MEAN DEVIATION

The mean deviation of a distribution is the average of the absolute deviations of the distribution's terms from its statistical mean (which can be the arithmetic mean, median, or mode).

- Mean deviation can be computed using any measure of central tendency. Nevertheless, mean deviation from the mean and median are frequently employed in statistical investigations.
- The mean deviation around the median is minimized.

1 Mean deviation for ungrouped data

Consider $x_1, x_2, x_3,\ x_n$ as n values of a variable X, and let k be the statistical mean (Arithmetic Mean, Median, Mode) for which we need to determine the mean deviation. The mean deviation (M.D.) around k is calculated by:

$$\text{M.D. }(k) = \frac{|x_1-k|+|x_2-k|+|x_3-k|+\cdots..+|x_n-k|}{n} = \frac{\sum_{i=1}^{n}|x_i-k|}{n}$$

Ex. Determine the mean deviation around the mean for the given data:

12, 3, 18, 17, 4, 9, 17, 19, 20, 15, 8, 17, 2, 3, 16, 11, 3, 1, 0, 5

Sol. We have, $\bar{x} = \frac{1}{20}\sum_{i=1}^{20} x_i = \frac{200}{20} = 10$

The absolute values of the deviations from the mean, specifically, $\left|x_i - \bar{x}\right|$ are

2,7,8,7,6,1,7,9,10,5,2,7,8,7,6,1,7,9,10,5

$$\sum_{i=1}^{20}|x_i - \bar{x}| = 124$$

$$\text{M.D. }(\bar{x}) = \frac{124}{20} = 6.2$$

Ex. The marks obtained by 9 students in an examination are as follows. Calculate their mean deviation from the median. 49,68,21,32,54,38,41,66,59

Sol. Organizing the observations in ascending order,

We have, 21,32,38,41,49,54,59,66,68

Number of observations = 9

Median = 5^{th} term = 49

Calculation of mean deviation:

x_i	$\lvert d_i\rvert = \lvert x_i - 49\rvert$
21	28
32	17
38	11
41	8
49	0
54	5
59	10
66	17
68	19
Total	**115**

$$\text{M.D. } = \frac{1}{n}\sum|d_i| = \frac{115}{9} = 12.78$$

Ex. If $\bar{x}$ is the mean and the mean deviation from the mean is M. D$(\bar{x})$, determine the number of observations lying between $\bar{x}$ – M.D. $(\bar{x})$ And $\bar{x}$ + M.D. $(\bar{x})$ For the following data:

34,66,30,38,44,50,40,60,42,51.

Sol. The provided data, arranged in ascending order, is:

$$30,34,38,40,42,44,50,51,60,66$$

$$\overline{x} = \frac{30+34+38+40+42+44+50+51+60+66}{10}$$

$$\overline{x} = \frac{455}{10} = 45.5$$

$$|x_i - \overline{x}| = 15.5,11.5,7.5,5.5,3.5,1.5,4.5,5.5,14.5,20.5$$

Mean deviation from the mean

$$M.D.(\overline{x}) = \frac{\sum|(x_i - \overline{x})|}{10}$$

$$= \left(\frac{1}{10}\right)[15.5 + 11.5 + 7.5 + 5.5 + 3.5 + 1.5 + 4.5 + 5.5 + 14.5 + 20.5]$$

$$= \frac{1}{10}[90.0] = 9$$

$$\overline{x} - M.D.(\overline{x}) = 45.5 - 9 = 36.5$$

$$\overline{x} + M.D.(\overline{x}) = 45.5 + 9 = 54.5$$

Given observations which lie between 36.5 and 54.5 are 38,40,42,44,50,51,, which are six in number.

