

Capability Ratios Vary

The key to using any number is to understand its limitations

Donald J. Wheeler

Capability ratios are widely used and sometimes misunderstood. The computer will gladly offer up values of each of the commonly used capability and performance indexes. Yet there is little appreciation of the inherent uncertainty contained in each of these numbers. Here we shall look at how to quantify these uncertainties and how to interpret the ratios.

All capability ratios are statistics. This means that, like all other statistics, capability ratios will vary from computation to computation even when the underlying process does not change. The following examples will illustrate the intrinsic and extrinsic components of this variation.

A PREDICTABLE PROCESS

Figure 1 shows 100 subgroup averages and ranges for subgroups of size 5 based on 500 values obtained from a bead board (or quincunx) while no changes were made in the configuration of the bead board. When operated in this manner the bead board serves as a prototype for a predictable process. The grand average is 10.084 and the average range is 4.11.

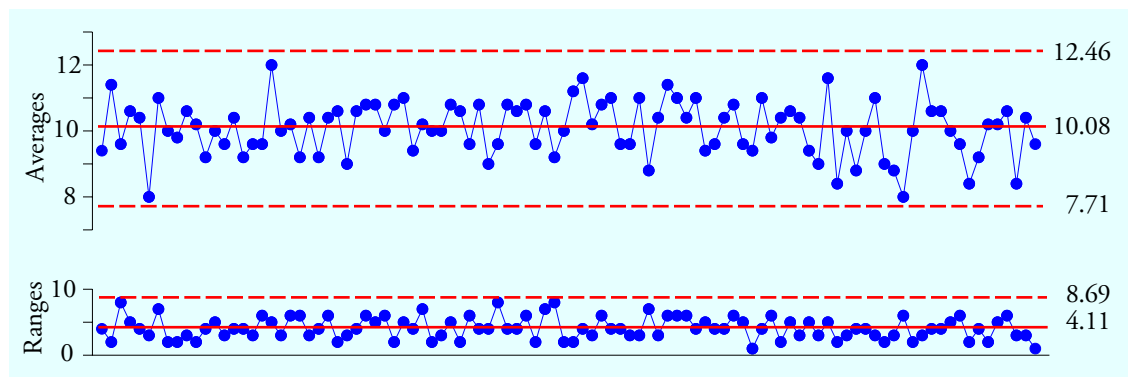


Figure 1: Average and Range Chart for Figure 10 Data.

As shown in Figure 2, assume the watershed specifications for the process characterized by Figure 1 are 2.5 to 17.5. The natural process limits are 4.78 to 15.39. So, this process is centered, the distance to the nearer specification is $DNS = [17.5 - 10.084] = 7.42$ units, and the capability indexes are:

$$C_p = \frac{USL - LSL}{6 \text{ Sigma}(X)} = \frac{17.5 - 2.5}{6 (1.767)} = 1.42$$

$$C_{pk} = \frac{2 \text{ DNS}}{6 \text{ Sigma}(X)} = \frac{2 (7.42)}{6 (1.767)} = 1.40$$

The global standard deviation statistic for the 500 values in Figure 2 is $s = 1.814$. Using this value we find performance indexes of:

