

Short Run SPC Part Three

The Robustness of Process Behavior Charts

Donald J. Wheeler

Parts one and two showed how to use zed charts and difference charts to track the underlying process while making different products. This part will illustrate both the robustness of the zed chart and an incorrect way of standardizing the data from the different products.

THE ROBUSTNESS OF PROCESS BEHAVIOR CHARTS

In this section we shall revisit the zed chart from the plant in Europe that was featured in Part One. We do this because, unlike the other examples in parts one and two, this process was being operated unpredictably. This unpredictability introduces some additional steps in the use of the *ANOMmR* chart. In addition, in retelling this story we shall also discover just how robust the process behavior chart is in practice.

Recall that three products labeled Red, Blue, and Green were produced in one production unit. The data are shown in Figure 1 in production order.

Batch	Prod	Value	Batch	Prod	Value	Batch	Prod	Value
1	Red	55.8	23	Blue	36.5	45	Red	56.4
2	Red	38.5	24	Blue	35.4	46	Red	59.1
3	Red	63.5	25	Blue	35.0	47	Blue	40.5
4	Blue	38.7	26	Red	52.5	48	Blue	39.6
5	Blue	40.4	27	Red	56.4	49	Blue	39.6
6	Blue	42.2	28	Red	58.8	50	Green	57.0
7	Red	53.1	29	Blue	43.6	51	Green	29.4
8	Red	55.7	30	Blue	37.3	52	Green	61.5
9	Red	52.5	31	Blue	40.3	53	Green	31.8
10	Green	30.4	32	Red	61.5	54	Red	65.1
11	Green	30.6	33	Red	83.5	55	Red	66.2
12	Green	28.2	34	Red	81.0	56	Red	66.8
13	Red	60.0	35	Green	29.0	57	Red	66.4
14	Red	65.1	36	Green	29.9	58	Red	66.9
15	Red	59.1	37	Green	32.6	59	Red	63.3
16	Green	29.3	38	Red	60.6	60	Blue	40.0
17	Green	27.2	39	Red	60.6	61	Blue	40.0
18	Green	29.0	40	Red	61.7	62	Blue	52.0
19	Green	29.0	41	Green	30.8	63	Green	29.1
20	Red	59.1	42	Green	30.8	64	Green	30.8
21	Red	58.4	43	Green	60.0	65	Green	29.8
22	Red	63.6	44	Red	56.6			

Figure 1: Data from Plant in Europe

When the plant personnel tried to create a standardized chart to visualize the process they used the traditional standardization transformation.

$$\text{Standardization Transformation} = Z = \frac{\text{Value} - \text{Average}}{\text{Global Standard Deviation Statistic}}$$

Each product value had the product average subtracted off, and the differences were divided by the global standard deviation for that product. The average for product Red is 60.92 and the standard deviation statistic is 8.17. The average for product Blue is 40.07 and the standard deviation statistic is 4.06. The average for product Green is 34.31 and the standard deviations statistic is 10.95. Using these values with the traditional standardization transformation above we get the “standardized” chart in Figure 2.

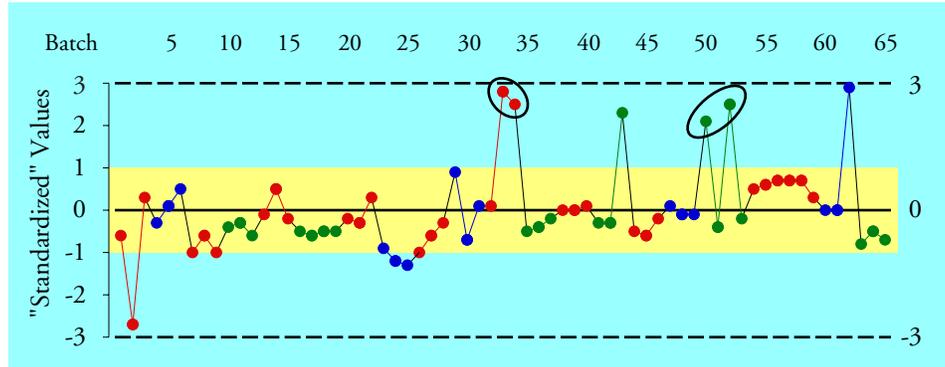


Figure 2: “Standardized” Chart for 65 Batches of Three Products

Figure 2 shows two runs beyond two sigma (points 33 and 34 and points 50 and 52). So while this chart does contain signals of unpredictable operation, it does not begin to show all the signals we find on the product-specific charts in Figure 3.

