Capability analysis

Menu:	QCExpert	Capability

This module computes the capability index, c_p and the performance index, p_p, based on data and user-specified limits. Additional values, like ARL are given as well. The module allows for one-sided specifications as well as for asymmetric (non-normal) distribution.

Data and parameters

Measured values of the quality characteristics of interest are entered as data. The module expects that the data are in one column. Target value has to be specified in the Dialog panel. In addition, at least one of the specification limits has to be entered (LSL, *Lower Specification Limit* or USL, *Upper Specification Limit*). In the *Columns* field, data column is selected. One can specify, whether all data, marked data, or unmarked data will be used for computations in the *Data* field. In the *Plots* field, graphical output requests are to be specified. List of available plots appears in the paragraph 0 below.

When only one specification limit is defined for the process under control (no matter whether lower or upper), it is entered in the dialog panel, while the field for the other limit is left blank. A desired *Confidence level* can be specified as well. It is used subsequently for calculation of confidence interval, capability and performance indexes. *Cp limit* is the value, below which we consider the process as being not capable. In the Protocol output, all index values and their confidence interval limits smaller than *Cp* are marked in red. Usually, 1 is selected for *Cp*.

If the *Classical indexes* field is marked, classical capability and performance indexes: c_p, c_{pk}, c_{pm}, p_p, p_{pk}, p_{pm} are computed, together with additional characteristics. Definitions of these indexes are shown below. When only one specification limit is specified, classical indexes are not computed. Then, one has to mark the *General indexes* selection – and the c_{pk}^* index (based on probabilistic grounds) is computed. This generalized index can be used for one-sided specification limit or asymmetric data, violating the usual normality assumption. (Distributional normality test can be found in the *Elementary statistics* module. If the *Asymmetric data distribution* selection is checked, the software allows for a possibility that the data come form asymmetric (skewed) distribution. The c_{pk}^* calculation is then adjusted via preliminary application of the exponential transformation of the data. Quantile function F^{-1} value (needed during c_{pk}^* calculations) is computed after the transformation. Warning: if the *Asymmetric data distribution* selection is not checked, the software goes straight ahead and uses „forcefully“ normal model, even though the true data generating distribution is not normal. Hence, if one is not sure about distributional symmetry, it is a good strategy to leave the selection checked. Further detail about properties and motivation of the exponential data transformation can be found in the manual for the Transformation module. If the data are not normal, classical indexes commonly give unrealistically optimistic impressions. They often overestimate true values (although they can be underestimate as well). Hence, if the data are not approximately normal, the classical indexes should not be used.

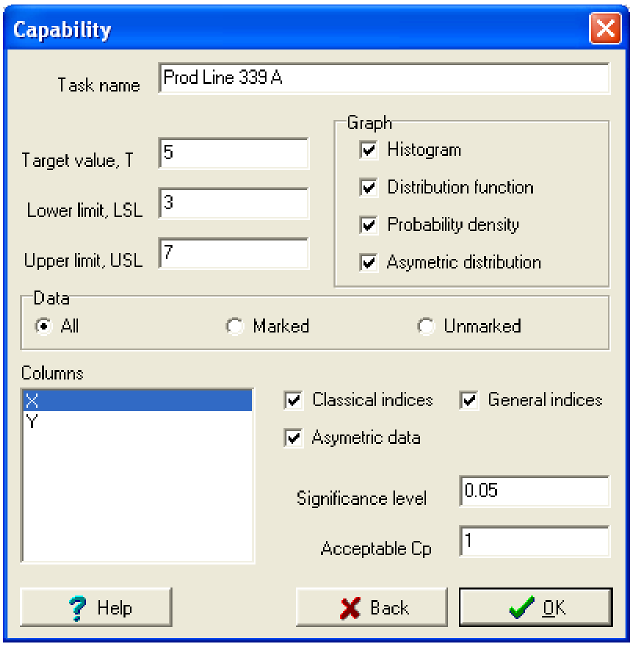


Fig. 1 Dialog panel for Capability

$$c_p = \frac{USL - LSL}{6\sigma_C}, \quad c_{pk} = \frac{\min\left(USL - x, x - LSL\right)}{3\sigma_C}, \quad c_{pm} = \frac{\left(USL - LSL\right)}{6\sqrt{\sigma_C{}^2 + \left(x - T\right)^2}}$$

