

## The Problem of Long-Term Capability

Poor labels lead to incorrect ideas

Donald J. Wheeler

Based on some recent inquiries there seems to be some need to review the four capability indexes in common use today. A clear understanding of what each index does, and does not do, is essential to clear thinking and good usage. To see how to use the four indexes to tell the story contained in your data and to learn how to avoid a common pit-fall read on.

### THE FOUR INDEXES

Four indexes in common use today are the capability ratio,  $C_p$  the performance ratio,  $P_p$ , the centered capability ratio,  $C_{pk}$  and the centered performance ratio,  $P_{pk}$ . The formulas for these four ratios are:

$$\text{Capability Ratio} = C_p = \frac{USL - LSL}{6 \text{ Sigma}(X)}$$

$$\text{Performance Ratio} = P_p = \frac{USL - LSL}{6 s}$$

$$\text{Centered Capability Ratio} = C_{pk} = \frac{2 \text{ DNS}}{6 \text{ Sigma}(X)}$$

$$\text{Centered Performance Ratio} = P_{pk} = \frac{2 \text{ DNS}}{6 s}$$

To understand these ratios we need to understand the four components used in their construction. The difference between the specification limits,  $USL - LSL$ , is the specified tolerance. It defines the *total space available* for the process.

The distance to the nearer specification,  $DNS$ , is the distance from the average to the nearer specification limit. Operating with an average that is closer to one specification than the other effectively narrows the space available to the process. It is like having a process that is centered within limits that have a specified tolerance =  $2 \text{ DNS}$ . Thus, the numerator of both the centered capability ratio and the centered performance ratio characterizes the *effective space available* due to the fact that the process is not centered within the actual specification limits.

$\text{Sigma}(X)$  denotes any one of several within-subgroup measures of dispersion. One such measure would be the average of the subgroup ranges divided by the appropriate bias correction factor. Another such measure is the average of the subgroup standard deviation statistics divided by the appropriate bias correction factor. The quantity denoted by  $6 \text{ Sigma}(X)$  represents the *generic space required by a process* when that process is operated up to its full potential.

The global standard deviation statistic,  $s$ , is the descriptive statistic introduced in every

statistics class. Since it is computed using all of the data, it effectively treats the data as one homogeneous group of values. This descriptive statistic is useful for summarizing the past, but if the process is not being operated up to its full potential the changes in the process will tend to inflate this global measure of dispersion. Thus, this measure of dispersion simply describes the past without respect to whether the process has been operated up to its full potential or not. The denominators of  $6s$  define the *space used by the process in the past*.

A glance at the formulas above will reveal that the only difference between the capability indexes and the corresponding performance indexes is simply which measure of dispersion is used. The performance indexes use the global standard deviation statistic to describe the past. The capability indexes use a within-subgroup measure of dispersion to approximate the process potential. Whenever and wherever this profound difference between these measures of dispersion is not appreciated it is inevitable that capability confusion will follow.

Depending upon what is happening with the underlying process, the four indexes above can be four estimates of one quantity, four estimates of two different quantities, or even four estimates of four different quantities. This variable nature of what these index numbers represent has complicated their interpretation in practice. As a result, many different explanations have been offered. Unfortunately, some of these explanations have been flawed and even misleading.

#### WHAT THE FOUR INDEXES MEASURE

Using these four components defined above, we see that the capability ratio,  $C_p$ , expresses the space available within the specifications as a multiple of the space required by the process when it is centered within the specifications and is operated predictably. It is the *space available* divided by the *space required* under the best possible circumstances.

The performance ratio,  $P_p$ , expresses the *space available* within the specifications as a multiple of the *space used in the past* by this process. If the process has been operated up to its full potential, the space used in the past and the space required by the process will be essentially the same, and the performance ratio will be quite similar to the capability ratio. If the process has not been operated up to its full potential then the space used by the process in the past will always exceed the space required by the process, and the performance ratio will be smaller than the capability ratio. Thus, the agreement between the capability ratio and the performance ratio will characterize the extent to which the process is, or is not, being operated predictably.

The centered capability ratio,  $C_{pk}$ , expresses the *effective space available* as a multiple of the *space required* by the process when it is operated predictably at the current average. It is the effective space available divided by the space required. The extent to which the centered capability ratio is smaller than the capability ratio will characterize how far off-center the process is operating.

