

## Determination of Standard Deviation and Assigned Value using Algorithm A


Algorithm is a procedure for solving a mathematical problem in a finite number of steps that frequently involves recursive operations.

Algorithm-A is a tool to identify standard deviation as well as assigned value in proficiency testing. This tool is used to determine an assigned value and standard deviation for a near normal data set with tentatively less than 20% outlier presence.


Before, we go into Algorithm-A. Let's understand little about "Winsorisation" technique which is used in Algorithm-A.

In this technique, the probable outliers are replaced with the nearest non-outlier value instead of removing straightway. By this technique, extreme values in the statistical data set can be limited to reduce the effect of possibly spurious outliers. Here a specified percentile of outlier is replaced by closest data in the data set at both upper and lower end.

Let's explain with below data set about how to use Winsorisation technique:

Data Set  5, 8, 5, 7, 11, 17, 18, 20, 20, 25, 28, 30, 38, 55, 8, 25

Let's sort the above data set with smallest to largest:

Sorted Data Set  5, 5, 7, 8, 8, 11, 17, 18, 20, 20, 25, 25, 28, 30, 38, 55

Now assume that at the largest side, 55 is probable outlier, so, instead of removing 55 straightway, we need to replace with nearest value, (i.e., 38), so the new range will be:

5, 5, 7, 8, 8, 11, 17, 18, 20, 20, 25, 25, 28, 30, 38, **38** (here 55 is replaced with 38!)

Similarly, we can apply the above concept in the other side (i.e., 5 can be replaced with 7).

So, we will get the below winsorized data set:

**7, 7, 7, 8, 8, 11, 17, 18, 20, 20, 25, 25, 28, 30, 38, 38** (here 5, 5 are replaced with 7, 7!)

(So, finally we have obtained a more purified data set than the original data, this process may require multiple iteration!)

### **Important Note:**

Here a very important point to understand is identification of possible outlier in a data set, since outlier identification requires different techniques. Also, there may be required specific knowledge to judge the nature of data set going to be analyzed.

Above example is a simplified example to understand the technique, Actual application of this technique happens in better way.

Now, see below excel example (try yourself to understand from left to right in excel worksheet):



**Step2:**

Calculate initial  $x^*$  - i.e. median of the results of all participants. Record this in the end of the column in which participant's results are given:

$\text{Initial } X^* = \text{median (range of data)} = 148.2$

**Step3:**

Calculate ABS of difference between each result and  $x^*$  calculated earlier, For each participant, record this in the next column:

$X_i - X^* =$

7.45
4.55
0.65
0.35
0.35
2.05
2.35
2.95

**Step4:**

Calculate the median of this data (i.e., median of the deviations of all participants from  $x^*$ ):

$=\text{median of } |X_i - X^*| = 2.2$

**Step5:**

Calculate initial  $s^*$  using formula "1.483 x median calculated in previous step). Record this in the end of the column in which participant's results are given:

$\text{initial } S^* = 1.483 * 2.2 = 3.263$

**Step6:**

Copy the origin results of the participants for iterations 1 to 6 or 14 in subsequent columns.

**Step7:**

For Iteration 1:

Calculate  $d$  (delta) using formula  $1.5 \times s^*$

Calculate  $x^* - d$  using previous value of  $x^*$

Calculate  $x^* + d$  using previous value of  $x^*$

**Step8:**

Examine the result of each participant. If it is less than  $x^* - d$  replace it with  $x^* - d$ ; If it is more than  $x^* + d$  replace it with  $x^* + d$ .

**Step9:**

Calculate new  $x^*$  - i.e. the average of the revised results of the participants. Record this in the bottom of this column.

