LOGICAL PROCEDURE FOR DETERMING THE APPROPRIATE METHOD OF CALCULATING THE PROCESS CAPABILITY

Sandro Doboviček, Tonči Mikac, Dani Damiani

Various types of influences that operate within a relatively simple production systems lead to the fact that, in these processes, a complex system of governing actions on the output parameters exists. Predicting output parameters in the future is a challenge and requires the use of many statistical tools as an important part of quality assurance in production systems. Various mathematical probability distributions that can describe the output parameters often make it difficult to assess process capability. Selection of process capability calculating formula requires data pre-processing with which it is possible to reach concise and accurate conclusions about the state of the process. This paper proposes the procedure for data pre-processing based on which it is possible to choose the appropriate formula for the calculation of process capability. Mathematical methods that are an integral part of the process are presented as well as their application in the example from the automotive industry.

Keywords: logical procedure, process capability, production systems

Logički postupak određivanja prikladne medode za računanje sposobnosti proizvodnog procesa

Različite vrste utjecaja koji djeluju unutar relativno jednostavnih proizvodnih procesa dovode do toga da u tim procesima vlada složeni sustav djelovanja na izlazne parametre procesa. Predviđanje izlaznih parametara u budućnosti predstavlja izazov i zahtijeva upotrebu niza statističkih alata kao važnih dijelova sustava osiguranja kvalitete unutar proizvodnih sustava. Različite matematičke distribucije vjerojatnosti kojima se mogu opisati izlazni parametri vrlo često otežavaju procjenu sposobnosti procesa. Odabir načina izračuna sposobnosti procesa zahtijeva pred-procesiranje podataka s kojima je moguće doći do konciznog i preciznog zaključka o stanju procesa. U ovom je radu predložen postupak pred-procesiranja podataka temeljem kojeg je moguće odabrati prikladnu formulu za izračun sposobnosti procesa. Prikazane su matematičke metode koje su sastavni dio postupka, kao i njihova primjena na primjeru iz automobilske industrije.

Ključne riječi: logički postupak, proizvodni sustavi, sposobnost procesa

1 Introduction

Manufacturers are faced with increasing demands for product quality. Those demands come from many sides: from the formal requirements such as ISO 9000 ff standards, specific standards such as ISO/TS 16949 for automotive industry and directly from customers. Methods of quality control and its statistical prediction play a key role for making reasonable and cost-effective decisions in quality management.

The statistical process control (SPC) is a methodology mainly used as a tool to understand, model and reduce the variability of an industrial process over time. SPC has the goal of reducing variation and of analyzing and improving process stability and process capability.

As SPC activities evolve from short term to long term, assignable causes contribute less and less to total process variation. A measure for evaluation of process state is the ratio of inherent process variation to total process variation [1].

Usually based on graphical tools such as control charts, the SPC methodology uses inference statistical analysis of monitored time series data to distinguish between common causes of variation due to the inherent nature of the process and assignable causes due to unusual shock or disruption [2].

Capable process is one that will produce the work within the limits of tolerance with some certainty. The process of computing capacity is based on three key parameters: average value of sample $\bar{x}$, sample standard deviation σ and known characteristics of the distribution.

Distribution of data and their position within the tolerance fields can be described with indices C_p and C_{pk}. While C_p is the quantity of variation given by standard deviation and an acceptable gap allowed by specified limits despite the mean (1), C_{pk} takes into account the distribution position within the tolerance limits (2), Table 1.

In the long term, it is expected that any process could change, to a greater or lesser extent. This means that over a longer period of data collection, the above-observed distribution is a subset of a larger distribution. This distribution represents a long-term ability of the observed process P_{pk}.

According to Six Sigma method and at the request of the majority of customers, the process is considered capable if the calculated parameter P_{pk} is greater than 1,67. Processes with a value of P_{pk} from 1,33 to 1,67 are usually conditionally accepted if a corrective actions plan is estimated to fix the value. Different systems can generate data that are non-normally distributed, but for which there is awareness of the expected form of distribution.

Existence of non-normal distribution of data is very frequent in engineering. Determining the parameters of process capability with such forms of distribution is done in a somewhat different manner since they have asymmetric control limits. Index C_{pk} is then calculated as (4) where x_p are percentiles, i.e. the value below which are the p (%) of all elements in the group. In order to precisely determine percentiles and indicators of process capability from non-normally distributed output data, it is necessary to approximate distribution in some of the non-normal distribution curves.

