15 Statistics

Lesson at a Glance

1. Arithmetic mean and median are some of the **measures of central tendency.**
2. **Formulae to compute arithmetic mean**

(i) A.M. of a series is given by $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

cross-multiplying $\sum_{i=1}^{n} x_i = n\,\bar{x}$

i.e., Sum of all observations = Number of observations × mean.

(ii) A.M. of a frequency distribution is given by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i \text{ where } N = \sum_{i=1}^{n} f_i .$$

(iii) Short cut method for finding mean of a frequency distribution is

$$\bar{x} = A + \frac{1}{N}\sum_{i=1}^{n} f_i d_i \quad \text{where } N = \sum_{i=1}^{n} f_i \text{ and } d_i = x_i - A$$

where A is assumed mean.

(iv) If the distribution is class frequency distribution, we take x_i as the mid-value of the class-interval and then use the formulae given in (ii) and (iii) above.

(v) If the class-intervals of a class frequency distribution are of equal size h say, then we form a column of values of $y_i = \frac{x_i - A}{h}$ and

use $\bar{x} = A + \frac{h}{N}\sum_{i=1}^{n} f_i y_i$ where $N = \sum_{i=1}^{n} f_i$.

3. Methods for finding the median

(*a*) **Median of ungrouped data (*i.e.*, series).** Arrange the data *in ascending* (or *descending*) *order* of magnitude.

Let n be the number of observations.

(*i*) If n **is odd,** then median is the value of the middle observation, *i.e.*, $\left(\frac{n+1}{2}\right)$th observation.

(*ii*) If n **is even,** then median is the **mean** of two middle observations, *i.e.*, $\left(\frac{n}{2}\right)$th and $\left(\frac{n}{2}+1\right)$th observations.

(*b*) **Median of discrete frequency distribution.** Arrange the data in *ascending order* of magnitude and then prepare a cumulative frequency column.

Let $N = \sum_{i=1}^{n} f_i$ be the total frequency.

(*i*) If N is odd, then median is the value of x for $\left(\frac{N+1}{2}\right)$th observation in the column of *c.f.*

(*ii*) If N is even, then median is the mean of values of x for $\left(\frac{N}{2}\right)$th and $\left(\frac{N}{2}+1\right)$th observations in the column of *c.f.*

(*c*) **Median of grouped frequency distribution.** If the classes are inclusive, (*i.e.*, of the type 10–19, 20–29, 30–39, etc.) convert them into **exclusive** classes (namely 9.5–19.5, 19.5–29.5, 29.5–39.5 etc.) and **prepare a cumulative frequency column.**

Let $N = \sum_{i=1}^{n} f_i$ be the total frequency.

Find the class in which $\left(\frac{N}{2}\right)$th observation lies, *i.e.*, **find the class which corresponds to c.f. just** $\geq \frac{N}{2}$. This class is called the **median class.** Then, we use the formula

$$\textbf{Median} = l + \frac{h}{f}\left(\frac{N}{2} - C\right)$$

where l = Lower limit of the median class.

h = Width of median class = (Upper limit – Lower limit) of median class.

f = Frequency of the median class.

$N = \sum_{i=1}^{n} f_i$ = Total frequency.

C = cumulative frequency of the class preceding the median class.

Note. The above formula is applicable even when the class-intervals are of unequal size.

4. Mean deviation, standard deviation, variance, range are some of the **measures of dispersion.**
5. **Mean deviation from the mean** for a series is $\frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|$.
6. **Mean deviation from the mean** for a frequency distribution is $\frac{1}{N}\sum_{i=1}^{n} f_i |x_i - \bar{x}|$ where $N = \sum_{i=1}^{n} f_i$.
7. For mean deviation from the mean of a class-frequency distribution, find the mid-values of the classes to obtain x_i and then use the formula given in result (6).

   Remark. To find mean for using in (5), (6) and (7), find mean $\bar{x}$ as explained in (2).
8. **Mean deviation from the median** for a **series** is $\frac{1}{n}\sum_{i=1}^{n} |x_i - M|$ where M is the median of the series.
9. **Mean deviation from the median** for a frequency distribution is

   $\frac{1}{N}\sum_{i=1}^{n} f_i |x_i - M|$ where $N = \sum_{i=1}^{n} f_i$ and M is median of frequency distribution.
10. For mean deviation from the median of a class-frequency distribution, find the mid-values of the classes to obtain x_i and use the formula given in result (9) above.

Remark. To find median for using in (8), (9) and (10); find median M as explained in (3).

11. **Variance** σ^2 for a series is defined as $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

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

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

12. **Standard deviation σ,** the positive square root of variance is called standard deviation.

13. Variance σ^2 for the frequency distribution is defined as

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

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

14. For finding variance of a class frequency distribution, find the mid-values of the classes to obtain x_i and then use the formula given in result (13) above.

15. **Short-cut methods for finding variance and standard deviation**

(*i*) **For an individual series**

If $d_i = x_i - A$, where A is assumed mean, then

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} d_i^2 - \bar{d}^2 \quad \text{where } \bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

(*ii*) **For a frequency distribution**

(*a*) If $d_i = x_i - A$ where A is assumed mean, then

$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i d_i^2 - \bar{d}^2 = = \frac{1}{N}\sum_{i=1}^{n} f_i d_i^2 - \left(\frac{\sum f_i d_i}{N}\right)^2$ because $\bar{d} = \frac{1}{N}\sum_{i=1}^{n} f_i d_i$.

Step deviation method

(*b*) If $y_i = \frac{x_i - A}{h}$ where h is the common difference of values of x, then

$$\sigma^2 = h^2 \left[\frac{1}{N} \sum_{i=1}^{n} f_i y_i^2 - \left(\frac{\sum_{i=1}^{n} f_i y_i}{N} \right)^2 \right]$$

(*iii*) **Grouped data (Class frequency distribution)**

Find the class-mark or mid-value of each class to obtain x_i. Now, use any of the two methods discussed in (*ii*) above.

16. **Coefficient of variation** of a distribution is denoted by C.V. and is defined as

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

TEXTBOOK QUESTIONS SOLVED

EXERCISE 15.1 (*Page No.: 360–61*)

Find the mean deviation about the mean for the data in Exercises 1 and 2.

1. **4, 7, 8, 9, 10, 12, 13, 17**

Sol. We make the following table from the given data:

x_i	$x_i - \bar{x} = x_i - 10$	$\lvert x_i - \bar{x} \rvert$
4	−6	6
7	−3	3
8	−2	2
9	−1	1
10	0	0
12	2	2
13	3	3
17	7	7
80		$\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert = 24$

Here $n = 8$, $\sum_{i=1}^{8} x_i = 80$.

$$\therefore \quad \bar{x} = \frac{1}{8}\sum_{i=1}^{8} x_i = \frac{1}{8}(80) = 10$$

We know that mean deviation about mean is given by

$$\text{M.D. } (\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}| = \frac{1}{8}\sum_{i=1}^{8} |x_i - \bar{x}|$$

$$= \frac{1}{8}(24) = 3.$$

2. 38, 70, 48, 40, 42, 55, 63, 46, 54, 44

Sol. Here n = Number of items = 10

$\bar{x}$ = Mean of the given items

$$= \frac{\sum x_i}{n}$$

$$= \frac{38+70+48+40+42+55+63+46+54+44}{10}$$

$$= \frac{500}{10} = 50$$

x_i	$x_i - \bar{x} = (x_i - 50)$	$\lvert x_i - \bar{x} \rvert$
38	– 12	12
70	20	20
48	– 2	2
40	– 10	10
42	– 8	8
55	5	5
63	13	13
46	– 4	4
54	4	4
44	– 6	6
$\Sigma x_i = 500$		$\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert = 84$

We know that M.D. from mean

$$= \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| = \frac{84}{10} = 8.4.$$

Find the mean deviation about the median for the data in Exercises 3 and 4.