Six entries of given data lie between $\overline{x}$ – M.D. ($\overline{x}$) and $\overline{x}$ + M.D. ($\overline{x}$)

2 Mean Deviation For Grouped Data

(a) Discrete Frequency Distribution:

Consider $X_1, X_2, X_3, \ldots\ldots\ldots, Xn_n$ be n observations with corresponding frequencies $f_1, f_2, f_{3,\ldots\ldots,} f_n$ and let K be the statistical mean (Arithmetic Mean, Median, Mode). The mean deviation (M.D.) around K is calculated by:

$$M.D.(k) = \frac{|x_1-k|f_1+|x_2-k|f_2+|x_3-k|f_3+\cdots..+|x_n-k|f_n}{f_1+f_2+f_3+\cdots\ldots+f_n}$$

$$= \frac{\sum_{i=1}^{n}|x_i-k|f_i}{\sum_{i=1}^{n}f_i} = \frac{\sum_{i=1}^{n}|d_i|f_i}{N},$$

Where $d_i = x_i - k$ and $N = \sum_{i=1}^{n} f_i =$ total frequency

The mean of given discrete frequency distribution is given by

$$\overline{x} = \frac{f_1x_1+f_2x_2+f_3x_3+\cdots..+f_nx_n}{f_1+f_2+f_3+\cdots..+f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}$$

To determine the median of a given discrete frequency distribution, arrange the observations in ascending order. Subsequently, calculate the cumulative frequencies. Identify the observation whose cumulative frequency is equal to or just greater than $\frac{N}{2}$, where N is the total number of observations. This identified observation, situated in the middle of the data, represents the required median.

Ex. Calculate the mean deviation from the median for the given frequency distribution:

Age (in years)	10	12	15	18	21	23
Frequency	3	5	4	10	8	4

Sol. Calculating the mean deviation from the median.

Age(in years) x_i	Frequency (f_i)	Cumulative frequency	$f_i x_i$	$\|x_i - 18\|$	$F_i \|x_i - 18\|$
10	3	3	30	8	24
12	5	8	60	6	30
15	4	12	60	3	12
18	10	22	180	0	0
21	8	30	168	3	24
23	4	34	92	5	20
Total	N = 34				110

Here, $N = \sum f_i = 34, \frac{N}{2} = 17$ and $\frac{N}{2} + 1 = 18$

$$\text{Median} = \frac{(\text{ value of 17 th term }) + (\text{ value of 18 th term })}{2}$$

$$= \frac{18+18}{2} = 18 \text{ years}$$

Mean deviation from the median $= \frac{\sum f_i |x_i - 18|}{N}$

$$= \frac{110}{34} = 3.23 \text{ years}$$

Ex. The mean of 4, 7, 2, 8, 6, and a is 7. Determine the mean deviation about the median of these observations.

Sol. In this case, the count of observations, denoted as n, is 6.

$$\frac{4+7+2+8+6+a}{6} = 7$$

$$27 + a = 42$$

$$a = 15$$

Arranging the observations in ascending order, we get 2, 4, 6, 7, 8, 15

Median, k = mean of $\frac{n}{2}^{th}$ observation

And $(\frac{n}{2} + 1)^{th}$ observation $= \frac{3 \text{ rd observations } + 4 \text{ th observation}}{2}$

$$= \frac{6+7}{2} = 6.5$$

Calculation of mean:.

x_i	$\lvert x_i - k \rvert$
2	4.5
4	2.5
6	0.5
7	0.5
8	1.5
15	8.5
Total	**18**

Mean deviation about median $= \frac{\sum |x_i - k|}{n} = \frac{18}{6} = 3$

Ex. Determine the mean deviation around the mean for the provided data.

x_i	1	4	9	12	13	14	21	22
f_i	3	4	5	2	4	5	4	3

Sol. To calculate the mean and subsequently find the deviation around the mean, we create the following table:

X_i	f_i	$x_i f_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
1	3	3	11	33
4	4	16	8	32
9	5	45	3	15
12	2	24	0	0
13	4	52	1	4
14	5	70	2	10
21	4	84	9	36
22	3	66	10	30
Total	**30**	**360**		**160**

Mean $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{360}{30} = 12$

M.D. (about mean) $= \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i} = \frac{160}{30} = 5.33$

(b) Continuous Frequency Distribution:

The mean of a continuous frequency distribution is computed under the assumption that the frequency in each class is cantered at its midpoint.