The centered performance ratio,  $P_{pk}$ , expresses the *effective space available* as a multiple of the *space used by the process in the past*. This ratio essentially describes the process as it is, where it is, without any consideration of what the process has the potential to do. The extent to which the centered performance ratio is smaller than the performance ratio is a characterization of how far off-center the process has been operated.

The relationship between these four indexes may be seen in Figure 1. There the top tier represents either the actual capability of a process that is operated predictably, or the

hypothetical capability of a process that is operated unpredictably. The bottom tier represents the actual performance of a process that is operated unpredictably. The left side represents what happens when the process is centered at the mid-point of the specifications, while the right side takes into account the effect of having an average value that is not centered at the mid-point of the specifications.

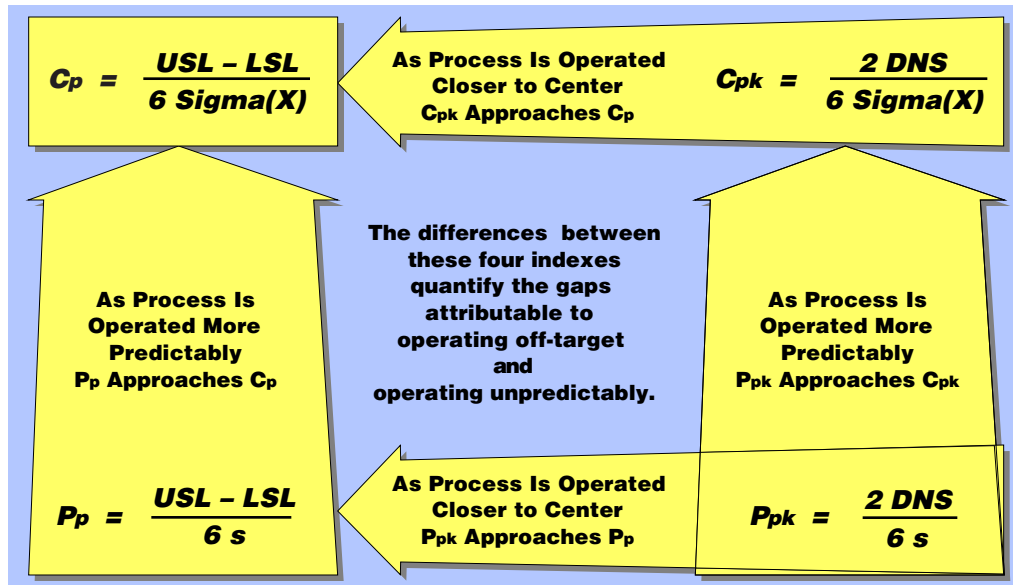


Figure 1: How the Capability and Performance Indexes Define the Gaps Between Performance and Potential

Thus, while the top tier of Figure 1 is concerned with the process potential, the bottom tier describes the process performance. As a process is operated ever more closely to its full potential, the values in the bottom tier will move up to be closer to those in the top tier.

While the left side implicitly assumes the process is centered within the specifications, the right side takes into account the extent to which the process may be off-center. As a process is operated closer to the center of the specifications the values on the right will move over to be closer to those on the left.

Thus, when a process is operated predictably and on target, the four indexes will be four estimates of the same thing. This will result in the four indexes being close to each other. (Since the indexes are all statistics, they will rarely be exactly the same.)

When a process is operated predictably but is not centered within the specifications, the discrepancy between the right and left sides of Figure 1 will quantify the effects of being off center. With a predictable process, the two indexes on the right side of Figure 1 will both estimate the same thing while the two indexes on the left side will be two estimates of another quantity.

When a process is operated unpredictably, the indexes in the bottom row of Figure 1 will be smaller than those in the top row, and these discrepancies will quantify the gap due to unpredictable operation.

When a process is operated unpredictably and off target, the four indexes will represent four different quantities.

Thus, while the Capability Ratio,  $C_p$ , is the best-case value, the Centered Performance Ratio,  $P_{pk}$ , is the worst-case value. The gap between these two values is the opportunity that exists for improving the current process by operating it up to its full potential.

