
WHY STANDARD DEVIATION IS AN IMPORTANT STATISTIC ?

FUNCTION F10 : SOEN 6011 VERSION 1.0

Basant Gera (40082433)*

Dept. of Computer Science & Software Engineering
Concordia University
basantgera29@gmail.com

Submitted to : Pankaj Kamthan* & Team*

Dept. of Computer Science & Software Engineering
Concordia University
kamthan@cse.concordia.ca

July 20, 2019

ABSTRACT

The software industry has grown to become the largest industry. However, it had humble beginnings. It is likely that calculator was the first 'educational' tool that many prospective economists, engineers, and scientists, were exposed to during their time in high school, and was a tool that could do something that they could not do, sometimes at all and other times at the same speed. It should therefore come as no surprise that many operating systems provide native support for a calculator. A question of interest for a software engineer is how a typical scientific calculator does what it is supposed to do. The question may be simple, but the answer is not (as simple). The answer is also crucial towards providing an insight as well as an in-depth understanding of how more challenging numerical or symbolic calculations are performed by the state-of-the-art scientific computing software.

1 Goal & objective.

To define requirements of Standard Deviation and how it works and explain their domain and co domain with their functionality how it is calculated with the help of some basic arithmetic symbols and calculate their mean value of data set and then calculating their variance followed by variance square root.

2 Calculating the standard deviation.

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

σ = Lower case sigma

Σ = Capital sigma

\bar{x} = x bar

Step I) Calculate the mean of following numbers or data set given to you. (Yes it can be negative or positive).
For e.g Numbers given to you can be (600,470,170,430,300).

Step II) After calculating the mean of the following number given to you Then for each number: subtract the Mean and square the result (the squared difference).

$$\text{Mean} = \frac{600+470+170+430+300}{5} = \frac{1970}{5} = 394$$

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

Step III) Then work out the average of those squared differences.

$$(Variance)^2 = \frac{206^2 + 762^2 + 224^2 + 36^2 + 94^2}{5} = \frac{42436 + 5776 + 50176 + 1296 + 8836}{5} = \frac{108520}{5} = 21704$$

So, therefore Variance is 21704 Now, Calculating Standard deviation.

Step IV) And the Standard Deviation is just the square root of Variance.(Which needs to be always Positive)

$$\text{Standard Deviation} = \sqrt{\sigma} = \sqrt{21704} = 147.32 = \mathbf{147} \text{ (to the nearest mm)}$$

3 Requirements for standard deviation.

Since there are two ways to calculate standard deviation which are as follows :

- Population standard deviation : for e.g analyzing test scores of a class
- Sample standard deviation : for e.g analyzing the effect of caffeine on reaction time on people age 18-25.

Our focus would be on first which is Population standard deviation and upon its requirements. The requirements based on priority for calculating standard deviation is as follows :

- ID:1 Since standard deviation tells you how close the numbers are w.r.t to mean so it deals with only numbers not with strings. [Requirement type :Functional,Interface] [priority :Highest] [Version :1.0]
- ID:2 For calculating the mean for natural and real numbers can be taken into account.[Requirement type :Functional] [priority :low] [Version :1.0]
- ID:3 Since we have square root of variance so therefore standard deviation can not have negative value.[Requirement type :Functional,Interface] [priority :high] [Version :1.0]
- ID:4 Remember as long as the variance exists, the standard deviation also exists.[Requirement type :Functional] [priority :Highest] [Version :1.0]Requirement type :Functional,Interface] [priority :Medium] [Version :1.0]
- ID:5 The problem here is that the standard deviations will vary from individual to individual. What we want is an estimate of the common within-subject variability.Provided the same number of results are obtained for each original sample, the common within-subject variability can estimated by estimating the mean of their individual variances - and then taking the square root of this, to give the within-subject standard deviation.[Requirement type :Functional] [priority :Highest] [Version :1.0]
- ID:6 Design should be simple enough and user friendly which user can understand and make some input..[Requirement type :Functional,Interface] [priority :Highest] [Version :1.0]
- ID:7 Number at the time of input should be without comma and with space so that standard deviation can be calculated. .[Requirement type :Functional] [priority : Very High] [Version :1.0]
- ID:8 Algorithm should be efficient and execute in small amount of time..[Requirement type :Functional] [priority :Highest] [Version :1.0]

4 Assumption for standard deviation.

The variance and standard deviation can be calculated for any variable - providing it can be ordered. But the standard deviation is only an appropriate measure of dispersion for a measurement variable, and only then if the data have a symmetrical distribution - and, in many cases, a normal one. Use of the standard deviation to display the variability of observations in range plots and box-and-whisker plots is misleading if these assumptions are not met. Assumptions about what proportion of observations are included within limits of agreement are also dependent upon this assumption.

For non-symmetrical (skewed) distributions there are two options :

- Use the median as the measure of location and the inadequately range as a measure of dispersion.
- Transform the data (often by taking logarithms) so that the distribution of values is symmetrical, and then work out the standard deviation of the transformed data..

5 Advantages and limitations of standard deviation.

Advantages are as follows:

- Shows how much data is clustered around a mean value.
- It gives a more accurate idea of how the data is distributed
- Not as affected by extreme values

Limitation are as follows:

- It doesn't give you the full range of the data.
- Only used with data where an independent variable is plotted against the frequency of it.
- Assumes a normal distribution pattern.
- It is very sensitive to outliers and does not use all the observations in a data set.

6 Conclusion for standard deviation

The higher the calculated value the more the data is spread out from the mean. As SD is used as a measure of dispersion where mean is used as measure of central tendency (ie, for symmetric numerical data).

References

- [1] <https://www.mathsisfun.com/data/standard-deviation.html>
- [2] https://influentialpoints.com/Training/variance_and_standard_deviation-principles-properties-assumpt.htm
- [3] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3198538/>