

## The Right and Wrong Ways of Computing Limits

How does your software measure up?

Donald J. Wheeler

Today virtually everyone uses software to create process behavior charts, yet the available software is notoriously unreliable in terms of the way the limits are computed. This column will explain and illustrate the difference between the correct and some of the incorrect ways of computing three-sigma limits for Average Charts. It will also provide a simple data set that can be used to evaluate the various options in your software so that, hopefully, you can select an option that uses one of the correct computations.

### THE PURPOSE OF ANALYSIS

The fundamental purpose of any statistical analysis is to separate the potential signals from the probable noise. Once we have said this, the immediate problem becomes how to use our data, which may contain signals, to compute a measure of dispersion that can be used to filter out the noise. When the signals are mixed up with the noise, the signals are likely to contaminate our filter and undermine our analysis.

This problem proved to be intractable until about 100 years ago. At that time a basic distinction was made which allowed us to reliably filter out the noise, and this distinction soon became the foundation of all modern techniques of statistical analysis. The basic idea is to use the context for the data to arrange the data into subsets *where each subset is logically homogeneous*. We can then use the variation *within* these subsets to characterize the probable noise. By summarizing this *within-subgroup variation* into a single measure of dispersion, we can filter out the probable noise, and anything that is left over will represent the potential signals within the data.

Thus, in performing statistical analyses, we want to have a computational approach that will minimize the impact of any signals upon our filter. Such computational approaches are said to be robust.

This use of the *within-subgroup variation* is the foundation for everything from the two-sample Student's *t*-test and ANOVA to Process Behavior Charts and the Analysis of Means. It lies at the heart of modern statistical analysis. And it also makes certain computations of dispersion inappropriate for use in statistical analysis. *Thus, since 1925, there have been right and wrong ways of computing measures of dispersion for use with statistical analysis techniques*. The complexity of most techniques keeps both students and programmers from straying from the correct computations. However, this is not the case with Process Behavior Charts where the apparent simplicity of the technique causes many to fall into the trap of using incorrect and non-robust computations for the limits.

AVERAGE CHARTS DONE RIGHT

When your data have been arranged into  $k$  logically homogeneous subgroups of size  $n$ , the default computation of limits for the Average Chart should be based upon either:

- (1) the Average of the  $k$  Subgroup Ranges or
- (2) the Average of the  $k$  Subgroup Standard Deviation Statistics.

(Computations based upon the  $k$  Subgroup Root Mean Square Deviations will be indistinguishable from those based upon the  $k$  Subgroup Standard Deviation Statistics, and so may be considered to be a subset of approach (2).)

