	90	35	45	70
50	9	140	44	70	88
60	6	200	50	100	100

Solution



Mean Deviation and its Calculations

Objective

After going through this lesson, you shall be able to understand the following concepts.

- **Mean Deviation**
- **Calculation of Mean Deviation in the following**
  - **Individual Series**
  - **Discrete Series**
  - **Continuous Series**

Introduction

In the previous lesson, we studied about two absolute measures of dispersion namely, Range and Quartile Deviation. These two measures of dispersion have two common limitations. First, they are not based on all the observations of the series. In other words, range is only based on the extreme values (highest value and the lowest value) while, quartile deviation is based on Q_1 and Q_3. In other words, all other observations in the series are ignored.

Second limitation of range and quartile deviation is that they do not reflect dispersion in the series around a central value (such as mean, median or mode). In this lesson, we will study about the other two measures of dispersion namely, mean deviation and standard deviation. These two measures of dispersion overcome the above said limitations.

Mean Deviation

Mean deviation is the arithmetic average of the deviations of all the values taken from some average value. This average value can either be mean, median or mode. Mean Deviation is also known as 'Average deviation'.

Calculation of Mean Deviation in Individual Series

The following steps are involved in the calculation of m**ean deviation in individual series.**

***Step* 1**: For the given series, calculate mean (or,median).

***Step* 2**: Take deviation of each observation form the mean (or, median), d = $|X - \overline{X}|$ or $|X - M|$. In the calculation of the deviations, the absolute value is taken while ignoring ± signs.

***Step* 3**: Obtain the sum of deviations obtained in ***step* 2** as $\Sigma|X - M|$.

Step* 4**: Divide the summation obtained in ***step 3 by the number of observations in the series.

$$MD_M = \frac{\sum|X - M|}{n} = \frac{\sum|D_M|}{n}$$

$$MD_{\overline{X}} = \frac{\sum|X - \overline{X}|}{n} = \frac{\sum|D_{\overline{X}}|}{n}$$

<u>Note</u>: In the calculation of deviations, all the deviations are taken as positive. The negative sign, if any, is ignored. This is because the sum of deviations taken from mean is zero. That is, if the negative sign is also considered in the calculation of deviation, the sum of deviations would be zero and thereby, the mean deviation would also be zero. Thus, it would lead to a wrong interpretation of the mean deviation.

Example: For the following data, calculate the Mean Deviation from-

i. Median

ii. Mean

X	12	15	16	18	19	20

Solution: Mean Deviation from Median

X	***|D|***
12	5
15	2
16	1
18	1
19	2
20	3
	Σ|*D*| = 14

$$\text{Median} = \frac{\text{Size of } \left(\frac{N}{2}\right)^{th} \text{item} + \text{Size of } \left(\frac{N}{2}+1\right)^{th}}{2}$$

$$= \frac{\text{Size of } \left(\frac{6}{2}\right)^{th} \text{item} + \text{Size of } \left(\frac{6}{2}+1\right)^{th}}{2}$$

$$= \frac{\text{Size of } 3^{rd}\text{item} + \text{Size of } 4^{th}}{2}$$

$$= \frac{16+18}{2} = \frac{34}{2} = 17$$

$$\text{Mean Deviation} = \frac{\Sigma|X-M|}{N} = \frac{\Sigma|D|}{N}$$

$$= \frac{14}{6} = 2.3$$

(ii) Mean Deviation from mean

X	***|D|***
12	4.66
15	1.66
16	0.66
18	1.34
19	2.34

20	3.34
	$\Sigma\|D\| = 14$

$$\text{Mean} = \frac{\Sigma X}{N} = \frac{100}{6} = 16.66$$

$$\text{Mean Deviation} = \frac{\Sigma|D|}{N} = \frac{14}{6} = 2.33$$

Calculation of Mean Deviation in Discrete Series

The following steps are involved in the calculation of m**ean deviation in discrete series.**

***Step* 1**: For the given series calculate mean (or, median).

***Step* 2**: Take deviation of each observation form the mean (or, median), *d* = $|X - \overline{X}|$ or $|X - M|$. In the calculation of the deviations the absolute values are taken ignoring ± signs.

Step* 3**: Multiply the deviations obtained in ***step 2 by their respective frequencies. Sum the column to obtain $\Sigma f|D|$.

***Step* 4**: Divide $\Sigma f|D|$ by the number of observations in the series. This gives the value of Mean Deviation.

$$MD_M = \frac{\Sigma f|X - M|}{N} = \frac{\Sigma f|D_M|}{N}$$

$$\text{or, } MD_{\overline{X}} = \frac{\Sigma f|X - \overline{X}|}{N} = \frac{\Sigma f|D_{\overline{X}}|}{N}$$

Example: Following data presents the marks of 20 students in a class. Calculate the mean deviation of marks from-

i. Median

ii. Mean

iii. Mode

X	**Frequency**
10	5

20	7
30	3
40	2
50	3

Solution

(i) *Mean Deviation from Median*

X	***f***	***c.f***	**From Median $\|D\|$**	***f|D|***
10	5	5	10	50
20	7	12	0	0
30	3	15	10	30
40	2	17	20	40
50	3	20	30	90
	N = 20		Σ|*D*| = 70	Σ*f*|*D*| = 210

$$M = \text{Size of} \left(\frac{N+1}{2}\right)^{\text{th}} \text{item}$$
$$= \text{Size of} \left(\frac{20+1}{2}\right)^{\text{th}} \text{item}$$
$$= 10.5^{\text{th}} \text{item}$$

The value of 10.5^{th} item is 20

Therefore, median = 20

$$\text{Mean Deviation from Median } MD_M = \frac{\Sigma f|D_M|}{N} = \frac{210}{20} = 10.5$$

(ii) *Mean Deviation form Mean*

X	***f***	***c.f***	***fx***	**From Mean $\|D\|$**	***f|D|***
10	5	5	50	15.5	77.5
20	7	12	140	5.5	38.5
30	3	15	90	4.5	13.5
40	2	17	80	14.5	29
50	3	20	150	24.5	73.5
	N = 20		Σ*fx* = 510		232

$$\text{Mean} = \frac{\Sigma X}{N} = \frac{510}{20} = 25.5$$

$$\text{Mean Deviation from Mean, } MD_{\overline{X}} = \frac{\Sigma f \left|D_{\overline{X}}\right|}{N} = \frac{232}{20} = 11.6$$

(iii) ***Mean Deviation from Mode***

Mode in this series is 20. So, in this case, the deviations will be the same as that from median.