If x_i represents the mid-value of the i^{th} class, fi is the frequency of the i^{th} class, and k is the statistical mean Arithmetic Mean, Median, Mode, then the mean deviation (M.D.) around k is expressed as:

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

$$d_i = x_i - k$$

$$N = \sum_{i=1}^{n} f_i = \text{ total frequency}$$

Shortcut method for calculating mean deviation about mean

For the provided continuous frequency distribution, the arithmetic mean can be computed using the shortcut (step-deviation) method. The remaining steps of the procedure remain unchanged. In this approach,

1. Select a presumed mean (either the midpoint or a value close to it).

2. Compute the deviations of the observations (or midpoints of classes) from the chosen assumed mean.

3. If all the deviations share a common factor, divide each of them by this common factor to simplify the deviations.

4. The expression for the arithmetic mean when employing the step-deviation method is now formulated as:

$$\bar{x} = a + \frac{\sum_{i=1}^{n} f_i d_i}{N} \times h,$$

Where $$d_i = \frac{x_i - a}{h}$$

a = assumed mean,

h = common factor

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

To calculate the median for a continuous frequency distribution

1. Compute. $\frac{N}{2}$, $(N = \sum_{i=1}^{n} f_i)$.

2. The class that corresponds to a cumulative frequency just exceeding $\frac{N}{2}$ is identified as the median class.

3. Median$= l + \frac{h}{f}(\frac{N}{2} - c)$,,

Where, l = lower limit of median class

f = frequency of the median class

h = width of the median class

c = cumulative frequency of the class just preceding the median class

Ex. Calculate the mean deviation from the mean for the given dataset:

Classes	Frequency
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. We create the table as follows:

Classes	x_i	$d_i = \frac{x_i - 350}{100}$	f_i	$f_i d_i$	$\|x_i - k\|$	$f_i \|x_i - k\|$
0-100	50	–3	4	–12	308	1232
100-200	150	–2	8	–16	208	1664
200-300	250	–1	9	–9	108	972
300-400	350	0	10	0	8	80
400-500	450	1	7	7	92	644
500-600	550	2	5	10	192	960
600-700	650	3	4	12	292	1168
700-800	750	4	3	12	392	1176
Total			$N = \sum f_i = 50$	$\sum f_i d_i = 4$		7896

Actual mean $k = A + \frac{\sum f_i d_i}{\sum f_i} \times h,$

Where, A = assumed mean and h = class interval

$$k = 350 + \frac{4}{50} \times 100 = 358$$

Now, mean deviation,

$$\text{M.D. }(k) = \frac{\sum f_i |x_i - k|}{\sum f_i} = \frac{7896}{50} = 157.92$$

Ex. Calculate the mean deviation of the given distribution from the median.

Classes	10-20	20-30	30-40	40-50	50-60	60-70
Frequencies	10	12	8	16	14	10

Sol. We have

Classes	x_i	f_i	Cumulative frequency	$\| x_i - 43.125\|$	$f_i \| x_i - 43.125\|$
10-20	15	10	10	28.125	281.250
20-30	25	12	22	18.125	217.500
30-40	35	8	30	8.125	65.000
40-50	45	16	46	1.875	300.000
50-60	55	14	60	11.875	166.250
60-70	65	10	70	21.875	218.750
Total		N = 70			1248.750

Here, $N = \sum f_i = 70, \frac{N}{2} = 35$

The class with a cumulative frequency just exceeding 35 is 40-50, making it the median class.

Now, Median $M = l + \frac{\frac{N}{2} - C}{f} \times h,,$

Here, l =40, N = 70, C = 30, h = 10, f = 16

$$M = 40 + \frac{35 - 30}{16} \times 10 = 43.125$$

Mean deviation from median= $\frac{\sum f_i |x_i - 43.125|}{N}$

$$= \frac{1248.750}{70} = 17.83$$

1. The total of absolute deviations from the mean is higher than the total of absolute deviations from the median. In fact, the mean deviation about the median is the least. Therefore, the mean deviation about the mean is not very appropriate.

2. In a series characterized by a high degree of variability, the median does not serve as a representative measure of central tendency. Similarly, the mean deviation about the mean is not a highly reliable measure of dispersion in such cases.

