Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 75 (2018) 385–390

www.elsevier.com/locate/procedia

15th CIRP Conference on Computer Aided Tolerancing – CIRP CAT 2018

Study of measurement process capability with non-normal data distributions

Gianfranco Genta*, Maurizio Galetto

Department of Management and Production Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

* Corresponding author. Tel.: +390110907257; fax: +390110907299. E-mail address: gianfranco.genta@polito.it

Abstract

The study of capability of measurement processes is becoming a common practice within Quality Management Systems. Actually, "Gauge R&R study" techniques are currently largely widespread. In general, most of the tools described in the scientific literature, prescribed by the international standards and/or implemented in practical applications are based on the normality assumption of data distributions. This work discusses some alternative approaches for "gauge R&R study" when non-normal data distributions are involved. In order to clarify the impact of this problem in practical applications, the techniques considered in the present paper are analyzed through computer simulation of real case studies in the field of dimensional tolerancing.

Peer-review under responsibility of the Scientific Committee of the 15th CIRP Conference on Computer Aided Tolerancing - CIRP CAT 2018.

Keywords: measurement process capability; gauge R&R study; non-normal distribution

1. Introduction

In the TQM (Total Quality Management) framework, pursuing quality means developing processes characterized by high efficiency and small errors, and carrying out cost reduction through the elimination of waste and unnecessary activities [1]. Consequently, the achieved product quality level is continuously improved with the aim of meeting the end customer needs. Determining the capability of the measurement process is an important aspect of many quality and process improvement activities. In general, in any operation involving measurements, part of the observed variability is due to the measured object(s) and part is due to the implemented measurement process [2]. ISO standards define a measurement process as a set of operations to determine the value of a quantity [3]. Such a process is typically based on the use of a measurement system, which consists in (at least) an instrument (or gauge) with the relevant measurement procedure, and is conditioned by a series of other elements, such as the operator(s), the secondary procedure(s) and the conditions in which the system is used. There may also be other factors that impact measurement system performance and, hence, the measurement process variability, such as setup or calibration operations.

Most of the studies on measurement system and process capability concerns:

- the evaluation of the percentage of total variability which is due to the system (or, more in detail, to the instrument/gauge);
- the separation of the measurement process variability components;
- the evaluation of system and process capability respectively (i.e. whether they are suitable for the considered application).

"Gauge R&R studies" [2] aim at evaluating the repeatability and reproducibility of a measurement process (and system) [4]. These studies are widely used in the Six Sigma DMAIC (Define, Measure, Analyze, Improve, Control) approach for continuous improvement of production processes [5].

However, a fundamental hypothesis on which "gauge R&R studies" are based, i.e. the normality of measurement processes, is often underestimated. It has been observed that,

2212-8271 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Scientific Committee of the 15th CIRP Conference on Computer Aided Tolerancing - CIRP CAT 2018.
10.1016/j.procir.2018.04.040

in several industrial cases, both production process and measurement process behaviors deviate from normal distribution, showing significant elements of asymmetry and kurtosis. In these situations, using an approach based on the hypothesis data normality can lead to gross errors.

For example, when measuring the (internal) diameter of a cylindrical hole, the measurements distribution may have a negative asymmetry due to the physical constraint for which it is not possible to measure beyond a certain limit (i.e., the actual diameter of the hole). In the same way, when measuring the (external) diameter of a cylindrical pin, a measurements distribution with a positive asymmetry is expected. Sometimes, it may also occur that the measurements distribution is given by a mixture of two or more normal distributions. This is typically due to discontinuities in the nature of the experimental data which may be related both to the measurement and the manufacturing. Let us consider, e.g., in the former case, measurements obtained with different instruments or by different operators, and, in the latter case, measurements performed on products originated from different production lines [6].

Therefore, an alternative approach for evaluating the capability of a measurement process in case of non-normal distributions is needed, since evaluation errors may bring to the rejection of suitable instruments and/or to the acceptance of inappropriate ones, determining the improper exclusion of conforming parts and/or the improper acceptance of defective parts.

2. Standards for capability study in process management

In order to evaluate the performance of a process, Juran combined the process parameters with product specifications and introduced process capability indices [7]. These give an indication of how much the process is “capable” to turn out products which satisfy the prescribed specifications and tolerances. This important property enabled their application in the industrial field for the purpose of continuous improvement. In order to standardize the definitions of the indices, the ISO 22514 family of standards, made up of 8 parts, was introduced.

