# 8

# MEASURES OF DISPERSION

## 8.1 INTRODUCTION

Although a good measure of central tendency represents the entire distribution but it has one major limitation that it does not study the variations in the distribution. The scatteredness or spreadness of the values around the central value is known as the *variation* or *dispersion*. The variation of the values is always studied with reference to the central value which may be either arithmetic mean or any other measure of central tendency. Following example will make this point clear :

| Series-X | Series-Y | Series-Z |
|---|---|---|
| 15 | 12 | 1 |
| 15 | 18 | 29 |
| 15 | 11 | 7 |
| 15 | 19 | 23 |
| 15 | 15 | 3 |
| 15 | 10 | 9 |
| 15 | 20 | 33 |
| $\Sigma X = 105$ | $\Sigma Y = 105$ | $\Sigma Z = 105$ |
| $\overline{X} = \frac{\Sigma X}{n} = \frac{105}{7} = 15$ | $\overline{Y} = \frac{\Sigma Y}{n} = \frac{105}{7} = 15$ | $\overline{Z} = \frac{\Sigma Z}{n} = \frac{105}{7} = 15$ |

The three series have the same average value. But the composition of each series is different. These series differ in respect of the variation of items. The extent of variability from the measure of central tendency constitutes the *measure of dispersion*. In other words, Measures of dispersion are the statistical tools to measure the variation or scatteredness or spread of the data about a measure of central tendency. Measures of dispersion are also called '*average of second order*'. This is so because these are based on the average of deviations of different data items from their average values.

## 8.2 DISPERSION–MEANING AND DEFINITIONS

Dispersion is an important statistical technique with the help of which we study the variability in a given series about its central value.

According to **A.L. Bowley**, ***"Dispersion is the measure of the variation of the items."***

According to **Brooks and Dick**, ***"Dispersion or spread is the degree of the scatter or variation of the variable about a central value."***

According to **M.R. Spiegel**, *"The degree to which numerical data tend to spread about an average value is called the variations or dispersion of the data."*

According to **L.R. Connor**, *"Dispersion is a measure of the extent to which the individual items vary."*

## 8.3 PROPERTIES OF A GOOD MEASURES OF DISPERSION

A good measure of dispersion should possess the following properties :

(*i*) It should be easy to understand and simple to calculate.

(*ii*) It should be rigidly defined.

(*iii*) It should be based on all the items of the series.

(*iv*) It should have the sampling stability.

(*v*) It should not be affected by the presence of extreme items in the data.

(*vi*) It should be capable of further algebraic or statistical treatments.

## 8.4 SIGNIFICANCE OR USES OR IMPORTANCE OF MEASURES OF DISPERSION

Measures of dispersion are useful in following ways :

**1. To Examine the Reliability of Average :** The measures of dispersion are used to judge the reliability of an average. If the variation or dispersion is small it shows a greater degree of uniformity in value of items and as such average may be regarded as highly representative. On the other hand if dispersion is large, it shows a lower degree of uniformity in the items and the average can not be regarded as true representative of the distribution.

**2. To Compare the Series on the Basis of Variability :** Relative measures of dispersion are used for comparing the two or more series in terms of their variability around the central value. The higher the variability in data, the lower the consistency in data and vice-versa.

**3. Helpful in Quality Control :** The measures of dispersion are helpful in statistical quality control. The extent of dispersion helps in determining the cause of variation in the quality of product. On the basis of this management can take up various steps to check the variations in quality of product.

**4. To Serve as Basis for Further Statistical Analysis :** The measures of dispersion are helpful in the study of other statistical measures such as correlation, regression, skewness, testing of hypothesis etc.

## 8.5 ABSOLUTE AND RELATIVE MEASURES OF DISPERSION

Measures of dispersion are of two types :

**(*i*) Absolute Measures :** The *absolute measures of dispersion* are always expressed in the same statistical units in which the original values are quoted. Absolute measures cannot be used for comparison of the variations in two or more series if they belong to different populations and have different units of measurement.

**(*ii*) Relative Measures :** The *relative measure of dispersion* is the ratio of a measure of absolute dispersion to an appropriate average and is called *Coefficient of dispersion*. It is independent of the unit of measurement because it is a pure number. Relative measures are used for making comparisons of variations of two or more than two series.

## 8.6 METHODS OF MEASURING DISPERSION

The methods to measure dispersion can be classified into following two categories :

**(A) Algebraic Methods :** The absolute measures and the corresponding relative measures for measuring dispersion are as follows :

| Absolute Measures | Relative Measures |
|---|---|
| 1. Range | 1. Coefficient of Range |
| 2. Semi Inter Quartile Range or Quartile Deviation | 2. Coefficient of Quartile Deviation |
| 3. Mean Deviation | 3. Coefficient of Mean Deviation |
| 4. Standard Deviation | 4. Coefficient of Standard Deviation |

The first two measures namely Range and Quartile deviation are also termed as *positional measures* as these measures depend upon the values of the variable at particular positions of the distribution.

**(B) Graphic Method :** Lorenz Curve

In the following sections we shall discuss each of these methods in detail.

## 8.7 RANGE

*Range* is defined as the difference between the value of largest and smallest item in the series. In other words, range is the difference between the two extreme observations of the distribution. It is the simplest measure of dispersion.

Thus, Range = Largest item − Smallest item

*i.e.* Range $$R = L - S \qquad ...(8.1)$$

where L is the value of largest item and S is the value of smallest item in the series.

The relative measure of range is called *coefficient of range* and is defined as

$$\text{Coefficient of Range} = \frac{L-S}{L+S} \qquad ...(8.2)$$

### Merits and Demerits of Range

**Merits :**

The following are the merits of range :

(*i*) Range is the easiest and simplest measure of dispersion

(*ii*) It is rigidly defined.

(*iii*) It gives us the broad picture of the variation present in the data.

(*iv*) It is very easy to calculate and understand.

(*v*) Range is useful in daily life in many ways such as (*a*) The study of quality control (*b*) Checking variations in stocks and share prices (*c*) forecasts etc.

**Demerits :**

The following are the limitations of range :

(*i*) It is a rough measure.

(*ii*) It is highly affected by the extreme values in the data.

(*iii*) It is not based on all the values of the distribution.

(*iv*) Range is too much affected by fluctuations in samples.

(*v*) Range cannot be used if we are dealing with open end classes.

Note In case of continuos frequency distribution, we can also take the smallest value as mid-value of smallest class-interval and the largest values as mid-value of largest class-interval instead of taking the smallest and largest values as lower limit of smallest class-interval and upper limit of highest class-interval respectively.

## CHECKPOINTS

1. What are the uses of measuring variation ? (G.N.D.U. B.C.A. April 2006)
2. What is meant by dispersion ? (P.U. B.C.A. Sept. 2006, 2007; G.N.D.U. B.Sc. I.T. April 2009)
3. Name various measures for finding dispersion. (G.N.D.U. B.C.A. April 2007, B.Sc. C.Sc. April 2005, I.T. April 2005)
4. What do you understand by Range ?
5. What are the merits and demerits of Range ?
6. Explain how Mean and Range help in interpretation of data ? (P.U. B.C.A. Sept. 2001)
7. Write a short note on different measures of dispersion. (Pbi.U. B.C.A. April 2006)

# ILLUSTRATIVE EXAMPLES

**Example 1.** The profit of a company for the last 8 years is given below. Calculate Range and coefficient of Range.

| Year : | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
|---|---|---|---|---|---|---|---|---|
| Profits : (in' 000 Rs.) | 40 | 30 | 80 | 100 | 120 | 90 | 200 | 230 |

**Sol.** Largest value in the data = L = 230

Smallest value in the data = S = 30

$$\text{Range} = L - S = 230 - 30 = 200$$

$$\text{Co-efficient of Range} = \frac{L-S}{L+S} = \frac{230-30}{230+30} = \frac{200}{260} = 0.77$$

Example 2. Given below are the heights (in cms.) of the students of two classes. Compare the Range of the heights

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Class I | 167 | 162 | 155 | 180 | 182 | 175 | 185 | 158 |
| Class II | 169 | 172 | 168 | 165 | 177 | 180 | 195 | 167 |

Sol. Class I

Largest value in the data = L = 185

Smallest value in the data = S = 155

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{185-155}{185+155} = \frac{30}{340} = 0.088$$

Class II

Largest value in the data = 195

Smallest value in the data = 165

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{195-165}{195+165} = \frac{30}{360} = 0.083$$

Clearly coefficient of range for heights of students of class I is more than that of class II. So heights of students of class I shows greater variation than that of class II.

Example 3. Find the Range and coefficient of dispersion of the following distribution

| Marks | No. of Students | Marks | No. of Students |
|---|---|---|---|
| 5 – 10 | 6 | 20 – 25 | 4 |
| 10 – 15 | 11 | 25 – 30 | 3 |
| 15 – 20 | 7 | 30 – 35 | 1 |

Sol. Here Largest value = L = 35

Smallest value = S = 5

$\therefore$ Range = L – S = 35 – 5 = 30

and Coefficient of range $= \frac{L-S}{L+S} = \frac{35-5}{35+5} = \frac{30}{40} = 0.75$

Alternatively, Smallest value = S = mid value of class interval 5 – 10 = 7.5

Largest value = L = mid value of class interval 30 – 35 = 32.5

Range = L – S = 32.5 – 7.5 = 25

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{32.5-7.5}{32.5+7.5} = \frac{25}{40} = 0.625$$

## EXERCISE 8.1

1. The following are the price of shares of a company from Monday to Saturday :

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

Calculate range and its coefficient.

2. Find the range and its coefficient from the following data :

| | | | | | | |
|---|---|---|---|---|---|---|
| Size : | 5 | 7 | 9 | 10 | 11 | 12 |
| Frequency : | 1 | 3 | 5 | 7 | 4 | 3 |

3. Calculate Range and its Coefficient from the following data :

| | | | | | |
|---|---|---|---|---|---|
| Class : | 1–10 | 11 – 20 | 21 – 30 | 31 – 40 | 41 – 50 |
| Frequency : | 3 | 7 | 20 | 15 | 6 |

4. Calculate the dispersion by Range method, from the data given as under

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Less than : | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 |
| Frequency : | 2 | 8 | 19 | 32 | 45 | 58 | 85 | 93 | 100 |

## ANSWERS

1. Rs. 90, 0.22 2. 7, 0.41 3. 50, 0.98 4. 9, 0.069

### 8.8 SEMI INTER QURTILE RANGE OR QUARTILE DEVIATION

*Quartile deviation* is an improved version of range. Range is very much affected by extreme values but quartile deviation involves two positional averages which are not affected by the extreme values in the data. It is based on lower quartile $Q_1$ and upper quartile $Q_3$. *Quartile range* is defined as the difference between third quartile ($Q_3$) and first quartile ($Q_1$).

*i.e.* $$\text{Quartile Range} = Q_3 - Q_1 \quad ...(8.3)$$

*Quartile deviation* is defined as the difference between upper quartile and lower quartile divided by 2.

*i.e.* $$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} \quad ...(8.4)$$

It is also called *Semi Inter Quartile Range*. The relative measure of quartile deviation is called *coefficient of quartile deviation* and is defined as

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \quad ...(8.5)$$

**Merits and Demerits of Quartile Deviation**

**Merits :**

The following are the merits of quartile deviation :

(*i*) Quartile deviation method of calculating variation is most appropriate method in case of open end frequency distributions.

(*ii*) Quartile deviation is easy to calculate and simple to understand.

(*iii*) Quartile deviation is more reliable than range.

(*iv*) Quartile deviation is particularly important and useful if we want to know the variability of the middle half of the data because 25 % of the items are less than $Q_1$ and 25% of the items are more than $Q_3$ thus including only 50 % in between $Q_1$ and $Q_3$.

**Demerits :**

The following are the demerits of quartile deviation :

(i) In the real sense coefficient of quartile deviation should not be termed as measure of dispersion, since it is concerned with the middle 50 % of the items. Therefore, it should be better known as measure of partition.

(ii) It is not suitable for further algebraic treatments because $Q_1$ and $Q_3$ are both positional measures.

(iii) It needs much labour for calculations but gives less dependable outcome in return.

(iv) It is too much affected by sampling fluctuations. Quartile deviation calculated for different samples taken from the same universe may vary considerably.

## CHECKPOINTS

1. What do you understand by Quartile Deviation ?
2. What are the merits and demerits of Quartile Deviation ?

# ILLUSTRATIVE EXAMPLES

**Example 1.** From following observations calculate quartile deviation and coefficient of quartile deviation :

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| X : | 59 | 62 | 65 | 64 | 63 | 61 | 60 | 56 | 58 | 66 |

**Sol.** First arranging the data in ascending order as follows :

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| X : | 56 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 |

Now $Q_1 = \text{Value of } \frac{n+1}{4} \text{th item}$

$= \text{Value of } \frac{10+1}{4} \text{th item} = \text{Value of } \frac{11}{4} \text{th item} = \text{Value of } 2.75\text{th item}$

$\therefore \quad Q_1 = 2\text{nd item} + 0.75\ (3\text{rd item} - 2\text{nd item})$

$= 58 + 0.75\ (59 - 58) = 58.75$

Also $Q_3 = \text{Value of } \frac{3(n+1)}{4} \text{th item}$

$= \text{Value of } \frac{3\,(10+1)}{4} = \text{Value of } \frac{33}{4} \text{th item} = \text{Value of } 8.25\text{th item.}$

$Q_3 = 8\text{th item} + 0.25\ (9\text{th item} - 8\text{th item})$

$= 64 + 0.25\ (65 - 64) = 64.25$

Now $Q.D. = \frac{Q_3 - Q_1}{2} = \frac{64.25 - 58.75}{2} = 2.75$

and Coefficient of Q.D. $= \frac{64.25 - 58.75}{64.25 + 58.75} = \frac{5.5}{123} = 0.045$ $\quad \frac{Q_3 - Q_1}{Q_3 + Q_1}$

**Example 2.** Compute quartile deviation and its coefficient from the following data :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Height (in inches) : | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 |
| No. of Students : | 15 | 20 | 32 | 35 | 33 | 22 | 20 | 10 | 8 |

**Sol.** First, we prepare the following table :

| Height X | No. of Students $f$ | $c.f.$ |
|---|---|---|
| 58 | 15 | 15 |
| 59 | 20 | 35 |
| 60 | 32 | 67 |
| 61 | 35 | 102 |
| 62 | 33 | 135 |
| 63 | 22 | 157 |
| 64 | 20 | 177 |
| 65 | 10 | 187 |
| 66 | 8 | 195 |
| | N = 195 | |

Now $Q_1$ = Value of $\frac{N+1}{4}$th item = Value of $\frac{195+1}{4}$th item = Value of 49th item.

