Understanding Process Capability Indices

Stefan Steiner, Bovas Abraham and Jock MacKay
Institute for Improvement of Quality and Productivity
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario N2L 3G1

1. Introduction

A process capability index is a numerical summary that compares the behavior of a product or process characteristic to engineering specifications. These measures are also often called capability or performance indices or ratios; we use capability index as the generic term. A capability index relates the voice of the customer (specification limits) to the voice of the process. A large value of the index indicates that the current process is capable of producing parts that, in all likelihood, will meet or exceed the customer's requirements. A capability index is convenient because it reduces complex information about the process to a single number. Capability indices have several applications, though the use of the indices is driven mostly by monitoring requirements specified by customers. Many customers ask their suppliers to record capability indices for all special product characteristics on a regular basis. The indices are used to communicate how well the process has performed. For stable or predictable processes, it is assumed that these indices also indicate expected future performance. Suppliers may also use capability indices for different characteristics to establish priorities for improvement activities. Similarly, the effect of a process change can be assessed by comparing capability indices calculated before and after the change.

Despite the widespread use of capability indices in industry, and some good review articles, such as Gunter (1989abcd), there is much confusion and misunderstanding regarding their interpretation and appropriate use. This problem is magnified because many quality programs, such as, for example, the automotive industry standard QS-9000 mandate the use of capability indices, but state the capability requirements in an overly complex and confusing way. The following three excerpts from the QS-9000 reference manuals illustrate the problem. The process capability requirements given in Section 4.9.3. of the manual "Quality Systems Requirements" (Automotive Industry Action Group (AIAG), 1995) state, in part:

> "Ongoing process performance requirements are defined by the customer. If no such requirements have been established, the following default values apply:
> - For stable processes and normally distributed data, a C_{pk} value ≥ 1.33 should be achieved.
> - For chronically unstable processes with output meeting specification and a predictable pattern, a P_{pk} value ≥ 1.67 should be achieved."

In the Advanced Product Quality Planning (APQP) manual (AIAG, 1995), the Ford Powertrain specific requirements for dynamic control plans (DCP, Appendix G) say, in part:

> "All processes must produce all characteristics to specification on a production basis. ... Significant Characteristics (SCs) must be in a state of statistical control with $P_{pk} \geq 1.67$ and $C_{pk} \geq 1.33$."

Finally, within the Production Part Approval Process (PPAP) manual (AIAG, 1995), the requirements that relate to process capability are given as follows:

"Calculate the P_{pk} index and take the following actions:

For Processes that Appear Stable

Results	Interpretation
P_p and $P_{pk} > 1.67$	The process probably meets customer requirements. After approval, begin production and follow the Approved Control plan.
$1.33 \leq P_{pk} \leq 1.67$	The process may not meet customer requirements. After part approval, begin production with additional attention to the characteristic until an ongoing $C_{pk} \geq 1.33$ is achieved.
$P_{pk} < 1.33$	The process is substandard for meeting customer requirements. Process improvements must be given high priority and documented in a corrective action plan. Increased inspection or testing is normally required until an ongoing C_{pk} of 1.33 is demonstrated. A revised control plan for these interim actions must be reviewed with and approved by the customer.

[Processes that appear unstable at the time PPAP approval is sought require special attention] ... until ongoing stability and an C_{pk} of 1.33 is demonstrated."

These excerpts from QS-9000 illustrate that often important decisions about part acceptance and the meeting of ongoing customer requirements are based on the value of a capability index. However, the excerpts also suggest many questions. For example,

Why does the standard refer to different indices?

How should the data used in the calculations be collected?

What is a chronically unstable, yet predictable process?

What is the importance of process stability?

Are capability indices comparable across processes?

The goal of this article is answer such questions by providing an illustration of the important issues related to capability indices. In addition, this article makes suggestions regarding the process information necessary to make appropriate use of capability indices. In the next section, the question of which capability index to use is addressed by contrasting the various common indices. It is shown that the index P_{pk} is always preferable. The third section discusses important issues such as the role of data collection and the importance of stability associated with the calculation and interpretation of a capability index. Finally, Section Four proposes clear guidelines for the appropriate use of process capability indices.

2. Definition of the Capability Indices

A capability index relates the engineering specification (determined by the customer) to the observed behaviour of the process. The capability of a process is defined as the ratio of the distance from the process center to the nearest specification limit divided by a measure of the process variability. The idea is illustrated graphically in Figure 1 that shows a histogram of the process output along with the specification limits.