The Capability Ratio,  $C_p$ , approximates what can be done without reengineering the process. If this best-case value is good enough, then the current process can be made to operate in such a way as to meet the process requirements. Experience has repeatedly shown that it is cheaper to learn how to operate the existing process predictably and on-target than it is to try to upgrade or reengineer that process.

Thus, by comparing the four capability and performance indexes you can quickly and easily get some idea about how a process is being operated. How close is it to being operated up to its full potential? Is it being operated on target? Will it be necessary to reengineer the process, or can it be made to meet the process requirements without the trouble and expense of reengineering?

#### EXAMPLE ONE

Figure 2 contains 260 observations from a predictable process. The corresponding average and range chart is shown in Figure 3. The specifications for this process are  $10.0 \pm 3.5$ .

Figure 2: 260 Observations from a Predictable Process (In Subgroups of Size 5)

No.	Values	$\bar{X}$	R	No.	Values	$\bar{X}$	R	No.	Values	$\bar{X}$	R
-1-	12 9 9 8 9	9.4	4	-19-	12 11 10 15 12	12.0	5	-36-	9 10 10 11 10	10.0	2
-2-	11 13 11 11 11	11.4	2	-20-	9 12 11 9 9	10.0	3	-37-	11 8 11 9 11	10.0	3
-3-	14 6 11 8 9	9.6	8	-21-	13 11 12 7 8	10.2	6	-38-	12 11 13 8 10	10.8	5
-4-	10 12 13 8 10	10.6	5	-22-	9 9 13 8 7	9.2	6	-39-	9 11 11 11 11	10.6	2
-5-	11 10 12 11 8	10.4	4	-23-	10 12 9 11 10	10.4	3	-40-	13 11 10 7 7	9.6	6
-6-	7 10 7 7 9	8.0	3	-24-	8 8 12 10 8	9.2	4	-41-	10 9 13 10 12	10.8	4
-7-	12 12 14 7 10	11.0	7	-25-	9 13 7 10 13	10.4	6	-42-	8 9 7 10 11	9.0	4
-8-	11 10 9 10 10	10.0	2	-26-	11 10 12 10 10	10.6	2	-43-	7 15 7 8 11	9.6	8
-9-	9 9 11 11 9	9.8	2	-27-	11 9 9 8 8	9.0	3	-44-	11 9 10 11 13	10.8	4
-10-	11 12 10 9 11	10.6	3	-28-	11 12 8 12 10	10.6	4	-45-	10 13 10 9 11	10.6	4
-11-	9 10 11 11 10	10.2	2	-29-	11 14 8 13 8	10.8	6	-46-	12 9 7 13 13	10.8	6
-12-	9 9 10 11 7	9.2	4	-30-	14 12 9 9 10	10.8	5	-47-	9 11 9 10 9	9.6	2
-13-	7 10 12 11 10	10.0	5	-31-	9 11 13 10 7	10.0	6	-48-	8 15 11 10 9	10.6	7
-14-	8 10 11 9 10	9.6	3	-32-	10 11 10 12 11	10.8	2	-49-	7 9 5 12 13	9.2	8
-15-	10 8 11 12 11	10.4	4	-33-	9 10 9 13 14	11.0	5	-50-	10 9 10 11 10	10.0	2
-16-	11 8 9 11 7	9.2	4	-34-	12 10 9 8 8	9.4	4	-51-	12 12 10 11 11	11.2	2
-17-	9 9 11 11 8	9.6	3	-35-	9 7 14 12 9	10.2	7	-52-	11 11 12 10 14	11.6	4
-18-	12 9 11 10 6	9.6	6								