X	***f***	**From Mode** **|*D*|**	***f*|*D*|**
10	5	10	50
20	7	0	0
30	3	10	30
40	2	20	40
50	3	30	90
	$N = 20$	$\Sigma\|D\| = 70$	$\Sigma f\|D\| = 210$

Mode = 20

$$\text{Mean Deviation from Mode, } MD_Z = \frac{\Sigma \left|D_Z\right|}{N} = \frac{210}{20} = 10.5$$

Calculation of Mean Deviation in Continuous Series

***Step* 1**: Obtain the mid-values of the class interval.

***Step* 2**: Calculate mean or median for the series.

***Step* 3**: Take deviations of the mid-values form the mean or median ignoring ± signs.

***Step* 4**: Multiply the deviations obtained in ***Step* 3** with their respective frequencies. Add the column to obtain $f|D|$

***Step* 5**: Divide $\sum f|D|$ by the sum of frequencies (i.e. $\sum f$ or N). This gives the value of the mean deviation.

<u>Example</u>: For the following data calculate the mean deviation from:

i. Median

ii. Mode

iii. Mean

Class Interval	Frequency
0-10	5
10-20	9
20-30	6
30-40	3
40-50	2

Solution:

(i) *Mean Deviation from Median*

X	*f*	*c.f*	Mid-Point (*m*)	From Median |*D*|	*f*|*D*|
0-10	5	5	5	5.33	26.65
10-20	9	14	15	4.67	42.03
20-30	6	20	25	14.67	88.02
30-40	3	23	35	24.67	74.01
40-50	2	25	45	34.67	69.34
	N = 25			Σ|*D*| = 80.01	Σ*f*|*D*|= 300.05

$$\text{Median} = \text{Size of } \left(\frac{N}{2}\right)^{th} \text{ item}$$

$$= \text{Size of } \left(\frac{25}{2}\right)^{th} \text{ item} = 12.5$$

This lies in the class interval $(10-20)$

$$\text{Median} = l + \frac{\frac{N}{2} - c.f.}{f} \times i$$

10 + 12.5 − 59×10 = 18.3310 + 12.5 - 59×10 = 18.33

$$MD_M = \frac{\sum f|D_M|}{N} = \frac{300.05}{25} = 12.002$$

(ii) *Mean Deviation from Mode*

X	*f*	*c.f*	Mid-point (*m*)	From Mode |D|	*f*|D|
0 – 10	5	5	5	10.71	53.55

10 – 20	9	14	15	0.71	6.39
20 – 30	6	20	25	9.29	55.74
30 – 40	3	23	35	19.29	57.87
40 – 50	2	25	45	29.29	58.58
	N = 25			Σ|*D*| = 69.29	Σ*f*|*D*| = 232.13

Here, the highest frequency class interval is 10 – 20

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 10 + \frac{9-5}{18-5-6} \times 10$$

$$= 10 + \frac{4}{7} \times 10$$

$$= 15.71$$

$$\text{Mean Deviation from Mode} = \frac{\Sigma f|D|}{n} = \frac{232.13}{25} = 9.28$$

(iii) ***Mean Deviation from Mean***

X	***f***	**Mid-point (*m*)**	***f.m***	**From Mean |*D*|**	***f*|*D*|**
0 – 10	5	5	25	15.2	76
10 – 20	9	15	135	5.2	46.8
20 – 30	6	25	150	4.8	28.8
30 – 40	3	35	105	14.8	44.4
40 – 50	2	45	90	24.8	49.6
	N = 25		Σ*fm*= 505	64.8	245.6

$$\text{Mean} = \frac{\Sigma fm}{N}$$

$$= \frac{505}{25} = 20.2$$

$$MD_{\overline{X}} = \frac{\sum f|D_{\overline{X}}|}{N} = \frac{245.6}{25} = 9.82$$

The formula for the calculation of Mean deviation can be summarised as follows

	When mean deviation are taken from **Median** (*M*)	When mean deviation are taken from **Arithmetic Mean** $(\overline{X})$	When mean deviation are taken from **Mode (*Z*)**
For individual series	$MD_M = \frac{\sum\lvert X-M\rvert}{n} = \frac{\sum\lvert D_M\rvert}{n}$	$MD_{\overline{X}} = \frac{\sum\lvert X-\overline{X}\rvert}{n} = \frac{\sum\lvert D_{\overline{X}}\rvert}{n}$	$MD_Z = \frac{\sum\lvert X-Z\rvert}{n} = \frac{\sum\lvert D_Z\rvert}{n}$
For discrete and continuous series	$MD_M = \frac{\Sigma f\lvert X-M\rvert}{N} = \frac{\Sigma f\lvert D_M\rvert}{N}$	$MD_{\overline{X}} = \frac{\Sigma f\lvert X-\overline{X}\rvert}{N} = \frac{\Sigma f\lvert D_{\overline{X}}\rvert}{N}$	$MD_Z = \frac{\sum f\lvert X-Z\rvert}{N} = \frac{\sum f\lvert D_Z\rvert}{N}$

Merits of Mean Deviation

i. ***Easy to compute***: Mean deviation is very easy to compute as well as easy to understand.

ii. ***Gives a precise value***: The value obtained after calculation of mean deviation is very definite and precise.

iii. ***Based on all items***: Mean deviation is based on all the items in the series. Thus, it is a better measure than range and quartile deviation.

iv. ***Less affected by extreme values***: The value of mean deviation is not much affected by the extreme values in the series.

Demerits of Mean Deviation

i. ***Ignores signs***: The deviation calculated under mean deviation ignores positive and negative signs.

ii. ***Cannot be used for further treatment***: Mean deviation cannot be used for further mathematical treatment.

iii. ***Involves heavy calculation****:* Mean deviation tedious and lengthy process if the values of mean, median on mode is in fraction.

iv. ***Unreliable results***: Mean deviation gives unreliable results if it is calculated from mode. This is because mode is difficult to calculate.

Standard Deviation and its Calculations

Objective

After going through this lesson, you shall be able to understand the following concepts.

- **Standard Deviation**
- **Calculation of Standard Deviation in**
  - **Individual Series**
  - **Discrete Series**
  - **Continuous Series**

Standard Deviation

Standard deviation is the most commonly used measure of dispersion. This was first used by 'Karl Pearson'. Standard deviation is the square root of the mean of the square of deviation of items from the mean value. It is also known as the 'Root Mean Square Deviation'.