3. 13, 17, 16, 14, 11, 13, 10, 16, 11, 18, 12, 17.

Sol. Arranging the data in ascending order, we have

10, 11, 11, 12, 13, 13, 14, 16, 16, 17, 17, 18

Number of observations is n = 12 which is even.

∴ Median is the mean of two middle terms, *i.e.*,

$$\frac{n}{2}\text{th} = 6\text{th and } \left(\frac{n}{2}+1\right)\text{th} = 7\text{th terms.}$$

$$\therefore \text{ Median M} = \frac{13+14}{2} = \frac{27}{2} = 13.5$$

Now, we prepare the following table:

x_i	$x_i - M = x_i - 13.5$	$\lvert x_i - M \rvert$
10	– 3.5	3.5
11	– 2.5	2.5
11	– 2.5	2.5
12	– 1.5	1.5
13	– 0.5	0.5
13	– 0.5	0.5
14	0.5	0.5
16	2.5	2.5
16	2.5	2.5
17	3.5	3.5
17	3.5	3.5
18	4.5	4.5
		28.0

Mean deviation about median is given by

$$\text{M.D. (M)} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \text{M}| = \frac{1}{12}\sum_{i=1}^{12}|x_i - \text{M}|$$

$$= \frac{1}{12}(28) = \frac{7}{3} = 2.33.$$

4. 36, 72, 46, 42, 60, 45, 53, 46, 51, 49.

Sol. Arranging the data in ascending order, we have

36, 42, 45, 46, 46, 49, 51, 53, 60, 72

Number of observations is n = 10 which is even.

∴ Median is the mean of two middle terms, *i.e.*,

$$\frac{n}{2}\text{th} = 5\text{th and } \left(\frac{n}{2}+1\right)\text{th} = 6\text{th terms}$$

$$\therefore \text{ Median M} = \frac{46+49}{2} = \frac{95}{2} = 47.5$$

Now, we prepare the following table:

x_i	$x_i - M = x_i - 47.5$	$\lvert x_i - M \rvert$
36	– 11.5	11.5
42	– 5.5	5.5
45	– 2.5	2.5
46	– 1.5	1.5
46	– 1.5	1.5
49	1.5	1.5
51	3.5	3.5
53	5.5	5.5
60	12.5	12.5
72	24.5	24.5
		70.0

Mean deviation about median is given by

$$\text{M.D. (M)} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \text{M}| = \frac{1}{10}\sum_{i=1}^{10} |x_i - \text{M}|$$

$$= \frac{1}{10}(70) = 7.$$

Find the mean deviation about the mean for the data in Exercises 5 and 6.

5.

x_i :	5	10	15	20	25
f_i :	7	4	6	3	5

Sol. We make the following table from the given data:

x_i	f_i	$f_i x_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
5	7	35	9	63
10	4	40	4	16
15	6	90	1	6
20	3	60	6	18
25	5	125	11	55
	25	350		158

Here, $N = \sum_{i=1}^{5} f_i = 25, \quad \sum_{i=1}^{5} f_i x_i = 350$

$\therefore \quad \bar{x} = \frac{1}{N} \sum_{i=1}^{5} f_i x_i = \frac{1}{25}(350) = 14$

and M.D. $(\bar{x}) = \frac{1}{N} \sum_{i=1}^{5} f_i \lvert x_i - \bar{x} \rvert = \frac{1}{25}(158)$

$= \frac{158 \times 4}{25 \times 4} = \frac{632}{100} = 6.32$ (From the table above)

6.

x_i :	10	30	50	70	90
f_i :	4	24	28	16	8

Sol. We make the following table from the given data:

x_i	f_i	$f_i x_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
10	4	40	40	160
30	24	720	20	480
50	28	1400	0	0
70	16	1120	20	320
90	8	720	40	320
	80	4000		1280

Here, $N = \sum_{i=1}^{5} f_i = 80, \quad \sum_{i=1}^{5} f_i x_i = 4000$

$\therefore \quad \bar{x} = \frac{1}{N} \sum_{i=1}^{5} f_i x_i$

$$= \frac{1}{80}(4000) = 50$$

and M.D. $(\bar{x}) = \frac{1}{N}\sum_{i=1}^{6} f_i |x_i - \bar{x}|$

$$= \frac{1}{80}(1280) \quad \text{(From the table above)}$$

$$= 16.$$

Find the mean deviation about the median for the data in Exercises 7 and 8.

7.

x_i :	5	7	9	10	12	15
f_i :	8	6	2	2	2	6

Sol. The given observations (*i.e.*, values of x) are already in ascending order. We make the following table from the given data:

x_i	f_i	c.f.	$\lvert x_i - M \rvert$	$f_i \lvert x_i - M \rvert$
5	8	8	2	16
7	6	14	0	0
9	2	16	2	4
10	2	18	3	6
12	2	20	5	10
15	6	26	8	48
	26			84

Here, $N = \sum_{i=1}^{6} f_i = 26$ which is even.

∴ Median is the mean of values of x for $\frac{N}{2} = \frac{26}{2}$th $= 13$th and the next 14th observations. Since values of x for all observations for c.f. from 9th to 14th are 7 each, therefore,

$$\text{Median } M = \frac{13\text{th observation} + 14\text{th observation}}{2}$$

$$= \frac{7+7}{2} = 7$$

From the table, $\sum_{i=1}^{6} f_i \mid x_i - M \mid = 84$

$\therefore$ M.D. (M) $= \frac{1}{N} \sum_{i=1}^{6} f_i \mid x_i - M \mid = \frac{1}{26}(84) = 3.23.$

8.

x_i :	15	21	27	30	35
f_i :	3	5	6	7	8

Sol. The given observations are already in ascending order. We make the following table from the given data:

x_i	f_i	c.f.	$\mid x_i - M \mid$	$f_i \mid x_i - M \mid$
15	3	3	15	45
21	5	8	9	45
27	6	14	3	18
30	7	21	0	0
35	8	29	5	40
	29			148

Here, $N = \sum_{i=1}^{5} f_i = 29$ which is odd.

$\therefore$ Median = value of x for $\left(\frac{29+1}{2}\right)$th observation = 15th observation in the column of c.f. = 30

(Since value of x for all observations from 15th to 21st is 30)

From the table, $\sum_{i=1}^{5} f_i \mid x_i - M \mid = 148$

$\therefore$ M.D. (M) $= \frac{1}{N} \sum_{i=1}^{5} f_i \mid x_i - M \mid = \frac{1}{29}(148) = 5.1.$

Find the mean deviation about the mean for the data in Exercises 9 and 10.

9.

Income (per day)	Number of persons
0-100	4
100-200	8
200-300	9
300-400	10
400-500	7
500-600	5
600-700	4
700-800	3

Sol. After writing the mid-values of class-intervals, and then taking the assumed mean a = 350 and h = 100, we form the following table:

Income per day	*Number of persons* f_i	*Mid-points* x_i	$u_i = \frac{x_i - 350}{100}$	$f_i u_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
0-100	4	50	– 3	– 12	308	1232
100-200	8	150	– 2	– 16	208	1664
200-300	9	250	– 1	– 9	108	972
300-400	10	350	0	0	8	80
400-500	7	450	1	7	92	644
500-600	5	550	2	10	192	960
600-700	4	650	3	12	292	1168
700-800	3	750	4	12	392	1176
	50			4		7896

Here, $N = \sum_{i=1}^{8} f_i = 50, \sum_{i=1}^{8} f_i u_i = 4$

$$\therefore \quad \bar{x} = a + \frac{\sum_{i=1}^{8} f_i u_i}{N} \times h = 350 + \frac{4}{50} \times 100 = 358$$

From the table, $\sum_{i=1}^{8} f_i |x_i - \bar{x}| = 7896$

$$\therefore \quad \text{M.D.}(\bar{x}) = \frac{1}{N}\sum_{i=1}^{8} f_i |x_i - \bar{x}| = \frac{1}{50}(7896) = \frac{7896\times 2}{50\times 2}$$

$$= \frac{15792}{100} = 157.92.$$

10.

