

Comparing the \bar{X} -R and \bar{X} -s Control Charts

When I was first introduced to control charts, the \bar{X} -R chart was used the most. This was some years ago – back in the early 1980s. The control charts we did back then were done manually – doing the calculations using a calculator and manually plotting the points, averages, and control limits. There is still something to be said for teaching control charts using this manual method. We definitely tried to keep things simple.



There were a number of control charts I was exposed to back then. One was the \bar{X} -s chart. This chart is similar to the \bar{X} -R chart, except that the subgroup standard deviation is used instead of the subgroup range for the within subgroup variation.

In this publication, we will compare the two charts to see when you use one or the other. Traditionally, it has been suggested that you use the \bar{X} -R chart with subgroup sizes of 9 or less; you use the \bar{X} -s chart with subgroup sizes larger than 9. For the purposes of this publication, the chart to use is the one that gives you the best estimate of the process standard deviation. A simulation was developed to help do this.

The basic conclusion is that there is no reason not to use the \bar{X} -s chart all the time.

In This Issue:

- [Review of \$\bar{X}\$ -R and \$\bar{X}\$ -s Control Charts](#)
- [When to Use the \$\bar{X}\$ -R and \$\bar{X}\$ -s Control Charts Historically](#)
- [Estimating Sigma, the Process Standard Deviation](#)
- [Data and Simulation](#)
- [Control Charts from a Stable Process](#)
- [Simulation with a Stable Process](#)
- [Simulation with a Mixed Process](#)
- [Simulation with Out of Control Points](#)
- [Summary](#)
- [Quick Links](#)

You may download the workbook with the simulation [at this link](#). This workbook contains the results of the simulation as well as the VBA code to let you run your own simulations. The data used in the control chart figures below are also included in the workbook

Review of \bar{X} -R and \bar{X} -s Control Charts

These two control charts are quite similar. Each is really two charts – one to track the between subgroup variation and the other to track the within subgroup variation. You start with the individual values and then form subgroups of size n . For example, suppose you are tracking the weight of bags containing sand. You are measuring the weight for 5 consecutive bags at the start of each hour. You take those 5 weights and form a subgroup. In this case, the subgroup size, n , is 5. The average of each subgroup is calculated using the following:

$$\bar{X} = \frac{\sum X_i}{n}$$

where X_i are the individual X results in the subgroup. This is repeated and the \bar{X} values are plotted over time. So, the \bar{X} chart monitors how the subgroup averages vary over time – this is why it is said to track the between subgroup variation.

The two charts differ in how they track the within subgroup variation. For the \bar{X} -R chart, the subgroup range is used. The range within a subgroup is defined as the largest value minus the smallest value:

$$\text{Subgroup Range} = R = X_{\max} - X_{\min}$$

This is a measure of the variation within a subgroup. The subgroup range is plotted on the R chart, which is used to measure the variation within a subgroup from subgroup to subgroup.

For the \bar{X} -s chart, the subgroup standard deviation (s) is used:

$$\text{Subgroup Standard Deviation} = s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}$$

The value of s is a measure of the variation within a subgroup – it essentially measures how far each individual X value in the subgroup is away from the subgroup average, \bar{X} . The s values are plotted on the s chart. Like the R chart, the s chart monitors the variation within a subgroup from subgroup to subgroup.

Note that the calculation of s involves each individual X value in the subgroup while the subgroup range, R, uses only two of the individual values, the maximum and the minimum, regardless of how large the subgroup is.