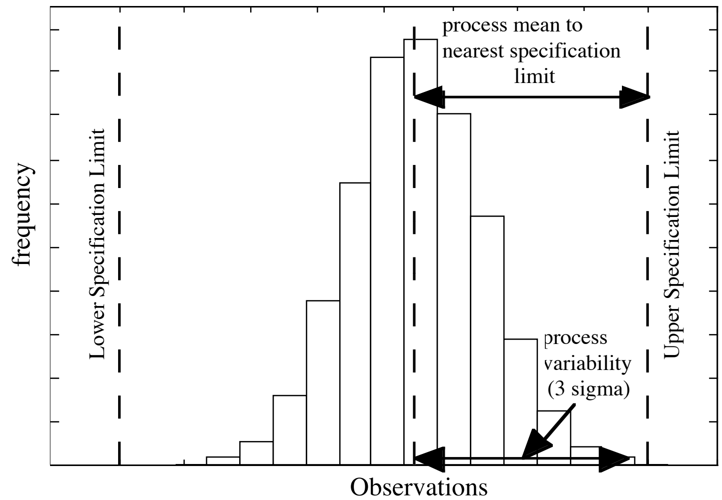

Figure 1: Graphical Illustration of Process Capability

In more mathematical terms,

$$\text{Process Capability} = \min\left(\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\right),$$

where USL and LSL are the upper and lower specification limits respectively, and μ and σ are the process mean and standard deviation respectively for individual measurements of the characteristic of interest. Calculating the process capability requires knowledge of the process mean and standard deviation, μ and σ. These values are usually estimated from data collected from the process.

Often the process data is collected in subgroups. Let X_{ij}, $i = 1,\ldots,m$ and $j = 1,\ldots,n$ represent the process data collected from the j^{th} unit in the i^{th} subgroup. Here, m equals the total number of subgroups, and n equals the subgroup sample size. The two most widely used capability indices are defined as:

$$P_{pk} = \min\left(\frac{USL-\bar{\bar{X}}}{3\hat{\sigma}_s}, \frac{\bar{\bar{X}}-LSL}{3\hat{\sigma}_s}\right) \quad (1)$$

$$C_{pk} = \min\left(\frac{USL-\bar{\bar{X}}}{3\hat{\sigma}_{\bar{R}/d_2}}, \frac{\bar{\bar{X}}-LSL}{3\hat{\sigma}_{\bar{R}/d_2}}\right), \quad (2)$$

where $\bar{\bar{X}}$, the overall average, is used to estimate the process mean μ, and $\hat{\sigma}_s$ and $\hat{\sigma}_{\bar{R}/d_2}$ are different estimates of the process standard deviation σ.

The estimate $\hat{\sigma}_s$ is the sample standard deviation $\sqrt{\sum_{j=1}^{n}\sum_{i=1}^{m}\left(X_{ij}-\overline{\overline{X}}\right)^2 \Big/ (nm-1)}$, whereas $\hat{\sigma}_{\overline{R}/d_2} = \overline{R}/d_2$ is an estimate derived using the subgroup ranges R_i, $i = 1,\ldots,m$. The parameter d_2 is an adjustment factor needed to estimate the process standard deviation from the average sample range. Since d_2 is also used in the derivation of control limits for $\overline{X}$ and R control charts it is tabulated in standard references on statistical process control, such as the QS-9000 SPC manual (AIAG, 1995) or Montgomery (1991). Large values of C_{pk} and P_{pk} should correspond to a capable process that produces the vast majority of units within the specification limits.

The index P_p, and the related index C_p, are similar to C_{pk} and P_{pk}. However, P_p and C_p ignore the current estimate of the process mean and relate the specification range directly to the process variation. In effect, C_p and P_p can be considered measures that suggest how capable the process could be if the process mean were centered midway between the specification limits. The indices P_p and C_p are not recommended for reporting purposes, since the information they provide to supplement C_{pk} and P_{pk} is also easily obtained from a histogram of the data. Histograms are preferable since they also provide other useful process information. As a result, only the indices C_{pk} and P_{pk} are considered in more detail in this article. For more information on other process capability measures see Kotz (1993).