Height (in cms)	Number of boys
95-105	9
105-115	13
115-125	26
125-135	30
135-145	12
145-155	10

Sol. After writing the mid-values of class-intervals and taking the assumed mean $a = 120$ and $h = 10$, we form the following table:

Height (in cms)	*Number of boys* (f_i)	*Mid-points* (x_i)	$y_i = \frac{x_i - 120}{10}$	$f_i y_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
95-105	9	100	– 2	– 18	25.3	227.7
105-115	13	110	– 1	– 13	15.3	198.9
115-125	26	120	0	0	5.3	137.8
125-135	30	130	1	30	4.7	141.0
135-145	12	140	2	24	14.7	176.4
145-155	10	150	3	30	24.7	247.0
	100			53		1128.8

Here, $N = \sum_{i=1}^{6} f_i = 100, \quad \sum_{i=1}^{6} f_i u_i = 53$

$$\therefore \quad \bar{x} = a + \frac{\sum_{i=1}^{6} f_i y_i}{N} \times h = 120 + \frac{53}{100} \times 10$$

$$= 120 + 5.3 = 125.3$$

From the table, $\sum_{i=1}^{6} f_i \,|\, x_i - \bar{x} \,| = 1128.8$

$$\therefore \quad \text{M.D. } (\bar{x}) = \frac{1}{N} \sum_{i=1}^{6} f_i \,|\, x_i - \bar{x} \,| = \frac{1}{100}(1128.8)$$

$$= 11.288 \text{ cm} \approx 11.29 \text{ cm.}$$

11. Find the mean deviation about median for the following data:

Marks	Number of girls
0-10	6
10-20	8
20-30	14
30-40	16
40-50	4
50-60	2

Sol. We form the following table from the given data:

Class	*Frequency* (f_i)	*Cumulative frequency* (*c.f.*)	*Mid-point* (x_i)	$\|x_i - M\|$	$f_i \|x_i - M\|$
0-10	6	6	5	22.86	137.16
10-20	8	14 = C	15	12.86	102.88
20-30	14 = f	28 ↑	25	2.86	40.04 ← Median
30-40	16	44	35	7.14	114.24 class
40-50	4	48	45	17.14	68.56
50-60	2	50	55	27.14	54.28
	50				517.16

Here, $N = \sum_{i=1}^{6} f_i = 50$

The class-interval containing $\frac{N}{2}$th *i.e.*, $\frac{50}{2}$th = 25th item *i.e.*, corresponding to c.f 28 just ≥ 25 is 20-30. Therefore, 20-30 is the median class.

Here, $l = 20, f = 14,$

$c = 14, h = 10$

and $N = 50$

$$\therefore \quad \text{Median } M = l + \frac{\frac{N}{2} - C}{f} \times h$$

$$= 20 + \frac{25 - 14}{14} \times 10$$

$$= 20 + \frac{55}{7}$$

$$= 20 + 7.86 = 27.86.$$

Thus, mean deviation about median is given by

$$\text{M.D. (M)} = \frac{1}{N} \sum_{i=1}^{6} f_i \, | x_i - M |$$

$$= \frac{1}{50}(517.16) = \frac{517.16 \times 2}{50 \times 2} = \frac{1034.32}{100}$$

$$= 10.3432 \approx 10.34 \text{ marks.}$$

12. Calculate the mean deviation about median age for the age distribution of 100 persons given below:

Age	Number
16-20	5
21-25	6
26-30	12
31-35	14
36-40	26
41-45	12
46-50	16
51-55	9

Sol. We shall have to find first the median of this grouped (class-interval) frequency distribution.

All the class-intervals here namely, 16-20, 21-25, etc., are discontinuous (inclusive), so firstly we shall have to make them continuous by subtracting 0.5 $\left(\frac{21-20}{2}=\frac{1}{2}=0.5\right)$ from lower limit and adding 0.5 to upper limit of each class-interval.

∴ New class-intervals are 15.5 – 20.5, 20.5 – 25.5, etc.

∴ We can write the given grouped (discontinuous class) frequency distribution in terms of continuous class frequency distribution so that we can find the median of the distribution:

Class-interval	*Frequency* (f_i)	*Cumulative frequency* (*c.f.*)
15.5-20.5	5	5
20.5-25.5	6	5+6=11
25.5-30.5	12	11+12=23
30.5-35.5	14	23+14=37↑=C
35.5-40.5	26=f	37+26=63←**Median class**
40.5-45.5	12	63+12=75
45.5-50.5	16	75+16=91
50.5-55.5	9	91+9=100
Total	$N=\Sigma f_i=100$	

Now, $\frac{N}{2}=\frac{100}{2}=50$; *c.f.* just ≥ $\frac{N}{2}=50$ is 63

The class-interval corresponding to this *c.f.* is 35.5-40.5.

∴ Median lies in the class-interval 35.5-40.5 (Median Class)

∴ l = 35.5, h = 5

f = frequency of median class = 26

C = cumulative frequency of the class preceding the median class = 37.

∴ Median $M = l + \frac{h}{f}\left(\frac{N}{2}-C\right)$

$$= 35.5 + \frac{5}{26}(50-37)$$

$$= 35.5 + \frac{5}{26}(13) = 35.5 + \frac{5}{2} = 35.5 + 2.5 = 38$$

Now let us find mean deviation from median

x_i (*Mid-value of class-interval*)	f_i	$\|x_i - M\| = \|x_i - 38\|$	$f_i \|x_i - M\|$
18	5	20	100
23	6	15	90
28	12	10	120
33	14	5	70
38	26	0	0
43	12	5	60
48	16	10	160
53	9	15	135
	$N = \Sigma f_i = 100$		$\Sigma f_i \|x_i - M\| = 735$

∴ By Formula, Mean deviation about Median

$$= \frac{1}{N} \sum_{i=1}^{8} f_i \; | x_i - M |$$

$$= \frac{735}{100} = 7.35.$$

EXERCISE 15.2 (*Page No.: 371–372*)

Find the mean and variance for each of the data in Exercises 1 to 5.

1. 6, 7, 10, 12, 13, 4, 8, 12.

Sol. From the given data, we form the following table:

x_i	$x_i - \bar{x} = x_i - 9$	$(x_i - \bar{x})^2$
6	– 3	9
7	– 2	4
10	1	1
12	3	9
13	4	16

4	−5	25
8	−1	1
12	3	9
72		74

Here $n = 8, \sum_{i=1}^{8} x_i = 72,$

$$\therefore \quad \bar{x} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{72}{8} = 9 \text{ is a natural number.}$$

From the table, $\sum_{i=1}^{8} (x_i - \bar{x})^2 = 74$

$$\therefore \quad \text{Variance } (\sigma^2) = \frac{1}{n} \sum_{i=1}^{8} (x_i - \bar{x})^2$$

$$= \frac{1}{8} \times 74 = 9.25.$$

2. First n natural numbers.

Sol. Here, x : 1, 2, 3, ..., n

$$\therefore \quad \Sigma x = 1 + 2 + 3 + \ldots + n = \frac{n(n+1)}{2}$$

$$\therefore \quad \text{Mean } \bar{x} = \frac{\Sigma x}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2}$$

Again $\Sigma x^2 = 1^2 + 2^2 + 3^2 + \ldots + n^2 = \dfrac{n(n+1)(2n+1)}{6}$

We know that

$$\textbf{Variance } \sigma^2 = \frac{1}{n}\Sigma x^2 - \bar{x}^2$$

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

$$= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$$

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

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

$$= \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}$$

Remark: The results of this Question is very important for I.I.T. entrance exam.

