2. MEASURES OF DISPERSION

Let's Study

- Meaning and Definition of Dispersion
- Absolute and Relative Measures of Dispersion
- Range of data.
- Semi Inter Quartile Range / Q.D.
- Variance and Standard Deviation
- Coefficient of Variation

Let's Recall

- ✓ Concept of Constant and Variable
- ✓ Concept of an Average
- ✓ Computation of Mean for Ungrouped and Grouped Data
- ✓ Quartiles

"An average does not tell the full story. It is hardly fully representative of a mass unless we know the manner in which the individual items scatter around it. A further description of the series is necessary if we are to gauge how representative the average is."

- George Simpson and Fritz Kafka

Let's Observe...

In the earlier classes we have learnt about the measures of central tendency namely mean, median and mode. Such an average tells us only about the central part of the data. But it does not give any information about the spread of the data. For example, consider the runs scored by 3 batsmen in a series of 5 One Day International matches.

Batsman	Runs scored	Total	Mean
X	90, 17, 104, 33, 6	250	50
Y	40, 60, 55, 50, 45	250	50
Z	112, 8, 96, 29, 5	250	50

Table 2.1

All the above series have the same size (n=5) and the same mean (50), but they are different in composition. Thus, to decide who is more consistent, the measure of central tendency are not sufficient. We need some other measure. One such measure is that of Dispersion.

In the above example, observations from series X and series Z are more scattered as compared to those in series Y. The scatter in observations is called dispersion. The amount by which the observations deviate from the average is called dispersion.

Let's Constuct...

Given two different series-

A : 0.5, 1, 1.5, 3, 4, 8

B : 2, 2.2, 2.6, 3.4, 3.8,

Find arithmetic means of the two series.

Plot the two series on the number line.

Observe the scatter of the data in each series and decide which series is more scattered.

Let's Learn

2.1 Definition :

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

- Spiegel

Measures of Dispersion :

Following measures of dispersion are the commonly used –

(i) Range

(ii) Semi Inter Quartile Range / Quartile Deviation

(iii) Variance

(iv) Standard deviation

2.2 Range :

Range is the simplest measure of dispersion. It is defined as the difference between the largest value and the smallest value in the data.

Thus,

Range = Largest Value – Smallest Value
= L – S

Where, L = Largest Value and S = Smallest Value.

SOLVED EXAMPLES

1. Following data gives weights of 10 students (in kgs) in a certain school. Find the range of the data.

70, 62, 38, 55, 43, 73, 36, 58, 65, 47

Solutions : Smallest Value = S = 36

Largest Value = L = 73

Range = L – S = 73 – 36 = 37

2. Calculate range for the following data.

Salary (00's ₹)	30-50	50-70	70-90	90-110	110-130	130-150
No. of Employees	7	15	30	24	18	11

Solution :

L = Upper limit of highest class = 150

S = Lower limit of lowest class = 30

∴ Range = L – S = 150 – 30 = 120

Let's Learn

2.3 Quartile Deviation (Semi - Inter Quartile Range) :

We observe the largest and smallest values in the data to find the range. If the data contains extreme values or outliers, then the range is an incorrect measure of variation. In such cases, we use a measure of dispersion based on quartiles to get rid of the effect of the extreme values. This measure is called quartile deviation or Semi-Inter Quartile Range (Semi-IQR) defined as follows.

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

$$= \frac{\textit{Interquartile Range}}{2}$$

Where $Q_3 - Q_1$, is called Interquartile Range.

Semi inter quartile range = $\frac{Q_3 - Q_1}{2}$

SOLVED EXAMPLES

Ex.1. Find Semi Inter Quartile Range for the following data.

Marks of 11 students in Mathematics : 67, 75, 84, 60, 72, 58, 61, 52, 79, 91, 56

Solution : Let us first arrange the given data in ascending order,

52, 56, 58, 60, 61, 67, 72, 75, 79, 84, 91

Here, n=11

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{ observation}$$

$$= \left(\frac{11+1}{4}\right)^{th} \text{ observation}$$

$$= 3^{rd} \text{ observation} = 58$$

$$Q_3 = \left[3\left(\frac{n+1}{4}\right)^{th}\right] \text{ observation}$$

$= \boxed{}$ observation

$= 9^{th}$ observation = 79

$\therefore$ Semi Interquartile range $= \dfrac{Q_3 - Q_1}{2} = \dfrac{79-58}{2} = \dfrac{21}{2}$

$= 10.5$

Ex.2. Find an appropriate measure of dispersion for the following data:

Income (Rs)	Less than 50	50-70	70-90	90-110	110-130	130-150	Above 150
No. of persons	59	102	135	330	200	122	52

Solution:

Since the classes given are with open end intervals, the only measure of dispersion that we can compute the Semi inter quartile range.

