Lecture - 08

- **Measures of Dispersion**

As the name suggests, the measure of dispersion shows the scatterings of the data. It tells the variation of the data from one another and gives a clear idea about the distribution of the data. The measure of dispersion shows the homogeneity or the heterogeneity of the distribution is the observations.

- **The measurement of the scattering of observation in a distribution about the average is called as measures of dispersion or variation.**
- **They are also called as Averages of Second Order.**
- **By dispersion we mean spreading or scatteredness or variation.**
- **When dispersion is significant, it means that the average is far from being a representative figure**

Some definitions

• "Dispersion is the measure of variation of the items"-A.L. Bowley

• "The measure of the scatteredness of the mass of figures in a series about an average is called the measure of variation or dispersion"- Simpson and Kafka

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

Properties of an ideal measure of Dispersion

• It should be easy to understand simple to calculate.

• It should be based on all observations.

• It should be capable of further algebraic treatment.

• It should be least affected by sampling fluctuations.

• It should not be unduly affected by extreme observations.

Characteristics of Measures of Dispersion

- A measure of dispersion should be rigidly defined
- It must be easy to calculate and understand
- Not affected much by the fluctuations of observations

- Based on all observations

Classification of Measures of Dispersion

The measure of dispersion is categorized as:

(i) An absolute measure of dispersion:

- The measures which express the scattering of observation in terms of distances i.e., range, quartile deviation.
- The measure which expresses the variations in terms of the average of deviations of observations like mean deviation and standard deviation.

(ii) A relative measure of dispersion:

We use a relative measure of dispersion for comparing distributions of two or more data set and for unit free comparison. They are the coefficient of range, the coefficient of mean deviation, the coefficient of quartile deviation, the coefficient of variation, and the coefficient of standard deviation.

- **Relative Measures of Dispersion**

Based on Selected Items

1. Coefficient of Range

2. Coefficient of Quartile Deviation

- **Based on All Items**

1. Coefficient of Mean Deviation

2. Coefficient of Standard Deviation

3. Coefficient of Variation

Merits

- They indicate the dispersal character of a statistical series.
- They speak of the reliability, or dependability of the average value of a series.

- They enable the statisticians for making a comparison between two or more statistical series with regard to the character of their stability or consistency.
- They facilitate in controlling the variability of a phenomenon under his purview.
- They facilitate in making further statistical analysis of the series through the devices like co-efficient of skewness, co-efficient of correlation, variance analysis etc.
- They supplement the measures of central tendency in finding out more and more information relating to the nature of a series.

Demerits

- They are liable to misinterpretations, and wrong generalizations by a statistician of based character.
- They are liable to yield inappropriate results as there are different methods of calculating the dispersions.
- Exception on or two, of the methods of dispersion involve complicated process of computation.
- They may give a value of variation, which may not be practically found with the items of the series.
- They, by themselves, cannot give any idea about the symmetricity, or skewed character of a series.
- Like the measures of central tendency, most of the measures of dispersion do not give a convincing idea about a series to a layman.

Lecture - 09

Range

A range is the most common and easily understandable measure of dispersion. It is the difference between two extreme observations of the data set. If X_{max} and X_{min} are the two extreme observations then

Range = $X_{max} - X_{min}$

Merits of Range

- It is the simplest of the measure of dispersion
- Easy to calculate
- Easy to understand
- Independent of change of origin

Demerits of Range

- It is based on two extreme observations. Hence, get affected by fluctuations
- A range is not a reliable measure of dispersion
- Dependent on change of scale

Quartile Deviation

The quartiles divide a data set into quarters. The first quartile, (Q_1) is the middle number between the smallest number and the median of the data. The second quartile, (Q_2) is the median of the data set. The third quartile, (Q_3) is the middle number between the median and the largest number.

Quartile deviation or semi-inter-quartile deviation is

$Q = ½ \times (Q_3 – Q1)$

Merits of Quartile Deviation

- All the drawbacks of Range are overcome by quartile deviation
- It uses half of the data
- Independent of change of origin
- The best measure of dispersion for open-end classification

Demerits of Quartile Deviation

- It ignores 50% of the data
- Dependent on change of scale
- Not a reliable measure of dispersion

Lecture - 10

Mean Deviation

Mean deviation is the arithmetic mean of the absolute deviations of the observations from a measure of central tendency. If x_1, x_2, … , x_n are the set of observation, then the mean deviation of x about the average A (mean, median, or mode) is

Mean deviation from average A = $1/n \; [\sum_i |x_i - A|]$

For a grouped frequency, it is calculated as:

Mean deviation from average A = $1/N \; [\sum_i f_i \, |x_i - A|]$, $N = \sum f_i$

Here, x_i and f_i are respectively the mid value and the frequency of the i^{th} class interval.