Once you have sufficient data, you can begin to calculate the averages and control limits. These calculations are given below:

\bar{X} -R Chart:

$$\bar{R} = \frac{\sum R_i}{k}, \quad UCLr = D_4\bar{R}, \quad LCLr = D_3\bar{R}$$

$$\bar{\bar{X}} = \frac{\sum \bar{X}_i}{k}, \quad UCL\bar{x} = \bar{\bar{X}} + A_2\bar{R}, \quad LCL\bar{x} = \bar{\bar{X}} - A_2\bar{R}$$

\bar{X} -s Chart:

$$\bar{s} = \frac{\sum s_i}{k}, \quad UCLs = B_4\bar{R}, \quad LCLs = B_3\bar{R}$$

$$\bar{\bar{X}} = \frac{\sum \bar{X}_i}{k}, \quad UCL\bar{x} = \bar{\bar{X}} + A_3\bar{s}, \quad LCL\bar{x} = \bar{\bar{X}} - A_3\bar{s}$$

where k = number of subgroups, \bar{R} = the average subgroup range, \bar{s} = the average subgroup standard deviation, $\bar{\bar{X}}$ = the overall average and D_4 , D_3 , A_2 , B_4 , B_3 and A_3 are all constants that depend on subgroup size.

For more information on the \bar{X} -R and \bar{X} -s Control Charts, please visit [our SPC Knowledge Base](#).

When to Use the \bar{X} -R and \bar{X} -s Control Charts Historically

One reason that the \bar{X} -R chart has been used more than the \bar{X} -s chart is that it is easier to explain a range than the standard deviation. Back in the calculator days, the range was a lot easier to calculate than the standard deviation. The use of software now makes that point rather moot.

Another reason is that as the subgroup size, n , increases, the standard deviation becomes a better estimator of the process standard deviation – the variation in the process. Is that true? Quite often, the suggestion is to use the \bar{X} -R chart when n is less than 8 to 10. Different sources suggest either 8, 9 or 10. Otherwise, for larger subgroups, use the \bar{X} -s chart. The rest of this publication takes a look at how accurate this is.

Estimating Sigma – The Process Standard Deviation

One key use of a control chart is to estimate the process standard deviation, σ . This value is used in other calculations such as process capability. The within subgroup variation control chart is used to estimate the process standard deviation. The process standard deviation is measuring the variation in the individual values. The equations for estimating σ are given below for the two chart:

\bar{X} -R Chart:

$$\sigma = \frac{\bar{R}}{d_2}$$

\bar{X} -s Chart:

$$\sigma = \frac{\bar{s}}{c_4}$$

where d_2 and c_4 are constants that depend on subgroup size. We will use the value of σ to help compare how well the \bar{X} -R chart and \bar{X} -s chart perform and when you should use one or the other.

Data and Simulation

To determine how well the \bar{X} -R chart and \bar{X} -s chart perform, a simulation was run multiple times. A “population” of 5000 randomly generated numbers was created using the random number generator in the SPC for Excel software. The normal distribution was used with an average of 100 and standard deviation of 10. The actual values for the 5000 points are:

Population Average = 100.0257

Population Standard Deviation = 10.0401

These 5000 points represent all the possible outcomes from a process. The simulation uses subgroup sizes from $n = 2$ to 20. The number of subgroups for each simulation is $k = 20$. The simulation works in the following way:

- Sort the 5000 points in random order
- Take the first nk points
 - For example, if $n = 3$ and $k = 20$, the first $3 \cdot 20 = 60$ points are taken to form the subgroups and do the calculations
- For each n value (from 2 to 20)
 - Form subgroups of size n
 - Calculate the subgroup range (R) and subgroup standard deviation (s) for each subgroup
 - Calculate the average subgroup range (\bar{R}) and the average subgroup standard deviation (\bar{s}) for the 20 subgroups
 - Repeat this 1000 times
 - After the 1000 times, calculate the overall average subgroup range and overall average subgroup standard deviation for the 1000 times for that value of n
- Calculate the process standard deviation (σ) from the overall average subgroup range and overall average subgroup standard deviation
- Compare the results to each other and to the actual process standard deviation from the 5000 data points

The results of the simulation are described below.

Control Charts from a Stable Process

Drawing random samples from the 5000 point population described above is a stable process. Taking samples from it and creating a control chart will produce a process that is very stable – it is in statistical control. Below are examples of the \bar{X} - R chart and \bar{X} - s chart for a random sample taken from the 5000 points with $n = 5$.

