11

STATISTICS

It is easy to lie with statistics. It is hard to tell the truth without it
-Andrejs Dunkels

- Introduction
- Measures of Dispersion
  - Range
  - Variance
  - Standard Deviation
- Coefficient of Variation

Karl Pearson
(1857-1936)
England

Karl Pearson, British statistician, is a leading founder of modern field of statistics. He established the discipline of mathematical statistics. He introduced moments, a concept borrowed from physics.

His book, 'The Grammar of Science' covered several themes that were later to become part of the theories of Einstein and other scientists.

11.1 Introduction

According to Croxton and Cowden, Statistics is defined as the collection, presentation, analysis and interpretation of numerical data. R.A. Fisher said that the science of statistics is essentially a branch of Mathematics and may be regarded as mathematics applied to observational data. Horace Secrist defined statistics as follows:

"By statistics we mean aggregates of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a pre-determined purpose and placed in relation to each other".

The word 'Statistics' is known to have been used for the first time in " Elements of Universal Erudiation" by J.F. Baron. In modern times, statistics is no longer merely the collection of data and their presentation in charts and tables - it is now considered to encompass the science of basing inferences on observed data and the entire problem of making decisions in the face of uncertainity.

We have already learnt about the measures of central tendency namely, Mean, Median and Mode. They give us an idea of the concentration of the observation (data) about the central part of the distribution.

The knowledge of measures of central tendency cannot give a complete idea about the distribution. For example, consider the following two different series (i) 82, 74, 89, 95 and (ii) 120, 62, 28, 130. The two distributions have the same Mean 85. In the former, the numbers are closer to the mean 85 where as in the second series, the numbers are widely scattered about the Mean 85. Thus the measures of central tendency may mislead us. We need to have a measure which tells us how the items are dispersed around the Mean.

11.2 Measures of dispersion

Measures of dispersion give an idea about the scatteredness of the data of the distribution. Range (R), Quartile Deviation (Q.D), Mean Deviation (M.D) and Standard Deviation (S.D) are the measures of dispersion. Let us study about some of them in detail.

11.2.1 Range

Range is the simplest measure of dispersion. Range of a set of numbers is the difference between the largest and the smallest items of the set.

$\therefore$ Range = Largest value − Smallest value

$= L - S.$

The coefficient of range is given by $\dfrac{L-S}{L+S}$

Example 11.1

Find the range and the coefficient of range of 43, 24, 38, 56, 22, 39, 45.

Solution Let us arrange the given data in the ascending order.

22, 24, 38, 39, 43, 45, 56.

From the given data the largest value, $L = 56$ and the smallest value, $S = 22$.

$$\therefore \quad \text{Range} = L - S$$

$$= 56 - 22 = 34$$

$$\text{Now the coefficient of range} = \frac{L-S}{L+S}$$

$$= \frac{56-22}{56+22} = \frac{34}{78} = 0.436$$

Example 11.2

The weight (in kg) of 13 students in a class are 42.5, 47.5, 48.6, 50.5, 49, 46.2, 49.8, 45.8, 43.2, 48, 44.7, 46.9, 42.4. Find the range and coefficient of range.

Solution Let us arrange the given data in the ascending order.

42.4, 42.5, 43.2, 44.7, 45.8, 46.2, 46.9, 47.5, 48, 48.6, 49, 49.8, 50.5

From the given data, the largest value $L = 50.5$ and the smallest value $S = 42.4$

$$\text{Range} = L - S$$

$$= 50.5 - 42.4 = 8.1$$

$$\text{The coefficient of range} = \frac{L-S}{L+S} = \frac{50.5-42.4}{50.5+42.4} = \frac{8.1}{92.9}$$

$$= 0.087$$

Example 11.3

The largest value in a collection of data is 7.44. If the range is 2.26, then find the smallest value in the collection.

Solution Range = largest value – smallest value

$\Longrightarrow$ 7.44 – smallest value = 2.26

$\therefore$ The smallest value = 7.44 – 2.26 = 5.18

11.2.2 Standard deviation

A better way to measure dispersion is to square the differences between each data and the mean before averaging them. This measure of dispersion is known as the Variance and the positive square root of the Variance is known as the Standard Deviation. The variance is always positive.

The term 'standard deviation' was first used by Karl Pearson in 1894 as a replacement of the term 'mean error' used by Gauss.

Standard deviation is expressed in the same units as the data. It shows how much variation is there from the mean. A low standard deviation indicates that the data points tend to be very close to the mean, where as a high standard deviation indicates that the data is spread out over a large range of values.