In Section 2.1 of the present paper, ISO 22514-1 is analyzed [8]. This standard illustrates the main concepts concerning the capability and performance of industrial processes. In Section 2.2, ISO 22514-7, which deals specifically with capability of the measurement processes, is described [3].

2.1. General principles and concepts

ISO 22514-1 describes the fundamental principles for the analysis of the capability and performance of manufacturing processes. It provides guidance on the study of capability when it is necessary to determine whether the output of an industrial process is acceptable according to certain previously specified criteria. This represents a common practice in quality control. These methods are very versatile and are applied in many industrial situations such as:

- assessing whether the characteristics of a product meet specific requirements;
- testing the batches of a supplier during acceptance;
- performing internal or external audits on product compliance;
- analyzing and evaluating the current production process for the continuous improvement.

Capability and performance indices describe the behavior of distribution tails of product characteristics. The variety of distribution families implies different behaviors and the estimated indices will depend significantly on the selected distribution. Therefore, an appropriate distribution has to be selected with great care. The first step is to determine the sample size and sampling rate. The sample size must be defined according to the desired level of confidence, accuracy and type of process under investigation. Generally a sample of about 100 observations is considered [8]. However, in case of analysis of measurement process capability, the sample size should be less than 100 [3]. If it is suspected that process data are not normally distributed, it is essential to increase the sample size by 50% with the aim of identifying the correct distribution [8].

In collecting measured data, it is important to define their reliability. The uncertainty of the measuring process is always present when the characteristics of interest are measured. This uncertainty must be evaluated referring to the product specification limits and the variation of the process. Therefore, the adopted measuring instruments/gauges should have specific metrological characteristics for the defined measurement activity [3]. As explained in the next subsection, there is a procedure to evaluate whether the uncertainty of the measuring process is acceptable.

2.2. Capability of measurement processes

ISO 22514-7 defines a procedure for validating measurement systems and processes. The aim of this procedure is determining whether the analyzed measurement process meets the requirements of a specific measurement activity based on acceptance criteria. These criteria are established through the use of capability indices (C_{MS} and C_{MP}) and performance ratios (Q_{MS} and Q_{MP}) [3], which will be defined later.

ISO 22514-7 follows the GUM (Guide to the Expression of Uncertainty in Measurement) approach [9]. It establishes a simplified procedure for defining and assessing the uncertainty related to the system and the measurement process. This procedure has been developed for unidimensional measurement processes, in which the uncertainties of method and technical specifications are lower than implementation uncertainty.

ISO 22514-7 defines the variability components which may influence the measuring system and the measurement process. In particular, the standard proposes a procedure based on ANOVA (Analysis of Variance) for estimating them, i.e. the so-called “gauge R&R study” (described in Section 3.1). Formulas for calculating the combined standard uncertainty of the measuring system u_{MS} and measurement process u_{MP} are also defined [3]. Moreover, performance ratios and capability indices are introduced. In detail, the performance ratio of the measuring system is:

$$Q_{MS} = \frac{2 \cdot U_{MS}}{U - L} \cdot 100\,(\%) \tag{1}$$

while the performance ratio of the measurement process is:

$$Q_{MP} = \frac{2 \cdot U_{MP}}{U - L} \cdot 100\,(\%) \tag{2}$$

where U and L are respectively the upper and lower specification limit, and U_{MS} and U_{MP} are the measuring system and measuring process expanded uncertainties given at 95% confidence level, i.e. $U_{MS} = 2 \cdot u_{MS}$ and $U_{MP} = 2 \cdot u_{MP}$.

However, when the measurement process is under statistical control, the expanded uncertainties may also be expressed at 99.73% confidence level, i.e. $U_{MS} = 3 \cdot u_{MS}$ and $U_{MP} = 3 \cdot u_{MP}$.

Then, the capability index of the measuring system is:

$$C_{MS} = \frac{0.3 \cdot (U - L)}{6 \cdot u_{MS}} \tag{3}$$

while the capability index of the measurement process is:

$$C_{MP} = \frac{0.3 \cdot (U - L)}{3 \cdot u_{MP}} \tag{4}$$

ISO 22514-7 recommends that Q_{MS} and Q_{MP} do not exceed respectively 15% and 30%, and both C_{MS} and C_{MP} exceed 1.33.