Look at Figures 1 and 3, which are the \bar{X} charts for the \bar{X} - R chart and \bar{X} - s chart, respectively. The plotted values are the same for each chart – the \bar{X} values. In addition, the overall average is the same for both charts. The only differences are the values of the upper control limit (UCL) and the lower control limit (LCL). The control limits in Figure 1 are based on the average subgroup range \bar{R} while the control limits in Figure 3 are based on the average subgroup standard deviation \bar{s} . Note that the values are close.

Figures 2 and 4 represent the within subgroup variation as defined by either the subgroup range or the subgroup standard deviation. These two will often follow a similar pattern, e.g., a large subgroup range will generate a large subgroup standard deviation. Usually, the \bar{X} - R chart and \bar{X} - s chart will look similar.

The control charts in Figures 1 to 4 are in statistical control – there are no points beyond the control limits or patterns in the data. The process standard deviation can be calculated for the two charts as shown below:

$$\sigma = \frac{\bar{R}}{d_2} = \frac{23.894}{2.326} = 10.272$$

Figure 1: \bar{X} Chart with Limits Based on the Average Subgroup Range

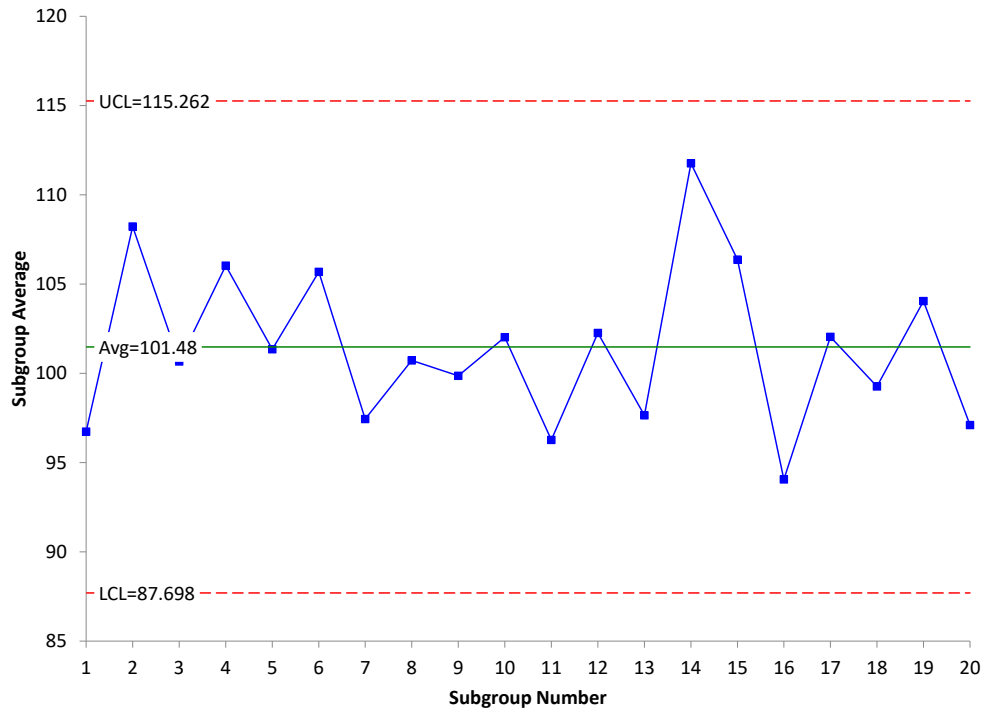


Figure 2: Subgroup Range Chart

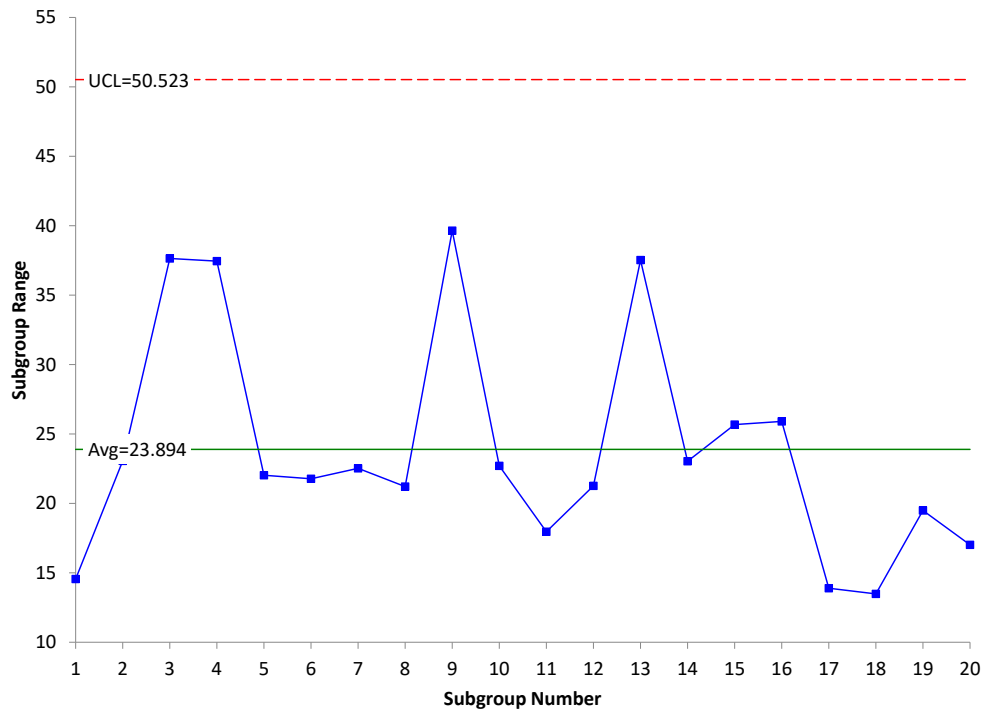


Figure 3 : \bar{X} Chart with Limits Based on the Average Subgroup Standard Deviation