	Potential capability C_p		Real capability C_{pk}	
Double-sides specification limits	$\dfrac{T_u - T_l}{6\sigma}$	(1)	$\min\left(\dfrac{\bar{x} - T_l}{3\sigma}, \dfrac{T_u - \bar{x}}{3\sigma}\right)$	(2)
Lower specification limit only (T_l)	N/A		$\dfrac{\bar{x} - T_l}{3\sigma}$	
Upper specification limit only (T_u)	N/A		$\dfrac{T_u - \bar{x}}{3\sigma}$	

	Potential capability C_p		Real capability C_{pk}	
Double-sides specification limits	$\dfrac{T_u - T_l}{x_{0,99865} - x_{0,00135}}$	(3)	$\min\left(\dfrac{x_{0,5} - T_l}{x_{0,5} - x_{0,00135}}, \dfrac{T_u - x_{0,5}}{x_{0,99865} - x_{0,5}}\right)$	(4)
Lower specification limit only	N/A		$\dfrac{x_{0,5} - T_l}{x_{0,5} - x_{0,00135}}$	
Upper specification limit only	N/A		$\dfrac{T_u - x_{0,5}}{x_{0,99865} - x_{0,5}}$	

2 Variety of expected non-normal distributions

Process of selection of the mathematical distribution that best fits the observed group of data can be very important for making correct decisions in the production system quality management.

Many non-normal distributions can be applied for that purpose but Weibull and lognormal distribution usually works very well.

2.1 Weibull distribution

The Weibull distribution is one of the most widely used lifetime distributions in reliability engineering. It has an asymmetric curve and it can be both positive and negative asymmetric, Fig 1.

Weibull probability distribution has three parameters: η, β and t_0. Like the exponential distribution, it can pose as the probability density of the first occurrence of the defect. The most general expression of the Weibull probability distribution function (pdf) is given by the three-parameter Weibull distribution expression, or [4]:

$$f_W(t) = \frac{\beta}{\eta}\left(\frac{t - t_0}{\eta}\right)^{\beta-1} e^{\left[-\left(\frac{t - t_0}{\eta}\right)^{\beta}\right]}, \qquad (5)$$

where

$$\eta > 0, \beta > 0, t > 0 \text{ and } -\infty < t_0 < \infty. \qquad (6)$$

Frequently, the position parameter t_0 is not used, and the value for this parameter can be set to zero. When this is the case, the pdf equation is reduced to that of the two-parameter Weibull distribution.

There is also a form of the Weibull distribution known as the one-parameter Weibull distribution. This in fact takes the same form as the two-parameter Weibull pdf, the only difference being that the value of β is

assumed to be known beforehand. This assumption means that only the scale parameter needs to be estimated, allowing for analysis of small data sets.
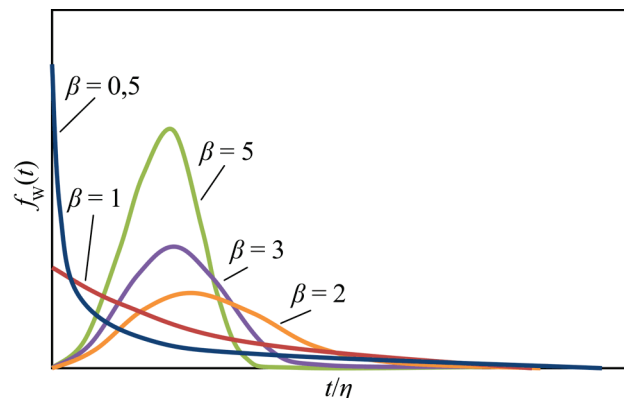

Figure 1 Three parameter Weibull distribution curve [4]

2.2 Lognormal distribution

The lognormal distribution is commonly used to model the lives of units whose failure modes are of a fatigue-stress nature. Since this includes most, if not all, mechanical systems, the lognormal distribution can have widespread application. Consequently, the lognormal distribution is a good companion to the Weibull distribution when attempting to model these types of units.

Probability density function of a three parameter lognormal distribution is described as:

$$f_L(t) = \frac{\rho}{\sqrt{2\pi}(t - t_0)} e^{\left\{-\frac{\left[\ln\left(\frac{t - t_0}{\theta}\right)^{\rho}\right]^2}{2}\right\}}, \qquad (7)$$

where

$\theta > 0$, $\rho > 0$, $t > 0$ and $-\infty < t_0 < \infty$.   (8)

In this case ρ is the shape parameter, θ a size parameter and t_0 a position parameter, Fig. 2. As may be surmised by the name, the lognormal distribution has certain similarities to the normal distribution. A random variable is lognormally distributed if the logarithm of the random variable is normally distributed [5]. Because of this, there are many mathematical similarities between the two distributions. For example, the mathematical reasoning for the construction of the probability plotting scales and the bias of parameter estimators is very similar for these two distributions.

Similarity with a normal distribution should not be a problem however, because every of the observed characteristics has its own shape that can be expected, i.e. probability curve for a hole diameter of a nominal size $\varnothing 12$ mm can be expected to be normally distributed. On the contrary, its position from a defined datum target defined as 0,2 max can be expected to be non-normally distributed because data can be only zero or positive. It can be concluded that there is a relatively high probability that position will be zero, but no chance that it can be negative. In such case, measured and collected data is not expected to be normally distributed.
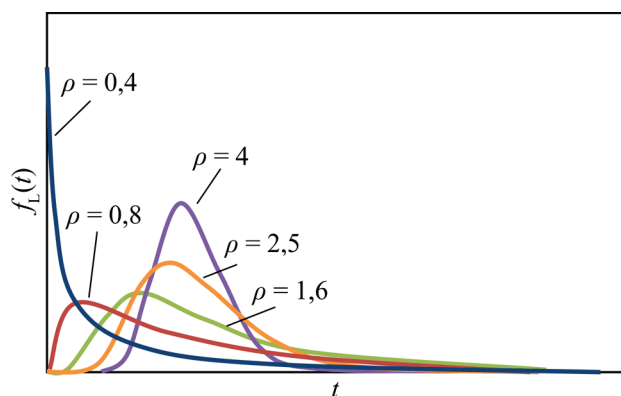

