Management Systems
in
Production Engineering

2012, No 3 (7), pp 26-30

IMPROVING QUALITY OF STATISTICAL PROCESS CONTROL BY DEALING WITH NON-NORMAL DATA IN AUTOMOTIVE INDUSTRY

Zuzana ANDRÁSSYOVÁ, Tomáš PAULIČEK, Pavol PICHŇA, Martin KOTUS
Slovak University of Agriculture in Nitra
Maria del Carmen BAS CERDÁ
Universidad Politécnica de Valencia

Abstract:
Study deals with an analysis of data to the effect that it improves the quality of statistical tools in processes of assembly of automobile seats. Normal distribution of variables is one of inevitable conditions for the analysis, examination, and improvement of the manufacturing processes (f. e.: manufacturing process capability) although, there are constantly more approaches to non-normal data handling. An appropriate probability distribution of measured data is firstly tested by the goodness of fit of empirical distribution with theoretical normal distribution on the basis of hypothesis testing using programme StatGraphics Centurion XV.II. Data are collected from the assembly process of 1st row automobile seats for each characteristic of quality (Safety Regulation -S/R) individually. Study closely processes the measured data of an airbag´s assembly and it aims to accomplish the normal distributed data and apply it the statistical process control. Results of the contribution conclude in a statement of rejection of the null hypothesis (measured variables do not follow the normal distribution) therefore it is necessary to begin to work on data transformation supported by Minitab15. Even this approach does not reach a normal distributed data and so should be proposed a procedure that leads to the quality output of whole statistical control of manufacturing processes.

Key words: *quality, statistical control, hypothesis testing, probability distribution, normality, data transformation*

INTRODUCTION

Manufacturing process quality of automobile seat in organization is monitored and evaluated by various statistical methods; one of them is the evaluation of manufacturing process capability on the basis of control charts and indexes of capability C_p and C_{pk}. Important condition for evaluating the process capability is a fact that studied variables follow the normal distribution. A medium-difficult statistical tools of quality for hypothesis testing about a goodness of fit (f. e.: Shapiro-Wilk, Kolmogorov-Smirnov, etc.) are used for an empirical distribution fit with theoretical one after the basic statistical tools for normality evaluation such as histogram, density function and probability plot. Contribution focuses on the application of tools for statistical process control in automotive industry where we can often find the non-normal data distributions. Measured values are analysed so it is possible to reject or accept the null hypothesis at the pre-determined level of significance α. Technique for the manufacturing process observation according to the quality of data entering into analysis is determined after the revealed results.

In case of acceptance of the null hypothesis, process capability evaluation of automotive seats would be carried on. Otherwise, it is necessary to approach to the transformation of data, to alteration of the indexes of manufacturing process capability C_p and C_{pk} for variables that do not follow the normal distribution or, as mentioned by Haridi (2011), to approach to a method which was developed by Ford Motor Company (estimated cumulative frequencies, mirroring technique).

OBJECT AND METHODS

Data collection

Measured data are collected from the process of assembly of automobile seats for AUDI Q7 and registered into a database of screwing torques of S/R quality characteristics. Case study closely focuses on the analysis of variables N = 250 (screwing torques, Nm) from the process of airbag assembly (figure 1) which are defined for the process capability observation and evaluation (table 1). Nominal value of the screwing torque for airbag is 8.1 ± 0.81 Nm with a tolerance.

