Some notes on the computation and interpretation of capability indices

Olle Carlsson and Sture Rydin

ABSTRACT: It is shown that the common capability index, C_{pk}, and the statistical tolerance interval are identical concepts. The capability index is affected by the method used to calculate the standard deviation for the data set, and this is demonstrated both practically and theoretically. It is also shown that the interpretation of a capability index depends on the sample size and the aggregation level of the original data.

KEYWORDS: Computation, indexing, information retrieval, measurement, quality control, sampling, specifications, standards, tolerance, variability.

The farther a measured value is from its target value, the more likely it is to cause problems for the consumer. Conventional quality requirements specify that the mean value for a sample should fall within a given interval. While this provides an assurance for the process level, it does not provide assurance for single items. In contrast, capability indices represent a summarized measure of both the process level and its variability. Used correctly, capability indices can provide information about the percentage of single items falling outside given limits.

Capability indices have become popular and are widely used. However, there are several important statistical issues regarding the correct use of these indices. While these issues have been addressed in the literature, the discussions have appeared only in advanced statistical journals, which are not well known by many users. A comprehensive review of the statistical aspects of the capability indices is available *(1)*. Critical studies were done by Gunter *(2–6)* and Chan *et al. (7)*.

Manufacturers seek to control the variability of their processes by requiring their suppliers to provide a capability index for incoming materials. In the pulp and paper industry, some customers require capability indices for a number of quality characteristics. The requirements are usually expressed so that the quality characteristics exceed a certain preset minimum value. In such cases, a lower capability index, C_{pl}, is commonly used. This capability index is defined as

$$C_{pl} = (\mu - L)/3\sigma \qquad (1)$$

where

μ = target value

σ = standard deviation

L = lower limit.

The lower limit L is the specification limit for a single item. The required size of C_{pl} is arbitrary, but in general, C_{pl} should be larger than 1.33.

Control plans based on the sample mean also involve upper and lower limits or single limits, which are the specification limits of the process level.

Tolerance interval

The tolerance interval is a familiar concept in the field of quality control. Standards such as MIL-STD-414 *(8)* and other specific standards, such as the specification for fiber building board *(9)*, are based on tolerance interval.

To introduce the concept of tolerance interval, first assume that X is a quality characteristic of interest that follows a normal distribution with a known mean value (process level), μ, and with a known standard deviation, σ. An individual observation, X, is said to belong to a lower tolerance interval if

$$X > \mu - k\sigma \qquad (2)$$

where the point $\mu - k\sigma$ is called the tolerance limit.

Carlsson is an associate professor in the Dept. of Statistics, University of Örebro, P.O. Box 923, S-701 30 Örebro, Sweden. Rydin is a systems engineer at Stora Cell Industries, S-814 89 Skutskär, Sweden.

The lower capability index can be rewritten

$$L = \mu - 3C_{pl}\sigma \quad (3)$$

so that this index defines a tolerance interval with $k = 3C_{pl}$. The probability of an individual item falling below L is $P(X < L)$. Using Eq. 3 and standardizing the expression gives

$$P(X<L) = \Phi(-k) \quad (4)$$

where Φ is the standardized normal distribution function.

A value of $C_{pu} = 1.33$ corresponds to a value of $k = 4$ and

$$P(X<L) = \Phi(-4) = 32 \times 10^{-6}$$

or, expressed another way, if $k = 4$, an average of only one out of a total of 31,500 units will not meet the specification.

The upper capability index, C_{pu}, is also a tolerance interval and is defined as

$$C_{pu} = (U - \mu)/3\sigma \quad (5)$$

where U is the upper tolerance limit.