Figure 2 Three parameter lognormal distribution curve [4]

3   Problem statement

In real life environment, characteristics which are expected to be normally distributed very often are not. The causes of this should not be sought in appearance of non-normal influences. Examples are many, and the causes may be from a completely different nature.

In mass production, various operators using multiple identical machines aim to make identical products, which all must meet set thresholds in characteristics such as shaft diameter and sheet thickness. Accordingly, disparities in operators' skills and machine performance are worth consideration when calculating C_{pk}. the overall histogram can signal bimodal distribution – the mixture of two normal distributions.

The root cause for such a phenomenon is the use of two spray guns [6]. Bimodal distribution is commonplace in processes where two pieces of manufacturing equipment are employed. Care should be exercised when multiple identical machines are allocated to produce the same parts because distribution may appear non-normal.

As a result, it is difficult to troubleshoot if data is not categorized by machines and operators.

On the other hand, statistical quality control when applied to a low volume, high gauge frequency machining process may also show signs of non-normality due to smaller sample size. Harris, Mynors and Wang (2009) propose data transformations in order to centralise data and the applicability of the capability improvements [7]. Historically SPC and capability analysis are performed on high volume processes with data being gathered on a sample basis. Their investigation applies the high volume theory to a low volume 100 % data gathering process. They also discussed the validity of capability analysis of this nature due to long cycle times and large gauge inspection frequency.

Generally, in the case of data distribution "non-normality", there must be a consciousness about what it could be caused by. If used properly, there is a great variety of mathematical and statistical tools to help decision making in the quality management. Knowing the specificity of the observed cases can help in the proper use of these tools.

3.1   Data transformations

Data transformations are commonly-used tools that can serve many functions in quantitative analysis of data, including improving normality of a distribution and equalizing variance to meet assumptions and improve effect sizes, thus constituting important aspects of data cleaning and preparing for statistical analyses. There are as many potential types of data transformations as there are mathematical functions. Some of the more commonly-discussed traditional transformations include: adding constants, square root, converting to logarithmic (e.g., base 10, natural log) scales, inverting and reflecting, and applying trigonometric transformations such as sine wave transformations.

While there are many reasons to utilize transformations, one of the most important is to improve normality of data, as both parametric and nonparametric tests tend to benefit from normally distributed data. However, a cautionary note is in order. While transformations are important tools, they should be utilized thoughtfully as they fundamentally alter the nature of the variable, making the interpretation of the results somewhat more complex (e.g., instead of predicting student achievement test scores, you might be predicting the natural log of student achievement test scores). Thus, some authors suggest reversing the transformation once the analyses are done for reporting of means, standard deviations, graphing, etc. This decision ultimately depends on the nature of the hypotheses and analyses, and is best left to the discretion of the researcher.

Many statistical procedures make two assumptions that are relevant to this topic: (a) an assumption that the variables (or their error terms, more technically) are normally distributed, and (b) an assumption of homoscedasticity or homogeneity of variance, meaning that the variance of the variable remains constant over the observed range of some other variable. In regression analyses this second assumption is that the variance

around the regression line is constant across the entire observed range of data. In ANOVA analyses, this assumption is that the variance in one cell is not significantly different from that of other cells. Most statistical software packages provide ways to test both assumptions. Osborne presents Box-Cox transformation (BCT) procedures to researchers as a potential best practice in data cleaning [8].

Wheeler also finds importance in context of data and finds transformation is considered reasonable only if the meaning takes into account the context and purpose of data analysis. According to them, the software transformation of data cannot be done automatically without the recognition of context. What software can do is try to transform the distribution of it to be more "normal". Such transformations are usually very complex and contain a non-linear, exponential, reverse exponential and logarithmic function [9].

Although its form was developed during the years to adopt transformation other authors have introduced modifications of this transformation for special applications and circumstances, Box-Cox transformation takes the following form:

$$y(\lambda) = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda}, & \text{if } \lambda \neq 0; \\ \lg y, & \text{if } \lambda = 0. \end{cases} \qquad (9)$$

BCT seeks to transform the optimal λ in the range $-5 < \lambda < 5$, and the failure of the transformation can be the cause of the following: data contains the value 0, optimal λ is not in the range $-5 < \lambda < 5$ as often happens when data are too harshly defined (measured), population contains multiple populations (camel-humps). The Box-Cox transformation is a complex but useful transformation that takes the original data and raises each data observation to the power λ. However, as is true of any transformation, one of the disadvantages of Box-Cox is the difficulty in interpreting the transformed data in terms of the original measurement units.

