Statistical Methods for Economists

Lecture 11 & 12

Taguchi methods

SILESIAN UNIVERSITY
SCHOOL OF BUSINESS ADMINISTRATION IN KARVINA

David Bartl
Statistical Methods for Economists
INM/BASTE

Outline of the lecture

- Introduction to Taguchi methods
- Taguchi Loss Function
- The expected loss
- Total Quality Costs
- Capability indices

Taguchi methods

Taguchi's methods – statistical methods developed by

Dr. Genichi Taguchi (*1st January 1924 †2nd June 2012)

to improve the quality of manufactured goods.

There are two classes of Taguchi's methods:

- online methods used in the production process
- offline methods used in the pre-production phases

The fundamental concept used of the online methods is the Loss Function.

Taguchi methods: Taguchi Loss Function

Assume that we manufacture some product.

We adopt the following assumptions:

1. The quality of a piece of the product is assessed according to the (numerical) value Y of some characteristic – such as the size, weight; mechanical, chemical, aesthetical or other characteristics.
2. The **Target Value** T of the above characteristics is given.
3. The lack of quality of the piece of the product is characterized by deviations of the characteristic from its target value T.
4. Every deviation of the characteristic from its target value T yields a financial

etc.

Taguchi methods: Taguchi Loss Function

The first three assumptions are usual. The fourth assumption, however, was new:

Until then, it was usual to consider that

- the pieces of product such that the characteristic is within some tolerance interval are considered acceptable (like the binary value "1"),
- the pieces of product such that the characteristic is out of the tolerance interval are considered inacceptable (like the binary value "0"),
- and there are no states in between.

The purpose of Taguchi's Loss Function is to evaluate every loss caused by

Taguchi methods: Taguchi Loss Function: Remarks

Taguchi's Loss Function evaluates every loss caused to the customer

by the supplier due to the (physical) impossibility to keep the absolute precision.

The link between the quality and the Loss Function

is the main contribution to the Quality Engineering;

it also allows to plan the costs.

The Loss Function measures the decreased quality within the tolerance interval.

The Loss Function

- Symmetric N-tolerance
- Non-symmetric N-tolerance
- S-type tolerance (small)
- L-type tolerance (large)

Taguchi Loss Function: Symmetric N-tolerance

Let

- Y denote the actual value of the characteristics of the quality;
  this is a random variable
- T denote the target value of the characteristics of the quality
- d be the tolerance limit
- k be a constant
- A be the loss if the value of Y is out of the tolerance interval $(T-d,\ T+d)$

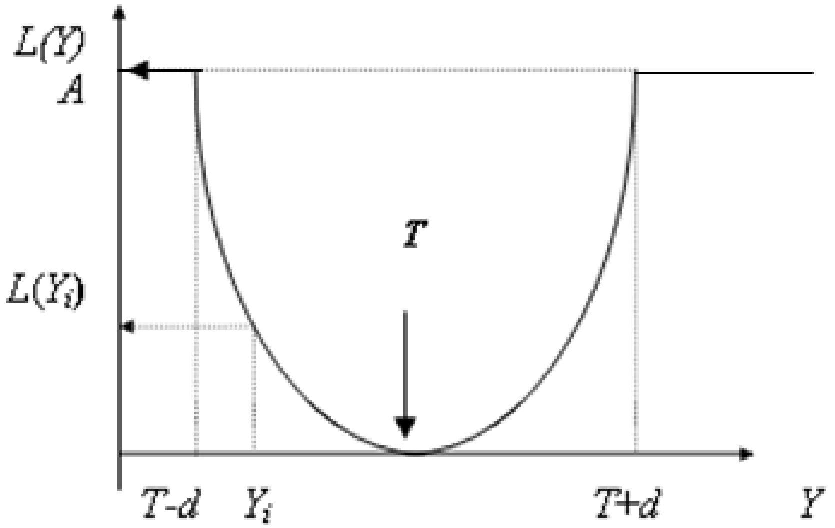
We then write $T \pm d$ and

$$(T-d,\ T+d)$$

Taguchi Loss Function: Symmetric N-tolerance

We then postulate that the Loss Function is continuous and of the form:

$$L(Y) = \begin{cases} k(Y-T)^2 & \text{if} \quad T-d \le Y \le T+d \\ A & \text{if} \quad Y \le T-d \quad \text{or} \quad T+d \le Y \end{cases}$$