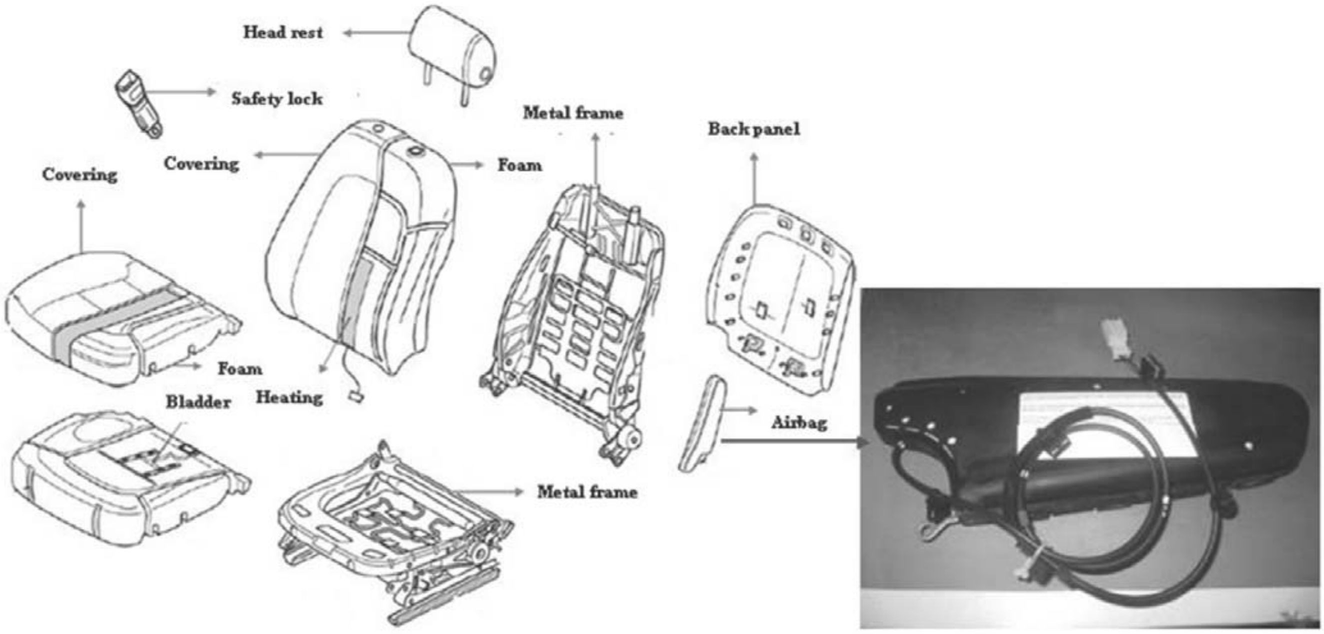

Fig. 1. Components of automobile seating

Table 1
Measured variables of screwing torque for airbag, Nm

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	8.141	8.141	8.108	8.132	8.213	8.108	8.141	8.213	8.165	8.213	8.132	8.165	8.124	8.149	8.181	8.173	8.100	8.149	8.141	8.141	8.108	8.132	8.213	8.108	8.141
2	8.181	8.189	8.124	8.124	8.100	8.108	8.149	8.157	8.197	8.108	8.157	8.173	8.222	8.173	8.197	8.181	8.100	8.108	8.181	8.189	8.124	8.124	8.100	8.108	8.149
3	8.100	8.213	8.181	8.173	8.108	8.132	8.124	8.132	8.173	8.124	8.100	8.116	8.124	8,181	8.132	8.108	8.116	8.116	8.100	8.213	8.181	8.173	8.108	8.132	8.124
4	8.173	8.149	8.230	8.222	8.124	8.205	8.124	8.108	8.116	8.157	8.108	8.132	8.181	8.124	8.132	8.124	8.116	8.116	8.173	8.149	8.230	8.222	8.124	8.205	8.124
5	8.197	8.157	8.197	8.116	8.124	8.100	8.100	8.181	8.116	8.100	8.165	8.149	8.197	8.141	8.124	8.197	8.124	8.116	8.197	8.157	8.197	8.116	8.124	8.100	8.100
6	8.181	8.222	8.181	8.116	8.100	8.149	8.141	8.246	8.132	8.141	8.124	8.165	8.157	8.157	8.173	8.165	8.149	8.116	8.181	8.222	8.181	8.116	8.100	8.149	8.141
7	8.116	8.181	8.189	8.132	8.108	8.132	8.181	8.124	8.141	8.197	8.173	8.197	8.100	8.116	8.141	8.213	8.157	8.141	8.116	8.181	8.189	8.132	8.108	8.132	8.181
8	8.124	8.157	8.165	8.181	8.141	8.124	8.141	8.141	8.173	8.100	8.173	8.181	8.157	8.108	8.108	8.213	8.173	8.149	8.124	8.157	8.165	8.181	8.141	8.124	8.141
9	8.141	8.116	8.213	8.157	8.132	8.149	8.157	8.124	8.132	8.141	8.149	8.116	8.157	8.108	8.181	8.116	8.116	8.124	8.141	8.116	8.213	8.157	8.132	8.149	8.157
10	8.141	8.108	8.157	8.173	8.173	8.189	8.108	8.246	8.124	8.100	8.165	8.108	8.149	8.108	8.189	8.173	8.108	8.165	8.141	8.108	8.157	8.173	8.173	8.189	8.108

