

A Practitioner's Guide to Process Capability Analysis

Cp, Cpk, Pp, Ppk from first principles. What each index measures, when assumptions are violated, why confidence intervals matter, and the common mistakes that produce misleading capability numbers.

Ref: AIAG SPC Manual, 2nd Ed. · Montgomery, 8th Ed. · ISO 22514-2 · ASTM E2281

I — Publication

A Practitioner's Guide to Process Capability Analysis

Cp, Cpk, Pp, Ppk from first principles. What each index measures, when assumptions are violated, why confidence intervals matter, and the common mistakes that produce misleading capability numbers.

Ref: AIAG SPC Manual, 2nd Ed. · Montgomery, 8th Ed. · ISO 22514-2 · ASTM E2281

Date 2026-01-15

Author Eric Wolters

Reading 15 min

SPC · Capability

II — Definitions

The Four Indices

Process capability analysis reduces the relationship between a process and its specification to a handful of ratios. Four indices dominate practice: Cp, Cpk, Pp, and Ppk. They share a common structure—each is the ratio of the specification width to some measure of process spread—but they answer different questions, and confusing them is one of the most common errors in quality engineering.

Cp measures potential capability. It compares the specification width (USL - LSL) to the process spread, defined as 6 times the within-subgroup standard deviation: $Cp = (USL - LSL) / 6\sigma_{\text{within}}$. Cp ignores the process mean entirely. A process centered perfectly between the specification limits and a process sitting on one limit can have the same Cp. This makes Cp a measure of what the process *could* achieve if it were centered, not what it actually achieves.

Cpk corrects for centering. It takes the distance from the process mean to the nearest specification limit and divides by half the process spread: $Cpk = \min((USL - \bar{x}) / 3\sigma_{\text{within}}, (\bar{x} - LSL) / 3\sigma_{\text{within}})$. When the process is perfectly centered, $Cpk = Cp$. When the process mean shifts toward a limit, Cpk decreases. Cpk is always less than or equal to Cp.

Pp and Ppk are the performance analogues. The formulas are identical to Cp and Cpk, but the standard deviation uses the overall (total) standard deviation computed from all individual observations, rather than the within-subgroup estimate. The within-subgroup estimate (typically from $\bar{R}/d_2$ or $\bar{S}/c_4$) captures only the short-term, common-cause variation present within rational subgroups. The overall standard deviation captures everything: common cause, between-subgroup shifts, operator differences, material lot variation, and every other source of variation that operates between subgroups.

Within vs. Overall: What the Gap Tells You

When Cp and Pp are nearly equal (equivalently, when Cpk and Ppk are nearly equal), the process has no significant between-subgroup variation. The within-subgroup estimate and the overall estimate agree. This is a stable process with no special causes operating between subgroups.

When Cp is substantially larger than Pp, additional variation is entering the process between subgroups. The process has short-term potential (captured by Cp) that it fails to sustain over time (captured by Pp). This gap is a diagnostic: it tells you that improving capability requires reducing the between-subgroup sources, not tightening the within-subgroup variation. The Cp-Pp gap is often more actionable than either index alone.

III — Assumptions

Normality

The standard capability formulas assume the data follow a normal distribution. This assumption is embedded in the interpretation: $Cpk = 1.00$ implies 2,700 ppm nonconforming (0.27%) only if the distribution is normal. For non-normal data, the actual nonconforming fraction at $Cpk = 1.00$ could be higher or lower than 2,700 ppm, depending on the shape of the distribution.

Many manufacturing processes produce non-normal data. Surface roughness, flatness, concentricity, and most geometric tolerances follow skewed distributions. Chemical concentrations, hardness values, and cycle times are often right-skewed. Applying normal-based indices to these data produces numbers that do not correspond to the actual fraction nonconforming.

Handling Non-Normal Data

Three approaches exist. The **transformation method** applies a Box-Cox or Johnson transformation to normalize the data, computes capability indices on the transformed scale, and reports the results. This works well when a simple transformation achieves normality, but the specification limits must also be transformed, and the interpretation becomes less intuitive.

