Measures of Dispersion

Dispersion is the extent to which values in a distribution differ from the average of the distribution.

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

DISPERSION: ABSOLUTE OR RELATIVE

The measure of dispersion can be either ***'absolute'*** or ***'relative'.***

1. **Absolute Measure**: ***The measures of dispersion which are expressed in terms of original units of a series are termed as Absolute Measures***.
   - For example, the dispersion of salaries about an average is measured in rupees.
   - Absolute measures are expressed in concrete units, i.e. units in terms of which the data has been expressed like rupees, centimetres, kilograms, etc.
   - Such measures are not suitable for comparing the variability of the two distributions which are expressed in different units of measurement.
2. **Relative Measure**: ***The measures of dispersion which are measured as a percentage or ratio of the average are termed as Relative Measures.***
   - Relative Measure is sometimes known as ***coefficient*** of dispersion as 'coefficient' means a pure number that is independent of the unit of measurement.
   - When two or more series have to be compared (whether expressed in same units or different units), relative dispersion is taken into account as the absolute dispersion may be erroneous or unfit for comparison if the series are originally in different units.

OBJECTIVES OF MEASURING DISPERSION

Following are some of the purposes for which measures of dispersion are needed.

1. ***To test reliability of an average:*** Measures of dispersion are used to test to what extent, an average represents the characteristic of a series.
   - A low value of dispersion implies that there is greater degree of uniformity among various items and, consequently, their average can be taken as more reliable or representative of the distribution.
   - On the other hand, if the value of dispersion is large, it means that the items in the series are more deviated from the central value, thereby implying that the average is not representative of the data and hence not quite reliable.
2. ***To compare the extent of variability in two or more distributions:*** It aims to find out degree of uniformity or consistency in two or more sets of data. A high degree of variation would mean little uniformity or consistency, whereas a low degree of variation would mean greater uniformity or consistency.
3. ***To control the variability itself:*** The study of variation helps to analyse the reasons and causes of variations. This may be helpful in controlling the variation itself.

4. ***To serve as the basis for further statistical analysis:*** Measures of dispersions are used in computations of various important statistical techniques like correlation, regression, test of hypothesis, etc.

CHARACTERISTICS OF A GOOD MEASURE OF DISPERSION

The requirements for an ideal measure of dispersion are:

1. It should be based on all observations.
2. It should be rigidly defined.
3. It should be easy to calculate and easy to understand.
4. It is not unduly affected by the fluctuations of sampling and also by extreme observations.
5. It should be capable of further mathematical/algebraic treatment.

STANDARD DEVIATION

Standard Deviation is the positive square root of the mean of squared deviations from mean.

- It is the most commonly used measure of dispersion as it satisfies most of the properties laid down for an ideal measure of dispersion.
- Standard deviation is also known as ***root mean square deviation*** because it is the square root of the mean of squared deviations from the arithmetic mean.
- The standard deviation is denoted by the Greek letter **'σ'**

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

Relative Measures of Standard Deviation

- ***Coefficient of Standard Deviation*** *(It is also known as '**standard coefficient of dispersion**')*

Coefficient of Standard Deviation $= \dfrac{\sigma}{\overline{X}}$

- ***Coefficient of Variation (CV)***

When two or more groups of similar data are to be compared with respect to stability (or uniformity or consistency or homogeneity), coefficient of variation is the most appropriate measure.

It indicates the relationship between the standard deviation and the arithmetic mean expressed in percentage terms.

$$\text{Coefficient of Variation} = \frac{\text{Standard Deviation}}{\text{Arithmetic Mean}} \times 100 \quad \text{i.e.} \quad \frac{\sigma}{\overline{X}} \times 100$$

The series for which coefficient of variation is greater is said to be more variable or in other words, less stable, less uniform, less consistent, less homogeneous.