The $c.f.$ next higher to 49 is 67.

$\therefore \quad Q_1 = 60$

Also $Q_3$ = Value of $\frac{3(N+1)}{4}$ th item = Value of $\frac{3(195+1)}{4}$ th item = Value of 147th item.

The $c.f.$ next higher to 147 is 157.

$\therefore \quad Q_3 = 63$

$\therefore$ Quartile deviation Q.D. = $\frac{Q_3 - Q_1}{2} = \frac{63-60}{2} = 1.5$

and Coefficient of Q.D $= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{63-60}{63+60} = \frac{3}{123} = 0.024$

**Example 3.** Compute the inter quartile range of marks of 59 students of English :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Marks : | 0–10 | 10–20 | 20–30 | 30–40 | 40–50 | 50–60 | 60–70 |
| No. of Students : | 4 | 8 | 11 | 15 | 12 | 6 | 3 |

(G.N.D.U. B.Sc. C.Sc. April 20

Sol. Let us denote the marks by X and number of students by $f$. First we shall prepare the following table :

| Marks X | No. of Students $f$ | $c.f.$ |
|---|---|---|
| 0 – 10 | 4 | 4 |
| 10 – 20 | 8 | 12 |
| $\boxed{20 - 30}$ | 11 | 23 |
| 30 – 40 | 15 | 38 |
| 40 – 50 | 12 | 50 |
| 50 – 60 | 6 | 56 |
| 60 – 70 | 3 | 59 |
| | N = 59 | |

Now $$Q_1 = \text{Value of } \frac{N}{4}\text{th item}$$

$$= \text{Value of } \frac{59}{4}\text{th item} = \text{Value of } 14.75\text{th item}$$

The $c.f.$ next higher to 14.75 is 23.

$\therefore$ $Q_1$ lies in the interval 20 – 30.

So $$Q_1 = L + \frac{\frac{N}{4} - c.f.}{f} \times i = 20 + \frac{14.75 - 12}{11} \times 10 = 22.5$$

$$Q_3 = \text{Value of } \frac{3N}{4}\text{th item} = \text{Value of } \frac{3\,(59)}{4}\text{th item}$$

$$= \text{Value of } 44.25\text{th item.}$$

The $c.f.$ next higher to 44.25 is 50.

$\therefore$ $Q_3$ lies in the interval 40–50

Now $$Q_3 = L + \frac{\frac{3N}{4} - c.f.}{f} \times i = 40 + \frac{44.25 - 38}{12} \times 10 = 45.21$$

So inter quartile range $= Q_3 - Q_1 = 45.21 - 22.5 = 22.71$

**Example 4.** Calculate quartile deviation and its relative measure for the following data :

| Variable | Frequency | Variable | Frequency |
|---|---|---|---|
| 20 – 29 | 306 | 50 – 59 | 96 |
| 30 – 39 | 182 | 60 – 69 | 42 |
| 40 – 49 | 144 | 70 – 79 | 24 |

**Sol.** First, we change the given data in the exclusive form of class intervals as follows :

| Variable (X) | Frequency ($f$) | $c.f.$ |
|---|---|---|
| 19.5 – 29.5 | 306 | 306 |
| 29.5 – 39.5 | 182 | 488 |
| 39.5 – 49.5 | 144 | 632 |
| 49.5 – 59.5 | 96 | 728 |
| 59.5 – 69.5 | 42 | 770 |
| 69.5 – 79.5 | 24 | 794 |
| | N = 794 | |

Now $Q_1$ = Value of $\frac{N}{4}$th item = Value of $\frac{794}{4}$ th item = Value of 198.5th item

The $c.f.$ next higher to 198.5 is 306

Therefore, $Q_1$ lies in 19.5 – 29.5

Now $Q_1 = L + \frac{\frac{N}{4} - c.f.}{f} \times i = 19.5 + \frac{10}{306}(198.5 - 0) = 19.5 + 6.49 = 25.99$

Also $Q_3$ = Value of $\frac{3N}{4}$th item = Value of $\frac{3\,(794)}{4}$ th item = Value of 595.5th item

The $c.f.$ next higher to 595.5 is 632

Therefore, $Q_3$ lies in 39.5 – 49.5

Now $Q_3 = L + \frac{\frac{3N}{4} - c.f.}{f} \times i = 39.5 + \frac{10}{144}(595.5 - 488) = 39.5 + 7.46 = 46.96$

$\therefore$ Quartile Deviation $= \frac{Q_3 - Q_1}{2} = \frac{46.96 - 25.99}{2} = 10.48$

and Coefficient of Q.D. $= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{46.96 - 25.99}{46.96 + 25.99} = 0.29$

# EXERCISE 8.2

1. Find out Quartile Deviation :

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| X : | 9 | 4 | 14 | 20 | 24 | 19 | 34 | 39 | 48 | 49 |

2. Calculate Semi-inter quartile Range and Quartile coefficient from the data :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Age (Years) : | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
| No. of Members : | 3 | 61 | 132 | 153 | 140 | 51 | 3 |

3. Calculate the semi-interquartile range and its coefficient from the following distribution :

| Marks | No. of Students | Marks | No. of Students |
|---|---|---|---|
| 0 – 4 | 4 | 8 – 12 | 12 |
| 4 – 6 | 6 | 12 – 18 | 7 |
| 6 – 8 | 8 | 18 – 20 | 2 |

4. Calculate Quartile deviation and its coefficient from the following data :

| | | | | | |
|---|---|---|---|---|---|
| Variable : | 20 – 30 | 30 – 40 | 40 – 50 | 50 – 60 | 60 – 70 |
| Frequency : | 4 | 6 | 8 | 6 | 4 |

5. Determine Quartile Deviation from following data :

| Marks | No. of Students | Marks | No. of Students |
|---|---|---|---|
| Below 30 | 69 | 60 and below 70 | 58 |
| 30 and below 40 | 167 | 70 and below 80 | 27 |
| 40 and below 50 | 207 | 80 and above | 10 |
| 50 and below 60 | 65 | | |

6. Determine Inter quartile range from the following data :

| | | | | | |
|---|---|---|---|---|---|
| Class Interval : | 1 – 5 | 6 – 10 | 11 – 15 | 16 – 20 | 21 – 25 |
| Frequency : | 5 | 10 | 15 | 6 | 4 |

## ANSWERS

1. 14.25 **2.** 10, 0.2 **3.** 2.915, 0.33 **4.** 10, 0.222
5. 8.26 **6.** 7.5

### 8.9 MEAN DEVIATION

While calculating the range we take into consideration only the largest and smallest items in the series ignoring all other items lying in between. Similarly in case of quartile deviation we study the variability of the middle 50% of the items. Thus, these two measures of dispersion are simply positional measures and not based on all the items. The *mean deviation* or *average deviation* is based on all the items.

The mean deviation is the mean of the deviations taken with positive signs only from the central value of the data which may be either mean or median or the mode depending upon the suitability of the problem under consideration. In simple words, mean deviations is defined as the arithmetic mean of the absolute deviations of values of a variable taken from either mean or median or mode of the variable.

**Merits and Demerits of Mean Deviation**

**Merits :** The following are the merits of mean deviation :

(i) It is easy to calculate and simple to understand.

(ii) It is a reliable measure of dispersion.

(iii) As compared to other measures of dispersion, mean deviation is not much affected by the extreme values.

(iv) It is based on all items in the series. While calculating mean deviation we give due consideration to all the items in the series.

(iv) It can be used for comparison purpose.

(v) It has much sampling stability as compared to range and quartile deviation.

**Demerits :** The following are the limitations of mean deviation :

(i) In the calculation of mean deviation we ignore the negative signs and take only positive signs. Thus, it makes the mean deviation incapable of further algebraic treatments.

(ii) Mean deviation is not much dependable if it is calculated from the mode, because mode itself is not a true representative of the items.

(iii) Its calculations are not easy in some cases. When the values of mean, median or mode are in fractions the calculations of mean deviation become difficult.

(iv) Mean deviation and its coefficient taken from mean, median or mode generally differ from each other.

## 8.9.1 CALCULATION OF MEAN DEVIATION IN CASE OF INDIVIDUAL SERIES

The following formulae are used to calculate mean deviation in case of individual series :

$$\text{Mean deviation from mean} = MD_{\overline{X}} = \frac{\Sigma|X-\overline{X}|}{n} \quad ...(8.6)$$

$$\text{Mean deviation from Median} = MD_{M} = \frac{\Sigma|X-M|}{n} \quad ...(8.7)$$

$$\text{Mean deviation from Mode} = MD_{Z} = \frac{\Sigma|X-Z|}{n} \quad ...(8.8)$$

The formulae for calculating coefficient of mean deviation for an individual series are as follows :

$$\text{Coefficient of mean deviation from mean} = \frac{MD_{\overline{X}}}{\overline{X}} \quad ...(8.9)$$

$$\text{Coefficient of mean deviation from median} = \frac{MD_{M}}{M} \quad ...(8.10)$$

$$\text{Coefficient of mean deviation from mode} = \frac{MD_{Z}}{Z} \quad ...(8.11)$$

## 9.2 CALCULATION OF MEAN DEVIATION IN CASE OF DISCRETE SERIES

Following are the formulae for calculating Mean Deviation from Mean, Median and Mode in case of discrete series :

$$\text{Mean deviation from mean} = MD_{\overline{X}} = \frac{\Sigma f|X-\overline{X}|}{N} \quad ...(8.12)$$

$$\text{Mean deviation from Median} = MD_M = \frac{\Sigma f|X-M|}{N} \quad ...(8.13)$$

$$\text{Mean deviations from Mode} = MD_Z = \frac{\Sigma f|X-Z|}{N} \quad ...(8.14)$$

where $N = \Sigma f$

Note For calculation of coefficient of mean deviation, the formulae are same as in case of individual series.

## 9.3 CALCULATION OF MEAN DEVIATION IN CASE OF CONTINUOUS SERIES

Calculations of mean deviation in the case of continuous series and discrete series is same except that we have to take the mid values of various class intervals *i.e.* $m$ in place of X.

### CHECKPOINTS

1. Define mean deviation. (G.N.D.U. B.Sc. I.T. April 2005)
2. Write down merits and demerits of mean deviation. (G.N.D.U. B.C.A. Sept. 2008)
3. Differentiate between range and mean deviation. (P.U. B.C.A. Sept. 2008)

# ILLUSTRATIVE EXAMPLES

**Example 1.** Calculate mean deviation from (*i*) arithmetic mean (*ii*) Median (*iii*) Mode, in respect of marks obtained by nine students given below and show that mean deviation from median is least.

Marks (out of 25) are 7, 4, 10, 9, 15, 12, 7, 9, 7.

**Sol.** Let us first calculate arithmetic mean, median and mode.

$$\text{Arithmetic Mean} = \overline{X} = \frac{7+4+10+9+15+12+7+9+7}{9} = \frac{80}{9} = 8.89$$

For median we arrange the marks in ascending order of magnitude as follows :

4, 7, 7, 7, 9, 9, 10, 12, 15

Now, $\quad \text{Median M} = \text{Value of } \frac{n+1}{2} \text{th item} = \text{Value of } \frac{9+1}{2} \text{th item} = \text{Value of 5th item}$

$\therefore \quad$ Median = 9

Further the value 7 occurs maximum number of times, therefore, Mode = 7.

Now for calculation of mean deviation from mean, median and mode we prepare the following table :

| Marks $X$ | $\bar{X} = 8.89$ $\lvert X - \bar{X} \rvert$ | $M = 9$ $\lvert X - M \rvert$ | $Z = 7$ $\lvert X - Z \rvert$ |
|---|---|---|---|
| 7 | 1.89 | 2 | 0 |
| 4 | 4.89 | 5 | 3 |
| 10 | 1.11 | 1 | 3 |
| 9 | 0.11 | 0 | 2 |
| 15 | 6.11 | 6 | 8 |
| 12 | 3.11 | 3 | 5 |
| 7 | 1.89 | 2 | 0 |
| 9 | 0.11 | 0 | 2 |
| 7 | 1.89 | 2 | 0 |
| $\Sigma X = 80$ | $\Sigma \lvert X - \bar{X} \rvert = 21.11$ | $\Sigma \lvert X - M \rvert = 21$ | $\Sigma \lvert X - Z \rvert = 23$ |

(*i*) $$MD_{\bar{X}} = \frac{\Sigma \lvert X - \bar{X} \rvert}{n} = \frac{21.11}{9} = 2.34$$

(*ii*) $$MD_M = \frac{\Sigma \lvert X - M \rvert}{n} = \frac{21}{9} = 2.33$$

(*iii*) $$MD_Z = \frac{\Sigma \lvert X - Z \rvert}{n} = \frac{23}{9} = 2.56$$

Clearly mean deviation from the median is least as compared to other values of mean deviations.