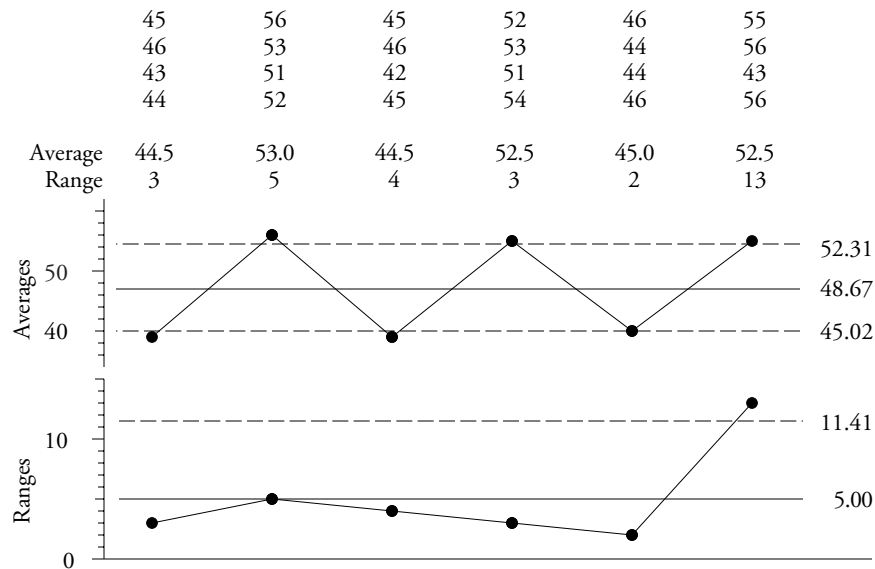


Figure 1 Average and Range Chart

To illustrate the correct and incorrect computations we will use a simple collection of  $k = 6$  subgroups of size  $n = 4$  that contains signals on both the Average Chart and the Range Chart. The values, averages, and ranges are shown along with the Average and Range Chart in Figure 1. The Grand Average is 48.67 while the Average Range is 5.00. Using the usual scaling factors  $A_2 = 0.729$  and  $D_4 = 2.282$  and the formulas found in all the textbooks we get the limits shown. All six subgroup averages and one subgroup range fall outside these limits.

Figure 2 shows the Average and Standard Deviation Chart. The Grand Average is still 48.67. The Average Standard Deviation Statistic is 2.33 Using the scaling factors  $A_3 = 1.628$  and  $B_4 = 2.266$  and the formulas found in all the textbooks we get the limits shown. Five subgroup averages and one subgroup range fall outside these limits.

Note that the limits on the Average Chart in Figure 2 are quite similar those in Figure 1. Also, note that the Range Chart and the Standard Deviation Chart are virtually the same except for the fact that they have a different vertical scale.

While the details change slightly between these two analyses, both tell the same story about the underlying process. Since the purpose of analysis is insight, we have to regard these two

approaches as essentially identical in results even though the details change.

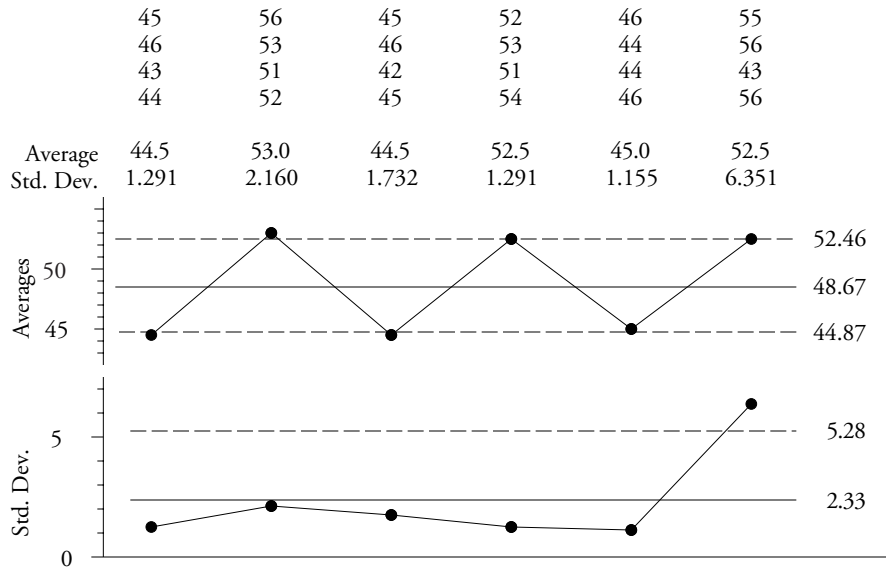


Figure 2 Average and Standard Deviation Chart

Until your subgroup size is greater than 10 there is no practical difference between the use of the Subgroup Ranges and the use of the Subgroup Standard Deviation Statistics. The choice of one over the other is simply a matter of preference. However, since you can always explain a range to anyone when the need arises, and since no one can really say exactly what the standard deviation statistic represents, I recommend using the Subgroup Ranges.

Approaches (1) and (2) are the default approaches. However, on those occasions when the average within-subgroup variation in approach (1) or (2) may have been inflated by some subgroups with excessive variation, it is appropriate to consider an alternative computation using either:

- (3) the Median of the  $k$  Subgroup Ranges or
- (4) the Median of the  $k$  Subgroup Standard Deviation Statistics.

