Statistics

Mean Deviation for Ungrouped Data

- Mean deviation about the mean ($\bar{x}$) of an ungrouped data x_1, x_2, x_3 ... x_n is given by the formula $\text{M.D.}(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$, where $\bar{x}$ is the mean of x_1, x_2, x_3 .... x_n.
- Mean deviation about the median (M) of an ungrouped data x_1, x_2, x_3 .... x_n is given by the formula $\text{M.D.}(\text{M}) = \frac{1}{n}\sum_{i=1}^{n}|x_i - \text{M}|$, where M is the median of x_1, x_2, x_3 .... x_n.
-

Example 1: Find the mean deviation about the median of the following data, if its mean deviation about the mean is $9.1\overline{6}$.

9, 17, 13, 19, 26, 31, 53, 42, *x*, 18, 21, 23

Where, the mean of this data lies between 23 and 26, and *x* lies between 27 and 31.

Solution:

To find the mean deviation about the median of the given data, first of all, we have to find the value of *x*.

Now, the mean of this data is given by

$$\bar{x} = \frac{9+17+13+19+26+31+53+42+x+18+21+23}{12} = \frac{272+x}{12}$$

Since mean ($\bar{x}$) of the given data lies between 23 and 26,

1. For each x_i less than or equal to 23, $|x_i - \bar{x}| = \bar{x} - x_i$
2. For each x_i greater than or equal to 26, $|x_i - \bar{x}| = x_i - \bar{x}$

Hence, the mean deviation about the mean is given by

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

$$= \frac{1}{12}\left[\begin{aligned}&\left|9-\left(\frac{272+x}{12}\right)\right| + \left|17-\left(\frac{272+x}{12}\right)\right| + \left|13-\left(\frac{272+x}{12}\right)\right| + \left|19-\left(\frac{272+x}{12}\right)\right| + \left|26-\left(\frac{272+x}{12}\right)\right| \\ &+\left|31-\left(\frac{272+x}{12}\right)\right| + \left|53-\left(\frac{272+x}{12}\right)\right| + \left|42-\left(\frac{272+x}{12}\right)\right| + \left|x-\left(\frac{272+x}{12}\right)\right| + \left|18-\left(\frac{272+x}{12}\right)\right| \\ &+\left|21-\left(\frac{272+x}{12}\right)\right| + \left|23-\left(\frac{272+x}{12}\right)\right|\end{aligned}\right]$$

$$= \frac{1}{12}\left[\begin{aligned}&\left\{\left(\frac{272+x}{12}\right)-9\right\} + \left\{\left(\frac{272+x}{12}\right)-17\right\} + \left\{\left(\frac{272+x}{12}\right)-13\right\} + \left\{\left(\frac{272+x}{12}\right)-19\right\} + \left\{26-\left(\frac{272+x}{12}\right)\right\} \\ &+\left\{31-\left(\frac{272+x}{12}\right)\right\} + \left\{53-\left(\frac{272+x}{12}\right)\right\} + \left\{42-\left(\frac{272+x}{12}\right)\right\} + \left\{x-\left(\frac{272+x}{12}\right)\right\} + \left\{\left(\frac{272+x}{12}\right)-18\right\} \\ &+\left\{\left(\frac{272+x}{12}\right)-21\right\} + \left\{\left(\frac{272+x}{12}\right)-23\right\}\end{aligned}\right]$$

$$= \frac{1}{12}\left[\frac{1}{12}\left\{\begin{aligned}&(164+x)+(68+x)+(116+x)+(44+x)+(40-x)+(100-x)+(364-x)+(232-x) \\ &+(11x-272)+(56+x)+(x+20)+(x-4)\end{aligned}\right\}\right]$$

$$= \frac{928+14x}{144}$$

It is given that $\text{M.D.}(\bar{x}) = 9.1\overline{6}$

$$\Rightarrow \frac{928+14x}{144} = 9.1\overline{6} = \frac{55}{6}$$

$$\Rightarrow 5568 + 84x = 7920$$

$$\Rightarrow 84x = 7920 - 5568 = 2352$$

$$\Rightarrow x = \frac{2352}{84} = 28$$

Hence, the given data is

9, 17, 13, 19, 26, 31, 53, 42, 28, 18, 21, 23

Arranging this data in the increasing order, we get

9, 13, 17, 18, 19, 21, 23, 26, 28, 31, 42, 53

The median of this data is given by

$$M = \frac{1}{2}(6^{th} \text{ observation} + 7^{th} \text{ observation}) = \frac{1}{2}(21+23) = 22$$

So, the mean deviation about the median is given by

$$\begin{aligned} M.D.(M) &= \frac{1}{12}\sum_{i=1}^{12}\left|x_i - M\right| \\ &= \frac{1}{12}\begin{bmatrix} |9-22|+|13-22|+|17-22|+|18-22|+|19-22|+|21-22|+|23-22|+|26-22| \\ +|28-22|+|31-22|+|42-22|+|53-22| \end{bmatrix} \\ &= \frac{1}{12}[13+9+5+4+3+1+1+4+6+9+20+31] \\ &= \frac{106}{12} \\ &= \frac{53}{6} \\ &= 8.83 \end{aligned}$$

Example 2: For what value of a natural number n are the values corresponding to 25 times the mean deviation about the median of the first $(2n - 1)$ even natural numbers and 24 times the mean deviation about the mean of the first $2n$ odd natural numbers equal?