**Example 2.** Calculate mean deviation about mean from the following data :

| | | | | |
|---|---|---|---|---|
| $X$ : | 10 | 12 | 14 | 15 |
| $f$ : | 5 | 14 | 20 | 11 |

(G.N.D.U. B.Sc. I.T. April 2005)

**Sol.** First, we prepare the following table :

| $X$ | $f$ | $fX$ | $\bar{X} = 13.26$ $\lvert X - \bar{X} \rvert$ | $f \lvert X - \bar{X} \rvert$ |
|---|---|---|---|---|
| 10 | 5 | 50 | 3.26 | 16.3 |
| 12 | 14 | 168 | 1.26 | 17.64 |
| 14 | 20 | 280 | 0.74 | 14.8 |
| 15 | 11 | 165 | 1.74 | 19.14 |
| | $N = 50$ | $\Sigma fX = 663$ | | $\Sigma f \lvert X - \bar{X} \rvert = 67.88$ |

Now Mean $$\bar{X} = \frac{\Sigma fX}{N} = \frac{663}{50} = 13.26$$

$$\therefore \quad MD_{\bar{X}} = \frac{\Sigma f \lvert X - \bar{X} \rvert}{N} = \frac{67.88}{50} = 1.36$$

**Example 3.** Find the mean deviation for the following grouped data :

| | | | | |
|---|---|---|---|---|
| Class Interval : | 2 – 4 | 4 – 6 | 6 – 8 | 8 – 10 |
| No. of Occurrence : | 3 | 4 | 2 | 1 |

(G.N.D.U. B.Sc. C.Sc. April 2004)

**Sol.** First we construct the following table :

| Class Interval (X) | Frequency ($f$) | Mid-value ($m$) | $fm$ | $\overline{X} = 5.2$ $\|m - \overline{X}\|$ | $f\|m - \overline{X}\|$ |
|---|---|---|---|---|---|
| 2 – 4 | 3 | 3 | 9 | 2.2 | 6.6 |
| 4 – 6 | 4 | 5 | 20 | 0.2 | 0.8 |
| 6 – 8 | 2 | 7 | 14 | 1.8 | 3.6 |
| 8 – 10 | 1 | 9 | 9 | 3.8 | 3.8 |
| | N = 10 | | $\Sigma fm = 52$ | | $\Sigma f\|m - \overline{X}\| = 14.8$ |

Now, Mean $\overline{X} = \dfrac{\Sigma fm}{N} = \dfrac{52}{10} = 5.2$

$\therefore$ Mean deviation about mean $MD_{\overline{X}} = \dfrac{\Sigma f|m - \overline{X}|}{N} = \dfrac{14.8}{10} = 1.48$

**Example 4.** Find mean deviation from median and its coefficient from the data :

| Class Interval | Frequency | Class Interval | Frequency |
|---|---|---|---|
| 0 – 1000 | 4 | 0 to – 1000 | 6 |
| 1000 – 2000 | 5 | – 1000 to – 2000 | 8 |
| 2000 – 3000 | 10 | – 2000 to – 3000 | 10 |
| 3000 – 4000 | 30 | | |
| 4000 – 5000 | 15 | | |
| 5000 – 6000 | 10 | | |

**Sol.** First, we write the given data in ascending order and prepare the following table :

| Class Interval (X) | Mid Value ($m$) | Frequency ($f$) | $c.f.$ | $m = 3200$ $\|m - M\|$ | $f\|m - M\|$ |
|---|---|---|---|---|---|
| –3000 to – 2000 | – 2500 | 10 | 10 | 5700 | 57000 |
| –2000 to –1000 | – 1500 | 8 | 18 | 4700 | 37600 |
| –1000 to 0 | – 500 | 6 | 24 | 3700 | 22200 |
| 0 to 1000 | 500 | 4 | 28 | 2700 | 10800 |
| 1000 to 2000 | 1500 | 5 | 33 | 1700 | 8500 |
| 2000 to 3000 | 2500 | 10 | 43 | 700 | 7000 |
| 3000 to 4000 | 3500 | 30 | 73 | 300 | 9000 |
| 4000 to 5000 | 4500 | 15 | 88 | 1300 | 19500 |
| 5000 to 6000 | 5500 | 10 | 98 | 2300 | 23000 |
| | | N = 98 | | | $\Sigma f\|m - M\| = 194600$ |

Now Median = Value of $\frac{N}{2}$th item = Value of $\frac{98}{2}$th item = Value of 49th item

The *c.f.* next higher to 49 is 73

∴ Median lies between 3000 – 4000

So $\text{Median} = L + \frac{\frac{N}{2} - c.f.}{f} \times i = 3000 + (49-43)\frac{1000}{30} = 3000 + 6 \times \frac{100}{3} = 3200$

∴ M.D. from Median $MD_M = \frac{\Sigma f\,|m - M|}{N} = \frac{194600}{98} = 1985.7$

and Coefficient of M.D. from Median $= \frac{MD_M}{\text{Median}} = \frac{1985.7}{3200} = 0.62$

# EXERCISE 8.3

1. Calculate mean deviation from median and coefficient of mean deviation from the following data :

20, 22, 25, 38, 40, 50, 65, 70, 75. (P.U. B.C.A. Sept. 2005)

2. With median as base, calculate mean deviation and compare the variability of two series A and B.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Series A : | 3484 | 4572 | 4124 | 3682 | 5624 | 4388 | 3680 | 4308 |
| Series B : | 487 | 508 | 620 | 382 | 408 | 266 | 186 | 218 |

3. Find the mean deviation from A.M. of the following data :

| | | | | | | |
|---|---|---|---|---|---|---|
| X : | 3 | 5 | 7 | 9 | 11 | 13 |
| F(x): | 2 | 7 | 10 | 9 | 5 | 2 |

(G.N.D.U. B.Sc. I.T. April 2007, B.C.A. April 2009)

4. Calculate Coefficient of mean deviation of the following data from mean, median and mode.

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

5. Calculate Mean deviation and its coefficient from the following information

| | | | | | | |
|---|---|---|---|---|---|---|
| Income (Rs.) : | 40–50 | 50–60 | 60–70 | 70–80 | 80–90 | 90–100 |
| No. of Workers: | 5 | 10 | 20 | 25 | 15 | 5 |

6. Compute Mean deviation from mean, median and mode from the following data

| | | | | | |
|---|---|---|---|---|---|
| Class | 0 – 10 | 10 – 20 | 20 – 30 | 30 – 40 | 40 – 50 |
| Frequency | 3 | 8 | 15 | 20 | 25 |
| Class | 50 – 60 | 60 – 70 | 70 – 80 | 80 – 90 | |
| Frequency | 10 | 9 | 6 | 4 | |

7. Calculate mean deviation from Median from the following data :

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Marks less than : | 80 | 70 | 60 | 50 | 40 | 30 | 20 | 10 |
| No. of Students : | 100 | 90 | 80 | 60 | 32 | 20 | 13 | 5 |

8. Find the mean deviation about the arithmetic mean for the following data

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Marks : | 30 – 39 | 40 – 49 | 50 – 59 | 60 – 69 | 70 – 79 | 80 – 89 | 90 – 99 |
| No. of Students : | 2 | 3 | 11 | 20 | 32 | 25 | 7 |

9. Calculate mean deviation from mean and median from the following data.

| | | | | | | |
|---|---|---|---|---|---|---|
| Class Interval : | 0 – 9 | 10 – 19 | 20 – 29 | 30 – 39 | 40 – 49 | 50 – 59 |
| Frequency : | 15 | 36 | 53 | 42 | 17 | 2 |

## ANSWERS

1. 17.22, 0.43
2. M.D. (Series A) = 490.25, M.D. (Series B) = 121.375,
   Series B has more variability as coefficient of M.D. is more for series B
3. 2.12 **4.** 0.239, 0.24, 0.24 **5.** 10.375, 0.144
6. 14.992, 15.072, 15 **7.** 14.286 **8.** 10.36 **9.** 9.22, 9.21

## 8.10 STANDARD DEVIATION

*Standard deviation* as a measure of dispersion was first given by **Prof. Karl Pearson** in 1893. It is considered as the best measure of dispersion and is widely used. Standard deviation is defined as the square root of the arithmetic mean of the squares of deviations of the items from arithmetic mean. Standard deviation has only arithmetic mean as its base, and as such is the most stable and rigidly defined measure of dispersion. It is denoted by symbol $\sigma$ (read as 'sigma').

*i.e.* S.D. $$\sigma = \sqrt{\frac{\Sigma(X-\overline{X})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n}} \quad \text{(for individual series)} \qquad ...(8.15)$$

where $x = X - \overline{X}$ and n = number of observations

and S.D. $$\sigma = \sqrt{\frac{\Sigma f\,(X-\overline{X})^2}{N}} \quad \text{(for a frequency distribution)} \qquad ...(8.16)$$

where $N = \Sigma f$ and X is the value of the variable (in case of discrete frequency distribution) or mid value of class interval (in case of continuous frequency distribution).

**Variance and Mean Square Deviation**

The *variance* is defined as the square of the standard deviation and is denoted by $\sigma^2$

*i.e.*
$$\sigma^2 = \frac{\Sigma\,(X-\overline{X})^2}{n} \quad \text{(for individual series)} \qquad ...(8.17)$$

and $$\sigma^2 = \frac{\Sigma f(X-\overline{X})^2}{N} \quad \text{(for a frequency distribution)} \qquad ...(8.18)$$

where $N = \Sigma f$

The *mean square deviation* is denoted by $S^2$ and is defined as

$$S^2 = \frac{\Sigma (X-A)^2}{n} \quad \text{(for individual series)} \qquad ...(8.19)$$

and $$S^2 = \frac{\Sigma f(X-A)^2}{N} \quad \text{(for a frequency distribution)} \qquad ...(8.20)$$

where $N = \Sigma f$ and A is any arbitrary number.

The square root of mean square deviation is called *root mean square deviation* and is denoted by S

*i.e.* $$S = \sqrt{\frac{\Sigma (X-A)^2}{n}} \quad \text{(for individual series)} \qquad ....(8.21)$$

and $$S = \sqrt{\frac{\Sigma f(X-A)^2}{N}} \quad \text{(for a frequency distribution)} \qquad ...(8.20)$$

where $N = \Sigma f$ and A is any arbitrary mean.

**Properties of Standard Deviation**

**Property I.** The standard deviation is independent of change of origin but not of scale *i.e.* if the variable X is transformed to a new variable U by using transformation $U = \frac{X-A}{h}$ then $\sigma_X = h\sigma_U$.

**Proof :** Since $$U = \frac{X-A}{h}$$

$\therefore$ $$X = A + hU$$

So $$\overline{X} = \frac{\Sigma X}{n} = \frac{\Sigma (A+hU)}{n} = \frac{\Sigma A + \Sigma hU}{n} = \frac{nA}{n} + \frac{h\Sigma U}{n} = A + h\overline{U}$$

*i.e.* $$\overline{X} = A + h\overline{U}$$

Now $$\sigma_X = \sqrt{\frac{1}{n}\Sigma(X-\overline{X})^2} = \sqrt{\frac{1}{n}\Sigma[A+hU-(A+h\overline{U})]^2}$$

$$= \sqrt{\frac{1}{n}\Sigma[h(U-\overline{U})]^2} = \sqrt{\frac{h^2}{n}\Sigma(U-\overline{U})^2}$$

$$= h\sqrt{\frac{1}{n}\Sigma(U-\overline{U})^2} = h\sigma_U$$

*i.e.* $$\sigma_X = h\sigma_U \qquad ...(8.23)$$

Thus standard deviation is independent of origin but not of scale.

**Property II. The standard deviation is minimum value of root mean square deviation.**

**Proof :** Consider mean square deviation $S^2 = \frac{1}{n}\Sigma(X-A)^2$, where A is any arbitrary constant.

Now
$$S^2 = \frac{1}{n}\Sigma[(X-\overline{X})+(\overline{X}-A)]^2$$
$$= \frac{1}{n}\Sigma[(X-\overline{X})^2+(\overline{X}-A)^2+2(X-\overline{X})(\overline{X}-A)]$$
$$= \frac{1}{n}\Sigma(X-\overline{X})^2+(\overline{X}-A)^2\frac{\Sigma 1}{n}+\frac{2}{n}(\overline{X}-A)\Sigma(X-\overline{X})$$
$$= \sigma^2+(\overline{X}-A)^2+0$$
$$\left[\because \Sigma(X-\overline{X}) = \Sigma X-\overline{X}\Sigma 1 = \Sigma X-\frac{\Sigma X}{n}.n = \Sigma X-\Sigma X = 0\right]$$

*i.e.* $$S^2 = \sigma^2+(\overline{X}-A)^2 \qquad ...(8.24)$$

*i.e.* $$S^2 = \sigma^2 + \text{a non-negative quantity}$$

*i.e.* $$S^2 \geq \sigma^2 \quad \text{or} \quad \sigma^2 \leq S^2$$

Further $\sigma^2 = S^2$ if an only if $(\overline{X}-A)^2 = 0$ (using equation (8.24))

*i.e.* if and only if $\overline{X} = A$

*i.e.* $S^2$ is least when $A = \overline{X}$ and in that case $S^2 = \sigma^2$

Thus mean square deviation and hence root mean square deviation is least when deviations are taken from mean and standard deviation is minimum value of root mean square deviation.