We use $\overline{x}$ and σ to denote the mean and the standard deviation of a distribution respectively. Depending on the nature of data, we shall calculate the standard deviation σ (after arranging the given data either in ascending or descending order) by different methods using the following formulae (proofs are not given).

Data	Direct method	Actual mean method	Assumed mean method	Step deviation method
Ungrouped	$\sqrt{\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2}$	$\sqrt{\frac{\sum d^2}{n}}$ $d = x - \overline{x}$	$\sqrt{\frac{\sum d^2}{n}-\left(\frac{\sum d}{n}\right)^2}$ $d = x - A$	$\sqrt{\frac{\sum d^2}{n}-\left(\frac{\sum d}{n}\right)^2}\times c$ $d = \frac{x-A}{c}$
Grouped		$\sqrt{\frac{\sum fd^2}{\sum f}}$	$\sqrt{\frac{\sum fd^2}{\sum f}-\left(\frac{\sum fd}{\sum f}\right)^2}$	$\sqrt{\frac{\sum fd^2}{\sum f}-\left(\frac{\sum fd}{\sum f}\right)^2}\times c$

Note

For a collection of n items (numbers), we always have $\sum(x-\overline{x}) = 0$, $\sum x = n\overline{x}$ and $\sum \overline{x} = n\overline{x}$.

(i) Direct method

This method can be used, when the squares of the items are easily obtained.

Example 11.4

The number of books read by 8 students during a month are

2, 5, 8, 11, 14, 6, 12, 10. Calculate the standard deviation of the data.

Solution

x	x^2
2	4
5	25
6	36
8	64
10	100
11	121
12	144
14	196
$\sum x = 68$	$\sum x^2 = 690$

Here, the number of items, $n = 8$

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

$= \sqrt{\frac{690}{8} - \left(\frac{68}{8}\right)^2}$

$= \sqrt{86.25 - (8.5)^2}$

$= \sqrt{86.25 - 72.25}$

$= \sqrt{14} \simeq 3.74$

(ii) Actual mean method

This method can be used when the mean is not a fraction.

The standard deviation, $\sigma = \sqrt{\frac{\sum (x - \overline{x})^2}{n}}$ or $\sigma = \sqrt{\frac{\sum d^2}{n}}$, where $d = x - \overline{x}$.

Example 11.5

A test in General Knowledge was conducted for a class. The marks out of 40, obtained by 6 students were 20, 14, 16, 30, 21 and 25. Find the standard deviation of the data.

Solution Now, A. M. $= \frac{\sum x}{n} = \frac{20 + 14 + 16 + 30 + 21 + 25}{6}$

$\Longrightarrow \overline{x} = \frac{126}{6} = 21.$

Let us form the following table.

x	$d = x - \overline{x}$	d^2
14	-7	49
16	-5	25
20	-1	1
21	0	0
25	4	16
30	9	81
$\sum x = 126$	$\sum d = 0$	$\sum d^2 = 172$

$\sigma = \sqrt{\frac{\sum d^2}{n}} = \sqrt{\frac{172}{6}}$

$= \sqrt{28.67}$

Thus, $\sigma \simeq 5.36$

(iii) Assumed mean method

When the mean of the given data is not an integer, we use assumed mean method to calculate the standard deviation. We choose a suitable item A such that the difference $x–A$ are all small numbers possibly, integers. Here A is an assumed mean which is supposed to be closer to the mean.

We calculate the deviations using $d = x - A$.

Now the standard deviation, $\sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$.

> *Note*
>
> Assumed mean method and step deviation method are just simplified forms of direct method.

Example: 11.6

Find the standard deviation of the numbers 62, 58, 53, 50, 63, 52, 55.

Solution Let us take $A = 55$ as the assumed mean and form the following table.

x	$d = x - A$ $= x - 55$	d^2
50	-5	25
52	-3	9
53	-2	4
55	0	0
58	3	9
62	7	49
63	8	64
	$\sum d = 8$	$\sum d^2 = 160$

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

$$= \sqrt{\frac{160}{7} - \left(\frac{8}{7}\right)^2}$$

$$= \sqrt{\frac{160}{7} - \frac{64}{49}}$$

$$= \sqrt{\frac{1056}{49}}$$

$$= \frac{32.49}{7}$$

$\therefore$ Standard deviation $\sigma \simeq 4.64$

(iv) Step deviation method

This method can be used to find the standard deviation when the items are larger in size and have a common factor. We choose an assumed mean A and calculate d *by* using $d = \frac{x - A}{c}$ where c is the common factor of the values of $x–A$.

We use the formula, $\sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2} \times c$.