Taguchi Loss Function: Symmetric N-tolerance

We then postulate that the Loss Function is continuous and of the form:

$$L(Y) = \begin{cases} k(Y-T)^2 & \text{if} \quad T-d \le Y \le T+d \\ A & \text{if} \quad Y \le T-d \quad \text{or} \quad T+d \le Y \end{cases}$$

It follows:

$$A = k\big((T-d)-T\big)^2$$

$$A = kd^2$$

$$k = \frac{A}{d^2}$$

Taguchi Loss Function: Non-symmetric N-tolerance

Let

- Y denote the actual value of the characteristics of the quality;

  this is a random variable
- T denote the target value of the characteristics of the quality
- k_1, k_2 be constants
- d_1, d_2 be tolerance limits
- A_1, A_2 be the losses outside of the tolerance interval $(T - d_1,\ T + d_2)$

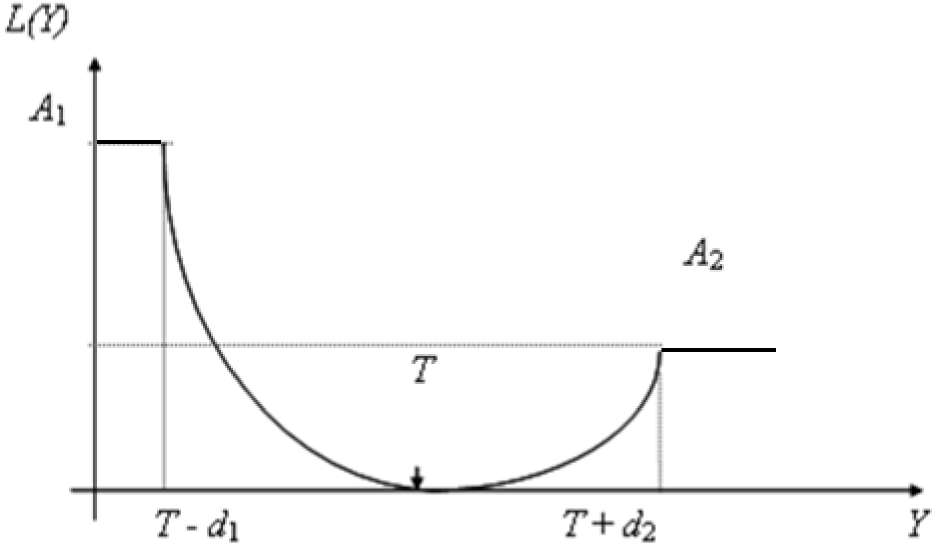
Then the tolerance interval is

$$(T - d_1,\ T + d_2)$$

Taguchi Loss Function: Symmetric N-tolerance

We then postulate that the Loss Function is continuous and of the form:

$$L(Y) = \begin{cases} A_1 & \text{if} \quad Y \le T - d_1 \\ k_1(Y-T)^2 & \text{if} \quad T - d_1 \le Y \le T \\ k_2(Y-T)^2 & \text{if} \quad T \le Y \le T + d_2 \\ A_2 & \text{if} \quad T + d_2 \le Y \end{cases}$$



Taguchi Loss Function: Symmetric N-tolerance

We then postulate that the Loss Function is continuous and of the form:

$$L(Y) = \begin{cases} A_1 & \text{if} \quad Y \le T - d_1 \\ k_1(Y-T)^2 & \text{if} \quad T - d_1 \le Y \le T \\ k_2(Y-T)^2 & \text{if} \quad T \le Y \le T + d_2 \\ A_2 & \text{if} \quad T + d_2 \le Y \end{cases}$$