Income (in ₹)	No. of persons	Less than c. f.
Less than 50	59	59
50-70	102	161
70-90	125	286
90-110	330	616
110-130	200	816
130-150	132	948
Above 150	52	1000

Table 2.2

Here, N = 1000,

$\therefore \dfrac{N}{4} = \dfrac{1000}{4} = 250$

$\therefore$ The first quartile class is : 70 – 90

$Q_1 = L + \dfrac{h}{f}\left(\dfrac{N}{4} - c.f.\right) = 70 + \dfrac{20}{125}(250 - 161) = 70 + 14.24 =$ Rs. 84.24

Now, $\dfrac{3N}{4} = 750$

$\therefore$ The third quartile class is : 110 – 130

$Q_3 = L + \dfrac{h}{f}\left(\dfrac{3N}{4} - c.f.\right) = 110 + \dfrac{20}{200}(750 - 616) = 110 + 13.4 =$ Rs. 123.4

$\therefore$ Semi Inter quartile range

$= \dfrac{Q_3 - Q_1}{2} = \dfrac{123.4 - 84.24}{2} = \dfrac{39.16}{2} = 19.58$

EXERCISE 2.1

1. Find range of the following data:
   575, 609, 335, 280, 729, 544, 852, 427, 967, 250

2. The following data gives number of typing mistakes done by Radha during a week. Find the range of the data.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
No. of mistakes	15	20	21	12	17	10

3. Find range for the following data.

Classes	62-64	64-66	66-68	68-70	70-72
Frequency	5	3	4	5	3

4. Find the Q. D. for the following data.
   3, 16, 8, 15, 19, 11, 5, 17, 9, 5, 3.

5. Given below are the prices of shares of a company for the last 10 days. Find Q. D.:
   172, 164, 188, 214, 190, 237, 200, 195, 208, 230

6. Calculate Q. D. for the following data.

X	24	25	26	27	28	29	30
F	6	5	3	2	4	7	3

7. Following data gives the age distribution of 250 employees of a firm. Calculate Q. D. of the distribution.

Age (In years)	20-25	25-30	30-35	35-40	40-45	45-50
No. of employees	30	40	60	50	46	14

8. Following data gives the weight of boxes. Calculate Q. D. for the data

Weights (kg.)	10-12	12-14	14-16	16-18	18-20	20-22
No. of boxes	3	7	16	14	18	2
c.f.	3	10	26	40	58	60

Let's Learn

2.5 VARIANCE and STANDARD DEVIATION:

The main drawback of the range and Q. D. is that both are based on only two values, and do not consider all the observations. The variance and standard deviation overcome this drawback as they are based on the deviations taken from the mean.

2.5.1 Variance:

The variance of a variable X is defined as the arithmetic mean of the squares of all deviations of X taken from its arithmetic mean.

It is denoted by Var(X) or σ^2 .

$$\text{Var}(X) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Let's Derive

We have, $V(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2\bar{x}.x_i + \bar{x}^2)$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}x_i^2 + \bar{x}^2\sum_{i=1}^{n}1 - 2\bar{x}\sum_{i=1}^{n}x_i\right]$$

$$= \frac{\sum_{i=1}^{n}x_i^2}{n} - \frac{2\bar{x}\sum_{i=1}^{n}x_i}{n} + \frac{\bar{x}^2\sum_{i=1}^{n}1}{n}$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 + \bar{x}^2\frac{n}{n} - 2\bar{x}\frac{\sum_{i=1}^{n}x_i}{n}$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 + \bar{x}^2 - 2\,\bar{x}\;\;\bar{x}$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 + \bar{x}^2 - 2\bar{x}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - (\bar{x})^2$$

2.5.2. Standard Deviation :

Standard Deviation is defined as the positive square root of the variance.