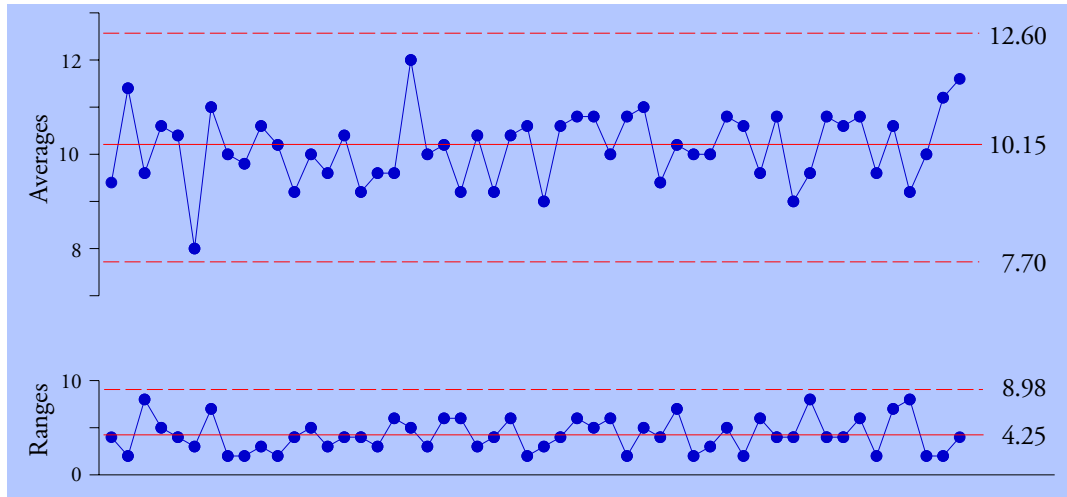


Figure 3: Average and Range Chart for Figure 2

This process has a grand average of 10.15. The specification limits are 6.5 and 13.5. Thus, the distance to nearer specification will be  $DNS = 13.5 - 10.15 = 3.35$ . The average range is 4.25. With subgroups of size 5 this latter value results in a value for  $Sigma(X)$  of  $4.25/2.326 = 1.83$ . Finally, the global standard deviation statistic is  $s = 1.847$ . Thus, the four capability and performance ratios are:

$$\text{Capability Ratio} = C_p = \frac{USL - LSL}{6 \text{Sigma}(X)} = \frac{13.5 - 6.5}{6 (1.83)} = 0.64$$

$$\text{Performance Ratio} = P_p = \frac{USL - LSL}{6 s} = \frac{13.5 - 6.5}{6 (1.85)} = 0.63$$

$$\text{Centered Capability Ratio} = C_{pk} = \frac{2 \text{DNS}}{6 \text{Sigma}(X)} = \frac{2 (3.35)}{6 (1.83)} = 0.61$$

$$\text{Centered Performance Ratio} = P_{pk} = \frac{2 \text{DNS}}{6 s} = \frac{2 (3.35)}{6 (1.85)} = 0.60$$

Here all four indexes tell the same story. They all might be taken to be estimates of the same quantity. Even without the average and range chart of Figure 3 we could tell that this process was being operated predictably and is fairly well-centered within the specifications. The fact that these indexes are all near 60% implies that this process is not capable of meeting the specifications even though it is being operated up to its full potential.

#### EXAMPLE TWO

Raw materials for a compound are dry-mixed in a pharmaceutical blender. The recipe calls for batches that are supposed to weigh 1000 kilograms. If the weight of a batch is off, then presumably the recipe is also off. As each batch is dumped out of the blender the weight is recorded. Figure 4 shows the weights of all 259 batches produced during one week. The values are in time order by rows. The  $XmR$  chart for these values is shown in Figure 5. The limits shown were based on the first 45 values. While there are points outside the limits within this baseline period, the process deteriorates as the week progresses.

Figure 4: Batch Weights in Kilograms for 259 Batches of a Compound

905	930	865	895	905	885	890	930	915	910	920	915	925	860	905
925	925	905	915	930	890	940	860	875	985	970	940	975	1000	1035
1020	985	960	945	965	940	900	920	980	950	955	970	970	1035	1040
1000	1000	990	1000	950	940	965	920	920	925	900	905	900	925	885
1005	1005	950	920	875	865	880	960	925	925	875	900	905	990	970
910	980	900	970	900	895	885	925	870	875	910	915	900	950	880
910	965	910	880	900	920	940	985	965	925	925	975	905	890	950
975	935	940	900	915	980	880	905	915	960	900	915	920	865	980
935	840	900	965	890	875	1020	780	900	900	800	960	845	820	910
885	940	930	925	850	965	1010	1030	980	1010	950	940	1005	880	930
845	935	905	965	975	985	975	950	905	965	905	950	905	995	900
840	1050	935	940	920	985	970	915	935	950	1030	875	880	955	910
1050	890	1005	915	1070	970	1040	770	940	950	1040	1035	1110	845	900
905	910	860	1045	820	900	860	875	1005	880	750	900	835	930	860
960	950	1020	975	950	960	950	880	1000	1005	990	1020	980	1020	920
960	1000	1000	860	1130	830	965	930	950	945	900	990	865	945	970
915	975	940	870	890	915	935	1060	1015	1100	810	1010	1140	805	1020
1110	975	970	1090											