Example 11.7

The marks obtained by 10 students in a test in Mathematics are :

80, 70, 40, 50, 90, 60, 100, 60, 30, 80. Find the standard deviation.

Solution We observe that all the data have 10 as common factor. Take A = 70 as assumed mean. Here the number of items, n = 10.

Take $c = 10$, $d = \frac{x - A}{10}$ and form the following table.

x	$d = \frac{x-70}{10}$	d^2
30	-4	16
40	-3	9
50	-2	4
60	-1	1
60	-1	1
70	0	0
80	1	1
80	1	1
90	2	4
100	3	9
	$\sum d = -4$	$\sum d^2 = 46$

Now $\sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2} \times c$

$= \sqrt{\frac{46}{10} - \left(\frac{-4}{10}\right)^2} \times 10$

$= \sqrt{\frac{46}{10} - \frac{16}{100}} \times 10 = \sqrt{\frac{460 - 16}{100}} \times 10$

$\therefore$ Standard deviation, $\sigma \simeq 21.07$

The standard deviation for a collection of data can be obtained in any of the four methods, namely direct method, actual mean method, assumed mean method and step deviation method.

As expected, the different methods should not give different answers for σ for the same data. Students are advised to follow any one of the above methods.

Results

(i) The standard deviation of a distribution remains unchanged when each value is added or subtracted by the same quantity.

(ii) If each value of a collection of data is multiplied or divided by a non-zero constant k, then the standard deviation of the new data is obtained by multiplying or dividing the standard deviation by the same quantity k.

Example: 11.8

Find the standard deviation of the data 3, 5, 6, 7. Then add 4 to each item and find the standard deviation of the new data.

Solution Given data 3, 5, 6, 7

Take $A = 6$

x	$d = x - 6$	d^2
3	– 3	9
5	–1	1
6	0	0
7	1	1
	$\sum d = -3$	$\sum d^2 = 11$

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

$$= \sqrt{\frac{11}{4} - \left(\frac{-3}{4}\right)^2}$$

$$\sigma = \sqrt{\frac{11}{4} - \frac{9}{16}} = \frac{\sqrt{35}}{4}$$

Let us add 4 to each term of the given data to get the new data 7, 9, 10, 11

Take $A = 10$

x	$d = x - 10$	d^2
7	– 3	9
9	–1	1
10	0	0
11	1	1
	$\sum d = -3$	$\sum d^2 = 11$

$$\text{Standard deviation}, \sigma_1 = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$$

$$= \sqrt{\frac{11}{4} - \left(\frac{-3}{4}\right)^2}$$

$$\sigma_1 = \sqrt{\frac{11}{4} - \frac{9}{16}} = \frac{\sqrt{35}}{4}$$

In the above example, the standard deviation remains unchanged even when each item is added by the constant 4.

Example 11.9

Find the standard deviation of 40, 42 and 48. If each value is multiplied by 3, find the standard deviation of the new data.

Solution Let us consider the given data 40, 42, 48 and find σ.

Let the assumed mean A be 44

x	$d = x - 44$	d^2
40	– 4	16
42	– 2	4
48	4	16
	$\sum d = -2$	$\sum d^2 = 36$

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

$$= \sqrt{\frac{36}{3} - \left(\frac{-2}{3}\right)^2}$$

$$\sigma = \frac{\sqrt{104}}{3}$$

When the values are multiplied by 3, we get 120,126,144. Let the assumed mean A be 132.

Let σ_1 be the S.D. of the new data.

x	$d = x - 132$	d^2
120	–12	144
126	– 6	36
144	12	144
	$\sum d = -6$	$\sum d^2 = 324$

$$\text{Standard deviation}, \sigma_1 = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$$

$$= \sqrt{\frac{324}{3} - \left(\frac{-6}{3}\right)^2}$$

$$\sigma_1 = \sqrt{\frac{312}{3}} = \sqrt{104}$$

In the above example, when each value is multiplied by 3, the standard deviation also gets multiplied by 3 .

Example 11.10

Prove that the standard deviation of the first n natural numbers is $\sigma = \sqrt{\frac{n^2-1}{12}}$.