Standard deviation can be calculated using the following three methods.

i. Actual Mean Method

ii. Direct Method

iii. Assumed Mean Method

Calculation of Standard Deviation in Individual Series

(i) Direct method

The following steps are involved in the calculation of standard deviation using the direct method.

***Step* 1**: Calculate mean for the given series.

***Step* 2**: From the mean, obtain deviations of each item and denote it by *x.*

Step* 3**: Square the deviations obtained in ***step 2 and obtain their summation (Σx^2).

***Step* 4**: Divide (Σx^2) by the number of observations to obtain the standard deviation.

$$\sigma = \sqrt{\frac{\Sigma x^2}{n}} = \sqrt{\frac{\Sigma\left(X - \overline{X}\right)^2}{n}}$$

Example: The following table presents the wages of 5 workers in a particular department of a factory. Calculate standard deviation of the wages.

Wage (Rs in '000)	3	4	6	8	9

Solution

Wage (Rs in' 000) (*X*)	$(X-\overline{X})$	$(X-\overline{X})^2$
3	– 3	9
4	– 2	4
6	0	0
8	2	4
9	3	9
N = 30		26

Solution

Here,

$$\overline{X}=\frac{\Sigma X}{n}=\frac{30}{5}=6$$

Now, Standard Deviation is

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

$$\text{or, } \sigma=\sqrt{\frac{26}{5}}=2.28$$

(ii) ***Actual mean method***

The following steps are involved in the calculation of standard deviation using the actual mean method.

***Step* 1**: Calculate mean for the given series.

***Step* 2**: Square each observation in the series.

***Step* 3**: Use the following formula to calculate standard deviation:

$$\sigma = \sqrt{\frac{\Sigma X^2}{n} - \left(\overline{X}\right)^2} = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$$

Considering the example given above, let us calculate standard deviation using the actual mean method.

Solution

X	X^2
3	9
4	16
6	36
8	64
9	81
30	206

$$\sigma = \sqrt{\frac{\sum X^2}{n} - \left(\overline{X}\right)^2}$$

$$= \sqrt{\frac{206}{5} - (6)^2} = \sqrt{41.2 - 36}$$

$$= 2.28$$

(iii) ***Assumed mean method***

The following steps are involved in the calculation of standard deviation using the assumed mean method.

***Step* 1**: Take an assumed mean and calculate the deviations of each item from this.

Step* 2**: Square the deviations obtained in ***step 1 and obtain the summation as (Σd^2)

***Step* 3**: Use the following formula for calculating standard deviation:

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

Continuing with the same example, let us calculate the standard deviation using the assumed mean method.

Solution

X	$d = (X - A)$	d^2
3	– 3	9
4	– 2	4
A = 6	0	0
8	2	4
9	3	9
		$\Sigma d^2 = 26$

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

$$= \sqrt{\frac{26}{5} - \left(\frac{0}{5}\right)^2}$$

$$= \sqrt{5.2} = 2.28$$

The following table summarises the formula for the calculation of standard deviation in individual series.

For Individual Series		
Actual Mean Method	$\sigma = \sqrt{\frac{\Sigma x^2}{n}} = \sqrt{\frac{\Sigma\left(X - \overline{X}\right)^2}{n}}$	x^2 = Sum total of square of deviation $\left(X - \overline{X}\right) = x =$ Deviation from mean $\sigma =$ Standard Deviation $n =$ Number of observations
Direct Method	$\sigma = \sqrt{\frac{\Sigma X^2}{n} - \left(\overline{X}\right)^2} = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$	$\Sigma X^2 =$ Sum of the square of values of X
Assumed Mean Method/Short-cut Method	$\sigma = \sqrt{\frac{\Sigma d_x^2}{n} - \left(\frac{\Sigma d_x}{n}\right)^2}$	$d_x = X - A$ $\Sigma d_x^2 =$ Sum total of square of deviation taken from assumed value (A) $\Sigma d_x =$ Sum of deviation taken from assumed value $d_x^2 = (d_x) \times (d_x)$

Calculation of Standard Deviation in Discrete Series

In discrete series, standard deviation can calculated by using the following methods.

i. Actual mean method

ii. Direct method

iii. Assumed mean/short-cut method

iv. Step deviation method

(i) *Actual mean method*

The following steps are involved in the calculation of standard deviation in discrete series using the actual mean method.

***Step* 1**: Calculate mean for the given series.

***Step* 2**: From the mean, obtain deviations of each item and denoted by *x.*

***Step* 3**: Multiply the deviations obtained in ***step* 2** by their respective frequencies and obtain the total as (Σfx^2).

***Step* 4**: Use the following formula to calculate standard deviation.

$$\sigma = \sqrt{\frac{\Sigma fX^2}{N}} = \sqrt{\frac{\Sigma f\left(X-\overline{X}\right)^2}{N}}$$

Example: For the following series calculate the standard deviation using actual mean method.

X	***f***
5	7
10	12
15	19
20	11
25	8
30	3

Solution

X	***f***	***fX***	***x*** $X-\overline{X}$	$\boldsymbol{x^2}$ $\left(X-\overline{X}\right)^2$	$\boldsymbol{fx^2}$ $f\left(X-\overline{X}\right)^2$
5	7	35	– 10.83	117.28	820.96
10	12	120	– 5.83	33.98	407.76
15	19	285	– 0.83	0.68	12.92
20	11	220	4.17	17.38	191.18

25	8	200	9.17	84.08	672.64
30	3	90	14.17	200.78	602.34
	Σf = 60	Σfx = 950			Σfx^2 = 2,707.8

Solution

$$\overline{X} = \frac{\Sigma fX}{\Sigma f} = \frac{950}{60} = 15.83$$

$$\sigma = \sqrt{\frac{\Sigma f\left(X - \overline{X}\right)^2}{n}} = \sqrt{\frac{\Sigma fx^2}{n}}$$

$$\sigma = \sqrt{\frac{2707.8}{60}} = \sqrt{45.13}$$

$$= 6.72$$

(ii) *Direct method*

The following steps are involved in the calculation of standard deviation in discrete series using the direct method.