3. First 10 multiples of 3.

Sol. We know that first 10 multiples of 3 are: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30

Their mean $\bar{x} = \frac{3+6+9+12+15+18+21+24+27+30}{10}$

$= \frac{165}{10} = 16.5$ is not a natural number.

So let us use step-deviation method to find variance.

Let us take A = 18 (say) as assumed mean.

x_i	$d_i = x_i - A = x_i - 18$	d_i^2
3	−15	225
6	−12	144
9	−9	81
12	−6	36
15	−3	9
18 = A	0	0
21	3	9
24	6	36
27	9	81
30	12	144
	$\Sigma d_i = -15$	$\Sigma d_i^2 = 765$

$\therefore \quad \bar{d} = \frac{\Sigma d_i}{n} = \frac{-15}{10} = -1.5$

We know that variance

$$\sigma^2 = \frac{1}{n}\Sigma d_i^2 - \bar{d}^2$$

$$= \frac{765}{10} - (-1.5)^2 = 76.5 - 2.25$$

$$= 76.50 - 2.25 = 74.25.$$

4.

x_i	6	10	14	18	24	28	30
f_i	2	4	7	12	8	4	3

Sol. Let the assumed mean A be 18. We obtain the following table from the given data.

x_i	f_i	$d_i = x_i - 18$	d_i^2	$f_i d_i$	$f_i d_i^2$
6	2	– 12	144	– 24	288
10	4	– 8	64	– 32	256
14	7	– 4	16	– 28	112
18	12	0	0	0	0
24	8	6	36	48	288
28	4	10	100	40	400
30	3	12	144	36	432
	40			40	1776

Here, $N = \sum_{i=1}^{7} f_i = 40, \ \sum_{i=1}^{7} f_i d_i = 40$

$$\therefore \quad \bar{x} = A + \frac{\sum_{i=1}^{7} f_i d_i}{N} = 18 + \frac{40}{40} = 19$$

and Variance $(\sigma^2) = \frac{1}{N}\sum_{i=1}^{7} f_i d_i^2 - \left(\frac{\sum_{i=1}^{7} f_i d_i}{N}\right)^2$

$$= \frac{1776}{40} - \left(\frac{40}{40}\right)^2$$

$$= \frac{177.6}{4} - (1)^2 = 44.4 - 1 = 43.4$$

OR

Second solution

We know that Mean $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{760}{40} = 19$ which is a natural number.

x_i	f_i	$f_i x_i$	$x_i - 19 = x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
6	2	12	–13	169	338
10	4	40	–9	81	324
14	7	98	–5	25	175
18	12	216	–1	1	12
24	8	192	5	25	200
28	4	112	9	81	324
30	3	90	11	121	363
	$N = \Sigma f_i = 40$	$\Sigma f_i x_i = 760$			1736

We know that Var (X) = $\frac{1}{N} \Sigma f_i (x_i - \bar{x})^2 = \frac{1736}{40} = 43.4$

Remark: $\therefore$ S.D. = $\sqrt{\text{Var.}(X)} = \sqrt{43.4} = 6.59.$

5.

x_i	92	93	97	98	102	104	109
f_i	3	2	3	2	6	3	3

Sol. Let the assumed mean A = 100. We obtain the following table from the given data.

x_i	f_i	$d_i = x_i - 100$	d_i^2	$f_i d_i$	$f_i d_i^2$
92	3	– 8	64	– 24	192
93	2	– 7	49	– 14	98
97	3	– 3	9	– 9	27
98	2	– 2	4	– 4	8
102	6	2	4	12	24
104	3	4	16	12	48
109	3	9	81	27	243
	22			0	640

Here, $N = \sum_{i=1}^{7} f_i = 22, \quad \sum_{i=1}^{7} f_i d_i = 0$

$$\therefore \quad \bar{x} = A + \frac{\sum_{i=1}^{7} f_i d_i}{N} = 100 + \frac{0}{22} = 100$$

$$\text{and Variance } (\sigma^2) = \frac{1}{N}\sum_{i=1}^{7} f_i d_i^2 - \left(\frac{\sum_{i=1}^{7} f_i d_i}{N}\right)^2$$

$$= \frac{640}{22} - \left(\frac{0}{22}\right)^2 = \frac{640}{22} = \frac{320}{11} = 29.09$$

6. Find the mean and standard deviation using short-cut method.

x_i	60	61	62	63	64	65	66	67	68
f_i	2	1	12	29	25	12	10	4	5

Sol. Let the assumed mean A = 64. We obtain the following table from the given data.

x_i	f_i	$d_i = x_i - 64$	d_i^2	$f_i d_i$	$f_i d_i^2$
60	2	– 4	16	– 8	32
61	1	– 3	9	– 3	9
62	12	– 2	4	– 24	48
63	29	– 1	1	– 29	29
64	25	0	0	0	0
65	12	1	1	12	12
66	10	2	4	20	40
67	4	3	9	12	36
68	5	4	16	20	80
	100			0	286

Here, $N = \sum_{i=1}^{9} f_i = 100, \quad \sum_{i=1}^{9} f_i d_i = 0$

$$\therefore \quad \bar{x} = A + \frac{\sum_{i=1}^{9} f_i d_i}{N} = 64 + \frac{0}{100} = 64$$

and Variance $(\sigma^2) = \frac{1}{N}\sum_{i=1}^{9} f_i d_i^2 - \left(\frac{1}{N}\sum f_i d_i\right)^2$

$$= \left[\frac{286}{100} - 0\right] = 2.86.$$

Standard deviation $(\sigma) = \sqrt{\text{variance}} = \sqrt{2.86} = 1.69.$

Find the mean and variance for the following frequency distributions in Exercises 7 and 8.

7.

Classes	Frequencies
0-30	2
30-60	3
60-90	5
90-120	10
120-150	3
150-180	5
180-210	2

Sol. Let us take A = 105 and h = 30 (= Difference of values of x)

Class-Interval	*Frequency* (f_i)	*Mid-value* (x_i)	$y_i = \frac{x_i - 105}{30}$	$f_i y_i$	$f_i y_i^2 = y_i f_i y_i$
0-30	2	15	–3	–6	18
30-60	3	45	–2	–6	12
60-90	5	75	–1	–5	5
90-120	10	105	0	0	0
120-150	3	135	1	3	3
150-180	5	165	2	10	20
180-210	2	195	3	6	18
	$N = \Sigma f_i = 30$			$\Sigma f_i y_i = 2$	$\Sigma f_i y_i^2 = 76$

$$N = 30, \ \Sigma f_i y_i = 2, \ \Sigma f_i y_i^2 = 76, \ h = 30$$

We know that Mean

$$\bar{x} = A + \frac{h}{N}\Sigma f_i\, y_i \;\Rightarrow\; \bar{x} = 105 + 30\left(\frac{2}{30}\right)$$

$$= 105 + 2 = 107$$

Also we know that

$$\text{Variance } (\sigma^2) = h^2\left[\frac{1}{N}\Sigma f_i y_i^2 - \left(\frac{1}{N}\Sigma f_i y_i\right)^2\right]$$

$$= 900\left[\frac{76}{30} - \left(\frac{2}{30}\right)^2\right]$$

$$= 900\left[\frac{76}{30} - \frac{4}{900}\right]$$

$$= 900\left[\frac{2280-4}{900}\right] = 2276.$$

8.

