8 Measures of Dispersion

Let's Study

- Meaning and Definition of Dispersion
- Range of data.
- 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

"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 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

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 dependable, the measure of mean is not sufficient. The spread of data or variation is a factor which needs our attention. To understand more about this we need some other measure. One such measure is Dispersion.

In the above example, observations from series X and series Z are more scattered as compared to those in series Y. So Y is more consistent. The extent of scatter in observations which deviate from mean is called dispersion.

Activity :

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

According to Spiegel:

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

8.1 Measures of Dispersion :

Following measures of dispersion are the commonly used –

(i) Range

(ii) Variance

(iii) Standard deviation

8.1.1 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.

Uses of Range:

1) It is used in stock market.
2) It is used in calculations of mean temperature of a certain place.

SOLVED EXAMPLES

Ex.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

Solution : Smallest Value = S = 36

Largest Value = L = 73

Range = L – S = 73 – 36 = 37

Ex.2. Calculate range for the following data.

Salary (00's Rs.)	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

$\therefore$ Range = L – S = 150 – 30 = 120

EXERCISE 8.1

1. Find range of the following data:
   19, 27, 15, 21, 33, 45, 7, 12, 20, 26

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

3. 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

4 Following results were obtained by rolling a die 25 times. Find the range of the data.

Score	1	2	3	4	5	6
Frequency	4	6	2	7	3	3

5. Find range for the following data.

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

Let's Learn

8.2 VARIANCE and STANDARD DEVIATION:

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

8.2.1 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 .

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

Note :

We have,

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

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

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

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

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

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

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

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

8.2.2 Standard Deviation :

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

It is denoted by σ (sigma) and $\sigma = \sqrt{Var(X)}$

(i) 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,

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

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

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

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

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

Where, $\bar{x} = \frac{\sum_{i=1}^{n}f_ix_i}{N}$, and $\sum f_i = N =$ Total frequency S. D. $= \sigma = \sqrt{Var(X)}$

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

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

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

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

Where,

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

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

8.2.3 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 the class mark of middle class interval if the number of classes is odd. If the number of classes is even, then there will be two middle class intervals and A is the class mark of the one having greater frequency.

   then $\sigma_d^2 = \sigma_x^2$ and $\sigma_d = \sigma_x$

   It means that if σ_x is standard deviation of the values $x_1, x_2, x_3, \ldots x_n$. The standard deviation σ_d, of $x_1 - A, x_2 - A, x_3 - A, \ldots\ldots, x_n - A$ is also same as that of σ_x.

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

Let u = $\frac{x-A}{h}$ where h is width of the class interval if given. If the class intervals are not given, then h is the difference (or distance) between the two consecutive value of x_i.

and $h \neq 0$, then $\sigma_x = h\, \sigma_u$

and $\sigma_x^2 = h^2\, \sigma_u^2$

It means that if σ_x is standard deviation of the values $x_1, x_2, \ldots x_n$. Then standard deviation σ_u of $\frac{x_1-A}{h}, \frac{x_2-A}{h}, \frac{x_3-A}{h}, \ldots, \frac{x_n-A}{h}$ is $\frac{1}{h}$ times of σ_x.

SOLVED EXAMPLES

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

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

Solution :

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**

Here, n = 7, $\bar{x} = \frac{\sum x_1}{n} = \frac{126}{7} = 18$

$$\text{Var}(X) = \sigma^2 = \frac{1}{n}\sum(x_i - \bar{x})^2 = \frac{252}{7} = 36$$

$$\text{S.D.} = \sigma = \sqrt{Var(X)} = \sqrt{36} = 6$$

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

Marks : 10, 13, 17, 20, 23

Solution : We use alternate 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**

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

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

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

$$= 297.4 - 275.56$$

$$= 21.84$$

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

$$= 4.67$$

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

Score (X)	1	2	3	4	5	6
Frequency (f)	2	6	2	5	10	5

Solution :