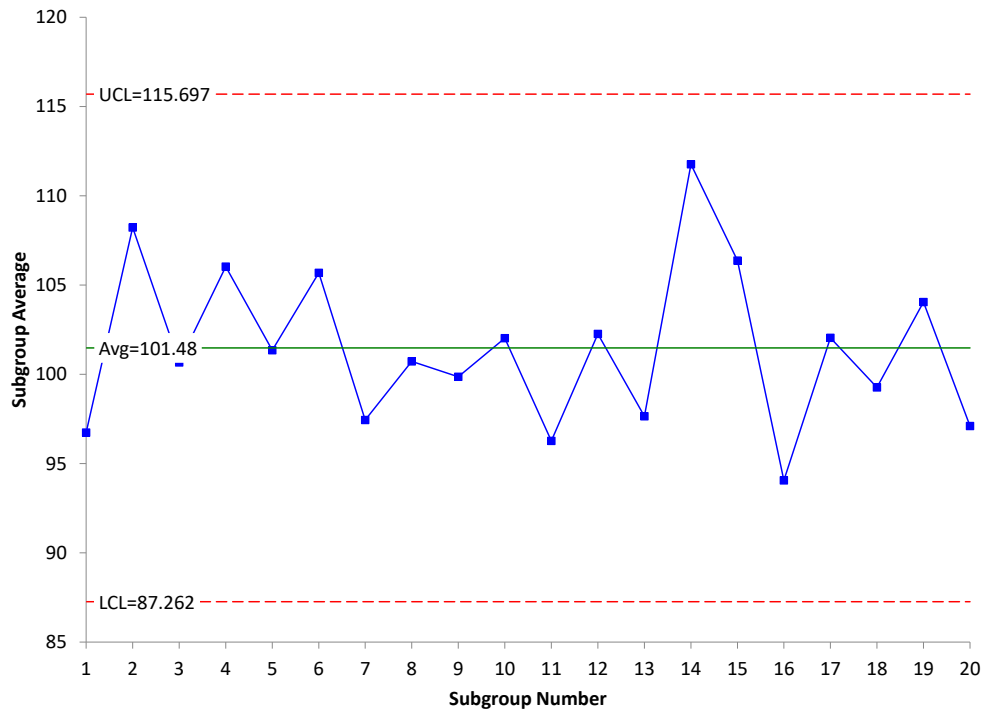
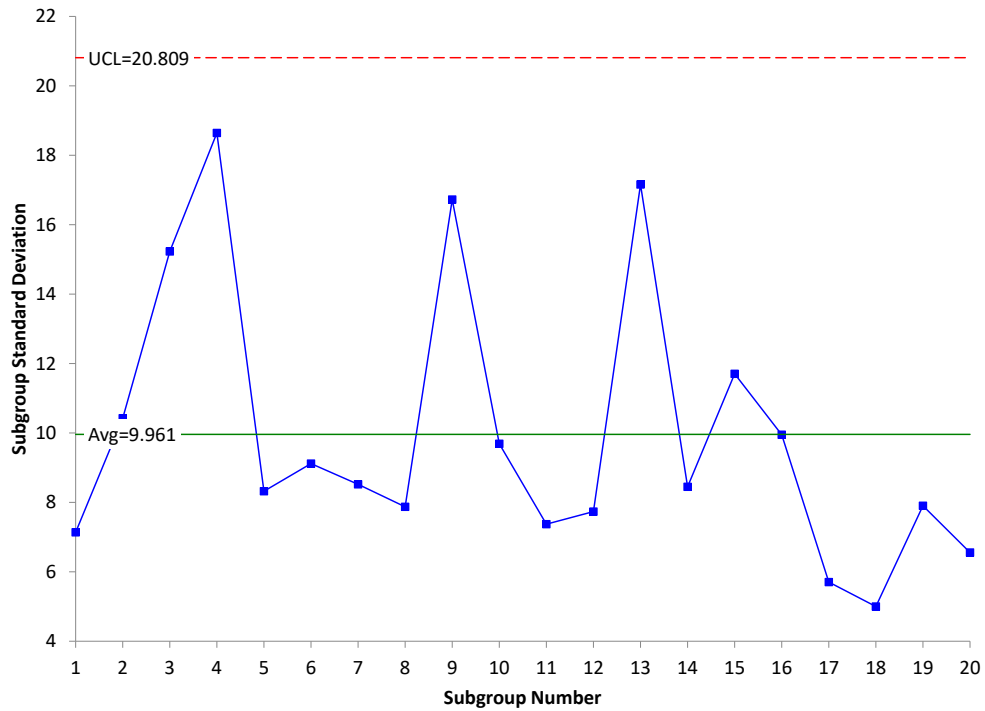


Figure 4: Subgroup Standard Deviation Chart



$$\sigma = \frac{\bar{s}}{c_4} = \frac{9.961}{0.94} = 10.597$$

The first one is based on the average subgroup range; the second on the average subgroup standard deviation. The two are similar. From above the “true” standard deviation is 10.0401. Both are near that as well.

Note that is one time going through the process of determining the process standard deviation. The simulation does this 1000 times for each subgroup size and chart and takes the overall average of the 1000 results for each chart to estimate σ .