$$P_p = \frac{USL - LSL}{6s} = \frac{17.5 - 2.5}{6(1.814)} = 1.38$$

$$P_{pk} = \frac{2 DNS}{6s} = \frac{2(7.42)}{6(1.814)} = 1.36$$

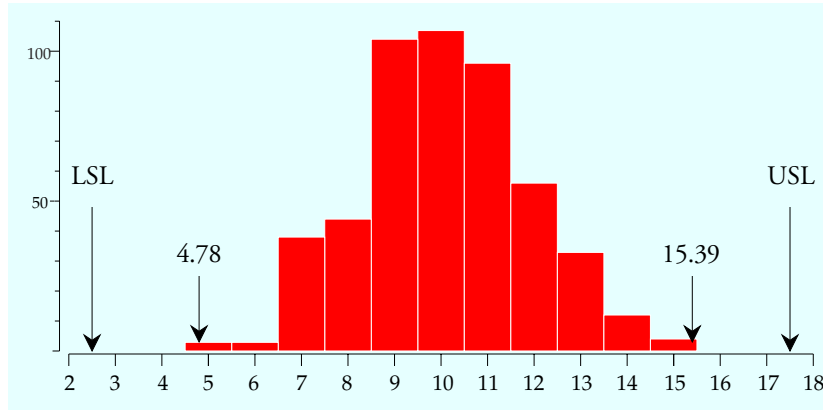


Figure 2: Histogram for 500 Data of Figure 1

Thus, the histogram and the indexes tell the same story. We have a process that is capable and well-centered within the specifications. Because this process is capable and on-target the four index numbers computed above are all estimates of the same underlying property—the capability of this production process. Yet the indexes are not all the same.

INTERVAL ESTIMATES FOR CAPABILITY AND PERFORMANCE INDEXES

Interval estimates can help us to understand the uncertainty for any statistic. We begin with the performance ratio, P_p . When the process is operated predictably it makes sense to use the computed statistics to extrapolate from the data to the underlying process. The uncertainty of this extrapolation can be characterized by a 90% interval estimate (i.e. a 90% confidence interval). For the P_p statistic this 90% interval estimate for the process capability uses the fifth and ninety-fifth percentiles of a chi-square distribution:

$$P_p \sqrt{\frac{\chi_{.05}^2}{d.f.}} < \text{Process Capability} < P_p \sqrt{\frac{\chi_{.95}^2}{d.f.}}$$

$$P_p \mathbf{LB1} < \text{Process Capability} < P_p \mathbf{UB1}$$

where $d.f.$ is the degrees of freedom for the global standard deviation statistic used in the denominator of P_p . Selected values of **LB1** and **UB1** are tabled in Figure 3.

When all 500 data were used we found a performance ratio of $P_p = 1.38$. The degrees of freedom for the global standard deviation statistic used in computing P_p is equal to one less than the number of data used, which in this case is 499 degrees of freedom. Using the critical values for 500 degrees of freedom we find a 90% interval estimate for the process capability of 1.31 to 1.45.

$$P_p \text{ LB1} = 1.38 (0.948) = 1.31 < \text{Process Capability} < P_p \text{ UB1} = 1.38 (1.052) = 1.45$$

<i>d.f.</i>	<i>LB1</i>	<i>UB1</i>	<i>d.f.</i>	<i>LB1</i>	<i>UB1</i>	<i>d.f.</i>	<i>LB1</i>	<i>UB1</i>	<i>d.f.</i>	<i>LB1</i>	<i>UB1</i>
5	.479	1.488	20	.736	1.253	55	.842	1.155	250	.926	1.073
6	.522	1.449	21	.743	1.247	60	.848	1.148	300	.933	1.067
7	.556	1.418	22	.749	1.242	65	.854	1.142	350	.938	1.062
8	.584	1.392	23	.754	1.236	70	.860	1.137	400	.942	1.058
9	.608	1.371	24	.760	1.232	75	.865	1.133	450	.945	1.055
10	.628	1.353	25	.764	1.227	80	.869	1.128	500	.948	1.052
11	.645	1.338	26	.769	1.223	85	.873	1.125	550	.950	1.049
12	.660	1.323	27	.773	1.219	90	.876	1.121	600	.952	1.047
13	.673	1.311	28	.777	1.215	100	.883	1.115	650	.954	1.045
14	.685	1.300	29	.781	1.211	110	.888	1.110	700	.956	1.044
15	.696	1.291	30	.785	1.208	120	.893	1.105	750	.957	1.042
16	.705	1.282	35	.801	1.193	140	.901	1.097	800	.959	1.041
17	.714	1.274	40	.814	1.181	160	.907	1.091	850	.960	1.040
18	.722	1.266	45	.825	1.170	180	.913	1.086	900	.961	1.039
19	.730	1.259	50	.834	1.162	200	.917	1.082	1000	.963	1.037