$$p_p = \frac{USL - LSL}{6\sigma_P}, \quad p_{pk} = \frac{\min\left(USL - x, x - LSL\right)}{3\sigma_P}, \quad p_{pm} = \frac{\left(USL - LSL\right)}{6\sqrt{\sigma_P{}^2 + \left(x - T\right)^2}}$$

$$\sigma_C = \sqrt{\frac{1}{n-1}\sum_1^n \left[x_i - x\right]^2}, \quad \sigma_P = \frac{1}{d_2}\frac{\sum_2^n \left|x_i - x_{i-1}\right|}{n-1}, \quad kde\ d_2 = 1.128$$

$$p_{zm} = F_N\left(\frac{x - LSL}{\sigma_C}\right) + 1 - F_N\left(\frac{USL - x}{\sigma_C}\right)$$

$$ARL = 1/p_{zm}$$

$$c^*_{pk} = -\frac{1}{3}F^{-1}\left\{1/ARL\right\},$$

where F^{-1} is the inverse function to the distribution function (or the quantile function) for the normal distribution.

Remark: Because the estimate is generally not equal to the true value, it is more appropriate not to concentrate only on the point estimate, but to consider associated confidence interval as well. A particular, rater more conservative strategy is typically suggested: behave as if the true the lower confidence interval limit was equal to the true index value. Remember that, for instance if the index comes out as c_p=1.001 with associated confidence interval stretching from 0.8 to 1.2, that the process is very likely not capable (c_p<1)! If the index comes out as c_p=1.2, with the confidence interval ranging from 1.0 to 1.4, it is very unlikely that the process is not capable.

Protocol

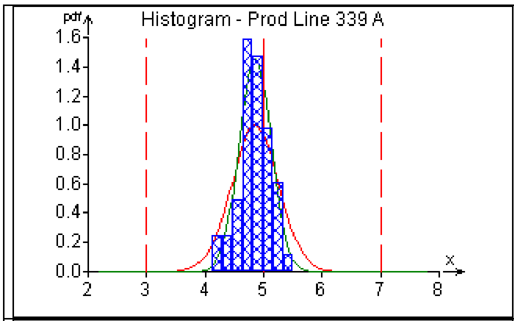
Capability and performance under the normality assumption	Normal distribution based calculations are performed only when both specification limits are given. When only one limit is given, report contains items from the „Cpk for asymmetrically distributed data“.
Project name	Name of the data spreadsheet
Target value	User-specified target parameter value.
Specification limits	
LSL	Lower specification limit (if specified by the user).
USL	Upper specification limit (if specified by the user).
CP limit	Lowest capability resp. performance index value that is acceptable. Values, smaller than the CP limit will be marked in red.
Capability indexes	
Arithmetic average	Arithmetic average computed from the data.
Standard deviation	Standard deviation, estimated from the data, σ_C
+/- 3sigma	Lower and upper limit of the $\pm 3\sigma_C$ interval around the arithmetic mean
Z-score	Z-scores correspond to the lower and upper data part
Index	
Cp	Classical capability index value, c_p computed from σ_C
Cpk	Classical capability index value, c_{pk} computed from σ_C
Cpm	Classical capability index value, c_{pm} computed from σ_C
Lower limit	Lower limit of the confidence interval for a particular index.
Upper limit	Upper limit of the confidence interval for a particular index.
Performance limits	
Arithmetic average	Arithmetic mean computed from the data
Standard deviation	Standard deviation computed from the data, σ_C
+/- 3sigma	Lower and upper limit of $\pm 3\sigma_C$ interval around the arithmetic mean
Z-score	Z-scores correspond to the lower and upper data part
Index	
Pp	Classical performance index value, p_p computed from σ_P
Ppk	Classical performance index value, p_{pk} computed from σ_P
Ppm	Classical performance index value, p_{pm} computed from σ_P
Lower limit	Lower limit of the confidence interval for a particular index.
Upper limit	Upper limit of the confidence interval for a particular index.
Probability of exceeding specification limits	Probability that the upper, or lower specification limit, p_{zm} will be exceeded. This number can be understood as the probability that the next measurement will fall above upper or below lower specification limit.
Expected relative frequency of exceeding in %	It can be understood as the expected number of the measurements falling above the upper or below the lower specification limit in the next 100 measurements taken under the same circumstances.
Expected relative frequency	It can be understood as the expected number of the measurements