Solution The first n natural numbers are 1,2,3, $\cdots$, n.

$$\text{Their mean,} \quad \bar{x} = \frac{\Sigma x}{n} = \frac{1+2+3+\cdots+n}{n}$$

$$= \frac{n(n+1)}{2n} = \frac{n+1}{2}.$$

Sum of the squares of the first n natural numbers is

$$\sum x^2 = \frac{n(n+1)(2n+1)}{6}.$$

$$\text{Thus, the standard deviation } \sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$= \sqrt{\frac{n(n+1)(2n+1)}{6n} - \left(\frac{n+1}{2}\right)^2}$$

$$= \sqrt{\frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2}$$

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

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

$$= \sqrt{\left(\frac{n+1}{2}\right)\left(\frac{4n+2-3n-3}{6}\right)}$$

$$= \sqrt{\left(\frac{n+1}{2}\right)\left(\frac{n-1}{6}\right)}$$

$$= \sqrt{\frac{n^2-1}{12}}.$$

Hence, the S.D. of the first n natural numbers is $\sigma = \sqrt{\frac{n^2-1}{12}}$.

Remarks

It is quite interesting to note the following:

The S.D. of any n successive terms of an A.P. with common difference d is, $\sigma = d\sqrt{\frac{n^2-1}{12}}$.

Thus,

(i) S.D. of $i, i+1, i+2, \cdots, i+n$ is $\sigma = \sqrt{\frac{n^2-1}{12}}, i \in \mathbb{N}$.

(ii) S.D. of any n consecutive even integers, is given by $\sigma = 2\sqrt{\frac{n^2-1}{12}}, n \in \mathbb{N}$.

(iii) S.D. of any n consecutive odd integers, is given by $\sigma = 2\sqrt{\frac{n^2-1}{12}}, n \in \mathbb{N}$.

Example 11.11

Find the standard deviation of the first 10 natural numbers.

Solution Standard deviation of the first n natural numbers = $\sqrt{\frac{n^2-1}{12}}$

Standard deviation of the first 10 natural numbers

$$= \sqrt{\frac{10^2-1}{12}} = \sqrt{\frac{100-1}{12}} \simeq 2.87$$

Standard Deviation of grouped data

(i) Actual mean method

In a discrete data, when the deviations are taken from arithmetic mean, the standard deviation can be calculated using the formula $\sigma = \sqrt{\frac{\sum fd^2}{\sum f}}$, where $d = x - \bar{x}$.

Example 11.12

The following table shows the marks obtained by 48 students in a Quiz competition in Mathematics. Calculate the standard deviation.

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

Solution Let us form the following table using the given data.

x	f	fx	$d = x - \bar{x}$ $= x - 9$	fd	fd^2
6	3	18	-3	-9	27
7	6	42	-2	-12	24
8	9	72	-1	-9	9
9	13	117	0	0	0
10	8	80	1	8	8
11	5	55	2	10	20
12	4	48	3	12	36
	$\sum f=48$	$\sum fx=432$	$\sum d = 0$	$\sum fd = 0$	$\sum fd^2 = 124$

Arithmetic mean, $\bar{x} = \frac{\sum fx}{\sum f} = \frac{432}{48} = 9.$

Standard deviation, $\sigma = \sqrt{\frac{\sum fd^2}{\sum f}}$

$$= \sqrt{\frac{124}{48}}$$

$$\sigma = \sqrt{2.58} \simeq 1.61$$

(ii) Assumed mean method

When deviations are taken from the assumed mean, the formula for calculating standard deviation is

$$\sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2}, \text{ where } d = x - A.$$

Example 11.13

Find the standard deviation of the following distribution.

x	70	74	78	82	86	90
f	1	3	5	7	8	12

Solution Let us take the assumed mean $A = 82$.

x	f	$d = x - 82$	fd	fd^2
70	1	-12	-12	144
74	3	-8	-24	192
78	5	-4	-20	80
82	7	0	0	0
86	8	4	32	128
90	12	8	96	768
	$\sum f = 36$		$\sum fd = 72$	$\sum fd^2 = 1312$

Standard deviation $\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$

$= \sqrt{\frac{1312}{36} - \left(\frac{72}{36}\right)^2}$

$= \sqrt{\frac{328}{9} - 2^2}$

$= \sqrt{\frac{328 - 36}{9}}$

$= \sqrt{\frac{292}{9}} = \sqrt{32.44}$

$\therefore \quad \sigma \simeq 5.7$

Example 11.14

Find the variance of the following distribution.

Class interval	3.5-4.5	4.5-5.5	5.5-6.5	6.5-7.5	7.5-8.5
Frequency	9	14	22	11	17

Solution Let the assumed mean A be 6.

class interval	x mid value	f	$d = x - 6$	fd	fd^2
3.5-4.5	4	9	-2	-18	36
4.5-5.5	5	14	-1	-14	14
5.5-6.5	6	22	0	0	0
6.5-7.5	7	11	1	11	11
7.5-8.5	8	17	2	34	68
		$\sum f = 73$		$\sum fd = 13$	$\sum fd^2 = 129$

$$\text{Now variance, } \sigma^2 = \frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2$$

$$= \frac{129}{73} - \left(\frac{13}{73}\right)^2 = \frac{129}{73} - \frac{169}{5329}$$

$$= \frac{9417 - 169}{5329} = \frac{9248}{5329}$$

Thus, the variance is $\sigma^2 \simeq 1.74$

(iii) Step deviation method

Example 11.15

The following table gives the number of goals scored by 71 leading players in International Football matches. Find the standard deviation of the data.