The **percentile method** (ISO 22514-2) estimates the 0.135th and 99.865th percentiles of the actual distribution and uses them in place of the $\pm 3\sigma$ spread. This method makes no distributional assumption and directly estimates the process spread that corresponds to the 3-sigma equivalent. It requires adequate sample size to estimate extreme percentiles reliably—typically $n > 100$.

The **distribution fitting method** fits a parametric distribution (Weibull, lognormal, gamma) to the data and computes the expected nonconforming fraction directly. This is the most accurate approach when the correct distribution family is identified, but it requires judgment in distribution selection and goodness-of-fit testing.

Stability

Capability indices are meaningless for an unstable process. If the process mean or variance is shifting over time, computing an index from a sample is equivalent to averaging across different process states. The resulting number does not describe the current process, the past process, or the future process. It describes nothing.

Always confirm statistical stability with a control chart before computing capability indices. This is not a suggestion. It is a requirement stated in the AIAG SPC Manual, in Montgomery, and in ISO 22514-2. The control chart comes first. If the process is not in control, the correct action is to identify and remove the special causes, not to compute a capability index and hope for the best.

Sample Size

The AIAG manual recommends a minimum of 25 subgroups for control chart construction and at least 100 individual observations for capability analysis. These are minimums, not targets. With $n = 30$ individual observations, the 95% confidence interval on Cpk is approximately ± 0.35 around the point estimate. A reported Cpk of 1.40 from 30 observations is consistent with true values anywhere from 1.05 to 1.75. This is not precision. This is a rough guess.

IV — Confidence Intervals

Every Capability Index Is an Estimate

The point estimate of Cpk is a function of the sample mean and sample standard deviation. Both are random variables. A different sample from the same process would produce a different Cpk. The point estimate alone, without a measure of its uncertainty, is incomplete. Yet most capability reports present only the point estimate.

The approximate 95% lower confidence bound for Cpk, assuming normality, is:

$$Cpk_{lower} = Cpk * (1 - z_{0.025} * \sqrt{1/(9*n*Cpk^2) + 1/(2*(n-1))})$$

where n is the sample size and $z_{0.025} = 1.96$. This formula, from Chou et al. (1990), provides a useful approximation. For Cpk = 1.33 with $n = 30$, the lower 95% bound is approximately 0.99. The data are consistent with a process that does not meet the 1.33 requirement.

Sample Size to Demonstrate Capability

Customers frequently require $Cpk \geq 1.33$ or $Cpk \geq 1.67$. The question is: how many observations are needed to demonstrate this with confidence? The answer depends on the true Cpk and the desired

confidence level. If the true Cpk is exactly 1.33, you need approximately 150 observations for the lower 95% confidence bound to exceed 1.00—and even then, you have not demonstrated 1.33 with confidence.

To demonstrate $Cpk \geq 1.33$ with 95% confidence (i.e., the lower bound exceeds 1.33), the true Cpk must be substantially higher than 1.33, or the sample must be large. If the true Cpk is 1.67, approximately 70 observations suffice. If the true Cpk is 1.50, approximately 200 observations are needed. These numbers surprise most quality engineers. The [sample size calculator](#) computes the exact requirement for any target Cpk and confidence level.

The Bayesian Alternative

Confidence intervals answer an indirect question: "What is the range of Cpk values consistent with the data at a given confidence level?" The [Bayesian approach](#) answers the direct question: "Given the data, what is the probability that Cpk exceeds 1.33?" For small samples and borderline results, this distinction changes decisions. See [WP-002](#) for a detailed treatment.

V — Common Mistakes

Reporting Cpk from an Unstable Process

This is the most consequential error. An out-of-control process has a non-stationary distribution. The sample statistics (mean, standard deviation) estimated from the data describe the historical average of a changing process. They do not predict the future. Reporting $Cpk = 1.45$ from a process with assignable causes is presenting a number that has no predictive value. It will not hold tomorrow. The only valid action is to stabilize the process first.