X	f	f.x	f.x²
1	2	2	2
2	6	12	24

3	2	6	18
4	5	20	80
5	10	50	250
6	5	30	180
Total	**30**	**120**	**554**

We get, $\bar{x} = \dfrac{\sum fx}{N} = \dfrac{120}{30} = 4$

Now, $\sigma_x^2 = \dfrac{\sum fx^2}{N} - \bar{x}^2$

$$= \frac{554}{30} - 4^2 = 18.47 - 16 = 2.47$$

Therefore, $\sigma_x = \sqrt{2.47} = 1.57$

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 $u = \dfrac{x-30}{5}$

X	u	f	f.u	f.u²
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**

We get, $\bar{u} = \dfrac{\sum f.u}{N} = \dfrac{4}{100} = 0.04$

Now, $\sigma_u^2 = \dfrac{\sum f.u^2}{N} - \bar{u}^2$

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

$$= 3.72 - 0.0016$$

$$= 3.7184$$

Therefore, $\sigma_u = \sqrt{3.1784} = 1.92$

$\sigma_x = h\,\sigma_u$

$= 5(1.92)$

$\therefore\ \sigma_x = 9.6$

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

C.I.	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 $u = \dfrac{X-90}{10}$.

Calculation of variance of u :

Class-intervals	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**

Now, $\bar{u} = \dfrac{\sum f_iu_i}{N} = \dfrac{-150}{100} = -1.5$

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

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

$$= 2.25$$

Thus, $Var(u) = 2.25$

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

$\therefore$ S.D. $= \sigma_x = \sqrt{Var(X)} = \sqrt{255} = 15$

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

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 $u = \dfrac{X - 32.5}{5}$.

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				**–120**	**1420**

Now, $\bar{u} = \dfrac{\sum f_i u_i}{N} = \dfrac{-120}{500} = -0.24$

$$Var(\text{u}) = \sigma_u^2 = \frac{\sum f_i u_i}{N} - \bar{u}^2$$

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

$$= 2.84 - 0.0576 = 2.7824$$

Thus, $Var(u) = 2.7824$

$\therefore$ $Var(X) = h^2 . Var(u) = 5^2 \times 2.7824 = 69.56 \text{ cm}^2$

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

EXERCISE 8.2

Q. Find variance and S.D. for the following set 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	12	14	16	18	20
F	8	10	10	7	6	4	3	4	2	6

4. Compute the variance and S.D.

X	31	32	33	34	35	36	37
Frequency	15	12	10	8	9	10	6

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

Age (In years)	16	17	18	19	20	21
No. of Students	20	7	11	17	30	15

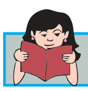
6. Find mean, variance and S.D. of the following data.

Class-es	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Freq.	7	14	6	13	9	15	11	10	15

7. Find the variance and S.D. of the following frequency distribution which gives the distribution of 200 plants according to their height.

Height (in cm)	14-18	19-23	24-28	29-33	34-38	39-43	44-48
No. of plants	5	18	44	70	36	22	5

8. The mean of 5 observations is 4.8 and the variance is 6.56. If three of the five observations are 1, 3 and 8, find the other two observations.



Let's :Learn

8.3 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.

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 :

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{10 \times 24 + 20 \times 45}{10 + 20}$$

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

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

$d_2 = \bar{x}_2 - \bar{x}_c = 45 - 38 = -7$

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

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

$$= \sqrt{190.67} = 13.4$$

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_c

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

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

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

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

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

8.3.1 Coefficient of Variation :

Standard deviation depends upon the unit of measurement. Therefore it cannot be used to compare two or more series expressed in different units. For this purpose coefficient of variation (C.V.) is used and is defined as,

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

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 and the series with larger C.V. is said to be more heterogeneous. The distribution with smaller C.V. is said to be more consistent.