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	8	12	17	14	9	7	4

Solution Let $A = 35$. In the 4th column, the common factor of all items, $c = 10$.

class interval	x mid value	f	$x-A$	$d = \frac{x-A}{c}$	fd	fd^2
0–10	5	8	–30	-3	-24	72
10–20	15	12	–20	-2	-24	48
20–30	25	17	–10	-1	-17	17
30–40	35	14	0	0	0	0
40–50	45	9	10	1	9	9
50–60	55	7	20	2	14	28
60–70	65	4	30	3	12	36
		$\sum f = 71$			$\sum fd = -30$	$\sum fd^2 = 210$

Standard deviation, $\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2} \times c$

$= \sqrt{\frac{210}{71} - \left(\frac{-30}{71}\right)^2} \times 10$

$= \sqrt{\frac{210}{71} - \frac{900}{5041}} \times 10$

$= \sqrt{\frac{14910 - 900}{5041}} \times 10$

$= \sqrt{\frac{14010}{5041}} \times 10 = \sqrt{2.7792} \times 10$

Standard deviation, $\sigma \simeq 16.67$

Example 11.16

Length of 40 bits of wire, correct to the nearest centimetre are given below. Calculate the variance.

Length cm	1-10	11-20	21-30	31-40	41-50	51-60	61-70
No. of bits	2	3	8	12	9	5	1

Solution Let the assumed mean A be 35.5

Length	mid value x	no. of bits (f)	$d = x - A$	fd	fd^2
1-10	5.5	2	−30	−60	1800
11-20	15.5	3	−20	−60	1200
21-30	25.5	8	−10	−80	800
31-40	35.5	12	0	0	0
41-50	45.5	9	10	90	900
51-60	55.5	5	20	100	2000
61-70	65.5	1	30	30	900
		$\sum f = 40$		$\sum fd = 20$	$\sum fd^2 = 7600$

Variance, $\sigma^2 = \frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2 = \frac{7600}{40} - \left(\frac{20}{40}\right)^2$

$= 190 - \frac{1}{4} = \frac{760 - 1}{4} = \frac{759}{4}$

$\therefore \quad \sigma^2 = 189.75$

11.2.3 Coefficient of variation

Coefficient of variation is defined as

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

where σ is the standard deviation and $\overline{x}$ is the mean of the given data. It is also called as a relative standard deviation.

Remarks

(i) The coefficient of variation helps us to compare the consistency of two or more collections of data.

(ii) When the coefficient of variation is more, the given data is less consistent.

(iii) When the coefficient of variation is less, the given data is more consistent.

Example 11.17

Find the coefficient of variation of the following data. 18, 20, 15, 12, 25.

Solution Let us calculate the A.M of the given data.

$$\text{A.M } \bar{x} = \frac{12 + 15 + 18 + 20 + 25}{5}$$

$$= \frac{90}{5} = 18.$$

x	$d = x - 18$	d^2
12	-6	36
15	-3	9
18	0	0
20	2	4
25	7	49
	$\sum d = 0$	$\sum d^2 = 98$

$$\sigma = \sqrt{\frac{\sum d^2}{n}} = \sqrt{\frac{98}{5}}$$

$$= \sqrt{19.6} \simeq 4.428$$

$\therefore$ The coefficient of variation $= \frac{\sigma}{\bar{x}} \times 100$

$$= \frac{4.428}{18} \times 100 = \frac{442.8}{18}.$$

$\therefore$ The coefficient of variation is 24.6

Example 11.18

Following are the runs scored by two batsmen in 5 cricket matches. Who is more consistent in scoring runs.