The common capability index, C_{pk}, is defined as

$$C_{pk} = \min(C_{pl}, C_{pu}) \quad (6)$$

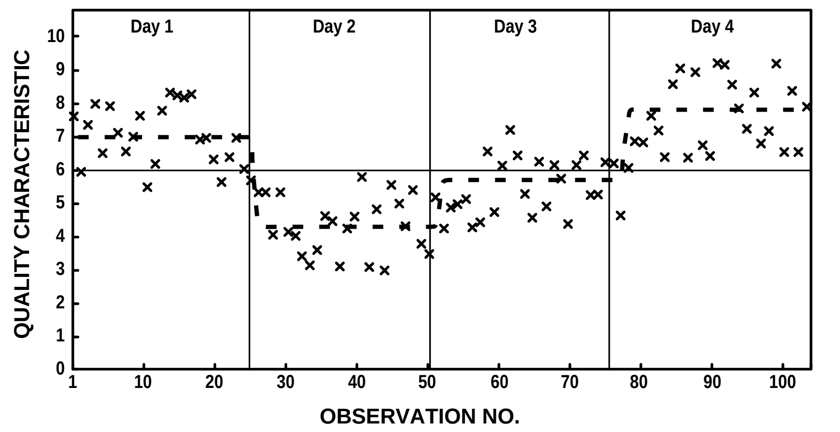
The C_{pk} is a measure of the capability of the production process. For the customer, it is considered to be a measure of both the location of the process level and the variability of the process. Most customers require knowledge about the C_{pk} of the lot.

However, μ and σ are unknown parameters in most cases and have to be estimated. The most common estimators are the sample mean $\bar{X}$ and the standard deviation S. In this case, the corresponding tolerance limit changes from $\mu - k\sigma$ to $\bar{X} - kS$. This limit is no longer a fixed point but a random variable, and the same is true for the capability index, here denoted as

$$C'_{pl} = (\bar{X} - L)/3S \quad (7)$$

It is not possible, however, to compute the probability of an item falling within the boundaries of the tolerance interval. For example, if C'_{pl} exceeds 1.33, it will not be a sufficient condition to guarantee that the fraction falling within the tolerance interval is equal to $\Phi(4) = 0.9999683$. Instead, statements similar to ordinary confidence intervals have to be made, e.g., "With a confidence of γ%, then $(1-P)$% of possible values for X will fall within that interval."

1. Time series with 100 observations from a lot manufactured during a period of four days ($L = 0.5$, $\mu = 6.0$)



I. Estimated process levels, standard deviations, and capability indices for each group of observations*

	Number of observations	*Mean*	*Standard deviation*	*Lower capability index*
Day 1	25	7.1	0.84	2.6
Day 2	25	4.5	0.88	1.5
Day 3	25	5.8	0.84	2.1
Day 4	25	7.8	1.11	2.2

*L = 0.5 and μ =6.0

Computing a capability index

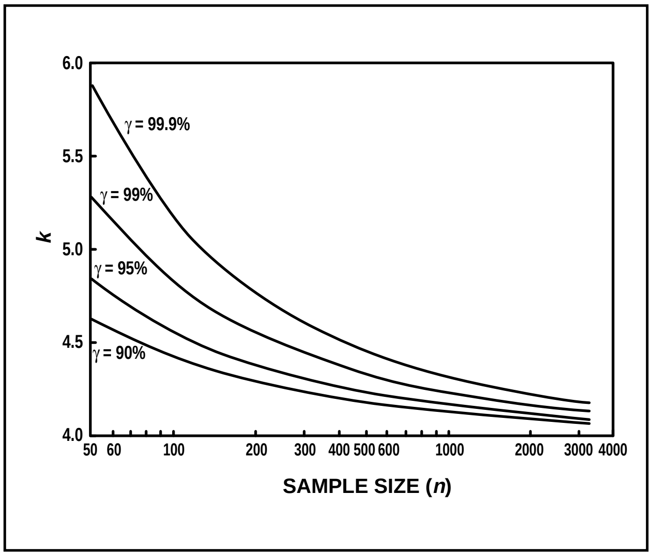
Assume a set of observations of a quality characteristic for a certain lot delivered to a customer, as shown in **Fig. 1**. The observations are assumed to originate either from measurements of the production process or from measurement of samples by the customer. Finally, assume that the customer demands that $C'_{pl} > 1.33$.

The solid line in Fig. 1 represents the setpoint of the process level, μ, while the dotted lines correspond to the computed daily means. The data points represent individual measurements. The daily means, the standard deviations, and the capability indices are given in **Table I**.