To illustrate the calculation of the estimated capability indices C_{pk} and P_{pk} we present a simple example. In this example, called the Pilot OD example, the diameter of the pilot on an output shaft is a special characteristic. The upper and lower specification limits for the diameter are $USL = 25$ and $LSL = -25$ respectively, when the measured quantity is the number of microns from nominal. A previous study verified that the measurement system utilized introduces very little measurement error. As part of a PPAP demonstration study, 300 units were produced. The data were classified into 25 subgroups of four observations each by measuring the diameters of the first four units in each batch of twelve units. Table 1 gives the 100 recorded data observations. Figure 2 is a histogram of the 100 data points, and shows an approximately normal shape with no observations outside the specification limits.

Table 1: Pilot OD Data

subgroup	1	2	3	4	$\overline{X}_i$	R_i
1	–10	–6	0	0	–4.0	10
2	–14	–4	–6	4	–5.0	18
3	–2	12	–2	8	4.0	14
4	–4	–6	–6	–2	–4.5	4
5	12	6	2	2	5.5	10
6	0	0	–6	–8	–3.5	8
7	2	–6	8	–6	–0.5	14
8	0	6	4	8	4.5	8
9	2	4	6	8	5.0	6
10	–8	0	–4	2	–2.5	10
11	4	2	2	6	3.5	4
12	–8	4	–14	6	–3.0	20
13	–10	2	–10	4	–3.5	14
14	–8	2	–4	4	–1.5	12
15	6	16	10	18	12.5	12
16	2	2	0	2	1.5	2
17	12	6	0	2	5.0	12
18	2	2	0	–8	–1.0	10
19	–6	–4	2	0	–2.0	8
20	–2	4	0	4	1.5	6
21	2	4	2	6	3.5	4
22	0	4	2	4	2.5	4
23	–2	4	–2	4	1.0	6
24	–10	4	–12	4	–3.5	16
25	6	8	–4	2	3.0	12

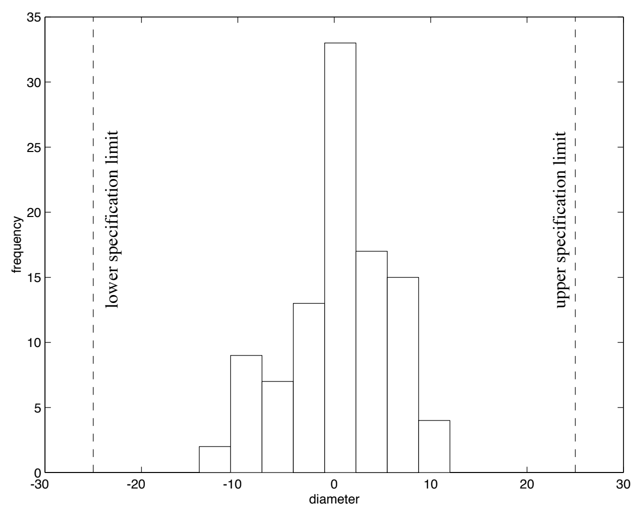

Figure 2: Histogram of Pilot OD data

Figure 3 shows the corresponding $\overline{X}$ and R control charts. There is one out-of-control point on the $\overline{X}$ chart corresponding to subgroup 15.