Batsman A	38	47	34	18	33
Batsman B	37	35	41	27	35

Solution

Batsman A

x	$d = x - \bar{x}$	d^2
18	-16	256
33	-1	1
$\boxed{34}$	0	0
38	4	16
47	13	169
170	0	442

Now $\bar{x} = \frac{170}{5} = 34$

$\sigma = \sqrt{\frac{\sum d^2}{n}}$

$= \sqrt{\frac{442}{5}} = \sqrt{88.4}$

$\simeq 9.4$

Coefficient of variation, C.V $= \frac{\sigma}{\bar{x}} \times 100$

$= \frac{9.4}{34} \times 100$

$= \frac{940}{34}$

$= 27.65$

$\therefore$ The coefficient of variation for the runs scored by batsman A is 27.65 (1)

Batsman B

x	$d = x - \bar{x}$	d^2
27	-8	64
35	0	0
$\boxed{35}$	0	0
37	2	4
41	6	36
175	0	104

$\bar{x} = \frac{175}{5} = 35$

$\sigma = \sqrt{\frac{\sum d^2}{n}}$

$= \sqrt{\frac{104}{5}} = \sqrt{20.8}$

$\simeq 4.6$

Coefficient of variation

$= \frac{\sigma}{\bar{x}} \times 100$

$= \frac{4.6}{35} \times 100$

$= \frac{460}{35} = \frac{92}{7} = 13.14$

$\therefore$ The coefficient of variation for the runs scored by batsman B is = 13.14 (2)

From (1) and (2), the coefficient of variation for B is less than the coefficient of variation for A.

$\therefore$ Batsman B is more consistent than the batsman A in scoring the runs.

Example 11.19

The mean of 30 items is 18 and their standard deviation is 3. Find the sum of all the items and also the sum of the squares of all the items.

Solution The mean of 30 items, $\bar{x} = 18$

The sum of 30 items, $\sum x = 30 \times 18 = 540$ $\left(\bar{x} = \frac{\sum x}{n} \right)$

Standard deviation, $\sigma = 3$

Now, $\sigma^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$

$$\Rightarrow \quad \frac{\sum x^2}{30} - 18^2 = 9$$

$$\Rightarrow \quad \frac{\sum x^2}{30} - 324 = 9$$

$$\Rightarrow \quad \sum x^2 - 9720 = 270$$

$$\sum x^2 = 9990$$

$$\therefore \quad \sum x = 540 \text{ and } \sum x^2 = 9990.$$

Example 11.20

The mean and the standard deviation of a group of 20 items was found to be 40 and 15 respectively. While checking it was found that an item 43 was wrongly written as 53. Calculate the correct mean and standard deviation.

Solution Let us find the correct mean.

$$\text{Mean of 20 items, } \bar{x} = \frac{\sum x}{n} = 40$$

$$\Rightarrow \quad \frac{\sum x}{20} = 40$$

$$\Rightarrow \quad \sum x = 20 \times 40 = 800$$

$$\text{corrected } \sum x = 800 - (\text{wrong value}) + (\text{correct value})$$

$$\text{Now, corrected } \sum x = 800 - 53 + 43 = 790.$$

$$\therefore \quad \text{The corrected Mean} = \frac{790}{20} = 39.5$$

$$\text{Variance, } \sigma^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = 225 \qquad \text{(given)}$$

$$\Rightarrow \quad \frac{\sum x^2}{20} - 40^2 = 225$$

$$\Rightarrow \quad \sum x^2 - 32000 = 225 \times 20 = 4500$$

$$\therefore \quad \sum x^2 = 32000 + 4500 = 36500$$

$$\text{corrected } \sum x^2 = 36500 - (\text{wrong value})^2 + (\text{correct value})^2$$

$$\text{corrected } \sum x^2 = 36500 - 53^2 + 43^2 = 36500 - 2809 + 1849$$

$$= 36500 - 960 = 35540.$$

$$\text{Now, the corrected } \sigma^2 = \frac{\text{Corrected } \sum x^2}{n} - (\text{Corrected mean})^2$$

$$= \frac{35540}{20} - (39.5)^2$$

$$= 1777 - 1560.25 = 216.75$$

$$\text{Corrected } \sigma = \sqrt{216.75} \simeq 14.72$$

$\therefore$ The corrected Mean = 39.5 and the corrected S.D. $\simeq$ 14.72

Example 11.21

For a collection of data, if $\sum x = 35$, $n = 5$, $\sum (x-9)^2 = 82$, then find $\sum x^2$ and $\sum (x-\overline{x})^2$.