Figure 3: Factors for 90% Interval Estimates for C_p and P_p

When a process is operated predictably both the performance ratio and the capability ratio are two estimates of the same quantity. For this reason, the critical values in Figure 3 can also be used to obtain an *approximate* 90% interval estimate based on the capability ratio C_p :

$$C_p \text{ LB1} < \text{Process Capability} < C_p \text{ UB1}$$

However, since the capability ratio uses a within-subgroup measure of dispersion, it will have a different number of degrees of freedom than does the performance ratio. Here the capability ratio was based on the average range, and the average range for $k = 100$ subgroups of size five has an effective degrees of freedom of:

$$\text{Effective } d.f. \text{ for Average Range} \approx \frac{k d_2^2}{2 d_3^2} + 0.2 = \frac{100 (2.326)^2}{2 (0.8641)^2} + 0.2 = 362.5$$

where d_2 and d_3 are the bias correction factors for ranges of subgroups of size five. Thus, using the critical values from Figure 3 for 350 degrees of freedom with the calculated C_p value of 1.42 we find an approximate 90% interval estimate for the process capability of 1.33 to 1.51.

$$C_p \text{ LB1} = 1.42 (0.938) = 1.33 < \text{Process Capability} < C_p \text{ UB1} = 1.42 (1.062) = 1.51$$

Thus, even when we use 500 data to compute a capability ratio or a performance ratio, the intrinsic uncertainty in our statistics will still be plus or minus five or six percent.

INTERVAL ESTIMATES FOR CENTERED INDEXES

The centered performance ratio, P_{pk} and the centered capability ratio, C_{pk} have the grand average in the numerator. This introduces additional uncertainty into these ratios and requires a different computation for their interval estimates. The centered performance ratio, P_{pk} is the minimum of two statistics where each statistic can be characterized by a non-central Student's t -distribution. Chou, Owen, and Borrego [JQT v.22, July 1990] show that the exact interval estimate depends strongly upon the degrees of freedom and depends weakly upon the value of the centered performance ratio. By ignoring the weak dependency we can create *approximate*

critical values that are suitable for use in practice. As before, when the process is operated predictably, an approximate 90% interval estimate for the centered process capability will be:

$$P_{pk} \text{ LB2} < \text{Centered Process Capability} < P_{pk} \text{ UB2}$$

The values for **LB2** and **UB2** will depend upon the degrees of freedom for the global standard deviation statistic used in the denominator of P_{pk} . Selected values of **LB2** and **UB2** are tabled in Figure 4.

When a process is operated predictably both the centered performance ratio and the centered capability ratio are two estimates of the same quantity. For this reason, the critical values in Figure 4 can also be used to obtain an approximate 90% interval estimate of the centered process capability based on the centered capability ratio C_{pk} :

$$C_{pk} \text{ LB2} < \text{Centered Process Capability} < C_{pk} \text{ UB2}$$

Here the values for **LB2** and **UB2** will depend upon the degrees of freedom for the average range, or other within-subgroup measure of dispersion, used in the denominator of C_{pk} .

<i>d.f.</i>	LB2	UB2	<i>d.f.</i>	LB2	UB2	<i>d.f.</i>	LB2	UB2	<i>d.f.</i>	LB2	UB2
5	.340	2.096	20	.665	1.386	55	.803	1.211	250	.910	1.092
6	.384	1.971	21	.673	1.376	60	.812	1.200	300	.920	1.082
7	.423	1.866	22	.680	1.368	65	.821	1.188	350	.927	1.075
8	.458	1.773	23	.686	1.358	70	.828	1.181	400	.933	1.069
9	.492	1.693	24	.693	1.351	75	.835	1.174	450	.936	1.066
10	.525	1.618	25	.698	1.342	80	.839	1.168	500	.939	1.063
11	.544	1.587	26	.704	1.335	85	.844	1.163	550	.941	1.060
12	.561	1.556	27	.710	1.328	90	.848	1.158	600	.942	1.058
13	.577	1.530	28	.715	1.320	100	.857	1.149	650	.944	1.056
14	.592	1.505	29	.720	1.313	110	.862	1.143	700	.946	1.055
15	.606	1.482	30	.725	1.309	120	.868	1.137	750	.947	1.053
16	.619	1.460	35	.746	1.281	140	.878	1.126	800	.949	1.052
17	.632	1.440	40	.764	1.259	160	.885	1.118	850	.950	1.051
18	.644	1.419	45	.780	1.238	180	.893	1.110	900	.951	1.049
19	.656	1.402	50	.793	1.222	200	.899	1.104	1000	.953	1.047