Confusing Cp with Cpk

$Cp = 2.0$ sounds excellent. It means the specification is 12 sigma wide relative to the process spread. But if the process mean has shifted 3 sigma from center, $Cpk = 1.0$ and the process is producing 1,350 ppm nonconforming on one side. Cp measures potential; Cpk measures reality. Report both. If Cp is high and Cpk is low, the corrective action is to center the process, not to reduce variation.

Using Short-Term Data for Long-Term Indices

A common mistake in new product introduction: collect 30 parts from a single production run, compute Cpk using the sample standard deviation, and report it as the process capability. This captures only the variation within one run. It misses lot-to-lot material variation, tool wear, operator rotation, environmental changes, and every other source that operates across runs. The reported Cpk is optimistic. It describes one run, not the process.

For long-term capability, collect data across multiple production runs, material lots, operators, and time periods. Use the within-subgroup standard deviation from a control chart for Cpk, and the overall standard deviation for Ppk. Compare the two. If they differ substantially, the between-subgroup sources are significant and the short-term study would have overstated the capability.

The "Cpk = 2.0 from 10 Samples" Illusion

Ten observations from a stable, centered process with true Cpk = 1.33 will occasionally produce a sample Cpk above 2.0. The sampling distribution of Cpk for small n is wide and right-skewed. The point estimate is unreliable. The 95% confidence interval from 10 observations spans roughly ± 0.7 around the point estimate. Presenting Cpk = 2.0 from $n = 10$ as evidence of a highly capable process is statistically indefensible.

VI — When Indices Mislead

Bimodal Distributions

A process that mixes output from two distinct states—two cavities, two spindles, two material lots—can produce a bimodal distribution. The overall standard deviation of the mixture is large, so Cpk looks poor. But each mode individually might be well within specification. The single Cpk number misrepresents the process. The correct approach is to stratify by the source of bimodality and compute capability separately for each stream. If the two modes have different means, the quality action is to reduce the between-stream difference, not to tighten within-stream variation.

Truncated Data

If 100% inspection removes nonconforming parts before capability data are collected, the data are truncated. The observed standard deviation is smaller than the true process standard deviation because the tails have been removed. Capability computed from truncated data is artificially inflated. This is common in industries with automated sorting—the data file looks clean, but the process is worse than the numbers suggest. If sorting is active, use the pre-sort data for capability analysis, or account for truncation in the statistical model.

Autocorrelated Data

Continuous process data (temperature, viscosity, thickness from a roll-to-roll process) are frequently autocorrelated: consecutive observations are not independent. The standard deviation computed from autocorrelated data underestimates the true variation when the autocorrelation is positive (which it usually is in manufacturing). The result is inflated capability indices. For autocorrelated data, use time-series methods to estimate the true process variation, or increase the sampling interval until autocorrelation is negligible.

What to Report Instead

When standard indices are inappropriate, report the estimated fraction nonconforming directly. This is the quantity that actually matters to the customer. Compute it from the fitted distribution (parametric or nonparametric), with a confidence interval. A statement like "estimated 45 ppm nonconforming above USL, 95% upper confidence bound 120 ppm" communicates more than "Cpk = 1.38" from a non-normal distribution where 1.38 does not correspond to any standard ppm interpretation.

The [Cpk calculator](#) computes all four indices with the option to handle non-normal data via the percentile method. Use it as a starting point, but understand what the numbers assume before presenting them.

VII — Related Instruments

INSTRUMENT	SPECIFICATION
Cpk / Ppk Calculator	Traditional process capability indices. Cp, Cpk, Pp, Ppk, sigma level, DPMO.
Bayesian Cpk Calculator	P(Cpk > 1.33) and credible intervals. Posterior distribution of capability.
Sample Size Calculator	Required n for capability demonstration at a given confidence level.

[← All Whitepapers](#)

SVEND is the consulting practice of Eric Wolters – operational excellence, quality management and strategic planning. ILSSI accredited, SCMEP Technical Partnership Provider.
svend.ai/whitepapers/process-capability-analysis/ · published 15 January 2026