It follows:

$$A_1 = k_1\big((T-d_1)-T\big)^2 \qquad A_2 = k_2\big((T-d_2)-T\big)^2$$

$$A_1 = k_1 d_1^2 \qquad A_2 = k_2 d_2^2$$

$$k_1 = \frac{A_1}{d_1^2} \qquad k_2 = \frac{A_2}{d_2^2}$$

Taguchi Loss Function: S-type tolerance (small)

Here, the smaller the characteristic Y, the better. The ideal target value is $T = 0$.

An example is the pollution (of air / water / etc.).

Let

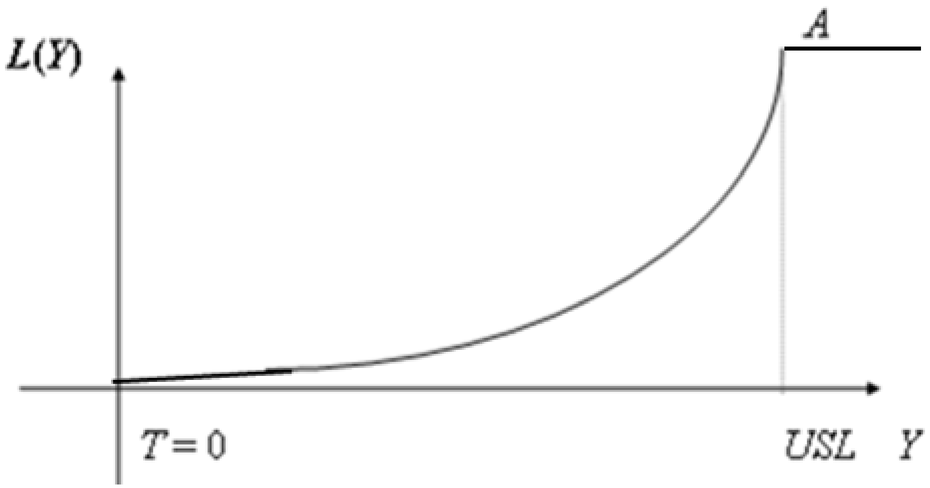
- Y denote the actual value of the characteristics of the quality; this is a random variable
- USL = Upper Specification Limit
- k be a constant
- A be the losses behind the tolerance interval $[0, \mathrm{USL})$

Then the tolerance interval is

Taguchi Loss Function: S-type tolerance (small)

We then postulate that the Loss Function is continuous and of the form:

$$L(Y) = \begin{cases} 0 & \text{if} \quad Y \le 0 \\ kY^2 & \text{if} \quad 0 \le Y \le \text{USL} \\ A & \text{if} \quad \text{USL} \le Y \end{cases}$$



Taguchi Loss Function: S-type tolerance (small)

We then postulate that the Loss Function is continuous and of the form:

$$L(Y) = \begin{cases} 0 & \text{if} \quad Y \le 0 \\ kY^2 & \text{if} \quad 0 \le Y \le \text{USL} \\ A & \text{if} \quad \text{USL} \le Y \end{cases}$$

It follows:

$$A = k \times \text{USL}^2$$

$$k = \frac{A}{\text{USL}^2}$$

Taguchi Loss Function: L-type tolerance (large)

Here, the larger the characteristic Y, the better. The ideal target value is $T = +\infty$.
Let

- Y denote the actual value of the characteristics of the quality; this is a random variable
- LSL = Lower Specification Limit
- k be a constant
- A be the losses forefront the tolerance interval $(\mathrm{LSL}, +\infty)$

Then the tolerance interval is

$$(\mathrm{LSL}, +\infty)$$

Taguchi Loss Function: L-type tolerance (large)