Figure 4: Factors for Approximate 90% Interval Estimates for C_{pk} and P_{pk}

For Figure 1 we found a centered performance ratio of $P_{pk} = 1.36$. With 499 degrees of freedom we use the critical values for 500 degrees of freedom of 0.939 and 1.063 to compute an approximate 90% interval estimate of the centered process capability of 1.28 to 1.45:

$$P_{pk} \text{ LB2} = 1.36 (0.939) = 1.28 < \text{Centered Process Capability} < P_{pk} \text{ UB2} = 1.36 (1.063) = 1.45$$

We also found $C_{pk} = 1.40$. With 363 degrees of freedom we might use the critical values for 350 degrees of freedom of 0.927 and 1.075 to obtain an approximate 90% interval estimate for the centered process capability of 1.30 to 1.51:

$$C_{pk} \text{ LB2} = 1.40 (0.927) = 1.30 < \text{Centered Process Capability} < C_{pk} \text{ UB2} = 1.40 (1.075) = 1.51$$

So, even when using 500 data, our two estimates of the effective space available for this process will have uncertainties of plus or minus six to seven percent. However, it is not common to use 500 values when computing these indexes. More often we use 30 to 50 data in these computations. So let us divide the 500 data of Figure 1 into ten blocks, each consisting of ten subgroups of size five, and compute the capability and performance indexes for each of these

blocks of 50 values. The resulting values are shown in Figure 5.

Subgroups Used	C_p	C_{pk}	P_p	P_{pk}
Subgroups 1 to 10	1.45	1.44	1.37	1.35
Subgroups 11 to 20	1.49	1.49	1.52	1.51
Subgroups 21 to 30	1.29	1.27	1.29	1.27
Subgroups 31 to 40	1.38	1.34	1.41	1.36
Subgroups 41 to 50	1.19	1.17	1.19	1.17
Subgroups 51 to 60	1.49	1.41	1.37	1.30
Subgroups 61 to 70	1.24	1.19	1.32	1.26
Subgroups 71 to 80	1.53	1.52	1.46	1.45
Subgroups 81 to 90	1.66	1.63	1.40	1.38
Subgroups 91 to 100	1.57	1.50	1.53	1.46

Figure 5: Ten Sets of Capability and Performance Indexes for the Figure 1 Data

This process is being operated predictably and it is well centered within the specifications. Under these conditions the capability and performance indexes in Figure 5 are forty estimates of the same quantity. Yet, like all statistics, these forty estimates differ. While this process is likely to meet the customer requirement of having a capability that exceeds 1.33, thirteen of the forty indexes in Figure 5 fall below 1.33. If we used one of these thirteen indexes to characterize this process it is likely that inappropriate and unnecessary process changes would be made. When we do not allow for the inherent variation in our capability indexes we can end up creating trouble.

Each of the capability and centered capability ratios of Figure 5 has 36.5 degrees of freedom. Interpolating in Figure 3 we find $LB1 = 0.805$ and $UB1 = 1.189$. Interpolating in Figure 4 we find $LB2 = 0.751$ and $UB2 = 1.274$. Using these critical values with each of the capability and centered capability ratios in Figure 5 we get the approximate 90% interval estimates shown in Figure 6.

Each of the performance and centered performance ratios of Figure 5 has 49 degrees of freedom. Interpolating in Figure 3 we find $LB1 = 0.832$ and $UB1 = 1.164$. Interpolating in Figure 4 we find $LB2 = 0.790$ and $UB2 = 1.225$. Using these critical values with the performance and centered performance ratios in Figure 5 we get the 90% interval estimates shown in Figure 6.