SOLVED EXAMPLES

Ex.1) The arithmetic mean of runs scored by 3 batsmen Varad, Viraj and Akhilesh in the same series are 50, 52 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 σ_1, σ_2, σ_3 be the means and standard deviations of the three batsmen Varad, Viraj and Akhilesh respectively.

Therefore, $\bar{x}_1 = 50$, $\bar{x}_2 = 52$, $\bar{x}_3 = 21$ 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 $= 100 \times \dfrac{\sigma_2}{\bar{x}_2}$

$= 100 \times \dfrac{16}{52} = 30.76$

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 (mean) 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 Y	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 $\sigma_x = \sqrt{7} = 2.64$, $\sigma_y = \sqrt{4} = 2$

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

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

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

The shares of company Y are more stable in value.

Activity : 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 8.3

1. The means of two samples of sizes 60 and 120 respectively are 35.4 and 30.9 and the standard deviations 4 and 5. Obtain the standard deviation of the sample of size 180 obtained by combining the two sample.

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

	X	Y
Mean	13	17
S. D.	3	2
Size	20	30

Obtain the combined standard deviation.

3. Calculate coefficient of variation of marks secured by a student in the exam, where the marks are: 85, 91, 96, 88, 98, 82

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

5. A group of 65 students of class XI have their average height is 150.4 cm with coefficient of variance 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 two classes. Calculate the coefficient of variation of the two distributions. Which series is more variable?

Weight (in kg)	Class A	Class B
30-40	22	13
40-50	16	10
50-60	12	17

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	14	10	14	16

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

Range :

Activity 1 :

The daily sale of wheat in a certain shop is given below.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Sale in Kg	135	39	142.5	78	120.5	93

Then L = ____________ S = ____________ Range L – S = ____________

Activity 2 :

Neeraj Chopra is an Indian track and field athlete, who competes in the Javelin throw.

The following data reveals his record of throws in Asian Championships (A.C.) and World Championships (W.C.)

Championship	2015 (A.C.)	2016 (A.C.)	2017 (A.C.)	2018 (A.C.)	2013 (W.C.)	2016 (W.C.)	2017 (W.C.)
Throw (Meters)	70.50	77.60	85.23	88.06	66.75	86.48	82.26

Then L = ____________ S = ____________ Range L – S = ____________

Variance and Standard Deviation :

Activity 1 :

The marks scored in a test by seven randomly selected students are

3	4	6	2	8	8	5

Find the Variance and Standard Deviation of these seven students.

Solution :

Mean $\overline{x} = \frac{3+4+\ \ +\ \ +\ \ +\ \ +5}{7} = \frac{36}{7}$

The deviation from mean for each observation is $(x - \overline{x})$

$3-\frac{36}{7}$				$8-\frac{36}{7}$		$5-\frac{\ }{7}$
$-\frac{15}{7}$				$\frac{20}{7}$		

The deviations squared are $(x - \overline{x})^2$

$\frac{225}{49}$				$\frac{400}{49}$		

Variance $= \frac{\sum (x-\overline{x})^2}{n}$ Standard Deviation $= \sqrt{\text{Variance}} = \sqrt{\qquad} =$

Activity 2 :

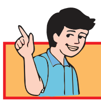
The number of centuries scored in a year by seven randomly selected batsmen are

3	5	6	3	7	6	4

Find the Variance and Standard Deviation of these seven batsmen.