We then postulate that the Loss Function is continuous and of the form:

$$L(Y) = \begin{cases} A & \text{if} \quad Y \leq \mathrm{LSL} \\ k\left(\dfrac{1}{Y}\right)^2 & \text{if} \quad \mathrm{LSL} \leq Y \end{cases}$$

Taguchi Loss Function: L-type tolerance (large)

We then postulate that the Loss Function is continuous and of the form:

$$L(Y) = \begin{cases} A & \text{if} \quad Y \leq \mathrm{LSL} \\ k\left(\frac{1}{Y}\right)^2 & \text{if} \quad \mathrm{LSL} \leq Y \end{cases}$$

It follows:

$$A = k\left(\frac{1}{\mathrm{LSL}}\right)^2$$

$$k = A \times \mathrm{LSL}^2$$

The expected loss

- Symmetric N-tolerance
- Non-symmetric N-tolerance
- S-type tolerance (small)
- L-type tolerance (large)

The expected loss

Recall that the quality of a product is measured by the characteristic Y, which is assumed to be a random variable.

The loss $L(Y)$ depends upon the deviation of the characteristic Y from its target value T.

We, however, are interested to know the **expected loss** $\mathrm{E}[L(Y)]$.

The expected loss: Symmetric N-tolerance

Assume that the Loss Function is of the form

$$L(Y) = \begin{cases} k(Y-T)^2 & \text{if} \quad T-d \le Y \le T+d \\ A & \text{if} \quad Y \le T-d \quad \text{or} \quad T+d \le Y \end{cases}$$

where

$$k = \frac{A}{d^2}$$

The expected loss: Symmetric N-tolerance

Assume for simplicity that the random variable

$$Y \in (T - d,\ t + d) \qquad \text{always}$$

Then:

$$\mathrm{E}[L(Y)] = \mathrm{E}[k(Y - T)^2] = k \times \mathrm{E}[(Y - T)^2]$$

If $\mathrm{E}[Y] = T$, then

$$\mathrm{E}[L(Y)] = k \times \mathrm{Var}(Y) = k\sigma^2$$

The expected loss: Symmetric N-tolerance

Assume for simplicity that the random variable

$$Y \in (T-d,\ t+d) \qquad \text{always}$$

Then:

$$\mathrm{E}[L(Y)] = \mathrm{E}[k(Y-T)^2] = k \times \mathrm{E}[(Y-T)^2]$$

Otherwise, in general:

$$\mathrm{E}[L(Y)] = k \times \mathrm{E}[(Y-\mathrm{E}[Y]+\mathrm{E}[Y]-T)^2] =$$

$$= k \times \mathrm{E}[(Y-\mathrm{E}[Y])^2 + 2(Y-\mathrm{E}[Y])(\mathrm{E}[Y]-T) + (\mathrm{E}[Y]-T)^2] =$$

$$= k \times (\mathrm{Var}(Y) + 2(\mathrm{E}[Y]-\mathrm{E}[Y])(\mathrm{E}[Y]-T) + (\mathrm{E}[Y]-T)^2) =$$

$$= k\sigma^2 + k(\mathrm{E}[Y]-T)^2$$

The expected loss: The estimates of E[Y] and Var(Y)

<u>Remark:</u>

The expected value $\mathrm{E}[Y]$ and the variance $\sigma^2 = \mathrm{Var}(Y)$ are unknown usually.

We then make experiments and use the

- the sample mean and
- the sample-population variance

to estimate these quantities.

The expected loss: The estimates of E[Y] and Var(Y)

Let us have a sample

$$y_1, y_2, \dots, y_n$$

of the observations of the characteristic Y, which is a random variable.