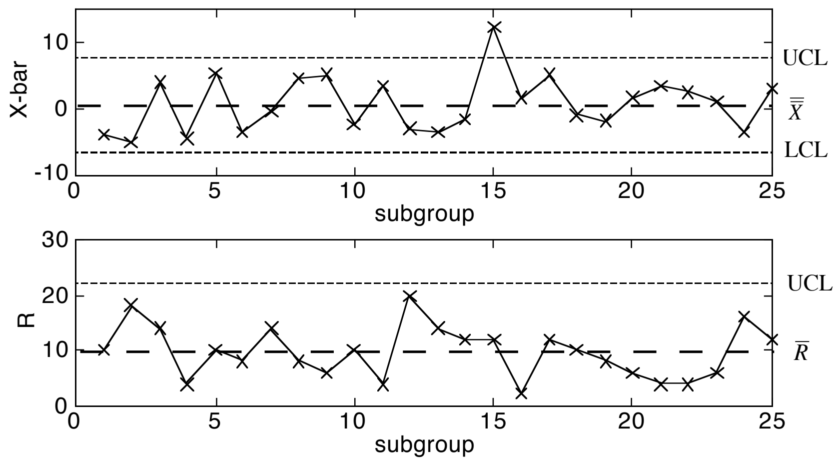

Figure 3: $\overline{\mathrm{X}}$ and *R* control charts from Pilot OD example

Based on the data in Table 1 we calculate the following quantities: $\overline{\overline{X}} = 0.74$, $\overline{R} = 9.76$, and $\hat{\sigma}_s = 6.11$. Since, in this example, the subgroup size equals four, $d_2 = 2.059$ and thus $\hat{\sigma}_{\overline{R}/d_2} = 4.74$. Using the definitions (1) and (2) yields C_{pk} = min(1.81, 1.71) = 1.71 and P_{pk} = min(1.40, 1.32) = 1.32. In this case, C_{pk} and P_{pk} are quite different, and, in fact, lie on different sides of the key cutoff values 1.33 and 1.67 given in QS-9000. Which capability index is better in this example?

As shown in (1) and (2), the measures C_{pk} and P_{pk} differ only in the estimate of the process standard deviation used in the denominator. As a result, to compare the two capability measures we need to compare the two standard deviation estimates $\hat{\sigma}_{\overline{R}/d_2}$ and $\hat{\sigma}_s$.

There is one important differences between $\hat{\sigma}_{\overline{R}/d_2}$ and $\hat{\sigma}_s$. Since the range-based estimate $\hat{\sigma}_{\overline{R}/d_2}$ is calculated based on subgroup ranges, it uses only the variability within each subgroup to estimate the process standard deviation. The sample standard deviation-based estimate $\hat{\sigma}_s$, on the other hand, combines all the data together, and thus uses both the within subgroup and between subgroup variability. The total variation in the Pilot OD process is the sum of the within subgroup and between subgroup variability. As a result, $\hat{\sigma}_s$ estimates the total variation present in the process while $\hat{\sigma}_{\overline{R}/d_2}$ estimates only the within subgroup variation.

The question of which estimate provides a more appropriate measure of process variability to use in process capability calculations can be answered by taking a customer perspective. Customers are concerned about all the variation in the process output, regardless of its source. As a result, the capability of a process should be based on the process' total variation, i.e. we should use the capability index P_{pk}. C_{pk} seriously underestimates the total variation if the between subgroup variability is substantial. This is illustrated in the Pilot OD example where the lack of stability, shown by the out-of-control point on the $\overline{X}$ chart, is evidence of substantial between subgroup variability. The main purpose of using capability indices is for customer reporting, as such it makes sense to consider the indices in terms of what the customer wants to know.

Note that in all cases of practical interest the estimate $\hat{\sigma}_s$ is larger than $\hat{\sigma}_{\overline{R}/d_2}$, since $\hat{\sigma}_s$ includes the between subgroup variability in the calculations. Thus, P_{pk} tends to be smaller than C_{pk}, and using P_{pk} rather than C_{pk} makes the process "look worse." For this reason, suppliers may be reluctant to use P_{pk} rather than C_{pk}. However, it is beneficial for both parties to obtain a realistic view of the capability of the process to produce parts within specification.

To further illustrate the differences and similarities between C_{pk} and P_{pk} we consider the Pilot OD data with some small changes. By definition the subgroup range-based estimate $\hat{\sigma}_{\overline{R}/d_2}$ is unaffected by changes to the individual observations so long as the subgroup ranges do not change. For example, subtracting 12.5 from all observations in the 15th subgroup of the Pilot OD example has no effect on R_{15}, and thus has no effect on $\hat{\sigma}_{\overline{R}/d_2} = \overline{R}/d_2$. If, in addition, the global average $\overline{\overline{X}}$ is unchanged, the capability index C_{pk} given by (2) will be unaffected. Figure 4 shows the resulting $\overline{X}$ and R control charts when 12.5 is subtracted from all observations in subgroup 15, and 6.25 is added to all observations in subgroups 1 and 2. The control charts now suggest a stable process.