Subgroups Used	90% Intervals for C_p	90% Intervals for C_{pk}	90% Intervals for P_p	90% Intervals for P_{pk}
Subgroups 1 to 10	1.17 to 1.72	1.08 to 1.83	1.14 to 1.59	1.07 to 1.66
Subgroups 11 to 20	1.20 to 1.77	1.12 to 1.89	1.26 to 1.77	1.20 to 1.85
Subgroups 21 to 30	1.04 to 1.53	0.95 to 1.62	1.07 to 1.50	1.00 to 1.55
Subgroups 31 to 40	1.11 to 1.64	1.01 to 1.71	1.17 to 1.64	1.07 to 1.67
Subgroups 41 to 50	0.96 to 1.41	0.88 to 1.49	0.99 to 1.39	0.93 to 1.44
Subgroups 51 to 60	1.20 to 1.77	1.06 to 1.79	1.14 to 1.59	1.02 to 1.59
Subgroups 61 to 70	1.00 to 1.47	0.89 to 1.51	1.10 to 1.54	1.00 to 1.55
Subgroups 71 to 80	1.23 to 1.82	1.14 to 1.93	1.21 to 1.70	1.15 to 1.78
Subgroups 81 to 90	1.34 to 1.97	1.23 to 2.08	1.16 to 1.63	1.09 to 1.69
Subgroups 91 to 100	1.26 to 1.87	1.13 to 1.91	1.27 to 1.78	1.15 to 1.79

Figure 6: 90% Interval Estimates for Figure 5 Capability and Performance Indexes

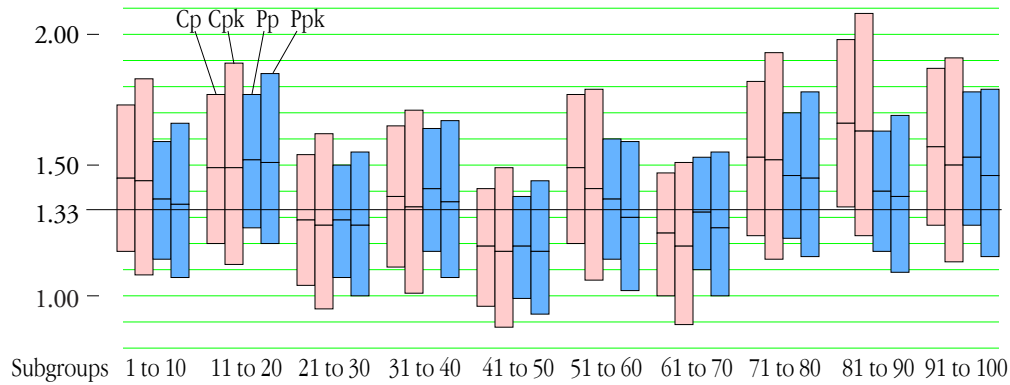


Figure 7: 90% Interval Estimates for Figure 5 Capability and Performance Indexes

Figure 7 shows the interval estimates for the capability and centered capability ratios in red and those for the performance and centered performance ratios in blue. Here these intervals overlap more than they differ, and they all tell the same story. While thirteen of the 40 indexes fell below 1.33, Figures 6 and 7 reveal that all of the indexes are consistent with having a process with a capability of 1.33 or more. (If we compare the intervals above with the indexes computed using all 500 data, thirty-nine of these forty 90% interval estimates contain the corresponding estimate of 1.42, 1.40, 1.38 or 1.36 found earlier.) Thus, these interval estimates show the inherent uncertainty attached to each statistic and also serve to characterize the underlying capability of this predictable process.

AN UNPREDICTABLE PROCESS

Figure 8 shows 100 averages and ranges for subgroups of size 5 based on 500 values obtained from a bead board while the configuration of the bead board was changed periodically. These data serve as a prototype for an unpredictable process. The grand average is 12.86, the average range is 3.95, and the global standard deviation statistic is $s = 3.326$. The specifications are still 2.5 to 17.5, while the natural process limits are 7.76 to 17.95.