***Step* 1**: Calculate mean for the given series.

***Step* 2**: Multiply each observation in the series by their respective frequencies and obtain the summation as Σfx.

***Step* 3**: Obtain the square of each observation and multiply by respective frequencies and obtain the summation as Σfx^2 .

***Step* 4**: Use the following formula to calculate standard deviation:

$$\sigma = \sqrt{\frac{\Sigma fX^2}{N} - \left(\overline{X}\right)^2} = \sqrt{\frac{\sum fX^2}{N} - \left(\frac{\sum fX}{N}\right)^2}$$

Continuing with the same example, let us calculate the Standard Deviation using the direct method.

X	f	X^2	fX^2	fX
5	7	25	175	35
10	12	100	1,200	120
15	19	225	4,275	285
20	11	400	4,400	220

25	8	625	5,000	200
30	3	900	2,700	90
	$\Sigma f = 60$		ΣfX^2=17,750	ΣfX=950

$$\overline{X} = \frac{\Sigma fX}{\Sigma f} = \frac{950}{60} = 15.83$$

$$\sigma = \sqrt{\frac{\Sigma fX^2}{n} - \left(\overline{X}\right)^2}$$

$$= \sqrt{\frac{17,750}{60} - 250.58} = \sqrt{45.25}$$

$$= 6.72$$

(iii) ***Assumed mean method***

Under the assumed mean method, the standard deviation is calculated using the assumed mean. The following are the steps involved in the calculation of Standard Deviation using the assumed mean method.

***Step* 1**: Take an assumed mean and calculate the deviations of each item from this.

Step* 2**: Multiply the deviations obtained in ***step 1 by their respective frequencies and obtain the summation as Σfd.

***Step* 3**: obtain the square of deviations and multiply by respective frequencies and obtain the summation as Σfd^2

***Step* 4**: Use the following formula to calculate standard deviation:

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

Continuing once again with the same example let us compute the Standard deviation using assumed mean method.

X	***f***	***d* (*X* – *A*)**	d^2	fd^2	***fd***
5	7	– 10	100	700	– 70
10	12	– 5	25	300	– 60
A = 15	19	0	0	0	0
20	11	5	25	275	55

25	8	10	100	800	80
30	3	15	225	675	45
	N = 60			Σfd^2= 2,750	Σfd=50

$$\overline{X} = A + \frac{\Sigma fd}{N}$$
$$= 15 + \frac{50}{60}$$
$$= 15.83$$
$$\sigma = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2}$$
$$= \sqrt{\frac{2750}{60} - \left(\frac{50}{60}\right)^2}$$
$$= \sqrt{45.14} = 6.72$$

(iv) *Step-Deviation method*

The calculation procedure in the step deviation method remains the same as that in assumed men method. However, in step-deviation method we divide the deviations by a common factor to further simplify the calculations.

The formula for calculating the standard deviation under this method is as follows.

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

In the above given example, standard deviation using the step deviation using the step-deviation method is calculated as follows.

X	*f*	*d* (*X* – *A*)	*d´* (*d/h*)	*d´*^{2}	*fd´*	*fd´*^{2}
5	7	– 10	– 2	4	– 14	28
10	12	– 5	– 1	1	– 12	12
A = 15	19	0	0	0	0	0
20	11	5	1	1	11	11
25	8	10	2	4	16	32
30	3	15	3	9	9	27
	60				10	110

$$\sigma=\sqrt{\frac{\Sigma fd_X'^2}{N}-\left(\frac{\Sigma fd_X'}{N}\right)^2}\times h$$

$$=\sqrt{\frac{110}{60}-\left(\frac{10}{60}\right)^2}\times 5$$

$$=\sqrt{1.83-0.02}\times 5$$
$$=1.34\times 5=6.72$$

The following table summarises the formula for the calculation of Standard Deviation in discrete series. **For Discrete Series**

Actual Mean Method	$\sigma=\sqrt{\frac{\Sigma fX^2}{N}}=\sqrt{\frac{\Sigma f\left(X-\overline{X}\right)^2}{N}}$	$\left(X-\overline{X}\right)=x=$ Deviation from mean $\sigma=$ Standard Deviation $f=$ frequency $fx^2=(f)\times(x^2)$ $\overline{X}=\frac{\sum fX}{N}$ $N=\sum f=$ Sum of frequencies
Direct Method	$\sigma=\sqrt{\frac{\Sigma fX^2}{N}-\left(\overline{X}\right)^2}=\sqrt{\frac{\Sigma fX^2}{N}-\left(\frac{\Sigma fX}{N}\right)^2}$	$\overline{X}=\frac{\Sigma fX}{N}$
Assumed Mean Method/Short-cut Method	$\sigma=\sqrt{\frac{\sum fd^2}{n}-\left(\frac{\sum fd}{n}\right)^2}$	$d_x=X-A$ $\Sigma d_x^2=$Sum total of square of deviation taken from assumed value (A) $\Sigma d_x=$Sum of deviation taken from assumed value $fd_x^2=(fd_x)\times(d_x)$
Step Deviation Method	$\sigma=\sqrt{\frac{\Sigma fd_X'^2}{N}-\left(\frac{\Sigma fd_X'}{N}\right)^2}\times h$	$d_x'=\frac{X-A}{h}$, $h=$common factor $\Sigma d_x=$Sum of deviation taken from assumed value $fd_x'^2=(fd_x')\times(d_x')$ $\overline{X}=A+\frac{\Sigma fd_x'}{N}\times h$ $N=$Sum of frequencies

Calculation of Standard Deviation in Continuous Series

Like discrete series, under continuous series as well, the standard deviation in continuous series can be calculated using the following four methods:

i. Actual Mean Method

ii. Direct Method

iii. Assumed Mean/Short Cut Method

iv. Step Deviation Method

(i) Actual mean method

The following steps are involved in the calculation of standard deviation in continuos series using the actual mean method.

***Step* 1**: Calculate mean for the given series.

***Step* 2**: From the mean, obtain deviations of mid points of each class interval and denoted by *x.*

***Step* 3**: Multiply the deviations by their respective frequencies and denote by Σfx .

***Step* 4**: Obtain the square of deviations and multiply by their respective frequencies and obtain the total as (Σfx^2).

***Step* 5**: Use the following formula to calculate standard deviation:

$$\sigma = \sqrt{\frac{\Sigma fm^2}{N}} = \sqrt{\frac{\Sigma f\left(m - \overline{X}\right)^2}{N}}$$

The following example illustrates the given calculation procedure.