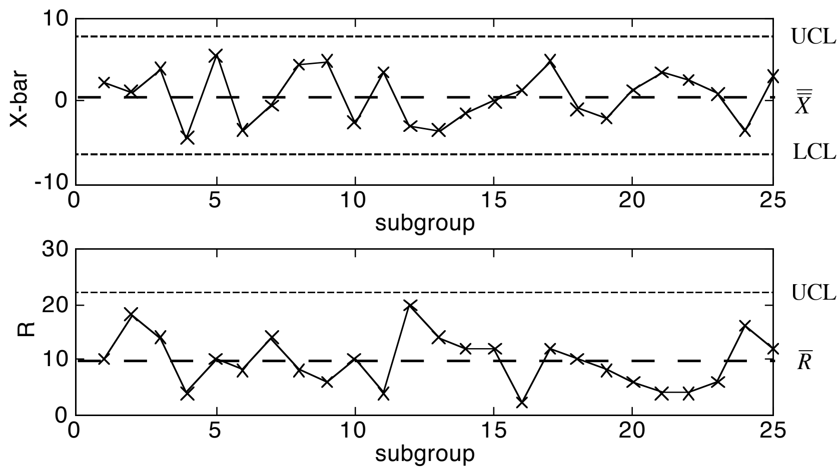

Figure 4: $\overline{\mathrm{X}}$ and *R* control charts from altered Pilot OD data

With the suggested changes to the data, $\overline{\overline{X}}$ is still 0.74, $\hat{\sigma}_{\overline{R}/d_2}$ still equals 4.74, and thus C_{pk} is unchanged at 1.71. However, now $\hat{\sigma}_s = 5.45$, and thus P_{pk} equals 1.48. C_{pk} and P_{pk} are closer since the between subgroup variability has been reduced. The small amount of between subgroup variation is also shown by the in-control $\overline{X}$ control chart. In general, for stable processes C_{pk} and P_{pk} will be similar. However, even for stable processes, P_{pk} is a better measure of capability since the small amount of between subgroup variability still contributes to the total variability in the process output.

The standard deviation estimates $\hat{\sigma}_s$ and $\hat{\sigma}_{\overline{R}/d_2}$ also differ in a less fundamental, but also important, way. The subgroup range-based approach yields estimates that are not as efficient as the sample standard deviation method even if the between subgroup variation is zero. For example, in a capability study that uses 100 observations divided into 25 subgroups the range-based method has an efficiency of only approximately 86% compared with the sample standard deviation method. This loss of efficiency results mostly from a loss of degrees of freedom, and means that when using the range method, process information is discarded needlessly. The less efficient range-based estimate is popular since it is used in control charts and can be calculated easily by hand.

3. Issues Relating to Capability Indices

This section provides a discussion of various important issues relating to the calculation and interpretation of capability indices.

3.1 Process View (Sampling Scheme)

As shown in (1) and (2), the capability of a process is estimated from collected data that represents a sample of the total production. Clearly, as a result, the capability indices C_{pk} and P_{pk} are greatly influenced by the way in which the process data are collected, what we will call the process view. A process view is defined by the time frame, and sampling method (sampling frequency, sample size, etc.) used to obtain the process data. Using an appropriate process view is crucial since different views can lead to very different conclusions. For example, in one view the process may appear stable, while in another the process appears unstable.

To define the process view, the first choice involves the time frame over which process data will be collected. Often the time frame is stipulated by the customer as a reporting interval. For example, the capability of each important process characteristic may be reported every quarter. In other situations, such as for characteristics subject to PPAP requirements, the time frame is restricted to a shorter interval, such as the production period needed to produce 300 units. To obtain a reasonable measure of the process capability, the length of the time frame should be chosen so that it is long enough to reflect all the substantial sources of variation in the process.

Defining the sampling method or procedure is also important. The process output should be sampled in such a way that we obtain a "fair" representation of the process over the chosen time frame. For the capability calculations, it is not necessary for the samples to be collected in subgroups. However, since subgroups can also be used to create control charts that may be helpful in managing the process, subgrouping of the data is recommended.

To illustrate the importance of the sampling procedure, or process view, consider a tool wear example. Figure 5 shows the output of the process over a selected time frame for every part produced. The tool replacement times are clearly visible. Although this process is unstable, we can sample in such a way that the process appears stable and the process capability appears high. For example, taking a subgroup of size five from the start of each tool cycle likely results in a stable control chart, since at the start of each tool cycle the process mean is close to ten. Calculating P_{pk} based on the data from this process view would lead to a large value, since the process variation at the start of each tool cycle is small compared with the distance to the upper specification limit.