The sample mean – estimate of $\mathrm{E}[Y]$:

$$\mathrm{E}[Y] \approx \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

The sample-population variance – estimate of $\mathrm{Var}(Y)$:

$$\sigma^2 \approx s^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2$$

The expected loss: The estimates of E[*Y*] and Var(*Y*)

<u>Another approach:</u>

- Assume that the random variable Y is **<u>normally</u>** distributed.
- This is the case, e.g., if the producer aims to achieve the target value T.
- Then the probability that $Y \in (\mathrm{E}[Y] - 3\sigma, \mathrm{E}[Y] + 3\sigma)$ is about 99.8%.
- Assuming that $T = \mathrm{E}[Y]$ and that all the production is within the tolerance interval $(T - d,\ T + d)$, we can estimate that

$$d \approx 3\sigma$$

$$\sigma \approx \frac{d}{3}$$

The expected loss: The estimates of E[Y] and Var(Y)

Yet another approach:

- Assume that the random variable Y is **uniformly** distributed.
- This is the case, e.g., if the producer aims to have all the production within the tolerance interval $(T-d,\ T+d)$ only; that is, does not aim to achieve the target value T.
- Assuming that $T=\mathrm{E}[Y]$, it follows

$$\sigma^2 \approx \frac{(2d)^2}{12} = \frac{4d^2}{12} = \frac{d^2}{3}$$

$$\sigma \approx \frac{d}{\sqrt{3}}$$

The expected loss: The cost of the quality

Recall that we have calculated:

$$\mathrm{E}[L(Y)] = k \times \mathrm{Var}(Y) = k\sigma^2 = \frac{A}{d^2}\sigma^2$$

Notice that

- the parameters A and d are given by the designer (design engineer), but
- the parameter σ^2 is given by the worker.

It follows that we wish to have d as large as possible and to minimize A and σ^2.

The expected loss: The cost of the quality

Example: Assume that

$$A = 40$$

$$d = 4$$

$$\sigma = 1.33$$

Then the expected loss is

$$\mathrm{E}[L(Y)] = \frac{A}{d^2}\sigma^2 = \frac{40}{4^2} \times 1.33^2 = 4.42225$$

monetary units per a piece of production.

The expected loss: The cost of the quality

Assume further that the variance $\sigma^2 = 1.33^2 = 1.7689$ can be decreased by 30 % to $0.70 \times \sigma^2 = 0.7 \times 1.7689 = 1.23823$, but the cost of this decrease is 0.50 monetary units per a piece of production. Then the total loss is:

$$\mathrm{E}[L_{\mathbf{new}}(Y)] + 0.50 = \frac{A}{d^2}\sigma^2_{\mathbf{new}} + 0.50 = \frac{40}{4^2} \times 1.23823 + 0.50 = 3.595575$$

which is a result better than the previous one (4.42225) with the original $\sigma^2 = 1.33^2$.

To conclude, this improvement can be recommended.

Total
Quality Costs

- Quality Cost Monitoring
- Taguchi's approach

  to Quality Cost Monitoring

Quality Cost Monitoring

We concentrate on Quality Cost Monitoring on the side of the manufacturer.

There are several approaches to Quality Cost Monitoring, including:

- monitoring based on PAF (Prevention, Appraisal, Failure) models
- monitoring based on process models
- monitoring based on Taguchi's approach

We here concentrate in Taguchi's approach.

Taguchi's approach to Quality Cost Monitoring

There are two main cases in Taguchi's approach to Quality Cost Monitoring:

- checking all (100%) pieces of production
- checking one out of every n pieces of production;
  - checking qualitative (non-measurable) characteristics of quality – attributes

Checking all (100%) pieces of production

Checking all (100%) pieces of production

Let us have a sample

$$y_1, y_2, \dots, y_n$$

of the observations of the characteristic Y, which is a random variable.