Classes	0-10	10-20	20-30	30-40	40-50
Frequencies	5	8	15	16	6

Sol. After writing the mid-values of class-intervals, let us take assumed mean A = 25. Here h = 10. We obtain the folloiwng table from the given data:

Class	*Frequency* (f_i)	*Mid-points* (x_i)	$y_i = \frac{x_i - 25}{10}$	y_i^2	$f_i y_i$	$f_i y_i^2$
0-10	5	5	– 2	4	– 10	20
10-20	8	15	– 1	1	– 8	8
20-30	15	25	0	0	0	0
30-40	16	35	1	1	16	16
40-50	6	45	2	4	12	24
	50				10	68

Here, $N = \sum_{i=1}^{5} f_i = 50, \; \sum_{i=1}^{5} f_i y_i = 10$

$$\therefore \qquad \bar{x} = A + \frac{\sum_{i=1}^{5} f_i y_i}{N} \times h$$

$$= 25 + \frac{10}{50} \times 10$$

$$= 25 + 2 = 27$$

$$\text{Variance } (\sigma^2) = h^2\left[\frac{1}{N}\sum_{i=1}^{5} f_i y_i^2 - \left(\frac{1}{N}\sum_{i=1}^{5} f_i y_i\right)^2\right]$$

$$= \frac{h^2}{N^2}\left[N\sum_{i=1}^{5} f_i y_i^2 - \left(\sum_{i=1}^{5} f_i y_i\right)^2\right]$$

$$= \frac{(10)^2}{(50)^2}[50 \times 68 - (10)^2]$$

$$= \frac{1}{25}(3400 - 100)$$

$$= \frac{3300}{25} = 132.$$

9. Find the mean, variance and standard deviation using short-cut method.

Height (in cms)	Number of children
70-75	3
75-80	4
80-85	7
85-90	7
90-95	15
95-100	9
100-105	6
105-110	6
110-115	3

Sol. After taking the mid-values of class-intervals, let us take the assumed mean A = 92.5. Here $h = 5$

We obtain the following table from the given data.

Height (in cms)	*No. of children* (f_i)	*Mid-point* (x_i)	$y_i = \frac{x_i - 92.5}{5}$	y_i^2	$f_i y_i$	$f_i y_i^2$
70-75	3	72.5	– 4	16	– 12	48
75-80	4	77.5	– 3	9	– 12	36
80-85	7	82.5	– 2	4	– 14	28
85-90	7	87.5	– 1	1	– 7	7
90-95	15	92.5	0	0	0	0
95-100	9	97.5	1	1	9	9
100-105	6	102.5	2	4	12	24
105-110	6	107.5	3	9	18	54
110-115	3	112.5	4	16	12	48
	60				6	254

Here, $N = \sum_{i=1}^{9} f_i = 60, \; \sum_{i=1}^{9} f_i y_i = 6$

$$\therefore \quad \bar{x} = A + \frac{\sum_{i=1}^{9} f_i y_i}{N} \times h = 92.5 + \frac{6}{60} \times 5$$

$$= 92.5 + 0.5 = 93$$

and Variance $(\sigma^2) = h^2\left[\frac{1}{N}\sum f_i y_i^2 - \left(\frac{\sum f_i y_i}{N}\right)^2\right]$

$$= \frac{h^2}{N^2}\left[N\sum_{i=1}^{9} f_i y_i^2 - \left(\sum_{i=1}^{9} f_i y_i\right)^2\right]$$

$$= \frac{(5)^2}{(60)^2}[60 \times 254 - (6)^2] = \frac{1}{144}(15240 - 36)$$

$$= \frac{15204}{144} = \frac{1267}{12} = 105.58$$

Standard deviation (σ) = $\sqrt{105.58}$ = 10.27.

10. The diameters of circles (in mm) drawn in a design are given below:

Diameters	**33-36**	**37-40**	**41-44**	**45-48**	**49-52**
No. of circles	**15**	**17**	**21**	**22**	**25**

Calculate the standard deviation and mean diameter of the circles.

Sol. First of all, let us make the data continuous. [See solution of Example 12, Exercise 15.1, Page 470]

Let us make the data continuous by taking classes as 32.5-36.5, 36.5-40.5, 40.5-44.5, 44.5-48.5, 48.5-52.5.

Diameter	*Mid-value* (x_i)	*No. of Circles* (f_i)	$y_i = \frac{x_i - 42.5}{4}$	$f_i y_i$	$f_i y_i^2$
32.5-36.5	34.5	15	–2	–30	60
36.5-40.5	38.5	17	–1	–17	17
40.5-44.5	42.5	21	0	0	0
44.5-48.5	46.5	22	1	22	22
48.5-52.5	50.5	25	2	50	100
		$\Sigma f_i = 100$		$\Sigma f_i y_i = 25$	$\Sigma f_i y_i^2 = 199$

$$N = \Sigma f_i = 100,\ \Sigma f_i y_i = 25,\ \Sigma f_i y_i^2 = 199$$

$$A = 42.5,\ h = 4$$

We know that Variance

$$= h^2\left[\frac{1}{N}\sum f_i y_i^2 - \left(\frac{\sum f_i y_i}{N}\right)^2\right]$$

$$= 16\left[\frac{1(199)}{100} - \left(\frac{25}{100}\right)^2\right]$$

$$= 16[1.99 - 0.0625]$$

or $\sigma^2 = 16 \times 1.9275 = 30.84$

$\therefore$ S.D. $\sigma = \sqrt{30.84} = 5.55.$

Again we know that Mean

$$\bar{x} = A + h\frac{\Sigma f_i y_i}{\Sigma f_i}$$

$$= 42.5 + 4\left(\frac{25}{100}\right) = 42.5 + 1 = 43.5.$$

EXERCISE 15.3 *(Page No.: 375–76)*

1. From the data given below state which group is more variable, A or B?

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Group A	9	17	32	33	40	10	9
Group B	10	20	30	25	43	15	7

Sol. For group A

Marks	*Frequency* (f_i)	*Mid-values* (x_i)	$u_i = \frac{x_i - 45}{10}$, A = 45, h = 10	$f_i u_i$	$f_i u_i^2$
10-20	9	15	– 3	– 27	81
20-30	17	25	– 2	– 34	68
30-40	32	35	– 1	– 32	32
40-50	33	45	0	0	0
50-60	40	55	1	40	40
60-70	10	65	2	20	40
70-80	9	75	3	27	81
	N = 150			– 6	342

$$\text{Mean } \bar{x} = A + \frac{\Sigma f_i u_i}{N} \times h = 45 - \frac{6}{150} \times 10$$

$$= 45 - 0.4 = 44.6$$

$$\text{and Variance } \sigma_x^2 = h^2\left[\frac{1}{N}\sum f_i u_i^2 - \left(\frac{1}{N}\sum f_i u_i\right)^2\right]$$

$$= \frac{h^2}{N^2}\left[N\Sigma f_i u_i^2 - (\Sigma f_i u_i)^2\right]$$

$$= \frac{(10)^2}{(150)^2}[150 \times 342 - (-6)^2]$$

$$= \frac{1}{225}[51300 - 36]$$

$$= \frac{51264}{225} = \frac{51264}{225}\times\frac{4}{4} = \frac{205056}{900} = \frac{2050.56}{9}$$

$$= 227.84$$

$\therefore$ S.D. $(\sigma_x) = \sqrt{227.84} = 15.09$

For group B

Marks	*Frequency* (f_i)	*Mid-values* (y_i)	$u_i = \frac{y_i - 45}{10}$ A = 45, h = 10	$f_i u_i$	$f_i u_i^2$
10-20	10	15	– 3	– 30	90
20-30	20	25	– 2	– 40	80
30-40	30	35	– 1	– 30	30
40-50	25	45	0	0	0
50-60	43	55	1	43	43
60-70	15	65	2	30	60
70-80	7	75	3	21	63
	N = 150			– 6	366