<u>Example</u>:

X	***f***
10 – 20	5
20 – 30	9
30 – 40	13
40 – 50	15
50 – 60	6
60 – 70	2

Solution

X	***f***	***m***	***fm***	$\left(m - \overline{X}\right)$ ***x***	***x²*** $\left(m - \overline{X}\right)^2$	***fx²*** $f\left(m - \overline{X}\right)^2$

10 – 20	5	15	75	– 22.8	519.84	2,599.2
20 – 30	9	25	225	– 12.8	163.84	1,474.56
30 – 40	13	35	455	– 2.8	7.84	101.92
40 – 50	15	45	675	7.2	51.84	777.6
50 – 60	6	55	330	17.2	295.84	1,775.04
60 – 70	2	65	130	27.2	739.84	1,479.68
			Σfm= 1,890			Σfx^2=8,208

$$\overline{X}=\frac{\Sigma fm}{n}=\frac{1,890}{50}=37.8$$

$$\sigma=\sqrt{\frac{\Sigma fx^2}{n}}=\sqrt{\frac{\Sigma f\left(m-\overline{X}\right)^2}{n}}$$

$$=\sqrt{\frac{8,208}{50}}=\sqrt{164.16}$$

$$=12.81$$

(ii) *Direct method*

The following are the steps involved in the calculation of standard deviation using the direct method.

***Step* 1**: Calculate mean for the given series.

***Step* 2**: Multiply mid-value of each class interval in the series by their respective frequencies and obtain the summation as Σfm.

***Step* 3**: Obtain the square of each observation and multiply by respective frequencies and obtain the summation as Σfm^2 .

***Step* 4**: Use the following formula to calculate standard deviation:

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

Considering the example given above. Let us calculate standard deviation using the direct method.

X	***f***	***m***	***m*²**	***fm***	***fm*²**

10 – 20	5	15	225	75	1,125
20 – 30	9	25	625	225	5,625
30 – 40	13	35	1,225	455	15,925
40 – 50	15	45	2025	675	30,375
50 – 60	6	55	3,025	330	18,150
60 – 70	2	65	4,225	130	8,450
	50			Σfm=1,890	Σfm^2=79,650

$$\overline{X}=\frac{\Sigma fm}{n}=\frac{1,890}{50}=37.8$$

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

$$=\sqrt{\frac{79,650}{50}-1,428.84}$$

$$=\sqrt{1,593-1,428.84}$$

$$=\sqrt{164.16}$$

$$=12.81$$

(iii) *Assumed mean/Short cut method*

The following are the steps involved in the calculation of standard deviation using the assumed mean method.

***Step* 1**: Take an assumed mean and calculate the deviations of each item from this.

Step* 2**: Multiply the deviations obtained in ***step 1 by their respective frequencies and obtain the summation as Σfd.

***Step* 3**: obtain the square of deviations and multiply by respective frequencies and obtain the summation as Σfd^2

Step 4: Standard deviation is calculated using the following formula.

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

The following example makes the calculation procedure clearer.

X	***f***	***m***	**(*d*)** ***m – A***	***fd*** ***f*(*m – A*)**	***d*²**	***fd*²**
10-20	5	15	– 20	– 100	400	2,000
20-30	9	25	– 10	– 90	100	900
30-40	13	35 (*A*)	0	0	0	0
40-50	15	45	10	150	100	1,500
50-60	6	55	20	120	400	2,400
60-70	2	65	30	60	900	1,800
	50			Σfd=140		Σfd^2=8,600

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

$$= \sqrt{\frac{8,600}{50} - \left(\frac{140}{50}\right)^2}$$
$$= \sqrt{164.16}$$
$$= 12.81$$

(iv) *Step-Deviation method*

Steps involved in the calculation of standard deviation in the step-deviation method is the same as that in assumed mean method. However, to make the calculation procedure more simple, the deviations are divided by a common factor.

Considering again the example considered above, let us calculate the standard deviation using step-deviation method.

X	***f***	***m***	***d*** ***m*** ***– A***	***d´*** $\frac{m-A}{10}$	***fd´***	***d´²***	***fd´²***
10-20	5	15	– 20	– 2	– 10	4	20
20-30	9	25	– 10	– 1	– 9	1	9
30-40	13	(35) *A*	0	0	0	0	0
40-50	15	45	10	1	15	1	15

50-60	6	55	20	2	12	4	24
60-70	2	65	30	3	6	9	18
	50				*Σfd'*=14		*Σfd'*²=86

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

$$=\sqrt{\frac{86}{50}-\left(\frac{14}{50}\right)^2}\times 10$$

$$=\sqrt{1.6416}\times 10=12.81$$

The following table summarises the formula for the calculation of Standard Deviation in discrete series.

For Continuous Series

Actual Mean Method	$\sigma=\sqrt{\frac{\Sigma fm^2}{N}}=\sqrt{\frac{\Sigma f\left(m-\overline{X}\right)^2}{N}}$	$\left(m-\overline{X}\right)$=Deviation from mean σ=Standard Deviation f=frequency m=mid points of class intervals $\overline{X}=\frac{\Sigma fm}{N}$ N=Sum of frequencies
Direct Method	$\sigma=\sqrt{\frac{\Sigma fm^2}{N}-\left(\overline{X}\right)^2}=\sqrt{\frac{\Sigma fm^2}{N}-\left(\frac{\Sigma fm}{N}\right)^2}$	$\overline{X}=\frac{\Sigma fm}{N}$
Assumed Mean Method/Short-cut Method	$\sigma=\sqrt{\frac{\Sigma fd^2_m}{N}-\left(\frac{\Sigma fd_m}{N}\right)^2}$	$d_m = m-A$ $\Sigma d_m{}^2 =$ Sum total of square of deviation taken from assumed value (A) $\Sigma d_m =$ Sum of deviation taken from assumed value $fd_m{}^2 = (fd_m)\times(d_m)$ $N =$ Sum of frequencies
Step Deviation Method	$\sigma=\sqrt{\frac{\Sigma fd'^2_m}{N}-\left(\frac{\Sigma fd'_m}{N}\right)^2}\times h$	$d'_m=\frac{X-A}{h}$, h = common factor

		$\Sigma d_m =$ Sum of deviation taken from assumed value $fd'^{2}_m = (fd'_m) \times (d'_m)$ $\overline{X} = A + \frac{\sum fd'_m}{N} \times h$ $N =$ Sum of frequencies

Merits of Standard Deviation

i. ***Based on all values***: The calculation of standard deviation is based on all the values of the sample.

ii. ***Gives a precise value***: Standard deviation gives a definite and precise value, therefore, is a well defined and definite measure of dispersion.

iii. ***Reliable measure***: Standard deviation depicts a clear picture and is thus more reliable then range. This is because it is based on mean that is an ideal average.

iv. ***Not much affected by change in values***: As it is based on all the values of the sample, so it is not much affected by the change in the sample value.

v. ***Can be used for further treatment***: Unlike mean deviation, standard deviation can be used for further mathematical treatment.