Solution:

The first $(2n - 1)$ even natural numbers are

2, 4, 6, 8 …… $2(2n - 1)$

Since n is a natural number, $(2n - 1)$ is an odd number.

The median (M) of the data 2, 4, 6, 8 …… $2(2n - 1)$ is given by

$$M = \left[\frac{(2n-1)+1}{2}\right]^{th} \text{ observation} = n^{th} \text{ observation} = 2+(n-1)\times 2 = 2n$$

Hence, the mean deviation about the median of the first $(2n - 1)$ even natural numbers is given by

$$\text{M.D.}(\text{M}) = \frac{1}{2n-1}\sum_{i=1}^{2n-1}|x_i - \text{M}|$$

$$= \frac{1}{2n-1}\begin{bmatrix}|2-2n|+|4-2n|+\ldots+|2(n-2)-2n|+|2(n-1)-2n|+|2n-2n|+|2(n+1)-2n| \\ +|2(n+2)-2n|+\ldots+|2(2n-2)-2n|+|2(2n-1)-2n|\end{bmatrix}$$

$$= \frac{1}{2n-1}\times\left[(2n-2)+(2n-4)+\ldots+4+2+0+2+4+\ldots+(2n-4)+(2n-2)\right]$$

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

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

$$= \frac{4}{2n-1}\times\frac{n(n-1)}{2}$$

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

The first $2n$ odd natural numbers are

1, 3, 5 ..... $4n - 1$

The mean ($\bar{x}$) of these numbers is given by

$$\bar{x} = \frac{1+3+5+\ldots.+(4n-1)}{2n}$$

$$= \frac{(2\times1-1)+(2\times2-1)+(2\times3-1)+\ldots.+(2\times2n-1)}{2n}$$

$$= \frac{2(1+2+3+\ldots.+2n)-2n}{2n}$$

$$= \frac{\frac{2\times2n(2n+1)}{2}-2n}{2n}$$

$$= 2n$$

Hence, the mean deviation about the mean of the first $2n$ odd natural numbers is given by

$$\text{M.D.}(\bar{x}) = \frac{1}{2n}\sum_{i=1}^{2n}|x_i - \bar{x}|$$

$$= \frac{1}{2n}\left[\begin{array}{l}|1-2n|+|3-2n|+\ldots+|(2n-3)-2n|+|(2n-1)-2n|+|(2n+1)-2n|+|(2n+3)-2n| \\ +\ldots+|(4n-3)-2n|+|(4n-1)-2n|\end{array}\right]$$

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

$$= \frac{1}{2n}\times 2\times\left[1+3+5+\ldots(2n-1)\right]$$

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

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

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

$$= n$$

It is given that

$24 \times \text{M.D.}(\bar{x}) = 25 \times \text{M.D}(M)$

$$\Rightarrow 24n = 25\times\frac{2n(n-1)}{2n-1}$$

$$\Rightarrow 48n^2 - 24n = 50n^2 - 50n$$

$$\Rightarrow 48n^2 - 50n^2 - 24n + 50n = 0$$

$$\Rightarrow -2n^2 + 26n = 0$$

$$\Rightarrow -2n(n-13) = 0$$

$$\Rightarrow n = 0 \text{ or } 13$$

Since n is a natural number, $n = 13$.

Mean Deviation for Discrete Frequency Distribution

Mean Deviation about the Mean

- A discrete frequency distribution is given as

Value (x_i)	x_1	x_2	…	x_n

Frequency (f_i)	f_1	f_2	…	f_n

The mean deviation about the mean ($\bar{x}$) for such a discrete frequency distribution is calculated by using the formula $\text{M.D.}(\bar{x}) = \frac{1}{N}\sum_{i=1}^{n} f_i |x_i - \bar{x}|$, where N = $\sum_{i=1}^{n} f_i$ and $\bar{x}$ is the mean of the discrete frequency distribution.

- The mean ($\bar{x}$) of the data can be calculated by any of the following methods: direct method, assumed-mean method or step-deviation method.

Mean Deviation about the Median

- A discrete frequency distribution is given as

Value (x_i)	x_1	x_2	…	x_n
Frequency (f_i)	f_1	f_2	…	f_n

- The mean deviation about the median (M) for such a discrete frequency distribution is calculated by using the formula $\text{M.D.}(\text{M}) = \frac{1}{N}\sum_{i=1}^{n} f_i |x_i - \text{M}|$, where N = $\sum_{i=1}^{n} f_i$ and M is the median of the discrete frequency distribution.

Example 1: The following frequency distribution shows the marks obtained by 50 students in a 30-marks test.

Marks Obtained	9	11	12	16	18	25	27
No. of students	7	8	-	-	3	6	4

If the median of the marks obtained by the students is 14, then find the mean deviation of the marks obtained about the mean and median.