Histogram, probability density function and probability plot

Firstly, all measured variables are analysed to define the normal distribution by histogram and probability density function. Variables are fitted into identical intervals with a number of intervals *k* and an interval width *d*. Apart from calculation of absolute and relative frequencies, values of the probability density function have to be calculated. Probability density function is plotted on the histogram of frequencies, it links means of individual intervals and counted values *f(x)*. Probability plot and other resultant images show if normal or non-normal distribution.

Goodness of fit of empirical with theoretical distribution

Pearson, Shapiro-Wilk, Kolmogorov-Smirnov tests of goodness of fit are used for the statistical hypothesis H_0 testing about the goodness of fit of empirical distribution with theoretical normal distribution.

Pearson coefficient

$$\chi^2 = \sum_{i=1}^{k'} \frac{(m_i - m_{Ti})^2}{m_{Ti}} \tag{1}$$

where:
m_{Ti} - expected frequency of enlarged *i*-interval,
m_i - observed frequency of enlarged *i*-interval,

Shapiro-Wilk

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{2}$$

where:
$x_{(i)}$ - ordered sample values ($x_{(1)}$ is the smallest),
a_i - constant generated from the mean, variance and covariance of the order statistics of a sample of size *n* from a normal distribution,

Kolmogorov-Smirnov

$$Dn = sup_x \lfloor F_n(x) - F(x) \rfloor \tag{3}$$

where:
sup_x - supremum of the set of distances,
F(x) - cumulative distribution function,
$F_n(x)$ - empirical distribution function.

Level of significance is normally defined as α = 0.05 (5%) for hypothesis testing which divides the set of values into the rejected and accepted ones. A difference between statistical expected results for normal distribution and observed results would define a final statistical significance (probability) of goodness of fit.

Finally, formulated null hypothesis H_0 is rejected or accepted. So the measured variables are evaluated as following the normal distribution or not, next moves are discussed.

Transformation of non-normal data

The most common dealing with non-normal data is applying a transformation technique. Transformation changes the measured variable *x* into transformed variable *y* according to a selected method (logarithmic, square root, inverse, arcsine, etc.). Family of transformation Box Cox is considered, by Osborne (2010), to be the one which develops and improve conventional transformation methods (*log x, √x, x^{-1}, arcsine x*, etc.). So it is also used for an experiment depending on the estimated parameter λ (the most common from -5 up to 5).
Box Cox

$$\Phi_\lambda(x) = \frac{x^\lambda - 1}{\lambda}, \qquad \text{if } \lambda \neq 0 \tag{4}$$

$$\Phi_{(x)} = log(x), \qquad \text{if } \lambda = 0 \tag{5}$$

There are several methods but they are used on the basis of the distribution of analysed data. Phase of determination of measured data distribution and suitable transformation method is one of the most difficult. The programme StatGraphics is used for graphical image of function of measured data (lognormal, uniform, Weibull, etc.) which could help with a following decision. Then a selected transformation is applied on data in order to reach their normality for evaluation of manufacturing process capability of assembly of automobile seats and so improve the quality of application of statistical tools in automotive industry.