It is denoted by σ (sigma)

(i) Variance and Standard Deviation for raw data :

Let the variable X take the values $x_1, x_2, x_3, \ldots x_n$.

Let $\bar{x}$ be the arithmetic mean. Then,

$$\text{Var}(X) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{\sum_{i=1}^{n}x_i^2}{n} - \bar{x}^2$$

Where $\bar{x} = \frac{\sum_{i=1}^{n}x_i}{n}$

And $\text{S. D.} = \sigma = \sqrt{Var(X)}$

(ii) Variance and Standard Deviation for ungrouped frequency distribution :

Let $x_1, x_2, \ldots, x_n$ be the values of variable X with corresponding frequencies $f_1, f_2, \ldots, f_n$ respectively, then the variance of X is defined as

$$\text{Var}(X) = \sigma^2 = \frac{1}{N}\sum_{i=1}^{n}f_i(x_i - \bar{x})^2$$

$$= \frac{\sum_{i-1}^{n}f_i x_i^2}{N} - \bar{x}^2,$$

Where, $\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$, and $\sum_{i=1}^{n} f_i = N$ = Total frequency

$$S.D. = \sigma = \sqrt{Var(X)} = +\sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - (\bar{x})^2}$$

(iii) Variance and Standard Deviation for grouped frequency distribution :

Let $x_1, x_2, \ldots, x_n$ be the mid points and $f_1, f_2, \ldots, f_n$ are the corresponding class frequencies, then the Variance is defined as :

$$Var(X) = \sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i \left(x_i - \bar{x}\right)^2$$

$$= \frac{\sum_{i-1}^{n} f_i x_i^2}{N} - \bar{x}^2,$$

Where $\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$ and $\sum_{i=1}^{n} f_i = N$

= Total frequency

$$S.D. = \sigma = \sqrt{Var(X)} = +\sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - (\bar{x})^2}$$

Change of origin and scale:

1. The variance and consequently the standard deviation are independent of change of origin.

   That is, if d = X – A, where A is a constant, then $\sigma_x^2 = \sigma_d^2$.

2. The variance and consequently the standard deviation are not independent of change of scale.

   Let $u = \frac{X - A}{h}$, where A and h are constants and $h \neq 0$, then $\sigma_x^2 = h^2\sigma_u^2$

SOLVED EXAMPLES

Ex.1. Compute variance and standard deviation of the following data.

9, 12, 15, 18, 21, 24, 27

Solution :

Here, n = 7, $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{126}{7} = 18$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
9	-9	81
12	-6	36
15	-3	9
18	0	0
21	3	9
24	6	36
27	9	81
126		**252**

Table 2.3

Therefore, $Var(X) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$

$$= \frac{252}{7} = 36$$

$$S.D. = \sigma = \sqrt{Var(X)} = \sqrt{36}$$

$$= 6$$

Ex.2. Given below are the marks out of 25 of 5 students in a mathematics test. Calculate the variance and standard deviation of these observations.

Marks : 10, 13, 17, 20, 23

Solution : We use alternate (direct) method to solve this problem.

Calculation of variance :

x_i	x_i^2
10	100
13	169
17	289
20	400
23	529
83	**1487**

Table 2.4

Here, n = 5 and $\bar{x} = \frac{\sum_{i=1}^{n} xi}{n} = \frac{83}{5} = 16.6$

Therefore, $Var(X) = \sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - (\bar{x})^2$

$= \frac{1487}{5} - (16.6)^2$

$= 297.4 - 275.56$

$= 21.84$

S.D. $= \sigma = \sqrt{Var(X)} = \sqrt{21.84}$

Ex.3. A die is rolled 30 times and the following distribution is obtained. Find the variance and S.D,

Score	1	2	3	4	5	6
Frequency	2	6	2	5	10	5

Solution :

x	f	f.x	x^2	$f.x^2$
1	2	2	1	2
2	6	12	4	24
3	2	6	9	18
4	5	20	16	80
5	10	50	25	250
6	5	30	36	180
Total	**30**	**120**		**554**