Since these medians will have about two-thirds of the degrees of freedom of the averages in approaches (1) or (2), they will have about 20% greater uncertainty in the computed limits. This greater uncertainty that is inherent in approaches (3) and (4) makes them unsuitable as default computations. However, when approach (1) or (2) has inflated limits, the use of (3) or (4) will provide a more robust approach that will result in a more sensitive analysis, in spite of the lower efficiency of the median within-subgroup variation.

With our simple data set the Median Range is 3.50. With the scaling factors of  $A_4 = 0.758$  and  $D_6 = 2.375$  and the usual formulas we find the limits in Figure 3. All six subgroup averages and one subgroup range fall outside these limits.

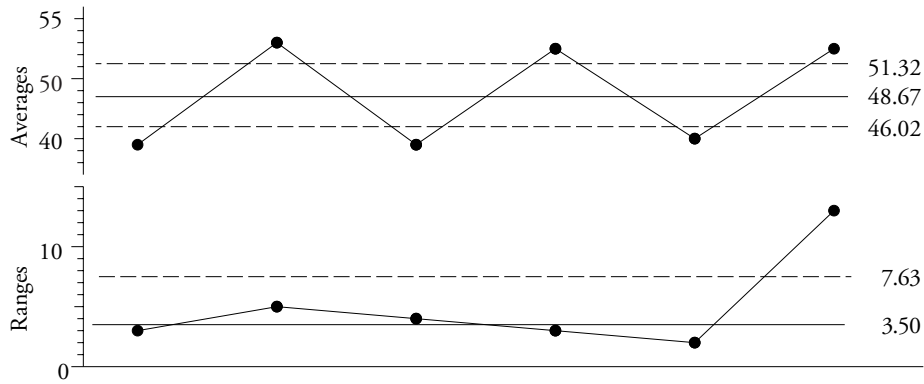


Figure 3 Average and Range Chart Based on Median Range

With our simple data set the Median Standard Deviation Statistic is 1.511. With the scaling factors of  $A_{10} = 1.689$  and  $B_{10} = 2.351$  and the usual formulas we find the limits in Figure 4. All six subgroup averages and one subgroup standard deviation fall outside these limits.

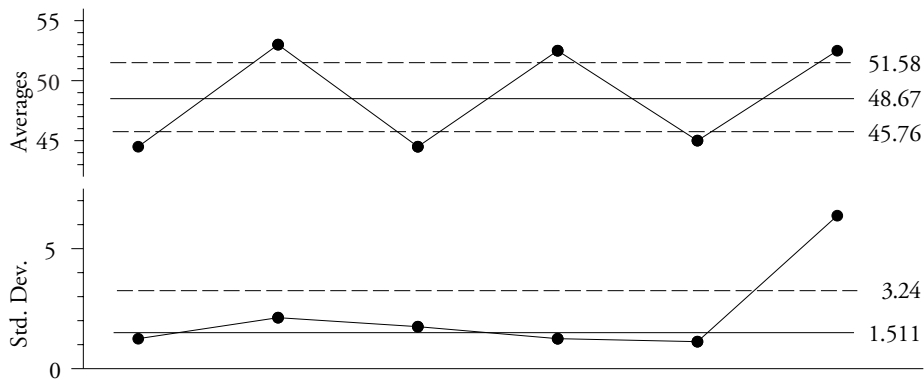


Figure 4 Average and Standard Deviation Chart Based on Median Standard Deviation

Thus, with only slight changes in the details, approaches (1), (2), (3), and (4) tell the same story about our simple data set. Because of the slight inflation in approaches (1) and (2), the alternate approaches (3) and (4) are more sensitive in this case.

AVERAGE CHARTS DONE ALMOST RIGHT

Some software defaults to another measure of within-subgroup variation known as the Pooled Variance. Specifically, the statistic used to compute limits would be:

- (5) the Square Root of the Average of the Subgroup Variance Statistics.

For our data set the Average of the Subgroup Variance Statistics (the Pooled Variance) is 8.778 and the square root of this value is 2.963. The Average Chart for our simple data set is shown in Figure 5 accompanied by a Variance Chart. The scaling factors for this chart are  $A_7 = 1.521$  and  $B_{12} = 3.449$ . (For a Standard Deviation Chart the scaling factor would be  $B_8 = 2.117$ .) These scaling factors depend upon both  $n$  and  $k$ . (Be careful, some software packages use the

wrong scaling factors here.) In Figure 5 we find only the one subgroup variance outside the limits. All of the subgroup averages fall inside the limits, which is completely different from the earlier results.