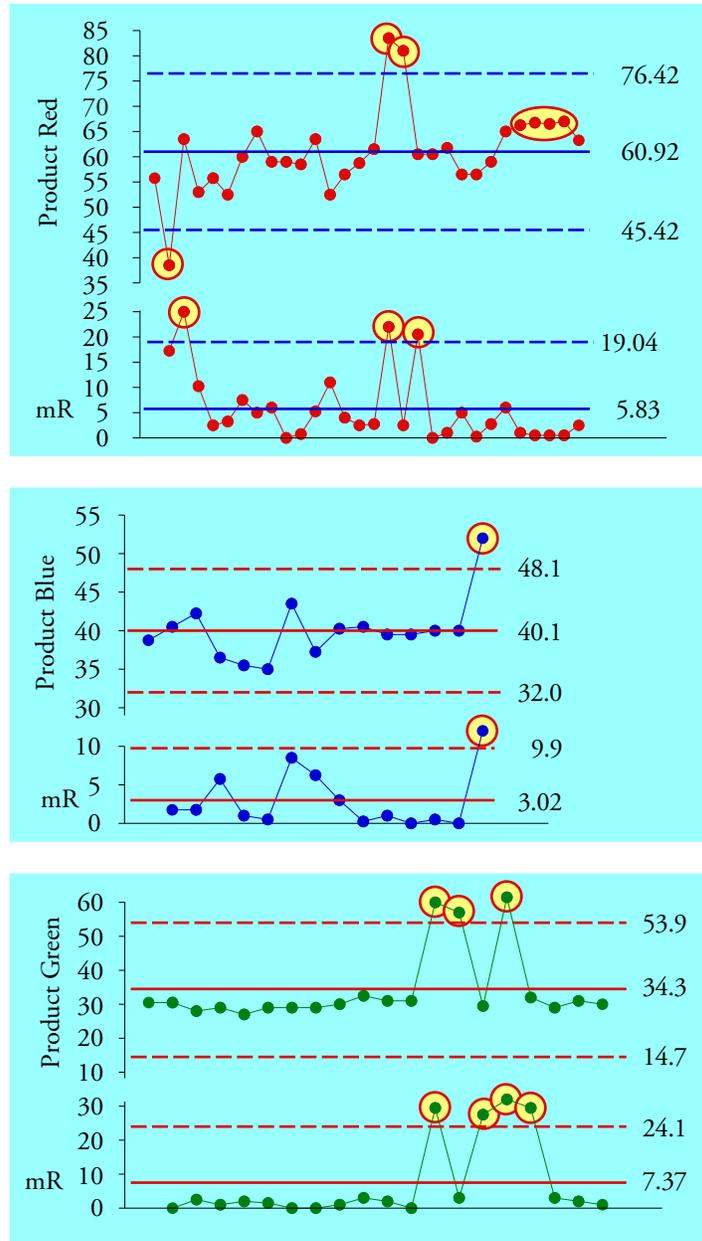


Figure 3: *XmR* Charts for Products Red, Blue, and Green

While the X-charts in Figure 3 have seven points outside the limits, these seven points all fall inside the limits in Figure 2. This happens because the global standard deviation statistic is the wrong measure of dispersion to use with process behavior charts. Whenever the standardization transformation is based on the global standard deviation most signals will be hidden.

To create a standardized process behavior chart we *must* use the zed transformation rather than the standardization transformation. The zed transformation for individual values uses a measure of dispersion known as *Sigma(X)* that is based upon either the average moving range or a median moving range:

$$\text{Zed Transformation} = \text{Zed} = \frac{\text{Value} - \text{Nominal}}{\text{Sigma}(X)}$$

The details were given in part one of this series, but to recap, the nominal values for Red Blue and Green are 60, 40, and 30 respectively. The average moving range for Product Red is 5.83, so $\text{Sigma}(X)$ is $5.83/1.128 = 5.17$ for Red. The average moving range for Product Blue is 3.02, so $\text{Sigma}(X)$ is $3.02/1.128 = 2.67$ for Blue. The average moving range for Product Green is 7.37, so $\text{Sigma}(X)$ is $7.37/1.128 = 6.53$ for Green. When these product specific zed transformations are applied to the 65 original values in Figure 1 we end up with the zed chart in Figure 4.