Table 2.5

$N = \sum f = 30$

We get, $\bar{x} = \frac{\sum_{i=1}^{n} fx}{N} = \frac{120}{30} = 4$

Now, $\sigma_x^2 = \frac{\sum_{i=1}^{n} fx^2}{N} - \bar{x}^2 = \frac{554}{30} - 4^2$

$= 18.47 - 16$

$= 2.47$

Therefore, $\sigma_x = \sqrt{2.47}$

Short cut method/Step deviation method/ Change of origin and scale method :

Let $u = \frac{X - A}{h}$

Then $Var(u) = \sigma_u^2 = \frac{\sum_{i=1}^{n} u_i^2}{n} - (\bar{u})^2$

And $Var(X) = h^2.Var(u)$

i.e. $\sigma_x^2 = h^2.\ \sigma_u^2$

S.D. is $\sigma_x = h.\sigma_u$

Ex. 4. Compute variance and standard deviation for the following data:

x	15	20	25	30	35	40	45
f	13	12	15	18	17	10	15

Solution:

Let $u = \frac{x - 30}{5}$

x	u	f	f.u	$f.u^2$
15	-3	13	-39	117
20	-2	12	-24	48
25	-1	15	-15	15
30	0	18	0	0
35	1	17	17	17
40	2	10	20	40
45	3	15	45	135
Total		**100**	**4**	**372**

Table 2.6

We get, $\bar{u} = \frac{\sum_{i=1}^{n} f.u.}{N} = \frac{4}{100} = 0.04$

Now, $\sigma_u^2 = \frac{\sum_{i=1}^{n} f.u^2}{N} - (\bar{u})^2$

$= \frac{372}{100} - (0.04)^2$

$= 3.72 - 0.0016$

$= 3.7184$

Therefore, $\sigma_x = 5 \times \sqrt{3.1784}$

Ex.5. Compute variance and standard deviation for the following data.

x	45-55	55-65	65-75	75-85	85-95	95-105	105-115	115-125
f	7	20	27	23	13	6	3	1

Solution :

Let us use short cut method to find the solution.

Let $u = \frac{X-90}{10}$. Here, A = 90 and h = 10

Calculation of variance of u :

Class	Mid value (x_i)	f_i	u_i	f_iu_i	$f_iu_i^2$
45-55	50	7	-4	-28	112
55-65	60	20	-3	-60	180
65-75	70	27	-2	-54	108
75-85	80	23	-1	-23	23
85-95	90	13	0	0	0
95-105	100	6	1	6	6
105-115	110	3	2	6	12
115-125	120	1	3	3	9
Total		**100**	-	**-150**	**450**

Table 2.7

Now, $\bar{u} = \frac{\Sigma f_iu_i}{N} = \frac{-150}{100} = -1.5$

$$Var(u) = \sigma_u^2 = \frac{\Sigma f_iu_i^2}{N} - \bar{u}^2$$

$$= \frac{450}{100} - (-1.5)^2 = 4.5 - 2.25$$

$$= 2.25$$

Thus, $Var(u) = 2.25$

$\therefore Var(X) = h^2.Var(u) = 10^2 \times 2.25 = 225$

$\therefore$ S.D. $= \sigma_x = \sqrt{Var(X)} = \sqrt{225} = 15$

Ex.6. Find the standard deviation of the following frequency distribution which gives distribution of heights of 500 plants in centimeters.

$(\sqrt{6956} = 83.4)$

Height of plants (in cm)	20-25	25-30	30-35	35-40	40-45	45-50
No. of plants	145	125	90	40	45	55

Solution : Let $u = \frac{X-32.5}{5}$. Here A = 32.5 and h = 5

Calculation of variance of u :

Class	Mid value (x_i)	f_i	u_i	f_iu_i	$f_iu_i^2$
20-25	22.5	145	-2	-290	580
25-30	27.5	125	-1	-125	125
30-35	32.5	90	0	0	0
35-40	37.5	40	1	40	40
40-45	42.5	45	2	90	180
45-50	47.5	55	3	165	495
Total		**500**		**-120**	**1420**

Table 2.8

$N = \sum f = 500$

Now, $\bar{u} = \frac{\Sigma f_iu_i}{N} = \frac{-120}{500} = -0.24$

$$Var(u) = \sigma_u^2 = \frac{\Sigma f_iu_i^2}{N} - (\bar{u})^2$$

$$= \frac{1420}{500} - (-0.24)^2$$

$$= 2.84 - 0.0576 = 2.7824$$

Thus, $Var(u) = 2.7824$

Therefore, $Var(X) = h^2.Var(u) = 5^2 \times 2.7824$ $= 69.56$ cm^2

$\therefore$ S.D. $= \sigma_x = \sqrt{Var(X)} = \sqrt{69.56} = 8.34$

EXERCISE 2.2

Find the variance and S.D. for the following sets of numbers.