All these definitions are given under the assumption of normal distributions of measurements. There are, therefore, criticalities in applying the "gauge R&R study" in presence of non-normal distributions. These problems will be highlighted in Section 3.1.

3. Evaluation of measurement process capability

3.1. Gauge R&R study

A "gauge R&R study" is a designed experiment applied to several parts (or workpieces) and inspectors (or operators). Specifically, it is a factorial experiment, so-called because each inspector (operator) measures all of the parts several times. The ANOVA can be applied in order to analyze the data from a "gauge R&R experiment" and to estimate the appropriate components of the measurement process variability [2].

If there are p randomly selected parts and o randomly selected operators and each operator measures every part n times, then the measurements (i = part, j = operator, k = measurement) could be represented by the following model [2]:

$$y_{ijk} = \mu + P_i + O_j + (PO)_{ij} + \varepsilon_{ijk} \qquad \begin{cases} i = 1, \ldots, p \\ j = 1, \ldots, o \\ k = 1, \ldots, n \end{cases} \tag{5}$$

where the model parameters P_i, O_j, $(PO)_{ij}$, and ε_{ijk} are all independent random variables that represent the effects of parts, operators, the interaction or joint effects of parts and operators, and random error . This is the reference model for a random effects model ANOVA. It is also sometimes called the standard model for a "gauge R&R experiment" [2]. We assume that the random variables P_i, O_j, $(PO)_{ij}$, and ε_{ijk} are normally distributed with mean zero and standard deviations σ_P, σ_O, σ_{PO} and σ. Therefore, the variance of any observation is:

$$\sigma^2_{total} = \sigma^2_P + \sigma^2_O + \sigma^2_{PO} + \sigma^2 \tag{6}$$

In particular, σ^2 is the repeatability variance component, while the reproducibility component is the sum of the operator and the part × operator variance components, i.e.:

$$\sigma^2_{reproducibility} = \sigma^2_O + \sigma^2_{PO} \tag{7}$$

Therefore:

$$\sigma^2_{gauge} = \sigma^2_{repeatability} + \sigma^2_{reproducibility} \tag{8}$$

Moreover, it results:

$$\sigma^2_{total} = \sigma^2_P + \sigma^2_{gauge} \tag{9}$$

It means that the analyzed variability concerns both the production process and measurement process.

"Gauge R&R analysis" allows to compare the estimate of gauge capability to the width of the specifications or the tolerance interval (U-L) relevant to the measured workpiece. Let us consider e.g. the precision-to-tolerance ratio:

$$\frac{P}{T} = \frac{k \cdot \sigma_{gauge}}{U - L} \tag{10}$$

Typically, it is assumed $k = 6$, i.e. the number of standard deviations between the usual natural tolerance limits of a normal population (at 99.73% confidence level) [2].

However, since the previous definitions are consistent only with the assumption of normal distribution of measured data, alternative approaches for calculating the performance ratio and the capability index of the measurement process in the case of non-normal distributions are needed. Regarding the production process, possible departures from normality are taken into account by ISO 22514-4 [10]. In the two following subsections, the two mainly applied approaches in these situations for production processes (i.e. data transformation and best fit) are described. Then, in Section 4.3, these two approaches are adapted for evaluating also measurement process capability in case of non-normality.

3.2. Transformation approach

In some circumstances, some transformations may be applied to the original measured data in order make their distribution more similar to a normal one. In particular, Johnson transformations can often bring to "acceptably normal" distributions [2]. Johnson developed a system of distributions basing on the method of moments [11]. The general formula of the transformation is:

$$z = \gamma + \eta \cdot k(x; \lambda, \varepsilon) \tag{11}$$

where z is a standard normal variable, γ and η are shape parameters, λ is a scale parameter, ε is a location parameter, and k is an arbitrary function. This function may assume three alternative expressions, leading to as many Johnson distribution families [11], i.e. S_U (unbounded) with:

$$k(x; \lambda, \varepsilon) = \ln\left(\frac{x-\varepsilon}{\lambda+\varepsilon-x}\right) \tag{12}$$

S_B (bounded on ε, ε+λ) with:

$$k(x; \lambda, \varepsilon) = \sinh^{-1}\left(\frac{x-\varepsilon}{\lambda}\right) \tag{13}$$

and S_L (log-normal) with:

$$k(x; \lambda, \varepsilon) = \ln\left(\frac{x-\varepsilon}{\lambda}\right) \tag{14}$$

However, the transformations may bring to complex calculations and to misinterpretations caused by the transformed data. For example, referring to transformed data, the evaluation of bias and dispersion may be hard to deal with.