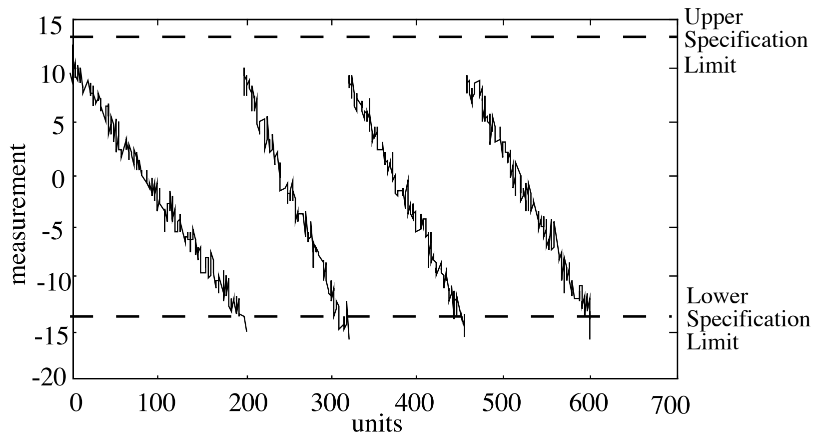

Figure 5: Tool Wear Example

However, sampling from the process in this way is clearly not a fair representation of the output of this process, and thus P_{pk} calculated from this data does not accurately represent the process capability. This is shown by the out–of–specification units produced at the end of each tool cycle. In this tool wear example, to obtain a "fair" representation of the process we should sample throughout one or more tooling cycles. Considering this example, it is clear that altering the process view can substantially change the conclusions about the process capability. As a result, more specific guidelines regarding the time frame, and the sampling method used to collect the data necessary for calculating capability should be given.

Another important issue related to the process view is the number of data points used in the estimation. P_{pk} is an estimate of the process capability, thus even if the process is unchanged, taking another sample and recalculating the index is unlikely to yield precisely the same result. The amount of uncertainty is based on both the properties of the process and the number of data observations used to calculate the capability index. Larger sample sizes provide more information and thus tend to lead to better estimates of the process capability. We recommend that a minimum of 100 observations be used to estimate the process capability.

3.2 Process Stability

A process is considered stable if all the points on its $\overline{X}$ and R control charts fall within the control limits, and there are no apparent patterns. The stability of a process is an important property since if the process is stable in the current time frame it is likely to also be stable in the future, assuming that no major changes occur. Thus, the total output of a stable process is, in some sense, predictable. If the output of a process is stable, then the process' capability is predictable from past performance. On the other hand, if the process output is not stable, it is still possible that over time the process capability index is stable. For example, a process subject to tool wear is not stable, but if worn tools are replaced well before non-conforming units are produced the process will likely produce no parts out-of-specification, and should be consider very capability. There are many applications that involve substantial and unavoidable tool wear, where, if properly managed, the process produces only parts well within the specification limits. Although not clearly specified in the QS-9000 standard, the chronically unstable, yet predictable process mentioned in the standard may refer to a process subject to tool wear whose output changes systematically.

To also account for capable processes subject to tool wear more reliable indicators of the predictability of the process capability can be obtained by considering the performance of the process in terms of its process capability over time. If the past process capability values exhibit a stable (or increasing if the quality is improving) pattern then we would have some confidence predicting future process capability indices.

In any event, setting aside the issue of process stability, we may examine the consequences of using the different capability indices C_{pk} and P_{pk}. As shown in the example, if the process is stable, C_{pk} is approximately equal to P_{pk}, since a stable process has little between subgroup variability. Thus, if the process is stable, it does not matter much which measure is used (though

P_{pk} is still preferred). On the other hand, if the process is unstable, there is substantial between subgroup variability, and C_{pk} is not equal to P_{pk}. In this case, C_{pk} overestimates the process capability since it does not include the between subgroup variability. The same thing applies if the process is chronically unstable and yet predictable. As a result, in all situations, P_{pk} provides a better measure of the process capability than C_{pk}.

3.3 Different Distributions

When interpreting and comparing capability indices from different processes or from the same process over time we must be careful. Process capability indices calculated from non–normal processes are not comparable with those from normal processes in terms of the proportion falling outside the specification limits. The reason for this difference is illustrated in Figure 6. The figure shows the distribution of the output of two processes which have equal process means and standard deviations, and thus equal process capabilities, but yield different proportions of nonconforming units. As shown, the skewed process produces more out-of-specification parts than the process whose output is normally distributed. This means that the capability indices of processes whose output distributions are not similar should not be directly compared. As a result, using capability indices to prioritize improvement efforts without checking the process distribution can lead to poor choices.