Solution:

Let the number of students corresponding to marks obtained as 12 and 16 be *x* and *y* respectively.

The cumulative frequency table of the distribution will be

Marks Obtained (x_i)	**No. of Students (f_i)**	**c.f.**
9	7	7
11	8	15
12	x	15 + x
16	y	15 + x + y
18	3	18 + x + y
25	6	24 + x + y
27	4	28 + x + y
Total	**28 + x + y**	-

Since there are 50 students,

$28 + x + y = 50$

$\Rightarrow x + y = 22$ ... (1)

Since there are 50 students, the median of the given frequency distribution will be

M = $\frac{1}{2}$ (25^{th} observation + 26^{th} observation)

Since M = 14, and 14 is the mean of 12 and 16, the 25^{th} and 26^{th} observations are 12 and 16 respectively.

$\therefore 15 + x = 25$

$\Rightarrow x = 10$

Substituting $x = 10$ in equation (1), we get

$y = 12$

We know the median of the discrete frequency distribution, i.e., M = 14. Now, we can find the mean deviation of the marks obtained by the students about the median as follows:

Marks Obtained (x_i)	**No. of Students (f_i)**	$\lvert x_i - M \rvert$	$f_i \lvert x_i - M \rvert$
9	7	5	35
11	8	3	24
12	10	2	20
16	12	2	24
18	3	4	12
25	6	11	66
27	4	13	52
Total	**50**	**-**	**233**

$$\therefore \text{M.D.(M)} = \frac{\sum_{i=1}^{n} f_i \lvert x_i - M \rvert}{\sum_{i=1}^{n} f_i} = \frac{233}{50} = 4.66$$

Thus, the mean deviation of the marks obtained by the students about the median is 4.66.

To find the mean deviation of the marks obtained by the students about the mean, first of all, we have to find the mean marks obtained by the students.

Marks Obtained (x_i)	No. of Students (f_i)	f_ix_i
9	7	63
11	8	88
12	10	120
16	12	192
18	3	54
25	6	150
27	4	108
Total	**50**	**775**

Mean $(\bar{x}) = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{775}{50} = 15.5$

Now,

Marks Obtained (x_i)	No. of Students (f_i)	$\left\lvert x_i - \bar{x} \right\rvert$	$f_i \left\lvert x_i - \bar{x} \right\rvert$
9	7	6.5	45.5
11	8	4.5	36
12	10	3.5	35

16	12	0.5	6
18	3	2.5	7.5
25	6	9.5	57
27	4	11.5	46
Total	**50**	**-**	**233**

$$\therefore \text{M.D.}(\bar{x}) = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i} = \frac{233}{50} = 4.66$$

Thus, the mean deviation of the marks obtained by the students about the mean is 4.66.

Example 2: In a given discrete frequency distribution, *n* distinct values 1, 2, 3 ... *n* occur with frequencies 1, 2, 3 ... *n* respectively. If the mean of the distribution is 5, then find the mean deviation of this data about its median.

Solution:

The mean of the given discrete frequency distribution can be calculated as

x_i	f_i	$f_i x_i$
1	1	1^2
2	2	2^2
.	.	.
.	.	.

n	n	n^2
Total	$\sum_{i=1}^{n} f_i = \frac{n(n+1)}{2}$	$\sum_{i=1}^{n} f_i x_i = \frac{n(n+1)(2n+1)}{6}$

Now,

$$\sum_{i=1}^{n} f_i = (1+2+3 \ldots + n) = \frac{n(n+1)}{2}$$ and

$$\sum_{i=1}^{n} f_i x_i = 1^2 + 2^2 + 3^2 \ \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

So,

Mean $$(\bar{x}) = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\frac{n(n+1)(2n+1)}{6}}{\frac{n(n+1)}{2}} = \frac{2n+1}{3}$$

It is given that the mean is 5.

$$\therefore \frac{2n+1}{3} = 5$$
$$\Rightarrow n = 7$$

To find the mean deviation about the median, first of all, we need to find the median of the given frequency distribution as follows:

x_i	f_i	c.f.
1	1	1
2	2	3
3	3	6

4	4	10
5	5	15
6	6	21
7	7	28
Total	**28**	-

The median (M) is given by

M = $\frac{1}{2}$ (14th observation + 15th observation) = $\frac{1}{2}$ (5 + 5) = 5

Now, the mean deviation about the median is calculated as follows:

-

x_i	f_i	$\|x_i - M\|$	$f_i\|x_i - M\|$
1	1	4	4
2	2	3	6
3	3	2	6
4	4	1	4
5	5	0	0
6	6	1	6
7	7	2	14
Total	**28**	-	**40**

$$\therefore \text{M.D.}(\text{M}) = \frac{\sum_{i=1}^{n} f_i |x_i - \text{M}|}{\sum_{i=1}^{n} f_i} = \frac{40}{28} = \frac{10}{7} = 1.43$$

Mean Deviation for Continuous Frequency Distribution

Mean Deviation about the Mean

Mean deviation about the mean ($\bar{x}$) for a continuous frequency distribution is calculated by using the formula $\text{M.D.}(\bar{x}) = \frac{1}{\text{N}} \sum_{i=1}^{n} f_i |x_i - \bar{x}|$, where N = $\sum_{i=1}^{n} f_i$, x_i's are the class marks and $\bar{x}$ is the mean of the frequency distribution.