Merits of Mean Deviation

- Based on all observations
- It provides a minimum value when the deviations are taken from the median
- Independent of change of origin

Demerits of Mean Deviation

- Not easily understandable
- Its calculation is not easy and time-consuming
- Dependent on the change of scale
- Ignorance of negative sign creates artificiality and becomes useless for further mathematical treatment

Lecture - 11

Standard Deviation

A standard deviation is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean. It is denoted by a Greek letter sigma, σ. It is also referred to as root mean square deviation. The standard deviation is given as

$$\sigma = [(\Sigma_i (y_i - \bar{y}) / n]^{1/2} = [(\Sigma_i y_i^2 / n) - \bar{y}^2]^{1/2}$$

Merits of Standard Deviation

- Squaring the deviations overcomes the drawback of ignoring signs in mean deviations
- Suitable for further mathematical treatment
- Least affected by the fluctuation of the observations
- The standard deviation is zero if all the observations are constant
- Independent of change of origin

Demerits of Standard Deviation

- Not easy to calculate
- Difficult to understand for a layman
- Dependent on the change of scale

Coefficient of Dispersion

Whenever we want to compare the variability of the two series which differ widely in their averages. Also, when the unit of measurement is different. We need to calculate the coefficients of dispersion along with the measure of dispersion. The coefficients of dispersion (C.D.) based on different measures of dispersion are

- Based on Range = $(X_{max} - X_{min}) / (X_{max} + X_{min})$.
- C.D. based on quartile deviation = $(Q_3 - Q1) / (Q_3 + Q1)$.
- Based on mean deviation = Mean deviation/average from which it is calculated.

- For Standard deviation = S.D. ⁄ Mean

Coefficient of Variation

100 times the coefficient of dispersion based on standard deviation is the coefficient of variation (C.V.).

C.V. = 100 × (S.D. / Mean) = $(\sigma/\bar{y}) \times 100$.

Unit - III

Measures of Variation (or) Dispersion of a data provide an idea of how observations spread out (or) scattered throughout the data.

Different Measures of Dispersion are

1. Range

2. Mean deviation

3. Quartile deviation

4. Standard deviation

5. Variance

6. Coefficient of Variation

1. Range

The difference between the largest value and the smallest value is called Range.

Range $R = L – S$

Coefficient of range = $(L – S) / (L + S)$

Where L - Largest value; S - Smallest value

Example 1 Find the range and coefficient of range of the following data: 25, 67, 48, 53, 18, 39, and 44.

Solution Largest value $L = 67$; Smallest value $S = 18$

Range R = $L - S = 67 - 18 = 49$

Coefficient of range = $(L – S) / (L + S)$

Coefficient of range = $(67 – 18) / (67 + 18) = 49/85$

$= 0.576$

Example 2 Find the range of the following distribution.

Age (in years)	16-18	18-20	20-22	22-24	24-26	26-28
Number of students	0	4	6	8	2	2

Solution Here Largest value $L = 28$

Smallest value $S = 18$

Range $R = L - S$

$R = 28 - 18 = 10$ Years

Example 3 The range of a set of data is 13.67 and the largest value is 70.08. Find the smallest value.

Solution

Range $R = 13.67$

Largest value $L = 70.08$

Range $R = L - S$

$13.67 = 70.08 - S$

$S = 70.08 - 13.67 = 56.41$

Therefore, the smallest value is 56.41.

2. Deviations from the mean

For a given data with n observations $x_1, x_2, ¼ x_n$, the deviations from the mean $\overline{x}$ are

$$x_1 - \overline{x},\ x_2 - \overline{x},\ \ldots,\ x_n - \overline{x}.$$

3. Squares of deviations from the mean

The squares of deviations from the mean $\overline{x}$ of the observations $x_1, x_2, \ldots, x_n$ are

$$(x_1 - \overline{x})^2, (x_2 - \overline{x})^2, \ldots, (x_n - \overline{x})^2 \text{ or } \sum_{i=1}^{n} (x_i - \overline{x})^2$$

4. Variance

The mean of the squares of the deviations from the mean is called Variance.

It is denoted by σ^2 (read as sigma square).

Variance = Mean of squares of deviations

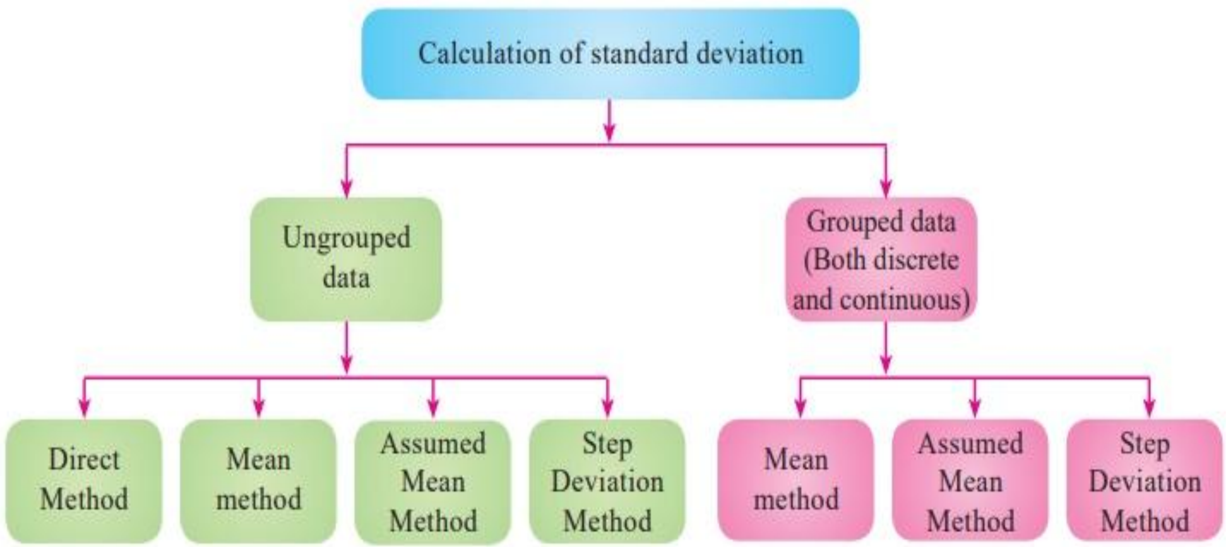
$$\begin{aligned} \text{Variance} &= \text{Mean of squares of deviations} \\ &= \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n} \\ \text{Variance } \sigma^2 &= \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n} \end{aligned}$$