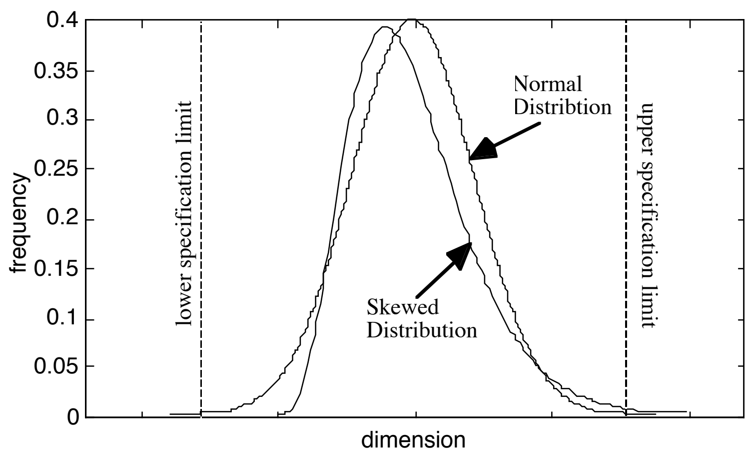

Figure 6: Normal versus Non–normal Distribution Plot

This is one reason why plotting a histogram of the data used to calculate the capability index, as given in Figure 2, is always useful. A histogram of the data will show whether the normal assumption is reasonable, and may also provide information regarding why the capability index is not higher. For example, it may be evident that the process is not centered, or that there are a few outliers that have a large influence on the index.

4. Proposal for the Appropriate Use of Capability Indices

Capability indices play an important part in quality reporting. The requirements given in quality standard, such as QS-9000, are confusing to suppliers and do not focus on all the important issues. To alleviate this confusion we recommend the following guidelines for the appropriate use of capability indices.

When reporting capability indices the supplier requires data to demonstrate both stability and capability of the process. To specify the monitoring scheme, the following questions must be considered:

- what is the time frame for the demonstration?
- what is the sampling method?
  - what is the appropriate subgrouping?
  - when and how often should these subgroups be collected?
  - how much data is needed?
- how should the data be reported?

A consistent time frame for monitoring is needed so that data are available to the customer on a timely basis while keeping the costs reasonable. Depending on volumes and past history of performance, a monthly or quarterly report may be sufficient.

The goal of the sampling method is to produce data that are representative of performance of the process over the specified time frame. Sufficient data are required to reduce the errors of estimation. At least 100 data points should be required and these points should be collected in subgroups spread out through the time frame in some rational way.

For variables data, the minimal reporting requirements should be a control chart with limits calculated internally to show the nature of stability over the time frame, a histogram to show shape of the process output distribution, and P_{pk} to compare performance to specifications. The index P_{pk} is preferable to C_{pk} since it captures all the process variation. A run chart of P_{pk} over all past reporting periods would also be useful to identify the consistency of the process capability and/or show process improvements. Minimal default capability requirements for most characteristics could be given in a simple statement such as $P_{pk} > 1.33$.

References

Automotive Industry Action Group, (1995) *Advanced Product Quality Planning and Control Plan: Reference Manual* .

Automotive Industry Action Group, (1995) *Production Part Approval Process*.

Automotive Industry Action Group, (1995) *Quality System Requirements QS-9000*.

Automotive Industry Action Group, (1995) *Statistical Process Control*.

Gunter, B.H. (1989a), "The Use and Abuse of C_{pk}" *Quality Progress*, January, 72-73.

Gunter, B.H. (1989b), "The Use and Abuse of C_{pk}, Part 2" *Quality Progress*, March, 108-109.

Gunter, B.H. (1989c), "The Use and Abuse of C_{pk}, Part 3" *Quality Progress*, May, 79-80.

Gunter, B.H. (1989d), "The Use and Abuse of C_{pk}, Part 4" *Quality Progress*, July, 86-87.

Kotz, S. (1993), "Process Capability Indices," Chapman and Hall, New York.

Montgomery, D.C. (1991), *Introduction to Statistical Quality Control*, Second Edition, John Wiley and Sons, New York.