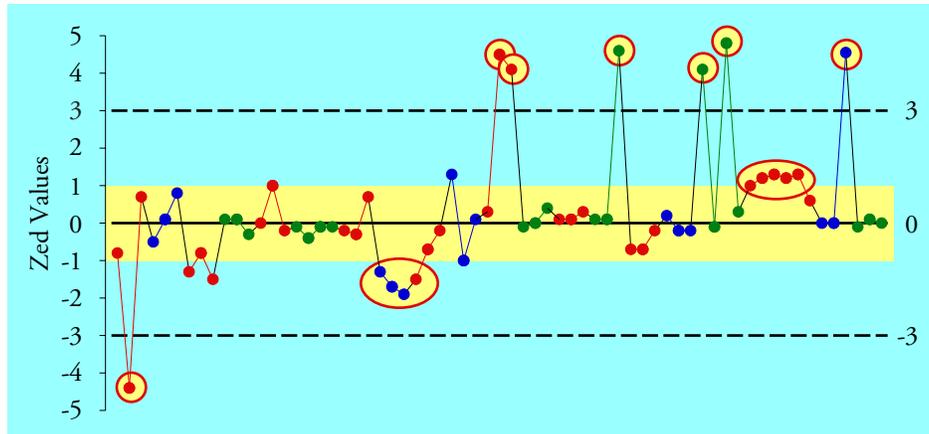


Figure 4: Zed Chart for 65 Batches of Three Products

Figure 4 shows the signals found in Figure 3: the seven points outside the limits and the run beyond one-sigma for Product Red. In addition it also shows another run beyond one-sigma for products Blue and Red that is not seen in Figure 3. So how we estimate dispersion can have a dramatic effect upon the chart.

Figure 5 compares the global standard deviation values used in creating Figure 2 with the $\text{Sigma}(X)$ values used for Figure 4. Since we divide by these measures of dispersion, the values in Figure 2 are deflated relative to those in Figure 4 as shown. It is this automatic deflation of the values that causes Figure 2 to hide the signals found in Figures 3 and 4.

Product	Fig. 2 Global Standard Deviation	Fig. 4 $mR\text{-bar}$ Based $\text{Sigma}(X)$	Deflation of Fig. 2 Values vs. Fig. 4
Red	8.17	5.17	63%
Blue	4.06	2.67	66%
Green	10.95	6.53	60%

Figure 5: Different Estimates of Dispersion Used

While Figure 4 does a much better job of revealing the process behavior than does Figure 2, Figure 4 is not perfect. The range charts in Figure 3 show that each of the average moving ranges has been inflated by one or more points above the upper range limits. This means that the $\text{Sigma}(X)$ values in Figure 5 are also inflated and the zed chart in Figure 4 is contaminated.

In spite of this contamination the zed chart is robust enough to separate the potential signals

from the probable noise. When we can do this we have an adequate basis for taking action to improve the process. Since the purpose for using a process behavior chart is to know when to take action, Figure 4 is good enough. Perfection is not required. As long as we follow the prescribed computations we can get limits that are good enough to identify the potential signals, and we can do this even when we use imperfect data to compute those limits. This robustness, this ability to get “good limits from bad data” is what makes the charts effective and useful for *process characterization*.

ESTIMATION VERSUS CHARACTERIZATION

But when it comes to estimating the process variation the contamination described above can be a problem. While the prescribed computations for the zed chart are robust enough for the charts to work as intended, the contamination introduced by the exceptional values can distort our estimates of process parameters. So when comparing variation from one product to another we may need to first decontaminate our statistics.

For product Red the average moving range is 5.828. But if we delete the four largest moving ranges and the last five small moving ranges (from the run beyond one sigma) the average of the 20 remaining moving ranges is 3.91.

For product Blue the average moving range was 3.016. But if we delete the one excessively large moving range the average of the 13 remaining moving ranges is 2.325.

For product Green the average moving range was 7.367. But if we delete the four excessively large moving ranges the average of the 15 remaining moving ranges is 1.421. These revised average moving ranges will provide better estimates of the process variation for the three products than the ones we used in creating the zed chart in Figure 4.

Before we create an *ANOMmR* chart to compare the variation across products we not only need to use decontaminated statistics, but we also need to have the same baseline lengths. Since our decontaminated estimate for product Blue has only 13 moving ranges, we truncate the other two sets of moving ranges to use only the first 13 values used above for the decontaminated estimates.