Solution :

x	3	5	6	3	7	6	4	Result
x^2	9							
$\sum x$	3 +	+	+	+	+	+		=
$\sum x$	81 +	+	+	+	49 +	+		=
Variance		$= \sigma^2$	$= \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$			$= \frac{\quad}{7} - \left(\frac{\quad}{7}\right)^2$		$= \frac{7(\quad) - \quad}{49}$ $=$
Standard Deviation		σ	$= \sqrt{\text{Variance}} = \sqrt{\qquad} =$					



Let's Remember

- **Range = Largest Value – Smallest Value = L – S**
- **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,

  $$\text{Var (X)} = \sigma^2$$

  $$= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{\sum_{i=1}^{n} x_1^2}{n} - \bar{x}^2$$

  Where , $\bar{x} = \frac{\sum x_i}{n}$

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

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

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

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

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

  and $\sum f_i = N =$ Total frequency

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

- **Change of origin and scale method :**

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

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

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

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

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

- **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 - 8

(I) Select the correct option from the given alternatives :

1) If there are 10 values each equal to 10, then S.D. of these values is : ————.

A) 100 B) 20
C) 0 D) 10

2) Number of patients who visited cardiologists are 13, 17, 11, 15 in four days then standard deviation (approximately) is

A) 5 patients B) 4 patients
C) 10 patients D) 15 patients

3) If the observations of a variable X are, –4, –20, –30, –44 and –36, then the value of the range will be:

A) -48 B) 40
C) –40 D) 48

4) The standard deviation of a distribution divided by the mean of the distribution and expressing in percentage is called:

A) Coefficient of Standard deviation
B) Coefficient of skewness
C) Coefficient of quartile deviation
D) Coefficient of variation

5) If the S.D. of first n natural numbers is $\sqrt{2}$, then the value of n must be ————.

A) 5 B) 4
C) 7 D) 6

6) The positive square root of the mean of the squares of the deviations of observations from their mean is called:

A) Variance B) Range
C) S.D. D) C.V.

7) The variance of 19, 21, 23, 25 and 27 is 8. The variance of 14, 16, 18, 20 & 22 is:

A) Greater than 8 B) 8
C) Less than 8 D) 8 – 5 = 3

8) For any two numbers SD is always

A) Twice the range B) Half of the range
C) Square of the range D) None of these

9) Given the heights (in cm) of two groups of students:

Group A : 131 cm, 150 cm, 147 cm, 138 cm, 144 cm

Group B : 139 cm, 148 cm, 132 cm, 151 cm, 140 cm

Which of the following is / are the true?

I) The ranges of the heights of the two groups of students are the same.

II) The means of the heights of the two groups of students are the same.

A) I only B) II only C) Both I and II
D) None

10) Standard deviation of data is 12 and mean is 72 then coefficient of variation is

A) 13.67% B) 16.67%
C) 14.67% D) 15.67%

(II) Answer the following :

1. 76, 57, 80, 103, 61, 63, 89, 96, 105, 72
Find the range for the following data.

2. 116, 124, 164, 150, 149, 114, 195, 128, 138, 203, 144

3. 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

4. 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

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

5. 25, 21, 23, 29, 27, 22, 28, 23, 21, 25

6. 125, 130, 150, 165, 190, 195, 210, 230, 245, 260

7. Following data gives no. of goals scored by a team in 100 matches. Compute the standard deviation.

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

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

X	62	30	64	47	63	46	35	28	60
F	5	8	3	4	5	7	8	3	7

9. Calculate S.D. from following data.

Age (In yrs)	20-29	30-39	40-49	50-59	60-69	70-79	80-89
Freq.	65	100	55	87	42	38	13

10. Given below is the frequency distribution of marks obtained by 100 students. Compute arithmetic mean and S.D.

Marks	40-49	50-59	60-69	70-79	80-89	90-99
No. of students	4	12	25	28	26	5

11. The arithmetic mean and standard deviation of a series of 20 items were calculated by a student as 20 cms and 5 cms respectively. But while calculating them, an item 13 was misread as 30. Find the corrected mean and standard deviation.