Solution Given that $\sum x = 35$ and $n = 5$.

$$\therefore \quad \overline{x} = \frac{\sum x}{n} = \frac{35}{5} = 7.$$

Let us find $\sum x^2$

Now, $\sum (x-9)^2 = 82$

$$\Longrightarrow \quad \sum (x^2 - 18x + 81) = 82$$

$$\Longrightarrow \quad \sum x^2 - \left(18\sum x\right) + \left(81\sum 1\right) = 82$$

$$\Longrightarrow \quad \sum x^2 - 630 + 405 = 82 \qquad \because \sum x = 35 \text{ and } \sum 1 = 5$$

$$\Longrightarrow \quad \sum x^2 = 307.$$

To find $\sum (x-\overline{x})^2$, let us consider

$$\sum (x-9)^2 = 82$$

$$\Longrightarrow \quad \sum (x-7-2)^2 = 82$$

$$\Longrightarrow \quad \sum [(x-7)-2]^2 = 82$$

$$\Longrightarrow \quad \sum (x-7)^2 - 2\sum [(x-7)\times 2] + \sum 4 = 82$$

$$\Longrightarrow \quad \sum (x-\overline{x})^2 - 4\sum (x-\overline{x}) + 4\sum 1 = 82$$

$$\Longrightarrow \quad \sum (x-\overline{x})^2 - 4(0) + (4\times 5) = 82 \qquad \because \sum 1 = 5 \text{ and } \sum (x-\overline{x}) = 0$$

$$\Longrightarrow \quad \sum (x-\overline{x})^2 = 62$$

$$\therefore \quad \sum x^2 = 307 \text{ and } \sum (x-\overline{x})^2 = 62.$$

Example 11.22

The coefficient of variations of two series are 58 and 69. Their standard deviations are 21.2 and 15.6. What are their arithmetic means?

Solution We know that coefficient of variation, $C.V = \frac{\sigma}{\overline{x}} \times 100$.

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

Mean of the first series, $\overline{x}_1 = \frac{\sigma}{C.V} \times 100$

$$= \frac{21.2}{58} \times 100 \qquad \because \quad C.V = 58 \text{ and } \sigma = 21.2$$

$$= \frac{2120}{58} = 36.6$$

Mean of the second series, $\overline{x_2} = \frac{\sigma}{C.V} \times 100$

$= \frac{15.6}{69} \times 100 \qquad \because \quad C.V = 69 \text{ and } \sigma = 15.6$

$= \frac{1560}{69}$

$= 22.6$

A.M of the I series = 36.6 and the A.M of the II series = 22.6

Exercise 11.1

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

   (i) 59, 46, 30, 23, 27, 40, 52,35, 29

   (ii) 41.2, 33.7, 29.1, 34.5, 25.7, 24.8, 56.5, 12.5

2. The smallest value of a collection of data is 12 and the range is 59. Find the largest value of the collection of data.

3. The largest of 50 measurements is 3.84kg. If the range is 0.46kg, find the smallest measurement.

4. The standard deviation of 20 observations is $\sqrt{5}$. If each observation is multiplied by 2, find the standard deviation and variance of the resulting observations.

5. Calculate the standard deviation of the first 13 natural numbers.

6. Calculate the standard deviation of the following data.

   (i) 10, 20, 15, 8, 3, 4 (ii) 38, 40, 34 ,31, 28, 26, 34

7. Calculate the standard deviation of the following data.

x	3	8	13	18	23
f	7	10	15	10	8

8. The number of books bought at a book fair by 200 students from a school are given in the following table.

No. of books	0	1	2	3	4
No of students	35	64	68	18	15

   Calculate the standard deviation.

9. Calculate the variance of the following data

x	2	4	6	8	10	12	14	16
f	4	4	5	15	8	5	4	5

10. The time (in seconds) taken by a group of people to walk across a pedestrian crossing is given in the table below.

Time (in sec.)	5-10	10-15	15-20	20-25	25-30
No. of people	4	8	15	12	11

Calculate the variance and standard deviation of the data.

11. A group of 45 house owners contributed money towards green environment of their street. The amount of money collected is shown in the table below.

Amount (₹)	0-20	20-40	40-60	60-80	80-100
No. of house owners	2	7	12	19	5

Calculate the variance and standard deviation.

12. Find the variance of the following distribution

Class interval	20-24	25-29	30-34	35-39	40-44	45-49
Frequency	15	25	28	12	12	8

13. Mean of 100 items is 48 and their standard deviation is 10. Find the sum of all the items and the sum of the squares of all the items.

14. The mean and standard deviation of 20 items are found to be 10 and 2 respectively. At the time of checking it was found that an item 12 was wrongly entered as 8. Calculate the correct mean and standard deviation.

15. If $n = 10$, $\overline{x} = 12$ and $\sum x^2 = 1530$, then calculate the coefficient of variation .

16. Calculate the coefficient of variation of the following data: 20, 18, 32, 24, 26.

17. If the coefficient of variation of a collection of data is 57 and its S.D is 6.84, then find the mean.

18. A group of 100 candidates have their average height 163.8 cm with coefficient of variation 3.2. What is the standard deviation of their heights?