1. 7, 11, 2, 4, 9, 6, 3, 7, 11, 2, 5, 8, 3, 6, 8, 8, 2, 6
2. 65, 77, 81, 98, 100, 80, 129

3. Compute variance and standard deviation for the following data:

x	2	4	6	8	10
f	5	4	3	2	1

4. Compute the variance and S.D.

x	1	3	5	7	9
Frequency	5	10	20	10	5

5. Following data gives age of 100 students in a school. Calculate variance and S.D.

Age (In years)	10	11	12	13	14
No. of Students	10	20	40	20	10

6. The mean and variance of 5 observations are 3 and 2 respectively. If three of the five observations are 1, 3 and 5, find the values of other two observations.

7. Obtain standard deviation for the following data :

Height (in inches)	60-62	62-64	64-66	66-68	68-70
Number of students	4	30	45	15	6

8. The following distribution was obtained by change of origin and scale of variable X.

d_i	−4	−3	−2	−1	0	1	2	3	4
f_i	4	8	14	18	20	14	10	6	6

If it is given that mean and variance are 59.5 and 413 respectively, determine actual class intervals.

Let's Learn

2.5.3 Standard Deviation for Combined data :

If σ_1, σ_2 are standard deviations and $\bar{x}_1, \bar{x}_2$ are the arithmetic means of two data sets of sizes n_1 and n_2 respectively ,then the mean for the combined data is :

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

And the Standard Deviation for the combined series is :

$$\sigma_c = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

Where, $d_1 = \bar{x}_1 - \bar{x}_c$ and $d_2 = \bar{x}_2 - \bar{x}_c$

SOLVED EXAMPLES

Ex.1. The means of two samples of sizes 10 and 20 are 24 and 45 respectively and the standard deviations are 6 and 11. Obtain the standard deviation of the sample of size 30 obtained by combining the two samples. ($\sqrt{19067} = 138$)

Solution :

Let $n_1 = 10$, $n_2 = 20$, $\bar{x}_1 = 24$, $\bar{x}_2 = 45$, $\sigma_1 = 6$, $\sigma_2 = 11$

Combined mean is :

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{10\times24 + 20\times45}{10+20}$$

$$= \frac{1140}{30} = 38$$

Combined standard deviation is given by,

$$\sigma_c = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

Where, $d_1 = \bar{x}_1 - \bar{x}_c$ and $d_2 = \bar{x}_2 - \bar{x}_c$

$\therefore d_1 = \bar{x}_1 - \bar{x}_c = 24 - 38 = -14$ and $d_2 = \bar{x}_2 - \bar{x}_c = 45 - 38 = 7$

$$\sigma_c = \sqrt{\frac{10(6^2 + (-14)^2) + 20(11^2 + 7^2)}{10+20}}$$

$$= \sqrt{\frac{2320 + 3400}{30}} = \sqrt{\frac{5720}{30}}$$

$$= \sqrt{190.67} = 13.8$$

Ex.2. The first group has 100 items with mean 45 and variance 49. If the combined group has 250 items with mean 51 and variance 130, find the mean and standard deviation of second group.

