Concordia University

Standard Deviation

Submitted by
Harjeet Kaur
Submitted to Professor Pankaj Kamthan

SOEN:6011 Software Engineering Processes

Engineering and Computer Science Department

July 19, 2019

1 Introduction

Standard deviation is a measure that is utilized to evaluate the measure of variation or dispersion of a set of information values. A low standard deviation demonstrates that the information tends to be near the mean (additionally called the expected value) of the set, while a high standard deviation shows that the information is spread out over a wider scope of values.[1]

The standard deviation of a random variable, statistical population, data set, or probability distribution is the square root of its variance. It is algebraically simpler, though in practice less robust, than the average absolute deviation. A useful property of the standard deviation is that, unlike the variance, it is expressed in the same units as the data. [Wikipedia]

In other words, the standard deviation σ (sigma) is the square root of the variance of X; i.e., it is the square root of the average value of $(X - \mu)^2$.

Population Standard Deviation

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

Sample Standard Deviation

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

1.1 Properties of Standard Deviation

- Standard deviation is only used to measure spread or dispersion around the mean of a data set.
- Standard deviation is never negative.
- Standard deviation is sensitive to outliers. A single outlier can raise the standard deviation and in turn, distort the picture of spread.
- For data with approximately the same mean, the greater the spread, the greater the standard deviation.

- If all values of a data set are the same, the standard deviation is zero (because each value is equal to the mean)[5]

2 Requirements and constraints

- R1—At least two numbers are required to calculate standard deviation.
- R2—Need to know if data set is a population data or sample data before calculation standard deviation.
- R3— Values in data set must be different (if identical , $\sigma = 0$)
- R4—User would enter data values separated by comma, otherwise it would be consider as a one value.
- R5—User would enter the numbers only(Separated by commas)
- R6—Clear Button to reset the value.

3 Pseudo code

Algorithm 1: standard deviation algorithm [1]

Result: standard deviation
Let n ← 0, Sum ← 0, SumSquare ← 0 size ← arraysize \\ initialization;
if *n==0* **then**
    return 0.0;
else
    while *n is not equal to size* **do**
        n ← n + 1 \\incrementing n;
        Sum ← Sum + x \\ adding all numbers in array to calculate mean;
        SumSquare ← SumSquare + x × x \\ adding square of all numbers;
    end
    if *Sample data* **then**
        Var = (SumSquare − (Sum × Sum) / n) / (n − 1) \\ if data is sample data set, division by n − 1 to calculate variance;
    else
        Var = (SumSquare − (Sum × Sum) / n) / (n) \\ if data set is population, divison by number of values in set to calculate variance;
    end
    Stdev = Square root(Var) \\ square root of variance;
end
[4]

3.1 Advantages

- Improves the readability of the approach to calculate standard deviation.
- Works as a rough documentation, so the program can be understood easily when a pseudo code is written out.
- Explain what exactly each line of a program should do with comments.
- It can be read and understood easily by non programmers.[3]

3.2 Disadvantages

- It was difficult to write a pseudo code before writing the actual code.

- Pseudo code does not covers the whole logic of the program.
- Difficult for non technical person to understand the flow of program(indication of next step from previous step)

References

[1] [Wikipedia]wikipedia.org

[2] [stackexchange]tex.stackexchange.com

[3] www.techwalla.com

[4] www.scribd.com

[5] www150.statcan.gc.ca