Mean $\bar{y} = A + \frac{\Sigma f_i u_i}{N} \times h = 45 - \frac{6}{150} \times 10$

$= 45 - 0.4 = 44.6$

and Variance $\sigma_y^2 = h^2\left[\frac{1}{N}\sum f_i u_i^2 - \left(\frac{1}{N}\sum f_i u_i\right)^2\right]$

$$= \frac{h^2}{N^2}[N\Sigma f_i u_i^2 - (\Sigma f_i u_i)^2]$$

$$= \frac{(10)^2}{(150)^2}[150 \times 366 - (-6)^2]$$

$$= \frac{1}{225}[54900 - 36] = \frac{54864}{225} = \frac{54864 \times 4}{225 \times 4}$$

$$= \frac{219456}{900} = \frac{2194.56}{9} = 243.84$$

$\therefore$ S.D. $(\sigma_y) = \sqrt{243.84} = 15.61.$

We know that coefficient of variation $= \frac{\sigma}{\bar{x}} \times 100$

Since the two groups have same mean, therefore, the group with greater standard deviation will be more variable.

Thus, the group B is more variable ($\because \sigma_y > \sigma_x$).

2. From the prices of shares X and Y below, find out which is more stable in value:

X	**35**	**54**	**52**	**53**	**56**	**58**	**52**	**50**	**51**	**49**
Y	**108**	**107**	**105**	**105**	**106**	**107**	**104**	**103**	**104**	**101**

Sol. For shares X

x_i	$d_i = x_i - 50$	d_i^2
35	– 15	225
54	4	16
52	2	4
53	3	9
56	6	36
58	8	64
52	2	4
50	0	0
51	1	1
49	– 1	1
$n = 10$	10	360

Mean $\bar{x} = A + \frac{\Sigma d_i}{n} = 50 + \frac{10}{10} = 50 + 1 = 51$

and Variance $\sigma_x^2 = \frac{1}{n}\sum d_i^2 - \left(\frac{\sum di}{n}\right)^2 = \frac{1}{n^2}[n\Sigma d_i^2 - (\Sigma d_i)^2]$

$$= \frac{1}{(10)^2}[10 \times 360 - (10)^2]$$

$$= \frac{1}{100}(3600 - 100)$$

$$= \frac{3500}{100} = 35$$

$\therefore$ S.D. $(\sigma_x) = \sqrt{35} = 5.92.$

For shares Y

y_i	$d_i = y_i - 104$	d_i^2
108	4	16
107	3	9
105	1	1
105	1	1
106	2	4
107	3	9
104	0	0
103	– 1	1
104	0	0
101	– 3	9
$n = 10$	10	50

Mean $\bar{y} = A + \frac{\Sigma d_i}{n} = 104 + \frac{10}{10} = 104 + 1 = 105$

and Variance $\sigma_y^2 = \frac{1}{n}\sum d_i^2 - \left(\frac{\sum di}{n}\right)^2 = \frac{1}{n^2}[n\Sigma d_i^2 - (\Sigma d_i)^2]$

$$= \frac{1}{(10)^2}[10 \times 50 - (10)^2]$$

$$= \frac{1}{100}(500 - 100)$$

$$= \frac{400}{100} = 4$$

$\therefore$ S.D. $(\sigma_y) = \sqrt{4} = 2$

Now C.V. for shares X = $\frac{\sigma_x}{\bar{x}} \times 100$

$$= \frac{5.92}{51} \times 100 = 11.61$$

C.V. for shares Y = $\frac{\sigma_y}{\bar{y}} \times 100$

$$= \frac{2}{105} \times 100 = 1.90$$

Since C.V. for shares Y is less than C.V. for shares X, therefore, share Y is more stable.

Remark: Y is more stable ⇒ X is more variable.

3. An analysis of monthly wages paid to workers in two firms A and B, belonging to the same industry, gives the following results:

	Firm A	**Firm B**
No. of wage earners	**586**	**648**
Mean of monthly wages	**₹ 5253**	**₹ 5253**
Variance of the distribution of wages	**100**	**121**

(i) Which firm, A or B, pays larger amount as monthly wages?

(ii) Which firm, A or B, shows greater variability in individual wages?

Sol. (i) Here $\bar{x} = \bar{y}$, each = ₹ 5253 (given)

$n_1 = 586,\quad n_2 = 648$

Amount paid by a firm as monthly wages = $n\bar{x}$ $\left(\because \bar{x} = \frac{\Sigma x}{n}\right)$

= No. of workers × Mean of monthly wages

∴ Firm A pays ₹ $n_1\bar{x}$ and Firm B pays ₹ $n_2\bar{y}$

Since $\bar{x} = \bar{y}$ and $n_2 > n_1$

∴ Firm B pays larger amount as monthly wages.

(ii) Here $\sigma_A^2 = 100,\quad \sigma_B^2 = 121$

⇒ $\sigma_A = 10,\quad \sigma_B = 11$

Since $\bar{x} = \bar{y}$ and $\sigma_B > \sigma_A$, therefore Firm B shows greater variability in individual wages.

Remark: More variablity in B ⇒ more stability in A.

4. The following is the record of goals scored by team A in a football session:

No. of goals scored	0	1	2	3	4
No. of matches	1	9	7	5	3

For the team B, mean number of goals scored per match was 2 with a standard deviation 1.25 goals. Find which team may be considered more consistent?

Sol. For team A

No. of goals scored (x_i)	*No. of matches* (f_i)	$f_i x_i$	$f_i x_i^2$
0	1	0	0
1	9	9	9
2	7	14	28
3	5	15	45
4	3	12	48
	N = 25	50	130

Mean $\bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{50}{25} = 2$

and Variance $\sigma_A^2 = \frac{1}{N}\sum f_i x_i^2 - \left(\frac{1}{N}\sum f_i x_i\right)^2$

$$= \frac{1}{N^2}[N\Sigma f_i x_i^2 - (\Sigma f_i x_i)^2]$$

$$= \frac{1}{(25)^2}[25 \times 130 - (50)^2]$$

$$= \frac{25}{(25)^2}[130 - 100] = \frac{30}{25} = 1.2.$$

For team B

Mean $\bar{y} = 2$,

Variance $\sigma_B^2 = (1.25)^2 = 1.5625 \approx 1.56$

Since $\bar{x} = \bar{y} = 2$ and $\sigma_A^2 < \sigma_B^2$, ($\Rightarrow \sigma_A < \sigma_B$), team A has smaller variability. Hence, team A is more consistent.

5. The sum and sum of squares corresponding to length x (in cm) and weight y (in g) of 50 plant products are given below:

$$\sum_{i=1}^{50} x_i = 212, \quad \sum_{i=1}^{50} x_i^2 = 902.8, \quad \sum_{i=1}^{50} y_i = 261,$$

$$\sum_{i=1}^{50} y_i^2 = 1457.6$$

Which is more varying, the length or weight?