**Property III. Standard deviation is always less than or equal to range.**

**Proof :** Suppose that we have $n$ observations $X_1, X_2, ......, X_n$ with range R.

Now
$$\sigma^2 = \frac{1}{n}(X-\overline{X})^2$$
$$= \frac{1}{n}[(X_1-\overline{X})^2+(X_2-\overline{X})^2+......+(X_n-\overline{X})^2$$
$$\leq \frac{1}{n}[R^2+R^2+......+R^2] \quad [\because X_1-\overline{X} \leq R,\ X_2-\overline{X} \leq R, ....., X_n-\overline{X} \leq R]$$
$$= \frac{1}{n}.nR^2 = R^2$$

*i.e.*
$$\Rightarrow \sigma^2 \leq R^2$$

Hence
$$\sigma \leq R \qquad ...(8.25)$$
$$\text{S.D.} \leq \text{Range}$$

**Property IV.** Standard deviation of first $n$ natural numbers is $\sqrt{\dfrac{n^2-1}{12}}$.

**Proof :** First $n$ natural numbers are 1, 2, 3, ......, $n$.

Mean of first $n$ natural numbers $\overline{X} = \dfrac{1+2+3+.....+n}{n} = \dfrac{1}{n} \times \dfrac{n(n+1)}{2} = \dfrac{n+1}{2}$

Further variance $\sigma^2 = \dfrac{1}{n}\Sigma(X-\overline{X})^2 = \dfrac{1}{n}\Sigma(X^2+\overline{X}^2-2X\overline{X})$

$$= \frac{1}{n}\Sigma X^2 + \overline{X}^2 \frac{\Sigma 1}{n} - 2\overline{X}\frac{\Sigma X}{n}$$

$$= \frac{1^2+2^2+.....+n^2}{n} + \left(\frac{n+1}{2}\right)^2 - 2\left(\frac{n+1}{2}\right)^2 \qquad \left[\because \ \Sigma 1 = n \text{ and } \frac{\Sigma X}{n} = \overline{X}\right]$$

$$= \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{(n+1)}{2}\left[\frac{2n+1}{3} - \frac{n+1}{2}\right]$$

$$= \frac{(n+1)}{2}\left[\frac{4n+2-3n-3}{6}\right] = \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}.$$

Hence $\sigma = \sqrt{\dfrac{n^2-1}{12}}$ ...(8.2

**Property V.** Standard deviation is suitable for further algebraic treatment.

*e.g.* if we have information about the size, mean and standard deviation of two or more groups th we can calculate the standard deviation of the group formed by combining all the groups.

(See Section 8.10.

**Property VI.** For a symmetrical bell shaped distribution, the standard deviation $\sigma$ and mean $\overline{X}$ satisfie approximately the following area properties :

(*a*) 67.45% observations lie in the range $\overline{X} \pm \sigma$

(*b*) 95.44% of the observations lie in the range $\overline{X} \pm 2\sigma$

(*c*) 99.73% of the observation lie in the range $\overline{X} \pm 3\sigma$

**Property VII.** The standard deviation is zero if and only if all the observations in a series are equal.

For example : Suppose that a variable X has a following 5 values :

4, 4, 4, 4, 4.

Here $\overline{X} = \dfrac{4+4+4+4+4}{5} = 4$

Now we prepare the following table :

| | | | | | |
|---|---|---|---|---|---|
| X | 4 | 4 | 4 | 4 | 4 |
| $X-\overline{X}$ | 0 | 0 | 0 | 0 | 0 |

$$\therefore \quad \text{S.D.} = \frac{1}{n}\Sigma(X-\overline{X})^2 = \frac{1}{5}\Sigma 0^2 = 0$$

Thus if all the observations are equal then standard deviation will be zero.

Conversely if the standard deviation of a given set of observations is zero then all the observations will be equal.

**Property VIII.** The approximate relationship between quartile deviation (Q.D.), mean deviation (M.D.) and standard deviation (S.D.) is

$$\text{Q.D.} = \frac{2}{3}\text{ S.D.} \quad \text{and} \quad \text{M.D.} = \frac{4}{5}\text{ S.D.}$$

$$\text{i.e.} \quad \text{Q.D. : M.D. : S.D. :: } 10 : 12 : 15 \qquad ...(8.27)$$

**Property IX.** For any distribution, standard deviation is not less than mean deviation from mean.

## Merits and Demerits of Standard Deviation

**Merits :**

The following are the merits of standard deviation :

(*i*) It is rigidly defined and as such it is quite stable and dependable

(*ii*) Since it is based upon the arithmetic mean, therefore, it has all the qualities which the arithmetic mean possesses.

(*iii*) Standard deviation is based upon all the observations given in the data. Each and every item in the data has its due share in determining its value.

(*iv*) It is the only dependable and reliable measure of dispersion for the purpose of comparing variability of two or more series.

(*v*) It is the only measure of dispersion which is used to find the combined measure of dispersion when the standard deviations of a few samples are given separately.

(*vi*) It is capable of further algebraic treatment.

(*vii*) Squaring the deviations $(X-\overline{X})$ removes the limitation of ignoring signs of deviations in calculating mean deviation.

**Demerits :**

The following are the demerits of standard deviation :

(*i*) Comparatively it is difficult method to understand and to compute.

(*ii*) It is too much affected by the extreme values in the data.

(*iii*) It is not familiar to common people.

**Different formulae for calculating Variance (and Standard deviation)**

We know that the variance of a variable X is given by

$$\sigma_X^2 = \frac{1}{n}\Sigma(X-\overline{X})^2 \quad \text{(for individual series)} \qquad ...(8.28)$$

and

$$\sigma_X^2 = \frac{1}{N}\Sigma f(X-\overline{X})^2 \quad \text{(for a frequency distribution)} \qquad ...(8.29)$$

If the arithmetic mean $\overline{X}$ for the given data comes out to be an integer then the standard deviation (and variance) can be calculated by using the formula (8.28) or (8.29).

But if $\overline{X}$ is in fractions then computation of $\sigma_X^2$ by above formula becomes very cumbersome and time consuming. In order to overcome this difficulty, we shall now develop different versions of formula (8.28) or (8.29) as follows :

**Formula 1.** $$\sigma_X^2 = \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2 \quad \text{(for individual series)} \qquad ...(8.30)$$

and

$$\sigma_X^2 = \frac{\Sigma fX^2}{N} - \left(\frac{\Sigma fX}{N}\right)^2 \quad \text{(for a frequency distribution)} \qquad ...(8.31)$$

where $N = \Sigma f$.

**Proof :** We know that for an individual series

$$\sigma_X^2 = \frac{1}{n}\Sigma(X-\overline{X})^2$$

$$= \frac{1}{n}\Sigma(X^2+\overline{X}^2-2X\overline{X})$$

$$= \frac{1}{n}\Sigma X^2 + \overline{X}^2\frac{\Sigma 1}{n} - 2\overline{X}\frac{\Sigma X}{n}$$

$$= \frac{1}{n}\Sigma X^2 + \overline{X}^2\frac{n}{n} - 2\overline{X}\,\overline{X} \qquad \left[\because \ \Sigma 1 = n \text{ and } \overline{X} = \frac{\Sigma X}{n}\right]$$

$$= \frac{1}{n}\Sigma X^2 - \overline{X}^2 = \frac{1}{n}\Sigma X^2 - \left(\frac{\Sigma X}{n}\right)^2$$

$$\therefore \quad \sigma_X^2 = \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2$$

Similarly for grouped data $\sigma_X^2 = \dfrac{\Sigma fX^2}{N} - \left(\dfrac{\Sigma fX}{N}\right)^2$, where $N = \Sigma f$.

**Note** The formula (8.30) or (8.31) are used if $\overline{X}$ comes out in fractions and the variable X does not assume large values.

**Formula 2.** $$\sigma_X^2 = \sigma_U^2 = \frac{\Sigma U^2}{n} - \left(\frac{\Sigma U}{n}\right)^2 \quad \text{(for individual series)} \qquad ...(8.32)$$

and $$\sigma_X^2 = \sigma_U^2 = \frac{\Sigma f U^2}{N} - \left(\frac{\Sigma f U}{N}\right)^2 \quad \text{(for a frequency distribution)} \qquad ...(8.33)$$

where $U = X - A$ and $N = \Sigma f$.

**Proof :** We know that the variance (and standard deviation) is independent of change of origin.

*i.e.* if the variable X is transformed to a new variable U by using the transformation $U = X - A$

then $\sigma_X^2 = \sigma_U^2$

$\therefore$ for individual series, $$\sigma_X^2 = \sigma_U^2 = \frac{\Sigma U^2}{N} - \left(\frac{\Sigma U}{N}\right)^2$$

Similarly for a frequency distribution, $$\sigma_X^2 = \sigma_U^2 = \sqrt{\frac{\Sigma f U^2}{N} - \left(\frac{\Sigma f U}{N}\right)^2}$$

where $N = \Sigma f$.

**Notes** (*a*) The formula (8.32) or (8.33) are used if the variable X assumes large values.

(*b*) From above formula, it immediately follows that $\text{Var}(X \pm A) = \text{Var}(X)$.

**Formula 3.** $$\sigma_X^2 = h^2\sigma_U^2 = h^2\left[\frac{\Sigma U^2}{n} - \left(\frac{\Sigma U}{n}\right)^2\right] \quad \text{(for individual series)} \qquad ...(8.34)$$

and $$\sigma_X^2 = h^2\sigma_U^2 = h^2\left[\frac{\Sigma f U^2}{N} - \left(\frac{\Sigma f U}{N}\right)^2\right] \quad \text{(for a frequency distribution)} \qquad ...(8.35)$$

where $U = \dfrac{X - A}{h}$ and $N = \Sigma f$.

**Proof :** We know that the variance (and standard deviation) is independent of change of origin but not scale.

*i.e.* if the variable X is transformed to a new variable by using the transformation $U = \dfrac{X - A}{h}$ then

$$\sigma_X^2 = h^2\sigma_U^2$$

$\therefore$ for individual series, $$\sigma_X^2 = h^2\sigma_U^2 = h^2\left[\frac{\Sigma U^2}{n} - \left(\frac{\Sigma U}{n}\right)^2\right]$$

Similarly for grouped data, $\sigma_X{}^2 = h^2\left[\frac{\Sigma f U^2}{N} - \left(\frac{\Sigma f U}{N}\right)^2\right]$

where $N = \Sigma f$.

**Notes** (*a*) The formula (8.34) and (8.35) are used if all the values of X–A have some common factor $h$.

(*b*) From above formula, it immediately follows that $\text{Var}(aX \pm b) = a^2\,\text{Var}(X)$.

### 8.10.1 CALCULATIONS FOR VARIANCE (AND STANDARD DEVIATION)

The methods discussed above for calculating variance (and standard deviation) can be classified into following three categories :

**1. Direct Methods :**

(*i*) If deviations are taken from actual mean then

$$\sigma^2 = \frac{\Sigma(X-\overline{X})^2}{n} = \frac{\Sigma x^2}{n} \quad \text{(for individual series)} \qquad \text{...(8.36)}$$

where $n$ = number of observations and $x = X - \overline{X}$.

and $$\sigma^2 = \frac{\Sigma f(X-\overline{X})^2}{N} = \frac{\Sigma f x^2}{N} \quad \text{(for a frequency distribution)} \qquad \text{...(8.37)}$$

where $N = \Sigma f$ and $x = X - \overline{X}$.

(*ii*) If no deviation are taken at all then

$$\sigma^2 = \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2 \quad \text{(for individual series)} \qquad \text{...(8.38)}$$

where $n$ = number of observations

and $$\sigma^2 = \frac{\Sigma f X^2}{N} - \left(\frac{\Sigma f X}{N}\right)^2 \quad \text{(for a frequency distribution)} \qquad \text{...(8.39)}$$

where $N = \Sigma f$.

**2. Short-cut Method :**

If the deviations are taken from assumed mean then

$$\sigma^2 = \frac{\Sigma dx^2}{n} - \left(\frac{\Sigma dx}{n}\right)^2 \quad \text{(for individual series)} \qquad \text{...(8.40)}$$

where $n$ = number of observations and $dx = X - A$

and $$\sigma^2 = \frac{\Sigma f\,dx^2}{N} - \left(\frac{\Sigma f\,dx}{N}\right)^2 \quad \text{(for a frequency distribution)} \qquad \text{...(8.41)}$$

where $N = \Sigma f$ and $dx = X - A$.

**3. Step deviation method :**

If some common factor can be taken from $dx$ then

$$\sigma^2 = c^2\left[\frac{\Sigma d'x^2}{n} - \left(\frac{\Sigma d'x}{n}\right)^2\right] \text{ (for individual series)} \qquad ...(8.42)$$

where $n$ = number of observations and $d'x = \dfrac{dx}{c}$

and
$$\sigma^2 = c^2\left[\frac{\Sigma f d'x^2}{N} - \left(\frac{\Sigma f d'x}{N}\right)^2\right] \text{ (for a frequency distribution)} \qquad ...(8.43)$$

where $N = \Sigma f$ and $d'x = \dfrac{dx}{c}$.

Note To find the variance or standard deviation for a continuous series, the procedure is same as that for a discrete series except that X is replaced by the mid values $m$ of the class-intervals.

## CHECKPOINTS

1. Define Standard Deviation.

   (G.N.D.U. B.Sc. I.T. April 2005)
2. What is the relationship between mean square deviation and variance ?
3. Is standard deviation independent of scale and origin ?
4. What is the standard deviation of first $n$ natural numbers ?
5. What is the relationship between quartile deviation, mean deviation and standard deviation ?
6. What are the merits and demerits of standard deviation ?
7. Compare range and standard deviation as measures of dispersion.

   (P.U. B.C.A. April 2007, 2008)
8. How mean and standard deviation help in interpretation of data ? (P.U. B.C.A. Sept. 2008)
9. Differentiate between M.D. and S.D. Which is a better measure and why ?

   (G.N.D.U. B.C.A. April 2008)

# ILLUSTRATIVE EXAMPLES

**Example 1.** The mean of 100 items is 50 and standard deviation is 5. Find the sum and sum of squares of all the 100 items.