Demerits of Standard Deviation

i. ***Affected by extreme values***: Standard deviation assigns more weights to the extreme values.

ii. ***Tedious calculation process***: It involves a complex and tedious calculation process and is also hard to understand.

Difference between Mean Deviation and Standard Deviation

The following are the two main differences between Mean Deviation and Standard Deviation

1. In mean deviation, deviations can be calculated from either of the three central measures namely, mean, median and mode. However, in the calculation of Standard Deviation, deviations are calculated only from mean.

2. In the calculation of deviations in mean deviation, the positive and negative signs are ignored. However, they are not ignored in standard deviation.

Relative Measures of Dispersion

Objective

After going through this lesson, you shall be able to understand the following topics.

- Coefficient of Range
- Coefficient of Quartile Deviation
- Coefficient of Mean Deviation
- Coefficient of Variance
- Coefficient of Standard Deviation

Introduction

In the previous two lessons, we studied the absolute measures of dispersion namely, range, quartile deviation, mean deviation and standard deviation. In this lesson, we will study the relative measures of dispersion. As we know, a relative measure of dispersion expresses the variability of data in relative value or percentage. Thus, relative measures of dispersion are used for the purpose of comparison.

Coefficient of Range

Coefficient of range is a relative measure of range. It used for the comparison of variability of the distributions independent of the units of measurement. The coefficient of range is calculated using the following formula.

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

where,

H represents the highest value, and

L represents the lowest value

Example: The following data presents the wages of a group of 6 workers. Calculate the range and the coefficient of the range.

S. No	Wage (Rs in' 000)
1	5
2	7

3	10
4	12
5	13
6	15

Solution

Range = $H - L$

Here,

Highest value = 15

Lowest value = 5

Range = 15 – 5 = 10

$$\text{Coefficient of Range} = \frac{H-L}{H+L} = \frac{15-5}{15+5} = 0.5$$

<u>Example</u>: The following table presents the wages of 80 workers in a factory. Calculate the range and the coefficient of range.

Wages (Rs in' 000)	Number of Workers
5	12
7	19
10	23
12	17
13	6
15	3
	80

Solution

As we know that in the calculation of range in discrete series, the frequencies of the observations are not taken into account.

Range = $H - L$

Here,

Highest value = 15

Lowest value = 5

Range = 15 – 5 = 10

$$\text{Coefficient of Range} = \frac{H-L}{H+L} = \frac{15-5}{15+5} = 0.5$$

Example: The following data presents the marks of 54 students in a class. Calculate the range of marks and its coefficient.

Marks	Number of Students
10-20	14
20-30	17
30-40	12
40-50	11

Solution

Range = Upper Limit of the Highest Class Interval - Lower Limit of the Lowest Class Interval

= 50 – 10 = 40

$$\text{Coefficient of Range} = \frac{H-L}{H+L} = \frac{50-10}{50+10} = 0.66$$

Example: Calculate the value of the smallest item in the series if, the value of the highest item is 60 and coefficient of range is 0.50. Also calculate the range of the series.

Solution

We know,

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

$$\text{or, } 0.50 = \frac{60-L}{60+L}$$

$$\text{or, } 0.50\,(60+L) = (60-L)$$

$$\text{or, } 30 + 0.5L = 60 - L$$

$$\text{or, } L = 20$$

Hence, lowest value in the series is 20.

Now,

Range = Highest Value – Lowest Value

= 60 – 20

= 40

Coefficient of Quartile Deviation

Coefficient of Quartile Deviation is the relative measure of Quartile deviation. It is obtained by dividing quartile deviation by the average value of the first quartile and the third quartile. Algebraically,

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

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

Example: Given below are the marks of 6 students in a class. Calculate the quartile deviation and the coefficient of quartile deviation.

S. No	Marks
1	8
2	13
3	17
4	14
5	11
6	15

Solution

We know,

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

Now,

$$Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{th} \text{ item}$$
$$= \left(\frac{6+1}{4}\right)^{th} = 1.75^{th} \text{ item}$$
$$= 13$$
$$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)^{th} \text{ item}$$
$$= 3\left(\frac{6+1}{4}\right)^{th} \text{ item} = 5.25^{th} \text{ item}$$
$$= 15$$

Therefore,

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$
$$= \frac{15-13}{2} = 1$$

Now,

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$
$$= \frac{15-13}{15+13}$$
$$= \frac{2}{28} = 0.07$$

Example: For the following data, calculate the coefficient of quartile deviation.

X	*f*
3	7
5	9
8	14
9	6
12	4

Solution

X	*f*	*c.f*
3	7	7
5	9	16
8	14	30
9	6	36
12	4	40

$$Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{th} \text{item}$$

$$= \left(\frac{40+1}{4}\right)^{th} \text{item} = 10.25^{th} \text{item} = 5$$

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

$$= 3\left(\frac{40+1}{4}\right)^{th} \text{item} = 30.75^{th} \text{ item} = 9$$

Now,

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

$$= \frac{9-5}{2} = 2$$

Now,

$$\text{Coefficient of Quartile Deviation} = \frac{QD.}{\frac{Q_3+Q_1}{2}} = \frac{2}{\frac{9+5}{2}} = 0.28$$

Example: Calculate quartile deviation and its coefficient from the following data.