Sol. For length (x)

$$\text{Mean } \bar{x} = \frac{\sum_{i=1}^{50} x_i}{50} = \frac{212}{50} = 4.24$$

$$\text{S.D. } (\sigma_x) = \sqrt{\frac{1}{50}\sum_{i=1}^{50} x_i^2 - (\bar{x})^2}$$

$$= \sqrt{\frac{1}{50}(902.8) - (4.24)^2}$$

$$= \sqrt{18.056 - 17.9776} = \sqrt{0.0784} = 0.28$$

$$\text{C.V. for } x = \frac{\sigma_x}{\bar{x}} \times 100 = \frac{0.28}{4.24} \times 100 = 6.60$$

For weight (y)

$$\text{Mean } \bar{y} = \frac{\sum_{i=1}^{50} y_i}{50} = \frac{261}{50} = 5.22$$

$$\text{S.D. } (\sigma_y) = \sqrt{\frac{1}{50}\sum_{i=1}^{50} y_i^2 - (\bar{y})^2} = \sqrt{\frac{1}{50}(1457.6) - (5.22)^2}$$

$$= \sqrt{29.152 - 27.2484} = \sqrt{1.9036} = 1.38$$

$$\text{C.V. for } y = \frac{\sigma_y}{y} \times 100 = \frac{1.38}{5.22} \times 100 = 26.44$$

Since C.V. for y > C.V. for x, therefore, weight is more varying.

MISCELLANEOUS EXERCISE ON CHAPTER 15

(Page No.: 380)

1. The mean and variance of eight observations are 9 and 9.25 respectively. If six of the observations are 6, 7, 10, 12, 12 and 13, find the remaining two observations.

Sol. Six of the observations are 6, 7, 10, 12, 12 and 13 (given)

Let the remaining two observations (out of $n = 8$) be a and b.

We know that mean $\bar{x} = \frac{\Sigma x}{n}$

But $\bar{x} = 9$ (given)

$\therefore \quad 9 = \frac{6+7+10+12+12+13+a+b}{8}$

$\Rightarrow (9 \times 8 =)\ 72 = 60 + a + b$ or $a + b = 12$ ...(*i*)

We also know that variance $\sigma^2 = \frac{1}{n}\Sigma x^2 - (\bar{x})^2$

Putting $\sigma^2 = 9.25$ (given), $\bar{x} = 9$ (given),
$n = 8$ and values of x, we have

$$9.25 = \frac{1}{8}(6^2 + 7^2 + 10^2 + 12^2 + 12^2 + 13^2 + a^2 + b^2) - 9^2$$

$$\Rightarrow 9.25 + 81 = \frac{1}{8}(36 + 49 + 100 + 144 + 144 + 169 + a^2 + b^2)$$

$\Rightarrow \quad 90.25 \times 8 = 642 + a^2 + b^2$

$\Rightarrow \quad 722 - 642 = a^2 + b^2$

$\Rightarrow \quad a^2 + b^2 = 80$ ...(*ii*)

From (*i*), $\quad b = 12 - a$

Putting in (*ii*), we have $a^2 + (12 - a)^2 = 80$

$\Rightarrow \quad a^2 + 144 + a^2 - 24a - 80 = 0$

$\Rightarrow \quad 2a^2 - 24a + 64 = 0$

$\Rightarrow$ Dividing by 2,

$a^2 - 12a + 32 = 0$

$\Rightarrow \quad (a - 4)(a - 8) = 0$

$\Rightarrow \quad a = 4, 8$

When $\quad a = 4, b = 12 - 4 = 8$

When $\quad a = 8, b = 12 - 8 = 4$

$\therefore$ The remaining two observations are 4 and 8.

2. The mean and variance of 7 observations are 8 and 16 respectively. If five of the observations are 2, 4, 10, 12, 14, find the remaining two observations.

Sol. Five of the observations are 2, 4, 10, 12 and 14. (given)

Let the remaining two observations (out of $n = 7$ (given)) be a and b.

We know that mean $\bar{x} = \frac{\Sigma x_i}{n}$

But $\bar{x} = 8$ (given)

$\therefore \quad 8 = \frac{2+4+10+12+14+a+b}{7}$

Cross-multiplying $56 = 42 + a + b$ or $a + b = 14$ ...(i)

We also know that variance $\sigma^2 = \frac{1}{n}\Sigma x^2 - \bar{x}^2$

Putting $\sigma^2 = 16$ (given), $\bar{x} = 8$ (given), $n = 7$ and values of x, we have

$$16 = \frac{1}{7}[2^2 + 4^2 + 10^2 + 12^2 + 14^2 + a^2 + b^2] - 64$$

Transposing, $80 = \frac{1}{7}[4 + 16 + 100 + 144 + 196 + a^2 + b^2]$

Cross-multiplying, $560 = 460 + a^2 + b^2$

or $a^2 + b^2 = 100$

Putting $b = 14 - a$ from (i), we have

$$a^2 + (14 - a)^2 = 100$$

or $a^2 + 196 + a^2 - 28a = 100$

or $2a^2 - 28a + 96 = 0$

Dividing by 2, $a^2 - 14a + 48 = 0$

or $a^2 - 6a - 8a + 48 = 0$

or $a(a - 6) - 8(a - 6) = 0$

or $(a - 6)(a - 8) = 0$

$\therefore \quad a = 6, \quad a = 8$

When $a = 6$, from (i) $b = 14 - a = 14 - 6 = 8$

When $a = 8$, from (i) $b = 14 - a = 14 - 8 = 6$.

$\therefore$ The remaining two observations are 6, 8 or 8, 6.

3. The mean and standard deviation of six observations are 8 and 4 respectively. If each observation is multiplied by 3, find the new mean and new standard deviation of the resulting observations.

Sol. Let the six observations be $x : x_1, x_2, x_3, x_4, x_5, x_6$

Mean $$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{6} = 8 \quad \text{(given)}$$

$$\therefore \quad \sum_{i=1}^{6} x_i = 8 \times 6 = 48 \quad ...(i)$$

S.D. $$\sigma_x = \sqrt{\frac{1}{6}\sum_{i=1}^{6}(x_i - \bar{x})^2} = 4 \quad \text{(given)}$$

Squaring both sides; $$\frac{1}{6}\sum_{i=1}^{6}(x_i - \bar{x})^2 = 16$$

$$\Rightarrow \quad \sum_{i=1}^{6}(x_i - \bar{x})^2 = 16 \times 6 = 96 \quad ...(ii)$$

When each observation is multiplied by 3, the new observations are

$y : 3x_1, 3x_2, 3x_3, 3x_4, 3x_5, 3x_6$

i.e., $y_i = 3x_i, i = 1, 2, ..., 6$

New mean $$\bar{y} = \frac{\sum_{i=1}^{6} y_i}{6} = \frac{\sum_{i=1}^{6} 3x_i}{6} = \frac{3\sum_{i=1}^{6} x_i}{6}$$

$$= \frac{3 \times 48}{6} \text{ [By (i)]} = 24 \quad \text{so that } \bar{y} = 3\bar{x}$$

New S.D. $$\sigma_y = \sqrt{\frac{1}{6}\sum_{i=1}^{6}(y_i - \bar{y})^2} = \sqrt{\frac{1}{6}\sum_{i=1}^{6}(3x_i - 3\bar{x})^2}$$

$[\because y_i = 3xi \text{ and } \bar{y} = 3\bar{x}]$

$$= \sqrt{\frac{1}{6}\sum_{i=1}^{6}9(x_i - \bar{x})^2} = \sqrt{\frac{9}{6}\sum_{i=1}^{6}(x_i - \bar{x})^2}$$

$$= \sqrt{\frac{3}{2} \times 96} \text{ [By (ii)]} = \sqrt{144} = 12.$$

4. Given that $\bar{x}$ is the mean and σ^2 is the variance of n observations $x_1, x_2, ..., x_n$. Prove that the mean and variance of the observations $ax_1, ax_2, ax_3, ..., ax_n$ are $a\bar{x}$ and $a^2\sigma^2$ respectively, ($a \neq 0$).