**Sol.** Given n = 100, $\overline{X} = 50, \sigma = 5$

We know $\overline{X} = \dfrac{\Sigma X}{n} \Rightarrow \Sigma X = n\overline{X} \Rightarrow \Sigma X = 100 \times 50 = 5000$

and $\sigma^2 = \dfrac{\Sigma X^2}{n} - \left(\dfrac{\Sigma X}{n}\right)^2 \Rightarrow \Sigma X^2 = n(\overline{X}^2 + \sigma^2)$

$$= 100\,[(50)^2 + 5^2] = 100\,(2525) = 252500$$

**Example 2.** Ten students of a class obtained marks in a subject out of 100 as follow :

| S. No. : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks : | 5 | 10 | 20 | 25 | 40 | 42 | 45 | 48 | 70 | 80 |

Find the standard deviation.

**Sol.** First, we prepare the following table :

| Marks (X) | $X^2$ |
|---|---|
| 5 | 25 |
| 10 | 100 |
| 20 | 400 |
| 25 | 625 |
| 40 | 1600 |
| 42 | 1764 |
| 45 | 2025 |
| 48 | 2304 |
| 70 | 4900 |
| 80 | 6400 |
| $\Sigma X = 385$ | $\Sigma X^2 = 20143$ |

Now $\text{S.D.} = \sqrt{\dfrac{\Sigma X^2}{n} - \left(\dfrac{\Sigma X}{n}\right)^2} = \sqrt{\dfrac{20143}{10} - \left(\dfrac{385}{10}\right)^2}$

$$= \sqrt{2014.3 - 1482.25} = \sqrt{532.05} = 23.07$$

**Example 3.** Find the standard deviation of the monthly salaries of 10 persons given below :

| Persons : | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| Salaries ('000 Rs.) | 120 | 110 | 115 | 122 | 126 | 140 | 125 | 121 | 120 | 131 |

Sol. First, we prepare the following table :

| Salaries ('000 Rs.) (X) | $\overline{X} = 123$ $x = X - \overline{X}$ | $x^2$ |
|---|---|---|
| 120 | – 3 | 9 |
| 110 | – 13 | 169 |
| 115 | – 8 | 64 |
| 122 | – 1 | 1 |
| 126 | 3 | 9 |
| 140 | 17 | 289 |
| 125 | 2 | 4 |
| 121 | – 2 | 4 |
| 120 | – 3 | 9 |
| 131 | 8 | 64 |
| $\Sigma X = 1230$ | | $\Sigma x^2 = 622$ |

Now $$\overline{X} = \frac{\Sigma X}{n} = \frac{1230}{10} = 123$$

∴ $$\text{S.D.} = \sqrt{\frac{\Sigma x^2}{n}} = \sqrt{\frac{622}{10}} = 7.89$$

**Example 4.** The following table shows the daily pocket expenses of 48 students of a college :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Daily pocket exp. (Rs.) : | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| No. of Students : | 3 | 6 | 9 | 13 | 8 | 5 | 4 |

Calculate standard deviation from the above data.

Sol. First, we prepare the following table :

| Daily Pocket exp. (Rs.) (X) | No. of Student ($f$) | A = 9 $dx = X - A$ | $f\,dx$ | $f\,dx^2$ |
|---|---|---|---|---|
| 6 | 3 | – 3 | – 9 | 27 |
| 7 | 6 | – 2 | – 12 | 24 |
| 8 | 9 | – 1 | – 9 | 9 |
| 9 | 13 | 0 | 0 | 0 |
| 10 | 8 | 1 | 8 | 8 |
| 11 | 5 | 2 | 10 | 20 |
| 12 | 4 | 3 | 12 | 36 |
| | N = 48 | | $\Sigma f\,dx = 0$ | $\Sigma f\,dx^2 = 124$ |

Now $\quad \text{S.D.} = \sqrt{\frac{\Sigma f dx^2}{N} - \left(\frac{\Sigma f dx}{N}\right)^2} = \sqrt{\frac{124}{48} - \left(\frac{0}{48}\right)^2} = \sqrt{2.58} = 1.61$

**Example 5.** Find the standard deviation and mean for the distribution of wages of 65 employees working in a factory :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Wages (Rs.) : | 55 | 65 | 75 | 85 | 95 | 105 | 115 |
| No. of Employees : | 8 | 10 | 16 | 14 | 10 | 5 | 2 |

**Sol.** First, we prepare the following table :

| Wages (X) | No. of Employees ($f$) | A = 85 $dx = X - A$ | $d'x$ $c = 10$ | $d'x^2$ | $f d'x$ | $f d'x^2$ |
|---|---|---|---|---|---|---|
| 55 | 8 | – 30 | – 3 | 9 | – 24 | 72 |
| 65 | 10 | – 20 | – 2 | 4 | – 20 | 40 |
| 75 | 16 | – 10 | – 1 | 1 | – 16 | 16 |
| 85 | 14 | 0 | 0 | 0 | 0 | 0 |
| 95 | 10 | 10 | 1 | 1 | 10 | 10 |
| 105 | 5 | 20 | 2 | 4 | 10 | 20 |
| 115 | 2 | 30 | 3 | 9 | 6 | 18 |
| | N = 65 | | | | $\Sigma f d'x = -34$ | $\Sigma f d'x^2 = 176$ |

Now $\quad \text{S.D.} = \sqrt{\frac{\Sigma f d'x^2}{N} - \left(\frac{\Sigma f d'x}{N}\right)^2} \times c = \sqrt{\frac{176}{65} - \left(\frac{-34}{65}\right)^2} \times 10 = \sqrt{2.71 - 0.27} \times 10$

$= \sqrt{2.44} \times 10 = 15.6$

and $\quad \overline{X} = A + \frac{\Sigma f d'x}{N} \times c = 85 - \frac{34}{65} \times 10 = 85 - 5.23 = 79.77$

**Example 6.** From the data given below find standard deviation :

| | | | | | |
|---|---|---|---|---|---|
| Age : | 5 – 7 | 8 – 10 | 11 – 13 | 14 – 16 | 17 – 19 |
| Students : | 7 | 12 | 19 | 10 | 2 |

**Sol.** First, we prepare the following table :

| Age (X) | Students ($f$) | Mid Value ($m$) | A = 12 $dx = X - A$ | $d'x$ $c = 3$ | $f d'x$ | $f d'x^2$ |
|---|---|---|---|---|---|---|
| 5 – 7 | 7 | 6 | – 6 | – 2 | – 14 | 28 |
| 8 – 10 | 12 | 9 | – 3 | – 1 | – 12 | 12 |
| 11 – 13 | 19 | 12 | 0 | 0 | 00 | 00 |
| 14 – 16 | 10 | 15 | 3 | 1 | 10 | 10 |
| 17 – 19 | 2 | 18 | 6 | 2 | 4 | 8 |
| | N = 50 | | | | $\Sigma f d'x = -12$ | $\Sigma f d'x^2 = 58$ |

Now, standard deviation $\sigma = \sqrt{\frac{\Sigma fd'x^2}{N} - \left(\frac{\Sigma fd'x}{N}\right)^2} \times c = \sqrt{\frac{58}{50} - \left(\frac{-12}{50}\right)^2} \times 3$

$$= \sqrt{1.16 - 0.0576} \times 3 = 3.15$$

**Example 7.** Find the standard deviation from the following table :

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Age under : | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
| No. of Persons Dying : | 15 | 30 | 53 | 75 | 100 | 110 | 115 | 125 |

(G.N.D.U. B.C.A. April 2006, 2008)

**Sol.** First we convert cumulative frequency distribution into simple frequency distribution as follows :

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Age : | 0 – 10 | 10– 20 | 20–30 | 30– 40 | 40–50 | 50–60 | 60–70 | 70–80 |
| No. of Persons Dying : | 15 | 15 | 23 | 22 | 25 | 10 | 5 | 10 |

Now we shall prepare the following table :

| Age (X) | No. of Persons Dying ($f$) | Mid Value ($m$) | A = 35 $dx = m - A$ | $c = 10$ $d'x = dx/c$ | $fd'x$ | $fd'x^2$ |
|---|---|---|---|---|---|---|
| 0 – 10 | 15 | 5 | – 30 | – 3 | – 45 | 135 |
| 10 – 20 | 15 | 15 | – 20 | – 2 | – 30 | 60 |
| 20 – 30 | 23 | 25 | – 10 | – 1 | – 23 | 23 |
| 30 – 40 | 22 | 35 | 0 | 0 | 0 | 0 |
| 40 – 50 | 25 | 45 | 10 | 1 | 25 | 25 |
| 50 – 60 | 10 | 55 | 20 | 2 | 20 | 40 |
| 60 – 70 | 5 | 65 | 30 | 3 | 15 | 45 |
| 70 – 80 | 10 | 75 | 40 | 4 | 40 | 160 |
| | N = 125 | | | | $\Sigma fd'x = 2$ | $\Sigma fd'x^2 = 488$ |

Now standard deviation $\sigma = \sqrt{\frac{\Sigma fd'x^2}{N} - \left(\frac{\Sigma fd'x}{N}\right)^2} \times c = \sqrt{\frac{488}{125} - \left(\frac{2}{125}\right)^2} \times 10$

$$= \sqrt{3.904 - 0.000256} \times 10 = 19.76$$

**Example 8.** The following table gives the distribution of marks obtained by 90 students in an examination

| | | | | | | |
|---|---|---|---|---|---|---|
| Marks : | 0 – 10 | 10– 20 | 20–30 | 30– 40 | 40–50 | 50–60 |
| No. of Students : | 4 | 10 | 20 | 35 | 15 | 6 |

Calculate the percentage of students lying within the range

(*i*) $\overline{X} \pm \sigma$ (*ii*) $\overline{X} \pm 2\sigma$

(G.N.D.U. B.C.A. April 2002)

**Sol.** Let us denote the marks by X and number of students by $f$.

First, we construct the following table :

| X | $f$ | Mid Value $(m)$ | $A = 25$ $dx = m - A$ | $c = 10$ $d'x = d/c$ | $d'x^2$ | $f\,d'x$ | $f\,d'x^2$ |
|---|---|---|---|---|---|---|---|
| 0 – 10 | 4 | 5 | – 20 | – 2 | 4 | – 8 | 16 |
| 10 – 20 | 10 | 15 | – 10 | – 1 | 1 | – 10 | 10 |
| 20 – 30 | 20 | 25 | 0 | 0 | 0 | 0 | 0 |
| 30 – 40 | 35 | 35 | 10 | 1 | 1 | 35 | 35 |
| 40 – 50 | 15 | 45 | 20 | 2 | 4 | 30 | 60 |
| 50 – 60 | 6 | 55 | 30 | 3 | 9 | 18 | 54 |
| | N = 90 | | | | | $\Sigma f d'x = 65$ | $\Sigma f d'x^2 = 175$ |

Now, Mean $\overline{X} = A + \dfrac{\Sigma f d'x}{N} \times c = 25 + \dfrac{65}{90} \times 10 = 32.22$

and Standard deviation $\sigma = \sqrt{\dfrac{\Sigma fd'x^2}{N} - \left(\dfrac{\Sigma fd'x}{N}\right)^2} \times c = \sqrt{\dfrac{175}{90} - \left(\dfrac{65}{90}\right)^2} \times 10 = 11.93$

(*i*) Here, we have to find number of students lying in the range $\overline{X} \pm \sigma$

*i.e.* 32.22 ± 11.93 or between 20.29 and 44.15.

For this, we make an assumption that the items are equally distributed within each class. In class 20–30 there are 20 items. So between 20–20.29, frequency would be $\dfrac{20}{10} \times 0.29$ *i.e.* 0.58. So frequency for interval 20.29–30 is $20 - 0.58 = 29.42 \approx 19$.

Further frequency for class 40–50 is 15. So between 40–44.15, frequency would be

$$\frac{15}{10} \times 4.15 = 6.22 \approx 6$$

So number of students in the range 20.29 to 44.15 = 19 + 35 + 6 = 60

∴ percentage of students lying between 20.29 to 44.15 = $\dfrac{60}{90} \times 100 = 66.67\%$

(*ii*) Now we have to find the number of students lying in the range $\overline{X} \pm 2\sigma$

*i.e.* 32.22 ± 2 (11.93) or between 8.36 and 56.08

For interval 0–10, frequency is 4. So in the interval 0 – 8.36, frequency would be $\dfrac{4}{10} \times 8.36 = 3.34$

So for the class 8.36 – 10, frequency is $4 - 3.34 = 0.66 \approx 1$

Similarly, for the interval 50 – 60, frequency is 6.

So, for class interval 50 – 56.08, frequency would be $\frac{6}{10} \times 6.08 = 3.648 \approx 4$

So number of students lying in the range 8.36 to 56.08 = 1 + 10 + 20 + 35 + 15 + 4 = 85

∴ percentage of students in the range 8.36 to 56.08 = $\frac{85}{90} \times 100 = 94.44\%$.

# EXERCISE 8.4

1. The mean of 50 items is 25 and their standard deviation is 2. Find the sum of the items and sum of the squares of all the items.

2. Calculate standard deviation and variance of the following items :

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| X | 5 | 7 | 11 | 16 | 15 | 12 | 18 | 12 |

3. Calculate standard deviation of the following grades obtained by 20 students in an examination :

   62, 85, 73, 81, 74, 58, 66, 72, 54, 84, 65, 50, 83, 62, 85, 52, 80, 86, 71, 75

4. The following data gives the annual crop in kilograms from 10 experimental farms :

   2020, 2100, 2040, 2030, 2070, 2060, 2080, 2060, 2110, 2090

   Find the mean and standard deviation of annual yield.