Simulation with a Stable Process

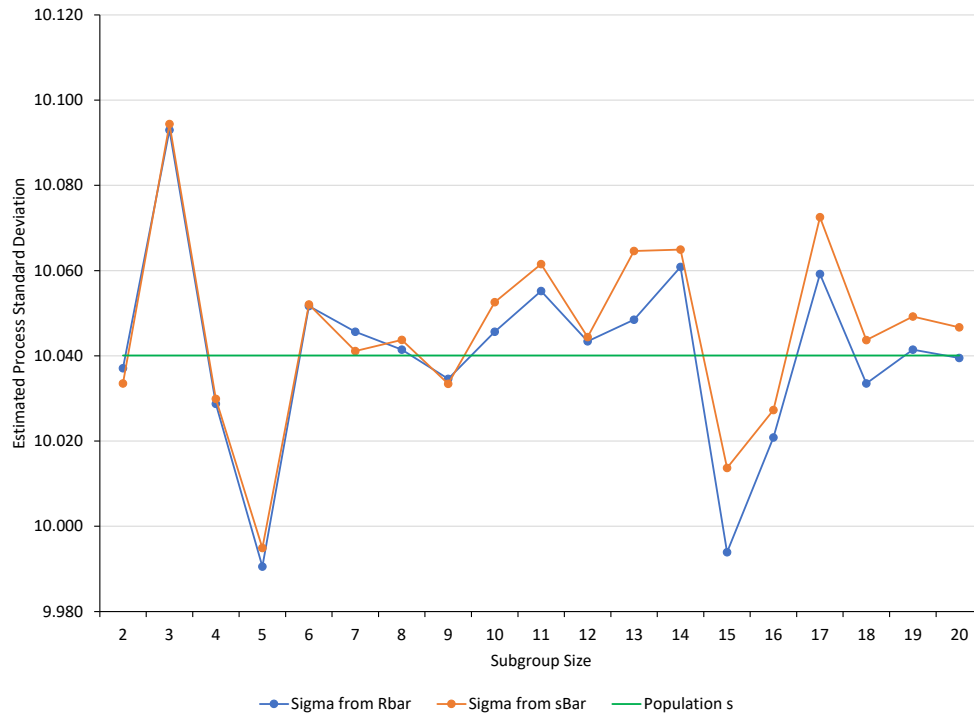
The simulation was run 1000 times to determine the average long-term value of σ as a function of subgroup size. The results are shown in Table 1.

Table 1: Simulation Results for a Stable Process

n	\bar{R}	\bar{s}	d_2	c_4	σ from \bar{R}	σ from \bar{s}
2	11.322	8.006	1.128	0.7979	10.0371	10.0335
3	17.087	8.946	1.693	0.8862	10.0930	10.0944
4	20.649	9.241	2.059	0.9213	10.0287	10.0299
5	23.238	9.395	2.326	0.9400	9.9906	9.9949
6	25.471	9.565	2.534	0.9515	10.0517	10.0520
7	27.163	9.633	2.704	0.9594	10.0456	10.0411
8	28.588	9.692	2.847	0.9650	10.0414	10.0437
9	29.803	9.725	2.97	0.9693	10.0346	10.0334
10	30.921	9.778	3.078	0.9727	10.0456	10.0526
11	31.905	9.814	3.173	0.9754	10.0552	10.0615
12	32.721	9.819	3.258	0.9776	10.0434	10.0444
13	33.522	9.857	3.336	0.9794	10.0485	10.0646
14	34.277	9.874	3.407	0.9810	10.0608	10.0649
15	34.699	9.836	3.472	0.9823	9.9939	10.0137
16	35.394	9.862	3.532	0.9835	10.0208	10.0273
17	36.092	9.916	3.588	0.9845	10.0592	10.0725
18	36.522	9.897	3.64	0.9854	10.0335	10.0437
19	37.043	9.911	3.689	0.9862	10.0414	10.0492
20	37.498	9.915	3.735	0.9869	10.0395	10.0467

This table contains the average of the 1000 estimates for \bar{R} and \bar{s} for the various subgroup sizes. It also includes the control chart constants needed to determine σ . The values of σ are included in the table. The values of σ are plotted in Figure 5 along with the “true” population average.

Figure 5: Values of σ from the Simulation for a Stable Process



The values of σ are close to the population value of 10.0401. For n values up to 9, the estimate of σ is about the same between the two charts. Sometimes the σ from \bar{R} is higher than the σ from \bar{s} . Other times it is not. Starting at $n = 10$, the value of σ from \bar{s} is always higher than the σ from \bar{R} . But, for a stable process, both give results fairly close to the population standard deviation. In fact, if you take the average of the σ results from $n = 10$ to $n = 20$, the average based on \bar{R} is closer to the population standard deviation than the average based on \bar{s} :

“True” population standard deviation = 10.0401

Based on $\bar{R} = 10.0402$ (for $n = 10$ to 20)

Based on $\bar{s} = 10.0492$ (for $n = 10$ to 20)

It appears, that for a stable process, it doesn’t matter whether you use the \bar{X} -R chart or the \bar{X} -s chart for subgroups $n = 2$ to 20.