When all (100%) pieces of production are checked,

the total cost of the quality control is:

$$L = \frac{Q}{R} + \frac{A}{d^2} s_0^2$$

Checking all (100%) pieces of production

The total cost of the quality control is:

$$L = \frac{Q}{R} + \frac{A}{d^2} s_0^2$$

where

- Q are the annual costs for 100% quality control
- R is the annual production = the number of the pieces of production per year
- d is the tolerance (see above)
- A is the loss when the tolerance is exceeded (see above)

$$s_0^2 = \frac{1}{n-1}\left[(y_2 - y_1)^2 + (y_3 - y_2)^2 + \cdots + (y_n - y_{n-1})^2\right] = \frac{1}{n-1}\sum_{i=1}^{n-1}(y_{i+1} - y_i)^2$$

Checking all (100%) pieces of production

Remark: The total variance

$$s_0^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(y_{i+1} - y_i)^2$$

consists of:

- s_p^2 = the production variance
- s_m^2 = the measurement variance

and it holds:

$$s_0^2 = s_\mathrm{p}^2 + s_\mathrm{m}^2$$

Checking one out of every *n* pieces of production

Checking one out of every *n* pieces of production

When n pieces of production are made between two quality checks, the total cost of the quality control is:

$$L = \frac{B}{n} + \frac{C}{u} + \frac{A}{d^2}\frac{D^2}{3} + \frac{A}{d^2}\frac{D^2}{u}\left(\frac{n+1}{2} + z\right) + \frac{A}{d^2}s_{\mathrm{m}}^2$$

where

- d is the tolerance (see above)
- A is the loss when the tolerance is exceeded (see above)
- B is cost of the check of one piece of production
- n is the number of pieces produced between two quality checks

Checking one out of every n pieces of production

When n pieces of production are made between two quality checks,

the total cost of the quality control is:

$$L = \frac{B}{n} + \frac{C}{u} + \frac{A}{d^2}\frac{D^2}{3} + \frac{A}{d^2}\frac{D^2}{u}\left(\frac{n+1}{2} + z\right) + \frac{A}{d^2}s_{\mathrm{m}}^2$$

where

- C are the repair costs of the production plant
- u is the average number of pieces produced between two failures of the production plant

Checking one out of every *n* pieces of production

The total cost of the quality control is:

$$L = \frac{B}{n} + \frac{C}{u} + \frac{A}{d^2}\frac{D^2}{3} + \frac{A}{d^2}\frac{D^2}{u}\left(\frac{n+1}{2} + z\right) + \frac{A}{d^2}s_{\mathrm{m}}^2$$

where

- d is the tolerance is defined by the customer (see above) – the tolerance within which the product is satisfactory in terms of its quality
- D is the tolerance defined by the producer; it holds

$$D \leq d$$

- z is the number of pieces of production made during the check

Checking one out of every *n* pieces of production

$$L = \frac{B}{n} + \frac{C}{u} + \frac{A}{d^2}\frac{D^2}{3} + \frac{A}{d^2}\frac{D^2}{u}\left(\frac{n+1}{2} + z\right) + \frac{A}{d^2}s_{\mathrm{m}}^2$$

where

- $\frac{B}{n}$ are the control costs per a piece of production
- $\frac{C}{u}$ are the repair costs per a piece of production
- $\frac{A}{d^2}\frac{D^2}{3}$ losses due to imprecise production
- $\frac{A}{d^2}\frac{D^2}{u}\left(\frac{n+1}{2} + z\right)$ losses due to production of defective pieces

Checking one out of every *n* pieces of production

$$L = \frac{B}{n} + \frac{C}{u} + \frac{A}{d^2}\frac{D^2}{3} + \frac{A}{d^2}\frac{D^2}{u}\left(\frac{n+1}{2} + z\right) + \frac{A}{d^2}s_{\mathrm{m}}^2$$

where

- $\frac{A}{d^2}\frac{D^2}{3}$ losses due to imprecise production
- $\frac{A}{d^2}\frac{D^2}{u}\left(\frac{n+1}{2} + z\right)$ losses due to production of defective pieces
- $\frac{A}{d^2}s_{\mathrm{m}}^2$ losses due to imprecise measurements