of exceeding in PPM	falling above the upper or below the lower specification limit in the next 1,000,000 measurements taken under the same circumstances.
Probability of being out of the SL	Probability that any of the specification limits will be exceeded. This number can be understood as the probability that the next measurement will fall beyond any of the specification limits.
Relative frequency of being out of the SL in %	It can be understood as the expected number of the measurements falling beyond any of the specification limits in the next 100 measurements taken under the same circumstances.
Relative frequency of being out of the SL in PPM	It can be understood as the expected number of the measurements falling beyond any of the specification limits in the next 1000000 measurements taken under the same circumstances.
ARL	*Average Run Length* is the expected number of measurements between two consecutive specification limit exceeding.
Cpk for asymmetrically distributed data	
Sample size	Number of the data points used for computations.
Corrected average	Expected value estimate, corrected for the data distribution skewness. When the data are symmetrically distributed, this characteristic is equal to the arithmetic average, see the *Transformation* module.
Target value	User-defined target.
CP limit	Lowest acceptable c_p value. Values lower than this limit will be marked in red.
Specification limits	User-supplied specification limits.
Probability of exceeding	Probability that the upper, or lower specification limit, p_{zm} will be exceeded. This number can be understood as the probability that the next measurement will fall above upper or below lower specification limit.
Expected relative frequency of exceedance in %	It can be understood as the expected number of the measurements falling above the upper or below the lower specification limit in the next 100 measurements taken under the same circumstances.
Expected relative frequency of exceedance in PPM	It can be understood as the expected number of the measurements falling above the upper or below the lower specification limit in the next 1,000,000 measurements taken under the same circumstances.
Probability of being out of the SL	Probability that any of the specification limits will be exceeded. This number can be understood as the probability that the next measurement will fall beyond any of the specification limits.
Relative frequency of being out of the SL in %	It can be understood as the expected number of the measurements falling beyond any of the specification limits in the next 100 measurements taken under the same circumstances.
Relative frequency of being out of the SL in PPM	It can be understood as the expected number of the measurements falling beyond any of the specification limits in the next 1000000 measurements taken under the same circumstances.
ARL	*Average Run Length* is the expected number of measurements between two consecutive specification limit exceedances.
Cpk	Generalized capability index estimate, c_{pk}^*, defined for both symmetrically and asymmetrically distributed data, both two sided and one-sided specification limits. This characteristic should be used whenever the data are not symmetrically distributed.

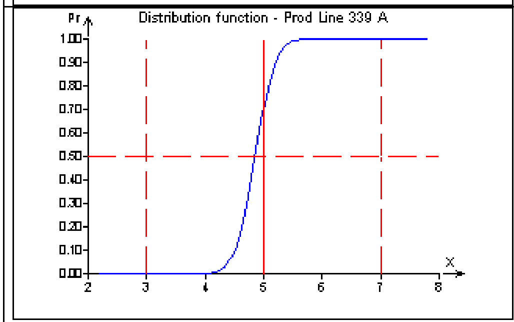
Cpk limits	Lower and upper confidence interval limits for c_{pk}^*.

