Standard deviation

It is quite obvious that Mean value/averages fail to tell us everything about a sample. Samples can either be very uniform with the data concentrated around the mean values or they can be spread out a long way from the mean.

The statistical tool that measures this spread of the values is called the **standard deviation**. The wider the spread of values observed, the larger is the value of standard deviation.

Definition: The standard deviation or mean error is defined as the absolute measure of **dispersion** of an individual series or a frequency distribution, is the square root of the arithmetic mean of the squares of deviations of values from their arithmetic mean.

The term 'root-mean square deviation' is the square root of the arithmetic mean of the squares of deviations of values from any arbitrary value, and so the term standard deviation' is a special case of the root-mean-square deviation when deviations are taken from the arithmetic mean.

Standard deviation is always represented by the small Greek letter sigma (σ). While calculating $\sigma \pm$ signs are taken into consideration.

In a discrete series, standard deviation is calculated by applying the following formula:

$\sigma = \sqrt{\sum (x - \overline{x})^2/N}$

where N is the number of observations.

Where the mean of a series is not a whole number, i.e. a fraction, the above formula will be difficult to work out and in that case the formula will be:

$\sigma = \sqrt{\sum dx^2/N - (\sum dx/N)^2}$

Where $\sum dx^2$ denotes the sum of squares of the deviations from the assumed mean.

The above two formulae, when applied to a frequency distribution will be written as follows:

$\sigma = \sqrt{\sum dx^2/N}$ when deviations are taken from actual mean.
$\sigma = \sqrt{\sum fdx^2/N - (\sum fdx^2/N)^2}$ when deviations are taken from an assumed mean.

In continuous series, calculations can be simplified if we divide the deviations of the midpoints by class Interval.

The standard error of a sample mean ($\bar{x}$) is given by the relation:
S.E. = $\sigma/\sqrt{n}$

Where n is the number of observations on which the mean has been calculated.

How do we calculate the Standard Deviation:

- First we calculate the mean ($\bar{x}$) of a set of data
- Then subtract the mean from each point of data to determine ($x-\bar{x}$). You'll do this for each data point, so you'll have multiple ($x-\bar{x}$).
- Square each of the resulting numbers to determine ($x-\bar{x}$)^2. As in step 2, you'll do this for each data point, so you'll have multiple ($x-\bar{x}$)^2.

- Add the values from the previous step together to get ∑(x-x̅)^2. Now you should be working with a single value.
- Calculate (n-1) by subtracting 1 from your sample size. Your sample size is the total number of data points you collected.
- Divide the answer from step 4 by the answer from step 5
- Calculate the square root of your previous answer to determine the standard deviation.
- The standard deviation data has the same number of units as your raw data, so you may need to round your answer.
- The standard deviation should have the same unit as the raw data you collected. For example, SD = +/- 0.5 cm.

Example:

Marks	*No. of students*
5	2
10	4
15	5
20	9
25	10
30	5
35	15

Solution : Here we assume 20 as the assumed average.

x	f	$dx = \frac{(x-20)}{5}$	fdx	dx^2	fdx^2
5	2	-3	-6	9	18
10	4	-2	-8	4	16
15	5	-1	-5	1	5
20	9	0	0	0	0
25	10	1	10	1	10
30	5	2	10	4	20
35	15	3	45	9	135
Total N = 50			46		204

$$\sigma = \sqrt{\frac{\Sigma fdx^2}{N} - \left(\frac{\Sigma fdx}{N}\right)^2} \times i$$

i = class interval, *i.e.* 5 in this case.

$$\sigma = \sqrt{\frac{204}{50} - \left(\frac{46}{50}\right)^2} \times 5$$

$$= \sqrt{4.08 - 0.85} \times 5 = 1.8 \times 5 = 9$$

Reference: Biologydiscussion.com