5. Standard Deviation

The positive square root of Variance is called Standard deviation. That is, standard deviation is the positive square root of the mean of the squares of deviations of the given values from their mean. It is denoted by σ.

Standard deviation gives a clear idea about how far the values are spreading or deviating from the mean.

$$\text{Standard deviation } \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}}$$



1. Calculation of Standard Deviation for ungrouped data

(i) Direct Method

$$\text{Standard deviation } \sigma = \sqrt{\frac{\Sigma(x_i - \overline{x})^2}{n}}$$

$$= \sqrt{\frac{\Sigma(x_i^2 - 2x_i\overline{x} + \overline{x}^2)}{n}}$$

$$= \sqrt{\frac{\Sigma x_i^2}{n} - 2\overline{x}\frac{\Sigma x_i}{n} + \frac{\overline{x}^2}{n} \times (1 + 1 + \cdots to\ n \text{ times})}$$

$$= \sqrt{\frac{\Sigma x_i^2}{n} - 2\overline{x} \times \overline{x} + \frac{\overline{x}^2}{n} \times n} = \sqrt{\frac{\Sigma x_i^2}{n} - 2\overline{x}^2 + \overline{x}^2} = \sqrt{\frac{\Sigma x_i^2}{n} - \overline{x}^2}$$

$$\text{Standard deviation, } \sigma = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$$

- While computing standard deviation, arranging data in ascending order is not mandatory.

- If the data values are given directly then to find standard deviation we can use the formula $\sigma = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$

- If the data values are not given directly but the squares of the deviations from the mean of each observation is given then to find standard deviation we can use the formula $\sigma = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}}$

Example 4 The number of televisions sold in each day of a week are 13, 8, 4, 9, 7, 12, 10.

Find its standard deviation.

Solution

x_i	x_i^2
13	169
8	64
4	16
9	81
7	49
12	144
10	100
$\Sigma x_i = 63$	$\Sigma x_i^2 = 623$

Standard deviation

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

$$= \sqrt{\frac{623}{7} - \left(\frac{63}{7}\right)^2}$$

$$= \sqrt{89 - 81} = \sqrt{8}$$

Hence, $\sigma \simeq 2.83$

(ii) Mean method

Another convenient way of finding standard deviation is to use the following formula.

Standard deviation (by mean method) $\sigma = \sqrt{\frac{\Sigma(x_i - \overline{x})^2}{n}}$

If $d_i = x_i - \overline{x}$ are the deviations, then $\sigma = \sqrt{\frac{\Sigma d_i^2}{n}}$

Example 5 The amount of rainfall in a particular season for 6 days are given as 17.8 cm, 19.2 cm, 16.3 cm, 12.5 cm, 12.8 cm and 11.4 cm. Find its standard deviation.

Solution Arranging the numbers in ascending order we get, 11.4, 12.5, 12.8, 16.3, 17.8, 19.2.

Number of observations $n = 6$

$$\text{Mean} = \frac{11.4 + 12.5 + 12.8 + 16.3 + 17.8 + 19.2}{6} = \frac{90}{6} = 15$$

x_i	$d_i = x_i - \overline{x} = x - 15$	d_i^2
11.4	−3.6	12.96
12.5	−2.5	6.25
12.8	−2.2	4.84
16.3	1.3	1.69
17.8	2.8	7.84
19.2	4.2	17.64
		$\Sigma d_i^2 = 51.22$

$$\text{Standard deviation } \sigma = \sqrt{\frac{\Sigma d_i^2}{n}}$$

$$= \sqrt{\frac{51.22}{6}} = \sqrt{8.53}$$

Hence, $\sigma \simeq 2.9$

(iii) Assumed Mean method

When the mean value is not an integer (since calculations are very tedious in decimal form) then it is better to use the assumed mean method to find the standard deviation.

Example 6 The marks scored by 10 students in a class test are 25, 29, 30, 33, 35, 37, 38, 40, 44, 48. Find the standard deviation.

