

# Scaling Factors for Process Behavior Charts

## A Quick Reference Guide

Donald J. Wheeler

In the 1940s the War Production Board trained approximately 50,000 individuals in how to use process behavior charts (also known as control charts). At that time the computations were done by hand, and the emphasis was on making things as easy as possible for those doing these computations. As a result, sets of scaling factors were created for use in computing limits for the different charts.

With this approach the number of scaling factors increased as the number of different charts increased, resulting in a veritable alphabet soup. Some of these scaling factors date back to 1935, while others are of more recent vintage. Some of these scaling factors were created for purely academic situations while others were created for use with the data. Since different references would give different collections of these scaling factors, it was inevitable that over time notational differences would occur. In 1995 I collected all 35 of these scaling factors together, along with their derivations and formulas, in my *Advanced Topics in Statistical Process Control*. In this article I have extracted the 22 scaling factors which can be used to compute correct limits directly from the data. Thus, this article is intended as a quick, yet complete, reference guide that will help the user to know when each scaling factor is to be used.

### WHICH CHART TO USE

Process behavior charts will generally consist of two parts: a chart for location and a chart for dispersion. The first step in determining which chart to use will depend upon how the data are organized. Here there are two major categories. The data may be organized into  $k$  rational subgroups of size  $n$ , where each subgroup is logically homogeneous, or the data may consist of a stream of  $k$  individual values where these values are logically comparable. These two categories are represented in the first column of Table 1.

When your data consist of a sequence  $k$  individual values you should place these values on a Chart for Individual Values and a Moving Range, as shown in the last row of Table 1.

When your data are arranged into rational subgroups you will have to determine what summary statistic will be used to characterize (a) the location and (b) the dispersion for each of the  $k$  subgroups.

Commonly we will use the  $k$  Within-Subgroup Averages to characterize the location of each subgroup. Since the idea of rational subgrouping is to have homogeneous subgroups, there is rarely any need to use the Within-Subgroup Medians. (Formerly we would occasionally use the Within-Subgroup Medians in order to minimize the arithmetic for the benefit of the operators, but software has essentially made this usage obsolete.) This choice for the Within-Subgroup measure of location is shown in Column Two of Table 1.

Next you will need to determine which Within-Subgroup Measure of Dispersion you are going to use. You might use the Subgroup Ranges, the Subgroup Standard Deviation Statistics,

or the Root Mean Square Deviation Statistics.

For the sake of clarity, the Root Mean Square Deviation (RMS Deviation) is defined by its name (read in reverse) and is the measure of dispersion with  $n$  in the denominator. The Standard Deviation Statistic differs from the Root Mean Square Deviation in that it has  $(n-1)$  in the denominator.

With subgroup sizes of  $n = 15$  or less there is little practical difference between the Within-Subgroup Range, the Within-Subgroup Standard Deviation, and the Within-Subgroup RMS Deviation. They work the same and result in comparable charts. Your choice is more a matter of your preferences than anything else. These choices are listed in the Third Column of Table 1. (However, as shown in Table 1, when using Subgroup Medians there is no reason to use anything other than the Subgroup Ranges.) Once you have made the choices offered in Columns Two and Three of Table 1, the type of chart you will need to use is listed in the last column of Table 1.

Table 1: Which Type of Chart to Use

How Are Your Data Organized?	Within-Subgroup Measures of Location	Within-Subgroup Measures of Dispersion	Type of Chart to Use
$k$ Subgroups of Size $n$	AVERAGES	RANGES	Average and Range Chart
$k$ Subgroups of Size $n$	AVERAGES	STD. DEV.	Average and Std. Dev. Chart
$k$ Subgroups of Size $n$	AVERAGES	RMS DEV.	Average and RMS Dev. Chart
$k$ Subgroups of Size $n$	MEDIANS	RANGES	Median and Range Chart
$k$ Individual Values	—	—	Individual and Range Chart

Once you know which type of chart you will be using, you will need to determine which Summary Within-Subgroup Measure of Dispersion will be used to obtain your limits. As may be seen in the second column of Table 2, your choice here will consist of (1) the Average of your Within-Subgroup Measures of Dispersion or (2) the Median of your Within-Subgroup Measures of Dispersion. In general, for the default computation, the Average of your Within-Subgroup Measures of Dispersion is recommended. However, in those cases where some extremely large values may have inflated this average, you may switch to using the Median of your Within-Subgroup Measures of Dispersion to improve the sensitivity of your analysis.