Similarly, Johnson's transformation transforms the distribution

$$\overline{X}_w = \frac{1}{n} \sum_{i=1}^{n} Z_i, \qquad (10)$$

where

$$Z_i = \begin{cases} X_{(e+1)} & \text{if } X_i \leq X_{(e+1)} \\ X_i & \text{if } X_{(e+1)} < X_i < X_{(n-e)} \\ X_{(n-e)} & \text{if } X_i \geq X_{(n-e)} \end{cases} \qquad (11)$$

and $e = \beta n$ where β is a proportion of distribution correction.

3.2 Finding the suitable distribution

Although the Weibull and lognormal distributions are the most commonly used non-normal distributions, there are a number of known mathematical distributions in

which set of observed data can be approximated by. Common way to find a suitable distribution is use of statistical hypothesis presented in a way that can be valued with statistical-analytical procedures.

A statistical hypothesis is a mathematical expression that represents the basis for the statistical test calculation. Hypothesis test is a statistical procedure that determines whether and how reliable the available data support the hypothesis set. Hypothesis testing and significance testing is basically the process of quantifying the impressions on these hypotheses. The null hypothesis, H_0, represents a theory that has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved. The significance level α of a statistical hypothesis test is a fixed probability of wrongly rejecting the null hypothesis H_0, if it is in fact true.

The Anderson-Darling test is widely used to test if a sample of data came from a population with a specific distribution. It is a modification of the Kolmogorov-Smirnov (K-S) test and gives more weight to the tails than does the K-S test. The K-S test is distribution free in the sense that the critical values do not depend on the specific distribution being tested. The Anderson-Darling test makes use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution.

The sequence of actions for hypothesis test is: (1) set up the null hypothesis and alternative hypothesis, (2) significance level α choice, (3) sample data collection, (4) calculating the value of a specific statistical test results for the null hypothesis H_0, (5) compare the results of the statistical test with the values of specific probability distributions for a given test, (6) result interpretation in terms of probability (P-value). Significance level α should be selected with regard to possible consequences, usually 0,05 or 0,01. Hypothesis testing can verify whether the observed distribution fits a defined, non-normal or normal distribution. If P-value exceeds the significance level α, hypothesis is accepted, otherwise rejected.

4 Logical procedure

Statistical analysis can easily lead to wrong conclusions if not used properly. A large number of mathematical and statistical models that are available can be confusing and not always give useful results. Also, statistical analysis and computing process capability indicator may show different values when used at the customer or the manufacturer's factory.

Proposed logical procedure summarizes available possibilities when observed data set shows signs of non-normality, Fig. 3. It also can serve as a guide in creating software solutions for online monitoring the process capability (online SPC). It should be noted that the awareness of the expected form the distribution of probability and context data are extremely important. It is desirable that the observed data have traceability marks with the time and machine of part creation, as well as time of making measurements. The input data set is first subjected to a test of normality in a way to test the

hypothesis of normality set with a certain significance level.