For product Red we use 13 moving ranges between batch 4 and batch 16 to get an average moving range of 4.673. And for product Green we use the first thirteen moving ranges within the limits to get an average of 1.426.

Using these three revised average moving ranges, our grand average moving range becomes 2.808. For $m = 3$ baselines with $k = 14$ values each our ten percent *ANOMmR* scaling factors from the table below are 0.593 and 1.450. The resulting *ANOMmR* chart is found in Figure 6.

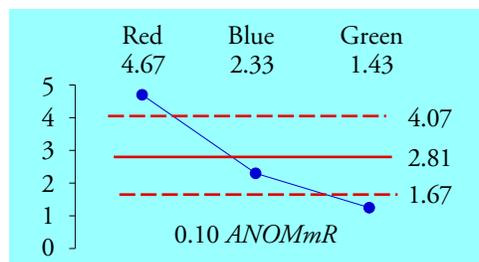


Figure 6: *ANOMmR* Chart for the Three Products

Thus, we find that the different products have detectably different amounts of variation, making the zed chart a necessity for tracking the underlying process. Dividing each of these average moving ranges by 1.128 gives the revised $\text{Sigma}(X)$ values in the last column of Figure 7. Comparing the first column with the last we find that the global standard deviation statistics are two to eight times larger than the revised estimates of dispersion.

	Fig. 2 Global Standard Deviation	Fig. 4 mR -bar Based $\text{Sigma}(X)$	Fig. 8 Revised $\text{Sigma}(X)$ $k = 14$
Product			
Red	8.17	5.17	4.14
Blue	4.06	2.67	2.06
Green	10.95	6.53	1.27

Figure 7: Different Estimates of Dispersion

POLISHING THE ZED CHART

Figure 2 shows a standardization chart done wrong. Figure 4 shows how the use of the zed transformations avoids burying the signals and properly separates the signals of process changes from the routine variation of the production process. But could we polish up the zed chart by using our less contaminated estimates of dispersion developed for the $ANOMmR$ chart? Yes, we could, but we seldom need to do so for the following reason.

Process behavior charts are intended for process *characterization* rather than *estimation*. If you have already found signals, then rather than revising the limits, the proper response is to take action to identify the assignable causes and then either control these causes or compensate for them in production. Any additional signals you might find by polishing the limits will always be the smaller signals. If you are already taking care of the larger signals, you will naturally recompute the limits as changes are made to the process, and eventually you will get around to the smaller signals. And if you are not taking care of the larger signals, then you certainly do not need to find the smaller signals. Thus, when the process is characterized as unpredictable, the revision of the limits is moot. Action is required.

On the other hand, when the process is reasonably predictable, your estimates of the parameters will already be as good as the amount of data will allow them to be, and there is no need to revise the limits.

Nevertheless, to further illustrate both the sensitivity of the zed transformation and the insensitivity of the traditional standardization transformation, we will polish the limits for our zed chart from Figure 4. Using the target values of 60, 40, and 30 respectively, and the revised $\text{Sigma}(X)$ values from Figure 7, we get the zed-chart in Figure 8.