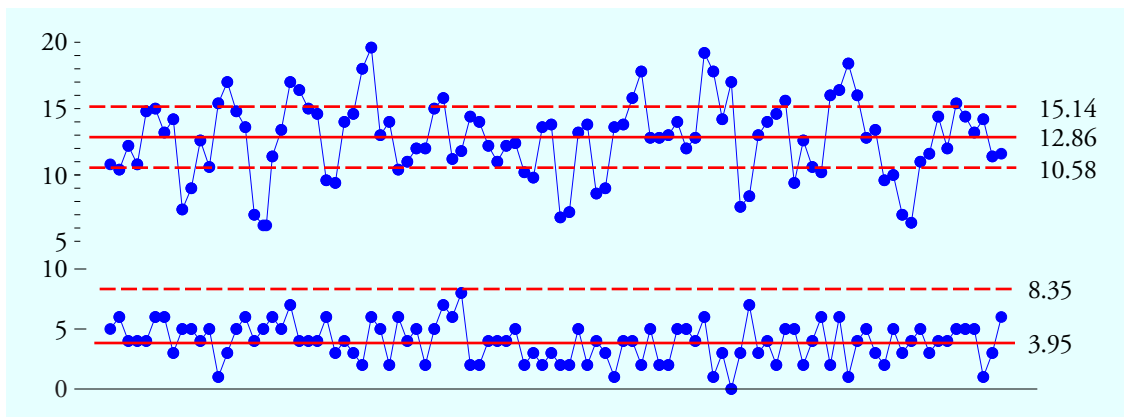


Figure 8: Average and Range Chart for an Unpredictable Process.

Not only is this process being operated unpredictably, but it also is being operated off-target. Thirty-seven of the 500 values are out of specification, which is 7.4% nonconforming. The histogram for these 500 data is shown in Figure 9.

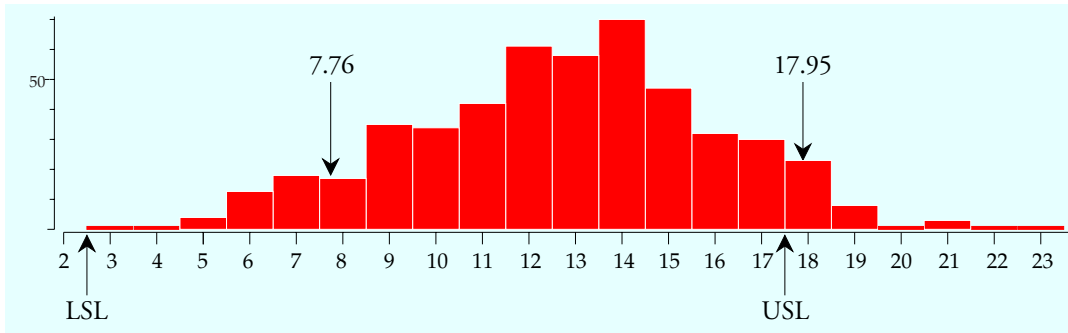


Figure 9: Histogram of the 500 Data Used for Figure 8

If we divide the 500 data used for Figure 8 into ten blocks, each consisting of ten subgroups of size five, we get the capability and performance indexes shown in Figure 10. Unlike the earlier values in Figure 5, the values in each column of Figure 10 are quantitatively different from those in the other columns.

Subgroups Used	C_p	C_{pk}	P_p	P_{pk}
Subgroups 1 to 10	1.21	0.92	0.82	0.63
Subgroups 11 to 20	1.32	0.93	0.66	0.47
Subgroups 21 to 30	1.35	0.48	0.70	0.25
Subgroups 31 to 40	1.16	0.76	0.95	0.62
Subgroups 41 to 50	1.88	1.29	1.26	0.86
Subgroups 51 to 60	2.01	1.48	0.66	0.49
Subgroups 61 to 70	1.76	0.69	0.90	0.35
Subgroups 71 to 80	1.42	1.12	0.81	0.64
Subgroups 81 to 90	1.66	1.09	0.59	0.38
Subgroups 91 to 100	1.42	0.87	1.15	0.70

Figure 10: Ten Sets of Capability and Performance Indexes for the Figure 9 Data

Since the degrees of freedom for the ratios in Figure 10 are the same as for those in Figure 5 we can use the critical values found earlier to compute 90% interval estimates for the ratios of Figure 10. When we do this we end up with the intervals shown in Figure 11. The red bars show the intervals for the capability and centered capability ratios, while the blue bars show the intervals for the performance and centered performance ratios.