3.3. Best fit approach

According to ISO 22514-4, the best fit method requires firstly to identify an appropriate family of distributions which may be associated to the experimental data, secondly to estimate the parameters of the distribution that best explain the data. Note that the standard does not recommend any specific procedure. It just requires using "some efficient estimation method" [10]. Finally, the quantiles of interest for the computation of the capability parameters are expressed in terms of the estimated parameters.

4. Simulative study

The presented methods for evaluating the measurement process capability have to be compared in order to define which is most appropriate. In the study performed in the present paper, the convergence to the theoretical indices was evaluated by means of a simulative study.

4.1. Data generation

The first step for the simulation was the definition of the statistical model, i.e.:

$$y_{ijk} = \mu + P_i + O_j + \varepsilon_{ijk} \qquad \begin{cases} i = 1,\ldots,p \\ j = 1,\ldots,o \\ k = 1,\ldots,n \end{cases} \tag{15}$$

This model is a simplified version of the one proposed in Equation (5), i.e. without considering interactions, which are often non-significant in application cases. In order to test the "gauge R&R study" with non-normal data, possible departures from normality of random variables P_i, O_j and ε_{ijk} were tested and are examined in Section 4.2.

An actual measurement of an external diameter of a mechanical component was considered as reference case study. According to prior experimental experience, the following parameters were known:

- expected value of the diameter μ=36.000 mm;
- standard deviation of the production process σ_P= 0.050 mm;
- standard deviation of reproducibility σ_O = 0.003 mm;
- standard deviation of repeatability σ= 0.001 mm.

The tolerance interval is (36.00 ± 0.01) mm.

In the simulation, a data set of 1000 observations was generated. It was supposed that there were o = 5 operators performing n = 4 replicated diameter measurements on p = 50 mechanical components. The statistical software *Minitab* was exploited for the data generation.

4.2. Gauge R&R approach

Three different combinations of the three involved random variables were examined in the simulation:

- case I: P_i, O_j and ε_{ijk} normal;
- case II: P_i and O_j normal, and ε_{ijk} lognormal;
- case III: P_i and O_j normal, and ε_{ijk} mixture of two normal distributions.

Examined conditions, i.e. effects of parts (P_i) and operators (O_j) normally distributed with error (ε_{ijk}) deviating from normality, are common in practical applications. Deviations from normality of the production process, i.e. effects of parts, may often be limited by controlling the production. In the same way, non-normality of the operators may be corrected, e.g., by training courses. Nevertheless, even with controlled productions and skilled operators, non-normality effects may occur in measurements, thus bringing to a non-normal error ε. This situation is inherent of measurements, but it is not adequately dealt with by standards and literature.

The three different distributions (i.e. normal, lognormal and mixture) considered for the repeatability of measurement process allow accounting for many experimental situations.

According to the simulation hypothesis made in Section 4.1, the random variables P_i, O_j and ε_{ijk} were assumed with null averages, and standard deviations respectively σ_P, σ_O and σ. Therefore, in the second and third cases, lognormal and mixture distributions were properly standardized. Then, by applying "gauge R&R study" through the "Quality Tools" of *Minitab*, the performance ratios and the capability indices shown in Table I were obtained.

Table 1. "Gauge R&R analysis" with simulated data.

Performance ratios and capability indices	Case I	Case II	Case III
Q_{MP} at 95%	36.2 %	35.8 %	34.5 %
Q_{MP} at 99.73%	54.3 %	53.7 %	51.8 %
C_{MP}	1.11	1.12	1.16

Obtained results showed that the "gauge R&R study" do not enable discriminate the three analyzed combinations of random variables, i.e. it does not identify non-normal distributions of cases II and III. For this reason, alternative approaches, developed in the next subsection, were considered.