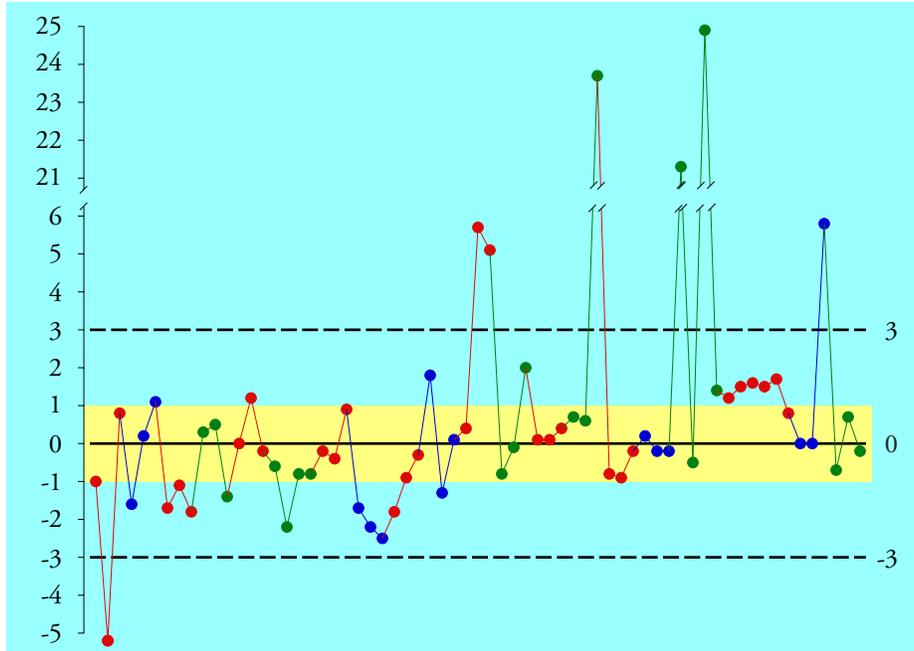


Figure 8: Zed Chart Using Revised Estimates of $\sigma(X)$

Figure 8 shows that this process is subject to excursions of up to 25 sigma on the high side and excursions of more than 5 sigma on the low side. Any process that meanders around over a range of 30 sigma represents a huge opportunity for improvement since no process needs elbow room in excess of six sigma when it is being operated predictably.

Yet Figure 2, reproduced here as Figure 9, shows all of these points falling within the three sigma limits. This illustrates how completely the traditional standardization transformation buries signals—a running record that spanned 30 standard deviations was compressed five-fold so that it would fit within a span of 6 standard deviations!

This is precisely what the global standard deviation is supposed to do. Since it is built on the assumption that the data are completely homogeneous, it will always do its best to shoe-horn every histogram into the interval [*Average* ± 3 *Global Standard Deviations*].

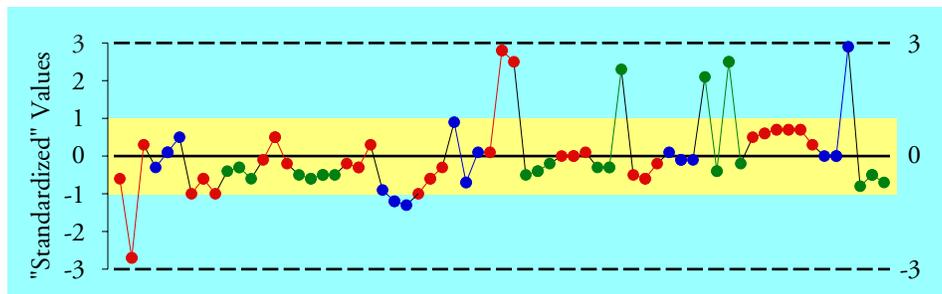


Figure 9: “Standardized” Chart for 65 Batches of Three Products

The traditional standardization transformation is a steam roller that will successfully flatten things out even when the data contain outliers that are up to 25 standard deviations away from

the average!

But rather than hiding the signals by sweeping them under the rug, the purpose of a process behavior chart is to examine the data for evidence of a lack of homogeneity. This is why *you should never, ever use a global standard deviation statistic with a process behavior chart*—it is simply incompatible with the objective of the charts.

The original zed chart in Figure 4 is not completely free of the contaminating effects of the many large ranges. However, in spite of this contamination, it correctly identified all of the potential signals. Polishing the zed transformations as we did in Figure 8 did not change the story already told by Figure 4, but it did show the magnitude of the process changes with greater clarity.

This is why we say the process behavior chart approach is robust. As long as we use the prescribed computations we will get a chart that is good enough to identify potential signals (e.g. Fig. 4). This will allow us to begin to operate our processes up to their full potential. We don't have to have the exact limits, and neither do we have to have the exact zed transformations in order to do this. As long as we use the two-point moving ranges to characterize dispersion when working with individual values, we can get good limits from bad data.

SUMMARY

When seeking to create a standardized chart it is important to avoid the proscribed computations. The traditional standardization transformation taught in introductory classes in statistics is inappropriate when creating a standardized process behavior chart. When working with individual values, the zed transformations will always use estimates of dispersion that are based on two-point moving ranges.

This is the reason standardized process behavior charts are called zed-charts. The label “z” is already attached to the traditional standardization transformation, and as we saw in Figure 9, this “z-transformation” is completely incorrect for use with any process behavior chart.