Graphs

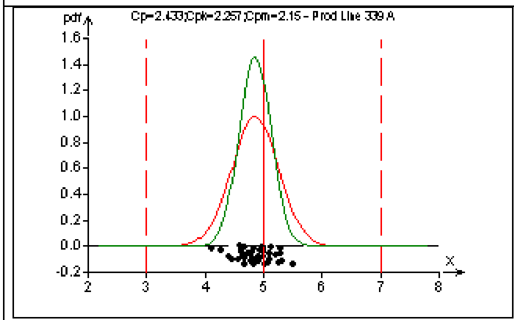
The capability module provides four plot types: three show density and one distribution function. The first three plots, that is: Histogram, Distribution function and a Density are plotted only when the *Classical indexes* selection is checked. The last plot: Density of the transformed data is plotted only when the *General index* selection is checked.



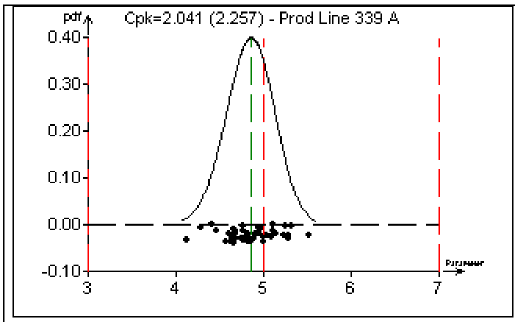
A simple graphical tool, which can be used to compare data against the specification limits. Data are summarized by the histogram, kernel density estimate (red curve) and fitted normal density (Gauss‘ curve) plotted as a green curve. Vertical lines show the target, lower and upper specification limit. The location of the fitted Gauss‘ curve maximum corresponds to the arithmetic mean of the data. It should be as close to the target value as possible.

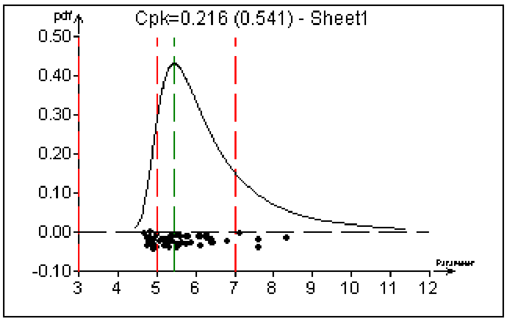


Normal cumulative distribution function (integrated probability density function), based on the parameters estimated form the data (under the normality assumption). Vertical lines correspond to the target and specification limits. Horizontal line corresponds to the probability of 0.5. This line intersects the cumulative distribution curve at the point, whose x-coordinate corresponds to the arithmetic mean of the data. This plot can be used to read probabilities of the normal variable being lower than or equal to a value given by the x-coordinate. The reading can be done more precisely using the *Detail* function in the interactive mode of this plot after a double-click.



Probability density curves. Red curve corresponds to the kernel estimate, Gauss‘ curve with parameters estimated from the data is plotted in green. If these two curves differ markedly, normality of the data is suspicious. Normality should be checked formally then, using an appropriate normality test. This test can be found in the *Elementary statistics*, module. Dashed lines correspond to the specification limits and to the target. Individual data points are plotted below the x-axis. For a better readability, the points are randomly scattered in the vertical direction (small amount of random jitter is added). Estimates of the classical indexes c_p, c_{pk} a c_{pm} are listed in the plot‘s header.





Probability density curve for the transformed data. The meaning of the plot is very similar to the previous one. The density is estimated via the exponential transformation. More details about the transformation can be found in the Transformation module. If the *Asymmetric data distribution* is not checked before the calculations, transformation is not used and normal density curve is plotted. Asymmetry of the data distribution can be checked graphically by inspection of the probability density curve on this plot. The c_{pk}^* index estimate is listed in the plot's header. When both specification limits are given, the classical c_{pk} index is listed in parenthesis as well. If the two values differ markedly, c_{pk}^* should be used. Illustrative examples on the left panel here show curve shapes for symmetric and asymmetric data distributions.