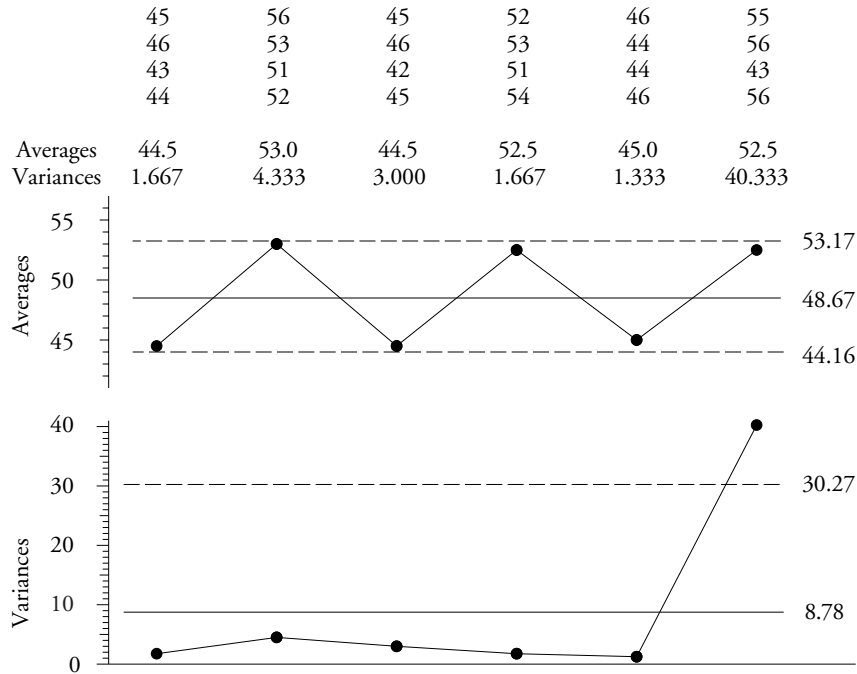


Figure 5 Average and Variance Chart

So what is the problem with using the Pooled Variance to compute limits?

The main problem with using the Pooled Variance to compute limits is the fact that the use of the Pooled Variance is inconsistent with the purpose of a process behavior chart. Specifically, by averaging the subgroup variances, this approach makes a strong assumption that all of the subgroups have the same amount of within-subgroup variation. This averaging of the subgroup *variances* means that any excessive within-subgroup variation will be squared before it is averaged in with the others. Thus, for any given data set, the Pooled Variance will always be more prone to being inflated than will the Average Range or the Average Standard Deviation. For example, the last variance in Figure 5 is almost 17 times the average of the first five variances. In comparison, the last Range in Figure 1 is only about 4 times the average of the first five ranges. Since the whole idea is to filter out the probable noise even though our computations may be contaminated by any signals that are present, the more robust approaches in Figures 1, 2, 3, and 4 are preferred over the less robust Pooled Variance approach in Figure 5.

But don't we use the Pooled Variance in ANOVA? Yes, we do. However, in ANOVA we are making a comparison between the Pooled Variance and the average of the *squared* signals. This average of squared signals is so aggregate and so inflated that any inflation in the Pooled Variance has minimal impact. Thus, while the Pooled Variance is correct and appropriate for ANOVA, and while it is theoretically acceptable for process behavior charts, it will always be less robust, and therefore less sensitive, than the first four approaches listed above. For this reason it

should not be used as the default computation when creating a process behavior chart.

But doesn't the Pooled Variance have more degrees of freedom than approaches (1) or (2)? Yes, it does. In this case the Pooled Variance has 18 degrees of freedom, the Average Range has 16.7 degrees of freedom, and the Average Standard Deviation has 17 degrees of freedom. So what does this mean? *In the case where there are no signals on the Range Chart*, these degrees of freedom would mean that the Pooled Variance would estimate the within-subgroup variation with 3% less uncertainty than would the Average Range! However, since we have a signal on the Range Chart, we know that the Pooled Variance, the Average Range, and the Average Standard Deviation are all inflated. This makes the degrees of freedom irrelevant.