5. Find out mean and standard deviation of the following distribution :

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of Accidents : | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Persons involved : | 16 | 16 | 21 | 10 | 16 | 8 | 4 | 2 | 1 | 2 | 2 | 0 | 2 |

6. Find out Standard Deviation from the following :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| X : | 12 | 14 | 16 | 18 | 20 | 22 | 24 |
| $f$ : | 6 | 12 | 18 | 26 | 16 | 10 | 8 |

7. Compute the standard deviation for the following frequency distribution of wage earners in a factory :

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Wage per week ('00 Rs.) : | 9 | 12 | 15 | 18 | 21 | 24 | 27 | 30 |
| No. of wage earners : | 20 | 60 | 150 | 250 | 200 | 120 | 50 | 40 |

8. Find S.D. and variance for the following data :

| | | | | |
|---|---|---|---|---|
| Class : | 1 – 3 | 3 – 5 | 5 – 7 | 7 – 9 |
| Frequency : | 40 | 30 | 20 | 10 |

(G.N.D.U. B.Sc. C.Sc. April 2004)

9. Compute standard deviation of the marks obtained by 100 students in an examination.

| | | | | | |
|---|---|---|---|---|---|
| Marks : | 0 – 10 | 10 – 20 | 20 – 30 | 30 – 40 | 40 – 50 |
| Students : | 12 | 21 | 23 | 34 | 10 |

10. Calculate the arithmetic mean and standard deviation from the following data :

| | | | | | | |
|---|---|---|---|---|---|---|
| X : | 5 – 10 | 10 – 15 | 15 – 20 | 20 – 25 | 25 – 30 | 30 – 35 |
| $f$ : | 2 | 9 | 29 | 24 | 11 | 6 |

11. Calculate Standard deviation from the following :

| Profit (in Rs.) | No. of Firms | Profit (in Rs.) | No. of Firms |
|---|---|---|---|
| 5000 to 6000 | 10 | 0 to 1000 | 4 |
| 4000 to 5000 | 15 | – 1000 to 0 | 6 |
| 3000 to 4000 | 30 | – 2000 to –1000 | 8 |
| 2000 to 3000 | 10 | – 3000 to – 2000 | 10 |
| 1000 to 2000 | 05 | | |

12. Calculate the standard deviation from the following data :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Class : | 10 – 29 | 30– 49 | 50 – 69 | 70 – 89 | 90 – 109 | 110 –129 | 130 – 149 |
| Frequency : | 2 | 3 | 12 | 22 | 20 | 14 | 2 |

(G.N.D.U. B.Sc. I.T. April 2005)

13. Calculate mean and standard deviation from the following :

| | | | | | | |
|---|---|---|---|---|---|---|
| Age below : | 10 | 20 | 30 | 40 | 50 | 60 |
| No. of Persons : | 15 | 32 | 51 | 78 | 97 | 109 |

14. The following table gives the data regarding the length of life of 200 persons :

| | | | | | |
|---|---|---|---|---|---|
| Age : | 0 – 9 | 10 –19 | 20 – 29 | 30 – 39 | 40 – 49 |
| No. of Persons : | 6 | 15 | 33 | 39 | 45 |
| Age : | 50 – 59 | 60 – 69 | 70 – 79 | 80 – 89 | |
| No. of Persons : | 27 | 18 | 10 | 7 | |

Calculate (*i*) Mean age (*ii*) Standard deviation (*iii*) Percentage number of persons of age where length of life of a person falls within $\overline{X} \pm 2\sigma$.

## ANSWERS

| | | | |
|---|---|---|---|
| **1.** 1250, 31450 | **2.** 4.12, 16.97 | **3.** 11.45 | **4.** 2066, 28.35 |
| **5.** 3, 2.65 | **6.** 3.21 | **7.** 4.65 | **8.** 2, 4 |
| **9.** 11.92 | **10.** 20.65, 5.74 | **11.** 2530.63 | **12.** 26.53 |
| **13.** 29.954, 15.48 | **14.** (*i*) 41.85 | (*ii*) 18.56 | (*iii*) 95% |

## 8.10.2 COMBINED STANDARD DEVIATION

As compared to the other measures of dispersion, S.D. has the benefit of getting a combined standard deviation on the basis of means and standard deviations given for various sub-groups or samples separately. For example, if we have three samples with number of items as $n_1$ , $n_2$ , $n_3$ , means as $\overline{X}_1, \overline{X}_2, \overline{X}_3$ and standard deviations as $\sigma_1$ , $\sigma_2$ , $\sigma_3$ respectively, then combined S.D. is calculated by formula :

$$\sigma_{123} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_3\sigma_3^2 + n_1 d_1^2 + n_2 d_2^2 + n_3 d_3^2}{n_1 + n_2 + n_3}} \qquad ...(8.44)$$

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

and $$\overline{X}_{123} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + n_3\overline{X}_3}{n_1 + n_2 + n_3}$$

# ILLUSTRATIVE EXAMPLES

**Example 1.** The mean weight of 150 students is 60 kg. The mean weight of boys is 70 kg. with a standard deviation of 10 kg. For the girls, the mean weight of 55 kg. and the standard deviation is 15 kg. Find the number of boys and the combined standard deviation.

(P.U. B.C.A. Sept. 2005; G.N.D.U. B.C.A. Sept. 2007)

**Sol.** Let $\overline{X}_1$ and $\overline{X}_2$ be the mean weight of boys and girls respectively, $\sigma_1$ and $\sigma_2$ be the standard deviation in the weight of boys and girls respectively, $\overline{X}_{12}$ be the combined mean of all the students, $\sigma_{12}$ be the combined standard deviation of all the students and $n_1$ and $n_2$ be the number of boys and girls respectively.

$\therefore \quad \overline{X}_1 = 70,\ \overline{X}_2 = 55,\ \sigma_1 = 10,\ \sigma_2 = 15$ and $\overline{X}_{12} = 60$

Also $\quad n_1 + n_2 = 150 \qquad ...(i)$

Now, $$\overline{X}_{12} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}$$

$$\Rightarrow \quad 60 = \frac{70\,n_1 + 55\,n_2}{150} \quad \Rightarrow \quad 70\,n_1 + 55\,n_2 = 9000$$

$$\Rightarrow \quad 14\,n_1 + 11\,n_2 = 1800 \qquad ...(ii)$$

Solving (i) and (ii) for $n_1$ and $n_2$, we get

$$\therefore \quad n_1 = 50 \text{ and } n_2 = 100$$

and
$$d_1 = \overline{X}_1 - \overline{X}_{12} = 70 - 60 = 10$$
$$d_2 = \overline{X}_2 - \overline{X}_{12} = 55 - 60 = -5$$

Further combined standard deviation

$$\sigma_{12} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$$

$$= \sqrt{\frac{50\,(10)^2 + 100\,(15)^2 + 50\,(10)^2 + 100\,(-5)^2}{50 + 100}}$$

$$= \sqrt{\frac{5000 + 22500 + 5000 + 2500}{150}} = 15.28$$

Hence, number of boys is 50 and combined standard deviation = 15.28.

**Example 2.** The first of two samples has 100 items with mean 15 and standard deviation 3. If the while group has 250 items with mean 15.6 and standard deviation of $\sqrt{13.44}$ find the standard deviation of the second group. (P.U. B.C.A. April 2004)

**Sol.** Given $\quad n_1 = 100,\ \overline{X}_1 = 15,\ \sigma_1 = 3$

$$n_1 + n_2 = 250,\ \overline{X}_2 = 15.6,\ \sigma_{12} = \sqrt{13.44}$$

Since $\quad \overline{X}_{12} = \dfrac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}$

$\Rightarrow \quad 15.6 = \dfrac{100 \times 15 + 150\,\overline{X}_2}{250} \qquad [\because n_2 = 250 - n_1 = 250 - 100 = 150]$

$\Rightarrow \quad 3900 = 1500 + 150\,\overline{X}_2$

$\Rightarrow \quad \overline{X}_2 = 16$

Now $\quad d_1 = \overline{X}_1 - \overline{X}_{12} = 15 - 15.6 = 0.6$

and $\quad d_2 = \overline{X}_2 - \overline{X}_{12} = 16 - 15.6 = 0.4$

Since $\quad \sigma_{12} = \sqrt{13.44}$

$\therefore \quad \sigma_{12}^{\,2} = 13.44$

$\Rightarrow \quad \dfrac{n_1\sigma_1^{\,2} + n_2\sigma_2^{\,2} + n_1 d_1^{\,2} + n_2 d_2^{\,2}}{n_1 + n_2} = 13.44$

$\Rightarrow \quad \dfrac{100\,(3)^2 + 150\,\sigma_2^{\,2} + 100\,(0.6)^2 + 150\,(0.4)^2}{250} = 13.44$

$\Rightarrow \quad 900 + 150\,\sigma_2^{\,2} + 36 + 24 = 3360$

$\Rightarrow \quad \sigma_2^{\,2} = 16 \Rightarrow \sigma_2 = 4$

Hence, standard deviation of second group = 4.

Example 3. Find the missing information from the following data :

| | | Group I | Group II | Group III | Combined |
|---|---|---|---|---|---|
| Number of items | : | 50 | ? | 90 | 200 |
| Standard Deviation | : | 6 | 7 | ? | 7.746 |
| Arithmetic Mean | : | 113 | ? | 115 | 116 |

Sol. Given $n_1 = 50, \quad n_2 = ?, \quad n_3 = 90 \qquad \text{Total} = 200$

$\overline{X}_1 = 113, \quad \overline{X}_2 = ?, \quad \overline{X}_3 = 115 \qquad \overline{X}_{123} = 116$

$\sigma_1 = 6, \quad \sigma_2 = 7, \quad \sigma_3 = ? \qquad \sigma_{123} = 7.746$

Now $\quad n_1 + n_2 + n_3 = 200$

$\therefore \quad 50 + n_2 + 90 = 200 \quad \Rightarrow \quad n_2 = 60$

Again $$\overline{X}_{123} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + n_3\overline{X}_3}{n_1 + n_2 + n_3}$$

$$\Rightarrow \quad 116 = \frac{50\times113 + 60\times\overline{X}_2 + 90\times115}{200}$$

$$\Rightarrow \quad 116 = \frac{565 + 6\overline{X}_2 + 1035}{20}$$

$$\Rightarrow \quad 2320 = 1600 + 6\,\overline{X}_2 \quad \Rightarrow \quad \overline{X}_2 = 120$$

$$d_1 = \overline{X}_{123} - \overline{X}_1 = 116 - 113 = 3$$

$$d_2 = \overline{X}_{123} - \overline{X}_2 = 116 - 120 = -4$$

$$d_3 = \overline{X}_{123} - \overline{X}_3 = 116 - 115 = 1$$

$$\therefore \text{ Combined S.D. } \sigma_{123} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_3\sigma_3^2 + n_1 d_1^2 + n_2 d_2^2 + n_3 d_3^2}{n_1 + n_2 + n_3}}$$

Squaring both sides

$$(7.746)^2 = \frac{50\times36 + 60\times49 + 90\sigma_3^2 + 50\times9 + 60\times16 + 90\times1}{200}$$

$$\Rightarrow \quad 60 = \frac{180 + 294 + 9\sigma_3^2 + 45 + 96 + 9}{20} \quad \Rightarrow \quad 60 = \frac{9\sigma_3^2 + 624}{20}$$

$$\Rightarrow \quad 1200 = 9\sigma_3^2 + 624 \quad \Rightarrow \quad 9\sigma_3^2 = 576 \quad \Rightarrow \quad \sigma_3^2 = 64$$

$$\therefore \quad \sigma_3 = 8$$

# EXERCISE 8.5

1. If $\overline{X}_1 = 10, \sigma_1^2 = 4, n_1 = 10$ and $\overline{X}_2 = 12, \sigma_2^2 = 9, n_2 = 20$

   Find combined standard deviation.

2. The first of the two samples has 50 items with mean 10 and standard deviation 2. If the whole group has 150 items with mean 12.5 and variance 8.11, find the standard deviation of the second group. (P.U. B.C.A. Sept. 2003)

3. For two groups of observations, the following data is available :

   Group I : $\Sigma (X-5) = 3, \quad \Sigma (X-5)^2 = 43, \quad n = 18$

   Group II : $\Sigma (X-8) = -11, \quad \Sigma (X-8)^2 = 76, \quad n = 17$

   Find the mean and standard deviation of 35 observations by combining two groups.

4. Find the combined S.D. of three distributions in parts given below :

| Parts | No. of Items | Mean | S.D. |
|---|---|---|---|
| 1 | 200 | 25 | 3 |
| 2 | 250 | 10 | 4 |
| 3 | 300 | 15 | 5 |

5. From the data given below, giving the arithmetic means and standard deviations of four sub-groups, calculate the average and standard deviation of the whole group.

| Sub-group | No. of items | Arithmetic Mean | Standard Deviation |
|---|---|---|---|
| A | 50 | 61.0 | 8 |
| B | 100 | 70.0 | 9 |
| C | 120 | 80.5 | 10 |
| D | 30 | 83.0 | 11 |

## ANSWERS

1. 2.87 2. 2.34 3. 6.23, 2.09 4. 7.19 5. 74, 12.18

### 8.10.3. CORRECTING INCORRECT STANDARD DEVIATION

The following examples illustrate the procedure of correcting incorrect standard deviation.