Solution :

Given, $n_1 = 100$, $\bar{x}_1 = 45$, $\sigma_1^2 = 49$

For combined group, $n = 250$, $\bar{x}_c = 51$,

$\sigma_c^2 = 130,$

To find : $\bar{x}_2$ and σ_2

$n = n_1 + n_2 \rightarrow 250 = 100 + n_2 \rightarrow n_2 = 150$

We have, $\bar{x}_c = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

$$51 = \frac{100 \times 45 + n_2\bar{x}_2}{100 + 150}$$

$$12750 = 4500 + 150.\ \bar{x}_2$$

Therefore, $\bar{x}_2 = 55$

Combined standard deviation is given by,

$$\sigma_c^2 = \frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}$$

Where, $d_1 = \bar{x}_1 - \bar{x}_c = 45\text{-}51 = -6$

and $d_2 = \bar{x}_2 - \bar{x}_c = 55\text{-}51 = 4$

$$130 = \frac{100(49 + 36) + 150(\sigma_2^2 + 16)}{100 + 150}$$

$$32500 = 150\sigma_2^2 + 10900$$

$$\sigma_2^2 = 144$$

Therefore, $\sigma_2 = \sqrt{144} = 12$

$\therefore$ S.D. of second group = 12

Let's Learn

2.6 Coefficient of Variation :

Standard deviation depends on the unit of measurements. It is therefore not useful for comparing series of data measured in different units **Coefficient of Variation (C.V.)** is independent of the unit of measurement, and it is defined by,

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

Coefficient of Variation is used to compare the variability of two distributions. A distribution with smaller C.V. is said to be more homogenous or compact, more consistent or steadier and a distribution with larger C.V. is said to be more heterogeneous or more variable.

SOLVED EXAMPLES

Ex.1. The arithmetic mean of runs scored by 3 batsmen Varad, Viraj and Akhilesh in the same series are 50, 58 and 21 respectively. The standard deviation of their runs are 11, 16 and 5 respectively. Who is the most consistent of the three? If one of the three is to be selected, who will be selected?

Solution :

Let $\bar{x}_1, \bar{x}_2, \bar{x}_3$ and $\sigma_1, \sigma_2, \sigma_3$ be the means and standard deviations of the three batsmen Varad, Viraj and Akhilesh respectively.

Therefore, $\bar{x}_1 = 50$, $\bar{x}_2 = 58$, $\bar{x}_3 = 21$ and $\sigma_1 = 11$, $\sigma_2 = 16$, $\sigma_3 = 5$

Now,

C. V. of runs scored by Varad $= 100 \times \dfrac{\sigma_1}{\bar{x}_1} = 100 \times \dfrac{11}{50} = 22\%$

C. V. of runs scored by Viraj $= 100 \times \dfrac{\sigma_2}{\bar{x}_2} = 100 \times \dfrac{16}{58} = 27.59\%$

C. V. of runs scored by Akhilesh $= 100 \times \dfrac{\sigma_3}{\bar{x}_3} = 100 \times \dfrac{5}{21} = 23.81\%$

(i) Since the C. V. of the runs is smaller for Varad, he is the most consistent player.

(ii) To take decision regarding the selection, let us consider both the C.V.s and means.

(a) Based on consistency :

Since C.V. of Varad is smallest, he is more consistent and hence is to be selected.

(b) Based on expected score :

If the player with highest expected score is to be selected, then Viraj will be selected.

Ex.2. The following values are calculated in respect of prices of shares of companies X and Y. State the share of which company is more stable in value.

	Share of X	Share of Y
Mean	50	105
Variance	7	4

Solution :

Here, $\sigma_x^2 = 7$, $\sigma_y^2 = 4$, $\bar{x} = 50$, $\bar{y} = 105$

Therefore $\sigma_x = \sqrt{7} = 2.64$, $\sigma_y = 2$,

$$\text{C.V.}(X) = 100 \times \frac{\sigma_x}{\bar{x}} = 100 \times \frac{2.64}{50} = 5.28\%$$

$$\text{C.V.}(Y) = 100 \times \frac{\sigma_y}{\bar{y}} = 100 \times \frac{2}{105} = 1.90\%$$

Since C.V.(Y) < C.V.(X),

The shares of company Y are more stable in value.

Let's Construct…

Construct the table showing the frequencies of words with different number of letters occurring in the following passage, omitting punctuation marks. Take the number of letters in each word as one variable and obtain the mean, S.D. and the coefficient of variation of its distribution.

"Take up one idea. Make that one idea your life – think of it, dream of it, live on that idea. Let the brain, muscles, nerves, every part of your body, be full of that idea, and just leave every other idea alone. This is way to success."

EXERCISE 2.3

1. Mean and standard deviation of two distributions of 100 and 150 items are 50, 5 and 40, 6 respectively. Find the mean and standard deviations of all the 250 items taken together.