12. The mean and S.D. of a group of 50 observation are 40 and 5 respectively. If two more observations 60 and 72 are added to the set, find the mean and S.D. of 52 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, $\sum_{i=1}^{n} x_i = 30$, $\sum_{i=1}^{n} y_i = 40$, $\sum_{i=1}^{n} x_i^2 = 220$,

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

15. Calculate coefficient of variation of the following data.

23, 27, 25, 28, 21, 14, 16, 12, 18, 16

16. Following data relates to the distribution of weights of 100 boys and 80 girls in a school.

	Boys	Girls
Mean	60	47
Variance	16	9

Which of the two is more variable?

17. The mean and standard deviations of two bands of watches are given below :

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

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

18. Calculate coefficient of variation for the data given below

Size (cm)	5-8	8-11	11-14	14-17	17-20	20-23	23-26
No of items	3	14	13	16	19	24	11

19. Calculate coefficient of variation for the data given below

Income (Rs.)	3000-4000	4000-5000	5000-6000	6000-7000	7000-8000	8000-9000	9000-10000
No. of families	24	13	15	28	12	8	10

20. Compute coefficient of variations for the following data to show whether the variation is greater in the yield or in the area of the field.

Year	Area (in acres)	Yield (in lakhs)
2011-12	156	62
2012-13	135	70
2013-14	128	68
2014-15	117	76
2015-16	141	65
2016-17	154	69
2017-18	142	71

21. There are two companies U and V which manufacture cars. A sample of 40 cars each from these companies are taken and the average running life (in years) is recorded.

Life (in years)	No of Cars	
	Company U	Company V
0-5	5	14
5-10	18	8
10-15	17	18

Which company shows greater consistency?

22. The means and S.D. of weights and heights of 100 students of a school are as follows.

	Weights	Heights
Mean	56.5 kg	61 inches
S.D.	8.76 kg	12.18 inches

Which shows more variability, weights or heights?

15) $e = \pm\frac{1}{\sqrt{3}}$ 17) $y = \pm 4$ or $y = 8x + 2\sqrt{11}$

18) $2x + 3y = 25$ 19) (1,2)

20) $x^2 - xy - 5 = 0$

22) i) $\frac{x^2}{36} - \frac{4y^2}{25} = 1$ ii) $\frac{x^2}{16} - \frac{y^2}{20} = 1$

iii) $\frac{x^2}{4} - \frac{4y^2}{9} = 1$

23) i) $17x - 2y + 17 = 0$ ii) $10x - 3\sqrt{3}y = 15$

iii) $32x - 25y = 400\sqrt{3}$

24) (3,2) 25) $y = 2x \pm 4$

26) $k(x^2 - a^2) = 2xy$

8. MEASURES OF DISPERSION

Exercise : 8.1

1) 38 2) 717 3) 11 4) 5 5) 10

Exercise : 8.2

1) $\sigma^2 = 8$; $\sigma = 2.82$
2) $\sigma^2 = 380$; $\sigma = 19.49$
3) $\sigma^2 = 32.39$; $\sigma = 5.69$
4) $\sigma^2 = 4.026$; $\sigma = 2.006$
5) $\sigma^2 = 3.0275$; $\sigma = 1.74$
6) $x = 58.2$; $\sigma^2 = 653.76$; $\sigma = 25.56$
7) $\sigma^2 x = 41.25$; $\sigma x = 6.42$
8) 5 and 7

Exercise : 8.3

1) $\sigma_c = 5.15$
2) $\sigma_c = 3.14$
3) C.V. = 6.32
4) C.V. = 20
5) S.D. = 3.76
6) $(C.V.)_P = 27.27$; $(C.V.)_Q = 33.33$;
   i) Worker P is more consistent.
   ii) Worker Q seems to be faster in completing the job.
7) $(C.V.)_1 = 1.07$ $(C.V.)_2 = 2.5$
   i) First department has larger bill
   ii) Second department has larger variability in wages.
8) $(C.V)_A = 18.6$; $(C.V)_B = 18.7$
   Series B is more variable
9) $(C.V)_A = 80$; $(C.V)_B = 74.5$
   Team B is more consistent.
10) $(C.V)_M = 10$; $(C.V)_S = 12$
   The subject Statistic shows higher vairablility in marks.