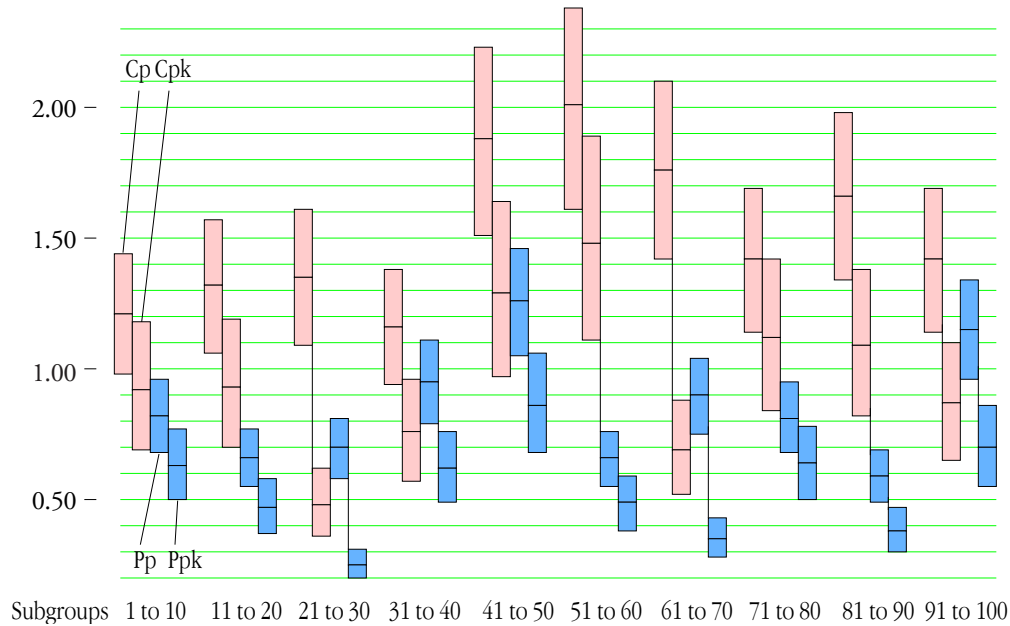


Figure 11: 90% Intervals for Figure 10 Capability and Performance Indexes

The intervals in Figure 11 do not overlap with each other in the same way or to the same extent that the intervals of Figure 7 did. Here there are also substantial differences in the intervals as we go from one block of fifty data to another.

For each block of data the first interval is for the capability ratio. These are always the highest values for that block because this statistic characterizes what the process has the ability to do if and when it is operated predictably and on-target. It characterizes the *hypothetical* capability.

For each block the second interval is for the centered capability ratio. These intervals are lower than the preceding interval because this statistic characterizes what the process is likely to do if it is operated predictably at the historical average. It characterizes another hypothetical capability.

The third interval for each block is for the performance ratio. These intervals characterize what the capability of the past production would have been if the process was centered within the specifications. This is yet another hypothetical measure of capability.

For each block of data the last interval is for the centered performance ratio. These are always the lowest values for each block because this statistic summarizes the past performance of the process *while that block of data was collected*. It describes the process as it was, where it was, with all of its warts and pimples included. While these statistics describe the past performance, and while these intervals describe the inherent uncertainty in the statistics, they cannot be relied upon to characterize the future when the underlying process is unpredictable.

So, when the process is unpredictable, three of the indexes are hypothetical and the other one describes the past. However, neither the statistics nor their intervals can deal with the extrinsic uncertainty introduced by the unpredictable nature of the underlying process. If we use all 500 data from Figure 8 we find a performance ratio of 0.75. Yet only six of the ten intervals for the performance ratios in Figure 11 include the value of 0.75. Is that what you would expect from a

set of 90% interval estimates?