The mean ($\bar{x}$) of the data can be calculated by any of the following methods: direct method, assumed-mean method or step-deviation method.

Mean Deviation about the Median

Mean deviation about the Median (M) for a continuous frequency distribution is calculated by using the formula $\text{M.D.}(\text{M}) = \frac{1}{\text{N}} \sum_{i=1}^{n} f_i |x_i - \text{M}|$, where N = $\sum_{i=1}^{n} f_i$ and x_i's are class marks.

Median, M of the continuous frequency distribution is calculated by using the formula,

Median (M) $= l + \frac{\frac{\text{N}}{2} - \text{C}}{f} \times h$, where median class is the class interval whose cumulative frequency is just greater than or equal to $\frac{\text{N}}{2}$.

l = Lower limit of the median class
N = Total number of observations
C = Cumulative frequency of the class just preceding the median class
f = Frequency of the median class
h = Class size

Example 1: The following table shows the number of students of different height groups, with two missing values.

Height (in cm)	130 – 140	140 – 150	150 – 160	160 – 170	170 – 180
No. of Students	11	-	10	-	7

If the median and the mean of this data are 151 and 152.6 respectively, then find the mean deviations about the mean and the median.

Solution:
Let the number of students of height groups 140 – 150 and 160 – 170 be x and y respectively.

To solve this question, first of all, we have to find the values of x and y.

The median and the mean of the data can be found by using the frequency distribution table as follows:

Height (in cm)	No. of Students (f_i)	c.f.	x_i	$f_i x_i$
130 – 140	11	11	135	1485
140 – 150	x	11 + x	145	145x
150 – 160	10	21 + x	155	1550
160 – 170	y	21 + x + y	165	165y
170 – 180	7	28 + x + y	175	1225
Total	**28 + x + y**	**-**	**-**	**4260 + 145x + 165y**

Since the median (M) is 151, the median class is 150 – 160.

Here, $l = 150$, $N = 28 + x + y$, $C = 11 + x$, $f = 10$, $h = 10$

We know that

Median (M) $= l + \dfrac{\dfrac{N}{2} - C}{f} \times h$

$$\Rightarrow 151 = 150 + \frac{\frac{28+x+y}{2} - (11+x)}{10} \times 10$$

$$\Rightarrow 1 = \frac{6-x+y}{2}$$

$$\Rightarrow x - y = 4$$

$$\Rightarrow x = 4 + y \qquad \dots (1)$$

It is also given that the mean $(\bar{x})$ of the data is 152.6. We know that

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

$$\Rightarrow 152.6 = \frac{4260 + 145x + 165y}{28 + x + y}$$

$$\Rightarrow 152.6 = \frac{4260 + 145(4+y) + 165y}{28 + (4+y) + y} \qquad [\text{Using equation (1)}]$$

$$\Rightarrow 4883.2 + 305.2y = 4840 + 310y$$

$$\Rightarrow 4.8y = 43.2$$

$$\Rightarrow y = 9$$

On substituting y = 9 in equation (1), we obtain

x = 13

Now, we construct the following table to find the mean deviations about the mean and the median.

Height (in cm)	**No. of Students (f_i)**	**x_i**	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$	$\lvert x_i - M \rvert$	$f_i \lvert x_i - M \rvert$
130 – 140	11	135	17.6	193.6	16	176

140 – 150	13	145	7.6	98.8	6	78
150 – 160	10	155	2.4	24.0	4	40
160 – 170	9	165	12.4	111.6	14	126
170 – 180	7	175	22.4	156.8	24	168
Total	**50**	**-**	**-**	**584.8**	**-**	**588**

Now, $$\text{M.D.}(\bar{x}) = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i} = \frac{584.8}{50} = 11.696$$

And $$\text{M.D.}(\text{M}) = \frac{\sum_{i=1}^{n} f_i |x_i - \text{M}|}{\sum_{i=1}^{n} f_i} = \frac{588}{50} = 11.76$$

Variance and Standard Deviation of Ungrouped Data

- The variance of a data is defined as the mean of the squares of deviations from the mean, and it is denoted by the symbol σ^2.
- For an ungrouped data $x_1, x_2, x_3 \ldots x_n$, the variance is calculated by the formula $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, where $\bar{x}$ is the mean of $x_1, x_2, x_3 \ldots x_n$.
- The mean of the data can be calculated by any of the following methods: direct method, assumed-mean method or step-deviation method.
- The square root of the variance is called the standard deviation. It is denoted by the symbol σ and calculated by using the formula $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$.

- The variance can also be found by using another formula, which is given by

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

.

Solved Examples

Example 1: Find the value of a natural number *x*, for which the standard deviation of the following ungrouped data is 4.

15, $2x - 1$, $3x + 2$, 13, 21

Solution: Mean $(\bar{x})$ of the given data is given by

$$\bar{x} = \frac{15 + (2x-1) + (3x+2) + 13 + 21}{5} = \frac{5x+50}{5} = x + 10$$

It is given that standard deviation (σ) is 4.