MISCELLANEOUS EXERCISE - 8

(I)

1	2	3	4	5	6	7	8	9	10
C	A	B	D	A	C	B	B	C	B

(II)

1) Range = 48
2) Range = 89
3) Range = Rs. 30
4) Range = 60
5) $\sigma = 2.72$
6) S. D. = 14.14
7) S. D. = 1.48
8) S. D. = 13.42
9) S. D. = 16.85
10) A. M. = 72; S. D. = 13.92
11) Mean = 19.15; S. D. = 4.66
12) Mean = 41; S. D. = 7.1
13) Number of boys = 75
    combined S. D. = 10.07

14) combined S. D. = 2.65

15) C.V. = 26.65

16) $(C.V.)_B = 6.67$ $(C.V.)_G = 6.38$
Series of boys is more variable

17) $(C.V.)_I = 22.22$ $(C.V.)_{II} = 20.83$
Brand-I is more variable

18) C.V. = 29.76

19) C.V. = 31.35

20) $(C.V.)_x = 9.21$; $(C.V.)_Y = 5.91$
The variation is greater in the area of the field.

21) $(C.V.)_U = 37.67$; $(C.V.)_V = 55.5$
i) Company U gives higher average life
ii) Company U shows greater consistency in performance.

22) $(C.V.)_1 = 15.50$ $(C.V.)_2 = 19.96$
Height shows more variability

9. PROBABILITY

Exercise : 9.1

1) S ={RR, GR, BR, PR, RG, GG, BG, PG, RB, GB, BB, PB, RP, GP, BP, PP}

a) A = {RR, GR, RB, RP, GR, BR, PR}

b) B = {RG, RB, RP, GR, GB, GP, BR, BG, BP, PR, PG, PB}

2) S = {(H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6), (T, 1), (T, 2), (T, 3), (T, 4), (T, 5), (T, 6)

a) A = {(T, 1), (T, 3), (T, 5)}

b) B = (H, 2), (H, 3),(H, 5),(T, 2), (T, 3), (T, 5),
c) C = (H, 1), (H, 4),

3) i) 56 ii) 120 iii) 720 iv) 1140

4) S = {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

a) A : {(1, 2), (2, 1), (1, 3), (2, 2), (3, 1), (1, 5), (2, 4), (3, 3), (4, 2), (5, 1), (2, 6), (3, 5), (4, 4), (5, 3), (6, 2), (3, 6), (4, 5), (5, 4), (6, 3), (6, 6)

b) B : {(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)

c) C : {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6)}

d) D : {(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

c) A and B are mutually exclusive and exhaustive.

f) C and Dare mutually exclusive and exhaustive.

5) a) S = {(5, 5), (5, 6), (5, 7), (5, 8), (6, 5), (6, 6), (6, 7), (6, 8), (7, 5), (7, 6), (7, 7), (7, 8), (8, 5), (8, 6), (8, 7), (8, 8)}

b) S = {(5, 6), (5, 7), (5, 8), (6, 5), (6, 7), (6, 8), (7, 5), (7, 6), (7, 8), (8, 5), (8, 6), (8, 7),}

6) a) $\frac{1}{9}$ b) 5/12 c) 1/6 d) 1/9

7) a) $\frac{8}{221}$ b) $\frac{13}{102}$ c) $\frac{12}{51}$ d) $\frac{25}{102}$ e) $\frac{13}{34}$

8) a) $\frac{6}{5525}$ b) $\frac{997}{1700}$ c) $\frac{22}{425}$ d) $\frac{16}{5525}$

9) a) 1/2 b) 1/2 c)7/10

10) a) 4/25 b) 8/75 c) 7/25 d) 1/15

11) a) 2/7 12) i) 25/81 ii) 5/18

13) i) 1/6 ii) $\frac{5}{6}$