Standard Deviation

By- Dr. Ekta Khare

Standard Deviation

- Standard deviation is commonly abbreviated as SD and denoted by 'σ’ and it tells about the value that how much it has deviated from the mean value.
- It is the measure of the dispersion of statistical data.
- If we get a low standard deviation then it means that the values tend to be close to the mean whereas a high standard deviation tells us that the values are far from the mean value.
- Standard deviation is the positive square root of the variance.
- Variance is defined as “The average of the **squared** differences from the Mean.”

Variance and Standard Deviation Formula

The formulas for the variance and the standard deviation is given below:

Standard Deviation Formula

The population standard deviation formula is given as:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2}$$

Here,

σ = Population standard deviation

N = Number of observations in population

Xi = ith observation in the population

μ = Population mean

... Variance and Standard Deviation Formula

Similarly, the sample standard deviation formula is:

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

Here,

s = Sample standard deviation

n = Number of observations in sample

xi = ith observation in the sample

$\overline{x}$
= Sample mean

Variance Formula:

The population variance formula is given by:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2$$

The sample variance formula is given by:

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

Standard Deviation Example

Let's calculate the standard deviation for the number of gold coins on a ship run by pirates.

There are a total of 100 pirates on the ship. Statistically, it means that the population is 100. We use the standard deviation equation for the entire population if we know a number of gold coins every pirate has.

Statistically, let's consider a sample of 5 and here you can use the standard deviation equation for this sample population.

This means we have a sample size of 5 and in this case, we use the standard deviation equation for the sample of a population.

Consider the number of gold coins 5 pirates have; 4, 2, 5, 8, 6.

Mean:

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

$$= \frac{x_1 + x_2 + x_3 + x_4 . \ldots .. + x_n}{n}$$

= (4 + 2 + 5 + 6 + 8) / 5

= 5

$x_n - \bar{x}$ *for every value of the sample* :

$$x_1 - \bar{x} = 4–5 = -1$$

$$x_2 - \bar{x} = 2–5 = -3$$

$$x_3 - \bar{x} = 5–5 = 0$$

$$x_4 - \bar{x} = 8–5 = 3$$

$$x_5 - \bar{x} = 6–5 = 1$$

$$\sum (x_n - \bar{x})^2$$

$$= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_5 - \bar{x})^2$$

$$= (-1)^2 + (-3)^2 + 0^2 + 3^2 + 1^2$$

= 20

Standard deviation:

$$S.D = \sqrt{\frac{\sum(x_n - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{20}{4}}$$

= √5

= 2.236

Computation of standard deviation for grouped data [(*discrete series*)]

$$S = \sqrt{\frac{\Sigma f.x^2}{\Sigma f}} \quad \text{or} \quad \hat{S} = \sqrt{\frac{\Sigma f.x^2}{\Sigma f - 1}}$$

Worked example : Weight of ovary of 50 fishes is given below in a simple frequency distribution table. Find the standard deviations.

Ovary Weight	2	2.5	2.7	2.9	3	3.1	3.3	3.7	3.9	4	4.6	4.8	4.9	5	5.5	5.9	6	6.1	6.7	6.9
Freq.	2	1	1	2	3	1	3	2	4	3	2	3	3	3	2	3	3	3	3	3

Solution : A Table of six columns is prepared :

First column for Variable i.e., ovary wieght. 2nd column for frequency. 3rd column for frequency multiplied by variable. 4th column for deviation obtained by $(X - \overline{X})$ formula. 5th column for deviation square. 6th column for frequency multiplied by deviation square.

Table 6.5

X	f	f.X	x	x^2	fx^2
2	2	4	–2.62	6.864	13.736
2.5	1	2.5	–2.12	4.494	4.494
2.7	1	2.7	–1.92	3.841	3.841
2.9	2	5.8	–1.72	2.958	5.916
3	3	9.0	–1.62	2.624	7.872
3.1	1	3.1	–1.52	2.310	2.310
3.3	3	9.9	–1.32	1.742	5.226
3.7	2	7.4	–0.92	0.846	1.692
3.9	4	15.6	–0.72	0.518	2.072
4.0	3	12.0	–0.62	0.384	1.152
4.6	2	9.2	–0.02	0.0004	0.0008
4.8	3	14.4	0.17	0.028	0.084
4.9	3	14.7	0.28	0.078	0.234
5.0	3	15.0	0.37	0.136	0.408
5.5	2	11.0	0.87	0.756	1.512
5.9	3	17.7	1.27	1.612	4.836

(Contd.)

X	f	f.X	x	x^2	fx^2
6.0	3	18.0	1.37	1.876	5.628
6.1	3	18.3	1.47	2.160	6.48
6.7	3	20.1	2.07	4.284	12.852
6.9	3	20.7	2.68	7.182	21.546
		$\Sigma f.X = 231.1$			$\Sigma f.x^2 = 101.89$

Solution :

$$\text{Mean} = \frac{\Sigma fX}{\Sigma f} = \frac{231.1}{50} = 4.622$$

$$S = \sqrt{\frac{\Sigma fx^2}{\Sigma f}} = \sqrt{\frac{101.89}{50}} = \sqrt{2.0378} = 1.427 \textbf{ Ans.}$$

Computation of standard deviation by direct method

Standard deviation (S) can also be calculated by following formula where we do not need to obtain deviation : $S = \sqrt{\frac{\Sigma f X^2 - \overline{X}^2}{\Sigma f}}$

Worked example : A sample of ten common limpet shells (*Patella vulgaris*) from a rocky shore haying the following maximum basal diameters in millimeters : 36, 34, 41, 39, 37, 43, 36, 37, 41, 39. Find out basal diameter and St. deviation of the shell.