2. For a certain bivariate data, following information is available.

	X	Y
Mean	13	17
S. D.	3	2
Size	10	10

Obtain the combined standard deviation.

3. Calculate coefficient of variation of marks secured by a student in the exam, where the marks are: 2, 4, 6, 8, 10.

(Given : $\sqrt{2} = 1.41$)

4. Find the coefficient of variation of a sample which has mean 25 and standard deviation 5.

5. A group of 65 students of class XI have their average height 150.4 cm with coefficient of variation 2.5%. What is the standard deviation of their height?

6. Two workers on the same job show the following results:

	Worker P	Worker Q
Mean time for completing the job (hours)	33	21
Standard Deviation (hours)	9	7

(i) Regarding the time required to complete the job, which worker is more consistent?

(ii) Which worker seems to be faster in completing the job?

7. A company has two departments with 42 and 60 employees respectively. Their average weekly wages are Rs. 750 and Rs. 400. The standard deviations are 8 and 10 respectively.

(i) Which department has a larger bill?

(ii) Which department has larger variability in wages?

8. The following table gives weights of the students of class A. Calculate the coefficient of variation (Given : $\sqrt{0.8} = 0.8944$)

Weight (in kg)	Class A
25-35	8
35-45	4
45-55	8

9. Compute coefficient of variation for team A and team B.

No. of goals	0	1	2	3	4
No. of matches played by team A	19	6	5	16	14
No. of matches played by team B	16	16	5	18	15

Which team is more consistent?

10. Given below is the information about marks obtained in Mathematics and Statistics by 100 students in a class. Which subject shows the highest variability in marks?

	Mathematics	Statistics
Mean	20	25
S.D.	2	3

Let's Remember

- **Range = Largest Value – Smallest Value = L – S**
- **Inter quartile range = $Q_3 - Q_1$**
- **Quartile deviation = $\frac{Q_3 - Q_1}{2}$**
- **Variance and Standard Deviation for raw data :**

Let the variable X takes the values $x_1, x_2, x_3, \ldots, x_n$.

Let $\bar{x}$ be the arithmetic mean. Then,

$$\text{Var}(X) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\bar{x}\right)^2$$

Where $\bar{x} = \frac{\Sigma x_i}{n}$

And S. D. $= \sigma_x = \sqrt{Var(X)}$

- **Variance and Standard Deviation for frequency distribution :**

$$\text{Var}(X) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2$$

$$= \sum_{i=1}^{n} f_i x_i^2 - (\bar{x})^2$$

Where, $\bar{x} = \frac{\Sigma f_i x_i}{N}$

and $\Sigma f_i = N =$ Total frequency

S. D. $= \sigma = \sqrt{Var(X)}$

- **Step - deviation Method : (Change of origin & scale method)**

Let $u = \frac{X - A}{h}$,

Where A is assumed mean and h is width of the class (common multiple)

Then $Var(u) = \sigma_u^2 = \frac{\Sigma f_i u_i^2}{n} - \left(\bar{u}\right)^2$

And $Var(X) = h^2 . Var(u)$

i.e. $\sigma_x^2 = h^2 \sigma_u^2$

S.D. is $\sigma_x = h.\ \sigma_u$

- **Standard Deviation for Combined data :**

$$\sigma_c = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

Where, $d_1 = \bar{x}_1 - \bar{x}_c$ and $d_2 = \bar{x}_2 - \bar{x}_c$

- **Coefficient of Variation :**

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

MISCELLANEOUS EXERCISE-2

Q. Find the range for the following data :

1. 116, 124, 164, 150, 149, 114, 195, 128, 138, 203, 144

2. Given below the frequency distribution of weekly wages of 400 workers. Find the range.

Weekly wages (in '00 Rs.)	10	15	20	25	30	35	40
No. of workers	45	63	102	55	74	36	25

3. Find the range of the following data

Classes	115-125	125-135	135-145	145-155	155-165	165-175
Frequency	1	4	6	1	3	5

4. The city traffic police issued challans for not observing the traffic rules:

Day of the Week	Mon	Tue	Wed	Thus	Fri	Sat
No. of Challans	40	24	36	58	62	80

Find Q.D.