3. Mean deviation is computed using the absolute values of deviations, making it unsuitable for further algebraic manipulation.

Important formula / points

- The smallest mean deviation occurs around the median.
- Mean deviation

  Ungrouped data: M.D. $= \frac{\sum_{i=1}^{n}|x_i - k|}{n}$; K be the statistical mean (A.M., median, mode)

  Discrete / continuous frequency distribution: :M.D. $= \frac{\sum_{i=1}^{n}|x_i - k| f_i}{\sum_{i=1}^{n} f_i}$

  Where, x_i = observations / mid – value of i^{th} class

  f_i = frequency of i^{th} observation / class

  n = number of observations

  k = statistical mean (A.M., Median, Mode)

Limitation of mean derivation

1. The mean deviation from the median may not be a fully reliable measure of dispersion in cases of high variability, given that the median may not adequately represent the central tendency.
2. The total of absolute deviations from the mean exceeds the total of absolute deviations from the median. This makes it less reliable in certain scenarios.
3. Mean deviation involves the use of absolute values for deviations, restricting further algebraic manipulations.

VARIANCE AND STANDARD DEVIATION

When computing the mean deviation, absolute values of deviations are utilized to mitigate issues arising from deviation signs. Alternatively, another approach involves squaring all the deviations.

1 Variance

The variance of a variable is the average of the squared deviations from the mean (A.M.) and is symbolized as or var (x).

Let there $X_1, X_2, X_3, \ldots \ldots .. x_n$ be n given values of a variable, and let $\bar{x}$ represent their mean. Then,

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

2 Standard Deviation

When computing variance, the units of individual observations and the unit of their mean ($\bar{x}$) differ from that of variance. The appropriate measure of dispersion around the mean of a set of observations is represented by the positive square root of the variance, known as standard deviation(σ).

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

3 Variance And Standard Deviation In Different Cases

(a) In case of individual series (ungrouped data):

Let $x_1, x_2, x_3 \ldots x_n$ are n values of a variable x, then by definition

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

(Where $\bar{x}$ is A.M. of $x_1, x_2, x_3, \ldots \ldots x_n$ i.e., $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$)

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

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

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

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

And standard deviation,

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

(b) In case of discrete frequency distribution:

Consider $x_1, x_2, x_3 \ldots x_n$ as n observations with frequencies$f_1, f_2, f_3 \ldots f_n$, respectively.

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

(Where $\bar{x}$ is A.M. of $x_1, x_2, x_{3,\ldots\ldots\ldots n} x_n$ i.e., $\bar{x} = \frac{\sum x_i f_i}{N}$ and $N = \sum_{i=1}^{n} f_i$)

$$= \frac{1}{N}\sum(x_i^2 - 2\bar{x}x_i + (\bar{x})^2)f_i$$

$$= \frac{\sum x_i^2 f_i}{N} - 2\bar{x}\frac{\sum x_i f_i}{N} + (\bar{x})^2\frac{\sum f_i}{N}$$

$$= \frac{\sum x_i^2 f_i}{N} - 2\overline{x}\overline{x} + (\bar{x})^2$$

$$= \frac{\sum x_i^2 f_i}{N} - (\bar{x})^2$$

Hence standard deviation, $\sigma = \sqrt{\frac{\sum x_i^2 f_i}{N} - (\bar{x})^2}$

(c) In case of continuous frequency distribution:

Let x_i = mid-value of i^{th} class

f_i= frequency of i^{th} class

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

$\bar{x}$ = A.M. of given observations

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

$$= \frac{\sum x_i^2 f_i}{N} - (\bar{x})^2$$

And standard deviation, $\sigma = \sqrt{\frac{\sum x_i^2 f_i}{N} - (\bar{x})^2}$

$$= \frac{1}{N}\sqrt{N\sum_{i=1}^{n} f_i x_i^2 - (\sum_{i=1}^{n} f_i x_i)^2},,$$

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

Ex. Determine the standard deviation of the initial n natural numbers.