Sol. Let $\bar{x}$ and σ^2 be the mean and variance of the observations $x_1, x_2, ..., x_n$. Then

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad ...(i)$$

When each observation is multiplied by a ($\neq$ 0), then new observations are $ax_1, ax_2, ..., ax_n$, *i.e.*, ax_i $(i = 1, 2, ..., n)$

$\therefore$ Mean of new observations

$$ax_1, ax_2, ..., ax_n \text{ is } \frac{ax_1 + ax_2 + ... + ax_n}{n}$$

$$= \frac{a(x_1 + x_2 + ... + x_n)}{n} = a\bar{x} \qquad \text{[By (i)]}$$

$\therefore$ New mean $= a\bar{x}$

Changing x_i to ax_i and $\bar{x}$ to $a\bar{x}$ in eqn. (*i*), we have

$$\text{New variance} = \frac{1}{n}\sum_{i=1}^{n}(ax_i - a\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} a^2(x_i - \bar{x})^2$$

$$= a^2\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right) = a^2\sigma^2. \qquad \text{[Using (i)]}$$

Note. Similarly, if each of the n given numbers (observations) of a series having variance σ^2 is divided by a $(a \neq 0)$, then the new mean is $\frac{\bar{x}}{a}$ and the new variance is $\frac{\sigma^2}{a^2}$.

Remark. The results of this question and the above Note are very useful for I.I.T. entrance examination.

5. The mean and standard deviation of 20 observations are found to be 10 and 2 respectively. On rechecking, it was found that an observation 8 was incorrect. Calculate the correct mean and standard deviation in each of the following cases:

(*i*) If wrong item is omitted. (*ii*) If it is replaced by 12.

Sol. Here $n = 20$, $\bar{x} = 10$, and S.D., $\sigma = 2$ (given)

We know that $\bar{x} = \frac{\Sigma x_i}{n}$

$\therefore$ $\Sigma x_i = n\bar{x} = 20 \times 10 = 200$ ...(*i*)

It is given that S.D. $\sigma = 2$

$\therefore$ $\sigma^2 = 4$

Using formula $\frac{1}{n}\Sigma x_i^2 - \bar{x}^2 = 4.$

Putting values of n and $\bar{x}$,

$$\frac{1}{20}\Sigma x_i^2 - (10)^2 = 4$$

or $$\frac{1}{20}\Sigma x_i^2 = 100 + 4 = 104$$

$\therefore$ $\Sigma x_i^2 = 104 \times 20 = 2080$ ...(*ii*)

(*i*) Wrong item 8 is omitted

$\therefore$ From (*i*) new $\Sigma x_i = 200 - 8 = 192$

and from (*ii*) new $\Sigma x_i^2 = 2080 - (8)^2$

$= 2080 - 64 = 2016$

and new $n = 20 - 1 = 19$

$\therefore$ New (correct) mean $= \frac{\text{New } \Sigma x_i}{n} = \frac{192}{19} = 10.1$

and New (correct) variance

$$= \frac{1}{n}(\text{New } \Sigma x_i^2) - (\text{New mean})^2$$

$$= \frac{1}{19}(2016) - \left(\frac{192}{19}\right)^2 = \frac{2016}{19} - \frac{(192)^2}{361}$$

$$= \frac{1}{361}(19 \times 2016 - 192 \times 192)$$

$$= \frac{1}{361}(38304 - 36864)$$

$$= \frac{1440}{361} = 3.99$$

$\therefore$ New S.D. $= \sqrt{3.99} = 1.99.$

(*ii*) Wrong entry 8 is replaced by correct entry 12

$\therefore$ Corrected Σx_i = Incorrect Σx_i from (*i*)

$-$ Incorrect entry + Correct entry

$= 200 - 8 + 12 = 204$

But new n again remains 20 ($\because$ No item is omitted)

$\therefore$ Corrected mean $= \dfrac{\text{Corrected } \Sigma x_i}{n}$

$= \dfrac{204}{20} = \dfrac{102}{10} = 10.2$

Again corrected Σx_i^2 = Incorrect Σx_i^2 from (*ii*)

$-$ (Incorrect value)^2 + (Correct value)^2

$= 2080 - (8)^2 + (12)^2$

$= 2080 - 64 + 144 = 2160$

$\therefore$ Corrected $\sigma^2 = \dfrac{1}{n}$ Correct $\Sigma x_i^2 -$ (Corrected $\bar{x}$)^2

$= \dfrac{2160}{20} - (10.2)^2 = 108 - 104.04 = 3.96$

$\therefore$ Corrected S.D. $\sigma = \sqrt{3.96} = 1.99$.

6. The mean and standard deviation of marks obtained by 50 students of a class in three subjects, Mathematics, Physics and Chemistry are given below:

Subject	Mathematics	Physics	Chemistry
Mean	42	32	40.9
Standard deviation	12	15	20

which of the three subjects shows the highest variability in marks and which shows the lowest?

Sol. C.V. for Mathematics $= \dfrac{\text{S.D.}}{\text{Mean}} \times 100 = \dfrac{12}{42} \times 100 = 28.57$

C.V. for Physics $= \dfrac{\text{S.D.}}{\text{Mean}} \times 100 = \dfrac{15}{32} \times 100 = 46.88$

C.V. for Chemistry $= \dfrac{\text{S.D.}}{\text{Mean}} \times 100 = \dfrac{20}{40.9} \times 100 = 48.9$

$\therefore$ Chemistry has the greatest C.V. and hence the highest variability. Mathematics has the least C.V. and hence the lowest variability.

7. The mean and standard deviation of a group of 100 observations were found to be 20 and 3 respectively. Later on it was found that three observations were incorrect, which were recorded as 21, 21 and 18. Find the mean and standard deviation if the incorrect observations are omitted.

Sol. $n = 100$, mean $\bar{x} = 20$

We know that $\bar{x} = \dfrac{\Sigma x_i}{n}$

$\therefore \quad \Sigma x_i = n\bar{x} = 100 \times 20 = 2000$

Three observations 21, 21 and 18 were found to be incorrect. (given)

$\therefore$ New Σx_i (after omitting incorrect observations as required)

$= \Sigma x_i - (21 + 21 + 18) = 2000 - 60 = 1940$

$\therefore$ New $n = 100 - 3$ incorrect observations

$= 97$

$$\therefore \text{ New mean} = \frac{\text{New } \Sigma x_i}{\text{New } n} = \frac{1940}{97} = 20$$

S.D. = 3 (given)

$$\Rightarrow \quad \sigma = 3 \quad \Rightarrow \quad \sigma^2 = 9 \quad \Rightarrow \quad \frac{1}{n}\Sigma x_i^2 - \bar{x}^2 = 9$$

Putting originally given values of n and $\bar{x}$, we have

$$\Rightarrow \quad \frac{1}{100}\Sigma x_i^2 - 400 = 9$$

$$\Rightarrow \quad \frac{1}{100}\Sigma x_i^2 = 409$$

$\therefore \quad \Sigma x_i^2 = 40900$

$\therefore$ Corrected $\Sigma x_i^2 = 40900 - (21)^2 - (21)^2 - (18)^2$

$= 40900 - 441 - 441 - 324$

$= 40900 - 1206 = 39694$

$$\therefore \text{ Corrected } \sigma^2 = \frac{1}{n} \text{ Corrected } \Sigma x_i^2 - (\text{Corrected } \bar{x})^2$$

$$= \frac{1}{97}(39694) - (20)^2$$

$= 409.216 - 400 = 9.216$

$\therefore$ Corrected $\sigma = \sqrt{9.216} = 3.036.$