So, before you use any software for “standardized charts” you need to make sure that the estimates of dispersion are not based on product-by-product global standard deviation statistics. A good way to discover if the software does standardized charts correctly is to use the data in Figure 1 with the software and see whether you end up with Figure 2 or Figure 4.

$$z = \frac{\text{Original Value} - \text{Average}}{\text{Standard Deviation Statistic}}$$


Analysis of Mean Moving Ranges
ANOMmR Scaling Factors
 $\alpha = 0.10$

For comparing m Average Moving Ranges,
each of which comes from an XmR chart for k original values:
Multiply the Grand Average Moving Range by the following ANOMmR scaling factors
to obtain upper and lower detection limits.

Average Moving Ranges that fall outside the detection limits
are detectably different from the Grand Average Moving Range.

Use this table for an Exploratory analysis with a ten percent overall risk of a false alarm.

0.10 ANOMmR

	m=2		m=3		m=4		m=5		m=6		m=7	
	LL	UL										
k=5	0.513	1.487	0.349	1.787	0.307	1.922	0.281	2.018	0.262	2.091	0.248	2.151
k=6	0.556	1.444	0.398	1.711	0.357	1.828	0.330	1.911	0.312	1.976	0.297	2.026
k=7	0.590	1.410	0.440	1.651	0.397	1.756	0.370	1.830	0.352	1.886	0.337	1.933
k=8	0.618	1.382	0.473	1.606	0.432	1.700	0.406	1.767	0.387	1.819	0.373	1.859
k=9	0.640	1.360	0.500	1.568	0.461	1.654	0.435	1.717	0.416	1.764	0.402	1.802
k=10	0.656	1.344	0.523	1.539	0.484	1.620	0.458	1.677	0.441	1.720	0.427	1.757
k=11	0.673	1.327	0.543	1.512	0.506	1.587	0.482	1.640	0.463	1.681	0.450	1.715
k=12	0.689	1.311	0.564	1.486	0.527	1.558	0.502	1.609	0.485	1.648	0.471	1.679
k=13	0.700	1.300	0.578	1.467	0.541	1.535	0.519	1.582	0.501	1.619	0.489	1.653
k=14	0.712	1.288	0.593	1.450	0.557	1.514	0.534	1.559	0.518	1.596	0.504	1.622
k=15	0.722	1.278	0.606	1.433	0.571	1.495	0.548	1.537	0.532	1.571	0.520	1.598
k=16	0.731	1.269	0.618	1.418	0.583	1.477	0.561	1.518	0.545	1.550	0.532	1.576
k=17	0.739	1.261	0.629	1.405	0.596	1.461	0.575	1.501	0.560	1.530	0.547	1.554
k=18	0.747	1.253	0.640	1.393	0.608	1.446	0.585	1.485	0.572	1.515	0.559	1.537
k=19	0.754	1.246	0.649	1.382	0.618	1.434	0.596	1.471	0.582	1.501	0.569	1.522
k=20	0.760	1.240	0.659	1.370	0.626	1.424	0.606	1.457	0.591	1.485	0.579	1.506
k=21	0.766	1.234	0.666	1.360	0.635	1.412	0.615	1.444	0.600	1.472	0.588	1.493
k=22	0.771	1.229	0.673	1.351	0.642	1.401	0.622	1.432	0.609	1.459	0.598	1.479
k=23	0.777	1.223	0.682	1.343	0.650	1.392	0.631	1.421	0.617	1.447	0.606	1.468
k=24	0.781	1.219	0.688	1.336	0.658	1.383	0.638	1.413	0.624	1.437	0.613	1.457
k=25	0.786	1.214	0.694	1.329	0.664	1.375	0.645	1.403	0.632	1.427	0.620	1.447
k=26	0.790	1.210	0.699	1.323	0.672	1.365	0.653	1.395	0.638	1.419	0.628	1.436
k=27	0.794	1.206	0.704	1.317	0.677	1.358	0.658	1.387	0.644	1.412	0.633	1.428
k=28	0.798	1.202	0.710	1.311	0.683	1.350	0.664	1.379	0.650	1.401	0.640	1.418
k=29	0.801	1.199	0.714	1.304	0.687	1.345	0.669	1.373	0.656	1.394	0.645	1.411
k=30	0.805	1.195	0.719	1.299	0.693	1.338	0.674	1.366	0.661	1.387	0.650	1.403
k=40	0.840	1.160	0.765	1.254	0.741	1.289	0.724	1.313	0.713	1.332	0.702	1.345
k=50	0.856	1.144	0.789	1.223	0.767	1.254	0.752	1.274	0.741	1.291	0.731	1.303

For additional tables see "Short Run SPC Part Two."