RESULTS AND DISCUSSION

Calculated frequencies were used for plotting on histogram and for a visual presentation of distribution of the airbag's measured data by programme StatGraphics. The average value were counted x = 8.14932 and the standard deviation σ = 0.0353557, also the number of intervals k = 16 and the range of intervals d = 0.0101. Probability density function *f(x)* for normal distribution was plotted on the histogram. Histogram, probability density function and probability plot for normal distribution shows that measured variables of airbag's screwing torques (figure 2) do not following the normal distribution. In this case, it was necessary to validate this result by the test of goodness of fit for normal distribution.

Hypothesis testing about the goodness of fit

The null hypothesis H_0 was formulated as a no-significant difference between the expected and observed frequencies. Data follow the normal distribution.

→ H_0: $F(x_e) = F(x_t)$

Alternative hypothesis H_1 was formulated as a significant difference between the expected and observed frequencies. Data do not follow the normal distribution.

→ H_1: $F(x_e) \neq F(x_t)$

Level of significance were defined as α = 0.05.

Because the calculated values of resultant probability are $p < \alpha$, the null hypothesis H_0 was rejected at the level of significance 5% (table 2). Detected difference is very large and so it cannot be random (chance), it is proved as the statistical significant difference. Therefore, the alternative hypothesis about the non-normal distributed data was accepted at the level of significance 95%.

The fact, that measured variables do not follow the normal distribution and so they do not meet the condition of normality for the capability evaluation of airbag's assembly by control charts and indexes of manufacturing process capability C_p and C_{pk}, urges to propose next steps for studying and evaluating the process.

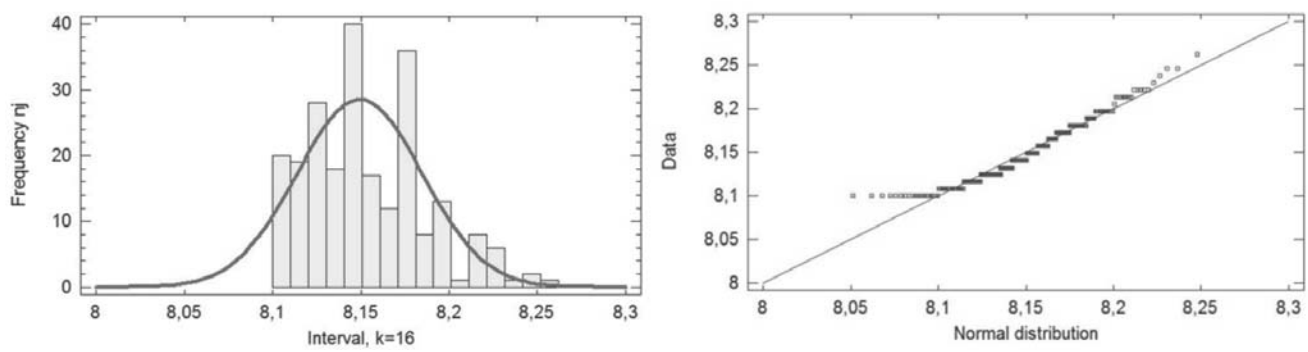

Fig. 2. Histogram, probability density function and probability plot for measured variables (StatGraphics)

Table 2
Results of the test of normality for measured data of screwing torque (StatGraphics)

Type of the test	Statistics	Probability p
Shapiro-Wilk	W = 0.933168	1.89e-15
Kolmogorov-Smirnov	Dn = 0.111087	0.00418192
Pearson	χ2 = 51.7028	6.99822e-07