## THE COMPUTATIONS

The limits for the Chart for Subgroup Averages and the limits for the Chart for Subgroup Medians will be found using Equation One:

$$\text{Grand Average} \pm \text{Scaling Factor} * \text{Summary Within Subgroup Measure of Dispersion} \quad (1)$$

The Grand Average is the typical Central Line for the Average Chart and may also be used for a Median Chart. Occasionally it may be replaced by the Median of the Subgroup Averages or even, in rare cases, by a nominal value. The scaling factors denoted by the letter  $A$  are used with Equation One.

The Central Line for the Chart for Dispersion is simply the Summary Within-Subgroup Measure of Dispersion used. The limits for this chart are found using Equations Two and Three:

$$\text{Lower Limit} = \text{Lower Scaling Factor} * \text{Summary Within Subgroup Measure of Dispersion} \quad (2)$$

$$\text{Upper Limit} = \text{Upper Scaling Factor} * \text{Summary Within Subgroup Measure of Dispersion} \quad (3)$$

Depending upon the statistic used, these scaling factors are denoted by the letters  $B$  or  $D$ . These scaling factors come in pairs. When the lower limit would be smaller than zero, Table 3 will simply not list a value for the lower scale factor.

When working with subgrouped data there will be occasions where the limits for individual values will also be required. (For example, these limits may be needed for comparison with the specification limits.) These values may also be computed from the same information used to compute the limits defined above. Limits for Individual Values may be found using Equation Four:

$$\text{Average} \pm \text{Scaling Factor} * \text{Summary Within Subgroup Measure of Dispersion} \quad (4)$$

The Average of the original data is the Central Line for these limits. The scaling factors for use with Equation Four are denoted by the letter  $E$ .

Table 2: Which Scaling Factors To Use

Type of Chart	Summary Within-Subgroup Measure of Dispersion Used	Scaling Factors			
		to use with Eq. 1	to use with Eq. 2	to use with Eq. 3	to use with Eq. 4
Average and Range Chart	Average Range	$A_2$	$D_3$	$D_4$	$E_2$
Average and Range Chart	Median Range	$A_4$	$D_5$	$D_6$	$E_5$
Average and Std. Dev. Chart	Average Std. Dev.	$A_3$	$B_3$	$B_4$	$E_3$
Average and Std. Dev. Chart	Median Std. Dev.	$A_{10}$	$B_9$	$B_{10}$	$E_6$
Average and RMS Dev. Chart	Average RMS Dev.	$A_1$	$B_3$	$B_4$	$E_1$
Average and RMS Dev. Chart	Median RMS Dev.	$A_5$	$B_9$	$B_{10}$	$E_4$
Median and Range Chart	Average Range	$A_6$	$D_3$	$D_4$	$E_2$
Median and Range Chart	Median Range	$A_9$	$D_5$	$D_6$	$E_5$
Individual and Range Chart	Average Moving Range *	—	—	$D_4$	$E_2$
Individual and Range Chart	Median Moving Range *	—	—	$D_6$	$E_5$

\* use scaling factor value for  $n = 2$

When these generic formulas are used with the scaling factors given in Table 3 they will all yield the appropriate three-sigma limits. These limits will filter out virtually all of the routine variation (regardless of the shape of the histogram) allowing you to identify any potential signals contained within the data.

The values for  $A_6$  and  $A_9$  are only given for odd values of  $n$ . This is because the rationale for using a Median Chart precludes the use of even values of  $n$ .

As noted earlier, there are 13 additional scaling factors. The scaling factors  $A$ ,  $A_8$ ,  $D_1$ ,  $D_2$ ,  $B_1$ ,  $B_2$ ,  $B_5$ , and  $B_6$  are not intended for use with data. Rather they are for the academic situation where the standard deviation of the process is said to be known. Since this essentially never happens in practice these scaling factors were not included in this summary.