4.3. Alternative approaches

In order to correctly evaluate measurement process capability in case of non-normality, application of transformation and best fit approaches (described in Sections 3.2 and 3.3) adequately adapted were tested. To this aim, the single components of the simulated data according to Equation (15) required to be analyzed separately. In particular, the analysis was focused on the repeatability of measurement process. The following model was exploited:

$$\varepsilon_{ijk} = y_{ijk} - (\mu + P_i) - (\mu + O_j) + \mu \quad (16)$$

As mentioned in Section 3.2, data non-normally distributed can be transformed in "acceptably" normal data by Johnson transformations and then analyzed according to standard approaches. As well as "gauge R&R study", the Johnson transformations may be applied through the "Quality Tools" of *Minitab*. This approach was implemented both in case of random variable ε_{ijk} distributed according a lognormal distribution (case II) and mixture distribution (case III). In case of lognormal distribution an appropriate Johnson transformation was identified, while in case of mixture distribution no Johnson transformation was found. Therefore, in the latter situation, it was not possible to calculate performance ratios and capability indices.

Alternatively, the best-fit approach, described in Section 3.3, may be implemented. In case of lognormal distribution (case II), the best-fit distribution was identified through the statistical software *Easyfit*. This software selects the parent distribution out of a comprehensive list and assesses its adequacy to describe the sample at hand by means of goodness of fit tests. However, in case of mixture distribution (case III), it was not possible to exploit *Easyfit*, since it does not include this kind of distributions. Therefore, the authors developed specific *MATLAB* routines to estimate mixture parameters, and then the quantiles needed for calculating performance ratios and capability indices [6].

4.4. Results comparison

The presented methods for the calculation of the index C_{MP} limited to the repeatability of the measurement process were compared. First of all, the simulation was extended by considering smaller samples sizes, i.e. 50, 100 and 500 observations. In fact, in the experimental practice, it is often not feasible to perform 1000 measurements. The results are shown in Fig. 1 for the case of lognormal distribution and in Fig. 2 for the case of mixture distribution. The estimates of C_{MP} obtained with the simulated data were compared with the theoretical (correct) values obtained using the exact values of parameters of the distributions used for the simulations.

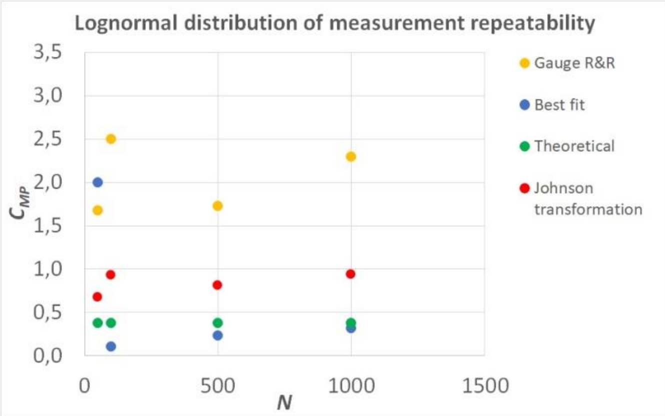

Fig. 1. Indices C_{MP} relevant to repeatability in case of lognormal distribution versus sample sizes *N*.

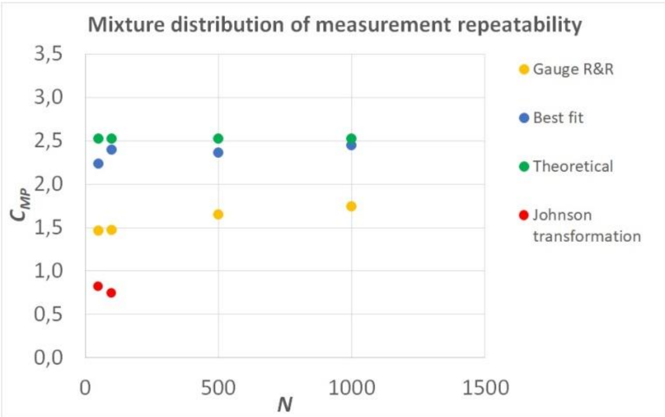

Fig. 2. Indices C_{MP} relevant to repeatability in case of mixture distribution versus sample sizes *N*.

Problems linked to "gauge R&R study" are evident, since it overestimates C_{MP} in case of lognormal distribution and underestimates C_{MP} in case of mixture distribution. Johnson transformation approach has also problems of overestimation in case of lognormal distribution, besides the estimates do not improve when the sample size is increased. Instead, the transformation approach does not work in case of mixture of distributions. In fact, the C_{MP} values obtained for samples sizes 50 and 100 have no meaning, while no transformation was found in case of sample sizes 500 and 1000. On the contrary, best-fit approach, in both cases (lognormal and mixture distribution), brings to estimates of C_{MP} close to the theoretical indices. It is also worth noting that estimates improve when the sample size is increased.