X	*f*
0 – 10	4
10 – 20	9
20 – 30	15
30 – 40	18
40 – 50	20

Solution

X	*f*	*c.f*	
0 – 10	4	4	
10 – 20	9	13	
20 – 30	15	28	– Q_1
30 – 40	18	46	
40 – 50	20	66	– Q_3

Solution

$$Q_1 = \text{Size of } \left(\frac{N}{4}\right)^{th} \text{ item}$$
$$= \left(\frac{66}{4}\right)^{th} \text{ item}$$
$$= 16.5^{th} \text{ item}$$
Thus, it lies in class interval 20−30
$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$$
$$= 20 + \frac{16.5 - 13}{15} \times 10 = 22.33$$
$$Q_3 = \text{Size of } 3\left(\frac{N}{4}\right)^{th} \text{ item} = 3\left(\frac{66}{4}\right)^{th} \text{ item} = 49.5 \text{ item}$$
$$Q_3 = l_1 + \frac{3\left(\frac{N}{4}\right) - c.f.}{f} \times i$$
$$= 40 + \frac{49.5 - 46}{20} \times 10 = 41.75$$
Hence, Quartile Deviation
$$= \frac{Q_3 - Q_1}{2}$$
$$= \frac{41.75 - 22.33}{2} = 9.71$$
Coefficient of Quartile Deviation
$$= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{41.75 - 22.33}{41.75 + 22.33} = \frac{19.42}{64.08} = 0.30$$

Example: For a particular series, the Quartile Deviation is 10 and its coefficient is 0.8. Calculate the value of the first quartile and the third quartile.

Solution

We know,

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$
$$\text{or, } 10 = \frac{Q_3 - Q_1}{2}$$

or, $Q_3 - Q_1 = 20$ …..(1)

Also,

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

$$\text{or, } 0.8 = \frac{20}{Q_3 + Q_1}$$

or, $Q_3 + Q_1 = 25$ ……(2)

Adding equations (1) and (2)

$$\begin{aligned} Q_3 - Q_1 &= 20 \\ Q_3 + Q_1 &= 25 \\ \hline 2Q_3 + 0 \quad &= 45 \end{aligned}$$

or, $Q_3 = 22.5$

Substituting the value of Q_3 in equation (1), we get
$22.5 – Q_1 = 20$
or, $Q_1 = 2.5$

Hence, the value of third quartile is 22.5 and the value of first quartile is 2.5.

Example: In a class, 25% of the students scored more than 80 marks and 75% of the students scored more than 30 marks. Calculate the absolute and the relative values of dispersion.

Solution

(i) The absolute value of dispersion is the Quartile Deviation.
It is given that 25% of the students score more than 80 marks.
$\therefore Q_3 = 80$

Similarly, 75% of the students score more than 30.
$\therefore Q_1 = 30$

Now,

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{80 - 30}{2} = 25$$

(ii) The relative value of dispersion is the coefficient of Quartile Deviation

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{80 - 30}{80 + 30} = \frac{50}{110} = 0.45$$

Coefficient of Mean Deviation

The relative measure of mean deviation is coefficient of mean deviation. The following is the formula for the calculation of coefficient of mean deviation.

$$\text{Coefficient of } M.D. = \frac{\text{Mean Deviation}}{\text{Central Value}}$$

Example: For the given data, calculate the Mean Deviation from median and mean and their respective coefficients.

S. No	***X***
1	10
2	13
3	17
4	19
5	20
6	21
7	25

Solution

S. No	***X***	**From Median |*D*|** ($X - M$)	**From Mean |*D*|** $\left(X - \overline{X}\right)$
1	10	9	7.85
2	13	6	4.85
3	17	2	0.85
4	19	0	1.15
5	20	1	2.15
6	21	2	3.15
7	25	6	7.15
	125	$\Sigma\|D\| = 26$	$\Sigma\|D\| = 27.15$

$$\text{Median} = \text{Size of } \left(\frac{7+1}{2}\right)^{th} \text{ item} = 4^{th} \text{item}$$

or, Median $= 19$

$$\text{Mean} = \frac{\Sigma X}{n}$$

$$= \frac{125}{7}$$

$$\overline{X} = 17.85$$

Mean Deviation from Median

$$= \frac{\Sigma|D|}{N}$$

$$= \frac{26}{7}$$

$$= 3.71$$

Mean Deviation from Mean

$$= \frac{\Sigma|D|}{N}$$

$$= \frac{27.15}{7}$$

$$= 3.87$$

Coefficient of Mean Deviation from Median

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

$$= \frac{3.71}{19}$$

$$= 0.195$$

Coefficient of Mean Deviation from Mean

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

$$= \frac{3.87}{17.85}$$

$$= 0.216$$

Example: For the following series calculate mean deviation from median and coefficient of mean deviation.

X	***f***
8	4
10	7
13	8
15	6
20	5

Solution

X	*f*	*c.f*	From Median |*D*|	*f* |*D*|
8	4	4	5	20
10	7	11	3	21
13	8	19	0	0
15	6	25	2	12
20	5	30	7	35
	30			Σ*f*|*D*| =88

Solution

$$= \text{Size of } \left(\frac{N+1}{2}\right)^{th} \text{ item}$$

$$\text{Median} = \text{Size of } \left(\frac{30+1}{2}\right)^{th} = 15.5^{th} = 13$$

Mean Deviation from Median

$$= \frac{\Sigma f|D|}{N}$$

$$= \frac{88}{30}$$

$$= 2.93$$

Coefficient of Mean Deviation from Median

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

$$= \frac{2.93}{13}$$

$$= 0.225$$

Example: The given table presents the distribution of marks in a class. Calculate mean deviation from median and the coefficient of mean deviation.

X	*f*
0-10	5
10-20	9
20-30	6
30-40	4
40-50	6

Solution

X	*f*	*c.f*	*m*	*fm*	From median |*D*|	*f*|*D*|

0-10	5	5	5	25	16.66	83.3
10-20	9	14	15	135	6.66	59.94
20-30	6	20	25	150	3.34	20.04
30-40	4	24	35	140	13.34	53.36
40-50	6	30	45	270	23.34	140.04
	30			720		$\Sigma f\|D\|$ =356.68

Solution

$$\text{Median} = \text{Size of } \left(\frac{N}{2}\right)^{th} \text{item}$$

$$= \text{Size of } \left(\frac{30}{2}\right)^{th} \text{item} = 15^{th}\text{item}$$

This corresponds to the class interval $(20-30)$.

$$\text{Median} = l + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20 + \frac{15-14}{6} \times 10 = 21.66$$

$$\text{Mean Deviation from Median} = \frac{\Sigma f|D|}{N} = \frac{356.58}{30} = 11.88$$

$$\text{Coefficient of Mean Deviation} = \frac{\text{Mean Deviation}}{\text{Median}} = \frac{11.88}{21.66} = 0.548$$

Coefficient of Standard Deviation and Coefficient of Variation

The relative measure of standard deviation is known as coefficient of standard deviation. Coefficient of variation is used to measure relative variation of two or more series. It is the most popular relative measure of dispersion to compare the variability in two or more than two series. It is estimated by dividing standard deviation of the series by the arithmetic mean.