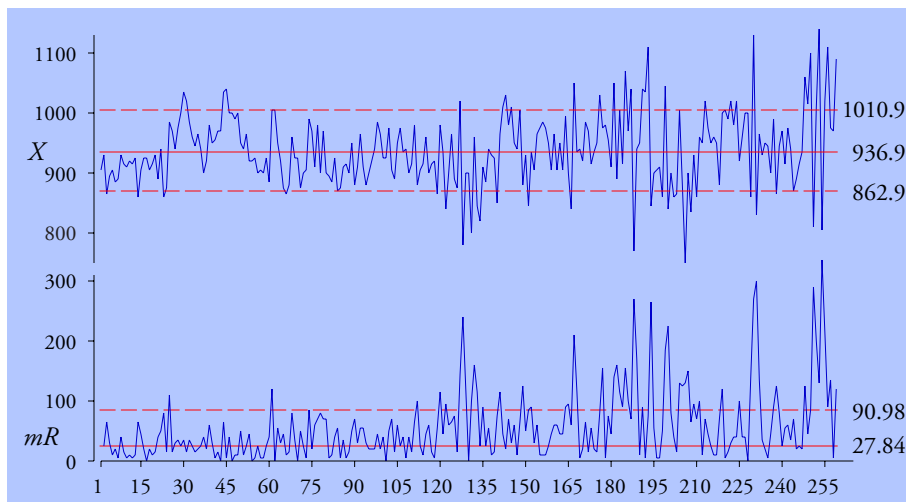


Figure 5: *XmR* Chart for the Batch Weight Data

The specifications for the batch weights are 900 kg. to 1100 kg. With an average moving range of 27.84 the value for  $\text{Sigma}(X)$  is  $27.84/1.128 = 24.7$  kg. The global standard deviation statistic for all 259 values is  $s = 61.3$  kg. With an average of 936.9, the *DNS* value is 36.9 kg. Thus, the four indexes are:

$$\text{Capability Ratio} = C_p = \frac{USL - LSL}{6 \text{Sigma}(X)} = \frac{1100 - 900}{6 (24.7)} = 1.35$$

$$\text{Performance Ratio} = P_p = \frac{USL - LSL}{6s} = \frac{1100 - 900}{6(61.3)} = 0.54$$

$$\text{Centered Capability Ratio} = C_{pk} = \frac{2 DNS}{6 \text{Sigma}(X)} = \frac{2(36.9)}{6(24.7)} = 0.50$$

$$\text{Centered Performance Ratio} = P_{pk} = \frac{2 DNS}{6s} = \frac{2(36.9)}{6(61.3)} = 0.20$$

The discrepancy between the capability ratio and the performance ratio shows that this process is being operated unpredictably. The discrepancy between the centered performance ratio and the performance ratio shows that the average is not centered within the specifications. The capability ratio describes what the current process is capable of doing when operated predictably and on target. The centered performance ratio describes the train wreck of what they actually accomplished during this week, and the gap between these two indexes describes the opportunity that exists for this process.

#### LONG-TERM CAPABILITY

As shown in these examples each of the four index numbers makes a specific comparison between the specified tolerance or the effective space available and either the within-subgroup variation or the global standard deviation statistic. In an effort to distinguish between the capability indexes and the performance indexes the performance indexes have sometimes been called "long-term capability indexes." This nomenclature is misleading and inappropriate.

The idea behind the terminology of "long-term capability" is that if you just collect enough data over a long enough period of time you will end up with a good estimate of the process capability. To illustrate how this is supposed to work we will use the data from example one to perform a sequence of computations using successively more and more data at each step. While we would not normally perform the computations in this way in practice, we do so here to see how increasing amounts of data affect the computation of performance and capability ratios.