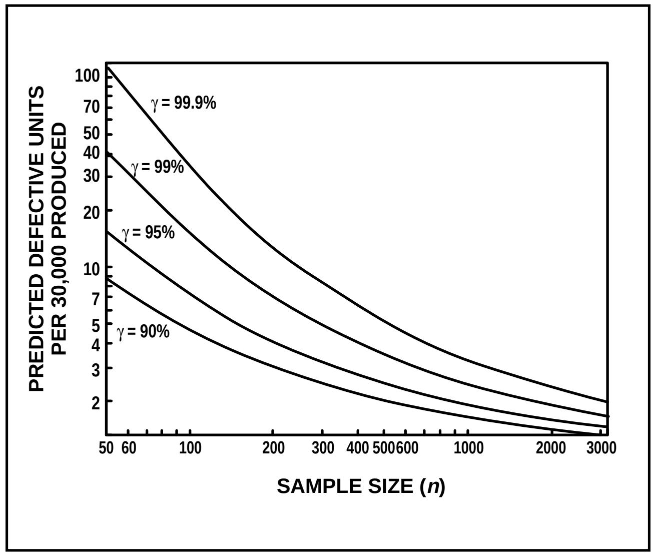
Now, suppose that the producer or the customer wants to compute a capability index that would represent the whole production period. This can be done in a number of ways.

First, compute the standard deviation of the whole production period from the sum of the daily variances, i.e., $S^2_{prod} = (0.84^2 + 0.88^2 + 0.84^2 + 1.11^2)/4 = 0.854$ and $S_{prod} = 0.92$. With a common mean of 6.3 for

2. Coefficient k as a function of sample size and confidence level when the probability of a defect is $P = 32 \times 10^{-6}$ (one defect per 30,000 units).



3. The expected number of defect units per 30,000 manufactured units when $k = 4$; μ and σ are unknown



II. Coefficient k for three sample sizes and three confidence levels at $P = 0.001$*

Sample size (n)	*Confidence level* γ *90%*	*95%*	*99%*
10	4.629	5.503	6.603
30	3.794	4.022	4.508
50	3.604	3.766	4.096

*$P = 0.001$ indicates the probability of a defective unit, i.e., 0.1% of the lot below the lower specification limit L

the four days of production, this gives $C'_{pl\,(prod)} = 2.10$.

Second, assume that the customer who receives these data with the shipment now wants to compute the $C'_{pl\,(cust)}$ of the lot. It is often said that capability indices should only be used when the process is stable, but the customer usually has no interest in knowing *when* the delivered items were produced. Rather, the customer wants to know on what process level they are produced and how homogenous the lot is. The customer computes the variance around the common mean, which is 6.3, with the standard deviation, $S_{cust} = 1.5$. The customer's capability index, $C'_{pl\,(cust)}$, becomes 1.29, which is smaller than 1.33.

Finally, the situation would have been even less favorable to the producer if the customer had used the target value $\mu = 6.0$ when computing the variance. In this case, the standard deviation S_{tot} becomes 1.59, and $C'_{pl\,(cust)}$ drops to 1.22.

From these calculations, one arrives at the following inequality:

$$S^2_{tot} \geq S^2_{cust} \geq S^2_{prod}$$

and as a consequence

$$C'_{pl\,(tot)} \leq C'_{pl\,(cust)} \leq C'_{pl\,(prod)}$$

This inequality is generally valid, as demonstrated in the appendix at the end of this paper.

Capability index and the importance of the sample size

In many practical situations, the parameter k is usually set to 4, or equivalently, $C'_{pl} = 1.33$, irrespective of the sample size, n. Such approximate tolerance intervals can result in serious erroneous conclusions, especially when the sample size is small.

The parameter k should be chosen with regard to the sample size, n, the confidence level, γ, and the proportion of product meeting specifications, $1 - P$, e.g., $k = k(P, \gamma, n)$. It is not possible to make an explicit computation of the coefficient k, but tables are available *(10)*. These tables show that the value of k will always be greater than those valid for known μ and σ. Examples of the relations between k, n, P, and γ are given in **Fig. 2** and **Table II**.

Assume that a producer of kraft liner has made the following agreement with a customer: the sample size is determined to be $n = 30$, with the requirement that no more than 0.1% of the delivered units are permitted to be below L = 660 kPa with a $\gamma = 99\%$ confidence level. Table II gives a value of $k = 4.508$, which implies that the agreement is fulfilled only if $\bar{X} - 4.508S > 660$ kPa or, equivalently, that $C'_{pl} > 1.50$.