CALCULATION OF STANDARD DEVIATION

- **UNGROUPED DATA / INDIVIDUAL SERIES**

Four alternative methods are available for the calculation of standard deviation of individual values. All these methods result in the same value of standard deviation. These are:

1. Actual Mean Method
2. Direct Method
3. Assumed Mean Method
4. Step-Deviation Method

Actual mean method	*Direct method*
$\sigma = \sqrt{\frac{\Sigma(X-\overline{X})^2}{n}}$	$\sigma = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2}$
Assumed mean method	***Step-deviation method***
$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$ where **d = X-A**	$\sigma = \sqrt{\frac{\Sigma d'^2}{n} - \left(\frac{\Sigma d'}{n}\right)^2} \times c$ where **d' = $\frac{(X-A)}{C}$**

For Example: Calculate Standard Deviation for the given data. X = 5, 10, 25, 30 and 50

1. ACTUAL MEAN METHOD

Solution: Actual Mean Method: $\overline{X} = \frac{5+10+25+30+50}{5} = \frac{120}{5} = 24$

X	$(X-\overline{X})$	$(X-\overline{X})^2$
5	–19	361
10	–14	196
25	+1	1
30	+6	36
50	+26	676
	0	**1270**

$$\sigma = \sqrt{\frac{\Sigma(X-\overline{X})^2}{n}} = \sqrt{\frac{1270}{5}} = \sqrt{254} = 15.937$$

2. DIRECT METHOD

X	X^2
5	25
10	100
25	625
30	900
50	2500
120	**4150**

$$\sigma = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2} = \sqrt{\frac{4,150}{5} - \left(\frac{120}{5}\right)^2} = \sqrt{254} = 15.937$$

3. ASSUMED MEAN METHOD (Let A = 25)

X	d = (X-25)	d^2
5	–20	400
10	–15	225
25	0	0
30	+5	25
50	+25	625
	–5	1275

$$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{1275}{5} - \left(\frac{-5}{5}\right)^2} = \sqrt{254} = 15.937$$

4. STEP-DEVIATION METHOD (Let A = 25)

X	d = (X-25)	d′ = (d/5)	d'^2
5	–20	–4	16
10	–15	–3	9
25	0	0	0
30	+5	+1	1
50	+25	+5	25
		–1	51

$$\sigma = \sqrt{\frac{\Sigma d'^2}{n} - \left(\frac{\Sigma d'}{n}\right)^2} \times c = \sqrt{\frac{51}{5} - \left(\frac{-1}{5}\right)^2} \times 5 = \sqrt{10.16} \times 5 = 15.937$$

Coefficient of Variation (C.V.) $= \frac{\sigma}{\overline{X}} \times 100 = \frac{15.937}{24} \times 100 = 66.4\%$

- **DISCRETE SERIES**

Actual mean method: $\sigma = \sqrt{\frac{\Sigma f(X-\overline{X})^2}{\Sigma f}}$ $\left(\text{we know that } \overline{X} = \frac{\Sigma fX}{\Sigma f}\right)$

Assumed mean method: $\sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2}$, where $d = X - A$

Direct method: $\sigma = \sqrt{\frac{\Sigma fX^2}{\Sigma f} - \left(\frac{\Sigma fX}{\Sigma f}\right)^2}$

Step deviation method: $\sigma = \sqrt{\frac{\Sigma fd'^2}{\Sigma f} - \left(\frac{\Sigma fd'}{\Sigma f}\right)^2} \times c$, where $d' = \frac{X - A}{c}$

For Example: Calculate Standard Deviation and Coefficient of Variation for the following data:

Marks	5	10	15	20
No. of students	2	1	4	3

1. *ACTUAL MEAN METHOD*

Marks (X)	No. of students (f)	fX	(X-14)	$(X-14)^2$	$f(X-14)^2$
5	2	10	–9	81	162
10	1	10	–4	16	16
15	4	60	+1	1	4
20	3	60	+6	36	108
	10	**140**			**290**

Mean, $\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{140}{10} = 14$

$\sigma = \sqrt{\frac{\Sigma f(X-\bar{X})^2}{\Sigma f}} = \sqrt{\frac{290}{10}} = 5.38$ Coefficient of Variation (C.V.) = $\frac{\sigma}{\bar{X}} \times 100 = \frac{5.38}{14} \times 100 = 38.43\%$

2. *DIRECT METHOD*

Marks (X)	No. of students (f)	fX	X^2	fX^2
5	2	10	25	50
10	1	10	100	100
15	4	60	225	900
20	3	60	400	1,200
	10	**140**		**2,250**

Standard deviation, $\sigma = \sqrt{\frac{\Sigma fX^2}{\Sigma f} - \left(\frac{\Sigma fX}{\Sigma f}\right)^2} = \sqrt{\frac{2,250}{10} - \left(\frac{140}{10}\right)^2} = \sqrt{225 - 196} = \sqrt{29} = 5.38$