19. Given $\sum x = 99$, $n = 9$ and $\sum (x - 10)^2 = 79$. Find $\sum x^2$ and $\sum (x - \overline{x})^2$.

20. The marks scored by two students A, B in a class are given below.

A	58	51	60	65	66
B	56	87	88	46	43

Who is more consistent?

Exercise 11.2

Choose the correct answer.

1. The range of the first 10 prime numbers 2, 3, 5, 7, 11, 13, 17, 19, 23 , 29 is

   (A) 28 (B) 26 (C) 29 (D) 27

2. The least value in a collection of data is 14.1. If the range of the collection is 28.4, then the greatest value of the collection is

   (A) 42.5 (B) 43.5 (C) 42.4 (D) 42.1

3. The greatest value of a collection of data is 72 and the least value is 28. Then the coefficient of range is

   (A) 44 (B) 0.72 (C) 0.44 (D) 0.28

4 For a collection of 11 items, $\Sigma x = 132$, then the arithmetic mean is

   (A) 11 (B) 12 (C) 14 (D) 13

5. For any collection of n items, $\Sigma(x - \overline{x}) =$

   (A) Σx (B) $\overline{x}$ (C) $n\overline{x}$ (D) 0

6. For any collection of n items, $(\Sigma x) - \overline{x} =$

   (A) $n\overline{x}$ (B) $(n-2)\overline{x}$ (C) $(n-1)\overline{x}$ (D) 0

7. If t is the standard deviation of $x, y. z$, then the standard deviation of $x+5,\ y+5,\ z+5$ is

   (A) $\frac{t}{3}$ (B) $t+5$ (C) t (D) $x\,y\,z$

8. If the standard deviation of a set of data is 1.6, then the variance is

   (A) 0.4 (B) 2.56 (C) 1.96 (D) 0.04

9. If the variance of a data is 12.25, then the S.D is

   (A) 3.5 (B) 3 (C) 2.5 (D) 3.25

10. Variance of the first 11 natural numbers is

    (A) $\sqrt{5}$ (B) $\sqrt{10}$ (C) $5\sqrt{2}$ (D) 10

11. The variance of 10, 10, 10, 10, 10 is

    (A) 10 (B) $\sqrt{10}$ (C) 5 (D) 0

12. If the variance of 14, 18, 22, 26, 30 is 32, then the variance of 28, 36,44,52,60 is

    (A) 64 (B) 128 (C) $32\sqrt{2}$ (D) 32

13. Standard deviation of a collection of data is $2\sqrt{2}$. If each value is multiplied by 3, then the standard deviation of the new data is

(A) $\sqrt{12}$ (B) $4\sqrt{2}$ (C) $6\sqrt{2}$ (D) $9\sqrt{2}$

14. Given $\sum (x-\overline{x})^2 = 48$, $\overline{x} = 20$ and $n = 12$. The coefficient of variation is

(A) 25 (B) 20 (C) 30 (D) 10

15. Mean and standard deviation of a data are 48 and 12 respectively. The coefficient of variation is

(A) 42 (B) 25 (C) 28 (D) 48

Points to Remember

- (i) Range $= L-S$, the difference between the greatest and the least of the observations.

  (ii) Coefficient of range $= \dfrac{L-S}{L+S}$.

- Standard deviation for an ungrouped data

  (i) $\sigma = \sqrt{\dfrac{\sum d^2}{n}}$, where $d = x - \overline{x}$ and $\overline{x}$ is the mean.

  (ii) $\sigma = \sqrt{\dfrac{\sum d^2}{n} - \left(\dfrac{\sum d}{n}\right)^2}$, where $d = x - A$ and A is the assumed mean.

- Standard deviation for a grouped data

  (i) $\sigma = \sqrt{\dfrac{\sum fd^2}{\sum f}}$, where $d = x - \overline{x}$ and $\overline{x}$ is the mean.

  (ii) $\sigma = \sqrt{\dfrac{\sum fd^2}{\sum f} - \left(\dfrac{\sum fd}{\sum f}\right)^2}$, where $d = x - A$ and A is the assumed mean.

- Standard deviation of a collection of data remains unchanged when each value is added or subtracted by a constant.

- Standard deviation of a collection of data gets multiplied or divided by the quantity k, if each item is multiplied or divided by k.

- Standard deviation of the first n natural numbers, $\sigma = \sqrt{\dfrac{n^2-1}{12}}$.

- Variance is the square of standard deviation.

- Coefficient of variation, C.V. $= \dfrac{\sigma}{\overline{x}} \times 100$. It is used for comparing the consistency of two or more collections of data.