Solution The mean of marks is 35.9 which is not an integer. Hence we take assumed mean, $A = 35$, $n = 10$.

x_i	$d_i = x_i - A$ $d_i = x_i - 35$	d_i^2
25	−10	100
29	−6	36
30	−5	25
33	−2	4
35	0	0
37	2	4
38	3	9
40	5	25
44	9	81
48	13	169
	$\Sigma d_i = 9$	$\Sigma d_i^2 = 453$

Standard deviation

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

$$= \sqrt{\frac{453}{10} - \left(\frac{9}{10}\right)^2}$$

$$= \sqrt{45.3 - 0.81}$$

$$= \sqrt{44.49}$$

$$\sigma \simeq 6.67$$

(iv) Step deviation method

Example 7 The amount that the children have spent for purchasing some eatables in one day trip of a school are 5, 10, 15, 20, 25, 30, 35, 40. Using step deviation method, find the standard deviation of the amount they have spent.

Solution We note that all the observations are divisible by 5. Hence we can use the step deviation method. Let the Assumed mean $A = 20$, $n = 8$.

x_i	$d_i = x_i - A$ $d_i = x_i - 20$	$d_i = \frac{x_i - A}{c}$ $c = 5$	d_i^2
5	−15	−3	9
10	−10	−2	4
15	−5	−1	1
20	0	0	0
25	5	1	1
30	10	2	4
35	15	3	9
40	20	4	16
		$\Sigma d_i = 4$	$\Sigma d_i^2 = 44$

Standard deviation

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

$$= \sqrt{\frac{44}{8} - \left(\frac{4}{8}\right)^2} \times 5 = \sqrt{\frac{11}{2} - \frac{1}{4}} \times 5$$

$$= \sqrt{5.5 - 0.25} \times 5 = 2.29 \times 5$$

$$\sigma \simeq 11.45$$

Example 8 Find the standard deviation of the following data 7, 4, 8, 10, 11. Add 3 to all the values then find the standard deviation for the new values.

Solution Arranging the values in ascending order we get, 4, 7, 8, 10, 11 and $n = 5$

x_i	x_i^2
4	16
7	49
8	64
10	100
11	121
$\Sigma x_i = 40$	$\Sigma x_i^2 = 350$

Standard deviation

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

$$= \sqrt{\frac{350}{5} - \left(\frac{40}{5}\right)^2}$$

$$\sigma = \sqrt{6} \simeq 2.45$$

When we add 3 to all the values, we get the new values as 7,10,11,13,14.

x_i	x_i^2
7	9
10	100
11	121
13	169
14	196
$\Sigma x_i = 55$	$\Sigma x_i^2 = 635$

Standard deviation

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

$$= \sqrt{\frac{635}{5} - \left(\frac{55}{5}\right)^2}$$

$$\sigma = \sqrt{6} \simeq 2.45$$

From the above, we see that the standard deviation will not change when we add some fixed constant to all the values.

Example 9 Find the standard deviation of the data 2, 3, 5, 7, 8. Multiply each data by 4. Find the standard deviation of the new values.

Solution Given, $n = 5$

x_i	x_i^2
2	49
3	9
5	25
7	49
8	64
$\Sigma x_i = 25$	$\Sigma x_i^2 = 151$

Standard deviation

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

$$\sigma = \sqrt{\frac{151}{5} - \left(\frac{25}{5}\right)^2} = \sqrt{30.2 - 25} = \sqrt{5.2} \simeq 2.28$$

When we multiply each data by 4, we get the new values as 8, 12, 20, 28, 32.

x_i	x_i^2
8	64
12	144
20	400
28	784
32	1024
$\Sigma x_i = 100$	$\Sigma x_i^2 = 2416$

Standard deviation

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

$$= \sqrt{\frac{2416}{5} - \left(\frac{100}{5}\right)^2} = \sqrt{483.2 - 400} = \sqrt{83.2}$$

$$\sigma = \sqrt{16 \times 5.2} = 4\sqrt{5.2} \simeq 9.12$$

From the above, we see that when we multiply each data by 4 the standard deviation also get multiplied by 4.

Calculation of Standard deviation for grouped data

(i) Mean method

Standard deviation $\sigma = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{N}}$

Let, $d_i = x_i - \bar{x}$

$$\sigma = \sqrt{\frac{\Sigma f_i d_i^2}{N}}, \text{ where } N = \sum_{i=1}^{n} f_i$$

Example 11

48 students were asked to write the total number of hours per week they spent on watching television. With this information find the standard deviation of hours spent for watching television.

x	6	7	8	9	10	11	12
f	3	6	9	13	8	5	4

Solution

x_i	f_i	$x_i f_i$	$d_i = x_i - \overline{x}$	d_i^2	$f_i d_i^2$
6	3	18	−3	9	27
7	6	42	−2	4	24
8	9	72	−1	1	9
9	13	117	0	0	0
10	8	80	1	1	8
11	5	55	2	4	20
12	4	48	3	9	36
	$N = 48$	$\Sigma x_i f_i = 432$	$\Sigma d_i = 0$		$\Sigma f_i d_i^2 = 124$