## ILLUSTRATIVE EXAMPLES

**Example 1.** For a group of 200 candidates, the mean and S.D. of scores were found to be 40 and 15 respectively. Later on it was discovered that the scores 43 and 35 were misread as 34 and 53 respectively. Find the corrected figures. (P.U. B.C.A. Sept. 2004)

... Given $n = 200$, incorrect $\overline{X} = 40$, incorrect $\sigma = 15$

Since $\overline{X} = \frac{\Sigma X}{n}$ so $\Sigma X = n\overline{X}$

$\therefore$ incorrect $\Sigma X = 200 \times 40 = 8000$

So correct $\Sigma X = 8000 - 34 - 53 + 43 + 35 = 7991$

$\therefore$ Correct $\overline{X} = \frac{7991}{200} = 39.955$

Now $\sigma = 15 \Rightarrow \sigma^2 = 225$

$$\Rightarrow \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2 = 225 \quad \Rightarrow \quad \frac{\Sigma X^2}{200} - \left(\frac{8000}{200}\right)^2 = 225$$

$$\Rightarrow \frac{\Sigma X^2}{200} = 225 + 1600$$

$\therefore$ incorrect $\Sigma X^2 = 365000$

So correct $\Sigma X^2 = 365000 - (34)^2 - (53)^2 + (43)^2 + (35)^2$

$= 365000 - 1156 - 2809 + 1849 + 1225 = 364109$

$$\therefore \text{ correct S.D.} = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2} = \sqrt{\frac{364109}{200} - \left(\frac{7991}{200}\right)^2} = 14.918$$

...ample 2. The mean and standard deviation of 100 items was 20 and 3 respectively. Later on it was found ...at observations 21, 21 and 18 were wrongly taken. What would be the mean and S.D. if wrong items are ...eleted ?

...l. Given $n = 100$, Incorrect $\overline{X} = 20$ and Incorrect S.D. = 3.

Now $\overline{X} = \frac{\Sigma X}{n} \quad \Rightarrow \quad 20 = \frac{\Sigma X}{100}$

$\therefore$ Incorrect $\Sigma X = 2000$

Corrected $\Sigma X = 2000 - 21 - 21 - 18 = 1940$

$$(\text{S.D.})^2 = \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2$$

$$\Rightarrow 9 = \frac{\Sigma X^2}{100} - \left(\frac{2000}{100}\right)^2 \quad \Rightarrow \quad 9 = \frac{\Sigma X^2}{100} - 400$$

$\therefore$ Incorrect $\Sigma X^2 = 40900$

Correct $\Sigma X^2 = 40900 - (21)^2 - (21)^2 - (18)^2 = 40900 - 1206 = 39694$

Now revised value of $n = 97$

$\therefore$ Corrected $\overline{X} = \frac{\Sigma X}{n} = \frac{1940}{97} = 20$

$$\text{Corrected S.D.} = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2} = \sqrt{\frac{39694}{97} - \left(\frac{1940}{97}\right)^2} = \sqrt{409.22 - 400} = \sqrt{9.22} = 3.04$$

# EXERCISE 8.6

1. Mean and standard Deviation of 25 items are found to be 15 and 2. But later on it was discovered that one item was taken as 20 instead of 30. Find the correct mean and correct standard deviation.
2. A student obtained the mean and standard deviation of 100 observations as 40 and 15.1 respectively. It was later discovered that he had wrongly copied down an observation as 50 instead of 40. Calculate correct mean and standard deviation.
3. The number of employees, wages per employee and variance of the wages are given below :

   Number of employees = 100

   Average wage per employee = 85

   Variance = 16

   Suppose the wages of an employee were wrongly noted as Rs. 120 instead of Rs. 100. What is the correct value of the variance ?
4. A study of the age of 100 film stars grouped in intervals of 10 – 12, 12 – 14, ...etc. revealed the mean and standard deviation to be 32.18 and 13.18 respectively. While checking it was discovered that the age 57 was misread as 27. Calculate correct mean age and standard deviation.
5. The arithmetic mean and variance of a set of 10 figures are known to be 17 and 33 respectively. Of the 10 figures one figure 26 was subsequently found inaccurate and was weeded out. What is the new mean and standard deviation ?

# ANSWERS

1. 15.4, 3.44 2. 39.9, 15.07 3. 5.96 4. 32.48, 13.40 5. 16, 5.16

## 8.10.4 COEFFICIENT OF STANDARD DEVIATION AND COEFFICIENT OF VARIATION

*Standard deviation* is an absolute measure of dispersion *i.e.* it expresses the measure of dispersion in terms of units of measurement of the variable. By using standard deviation we can not compare the variability of two or more series. So we have to use the relative measure of dispersion which is known as *coefficient of standard deviation* and is defined as

$$\text{coefficient of standard deviation} = \frac{\sigma}{\overline{X}}. \qquad ...(8.45)$$

The coefficient of standard deviation is a pure number *i.e.* independent of units of measurement and hence can be used to compare the variability of two or more series.

Prof. Karl Pearson suggested another relative measure namely *coefficient of variation.* According to him, "Coefficient of variation is the percentage variation in mean, standard deviation being considered as total variation in the mean".

He defined coefficient of variation (C.V.) as

$$\text{C.V.} = \frac{\sigma}{\overline{X}} \times 100 \qquad ...(8.46)$$

Cofficient of VARIATION

For comparing the variability of two or more series, we calculate the coefficient of variation for each series. The series with smaller coefficient of variation will be more consistent *i.e.* will have less variability than the other series and the series with larger coefficient of variation will be less consistent *i.e.* will have more variability than the other series.

## CHECKPOINTS

1. What is meant by relative dispersion ? Define coefficient of variation and explain its uses. (G.N.D.U. B.Sc. C.Sc. April 2006)
2. What is the relationship between standard deviation and C.V. ? (G.N.D.U. B.C.A. Sept. 2006)
3. Explain the significance of variance and coefficients of variation. (G.N.D.U. B.Sc. C.Sc. Sept. 2007)

# ILLUSTRATIVE EXAMPLES

**Example 1.** Calculate the C.V. for the following data :

| | | | | | | |
|---|---|---|---|---|---|---|
| X : | 5 | 15 | 25 | 35 | 45 | 55 |
| $f$ : | 12 | 18 | 27 | 20 | 17 | 6 |

(G.N.D.U. B.C.A. April 2005)

**Sol.** First we construct the following table :

| X | $f$ | A = 25<br>$dx = X-A$ | $c = 10$<br>$d'x = dx/c$ | $f\,d'x$ | $f\,d'x^2$ |
|---|---|---|---|---|---|
| 5 | 12 | –20 | –2 | –24 | 48 |
| 15 | 18 | –10 | –1 | –18 | 18 |
| 25 | 27 | 0 | 0 | 0 | 0 |
| 35 | 20 | 10 | 1 | 20 | 20 |
| 45 | 17 | 20 | 2 | 34 | 68 |
| 55 | 6 | 30 | 3 | 18 | 54 |
| | N = 100 | | | $\Sigma f\,d'x = 30$ | $\Sigma f\,d'x^2 = 208$ |

Now, Mean $\overline{X} = A + \frac{\Sigma f\,d'x}{N} \times c = 25 + \frac{30}{100} \times 10 = 28$

and S.D. $\sigma = \sqrt{\frac{\Sigma fd'x^2}{N} - \left(\frac{\Sigma fd'x}{N}\right)^2} \times c$

$$= \sqrt{\frac{208}{100} - \left(\frac{30}{100}\right)^2} \times 10$$

$$= \sqrt{2.08 - 0.09} \times 10$$

$$= 14.107$$

Hence C.V. $= \frac{\sigma}{\overline{X}} \times 100 = \frac{14.107}{28} \times 100 = 50.38$

**Example 2.** Consider the following two data sets :

| | | | | | |
|---|---|---|---|---|---|
| I : | 12 | 25 | 37 | 8 | 41 |
| II : | 19 | 32 | 44 | 15 | 48 |

Calculate standard deviation for each of these two data sets. Comment on the relationship between these two standard deviations. (G.N.D.U. B.C.A. April 2007)

**Sol.** Let us denote data set I by X and data set II by Y.

Now we prepare the following table :

| X | Y | $X^2$ | $Y^2$ |
|---|---|---|---|
| 12 | 19 | 144 | 361 |
| 25 | 32 | 625 | 1024 |
| 37 | 44 | 1369 | 1936 |
| 8 | 15 | 64 | 225 |
| 41 | 48 | 1681 | 2304 |
| $\Sigma X = 123$ | $\Sigma Y = 158$ | $\Sigma X^2 = 3883$ | $\Sigma Y^2 = 5850$ |

Now $\sigma_X = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2} = \sqrt{\frac{3883}{5} - \left(\frac{123}{5}\right)^2} = \sqrt{\frac{4286}{25}} = 13.09$

and $\sigma_Y = \sqrt{\frac{\Sigma Y^2}{n} - \left(\frac{\Sigma Y}{n}\right)^2} = \sqrt{\frac{5850}{5} - \left(\frac{158}{5}\right)^2} = \sqrt{\frac{4286}{25}} = 13.09$

Clearly $\sigma_X = \sigma_Y$ *i.e.* standard deviations of these two sets of data are equal.

Although the standard deviation of two sets of data are equal yet we can not say that the variation of two sets of data is same. This is because of the fact that standard deviation can not be used to compare the variability in the two sets of data.

To compare the variability of two series, we calculate coefficient of variance for the two series.

For series I, $\overline{X} = \frac{\Sigma X}{n} = \frac{123}{5} = 24.6$

For series II, $\overline{Y} = \frac{\Sigma Y}{n} = \frac{158}{5} = 31.6$

$$\text{C.V. for series I} = \frac{\sigma_X}{\overline{X}} \times 100 = \frac{13.09}{24.6} \times 100 = 53.21$$

and $$\text{C.V. for series II} = \frac{\sigma_Y}{\overline{Y}} \times 100 = \frac{13.09}{31.6} \times 100 = 41.42$$

Clearly, C.V. for series I is more than that of series II.

So there is more variability in series I than in series II.

**Example 3.** A factory produces two types of electric bulbs P and Q. In an experiment relating to their life, the following results were obtained.

| Length of life (hrs): | 500 – 700 | 700 – 900 | 900 – 1100 | 1100 – 1300 | 1300 – 1500 |
|---|---|---|---|---|---|
| No. of bulbs of P | 5 | 11 | 26 | 10 | 8 |
| No. of bulbs of Q | 4 | 30 | 12 | 8 | 6. |

Compare the variabilities of life of two varieties using coefficient of variation.

Sol. First, we prepare the following table :

| | | | | Bulb P | | | Bulb Q | | |
|---|---|---|---|---|---|---|---|---|---|
| Length of Life (X) | M.V. ($m$) | A = 1000 $dx = m - 1000$ | $d'x$ $c = 200$ | Freq. ($f$) | $fd'x$ | $fd'x^2$ | Freq. ($f$) | $fd'x$ | $fd'x^2$ |
| 500 – 700 | 600 | – 400 | – 2 | 5 | – 10 | 20 | 4 | – 8 | 16 |
| 700 – 900 | 800 | – 200 | – 1 | 11 | – 11 | 11 | 30 | – 30 | 30 |
| 900 – 1100 | 1000 | 000 | 0 | 26 | 00 | 00 | 12 | 00 | 00 |
| 1100 – 1300 | 1200 | 200 | 1 | 10 | 10 | 10 | 8 | 8 | 8 |
| 1300 – 1500 | 1400 | 400 | 2 | 8 | 16 | 32 | 6 | 12 | 24 |
| | | | | N = 60 | $\Sigma fd'x = 5$ | $\Sigma fd'x^2 = 73$ | N = 60 | $\Sigma fd'x = -18$ | $\Sigma fd'x^2 = 78$ |

$$\overline{X}(P) = A + \frac{\Sigma fd'x}{N} \times c = 1000 + \frac{5}{60} \times 200 = 1016.67$$

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

$$= \sqrt{\frac{73}{60} - \left(\frac{5}{60}\right)^2} \times 200 = \sqrt{1.22 - 0.007} \times 200 = 220.27$$

$$\text{C.V. (P)} = \frac{\sigma(P)}{\overline{X}(P)} \times 100 = \frac{220.27}{1016.67} \times 100 = 21.67\%$$

Also $$\overline{X}(Q) = A + \frac{\Sigma fd'x}{N} \times c = 1000 - \frac{18}{60} \times 200 = 940$$

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

$$= \sqrt{\frac{78}{60} - \left(\frac{-18}{60}\right)^2} \times 200 = \sqrt{1.3 - 0.09} \times 200 = 220$$

$$\text{C.V. (Q)} = \frac{\sigma(Q)}{\overline{X}(Q)} \times 100 = \frac{220}{940} \times 100 = 23.4\%$$

Comparing the two results, bulb of type Q has more variability of life.

# EXERCISE 8.7

1. Find the S.D. and C.V. for the following individual series :

   4, 4, 4, 4, 4, 4, 4

   (G.N.D.U. B.C.A. April 2004)

2. Following frequency distribution shows the monthly free market rates of US Dollar in Rupees per US Dollar.

| | | | | | | |
|---|---|---|---|---|---|---|
| Rs. per US Dollar : | 8.51–9.00 | 9.01–9.50 | 9.51–10.00 | 10.01–10.50 | 10.51–11.00 | 11.01–11.50 |
| Frequency : | 10 | 3 | 8 | 25 | 12 | 9 |
| Rs. per US Dollar : | 11.51–12.00 | 12.01–12.50 | 12.51–13.00 | | | |
| Frequency : | 9 | 9 | 2 | | | |

   (*a*) Calculate standard deviation of dollar series

   (*b*) Calculate coefficient of variation of the dollar series. (P.U. B.C.A. Sept. 2007)

3. The following are the scores made by two batsmen A and B in a series of innings

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A : | 12 | 115 | 6 | 73 | 7 | 19 | 119 | 36 | 84 | 29 |
| B : | 47 | 12 | 76 | 42 | 4 | 5 | 37 | 48 | 13 | 0 |

   Who is the better scorer ? Who is more consistent ?

4. Calculate the standard deviation of the following two series. Which shows greater variation ?

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Series A : | 192 | 288 | 236 | 229 | 184 | 260 | 348 | 291 | 330 | 243 |
| Series B : | 83 | 87 | 93 | 109 | 124 | 126 | 126 | 101 | 102 | 108 |

5. In two factories A and B engaged in the same industry in the area, the average weakly wages (in Rs.) and standard deviations are as follow :

| Factory | Average | Standard Deviations | No. of Employees |
|---|---|---|---|
| A | 35.5 | 5.0 | 476 |
| B | 28.5 | 4.5 | 524 |

(*i*) Which factory, A or B pays out larger amount as weakly wages ?