Previous observations concern point estimates of the indices C_{MP}. These results were confirmed by performing further simulations for each of the four considered sample sizes. Fig. 3 shows the results of this simulation for the case of lognormal distribution. The transformation approach was no more considered for the abovementioned reasons. Best fit approach brings to estimates of C_{MP} significantly closer to the theoretical values with respect to “gauge R&R study”.

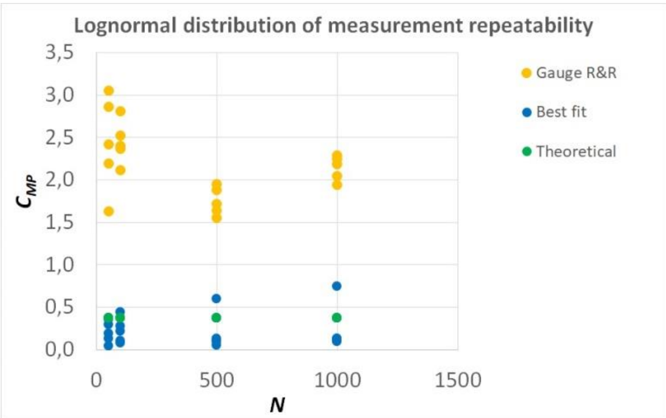

Fig. 3. Indices C_{MP} relevant to repeatability in case of lognormal distribution (five simulations for each approach) versus sample sizes N.

5. Conclusions

In Quality Management Systems, the evaluation of measurement process capability is required. To this aim, “gauge R&R studies” are the most used techniques. These methods are based on the assumption of normality of considered data distributions. Therefore, the application of these methods in presence of departures from normality may bring to significant errors. Experimental data of a “gauge R&R study” include a component related to the production process and another component related to the measurement process. Separating these two components allows discovering the origins of the non-normality. If the component related to the measurement process is non-normal, it is needed to refer to alternative approaches for evaluating measurement process capability. The application of Johnson transformation and best fit approaches, developed for the production process, was tested. The former approach may lead to biased results (e.g. in case of lognormal distributions) and is not applicable in several cases (e.g. in case of mixture distributions), while the latter approach generally brings to accurate results and allows waste reduction. However, the best fit approaches may require high sample sizes to correctly identify the experimental data distribution. If this is not feasible, in a first approximation, the standard approach of “gauge R&R study” (based on the assumption of normality) results to be better than the transformation approach.

Future work will focus on the development of a specific approach for measurement processes, characterized by non-normal distributions, in which the best fit method must be applied with a small number of data. Furthermore, the proposed procedures will be tested on the field with specific experimental case studies.

References

[1] Besterfield DH, Besterfield-Michna C, Besterfield G, Besterfield-Sacre M. Total Quality Management. 3rd ed. Upper Saddle River: Prentice Hall; 2002.
[2] Montgomery DC. Introduction to statistical quality control. 6th ed. Hoboken: John Wiley & Sons; 2008.
[3] ISO 22514-7:2012. Statistical methods in process management - Capability and performance - Capability of measurement processes.
[4] JCGM 200:2012. International vocabulary of metrology – Basic and general concepts and associated terms (VIM). 3rd ed.
[5] Hahn GJ, Hill WJ, Hoerl RW, Zinkgraf SA. The impact of Six Sigma improvement—A glimpse into the future of statistics. The American Statistician 1999; 53(3):208-215.
[6] Aggogeri F, Barbato G, Genta G, Levi R. Statistical modeling of industrial process parameters. Procedia CIRP 2015; 33:203-208.
[7] Juran JM, Gryna FM. Juran's quality control handbook. 4th ed. New York: McGraw-Hill; 1988.
[8] ISO 22514-1:2014. Statistical methods in process management - Capability and performance - General principles and concepts. 2nd ed.
[9] JCGM 100:2008. Evaluation of measurement data - Guide to the expression of uncertainty in measurement (GUM).
[10] ISO 22514-4:2016. Statistical methods in process management - Capability and performance - Process capability estimates and performance measures.
[11] Johnson NL. System of frequency curves generated by methods of translation. Biometrika 1949; 36:149–176.