We begin with the first eight subgroups. The global standard deviation statistic for these 40 values is 1.974. The specifications are 6.5 to 13.5, so our  $USL - LSL = 7.0$ . Using these values we get a performance ratio of 0.591. The average range for these eight subgroups is 4.375, so  $\text{Sigma}(X)$  is 1.881, and with this value we get a capability ratio of 0.620. It is instructive to note how close these values are to the values found using all the data in Example One above.

The first 12 subgroups contain 60 values. The global standard deviation statistic for these 60 values is 1.742. Using this value we get a performance ratio of 0.670. The average range for these 12 subgroups is 3.833, so  $\text{Sigma}(X)$  is 1.648, and with this value we get a capability ratio of 0.708.

The first 16 subgroups contain 80 values. The global standard deviation statistic for these 80 values is 1.678. Using this value we get a performance ratio of 0.691. The average range for these 16 subgroups is 3.875, so  $\text{Sigma}(X)$  is 1.666, and with this value we get a capability ratio of 0.700.

Continuing in this manner, adding 20 more values at each step, we get the performance ratios and capability ratios shown in Figure 6. There we see that as we use greater amounts of data in the calculations these ratios settle down and get closer and closer to a value near 0.640.

Data Used	Global $s$	$P_p$	$\text{Sigma}(X)$	$C_p$
1 to 40	1.974	0.591	1.881	0.620
1 to 60	1.742	0.670	1.648	0.708
1 to 80	1.687	0.691	1.666	0.700
1 to 100	1.732	0.674	1.698	0.687
1 to 120	1.763	0.662	1.756	0.665
1 to 140	1.753	0.666	1.735	0.672
1 to 160	1.787	0.653	1.761	0.662
1 to 180	1.801	0.648	1.781	0.655
1 to 200	1.792	0.651	1.775	0.657
1 to 220	1.823	0.640	1.810	0.645
1 to 240	1.834	0.636	1.829	0.638
1 to 260	1.847	0.632	1.821	0.641



Figure 6: Performance and Capability Ratios for Example One  
Converge with Increasing Amounts of Data

Of course, as may be seen above, when a process is operated predictably, the capability ratio and the performance ratio both estimate the same quantity. Thus, when a process is operated up to its full potential there is no distinction to be made between the “short-term” capability and the “long-term” capability. Both computations describe the actual capability of the predictable process.

The convergence of a statistic to some asymptotic value that occurs with increasing amounts of data that is seen in Figure 6 is the idea behind many things we do in statistics. Unfortunately, this convergence only happens when the data are homogeneous. In order to see what happens with a process that is not operated up to its full potential we shall repeat the exercise above using the data from Example Two.

The first 40 batch weights have a global standard deviation statistic of 41.60. The specifications are 900 to 1100, so our specified tolerance is  $USL - LSL = 200$ . Using these values we get a performance ratio of 0.801. The average moving range for these 40 values is 29.10, so  $\text{Sigma}(X)$  is 25.80, and with this value we get a capability ratio of 1.292.

The first 60 batch weights have a global standard deviation statistic of 44.20. Using this value we get a performance ratio of 0.754. The average moving range for these 40 values is 25.76 so  $\text{Sigma}(X)$  is 22.84, and with this value we get a capability ratio of 1.459.

Continuing in this manner, adding 20 more values at each step, we get the performance ratios and capability ratios shown in Figure 7. For the sake of comparison, both Figure 6 and Figure 7 use the same horizontal and vertical scales.



Data Used	Global $s$	$P_p$	$\Sigma(X)$	$C_p$
1 to 40	41.60	0.801	25.80	1.292
1 to 60	44.20	0.754	22.84	1.459
1 to 80	44.24	0.753	27.44	1.215
1 to 100	42.82	0.778	27.49	1.213
1 to 120	41.45	0.804	28.94	1.152
1 to 140	46.09	0.723	34.47	0.967
1 to 160	47.01	0.709	35.68	0.934
1 to 180	47.44	0.703	38.01	0.877
1 to 200	53.87	0.619	44.17	0.755
1 to 220	55.26	0.603	45.66	0.730
1 to 240	56.61	0.589	47.68	0.699
1 to 259	61.26	0.544	51.25	0.650