Logical procedure has two possible ends: calculation with formula for normal distribution and calculation with formula for some of the non-normal distributions. In case of confirmation of the hypothesis of normal distribution, the capability is calculated using the formula for the normal distribution (1) and (2). Alternative end of the procedure may occur in cases when observed set of data fits neither of mathematical distributions. If input set of data fails the normality test, and especially if normal distribution is anticipated, it is wise to check if there is more sub-groups in data. If subgroups existence is evident, subgroups should be separated. Re-verification of the hypothesis of normal distribution at each of them separately should be made. If data transformation is acceptable and possible, another hypothesis test of normal distribution should be made on the "normalized" set of data. All hypothesis tests are marked with (HT), Fig. 3. Impossibility of data normalization can be a warning of existence of subgroups inside the observed data and signal for data revision. Finding suitable non-normal distribution involves interpretation of P-value for a goodness of fit test, such as Anderson-Darling, when using Individual distribution identification. A P-value less than α suggests that the data do not follow that distribution [10].
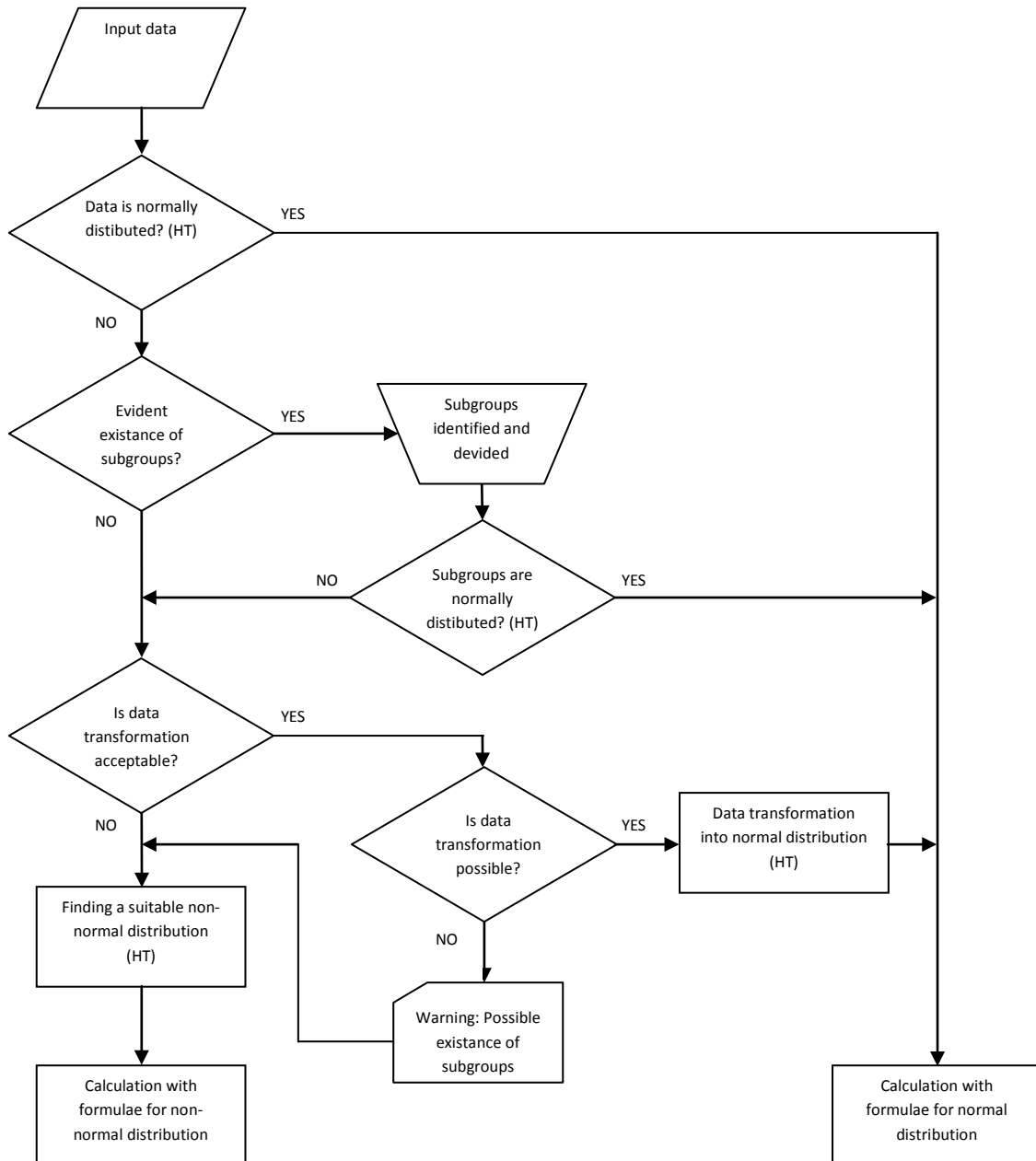

Figure 3 Logical procedure for determining the appropriate method of calculating the process capability

5   Applying logical procedure

One of the software solutions for calculating process capabilities and statistical analysis is Minitab. It is often used in conjunction with the implementation of Six Sigma, CMMI (Capability Maturity Model Integration) and other statistics-based process improvement methods. It includes various types of data transformations, numerous mathematical distributions with a possibility of testing hypothesis. This possibility makes Minitab a suitable tool for application of Logical procedure as shown below. Common case is that the work piece, whose

production requires more types of technology, takes on more productive sites - locations. The production process of levers as a part of a control mechanism of an automotive gearbox (Fig. 4) includes several technologies (operations), and complete production process takes place at four locations. It includes metal forming, machining, surface protection and mounting. The quality of the final product depends on the overall production system where mistakes are often multiplied and formation of subgroups inside the measured data is more likely. The product has a total of eight functionally defined important dimensions (special characteristics), Tab. 3. For dimensions with a boundary (as perpendicularity D2 and parallelism D3), distribution is expected to be skewed or non-normal.
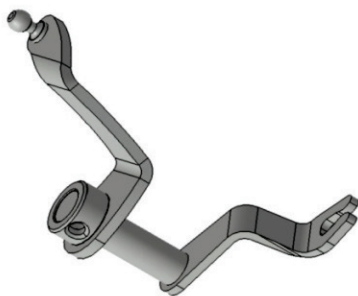

Figure 4 Part of gear lever mechanism

Two dimensions are observed, one that is expected to be normally distributed (D8) and one that is expected to be non-normally distributed (D3). As the manufacturing process is done at several locations and at different times (at suppliers factories), and since the dimensions are interdependent, although the measured sample of 40 pieces is taken from the assembly operations in a single shift, the observation cannot be regarded as short-term.