But what about unstable processes? Two cases are considered below.

Simulation with Two Processes Mixed

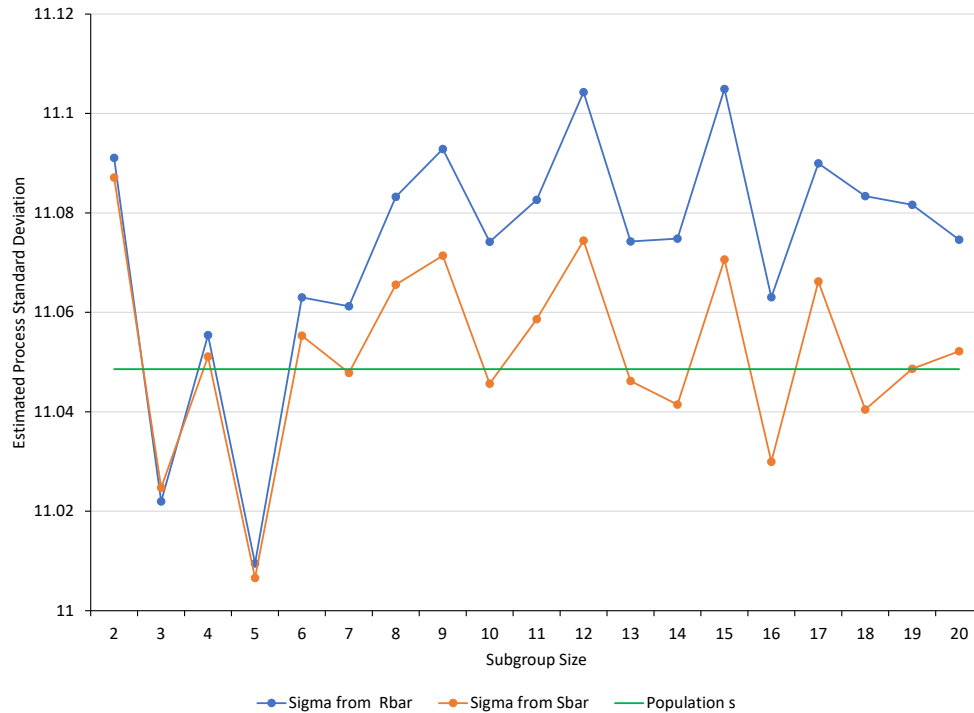
To simulate a less stable process, 500 of the original 5000 points were replaced with a distribution that had an average of 85 and a standard deviation of 10. So, 90% of this new distribution had an average of 100 and 10% had an average of 85. Both had a standard deviation of 10. The new distribution average and standard deviation were:

Population Average = 98.454

Population Standard Deviation = 11.0486

The simulation was run again using this distribution. The tabular results are not included here but are in the workbook containing the simulation [at this link](#). Figure 6 shows the results for σ .

Figure 6: Values of σ from the Simulation for Mixed Processes



For $n = 2$ to $n = 5$, the results are similar. But from $n = 6$ on, the value of σ estimated from \bar{R} is considerably higher than that estimated from \bar{s} . So, when there is some additional variation added to the mix, it appears that \bar{s} is a much better choice than \bar{R} to estimate σ .

Simulation with Out of Control Points Added

To include some out of control points, 200 points in the original distribution were deleted and 50 values each of 60, 65, 135 and 140 were added to the distribution. These represent out of control points since the original distribution was based on an average of 100 and a standard deviation of 10. About 4% of the points are beyond the 70 to 130 range defined by the original distribution. The new distribution had the following values:

Population Average = 100.025

Population Standard Deviation = 12.3904

The simulation was run again. The tabular results are not included here but are in the workbook containing the simulation [at this link](#). The results are shown in Figure 7.