Sol. Here, $x_i = i$ where $i = 1, 2 \ldots n$

$$\sigma = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1+2+3+\cdots\ldots+n}{n}$$

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

$$\sigma = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1+2+3+\cdots\ldots+n}{n}$$

$$\sum x_i^2 = 1^2 + 2^2 + 3^2 + \cdots\ldots\ldots+n^2 = \frac{n(n+1)(2n+1)}{6}$$

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

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

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

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

Ex. If the heights of 8 students are given in centimeters, calculate their arithmetic mean and standard deviation.162,163,160,164,160,170,161,164

Sol. We have,

$$\text{Mean, } \bar{x} = \frac{162+163+160+164+160+170+161+164}{8} = 163$$

Height (in cm) x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
160	–3	9
160	–3	9
161	–2	4
162	–1	1
163	0	0
164	1	1
164	1	1
170	7	49
Total		**74**

Here n = 8 and $\sum(x_i - \bar{x})^2 = 74$

$$\sigma = \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}$$

$$= \sqrt{\frac{74}{8}} = \sqrt{9.25} = 3.04\text{cm}$$

Thus, mean = 163 and standard deviation = 3.04 cm

Ex. The mean and standard deviation of 20 observations were initially determined to be 10 and 2, respectively. Upon rechecking, it was discovered that an observation was incorrectly recorded as 8. Calculate the corrected mean and standard deviation in each of the following cases:

1. If the wrong item is omitted. **2.** If it is replaced by 12.

Sol. **1.** We have, $n = 20, \overline{x} = 10$ and $\sigma = 2$

$$\overline{x} = \frac{1}{n}\sum x_i$$

$$\sum x_i = n\overline{x} = 20 \times 10 = 200$$

Incorrect $\sum x_i = 200$

$$\sigma = 2$$

$$\sigma^2 = 4$$

$$\frac{1}{n}\sum x_i^2 - \overline{x}^2 = 4$$

$$\frac{1}{20}\sum x_i^2 - 100 = 4$$

$$\sum x_i^2 = 2080$$

Incorrect $\sum x_i^2 = 2080$

1. By excluding the incorrect item, 8, from the observations, we are left with a set of 19 observations.

Correct $\sum x_i + 8 =$ incorrect $\sum x_i$

Correct $\sum x_i = 200 - 8 = 192$

Correct mean $= \frac{192}{19} = 10.10$

And Correct $\sum x_i^2 = 2080 - 64 = 2016$

Correct variance $= \frac{1}{19}(\text{ correct } \sum x_i^2) - (\text{ correct mean })^2$

$$= \frac{2016}{19} - (\frac{192}{19})^2 = \frac{1440}{361}$$

Correct standard deviation $= \sqrt{\frac{1440}{361}} = 1.997$

2. If we replace the wrong item by 12

Incorrect $\sum x_i - 8 + 12 = \text{correct} \sum x_i$

Correct $\sum x_i = 200 + 4 = 204$ and correct mean $= \frac{204}{20} = 10.2$

Incorrect $\sum x_i^2 - 8^2 + 12^2 = 2160$

Correct variance $= \frac{1}{20}(\text{ correct } \sum x_i^2) - (\text{ correct mean })^2$

$$= \frac{2160}{20} - \left(\frac{204}{20}\right)^2 = \frac{1584}{400}$$

Correct standard deviation $= \sqrt{\frac{1584}{400}} = 1.9899$

Ex. Calculate the variance and standard deviation for the given frequency distribution:

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

Sol. Compute the variance and standard deviation:

x_i	f_i	$f_i x_i$	$x_i - 19$	$(x_i - 19)^2$	$f_i (x_i - 19)^2$
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
	$N = \Sigma f_i = 40$	$\Sigma f_i X_i = 760$			$\Sigma f_i(x_i - 19)^2 = 1736$

Here, $N = 40$, $\Sigma f_i x_i = 760$

Mean $\bar{x} = \frac{1}{N} \cdot \sum f_i x_i = \frac{760}{40} = 19.$

We have, $\sum f_i (x_i - \bar{x})^2 = 1736$

Variance, $\sigma^2 = \frac{1}{N} \cdot \sum f_i (x_i - \bar{x})^2 = \frac{1736}{40} = 43.4$