Table 3 Functionally important characteristics and their expected distributions

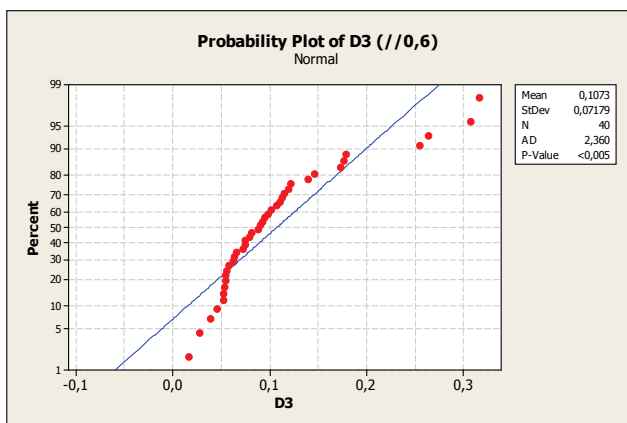| Index	Dimension (in mm)	Expected distribution
D1	length 29±0,3	normal
D2	perpendicularity 0,6 max.	non-normal
D3	parallelism 0,6 max.	non-normal
D4	length 47,48±2	normal
D5	length 83,3±1,5	normal
D6	length 90±0,5	normal
D7	length 90±0,5	normal
D8	length 69,2±0,35	normal

Figure 5 A-D normality test for characteristic D3

Anderson-Darling (A-D) normality test showed *P*-values for both distributions <0,05 (chosen significance level is α=0,05) which means that the null-hypothesis is not accepted and that none of the observed distributions can be accepted as normal, Figs. 5 and 6.

It can be concluded that the dimension D3 is distributed as expected, Fig. 7. Distribution histogram of characteristics D8 shows signs of sub-grouping. Although there is no precise mathematical solution to the identification and separation of these data sets, this indicates a considerable instability which can be attributed to the fact that it is a dimension, to which the impact is more productive sub-processes, Fig. 8. Their mathematical identification in this case is unknown.
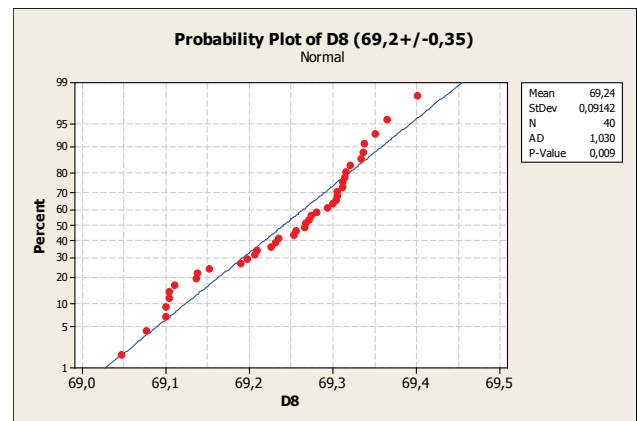

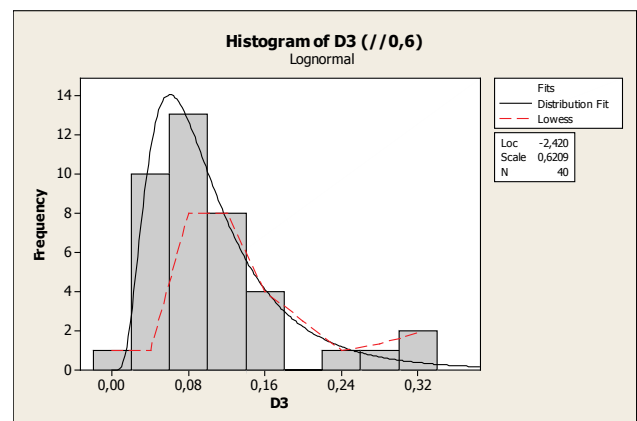**Figure 6** A-D normality test for characteristic D8

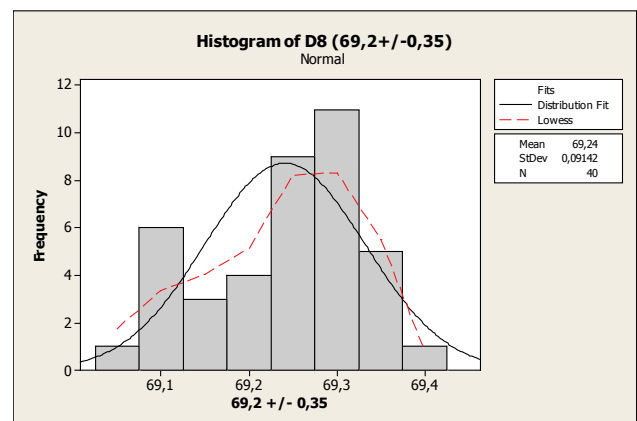**Figure 7** Histogram for characteristic D3