Coefficient of Variation (C.V.) = $\frac{\sigma}{\bar{X}} \times 100 = \frac{5.38}{14} \times 100 = 38.43\%$

3. *ASSUMED MEAN METHOD (Let A = 10)*

Marks (X)	No. of students (f)	d = (X – 10)	d^2	fd	fd^2
5	2	–5	25	–10	50
10	1	0	0	0	0
15	4	+5	25	20	100
20	3	+10	100	30	300
	10			**40**	**450**

Standard deviation, $\sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2} = \sqrt{\frac{450}{10} - \left(\frac{40}{10}\right)^2} = \sqrt{45 - 16} = \sqrt{29} = 5.38$

Coefficient of Variation (C.V.) = $\frac{\sigma}{\bar{X}} \times 100 = \frac{5.38}{14} \times 100 = 38.43\%$

4. STEP-DEVIATION METHOD (Let A = 10)

Marks X	No. of students (f)	d = X – 10	d′ = d/5	d'^2	fd′	fd'^2
5	2	–5	–1	1	–2	2
10	1	0	0	0	0	0
15	4	+5	+1	1	+4	4
20	3	+10	+2	4	+6	12
	10				+8	18

Standard deviation, $\sigma = \sqrt{\frac{\Sigma fd'^2}{\Sigma f} - \left(\frac{\Sigma fd'}{\Sigma f}\right)^2} \times c = \sqrt{\frac{18}{10} - \left(\frac{8}{10}\right)^2} \times 5 = \sqrt{1.8 - 0.64} \times 5 = \sqrt{1.16} \times 5 = 1.077 \times 5 = 5.38$

Coefficient of Variation (C.V.) = $\frac{\sigma}{\bar{X}} \times 100 = \frac{5.38}{14} \times 100 = 38.43\%$

NOTE: While comparing dispersion for two series, the series for which coefficient of variation is greater is said to be more variable or in other words, less stable, less uniform and less consistent.

- **CONTINUOUS SERIES**

Actual Mean Method	***Direct Method***
$\sigma = \sqrt{\frac{\Sigma fx^2}{N}}$ $x = m - \bar{X}$ $N = \Sigma f$	$\sigma = \sqrt{\frac{\Sigma fm^2}{N} - (\bar{X})^2}$ $\bar{X} = \frac{\Sigma fm}{\Sigma f}$; $N = \Sigma f$
Assumed Mean Method $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$ Where $d = m - A$ $N = \Sigma f$	***Step deviation Method*** $\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$ Where $d' = \frac{(m - A)}{C}$ $N = \Sigma f$

For Example: Calculate the standard deviation from the following data by:

(i) Actual Mean Method (ii) Direct Method
(ii) Assumed Mean Method (iv) Step-deviation Method

X (Marks)	0-10	10-20	20-30	30-40
Frequency	2	3	4	1

Actual Mean Method | ***Direct Method***

X	*f*	*m*	*fm*	$x = m - \bar{X}$	x^2	fx^2	*f*	*m*	*fm*	m^2	fm^2
0–10	2	5	10	– 14	196	392	2	5	10	25	50
10–20	3	15	45	– 4	16	48	3	15	45	225	675
20–30	4	25	100	+ 6	36	144	4	25	100	625	2,500
30–40	1	35	35	+ 16	256	256	1	35	35	1,225	1,225
	N = 10		**Σfm = 190**			**Σfx^2 = 840**	**N = 10**		**Σfm = 190**		**Σfm^2 = 4,450**

$$\bar{X} = \frac{\Sigma fm}{\Sigma f} = \frac{190}{10} = 19$$

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N}} = \sqrt{\frac{840}{10}}$$

$[x = m - \bar{X}]$

$$\sigma = \sqrt{84} = 9.165$$

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

$\left[\bar{X} = \dfrac{\Sigma fm}{\Sigma f}\right]$

$$\sigma = \sqrt{\frac{4{,}450}{10} - (19)^2}$$

$$\sigma = \sqrt{84} = 9.165$$

m = mid-points; **d** = m – A where, A = 15; $d' = \dfrac{m - A}{C}$ where, C = 10