Table 3: Values for the Scaling Factors

$n$	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$	$A_9$	$A_{10}$	$B_3$	$B_4$	$B_9$
2	3.760	1.880	2.659	2.224	4.447	—	—	3.143	—	3.267	—
3	2.393	1.023	1.954	1.091	2.547	1.187	1.265	2.082	—	2.568	—
4	1.880	0.729	1.628	0.758	1.951	—	—	1.689	—	2.266	—
5	1.595	0.577	1.427	0.594	1.638	0.691	0.712	1.465	—	2.089	—
6	1.410	0.483	1.287	0.495	1.438			1.313	0.030	1.970	0.031
7	1.277	0.419	1.182	0.429	1.297	0.509	0.520	1.201	0.118	1.882	0.120
8	1.175	0.373	1.099	0.380	1.190	—	—	1.114	0.185	1.815	0.188
9	1.095	0.337	1.032	0.343	1.107	0.412	0.419	1.044	0.239	1.761	0.242
10	1.028	0.308	0.975	0.314	1.039	—	—	0.985	0.284	1.716	0.287
11	0.973	0.285	0.927	0.290	0.982	0.350	0.356	0.936	0.322	1.678	0.324
12	0.925	0.266	0.886	0.270	0.933	—	—	0.893	0.354	1.646	0.357
13	0.884	0.249	0.850	0.253	0.891	0.308	0.312	0.856	0.382	1.619	0.384
14	0.848	0.235	0.817	0.239	0.854	—	—	0.823	0.407	1.593	0.409
15	0.816	0.223	0.789	0.226	0.821	0.276	0.280	0.794	0.428	1.572	0.431

  

$n$	$B_{10}$	$D_3$	$D_4$	$D_5$	$D_6$	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$
2	3.864	—	3.268	—	3.865	5.317	2.660	3.760	6.289	3.145	4.444
3	2.733	—	2.574	—	2.745	4.146	1.772	3.385	4.412	1.889	3.606
4	2.351	—	2.282	—	2.375	3.760	1.457	3.256	4.115	1.517	3.378
5	2.145	—	2.114	—	2.179	3.568	1.290	3.191	3.663	1.329	3.275
6	2.008	—	2.004	—	2.055	3.454	1.184	3.153	3.521	1.214	3.215
7	1.913	0.076	1.924	0.078	1.967	3.378	1.109	3.127	3.432	1.134	3.178
8	1.839	0.136	1.864	0.139	1.901	3.323	1.054	3.109	3.367	1.075	3.151
9	1.782	0.184	1.816	0.187	1.850	3.283	1.010	3.095	3.322	1.029	3.132
10	1.735	0.223	1.777	0.227	1.809	3.251	0.975	3.084	3.286	0.992	3.115
11	1.695	0.256	1.744	0.26	1.773	3.226	0.945	3.076	3.257	0.961	3.106
12	1.661	0.283	1.717	0.288	1.744	3.205	0.921	3.069	3.233	0.935	3.093
13	1.631	0.307	1.693	0.312	1.719	3.188	0.899	3.063	3.212	0.913	3.086
14	1.604	0.328	1.672	0.333	1.697	3.174	0.881	3.058	3.195	0.894	3.080
15	1.582	0.347	1.653	0.352	1.678	3.161	0.864	3.054	3.181	0.877	3.074

In my January 7 column we found the pooled variance approach to computing limits to be *almost* right. This approach uses scaling factors  $A_7$ ,  $B_7$ ,  $B_8$ ,  $B_{11}$ , and  $B_{12}$ . Since some of these scaling factors depend upon both  $n$  and  $k$ , and since the results are less robust than the approaches given above, these scaling factors were not included in this summary.

The software available today has added many different (and incorrect) options to the computation of limits for process behavior charts. The first foundation of the process behavior chart approach to making sense of your data is the use of within-subgroup measures of dispersion computed using rational (i.e. logically homogeneous) subgroups. The use of any other measure of dispersion is wrong. Likewise, irrational subgroupings are inappropriate. The second foundation of the process behavior chart is the use of three-sigma limits. The within-subgroup dispersion computed using rational subgroups creates the robustness that allows you

to get useful limits in spite of changes that may occur within your data. The three-sigma limits are sufficiently conservative to work with any type of data and still filter out virtually all of the routine variation, thereby minimizing false alarms while allowing you to detect those signals that are of economic importance.

Do not attempt to use any other measure of dispersion. Do not attempt to use something other than three-sigma limits. Virtually all of the correct ways of computing the appropriate limits are covered here. Use any other approach at your own peril.