Solution : In order to calculate the mean maximum basal diameters and the Standard deviation it is necessary to calculate $\Sigma f / f.X^2$ and $\overline{X}^2$ as mentioned in the Table 6.6.

Table 6.6

X	f	fX	$f.X^2$
34	1	34	1156
36	2	72	2592
37	2	74	2738
39	2	78	3042
41	2	82	3362
43	1	43	1849
	$\Sigma f = 10$	$\Sigma fX = 383$	$\Sigma fX^2 = 14739$

Therefore $\overline{X} = \frac{\Sigma fX}{\Sigma f} = \frac{383}{10} = 38.3$

$\therefore \quad \overline{X}^2 = 1466.9$

Since $$S = \sqrt{\frac{\Sigma f X^2}{\Sigma f} - \bar{X}^2}$$

$$= \sqrt{\left(\frac{14739}{10} - 1466.9\right)}$$

$$= \sqrt{1473.9 - 1466.9}$$

$$= \sqrt{7}$$

$$= 2.65 \text{ or } 2.7$$

Ans.

In this population of the *Patella vulgaris* the mean maximum basal diameter of the shell is 38.3 mm, with a standard deviation of 2.7 mm (correct to one decimal place). If these values are applied to a larger population of the *Patella* then it may be assumed, on statistical grounds, that approximately 68% of the population will have a basal diameter of the shell of 38.3 mm plus and minus 1 standard deviation (2.7 mm), that is they will lie within a range 35.6–41.0 mm; approximately 95% of the population will have a basal diameter of the shell of 38.3 plus and minus 2 Standard deviation (5.4 mm), that is they will lie within the range 32.9 — 43.7 mm. and practically 100% will lie within plus and minus 3 Standard deviations.

Computation of Standard deviation for grouped data

(Continuous series)

Worked example : Ovary wt. of 50 fishes and their frequency is given in class interval. Find Standard deviation.

Wt. of ovary	2–2.9	3–3.9	4–4.9	5–5.9	6–6.9
Frequency	6	13	11	8	12

Solution : It is computed by following six steps :

Step 1 : Find mid point (m) of each class interval.

Step 2 : Find mean of the series applying formula $\overline{X} = \frac{\Sigma fX}{\Sigma f}$ here, $\bar{x} = \frac{\Sigma f m}{\Sigma f}$

Step 3 : Find deviation of each observation $(X - \overline{X})$. here, $(m - \bar{x})$

Step 4 : Square each deviation.

Step 5 : Multiply each squared deviation with their frequency

Step 6 : Sum all the multiplied value of f and x^2.

Now a table of 7 columns is framed.

Table 6.7

$m - \bar{X} = x$

Class Interval	Mid point m	Frequency f	$f.m$	Deviation $X - \bar{X} = x$	Deviation Square x^2	$f.x^2$
2 - 2.9	2.45	6	14.7	–2.14	4.579	27.47
3 - 3.9	3.45	13	44.85	–1.14	1.299	16.88
4 - 4.9	4.45	11	48.95	–0.14	0.019	0.21
5 - 5.9	5.45	8	43.6	+0.86	0.739	0.91
6 - 6.9	6.45	12	77.4	+1.86	3.459	41.4
		$\Sigma f = 50$	$\Sigma f.m = 229.5$		$\Sigma x^2 = 10.088$	$\Sigma f.x^2 = 92.02$

$$\bar{X} = \frac{\Sigma f m}{\Sigma f} = \bar{X} = \frac{\Sigma f.X}{\Sigma f} = \frac{229.5}{50} = 4.59$$

$$S = \sqrt{\frac{\Sigma f x^2}{\Sigma f}} = \sqrt{\frac{92.02}{50}} = \sqrt{1.84} = 1.35 \textbf{ Ans.}$$

Co-efficient of standard deviation is calculated by following formula :

Co-efficient S = $S/\bar{X}$. Here co-efficient $S = \dfrac{1.35}{4.59} = 0.29$ **Ans.**

Merits and demerits of Standard deviation

Merits : (i) Standard deviation summarises the deviation of a large distribution from mean in one figure used as a unit of variation. (ii) It indicates whether the variation of difference of an individual from the mean is real or by chance. (iii) Standard Deviation helps in finding the suitable size of sampe for valid conclusions. (iv) It helps in calculating the Standard error.

Demerits : Standard deviation gives weightage to only extreme values. The process of squaring deviations and then taking square root involves lengthy calculations.

Differences between Mean deviation (δ) and Standard deviation (S)

(1) Algebraic signs are ignored while calculating the Mean deviation. But they are taken into account while calculating the Standard deviation.

(2) Mean deviation can be computed either from the mean or median. But the Standard deviation is always computed from the arithmetic mean. Because the sum of the squares of the deviations of items from arithmetic mean is the least.