Mean

$$\overline{x} = \frac{\Sigma x_i f_i}{N} = \frac{432}{48} = 9 \quad (\text{Since } N = \sum f_i)$$

Standard deviation

$$\sigma = \sqrt{\frac{\Sigma f_i d_i^2}{N}} = \sqrt{\frac{124}{48}} = \sqrt{2.58}$$

$$\sigma \simeq 1.6$$

(ii) Assumed Mean Medthod

Let $x_1, x_2, x_3, \ldots x_n$ be the given data with frequencies $f_1, f_2, f_3, \ldots f_n$ respectively. Let x be their mean and A be the assumed mean..

$$d_i = x_i - A$$

Standard deviation, $\sigma = \sqrt{\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2}$

Example 12

The marks scored by the students in a slip test are given below.

x	4	6	8	10	12
f	7	3	5	9	5

Find the standard deviation of their marks.

Solution

Let the assumed mean, $A = 8$

x_i	f_i	$d_i = x_i - A$	$f_i d_i$	$f_i d_i^2$
4	7	–4	–28	112
6	3	–2	–6	12
8	5	0	0	0
10	9	2	18	36
12	5	4	20	80
	$N = 29$		$\Sigma f_i d_i = 4$	$\Sigma f_i d_i^2 = 240$

Standard deviation

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

$$= \sqrt{\frac{240}{29} - \left(\frac{4}{29}\right)^2} = \sqrt{\frac{240 \times 29 - 16}{29 \times 29}}$$

$$\sigma = \sqrt{\frac{6944}{29 \times 29}}\ ; \quad \sigma \simeq 2.87$$

2. Calculation of Standard deviation for continuous frequency distribution

(i) Mean method

Standard deviation $\sigma = \sqrt{\frac{\Sigma f_i \left(x_i - \overline{x}\right)^2}{N}}$ where x_i = Middle value of the i th class.

f_i = Frequency of the i th class.

(ii) Shortcut method (or) Step deviation method

To make the calculation simple, we provide the following formula. Let A be the assumed mean, x_i be the middle value of the i^{th} class and c is the width of the class interval.

Let
$$d_i = \frac{x_i - A}{c}$$
$$\sigma = c \times \sqrt{\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2}$$

13Marks of the students in a particular subject of a class are given below.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Number of students	8	12	17	14	9	7	4

Find its standard deviation.

Solution

Let the assumed mean, A = 35, c = 10

Marks	Midvalue (x_i)	f_i	$d_i = x_i - A$	$d_i = \frac{x_i - A}{c}$	$f_i d_i$	$f_i d_i^2$
0-10	5	8	-30	-3	−24	72
10-20	15	12	-20	−2	−24	48
20-30	25	17	-10	−1	−17	17
30-40	35	14	0	0	0	0
40-50	45	9	10	1	9	9
50-60	55	7	20	2	14	28
60-70	65	4	30	3	12	36
		$N = 71$			$\Sigma f_i d_i = -30$	$\Sigma f_i d_i^2 = 210$

Standard deviation $\sigma = c \times \sqrt{\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2}$

$$\sigma = 10 \times \sqrt{\frac{210}{71} - \left(-\frac{30}{71}\right)^2} = 10 \times \sqrt{\frac{210}{71} - \frac{900}{5041}}$$
$$= 10 \times \sqrt{2.779} \; ; \quad \sigma \simeq 16.67$$

Example 14

The mean and standard deviation of 15 observations are found to be 10 and 5 respectively. On rechecking it was found that one of the observation with value 8 was incorrect. Calculate the correct mean and standard deviation if the correct observation value was 23?

Solution

$$n = 15\,,\ \overline{x} = 10\,,\ \sigma = 5\,;\quad \overline{x} = \frac{\Sigma x}{n};$$

$$\Sigma x = 15 \times 10 = 150$$

Wrong observation value = 8, Correct observation value = 23.

Correct total = 150 − 8 + 23 = 165

$$\text{Correct mean } \overline{x} = \frac{165}{15} = 11$$

$$\text{Standard deviation } \sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$\text{Incorrect value of } \sigma = 5 = \sqrt{\frac{\Sigma x^2}{15} - \left(10\right)^2}$$

$$25 = \frac{\Sigma x^2}{15} - 100 \text{ gives, } \frac{\Sigma x^2}{15} = 125$$

$$\text{Incorrect value of } \Sigma x^2 = 1875$$

$$\text{Correct value of } \Sigma x^2 = 1875 - 8^2 + 23^2 = 2340$$

$$\text{Correct standard deviation } \sigma = \sqrt{\frac{2340}{15} - \left(11\right)^2}$$

$$\sigma = \sqrt{156 - 121} = \sqrt{35} \qquad \sigma \simeq 5.9$$

1. Find the range and coefficient of range of the following data.

(i) 63, 89, 98, 125, 79, 108, 117, 68

(ii) 43.5, 13.6, 18.9, 38.4, 61.4, 29.8

(i) 63,89,98,125,79,108,117,68

Solution:

Let us arrange the given data in the ascending order 63,68,79,89,98,108,117,125 Largest value L=125 Smallest value S=63