If we use all 500 data from Figure 8 we find a centered performance ratio of 0.47, yet only three of the ten intervals for the centered performance ratios in Figure 11 include the value of 0.47. Once again, is this what you would expect from a set of 90% interval estimates?

Neither the indexes nor their “confidence intervals” can handle the extrinsic variation of an unpredictable process.

So when do these capability and performance indexes change from being hypothetical or descriptive measures and start being estimates of actual process capability? It is the predictability of the underlying process that turns the *statistics* into *estimates* of actual capability. As long as the process is operated unpredictably it does not possess a well-defined capability.

As Walter Shewhart put it, “measurements of phenomena in both social and natural science for the most part obey neither deterministic nor statistical laws until assignable causes of variability have been found and removed.” When there is no probability model there are no parameters to estimate. We may calculate as many statistics as we wish, but none of these will estimate anything real.

It is the *predictability* of Figure 1 that turns the intervals in Figures 6 and 7 into *estimates* of the process capability. The *unpredictability* of Figure 8 means that there is no well-defined process capability for the intervals of Figure 11 to estimate. So, while the statistics in Figure 10 describe hypothetical possibilities or summarize the past, and while the intervals of Figure 11 describe the intrinsic uncertainty in these statistics, the question of whether these indexes are hypothetical or actual, and whether they can be used to predict the future will depend upon whether or not the process is being operated predictably.

SUMMARY

Process behavior charts and histograms with specification limits allow you to compare the voice of the process with the voice of the customer. Numerical summaries complement these graphic comparisons but they cannot replace them. The main problem with numerical summaries is not their computation but rather their interpretation, since their meaning changes with the status of the process.

When the process is operated predictably the capability ratio C_p and the performance ratio P_p both characterize the same thing—the space available for the process. They both express the specified tolerance as a percentage of the space consumed by the natural process variation. Interval estimates for these ratios make sense.

When the process is operated predictably the centered capability ratio C_{pk} and the centered performance ratio P_{pk} both characterize the same thing—the effective space available for the process. They both compare twice the distance to the nearer specification (*DNS*) with the generic space required by natural process variation. If the process is well-centered within the specifications these ratios will converge to the two ratios above. Interval estimates for these ratios make sense.

When the process is operated unpredictably the capability ratio C_p and the performance ratio P_p describe different characteristics of the data.

When the process is operated unpredictably the capability ratio C_p provides a data-based

estimate of the *hypothetical* capability that the process is likely to achieve if and when the assignable causes of exceptional variation are found and made part of the set of control factors for the process. Interval estimates make hypothetical sense as a way to characterize the intrinsic uncertainty of this hypothetical value.

When the process is operated unpredictably the performance ratio P_p will describe the past performance of the process. It expresses the specified tolerance as a percentage of the generic space used by the past process outcomes. With an unpredictable process P_p will always be smaller than C_p because the global dispersion will be greater than the within-subgroup variation. Intervals are not estimates of any well defined property of the process, but do quantify the intrinsic uncertainty in the descriptive statistic.

When the process is operated unpredictably the centered capability ratio C_{pk} and the centered performance ratio P_{pk} describe different characteristics of the data.

When the process is operated unpredictably the centered capability ratio C_{pk} describes the hypothetical capability that the process is likely to achieve if the process is operated predictably at the historical average. Since we gain leverage for changing the process aim as we identify assignable causes and make them part of the set of control factors, this hypothetical value of little interest in most cases. Since the centered capability ratio is essentially a lower bound for the hypothetical capability, intervals are rarely of interest here.

When the process is operated unpredictably the centered performance ratio P_{pk} describes the past performance relative to the specifications. Intervals can describe the intrinsic uncertainty in the descriptive statistic, but they do not estimate any property of the process.

Because of the shifting meanings for these commonly computed indexes it is impossible to ever have a meaningful discussion using the capability or performance indexes alone. *You will always need to know whether or not the underlying process is being operated predictably before you know how to interpret and use a capability or performance index.*