And standard deviation $\sigma = \sqrt{43.4} = 6.59$

Ex. Determine the mean, variance, and standard deviation for the given distribution:

Class	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	3	7	12	15	8	3	2

Sol. Compute the mean, variance, and standard deviation:

Classes	Mid – value s (x_i)	f_i	$f_i x_i$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
30-40	35	3	105	729	2187
40-50	45	7	315	289	2023
50-60	55	12	660	49	588
60-70	65	15	975	9	135
70-80	75	8	600	169	1352
80-90	85	3	255	529	1587
90-100	95	2	190	1089	2178
		$N = \Sigma f_i = 50$	$\Sigma f_i X_i = 3100$		$\Sigma f_i (x_i - \bar{x})^2 = 10050$

Mean, $\bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{3100}{50} = 62$

Hence, variance $\sigma^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{N} = \frac{10050}{50} = 201$

And standard deviation, $\sigma = \sqrt{201} = 14.17$

Ex. The variance of n observations is denoted as σ^2. Demonstrate that if each observation is increased by a, the variance of the new set of observations remains σ^2.

Sol. Let the observations be $x_1, x_2, x_3, \dots \dots x_n$

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - [\frac{\sum_{i=1}^{n} x_i}{n}]^2 \quad \dots (1)$$

When each observation is incremented by a, the values of the observations become

$$x_1 + a, x_2 + a, x_3 + a, \dots \dots, x_n + a$$

Sum of these observations

$$= \sum_{i=1}^{n} (x_i + a) = \sum_{i=1}^{n} x_i + na \quad \dots (2)$$

And sum of squares of these observations

$$= \sum_{i=1}^{n} (x_i + a)^2 = \sum_{i=1}^{n} (x_1^2 + a^2 + 2x_i a)$$

$$= \sum_{i=1}^{n} x_i^2 + na^2 + 2a \sum_{i=1}^{n} x_i \quad \dots (3)$$

Hence, the variance of the new set of observations

$$= \frac{\sum_{i=1}^{n} x_i^2 + na^2 + 2a \sum_{i=1}^{n} x_i}{n} - [\frac{\sum_{i=1}^{n} x_i + na}{n}]^2 \quad \text{(Using (2) and (3))}$$

$$= \frac{\sum_{i=1}^{n} x_i^2}{n} + a^2 + 2a[\frac{\sum_{i=1}^{n} x_i}{n}] - [\frac{\sum_{i=1}^{n} x_i}{n}]^2 - a^2 - \frac{2na \sum_{i=1}^{n} x_i}{n^2}$$

$$= \frac{\sum_{i=1}^{n} x_i^2}{n} - [\frac{\sum_{i=1}^{n} x_i}{n}]^2 = \sigma^2 \quad \text{(Using (1))}$$

4 Shortcut Method To Find Variance and Standard Deviation:

(When x_i are large)

Consider the assumed mean as A and the class interval width as h.

$$u_i = \frac{x_i - A}{h}$$

$$x_i = A + hu_i \qquad \text{... (1)}$$

The arithmetic mean

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

$$= \frac{\sum f_i (A + hu_i)}{N}$$

$$\bar{x} = A + h\frac{\sum f_i u_i}{N} \qquad \text{... (2)}$$

From (1) and (2),

$$x_i - \bar{x} = h\left(u_i - \frac{\sum f_i u_i}{N}\right) \qquad \text{... (3)}$$

Now, variance

$$(\sigma^2) = \frac{\sum f_i (x_i - \bar{x})^2}{N}$$

$$= \frac{h^2}{N}\sum\left(u_i - \frac{\sum f_i u_i}{N}\right)^2 f_i \qquad \{\text{Using (3)}\}.$$

$$= \frac{h^2}{N}\sum(u_i - \bar{u})^2 f_i = h^2 \times (\text{ var iance of var iable } u_i)$$