$\Rightarrow$ Variance = σ^2 = 16

We know that variance of an ungrouped distribution is given by

$$\sigma^2 = \frac{1}{5}\sum_{i=1}^{5}(x_i - \bar{x})^2$$

$$\Rightarrow 16 = \frac{1}{5}\begin{bmatrix}\{15-(x+10)\}^2 + \{(2x-1)-(x+10)\}^2 + \{(3x+2)-(x+10)\}^2 \\ +\{13-(x+10)\}^2 + \{21-(x+10)\}^2\end{bmatrix}$$

$$\Rightarrow 80 = (5-x)^2 + (x-11)^2 + (2x-8)^2 + (3-x)^2 + (11-x)^2$$

$$\Rightarrow 80 = 8x^2 - 92x + 340$$

$$\Rightarrow 4(2x^2 - 23x + 65) = 0$$

$$\Rightarrow 2x^2 - 13x - 10x + 65 = 0$$

$$\Rightarrow x(2x-13) - 5(2x-13) = 0$$

$$\Rightarrow (2x-13)(x-5) = 0$$

$$\Rightarrow x = 5 \text{ or } \frac{13}{2}$$

Since *x* is a natural number, *x* = 5.

Example 2: If the mean of 5 numbers is 15 and their standard deviation is 4, then find the sum of the squares of these 5 numbers.

Solution:
We know that

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

$$\Rightarrow 4^2 = \frac{1}{5}\sum_{i=1}^{5} x_i^2 - 15^2$$

$$\Rightarrow \sum_{i=1}^{5} x_i^2 = 5(16+225) = 1205$$

Thus, the sum of the squares of the 5 numbers is 1205.

Example 3: A data contains 28 observations. If each observation is multiplied by 3 and then increased by 2, then the standard deviation of this data is 14 more than the original. Find the variance of the original 28 observations.

Solution: Let the 28 observations be $x_1, x_2, x_3 \ldots x_{28}$, and $\bar{x}$ be their mean. Let their standard deviation be σ.

$$\sigma = \sqrt{\frac{1}{28}\sum_{i=1}^{28}(x_i - \bar{x})^2},$$

Where, $\bar{x} = \frac{1}{28}\sum_{i=1}^{28} x_i$

If each observation x_i is multiplied by 3 and then 2 is added to each of them, then the new observations y_i are of the form

$$y_i = 3x_i + 2$$

Now,

$$\bar{y} = \frac{1}{28}\sum_{i=1}^{28} y_i = \frac{1}{28}\sum_{i=1}^{28}(3x_i+2) = 3\times\frac{1}{28}\sum_{i=1}^{28} x_i + \frac{1}{28}(2+2+2 \ldots 28 \text{ times}) = 3\bar{x}+2$$

The new standard deviation (σ') is given by

$$\sigma' = \sqrt{\frac{1}{28}\sum_{i=1}^{28}(y_i-\bar{y})^2} = \sqrt{\frac{1}{28}\sum_{i=1}^{28}[(3x_i+2)-(3\bar{x}+2)]^2} = \sqrt{\frac{1}{28}\sum_{i=1}^{28}[3(x_i-\bar{x})]^2} = 3\sqrt{\frac{1}{28}\sum_{i=1}^{28}(x_i-\bar{x})^2} = 3\sigma$$

But it is given that

$$\sigma' = \sigma + 14$$
$$\Rightarrow 3\sigma = \sigma + 14$$
$$\Rightarrow \sigma = 7$$

∴Original variance, $\sigma^2 = 7^2 = 49$

Variance and Standard Deviation of a Discrete Frequency Distribution

- A discrete frequency distribution is given as

Value (x_i)	x_1	x_2	…	x_n
Frequency (f_i)	f_1	f_2	…	f_n

- The variance is calculated by using the formula, $\text{Variance }(\sigma^2) = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2$, where $N = \sum_{i=1}^{n} f_i$ and $\bar{x}$ is the mean of the data.
- The standard deviation is calculated by using the formula, $\text{Standard Deviation }(\sigma) = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2}$, where $N = \sum_{i=1}^{n} f_i$ and $\bar{x}$ is the mean of the data.

- There is one more way to find the variance and the standard deviation of a given discrete frequency distribution. It is known as the shortcut method. In this method the formulae used are

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

- $$\text{Standard Deviation }(\sigma) = \frac{1}{N}\sqrt{N\sum_{i=1}^{n} f_i x_i^2 - \left(\sum_{i=1}^{n} f_i x_i\right)^2}$$

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

Example 1: If the mean and the variance of the following distribution are 18 and 17 respectively, then find the values of f_1 and f_2.

x_i	10	15	18	20	25
f_i	3	2	f_1	f_2	2

Solution:

The mean and the variance of the given discrete frequency distribution can be calculated as follows:

x_i	f_i	$f_i x_i$	x_i^2	$f_i x_i^2$
10	3	30	100	300
15	2	30	225	450
18	f_1	$18f_1$	324	$324f_1$
20	f_2	$20f_2$	400	$400f_1$
25	2	50	625	1250
Total	**$7 + f_1 + f_2$**	**$110 + 18f_1 + 20f_2$**	**-**	**$2000 + 324f_1 + 400f_2$**

It is given that the mean of this data is 18.

$$\therefore \frac{\sum_{i=1}^{5} f_i x_i}{\sum_{i=1}^{5} f_i} = \frac{110 + 18f_1 + 20f_2}{7 + f_1 + f_2} = 18$$

$$\Rightarrow 110 + 18f_1 + 20f_2 = 126 + 18f_1 + 18f_2$$

$$\Rightarrow 2f_2 = 16$$

$$\Rightarrow f_2 = 8$$

$$\text{Now, N} = \sum_{i=1}^{5} f_i = 7 + f_1 + f_2 = 7 + f_1 + 8 = 15 + f_1$$

$$\sum_{i=1}^{5} f_i x_i = 110 + 18f_1 + 20f_2 = 110 + 18f_1 + 20 \times 8 = 270 + 18f_1$$

$$\sum_{i=1}^{5} f_i x_i^2 = 2000 + 324f_1 + 400f_2 = 2000 + 324f_1 + 400 \times 8 = 5200 + 324f_1$$

It is also given that the variance of this data is 17.

$$\therefore \frac{1}{N^2}\left[N\sum_{i=1}^{5} f_i x_i^2 - \left(\sum_{i=1}^{5} f_i x_i\right)^2\right] = 17$$

$$\Rightarrow \frac{1}{(15+f_1)^2}\left[(15+f_1)(5200+324f_1) - (270+18f_1)^2\right] = 17$$

$$\Rightarrow \left(78000 + 4860f_1 + 5200f_1 + 324f_1^2\right) - \left(72900 + 324f_1^2 + 9720f_1\right) = 17\left(225 + f_1^2 + 30f_1\right)$$

$$\Rightarrow 17f_1^2 + 170f_1 - 1275 = 0$$

$$\Rightarrow 17(f_1^2 + 10f_1 - 75) = 0$$

$$\Rightarrow f_1^2 + 10f_1 - 75 = 0$$

$$\Rightarrow (f_1 + 15)(f_1 - 5) = 0$$

$$\Rightarrow f_1 = -15 \text{ or } 5$$

Since frequency is a non-negative integer, $f_1 = 5$.

Example 2: If the mean of the following frequency distribution is 9, then find the value of *n* and the standard deviation of this data.

x_i	1	2	3	...	r	...	2n – 2	2n – 1
f_i	2n – 1	2n – 2	2n – 3	...	2n – r	...	2	1

Solution:

The mean and the standard deviation of the given data can be calculated as follows:

x_i	f_i	$f_i x_i$	x_i^2	$f_i x_i^2$
1	2n – 1	$2n \times 1 - 1^2$	1^2	$2n \times 1^2 - 1^3$

2	2n – 2	$2n \times 2 - 2^2$	2^2	$2n \times 2^2 - 2^3$
3	2n – 3	$2n \times 3 - 3^2$	3^2	$2n \times 3^2 - 3^3$
...	...	...	...	...
r	2n – r	$2n \times r - r^2$	r^2	$2n \times r^2 - r^3$
...	...	...	...	...
2n – 2	2 = 2n – (2n – 2)	$2n\,(2n-2) - (2n-2)^2$	$(2n-2)^2$	$2n\,(2n-2)^2 - (2n-2)^3$
2n – 1	1 = 2n – (2n – 1)	$2n\,(2n-1) - (2n-1)^2$	$(2n-1)^2$	$2n\,(2n-1)^2 - (2n-1)^3$
Total	$\sum_{r=1}^{2n-1} r$	$\sum_{r=1}^{2n-1} (2nr - r^2)$	-	$\sum_{r=1}^{2n-1} (2nr^2 - r^3)$

$$\text{Now, N} = \sum_{i=1}^{2n-1} f_i = \sum_{r=1}^{2n-1} r = [1+2+3+\ldots+(2n-1)] = \frac{(2n-1)(2n-1+1)}{2} = n(2n-1)$$

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

$$= 2n\sum_{r=1}^{2n-1} r - \sum_{r=1}^{2n-1} r^2$$

$$= 2n \times n(2n-1) - [1^2 + 2^2 + 3^2 + \ldots + (2n-1^2)]$$

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

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

$$= \frac{1}{3}n(2n-1)(2n+1)$$

$$\sum_{i=1}^{2n-1} f_i x_i^2 = \sum_{r=1}^{2n-1}(2nr^2 - r^3)$$

$$= 2n\sum_{r=1}^{2n-1} r^2 - \sum_{r=1}^{2n-1} r^3$$

$$= 2n \times \frac{(2n-1)(2n)(4n-1)}{6} - [1^3 + 2^3 + 3^3 + \ldots + (2n-1^3)]$$

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

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

$$= \frac{1}{3}n^2(2n-1)(2n+1)$$

It is given that the mean of the given frequency distribution is 9.