In Fig. 2, the constant k (and, by extension, the capability index, since $k = 3C_{pl}$) for a tolerance interval is shown graphically as a function of the sample size, n. The curves are associated with different confidence levels, γ, and the probability of a defective unit, P, is less than 32 x 10^{-6}. It is apparent from this figure that the parameter k always has to be larger than four and, consequently, that C'_{pl} has to be larger than 1.33 for all n. If $n = 100$, then k should be 4.42 and 5.18 for confidence levels of 90% and 99.9%, respectively.

If, on the other hand, $k = 4$ (or $C'_{pl} = 1.33$), then the expected number of defective units will be about 1 per 30,000 manufactured units when μ and σ are known. In **Fig. 3**, the expected number of defectives per 30,000 units are given for different confidence levels and sample sizes when μ and σ are unknown and estimated by using $\bar{X}$ and S. As an example, it can be seen that for a sample size of $n = 100$ and for a tolerance interval of 90% or 99.9%, the expected number of defective units will be equal to 5 and 34, respectively.

The customer usually accepts the existence of a certain, natural variability that is taken into consideration when the capability indices are agreed upon. However, the customer can get a higher capability index than originally required—even if the producer has a higher process variability than accepted by the customer—if the customer is agreeable to the producer's raising of the process level. Of course, this is an unsatisfactory solution, since a higher process level usually means increased production costs. The relationship between production costs and process level within a pulp mill has been studied by the authors *(11)*.

Appendix

Let

$$\bar{X} = (\sum_j^p \sum_i^n Xji)/pn = (1/p) \sum_{j=1}^{p} \bar{X}j$$

where

$$\bar{X}j = (1/n) \sum_{j=1}^{n} Xji$$

and

$$j = 1, 2, \ldots, p$$

The sum of squares, SS, can be written

$$SS_{tot} = \sum_j^p \sum_i^n (Xji-\mu)^2 = \sum_j^p \sum_i^n (Xji-\bar{X})^2 + np(\bar{X}-\mu)^2$$

$$= SS_{cust} + np(\bar{X}-\mu)^2$$

and

$$SS_{cust} = \sum_j^p \sum_i^n (Xji-\bar{X})^2$$

$$= \sum_j^p \sum_i^n (Xji-\bar{X}j)^2 + n \sum_i^n (\bar{X}j-\bar{X}])^2$$

$$= SS_{prod} + n \sum_j^p (\bar{X}j-\bar{X})^2$$

Then

$$SS_{tot} \geq SS_{cust} \geq SS_{prod}$$

After dividing with the number of degrees of freedom for SS_{tot}, i.e., np, we get

$$S^2_{tot} \approx S^2_{cust} + (\bar{X}-\mu)^2$$

In the same way,

$$S^2_{cust} \approx S^2_{prod} + \{\sum_j^p (\bar{X}j-\bar{X})^2]/p\}$$

and, finally,

$$S^2_{tot} \geq S^2_{cust} \geq S^2_{prod}$$

Literature cited

1. *J. Quality Technol.* 24: 175(1992).
2. Gunter, B. H., *Quality Progress* 22: 72(1989).
3. Gunter, B. H., *Quality Progress* 22(2): 108(1989).
4. Gunter, B. H., *Quality Progress* 22(3): 79(1989).
5. Gunter, B. H., *Quality Progress* 22(4): 86(1989).
6. Gunter, B. H., *Quality Progress* 24(1): 90(1991).
7. Chan, L.K., Cheng, S.W., and Spiring, F.A., *et al.*, *J. Quality Technol.* 20: 162(1988).
8. "Sampling procedures and tables for inspection by variables for percent defective," MIL-STD 414, U.S. Govt. Printing Office, Washington, DC, 1957.
9. British Standard 1142:2, British Standards Institution, London, 1971.
10. Owen, D. B., *Technometrics* 10: 445(1968).
11. Carlsson, O. and Rydin, S., *Intern. Statistical Rev.* 61(1): 109(1993).

Received for review Oct. 11, 1993.

Accepted June 18, 1994.