Range = L-S

= 125-63=62

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

$$= \frac{125 \quad 63}{125 \quad 63} = \frac{62}{188} = 0.33$$

Answer R=62 Coefficient of

Range = 0.33

(ii) 43.5,13.6,18.9,38.4,61.4,29.8

Solution:

Let us arrange in ascending order 13.6,18.9,29.8,38.4,43.5,61.4 Largest value L=61.4 Smallest value S= 13.6

Range = L-S

61.4-13.6 = 47.8

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

$$= \frac{61.4-13.6}{61.4+13.6}$$

$$= \frac{47.8}{75} = 0.64$$

Answer: R=47.8 Coefficient of Range = 0.64

2. If the range and the smallest value of a set of data are 36.8 and 13.4 respectively, then find the largest value.

Solution:

Range R = 36.8

Smallest value S = 13.4

$$R = L-S$$

$$36.8 = L-13.4$$

L= 36.8+13.4=50.2

Answer: Largest Value = 50.2

3. Calculate the range of the following data.

Income	400-450	450-500	500-550	550-600	600-650
Number of workers	8	12	30	21	6

Solution:

Largest value L=650

Smallest value S =400

Range R = L-S

=650-450=250

Answer : R=250.

4. A teacher asked the students to complete 60 pages of a record note book. Eight students have completed only 32, 35, 37, 30, 33, 36, 35 and 37 pages. Find the standard deviation of the pages yet to be completed by them.

Solution:

By Assumed Mean Method :

Pages yet to be completed are 28, 25, 23, 30, 27, 24, 25, and 23

Assumed mean A=25 n = 8

x_i	$d_i=x_i-A$ $d_i=x_i-35$	d_i^2
23	-2	4
23	-2	4
24	-1	1
25	0	0
25	0	0
27	2	4
28	3	9
30	5	25
	$\sum d_i=5$	$\sum d_i^2=47$

Standard deviation

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

$$= \sqrt{\frac{47}{8} - \left(\frac{5}{8}\right)^2}$$

$$= \sqrt{\frac{47}{8} - \frac{25}{64}}$$

$$= \sqrt{\frac{351}{64}}$$

$$= \frac{18.735}{8}$$

$$\sigma = 2.34$$

Ans: S.D of the pages to be completed = 2.34

5. Find the variance and standard deviation of the wages of 9 workers given below: ₹310, ₹290, ₹320, ₹280, ₹300, ₹290, ₹320, ₹310, ₹280.

Solution Mean Method :

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{280+280+290+290+300+310+310+232+320}{9}$$

$$= \frac{2700}{9} = 300$$

x_i	$d_i = x_i - \bar{x}$ $= x_i - 300$	d_i^2
280	–20	400
280	–20	400
290	–10	100
290	–10	100
300	0	0
310	10	100
310	10	100
320	20	400
320	20	400
	$\Sigma d_i = -5$	$\Sigma d_i^2 = 2000$

Variance

$$\sigma^2 = \frac{\Sigma d_i^2}{n}$$

$$= \frac{2000}{9}$$

$$= 222.22$$

Standard deviation = 14.91

$$\sigma = \sqrt{\text{Variance}}$$

$$= \sqrt{222.22}$$

Answer : Variance = 222.22

S.D = 14.91

6. The rainfall recorded in various places of five districts in a week are given below.

Rainfall (in mm)	45	50	55	60	65	70
Number of places	5	13	4	9	5	4

Find its standard deviation.

Solution : Assumed mean method A = 60

x_i	f_i	$d_i = x_i - A = x_i - 60$	$f_i d_i$	$f_i d_i^2$
45	5	–15	–75	1125
50	13	–10	–130	1300
55	4	–5	–20	100
60	9	0	0	0
65	5	5	25	125
70	4	10	40	400
	N = 40		$\Sigma f_i d_i = -160$	$\Sigma f_i d_i^2 = 3050$

Standard deviation

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

$$= \sqrt{\frac{305\cancel{0}}{4\cancel{0}} - \left(\frac{-16\cancel{0}}{4\cancel{0}}\right)^2}$$

$$= \sqrt{76.25 - 16}$$

$$= \sqrt{60.25}$$

$$= 7.76$$

Answer : Standard deviation $\sigma \cong = 7.76$

7. In a study about viral fever, the number of people affected in a town were noted as

Age in years	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Number of people affected	3	5	16	18	12	7	4

Find its standard deviation.

Solution : Assumed mean method A = 35

Class Interval	Mid value x_i	f_i	$d_i = x_i - A$ $= x_i - 35$	$f_i d_i$	$f_i d_i^2$
0 – 10	5	3	–30	–90	2700
10 – 20	15	5	–20	–100	2000
20 – 30	25	16	–10	–160	1600
30 – 40	35	18	0	0	0
40 – 50	45	12	10	120	1200
50 – 60	55	7	20	140	2800
60 – 70	65	4	30	120	3600
		N = 65		$\Sigma f_i d_i = 30$	$\Sigma f_i d_i^2 = 13900$

Standard deviation

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

$$= \sqrt{\frac{13900}{65} - \left(\frac{30}{65}\right)^2}$$

$$= \sqrt{\frac{13900}{65} - \frac{900}{4225}} \quad = \sqrt{213.85 - 0.21}$$

$$= \sqrt{213.64}$$

$$= 14.62$$

Answer : S.D $\sigma \cong = 14.6$

8. The measurements of the diameters (in cms) of the plates prepared in a factory are given below. Find its standard deviation.

