STATISTICS
VARIANCE AND STANDARD DEVIATION

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,

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

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).

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

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

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

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

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

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

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

And standard deviation,

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

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

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

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

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

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

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

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

(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)

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

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

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

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

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

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

Ex.1 Determine the standard deviation of the initial n natural numbers.

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

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

$$\overline{x} = \frac{\sum x_i}{n} = \frac{1+2+3+\ldots..+n}{n}$$

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

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

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

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

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

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

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

Ex.2 If the heights of 8 students are given in centimetres, calculate their arithmetic mean and standard deviation.

162, 163, 160, 164, 160, 170, 161, 164

Sol. We have,

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.3 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:

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

Sol. (i) We have, n = 20 , $\bar{x}$ = 10 and $\sigma = 2$

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

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

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 $\sum x_i^2 = 2080$

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

$$\text{Correct } \sum x_i + 8 = \text{incorrect } \sum x_i$$

$$\text{Correct } \sum x_i = 200 - 8 = 192$$

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

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

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

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

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

(ii) If we replace the wrong item by 12

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

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

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

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

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

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

Ex.4 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 $\bar{x} = \frac{1}{N}.\sum f_i x_i = \frac{760}{40} = 19$.

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

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

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

Ex.5 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

Solution: 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, $$\overline{x} = \frac{\sum f_i x_i}{N} = \frac{3100}{50} = 62$$

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

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

Ex.6 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 $x_1, x_2, x_3, \ldots\ldots x_n$

$$\therefore \qquad \sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left[\frac{\sum_{i=1}^{n} x_i}{n}\right]^2 \qquad \ldots(i)$$

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

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

Sum of these observations

$$= \sum_{i=1}^{n}(x_i + a) = \sum_{i=1}^{n} x_i + na \qquad \ldots(ii)$$

and sum of squares of these observations

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

$$= \sum_{i=1}^{n} x_i^2 + na^2 + 2a\sum_{i=1}^{n} x_i \qquad \ldots(iii)$$

Hence, the variance of the new set of observations

$$= \frac{\sum_{i=1}^{n} x_i^2 + na^2 + 2a\sum_{i=1}^{n} x_i}{n} - \left[\frac{\sum_{i=1}^{n} x_i + na}{n}\right]^2 \qquad \text{(Using (ii) and (iii))}$$

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

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

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

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

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

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

$$\sigma_x = h\sigma_u \qquad \text{.......(4)}$$

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

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

Ex.7 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_i	Mid-point x_i	$y_i = \frac{x_i - 65}{10}$	y_i^2	$f_i y_i$	$f_i y_i^2$
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

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

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

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

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

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

Ex.8 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
	$\sum f_i = N = 100$			$\Sigma f_i d_i = 0$	$\Sigma f_i d_i^2 = 286$

Here, assumed mean = 64

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

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

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

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