<u>Example</u>: For the following data, calculate the coefficient of standard deviation.

X	*f*
8	5
11	8
12	3
14	7
15	4
17	3

Solution

x	***f***	***d* (*X*-12)**	***fd***	***d*²**	***fd*²**
8	5	– 4	– 20	16	80
11	8	– 1	– 8	1	8
A = 12	3	0	0	0	0
14	7	2	14	4	28
15	4	3	12	9	36
17	3	5	15	25	75
	30		$\Sigma fd = 13$		$\Sigma fd^2 = 227$

$$\text{Mean} = A + \frac{\sum fd}{N} = 12 + \frac{13}{30} = 12.43$$

$$\text{S.D.} = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

$$= \sqrt{\frac{227}{30} - \left(\frac{13}{30}\right)^2} = \sqrt{7.56 - 0.18} = \sqrt{7.38} = 2.71$$

$$\text{Coefficient of Standard Deviation} = \frac{S.D}{AM} = \frac{2.71}{12.43} = 0.218$$

Example: The following data presents the score of 70 students of a class in term 1 and term 2. Analyse in which term the overall performance of the class is better.

Marks	**Number of Students (Term 1)**	**Number of Students (Term 2)**
0 – 5	11	10

5 – 10	9	8
10 – 15	12	15
15 – 20	15	13
20 – 25	13	10
25 – 30	10	14
	70	70

Solution

Term 1

X	***f***	***m***	***d*** ***m – A***	***d´*** $\frac{m-A}{c}$	***fd´***	***d´²***	***\`fd´²***
0 – 5	11	2.5	– 15	3	33	9	99
5 – 10	9	7.5	– 10	2	18	4	36
10 – 15	12	12.5	– 5	1	12	1	12
15 – 20	15	(17.5) *A*	0	0	0	0	0
20 – 25	13	22.5	5	1	13	1	13
25 – 30	10	27.5	10	2	20	4	40
	70				$\Sigma fd'$=96		$\Sigma fd^{2\prime}$=200

Standard Deviation

$$\sigma = \sqrt{\frac{\sum fd'^2}{n} - \left(\frac{\sum fd'}{n}\right)^2} \times h$$

$$= \sqrt{\frac{200}{70} - \left(\frac{96}{70}\right)^2} \times 5$$

$$= \sqrt{0.98} \times 5$$

$$= 4.94$$

$$\text{Mean} = A + \frac{\sum fd'}{N} \times h$$

$$=17.5+\frac{96}{70}\times 5$$
$$=17.5+6.85$$
$$=24.35$$
$$\text{Coefficient of variance}=\frac{S.D}{\text{Mean}}=\frac{4.94}{24.35}=0.202$$

Term 2

X	***f***	***m***	***d*** ***m – A***	***d´*** $\frac{m-A}{c}$	***fd´***	***d´²***	***fd´²***
0-5	10	2.5	– 15	– 3	– 30	9	90
5-10	8	7.5	– 10	– 2	– 16	4	32
10-15	15	12.5	– 5	– 1	– 15	1	15
15-20	13	(17.5) A	0	0	0	0	0
20-25	10	22.5	5	1	10	1	10
25-30	14	27.5	10	2	28	4	56
	70				Σ*fd'*=–23		Σfd^{2}'=203

Standard Deviation

$$\sigma=\sqrt{\frac{\sum fd'^2}{n}-\left(\frac{\sum fd'}{n}\right)^2}\times h$$

$$=\sqrt{\frac{203}{70}-\left(\frac{-23}{70}\right)^2}\times 5$$
$$=\sqrt{2.8}\times 5$$
$$=8.36$$

$$\text{Mean}=A+\frac{\sum fd'}{N}\times h=17.5+\frac{-23}{70}\times 5=15.86$$

$$\text{Coefficient of Variance}=\frac{S.D}{\text{Mean}}=\frac{8.36}{15.86}=0.527$$

Since, the coefficient of variation in term 1 is less than the coefficient of variation in term 2. This suggests that the overall performance in term1 is better than that in term 2.

Example: Given below is the score of two batsmen *A* and *B* in 5 matches. Based on the data, determine which player is more consistent.

Player *A*	20	50	10	70	40
Player *B*	30	35	32	28	30

Solution

To determine which player is more consistent, we calculate the Coefficient of Variation for each player.

Player *A*

X	***d*** **(*X*– *A*)**	***d*²**
20	– 10	100
50	– 40	160
A = 10	0	0
70	60	360
40	30	90
Σ*X* = 190	Σ*d* = 40	Σ*d*² = 710

$$\text{Coefficient of Variation} = \frac{\sigma}{X} \times 100$$

Now,

$$\text{Mean}\left(\overline{X}\right) = \frac{\sum X}{N} = \frac{190}{5} = 38$$

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

$$= \sqrt{\frac{710}{5} - \left(\frac{40}{5}\right)^2}$$

$$= \sqrt{142 - 64} = 8.83$$

Substituting the values of mean and standard deviation in the formula for coefficient of variation.

$$\text{Coefficient of Variation} = \frac{\sigma}{X} \times 100$$

$$= \frac{8.83}{38} \times 100 = 23.2\%$$

Thus, coefficient of variation for player A is 23.2%

Player *B*

X	***d*** **(*X* – *A*)**	***d*²**
30	– 2	4
35	– 3	9
A = 32	0	0
28	4	16
30	2	4
ΣX = 155	Σd = 1	Σd^2 = 33

$$\text{Mean}\left(\overline{X}\right) = \frac{\sum X}{N} = \frac{155}{5} = 31$$

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

$$= \sqrt{\frac{33}{5} - \left(\frac{1}{5}\right)^2}$$

$$= \sqrt{6.6 - 0.04} = 2.56$$

$$\text{Coefficient of Variation} = \frac{\sigma}{\overline{X}} \times 100 = \frac{2.56}{31} \times 100 = 8.25\%$$

Thus, coefficient of variation for player B is 8.25%

Thus, we see that although the total score for players A is more than that for players B, the coefficient of variation for players *A* is less. This suggests that p***layers B is more consistent*** in performance.