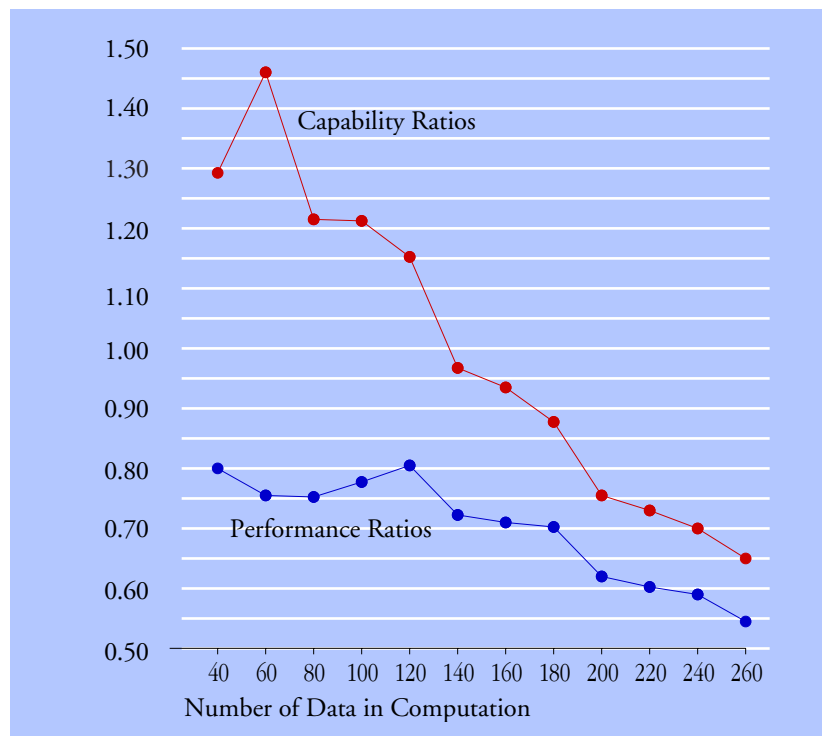


Figure 7: Neither the Performance Ratio Nor the Capability Ratio for Example Two Settles Down to Some Fixed Value with Increasing Amounts of Data

To what value is the performance ratio curve in Figure 7 converging? After 120 values it appears to be approaching 0.80, then with 20 additional values it suddenly drops down to the neighborhood of 0.70. After 180 values it seems to be approaching 0.70, then with 20 more values it drops down to the neighborhood of 0.60. After 240 values we are still in the vicinity of 0.60, but then with 259 values we drop down to 0.54. So which value are you going to use as your “long-term” capability? 0.80? 0.70? 0.60? or 0.54?

Here we see that even though we use ever greater amounts of data, the ratios do not settle down to any particular value. Neither do we see the agreement between the performance ratio and the capability ratio that was evident in Figure 6. Clearly these two ratios characterize different aspects of the data in this case. Both the migration and the estimation of different things

happen because this process is changing over time. Because of these changes there is no magic amount of data that will result in a “good number.” The computations are chasing a moving target. The question “What is the long-term capability of this process?” is meaningless simply because there is no such quantity to be estimated regardless of how many data we might use.

With an unpredictable process, as we use greater amounts of data in our computation we eventually combine values that were obtained while the process was acting differently. This combination of unlike values does not prevent us from computing our summary statistics, but it does complicate the interpretation of those statistics. With an unpredictable process there is no single value for the process average, or the process variation, or the process capability. All such notions of process characteristics become chimeras, and any attempt to use our statistics to estimate these nonexistent process characteristics is an exercise in frustration. This is why the idea of “long-term” capability is just so much nonsense.

However, once we understand that we are working with an unpredictable process, we are free to use our statistics to characterize different aspects of the *data* (as opposed to the process). As noted earlier, the capability ratio of 1.35 computed from the first 45 values of Example Two provides an approximation of what this process has the potential to do. In the same manner, the centered performance ratio of 0.20 describes what was done during this week. And the difference between these two statistics characterizes the gap between performance and potential. Thus, we may use the capability and performance indexes to identify opportunities even when they do not estimate fixed aspects of the underlying process.

Thus, referring to the performance indexes as “long-term” capabilities confuses the issue and misleads everyone. They are descriptive statistics that summarize the past. They do not estimate any fixed quantity unless the process is being operated predictably. And they definitely do not describe the indescribable “long-term capability of an unpredictable process.”