Assumed Mean Method ***(Let A = 15)*** ***Step-Deviation Method***

X	*f*	*m*	*d*	*fd*	d^2	fd^2	*f*	*m*	*d*	*d'*	*fd'*	d'^2	fd'^2
0–10	2	5	– 10	– 20	100	200	2	5	– 10	– 1	– 2	1	2
10–20	3	**15 (A)**	0	0	0	0	3	**15 (A)**	0	0	0	0	0
20–30	4	25	+ 10	+ 40	100	400	4	25	+ 10	+ 1	+ 4	1	4
30–40	1	35	+ 20	+ 20	400	400	1	35	+ 20	+ 2	+ 2	4	4
	N = 10			**Σfd = 40**		**Σfd^2 = 1,000**	**N = 10**				**Σfd' = 4**		**$\Sigma fd'^2$ = 10**

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

$[d = m - A]$

$$\sigma = \sqrt{\frac{1{,}000}{10} - \left(\frac{40}{10}\right)^2}$$

$$\sigma = \sqrt{84} = 9.165$$

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

$\left[d' = \dfrac{m - A}{C}\right]$

$$\sigma = \sqrt{\frac{10}{10} - \left(\frac{4}{10}\right)^2} \times 10$$

$$\sigma = \sqrt{0.84} \times 10 = 9.165$$

Variance is the square of Standard Deviation. It is another measure of calculating dispersion.

$$\text{Variance} = \sigma^2$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\text{Variance}}$$

PROPERTIES OF STANDARD DEVIATION

1. ***The sum of the square of the deviations of the items from their arithmetic mean is minimum.*** The sum is less than the sum of the square of the deviations of the items from any other value. $\Sigma(X - \bar{X})^2$ is always minimum.

2. ***Standard Deviation is independent of change of origin, i.e.*** value of standard deviation remains the same if, in a series, a constant is added (or subtracted) to all observations.

3. ***Standard Deviation is not independent of change of scale or is*** *affected by change of scale, i.e.* if all the observations are multiplied (or divided) by a constant, then the standard deviation also gets multiplied (or divided) by that constant.

MERITS AND DEMERITS OF STANDARD DEVIATION

Merits of Standard Deviation

1. **Based on all observations**: Standard Deviation considers every item of the series. So, a change in even one value affects the value of standard deviation.
2. **Rigidly defined**: Standard Deviation is by far the most important and widely used measure of dispersion. It is rigidly defined, i.e. it is a definite measure of dispersion.
3. **Less affected by fluctuations in sampling**: If several independent samples are drawn from the same population, it may be observed that standard deviation is least affected from sample to sample as compared with other measures of dispersions.
4. **Capable of further algebraic treatment:** Standard Deviation is capable of further algebraic treatment.

Demerits of Standard Deviation

1. **Difficult to compute**: Standard Deviation is more difficult to be measured as compared to other measures of dispersion.
2. **More stress on extreme items**: It gives more weightage to extreme values and less to those which are nearer to mean.
3. **Depend upon units of measurement:** It depends upon the units of measurement of the observations. So, it cannot be used to compare the dispersion of the distributions expressed in different units.

LIST OF FORMULAE

Standard Deviation

	Individual Series	*Discrete Series*	*Continuous Series*
Actual Mean Method	$\sigma = \sqrt{\frac{\Sigma x^2}{N}}$ $[x = X-\overline{X}]$	$\sigma = \sqrt{\frac{\Sigma fx^2}{N}}$ $[x = X-\overline{X}]$	$\sigma = \sqrt{\frac{\Sigma fx^2}{N}}$ $[x = m-\overline{X}]$
Direct Method	$\sigma = \sqrt{\frac{\Sigma X^2}{N} - (\overline{X})^2}$	$\sigma = \sqrt{\frac{\Sigma fX^2}{N} - (\overline{X})^2}$	$\sigma = \sqrt{\frac{\Sigma fm^2}{N} - (\overline{X})^2}$
Short-Cut Method	$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$	$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$	$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$
Step Deviation Method	$\sigma = \sqrt{\frac{\Sigma d'^2}{n} - \left(\frac{\Sigma d'}{n}\right)^2} \times c$	$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$	$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$
Variance = σ^2			**For grouped data: N = Σf**

Coefficient of Standard Deviation $= \frac{\sigma}{\overline{X}}$; Coefficient of Variation $= \frac{\sigma}{\overline{X}} \times 100$

Click on the following link for further explanation of the topics discussed above:
https://www.youtube.com/watch?v=E6tJ6tZ9caU (Standard Deviation)