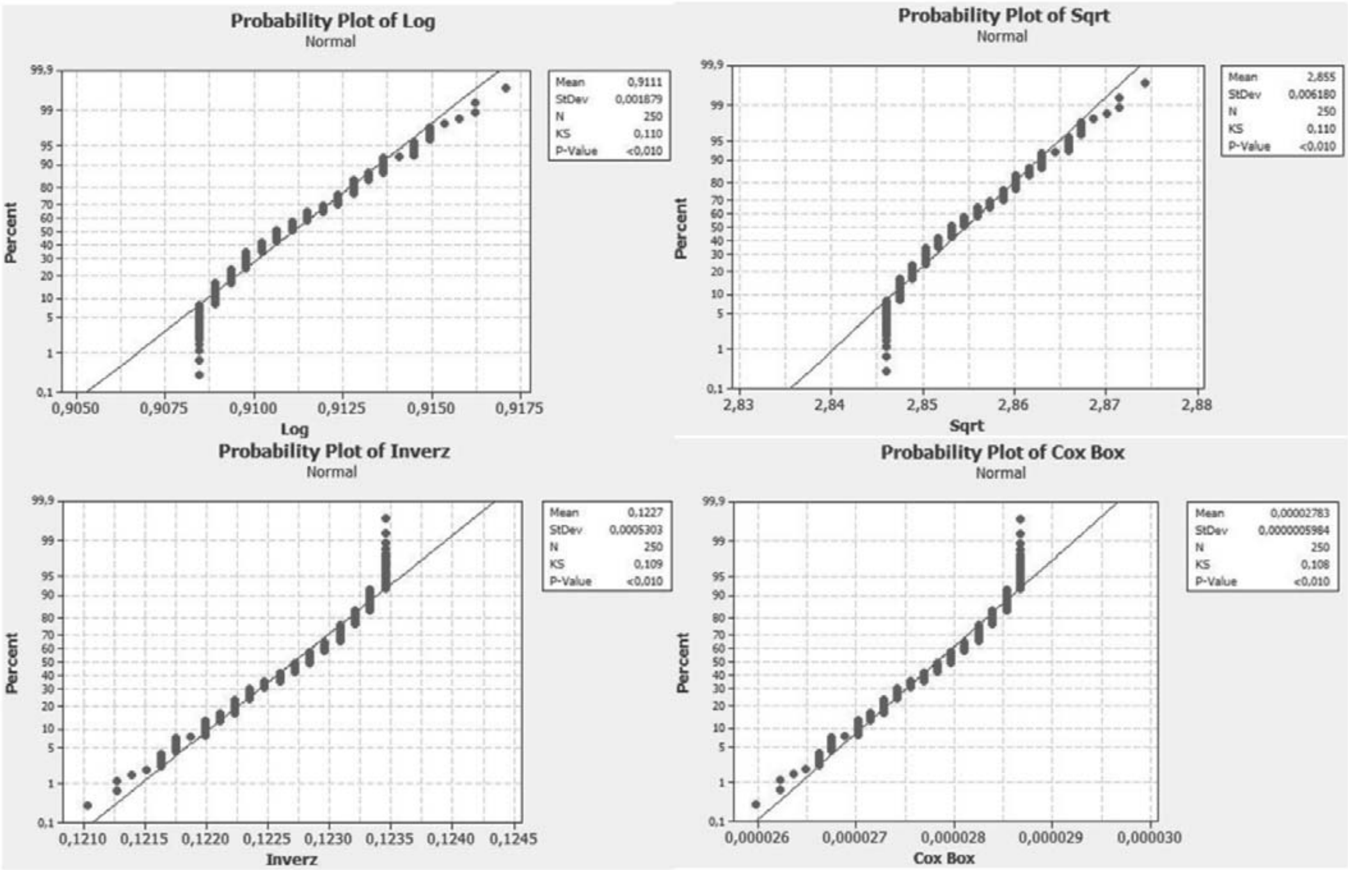

Fig. 3. Resulting probability plots of transformed data for each transformation method (Minitab15)

Data transformation

According to a fact that it was difficult to determine the distribution of measured data of screwing torque from the graphical image unequivocally, we applied the transformation methods for non-normal data which were the most fitted ones in practice: inverse, square root, logarithmic and family of Box Cox (λ = -5, estimated by software) with a support of software Minitab15. Further, the transformed data were analysed in view of their distribution (figure 3). None of transformed function followed the normal distribution. Results show that none of the transformation methods of this study can be used for measured data of screwing torque of airbag with aim to reach their normality. Transformation is not always the best choice for handling the non-normal data although it is the most common method. Therefore it is needed to consider an application of another methodology for non-normal data handling that are necessary for further statistical quality control of manufacturing processes.