$$\therefore \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\frac{1}{3}n(2n-1)(2n+1)}{n(2n-1)} = 9$$

$$\Rightarrow 2n+1 = 27$$

$$\Rightarrow n = 13$$

$$\therefore \text{Variance } (\sigma^2) = \frac{1}{N^2}\left[N\sum_{i=1}^{n} f_i x_i^2 - \left(\sum_{i=1}^{n} f_i x_i\right)^2\right]$$

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

$$= \frac{n(2n+1)}{3} - \frac{(2n+1)^2}{9}$$

$$= \frac{13 \times 27}{3} - \frac{27^2}{9}$$

$$= 36$$

$$\therefore \text{ Standard Deviation } (\sigma) = \sqrt{36} = 6$$

Variance and Standard Deviation for a Continuous Frequency Distribution

The variance (σ^2) and the standard deviation (σ) for a continuous frequency distribution (with uniform class size) are given by the formulae

- $\text{Variance } (\sigma^2) = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2$, where $N = \sum_{i=1}^{n} f_i$, x_i's are the class marks of intervals and $\bar{x}$ is the mean of the distribution

- $\text{Standard Deviation } (\sigma) = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}$, where $N = \sum_{i=1}^{n} f_i$, x_i's are the class marks of intervals and $\bar{x}$ is the mean of the distribution

There is one more way to find the variance and the standard deviation of a continuous frequency distribution. It is known as the shortcut method. The formulae used in this method are

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

- $\text{Standard Deviation } (\sigma) = \frac{h}{N}\sqrt{N\sum_{i=1}^{n} f_i y_i^2 - \left(\sum_{i=1}^{n} f_i y_i\right)^2}$

Here, $N = \sum_{i=1}^{n} f_i$, $y_i = \frac{x_i - a}{h}$, a is the assumed mean, h is the class size and x_i's are the class marks.

Example 1:

The mean and the median of the following frequency distribution with missing frequencies in intervals 11 – 19 and 29 – 37 are 25.125 and 24 respectively.

Class Intervals	2 – 10	11 – 19	20 – 28	29 – 37	38 – 46
Frequency	3	-	16	-	9

Find the missing frequencies of the class intervals 11 – 19 and 29 – 37. Also, find the variance and the standard deviation for the above frequency distribution.

Solution:

Let the frequencies of the class intervals 11 − 19 and 29 − 37 be x and y respectively.

The given frequency distribution is not continuous.

Now, 11 − 10 = 1 and $\frac{1}{2} = 0.5$

In order to make the given frequency distribution continuous, we have to subtract 0.5 from the lower limits, and add 0.5 to the upper limits of each class interval. Now, we construct a table as follows:

Class	**Frequency (f_i)**	**c.f.**	**Class Marks (x_i)**	**$f_i x_i$**
1.5 − 10.5	3	3	6	18
10.5 − 19.5	x	3 + x	15	15x
19.5 − 28.5	16	19 + x	24	384
28.5 − 37.5	y	19 + x + y	33	33y
37.5 − 46.5	9	28 + x + y	42	378
Total	**28 + x + y**	**-**	**-**	**780 + 15x + 33y**

Since the median (M) is 24, it lies in the interval 19.5 − 28.5. So, the median class is 19.5 − 28.5.

Here,

l = Lower limit of the median class = 19.5

N = Total number of observations = 28 + x + y

C = Cumulative frequency of the class preceding the median class = 3 + x

f = Frequency of the median class = 16

h = Class size = 9

We know that

Median (M) $= l + \dfrac{\dfrac{N}{2} - C}{f} \times h$

$$\Rightarrow 24 = 19.5 + \frac{\frac{28+x+y}{2} - (3+x)}{16} \times 9$$
$$\Rightarrow \frac{16}{9}(24 - 19.5) = \frac{28+x+y-6-2x}{2}$$
$$\Rightarrow 8 = \frac{22-x+y}{2}$$
$$\Rightarrow x - y = 6$$
$$\Rightarrow x = 6 + y \qquad \ldots (1)$$

It is also given that the mean $(\bar{x})$ of the data is 25.125.

$$\therefore \bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = 25.125$$
$$\Rightarrow \frac{780+15x+33y}{28+x+y} = \frac{25125}{1000}$$
$$\Rightarrow \frac{780+15(6+y)+33y}{28+(6+y)+y} = \frac{201}{8} \qquad \text{[Using equation (1)]}$$
$$\Rightarrow 6960 + 384y = 6834 + 402y$$
$$\Rightarrow -18y = -126$$
$$\Rightarrow y = 7$$

On substituting $y = 7$ in equation (1), we obtain

$x = 13$

To find the variance and standard deviation of the given frequency distribution by the shortcut method, let us make the following table, where A = assumed mean = 24 and $h = 9$.