Diameter(cm)	21-24	25-28	29-32	33-36	37-40	41-44
Number of plates	15	18	20	16	8	7

Solution :

Step Deviation method

$A = 30.5$ $\quad\quad C = 4$

$$d_i = \frac{x_i - 30.5}{4}$$

Diameter	Mid value x_i	f_i	$di = \frac{x_i - A}{c}$	$f_i d_i$	$f_i d_i^2$
21 – 24	22.5	15	–2	–30	60
25 – 28	26.5	18	–1	–18	18
29 – 32	30.5	20	0	0	0
33 – 36	34.5	16	1	16	16
37 – 40	38.5	8	2	16	32
41 – 44	42.5	7	3	21	63
		N = 84		$\Sigma f_i d_i = 5$	$\Sigma f_i d_i^2 = 189$

Standard deviation

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

$$= 4 \times \sqrt{\frac{189}{84} - \left(\frac{5}{84}\right)^2}$$

$$= 4 \times \sqrt{\frac{189}{84} - \frac{.25}{7059}} \quad = 4 \times \sqrt{2.25 - 0.0035}$$

$$= 4 \times \sqrt{2.2465}$$

$$= 4 \times 1.5 = 6.0$$

Answer : S.D $\sigma \cong = 6$

9. The time taken by 50 students to complete a 100 meter race are given below. Find its standard deviation.

Time taken(sec)	8.5-9.5	9.5-10.5	10.5-11.5	11.5-12.5	12.5-13.5
Number of students	6	8	17	10	9

Solution :

Step Deviation method

$A = 11$ $\quad C = 1$ $\quad d_i = \frac{x_i - A}{c} = \frac{x_i - 11}{1}$

Time taken	Mid value x_i	f_i	$d_i = \frac{x_i - A}{c}$	$f_i d_i$	$f_i d_i^2$
8.5 – 9.5	9	6	-2	-12	24
9.5 – 10.5	10	8	-1	-8	8
10.5 – 11.5	11	17	0	0	0
11.5 – 12.5	12	10	1	10	10
12.5 – 13.5	13	9	2	18	36
		N = 50		$\Sigma f_i d_i = 8$	$\Sigma f_i d_i^2 = 78$

Standard deviation

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

$$= 1 \times \sqrt{\frac{78}{50} - \left(\frac{8}{50}\right)^2}$$

$$= \sqrt{\frac{78}{50} - \frac{64}{2500}} \qquad = \sqrt{1.56 - 0.026}$$

$$= \sqrt{1.534}$$

Answer : S.D $\sigma \cong$ = 1.24

Coefficient of Variation

Comparison of two data in terms of measures of central tendencies and dispersions in some cases will not be meaningful, because the variables in the data may not have same units of measurement.

For example consider the two data

	Weight	Price
Mean	8 kg	₹ 85
Standard deviation	1.5 kg	₹ 21.60

Here we cannot compare the standard deviations 1. 5kg and ₹21.60. For comparing two or more data for corresponding changes the relative measure of standard deviation, called "Coefficient of variation" is used.

Coefficient of variation of a data is obtained by dividing the standard deviation by the arithmetic mean. It is usually expressed in terms of percentage. This concept is suggested by one of the most prominent Statistician Karl Pearson.

Thus, coefficient of variation of first data (C.V1) = $\sigma_1/x_1 \times 100\%$ $= \frac{\sigma_1}{\bar{x}_1} \times 100\%$

and coefficient of variation of second data (C.V2) = $\sigma_2/ x_2 \times 100\%$ $= \frac{\sigma_2}{\bar{x}_2} \times 100\%$

The data with lesser coefficient of variation is more consistent or stable than the other data. Consider the two data

A	500	900	800	900	700	400
B	300	540	480	540	420	240

Then, we get

	Mean	Standard deviation
A	700	191.5
B	420	114.9

If we compare the mean and standard deviation of the two data, we think that the two datas are entirely different. But mean and standard deviation of *B* are 60% of that of *A*. Because of the smaller mean the smaller standard deviation led to the misinterpretation.

To compare the dispersion of two data, coefficient of variation = σ/*x* ×100% $= \frac{\sigma}{\overline{x}} \times 100\%$

The coefficient of variation of *A* = 191. 5/700 ×100% = 27. 4%

The coefficient of variation of *B* = 114. 9/420 ×100% = 27. 4%

Thus the two data have equal coefficient of variation. Since the data have equal coefficient of variation values, we can conclude that one data depends on the other. But the data values of *B* are exactly 60% of the corresponding data values of *A*. So they are very much related. Thus, we get a confusing situation.

To get clear picture of the given data, we can find their coefficient of variation. This is why we need coefficient of variation.