It is important to remember that this computation is focused on how to filter out the noise. We are essentially using our estimate of dispersion to wrap up and carry out the garbage. Any package that doesn't leak will do the job. Since we are not trying to estimate some parameter value, we do not need to worry about slight differences in degrees of freedom. The robustness of the estimator is more important than the degrees of freedom. We can see this with approaches (3) and (4) which provide the most sensitive analyses of these data in spite of having, respectively, only 12 and 12.1 degrees of freedom. So while the Pooled Variance has the maximum number of degrees of freedom of any within-subgroup measure of dispersion, it is less robust than approaches (1) and (2). This is why no credible authority has ever recommended using the Pooled Variance as the default computation for a process behavior chart.

#### AVERAGE CHARTS DONE WRONG

A very common, and yet completely erroneous, method for computing the limits for a process behavior chart is to use:

- (6) the global standard deviation statistic.

This descriptive statistic is taught in virtually all classes in statistics, and is appropriate for summarizing a collection of values that are known to be homogeneous. However, the primary question addressed by a process behavior chart is whether or not the data show evidence of homogeneity. For our data set the global standard deviation statistic, computed using all 24 values together, is  $s = 4.860$ . Since this number is logically an estimate of the standard deviation of a single observation, we need to divide by the square root of the subgroup size to convert it into an estimate of the standard deviation of a subgroup average. Dividing by the square root of  $n = 4$ , and multiplying by 3, and then adding the result to the Grand Average, and subtracting the result from the Grand Average, we get the limits in Figure 6. While these limits can truly be said to be "three-standard-deviation limits" for the subgroup averages, they have been computed in a manner that is completely non-robust. This approach will always yield severely inflated limits in the presence of any signals on the Average Chart, and is therefore totally incorrect and inappropriate.

The mistake of using a descriptive statistic in an analytic procedure illustrated in Figure 6 is so fundamental that it has a name. It is known as the Quetelet Fallacy, after the Nineteenth Century social scientist Adolphe Quetelet. By using approach (6) many software programmers have unknowingly become disciples of Quetelet. The trap here is the fact that, *when the data are completely homogeneous*, this method of computing limits will approximate the limits found by

methods (1) and (2). However, as seen here, in the presence of signals, this approach is seriously non-robust. Shewhart rejected approach (6) on page 302 of *Economic Control of Quality of Manufactured Product*, published in 1931. It has been known to be incorrect ever since. Yet we still find it in the software.

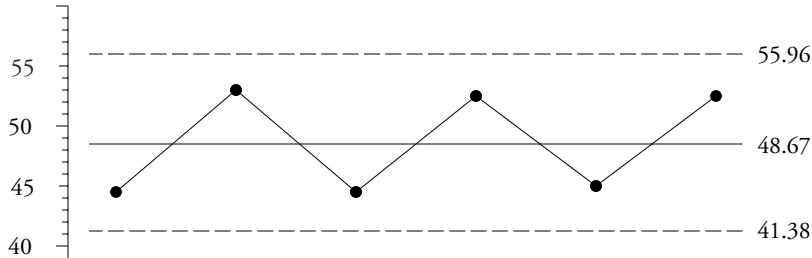


Figure 6 An Incorrect Average Chart

AVERAGE CHARTS DONE VERY WRONG

The next computational approach is in the same league as Hiebert’s Theorem, which states that:

$$[ \text{the number of toes per inch} ] \times 12 = [ \text{the number of toes per foot} ]$$

The major corollary of this theorem is: As long as you say it correctly, you can make patent nonsense sound plausible.

Some software options, and some authors, have suggested that since we are going to use three standard deviation limits, and since we are going to plot the subgroup averages, that we could use:

- (7) the standard deviation statistic computed from the subgroup averages.