Figure 7: Values of σ from the Simulation with Out of Control Points Added

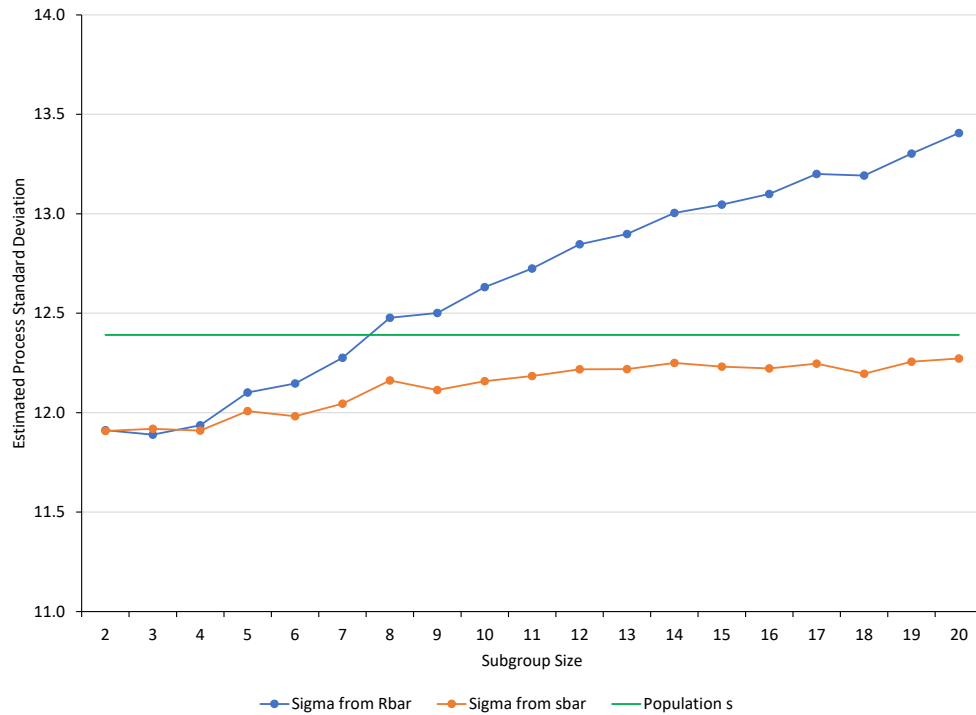


Figure 7 shows that the estimated process standard deviation from the two methods diverge considerably. The value of σ estimated by \bar{R} is inflated while the value of σ estimated by \bar{s} is deflated but does begin to approach the population standard deviation as n increases.

If the process is out of control, then neither method is that accurate although the \bar{s} method appears to be a better choice as n increases.

Summary

This publication examined using the estimated process standard deviation (σ) to determine if it is better to use the \bar{X} -R chart or the \bar{X} -s chart. Historically, if your subgroup size is less than 9, you would use the \bar{X} -R chart; if larger than 9, you would use the \bar{X} -s chart.

A simulation was used that examined the values of σ obtained by both charts using 20 subgroups with subgroup sizes (n) of 2 to 20. The simulation was run 1000 times for each subgroup size under three different scenarios.

The first scenario was a stable process. In this scenario, there was little different between the values of σ obtained from the two different charts. It really doesn't matter that much which chart you used.

The second scenario used a mixed process. In this scenario, using \bar{s} gave better results. \bar{R} seemed to overestimate the value of σ .

The third scenario added some out of control points to the process. Neither method predicted σ very well. This is not surprising since the process is out of control. However, it did appear that \bar{R} definitely inflated the value of σ .

Final conclusion: it appears to me that you might as well use the \bar{X} -s chart all the time. It works just as well as the \bar{X} -R chart for the smaller subgroup sizes and it seems to handle out of control situations better overall.

Quick Links

[Visit our home page](#)

[SPC for Excel Software](#)

[SPC Training](#)

[SPC Consulting](#)

[SPC Knowledge Base](#)

[Ordering Information](#)

Thanks so much for reading our publication. We hope you find it informative and useful. Happy charting and may the data always support your position.

Sincerely,

Dr. Bill McNeese
BPI Consulting, LLC