5. Calculate Q.D. from the following data.

X (less than)	10	20	30	40	50	60	70
Frequency	5	8	15	20	30	33	35

6. Calculate the appropriate measure of dispersion for the following data.

Wages (In Rs.)	Less than 35	35-40	40-45	45-50	50-55	55-60
No. of Workers	15	50	85	40	27	33

7. Calculate Q.D. for the following data.

height of plants (in feet)	2-4	4-6	6-8	8-10	10-12	12-14	14-16
No. of plants	15	20	25	12	18	13	17

8. Find variance and S.D. for the following set of numbers.

25, 21, 23, 29, 27, 22, 28, 23, 27, 25

(Given $\sqrt{6.6} = 2.57$)

9. Following data gives no. of goals scored by a team in 100 matches.

No. of Goals Scored	0	1	2	3	4	5
No. of matches	15	20	25	15	20	5

Compute variance and standard deviation for the above data.

10. Compute arithmetic mean and S.D. and C.V. (Given $\sqrt{296} = 17.20$)

C.I.	45-55	55-65	65-75	75-85	85-95	95-105
f	4	2	5	3	6	5

11. The mean and S.D. of 200 items are found to be 60 and 20 respectively. At the time of calculation two items were wrongly taken as 3 and 67 instead of 13 and 17. Find the correct mean and variance.

12. The mean and S.D. of a group of 48 observation are 40 and 8 respectively. If two more observations 60 and 65 are added to set, find the mean and S.D. of 50 items.

13. The mean height of 200 students is 65 inches. The mean heights of boys and girls are 70 inches and 62 inches respectively and the standard deviations are 8 and 10 respectively. Find the number of boys and the combined S.D.

14. From the following data available for 5 pairs of observations of two variables x and y, obtain the combined S.D. for all 10 observations.

Where, $\sum_{i=1}^{n} x_i = 30$, $\sum_{i=1}^{n} y_i = 40$, $\sum_{i=1}^{n} x_i^2 = 225$,

$\sum_{i=1}^{n} y_i^2 = 340$

15. The mean and standard deviations of two brands of watches are given below :

	Brand-I	Brand-II
Mean	36 months	48 months
S.D.	8 months	10 months

Calculate coefficient of variation for the two brands and interpret the results.

16. Calculate coefficient of variation for the data given below [Given : $\sqrt{3.3} = 1.8161$]

C.I.	5-15	15-25	25-35	35-45	45-55	55-65	65-75
f	6	7	15	25	8	18	21

Activity 2.1

Range :

The daily sale of sugar in a certain grocery shop is given below :

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
120kg	75kg	33kg	140.5kg	50kg.	70.5kg

Then L = ________, S = ________,

Range = L − S = __________kg

Activity 2.2

Inter quartile range :

Consider the following data :

2, 4, 5, 6, 8, 10, 12, 14, 20, 30, 60

Here, n = 11

Q_1 = Value of $\left(\frac{n+1}{4}\right)^{th}$ observation

= Value of 3rd observation = __________

Q_3 = Value of 3 $\left(\frac{n+1}{4}\right)^{th}$ observation

= Value of ____th observation = _________

∴ Inter Quartile Range = Q_3 − ____ = _______

Activity 2.3

Variance and Standard Deviation :

Suppose the returns on an investment for 4 years are :

Rs.1000, Rs.3000, Rs. 4500 & Rs.5000.

Then, mean = $\bar{x}$ = _______

$x - \bar{x}$: _____, _______, ______, ______

$(x - \bar{x})^2$: _____, _______, ______, ______

$\sum (x - \bar{x})^2$: ___________

$$\text{Var}(X) = \frac{\sum(x-\bar{x})^2}{n}$$

$\text{S.D.} = \sqrt{\square}$

Activity 2.4

Select 4 grocery items, like Rice, Wheat, Sugar and Tur-Dal, Make 4 groups of students

1st group will record price of Rice in 10 shops

2nd group will record price of Wheat in 10 shops

3rd group will record price of Sugar in 10 shops

4th group will record price of Tur-Dal in 10 shops. Every group will find mean & S.D. for each of the grocery items.

Use same activity to calculate all partition values.

❖ ❖ ❖