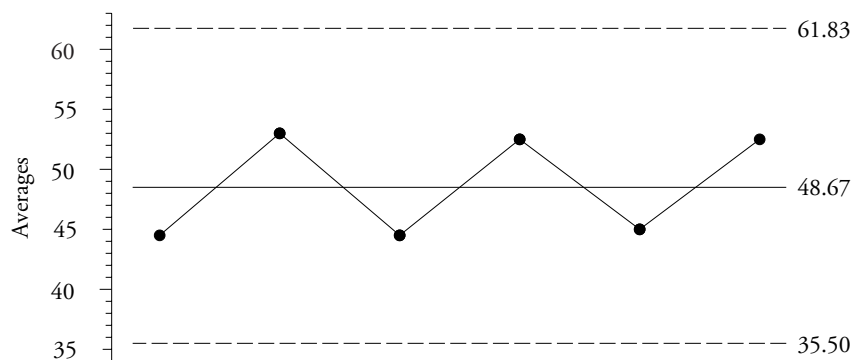


Figure 7 Another Incorrect Average Chart

Using the six subgroup averages in our data set we compute a standard deviation statistic of 4.389. Since this statistic was computed using the subgroup averages it is a logical estimate of the standard deviation of the subgroup averages. We do not need to adjust it by the square root of the subgroup size as we did with approach (6). The limits in Figure 7 are found by simply

multiplying the value above by 3.0 and then adding the result to the Grand Average and subtracting the result from the Grand Average. While these limits are “three-standard-deviation limits” for averages, they have been computed in a manner that is completely non-robust. This approach will always yield wildly inflated limits in the presence of any signals on the Average Chart, and is therefore totally incorrect and inappropriate.

The use of the global standard deviation statistic computed from the subgroup averages is very wrong. It has always been very wrong, and it will always be very wrong. Yet I found a plant in Indiana whose software used this method. I asked them how it worked. They said that while they never found a point outside the limits, their customers kept telling them that their processes were unpredictable! It is a sad day when your customer knows more about your process than you do.

Notice that neither approach (6) or approach (7) use the within-subgroup variation. Both of these approaches display the consequences of ignoring this foundation of modern statistical analysis.

Summary of Right and Wrong Ways to Compute Limits for Average Charts

Statistic Used for Limits	Lower Average Limit	Upper Average Limit	Result for Average Chart	Signals Between Subgroups	Signals Within Subgroups	Summary
(1) Average Range	45.02	52.31	6 of 6 Out	Robust	Robust	DEFAULT METHOD
(2) Average Std. Dev.	44.87	52.46	5 of 6 Out	Robust	Robust	DEFAULT METHOD
(3) Median Range	46.02	51.32	6 of 6 Out	Robust	Robust	ALTERNATE METHOD
(4) Median Std. Dev.	45.76	51.58	6 of 6 Out	Robust	Robust	ALTERNATE METHOD
(5) Pooled Variance	44.16	53.17	None Out	Robust	<b>More Inflated Than (1) or (2)</b>	ALMOST RIGHT
(6) Global Std. Dev. (X)	40.61	55.19	None Out	<b>Not Robust</b>	<b>Not Robust</b>	<b>WRONG</b>
(7) Global Std. Dev. (Avg.)	35.50	61.83	None Out	<b>Not Robust</b>	<b>Not Robust</b>	<b>VERY WRONG</b>

## SUMMARY

The use of wrong methods will undermine any analysis. You can avoid these and other incorrect methods by using the simple data set given here to evaluate the options in your software. By matching the limits given by your software to those in the following table you can determine which of these known methods is used by each option in your software. Of course, some options may not match up with any of the results in the table below. Options that do not match approaches (1), (2), (3), or (4) should be quarantined and avoided.



Be aware that all seven of the approaches described here will give substantially the same results when used with a homogeneous set of values, so do not attempt to evaluate alternative computations using a data set that does not contain signals. It is only in the presence of signals on the Average Chart or signals on the Range Chart that the various approaches will begin to yield different results.

The first four approaches allow you to get good limits from data containing signals. If you are interested in an effective analysis you need go no further. Playing around with other approaches will only get you in trouble.