Example 15

The mean of a data is 25.6 and its coefficient of variation is 18.75. Find the standard deviation.

Solution

Mean $\overline{x}$ = 25. 6 , Coefficient of variation, C.V. = 18.75

Coefficient of variation, C.V. = σ/ $\overline{x}$ ×100%

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

$$18.75 = \frac{\sigma}{25.6} \times 100 \ ; \qquad \sigma = 4.8$$

Example 16

The following table gives the values of mean and variance of heights and weights of the 10th standard students of a school.

	Height	Weight
Mean	155 cm	46.50 kg^2
Variance	72.25 cm^2	28.09 kg^2

Which is more varying than the other?

Solution

For comparing two data, first we have to find their coefficient of variations

Mean $\overline{x}$ 1= 155cm, variance σ_1^2 = 72. 25 cm^2

Therefore standard deviation σ_1 = 8. 5

Coefficient of variation

$$C.V_1 = \frac{\sigma_1}{\overline{x}_1} \times 100\%$$

$$C.V_1 = \frac{8.5}{155} \times 100\% = 5.48\% \qquad \text{(for heights)}$$

Mean $\overline{x}$ 2 = 46.50 kg, Variance σ_2^2 = 28.09 kg2

Standard deviation σ_2 = 5. 3*kg*

Coefficient of variation

$$C.V_2 = \frac{\sigma_2}{\overline{x}_2} \times 100\%$$

$$C.V_2 = \frac{5.3}{46.50} \times 100\%$$

= 11. 40% (for weights)

$C.V_1$ = 5.48% and $C.V_2$ = 11.40%

Since $C.V_2 > C.V_1$, the weight of the students is more varying than the height.

17 The consumption of number of guava and orange on a particular week by a family are given below.

Number of Guavas	3	5	6	4	3	5	4
Number of Oranges	1	3	7	9	2	6	2

Which fruit is consistently consumed by the family?

First we find the coefficient of variation for guavas and oranges separately.

Number of guavas, $n = 7$

x_i	x_i^2
3	9
5	25
6	36
4	16
3	9
5	25
4	16
$\Sigma x_i = 30$	$\Sigma x_i^2 = 136$

$$\text{Mean } \bar{x}_1 = \frac{30}{7} = 4.29$$

$$\text{Standard deviation } \sigma_1 = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$$

$$\sigma_1 = \sqrt{\frac{136}{7} - \left(\frac{30}{7}\right)^2} = \sqrt{19.43 - 18.40} \simeq 1.01$$

Coefficient of variation for guavas

$$\text{C.V}_1 = \frac{\sigma_1}{\bar{x}_1} \times 100\% = \frac{1.01}{4.29} \times 100\% = 23.54\%$$

Number of oranges $n = 7$

x_i	x_i^2
1	1
3	9
7	49
9	81
2	4
6	36
2	4
$\Sigma x_i = 30$	$\Sigma x_i^2 = 184$

$$\text{Mean } \bar{x}_2 = \frac{30}{7} = 4.29$$

$$\text{Standard deviation } \sigma_2 = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$$

$$\sigma_2 = \sqrt{\frac{184}{7} - \left(\frac{30}{7}\right)^2} = \sqrt{26.29 - 18.40} = 2.81$$

Coefficient of variation for oranges

$$C.V_2 = \frac{\sigma_2}{\bar{x}_2} \times 100\% = \frac{2.81}{4.29} \times 100\% = 65.50\%$$

$C.V_1 = 23.54\%$, $C.V_2 = 65.50\%$ Since, $C.V_1 < C.V_2$, we can conclude that the consumption of guavas is more consistent than oranges.

Solution :

Find the coefficient of variation for city A and city B.

City A

$$\bar{x} = \frac{18+20+22+24+26}{5}$$

$$= \frac{110}{5} = 22$$

$\bar{x} = 22$

x_i	$d_i = x_i - \bar{x} = x_i - 22$	d_i^2
18	−4	16
20	−2	4
22	0	0
24	2	4
26	4	16
	$\Sigma d_i = 0$	$\Sigma d_i^2 = 40$

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

$$= \sqrt{\frac{40}{5}} = \sqrt{8}$$

$\sigma = 2.828$

$$C.V = \frac{\sigma}{x} \times 100$$

$$= \frac{2.828}{22} \times 100$$

$= 12.85\ \%$

City B

$$\bar{x} = \frac{11+14+15+17+18}{5}$$

$$= \frac{75}{5} = 15$$

$\bar{x} = 15$

x_i	$d_i = x_i - \bar{x} = x_i - 22$	d_i^2
18	−4	16
20	−2	4
22	0	0
24	2	4
26	4	16
	$\Sigma d_i = 0$	$\Sigma d_i^2 = 40$

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

$$= \sqrt{\frac{30}{5}} = \sqrt{6}$$

$\sigma = 2.449$

$$C.V = \frac{\sigma}{x} \times 100$$

$$= \frac{2.449}{15} \times 100$$

$= 16.33\ \%$

C.V of city A < C.V of city B.

Answer : City A is more consistent.