Class	Frequency (f_i)	Class Marks (x_i)	$y_i = \frac{x_i - 24}{9}$	y_i^2	$f_i y_i$	$f_i y_i^2$
1.5 – 10.5	3	6	−2	4	−6	12
10.5 – 19.5	13	15	−1	1	−13	13
19.5 – 28.5	16	24	0	0	0	0
28.5 – 37.5	7	33	1	1	7	7
37.5 – 46.5	9	42	2	4	18	36
Total	**N = 48**	-	-	-	**6**	**68**

$$\therefore \text{Variance } (\sigma^2) = \frac{h^2}{N^2}\left[N\sum_{i=1}^{n} f_i y_i^2 - \left(\sum_{i=1}^{n} f_i y_i\right)^2\right] = \frac{9^2}{48^2}\left[48\times 68 - 6^2\right] = 113.484 \text{ (approx.)}$$

$\Rightarrow$ Standard Deviation (σ) = $\sqrt{\text{Variance}} = \sqrt{113.484} = 10.652$ (approx.)

Coefficient of Variation

- The coefficient of variation (C.V.) of a data is defined as $\text{C.V.} = \frac{\sigma}{\bar{x}} \times 100$, where σ and $\bar{x}$ are respectively the standard deviation and the mean of the data.
- In the comparison of two or more frequency distributions, the distribution having the greater C.V. is said to be more variable or dispersed than the other, whereas the distribution having the lesser C.V. is said to be more consistent.

Example 1: The mean and the coefficient of variation of five numbers are 15 and $26.\overline{6}$ respectively. If three of the numbers are 9, 15 and 21, then find the other two numbers.

Solution:

Let the remaining two numbers be *x* and *y*.

It is given that the mean of the five numbers is 15.

$$\therefore \frac{1}{5}(9+15+21+x+y)=15$$
$$\Rightarrow x+y=75-45=30$$
$$\Rightarrow y=30-x \qquad \ldots (1)$$

It is given that the coefficient of variation is $26.\overline{6}$.

$$\frac{\sigma}{\overline{x}}\times 100=26.\overline{6}$$
$$\Rightarrow \frac{\sigma}{15}\times 100=\frac{80}{3}$$
$$\Rightarrow \sigma=\frac{15\times 80}{100\times 3}=4$$
$$\Rightarrow \sigma^2=16$$
$$\Rightarrow \frac{1}{5}\sum_{i=1}^{5}x_i^{\,2}-\overline{x}^{\,2}=16$$
$$\Rightarrow \frac{1}{5}(9^2+15^2+21^2+x^2+y^2)-15^2=16$$
$$\Rightarrow 81+225+441+x^2+y^2=5(225+16)=1205$$
$$\Rightarrow x^2+y^2=1205-747=458$$
$$\Rightarrow x^2+(30-x)^2=458$$
$$\Rightarrow x^2+(900+x^2-60x)=458$$
$$\Rightarrow 2(x^2-30x+221)=0$$
$$\Rightarrow x^2-30x+221=(x-13)(x-17)=0$$
$$\Rightarrow x=13 \text{ or } 17$$

If x = 13, then y = 17, and if x = 17, then y = 13.

Thus, the remaining two numbers are 13 and 17.

Example 2: In a school,the ratio of the number of girls to the number of boys is 2:3. During a certain observation, it is found that the mean age of the girls and the mean age of the students are 13.5 years and 14.4 years respectively. If the ratio of variance of the girls to that of the boys is 9:25, then find the ratio of the coefficient of variation of the girls to that of the boys. Between the girls and the boys, who show a greater variability?

Solution:

Let the number of girls and the number of boys in the school be $2n$ and $3n$ respectively.

∴Total number of students in the school = $2n + 3n = 5n$

Let σ_1 and $\overline{x}_1$ respectively be the standard deviation and the mean age of the girls.

Let σ_2 and $\overline{x}_2$ respectively be the standard deviation and the mean age of the boys.

It is given that $\overline{x}_1$ = 13.5 years

It is also given that

$$\frac{\sigma_1^2}{\sigma_2^2} = \frac{9}{25}$$
$$\Rightarrow \frac{\sigma_1}{\sigma_2} = \frac{3}{5}$$

Also, mean age of the students is 14.4 years.

$$\therefore \frac{2n\overline{x}_1 + 3n\overline{x}_2}{5n} = 14.4$$
$$\Rightarrow \frac{2\times 13.5 + 3\overline{x}_2}{5} = 14.4$$
$$\Rightarrow 3\overline{x}_2 = 72 - 27 = 45$$
$$\Rightarrow \overline{x}_2 = 15$$

Ratio of coefficient of variation of the girls ($C.V_1$) to the coefficient of variation of the boys ($C.V_2$) is given as

$$\frac{C.V_1}{C.V_2} = \frac{\frac{\sigma_1}{\overline{x}_1}\times 100}{\frac{\sigma_2}{\overline{x}_2}\times 100} = \frac{\sigma_1}{\sigma_2}\times\frac{\overline{x}_2}{\overline{x}_1} = \frac{3}{5}\times\frac{15}{13.5} = \frac{3}{5}\times\frac{10}{9} = \frac{2}{3}$$

In the above expression, we have

$$\frac{C.V_1}{C.V_2} = \frac{2}{3}$$
$$\Rightarrow C.V_2 = \frac{3}{2}C.V_1$$

Clearly, coefficient of variation of the boys is more than that of the girls. So, the boys show greater variability.