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

$$\sigma_x = h\sigma_u \qquad \text{... (1)}$$

$$\sigma_x = \frac{h}{N}\sqrt{N\sum f_i u_i^2 - (\sum f_i u_i)^2}$$

$$\sigma = \frac{1}{N}\sqrt{N\sum f_i x_i^2 - (\sum f_i x_i)^2}$$

Ex. Determine the mean, variance, and standard deviation for the given distribution:

Class	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	3	7	12	15	8	3	2

Sol. Let the assumed mean A = 65.
Here h = 10
We obtain the following table:

Class	**Frequency f_1**	**Mid – Point x_1**	$y_i = \frac{x_i - 65}{10}$	y_1^2	$f_i y_i$	$f_i y_1^2$
30-40	3	35	–3	9	–9	27
40-50	7	45	–2	4	–14	28
50-60	12	55	–1	1	–12	12
60-70	15	65	0	0	0	0
70-80	8	75	1	1	8	8
80-90	3	85	2	4	6	12
90-100	2	95	3	9	6	18
	N = 50				–15	105

$$\bar{x} = A + \frac{\sum f_i y_i}{50} \times h$$

$$= 65 - \frac{15}{50} \times 10 = 62$$

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

$$= \frac{(10)^2}{(50)^2} \times (50 \times 105 - (-15)^2) = 201$$

And standard deviation $(\sigma) = \sqrt{201} = 14.18$

Ex. Compute the mean and standard deviation for the given data using the shortcut method.

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

Sol. Compute the variance and standard deviation:

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
	$\sum f_i = N = 100$			$\sum f_i d_i = 0$	$\sum f_i d_i^2 = 286$

Here, assumed mean = 64

Actual mean, $\bar{x} = a + \frac{1}{N}\sum f_i d_i$

$= 64 + \frac{0}{100} = 64$

Standard deviation, $\sigma = \sqrt{\frac{1}{N}\sum f_i d_i^2 - \left(\frac{\sum f_i d_i}{N}\right)^2}$

$= \sqrt{\left(\frac{286}{100}\right) - 0} = \sqrt{2.86} = 1.69$

ANALYSIS OF FREQUENCY DISTRIBUTIONS

To assess the variability of two series with the same mean, measured in different units, it is insufficient to merely calculate measures of dispersion. Instead, we need measures that are independent of units. The measure of variability that is unit-independent is known as the coefficient of variation (C.V.) and is defined as

$$\text{C.V.} = \frac{\sigma}{\bar{x}} \times 100, \bar{x} \neq 0$$

Where σ and $\bar{x}$ represent the standard deviation and mean of the data, respectively. The series with a higher coefficient of variation (C.V.) is considered more variable than the other series, while the series with a lower C.V. is considered more consistent.

Comparison Of Two Frequency Distribution With Same Mean

Let σ_1 and σ_2 be the standard deviations of two series with a common mean $\bar{x}$, then

$$\text{C.V. (Ist Distribution)} = \frac{\sigma_1}{\bar{x}} \times 100, \bar{x} \neq 0$$

$$\text{C.V. (IInd Distribution)} = \frac{\sigma_2}{\bar{x}} \times 100, \bar{x} \neq 0$$

Thus, the comparison of the two coefficient of variations (C.V.) is based solely on the values of σ_1 and σ_2.

Therefore, when two series have equal means, the series with a higher standard deviation (or variance) is considered more variable or dispersed than the other. Conversely, the series with a lower standard deviation (or variance) is deemed more consistent than the other.

Ex. The One Day International (ODI) performance of two cricket players from a cricket team is outlined as follows:

Player	Runs in last 10 ODI matches									
Rahul	27	45	31	46	23	87	101	78	24	11
Sachin	43	95	5	78	88	103	23	01	41	52

Which of these two is more dependable?

Sol. It is evident that Sachin scored significantly more runs than Rahul, with 529 compared to 473, resulting in a higher average of 52.9 versus 47.3. However, to assess the reliability of the two datasets, we must calculate the standard deviation.

S.D. of Rahul: $s = 30.8$

S.D. of Sachin: $s = 36.9$

C.V For Sachin: $= 0.698$

C.V. for Rahul: $= 0.651$

Decision: Since the C.V. of Rahul is less, he is more reliable than Sachin.