Figure 8 Histogram for characteristic D8

Assuming acceptance of the calculation based on the ability of transformed data, the Box-Cox transformation

of the D3 characteristic distribution finds the optimal values of λ in the $\lambda=-0,02$, Fig. 9. The same method of transformation does not find the optimal λ in the range $-5<\lambda<5$ for the distribution of characteristic D8 as confirmed by the lack of such methods in cases when the distribution contains several sub-groups (camel hump shape, Fig. 10.
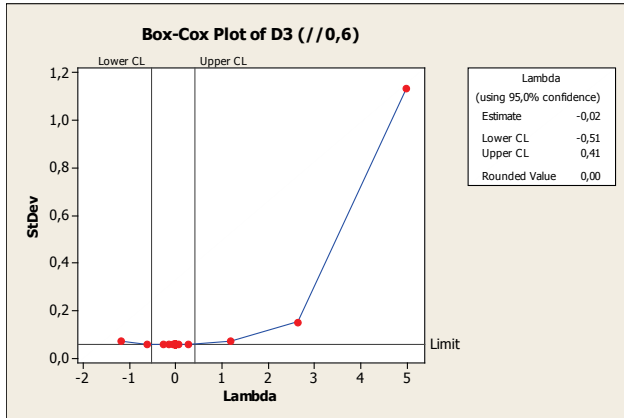

Figure 9 Box-Cox transformation plot for characteristic D3
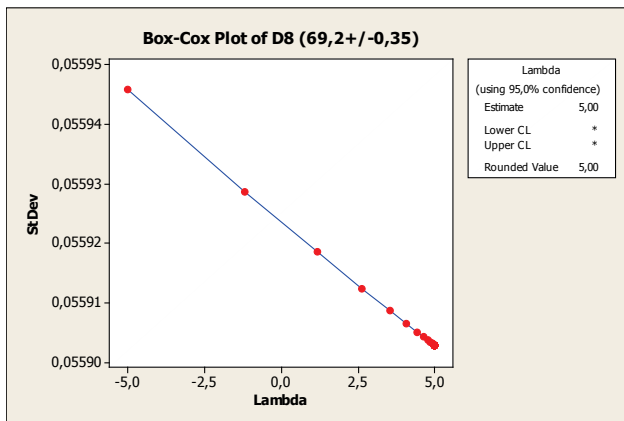

Figure 10 Box-Cox transformation plot for characteristic D8

Johnson transformation is indicated as a more powerful tool in this case. Johnson distribution function for the dimension of characteristic D3 is then:

$$f(x) = -1{,}166 + 0{,}9907 \cdot \operatorname{asinh}\left(\frac{x - 0{,}0443}{0{,}0250}\right) \qquad (12)$$

and for characteristic D8 (Fig. 11):

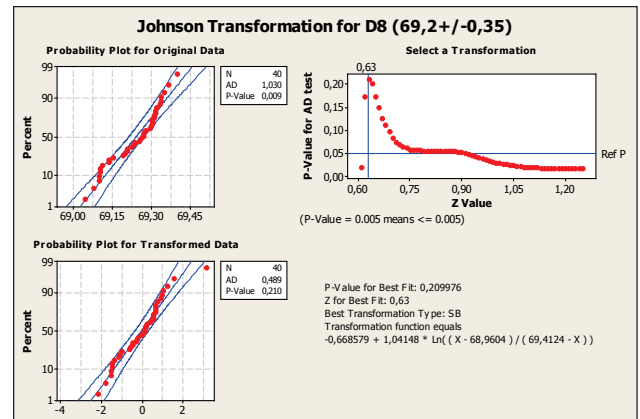$$f(x) = -0{,}668 + 1{,}041 \cdot \ln\left(\frac{x - 68{,}960}{69{,}4124 - x}\right). \qquad (13)$$

Figure 11 Johnson transformation plot for characteristic D8

Normality test of distribution of characteristic D8 after Johnson transformation shows P-value$=0{,}936$ which indicates that the null hypothesis of normal distribution is accepted. Considering the context of data that defines the distribution of characteristic D3, the process capability is useful to calculate on the non-transformed data.

Basis for such calculation is finding a non-normal distribution that fits the observed distribution i.e. individual distribution identification. In the case of characteristic D3, the test shows that the significance level $\alpha=0{,}05$:

- lognormal distribution ($P=0{,}517$)
- 3-parameter Weibull distribution ($P=0{,}070$)
- gamma distribution ($P=0{,}082$) and
- loglogistic distribution ($P>0{,}250$)

can be accepted, Tab. 4. The Likelihood Ratio Test (LRT P) rates significance of 3-parameter distribution from a 2-parameter distribution. If the P-value is less than α, then the improvement by using a 3-parameter distribution (instead of a 2-parameter distribution) is large enough to be statistically significant. Therefore, the 3-parameter lognormal distribution (LRT $P=0{,}815$) and 3-parameter gamma distribution (LRT $P=0{,}082$) can also be accepted.

Table 4 Goodness of fit test and distribution parameters for characteristic D3 (Minitab)

Distribution	A-D	P-value	LRT P	Location	Shape	Scale	Threshold
Normal	2,360	<0,005		0,1073		0,07179	
Box-Cox transformation	0,322	0,517		−2,41985		0,62092	
Lognormal	0,322	0,517		−2,41985		0,62092	
3-Parameter Lognormal	0,328		0,815	−2,37907		0,58821	-0,00309
Exponential	3,542	<0,003				0,10730	
2-Parameter Exponential	2,256	<0,010	0,001			0,09262	0,01468
Weibull	1,052	<0,010			1,65390	0,12101	
3-Parameter Weibull	0,709	0,070	0,049		1,36800	0,10059	0,01551
Smallest Extreme Value	4,026	<0,010		0,14726		0,09004	
Largest Extreme Value	0,746	0,047		0,07793		0,04544	
Gamma	0,685	0,082			2,81915	0,03806	
3-Parameter Gamma	0,543		0,280		2,03494	0,04668	0,01232
Logistic	1,503	<0,005		0,09549		0,03570	
Loglogistic	0,237	>0,250		−2,43112		0,34188	
3-Parameter Loglogistic	0,222		0,773	−2,48587		0,36264	0,00435
Johnson Transformation	0,165	0,936		0,09058		0,95918	