Johnson (2007) introduces a usage of Pearson's resp. Johnson technique for goodness of fit and a determination of process capability with suitable percentage points of distribution in his study. He also states that whether a transformation method for non-normal data has to be selected or non-normal distribution model is going to be identified it is useful to choose a Statistical Software (f. e.: Minitab can be used to accurately verify process stability and calculate process capability for non-normal quality characteristics) for that investigation.

The same manufacturing processes usually demonstrate very similar data behaviour. If data distribution is determined correctly and transformation applied with success, it is possible to use this transformation method for other similar manufacturing processes without the closer analysis of non-normal data. Sharman (2012) states two types of non-normal data: data that follows the other distribution and data that contents mixture or multiple distributions. Osborne (2002) depicts effects of application of the data transformation according to selected transformation method.

CONCLUSION

Determination of the measured data distribution by the basic statistical tools of quality (histogram) shows that values do not follow the normal distribution. This fact is validated by hypothesis testing about the goodness of fit of empirical and theoretical normal distribution, the null hypothesis H_0 is rejected. Therefore, further study of transformation of non-normal data into data following the normality is needed for Shewhart's control charts and indexes of manufacturing process capability C_p and C_{pk} for the statistical quality control. Procedure for normality satisfaction resp. calculation of indexes of process capability for non-normal data distribution is going to be defined after a closer analysis of measured variables. Because there were not accomplished the aim to reach the condition for statistical process control used in the organization, another technique for normality or modification of indexes of process capability for non-normal distributed data.

REFERENCES

[1] Haridy S., Wua Z., Castagliola P.: Univariate and multivariate approaches for evaluating the capability of dynamic-behaviour processes. Statistical Methodology. Vol. 8, Issue 2, 2011, pp 185–203.
[2] Sherman J. P.: Tips for Recognizing and Transforming Non-normal Data. iSixSigma. http://www.isixsigma.com/tools-templates/normality/tips-recognizing-and-transforming-non-normal-data/ (23.03. 2012).
[3] Osborne J.: Notes on the use of data transformations. Practical Assessment, Research & Evaluation. Vol. 8, Issue 6, 2002. http://PAREonline.net/getvn.asp?v=8&n=6 (25.03.2012).
[4] STN ISO 8258: 1995 (010271) Shewhart control charts. SR: Slovak Standards Institute (SUTN). 1995.
[5] Johnson L.: Modelling non-normal data using statistical software. Process control and process capability can now be modelled using non-normal distributions. R&D Magazine. http://rdmag.com/Featured-Articles/2007/08/Modeling-Non-Normal-Data-Using-Statistical-Software/ (26.03.2012).
[6] Osborne, J.: Improving your data transformations: Applying the Box-Cox transformation. Practical Assessment, Research & Evaluation. Vol. 15, Issue 12, 2010. http://pareonline.net/pdf/v15n12.pdf (18.05.2012).

Ing. Zuzana Andrássyová
Ing. Tomáš Pauliček
Ing. Pavol Pichňa
Ing. Martin Kotus, PhD.
Slovak University of Agriculture, Faculty of Engineering
Department of Quality and Engineering Technologies
Tr. A. Hlinku 2, 949 76 Nitra, SLOVAKIA
e-mail: xandrassyova@is.uniag.sk

Maria del Carmen Bas Cerdá
Universidad Politécnica de Valencia
Centre for Quality and Change Management
Camino de Vera, s/n 46022 Valencia, SPAIN
e-mail: maibacer@doctor.upv.es