(*ii*) Which factory A or B has greater variability in individual wages ?

6. The following table gives the mean and variance of the heights of boys and girls studying in a college. Find

(*i*) the standard deviation of the height of boys and girls taken together and

(*ii*) which groups has more variability in heights ?

| | Boys | Girls |
|---|---|---|
| Number | 400 | 100 |
| Average Height | 68 inches | 65 inches |
| Variance | 9 (inches)$^2$ | 4 (inches)$^2$ |

(G.N.D.U. B.C.A. April 2003, 2006)

7. Particulars regarding the income of the localities A and B are given below :

| | A | B |
|---|---|---|
| No. of Persons | 600 | 500 |
| Average income (Rs.) | 175 | 186 |
| Variance (Rs.) | 100 | 81 |

(*i*) In which locality is the variation of income greater ?

(*ii*) What is the total income of both the localities taken together ?

(*iii*) What is the average income of the two localities taken together ?

(*iv*) What is the combined standard deviation ?

(P.U. B.C.A. April 2002)

8. Coefficient of variation of wages of skilled and unskilled workers were 55% and 70% respectively. If the standard deviations were 20.35 and 14 respectively. Calculate the combined average wage and combined S.D. of all workers if 75% are skilled workers.

# ANSWERS

1. 0, 0 2. 1.06, 10
3. Batsman A is better scorer and more consistent than Batsman B
4. S.D. (A) = 51.59, S.D. (B) = 14.96, Series A shows greater variation.
5. (*i*) Factory A (*ii*) Factory B 6. (*i*) 3.07 (*ii*) Boys
7. (*i*) Locality A (*ii*) Rs. 198000 (*iii*) Rs. 180 (*iv*) 11.02
8. 32.42, 20.56

## 8.11 COMPARISON OF MEAN DEVIATION AND STANDARD DEVIATION

1. Both measures of dispersion namely mean deviation and standard deviation are based on all the observations of data.
2. Mean deviation can be calculated from either arithmetic mean or median or mode. Standard deviation is always calculated by taking deviations from arithmetic mean only.
3. Signs of deviations are ignored *i.e.* absolute value of deviations is taken while computing mean deviation. On the other hand, standard deviation does not ignore the signs of deviations.
4. Standard deviation is capable of further algebraic treatment whereas mean deviation is not.
5. Mean deviation has no further application in advanced statistical analysis, whereas standard deviation is used in statistical tools like skewness, correlation, regression, analysis of variance, testing of hypothesis etc.

## 8.12 GRAPHICAL METHOD OF DISPERSION (LORENZ CURVE)

**Dr. Max. O. Lorenz**, a famous economist and statistician first used this curve for studying the distribution of wealth and income. It is a graphic method of studying dispersion. The curve is most important in comparing the distribution of income, wealth, profits etc. It helps in studying comparative value of variability in different components of the distribution particularly economic aspects of the problem under study. These are all compared with the absolutely equal distribution of income, wealth, profits etc. as the case may be , which is represented by a line with zero variability having slope $\pm$ 1 *i.e.*, making an angle of 45° with the horizontal line. The farther the curve from this line of equality the greater the extent of variability in the distribution. Following are the steps for the construction of Lorenz curve :

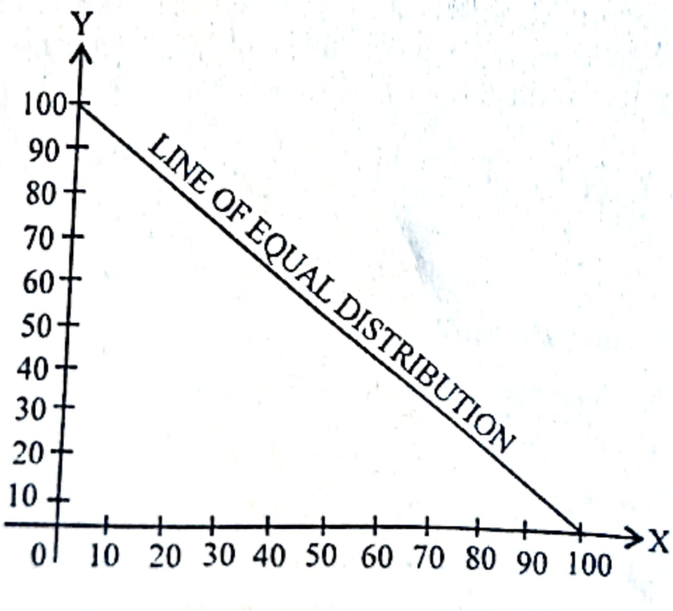
(*i*) Take the cumulative frequencies of both the number of items and the values of items.

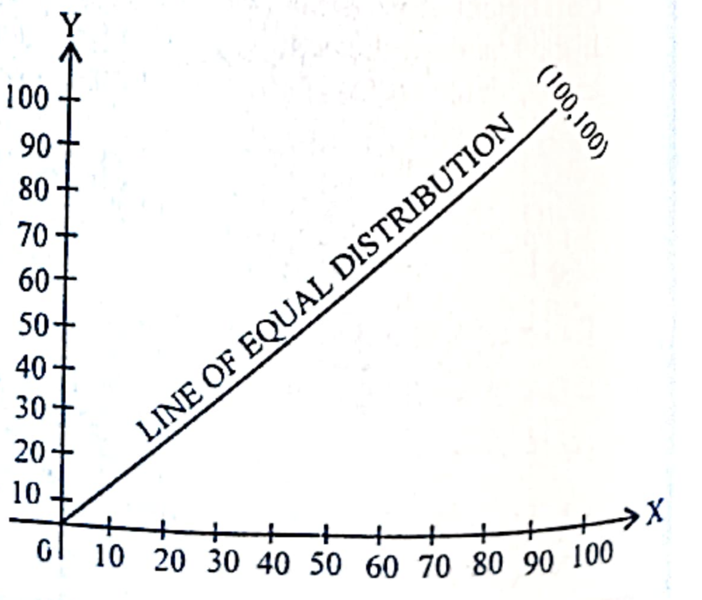
(*ii*) Convert these cumulative frequencies as percentage of the totals. As such the sum of these percentage values will be 100 each.

(*iii*) Mark the values from 0 to 100 of percentage of variables *i.e.* values of the frequencies along $x$-axis.

(*iv*) Mark the values 0 to 100 of percentage of the number of items according to the *same scale* along $y$ – axis.

(*v*) Now join the points with co-ordinates (0, 0) and (100, 100) or (0, 100) with (100, 0). This line will have slope 1 (or – 1) *i.e.*, it will make 45° (or 135) with $x$ axis and will be known as the '*line of equal distribution*' or '*line of zero dispersion*'.



Now, we shall illustrate this method with the help of following example.

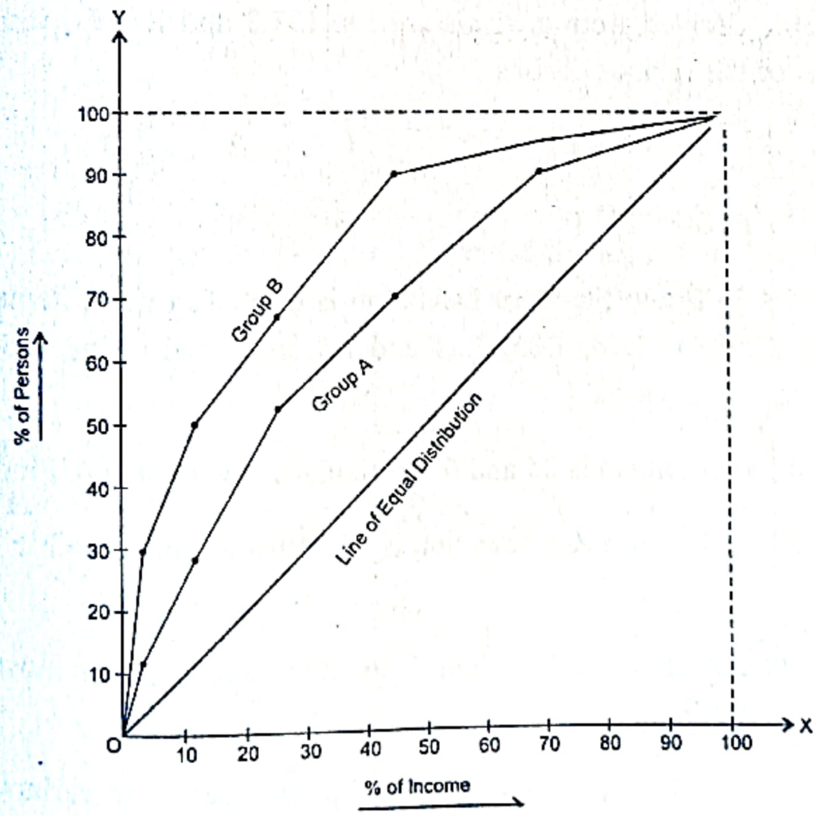
Consider the following data relating to annual income of two group of persons :

| Annual income (Rs.) | No of Persons Group A | No. of Persons Group B |
|---|---|---|
| 0– 1000 | 6000 | 15000 |
| 1000 – 2000 | 8000 | 10000 |
| 2000 – 3000 | 12000 | 9000 |
| 3000 – 4000 | 9000 | 11000 |
| 4000 – 5000 | 10000 | 3000 |
| 5000 – 6000 | 5000 | 2000 |

First, we prepare the following table :

| Class Interval | Mid value | Cum. Totals | Cum.% | Group A | | | Group B | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | No. of Persons | Cum. Total | Cum. % | No. of Persons | Cum. Total | Cum. % |
| 0 – 1000 | 500 | 500 | 2.8 | 6000 | 6000 | 12 | 15000 | 15000 | 30 |
| 1000 – 2000 | 1500 | 2000 | 11.1 | 8000 | 14000 | 28 | 10000 | 25000 | 50 |
| 2000 – 3000 | 2500 | 4500 | 25 | 12000 | 26000 | 52 | 9000 | 34000 | 68 |
| 3000 – 4000 | 3500 | 8000 | 44.4 | 9000 | 35000 | 70 | 11000 | 45000 | 90 |
| 4000 – 5000 | 4500 | 12500 | 69.4 | 10000 | 45000 | 90 | 3000 | 48000 | 96 |
| 5000 – 6000 | 5500 | 18000 | 100 | 5000 | 50000 | 100 | 2000 | 50000 | 100 |

Now we sketch the following figure :




Since curve of group B is away from the line of zero dispersion as compared to the curve of group A, therefore group A is more consistent.

# MISCELLANEOUS EXERCISE

1. Given, largest value = 80, Coefficient of Range = 0.6 Find the smallest value in the data.

2. From the following data calculate the percentage of workers getting wages

(*i*) more than Rs. 44 (*ii*) between Rs. 22 and 58 (*iii*) the quartile deviation.

| Wages (Rs.) : | 0 – 10 | 10– 20 | 20– 30 | 30– 40 | 40– 50 | 50– 60 | 60– 70 | 70– 80 |
|---|---|---|---|---|---|---|---|---|
| No. of Workers : | 20 | 45 | 85 | 160 | 70 | 55 | 35 | 30 |

3. If $\delta_x$ and $\delta_y$ are standard deviations of two uncorrelated statistical variables $x$ and $y$ prove that the standard deviation of $ax + by$ is square root of $a^2\delta_x{}^2 + b^2\delta_y{}^2$.

(G.N.D.U. B.Sc. C.Sc. April 2003)

4. Calculate standard deviation from the following data :

| Mid Points : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Frequency : | 2 | 60 | 101 | 152 | 205 | 155 | 79 | 40 | 1 |

5. The values of the arithmetic mean and standard deviation of the following frequency distribution of a continuous variable derived from analysis are Rs. 135.3 and Rs. 9.6 respectively. Find the upper and lower limits of the various classes.

| $d'x$ : | – 4 | – 3 | – 2 | – 1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|
| $f$ : | 2 | 5 | 8 | 18 | 22 | 13 | 8 | 4 |

6. The mean of 50 items is 7·43 and Standard Deviation is 0.28. Ten more items whose values are 6.8, 7.81, 7.58, 7.7, 8.05, 6.98, 7.78, 7.85, 7.21 and 7.4 are added to the data. Find the revised value of mean and standard deviation.

7. The arithmetic mean of two numbers is 25 and their standard deviation is 4. Find the numbers.

8. In a series the most likely value of mean deviation is 12. What are the possible values of Q.D. and S.D. ?

9. For a group, quartile deviation is 20 and median is 36. What would be the most probable value of coefficient of mean deviation ?

10. The variance of a distribution is 6. What will be the new value of variance if each item is (*i*) multiplied by 2, (*ii*) divided by 2, (*iii*) increased by 2, (*iv*) decreased by 2 ?

11. From the following figures determine the percentage of cases which lie outside the Range $\overline{X} \pm \sigma$, $\overline{X} \pm 2\sigma$ and $\overline{X} \pm 3\sigma$.

115, 117, 121, 125, 116, 120, 118, 117, 119, 116, 122, 124, 123, 118, 120, 118, 126, 127, 122, 123

12. What do you understand by dispersion ? Explain the relative merits and demerits of various measures of dispersion.

(Pbi. U, April 2005)

# ANSWERS

1. 20
2. (*i*) 32.4% (*ii*) 68.4% (*iii*) 11.12
4. 1.57
5. 109.5 – 115.5, 115.5 – 121.5, ......, 151.5 – 157.5
6. 7.44, 0.302
7. 21, 29
8. 10, 15
9. 2/3
10. (*i*) 24 (*ii*) 1.5 (*iii*) 6
11. 35%, 0%, 0%