Multiple distributions that can be accepted can be explained by the smaller sample size. Process capability analysis of characteristic D3 for each distribution with a *P*-value greater than a given significance level α (α=0,05) results with a different capability indices, Tab. 5. This shows the importance of this step where decision on the selected distribution can be influential on quality management decisions.

Johnson transformation of distribution D8 increases P_{pk} indices which can be significant for determining process status, Tab. 6. Transformation can be justified with expected normal distribution for this characteristic.

Table 5 Process capability indices for characteristic D3 (Minitab)

Distribution	LB	USL	Target	C_{pk}
3-Parameter Lognormal	0	0,60	0	1,14
3-Parameter Weibull	0	0,60	0	1,57
3-Parameter Gamma	0	0,60	0	1,50
3-Parameter Loglogistic	0	0,60	0	0,62
Box-Cox transformation	−6,91*	−0,51	−6,91	1,04
Johnson transformation	−2,49	2,59	−2,49	0,87

*LSL

Table 6 Process capability indices for characteristic D8 (Minitab)

Distribution	LSL	USL	Target	C_{pk}	P_{pk}
Normal	68,85	69,55	69,200	1,84	1,13
Johnson transformation	-	-	−0,543	-	1,57

6 Conclusion

When the distribution of a process characteristic is non-normal. capability indices C_p and C_{pk} calculated using conventional methods often lead to wrong interpretation of the process's capability. Non-normality of the distribution of observed data may be the cause of the nature of their occurrence (distribution is expected to be non-normal) or in multiple populations occurred in the manufacturing process (parallel streamlines).

Proposed logical procedure summarizes possibilities in case of non-normality. Using of logical procedure assumes the knowledge of the expected distribution of observed characteristics.
Procedure is shown in two examples with different forms of the expected distribution. In case of characteristic for which the expected distribution is in normal form normalization of data was justified and done with one of the available methods of normalization. In doing so the result of Box-Cox method of normalization can be a signal that the distribution is made of several subsets.

The choice of mathematical distribution that fits the observed distribution in observed case shows significant differences in capability indices. This shows the importance of this step in a logical procedure. A larger sample of observed data can certainly make this step more accurate by reducing the number of possible mathematical distributions. It can be concluded that this step is important in complementing the new data for a given data set where the "new" information may affect the change in Goodness of fit test results and thus change the capability indices.

7 References

[1] Grabov, P. SPC's Effectiveness Prediction and Analysis. International Congress Annual Quality Congress - Proceedings. Philadelphia, 1998.
[2] Fournier, B.; Rupin, N.; Biberelle, M.; Najjar, D.; Iost, A. Application of the generalized lambda distributions in a statistical process control methodology. // Journal of Process Control. 16, (2006).
[3] Czarski, A. Estimation of process capability indices in case of distribution unlike the normal one. // Archives of Materials Science and Engineering. 34, 1(2008), pp. 39-42.
[4] Liu, C.-C.; Willey, P. C. T. The approximate equations between the Weibull and lognormal distributions using MRR. http://academic-papers.org/ocs2/-session/Papers/F3/ 747-2034-2-DR.doc
[5] Zongming, L. How to Avoid Common Mistakes When Measuring Performance. http://www.isixsigma.com/. 2012.
[6] Harris, M. A.; Mynors, D. J.; Wang, C. J. A low volume production process capability analysis. // Applied Mechanics and Materials. 16, 19(2009), pp. 1038-1042.
[7] Osborne, J. W. Improving your data transformations: Applying the Box-Cox transformation: Practical Assessment. // Research & Evaluation. (2010), pp. 15-12.
[8] Wheeler, D. J. Understanding Variation – The Key to Managing Chaos. SPC Press, 2000.
[9] NIST/SEMATECH e-Handbook of Statistical Methods. http://www.itl.nist.gov/div898/handbook/. 2012.
[10] Minitab Support, http://www.minitab.com/en-US/support/

Authors' addresses

Sandro Doboviček
Faculty of Engineering
University of Rijeka
Vukovarska 58
HR-51000 Rijeka, Croatia
E-mail: sandro.dobovicek@riteh.hr

Tonči Mikac, Full Professor Ph.D.
Faculty of Engineering
University of Rijeka
Vukovarska 58
HR-51000 Rijeka, Croatia
E-mail: tmikac@riteh.hr

Dani Damiani
P. P. C. Buzet d.o.o.
Most 24
HR-52420 Buzet, Croatia
E-mail: dani.damiani@cimos.eu