Remarks

- Dr. Genichi Taguchi **proposed** the above formula based on his long experience.
- That is, the formula expresses Dr. Taguchi's experience in mathematical terms.
- That is, there is no mathematical justification of the above formula.
- Notice, however, that the last three terms are mathematically justified because they are based on the loss function $L(Y)$.
- By letting

$$\frac{\partial L}{\partial n} = 0 \qquad \text{or} \qquad \frac{\partial L}{\partial D} = 0$$

we obtain:

Remarks

What is the optimal inspection interval?

$$n^* = \sqrt{\frac{2uB}{A}\frac{d}{D}}$$

What is the optimal tolerance defined by the producer?

$$D^* = \sqrt[4]{\frac{3CD_0^2 d^2}{Au}} \qquad (?)$$

where D_0 is the prior (previous) tolerance defined by the producer.

Remarks

The <u>new</u> average number of pieces produced between two failures of the production plant:

$$u = \frac{D^{*2}}{D_0^2} u_0 \qquad (?)$$

where

- D_0 is the prior (previous) tolerance defined by the producer
- u_0 is the prior (previous) number of pieces produced between two failures of the production plant

Checking attributes

When checking qualitative (non-measurable) characteristics of quality – attributes, the total cost of the quality control is:

$$L = \frac{B}{n} + \frac{C}{u} + \frac{A}{u}\left(\frac{n+1}{2} + z\right)$$

where

- A is the loss when the piece of production is defective (see above)
- B is cost of the check of <u>one</u> piece of production
- C are the repair costs of the production plant

Checking attributes

When checking qualitative (non-measurable) characteristics of quality – attributes, the total cost of the quality control is:

$$L = \frac{B}{n} + \frac{C}{u} + \frac{A}{u}\left(\frac{n+1}{2} + z\right)$$

where

- n is the number of pieces produced between two quality checks
- u is the average number of pieces produced between two failures of the production plant
- z is the number of pieces of production made during the check

Capability indices

- Capability index C_p
- Capability index C_{pK}

Capability indices

Recall the notation:

- Y is the (numerical) value of some characteristic – a random variable – by using which we measure the quality
- T is the Target value
- USL is the Upper Specification Limit
- LSL is the Lower Specification Limit

We have:

$$\mathrm{LSL} \leq T \leq \mathrm{USL}$$

Capability indices

Assume that the random variable Y is normally distributed:

$$Y \sim \mathcal{N}(\mu, \sigma^2) \qquad \text{for some} \quad \mu \in \mathbb{R} \quad \text{and} \quad \sigma^2 > 0$$

<u>Remarks:</u>

- use a large sample
- use a statistical test for normality
- remove outlying values from the sample
- μ is approximated by the sample mean $\bar{y} = \sum_{i=1}^{n} y_i / n$
- σ^2 is approximated by the sample variance $s^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 / (n-1)$

Capability indices

If the characteristic is qualitative (an attribute), the capability index is the ratio

$$C = \frac{\text{the number of good pieces of production}}{\text{the total number of pieces of production}}$$

In general, if the characteristic Y is quantitative, the capability index is

$$C = \frac{\text{the length of the tolerance interval, where all the values should be}}{\text{the length of the interval where the observed values actually are}}$$

Capability index C_p

Recall the six-sigma rule for the normally distributed random variable

$$Y \sim \mathcal{N}(\mu, \sigma^2)$$

Under the assumption that the target value is $T = \mu$,

the capability index C_p is defined:

$$C_p = \frac{USL - LSL}{6\sigma}$$

or

$$C_p = \frac{USL - LSL}{6s}$$

Capability index C_{pK}

Assuming that $Y \sim \mathcal{N}(\mu, \sigma^2)$, the capability index C_{pK} is defined:

$$C_{pK} = \min\left\{\frac{\mathrm{USL} - \mu}{3\sigma}, \frac{\mu - \mathrm{LSL}}{3\sigma}\right\}$$