Software Implementation and Visualization for Process Capability of One-dimensional Non-normal Data

Geng WANG\textsuperscript{a,}\textsuperscript{*}, Zheng-Quan ZHOU

Department of Statistics, Nanjing Finance & Economics University, Wenyuan Road, Qixia District, Nanjing, Jiangsu, China
\textsuperscript{a}wgwgw0912@126.com
\textsuperscript{*}Corresponding author

Keywords: Process Capability, the Generalized Process Capability Index, Data Conversion.

Abstract. For one-dimensional non-normal Data measurement data process ability, this paper used JMP software cleverly to calculate the generalized capability index formula and new abnormal data capability index formula. Finally, use non parametric software package to calculate the process capability index and implement in the software Minitab. The example showed the good characteristics of JMP visualization in the calculating of process capability adequately.

Introduction

Usually, the process capability indices calculation formula is designed according to the probability characteristic of normal distribution. In practical production, a lot of the stable process can not meet the needs of normal distribution hypothesis. The reason is that not only systemic factors which can be identified can affect the course of processing, but also random factors that excluding special reasons can make a difference. When the process is only affected by random factors, says the process is in a state of statistical control, Process the output data obey the stable normal distribution, when the system factors exist in the process, In a state of statistical control process, the output of the process will not be normal at this time.

Systemic factors refer to those relatively stable factors that have an impact on the process of mass production rather than a single product. Additionally, when the processing operator have a subjective tendency during operation, also can produce the output of the non-normal distribution. The above phenomenon often occurs in many parts processing enterprises, such as petroleum, chemical, metallurgical and so on. These enterprises’ production process are often autocorrelation, and there exists a cycle fluctuation during production process. At this point, the normality assumption is based on the process capability index has lost its original meaning, Process capability index does not make sense. The estimation of the fraction defective will have serious deviation. In this context, how to calculate CP under the condition of non-normal distribution is one of the hot spots of process capability analysis.

Under the condition of non-normal distribution, there are four methods to calculate CP(Process capability index).The first method is data transformation. Through a variety of data conversion function will non-normal data into a normal data. The second method is curve fitting, this method mainly uses Pearson family of curves and Johnson family of curves to fit the frequency distribution of the actual data. The third method is generalized CP calculation, this method redefine the way to calculate CP by generalizing process which obey normal distribution. After redefinition, we can use the new definition to calculate CP under different
random distribution conditions. The last method is empirical distribution method. We usually use Minitab15.0 and JMP9.02.

**Method 1**

Data transformation method. We use common transform (logarithmic transformation, square root transformation, inverse transformation, power conversion) and the Box - Cox transformation method or Johnson transformation method to transform statistical data with non-normal distribution into normal distribution. Then we can calculate CP on the basis of normal distribution. Data transformation method can be realized by JMP or Minitab, Minitab also offers users related algorithms. However, since the essence of these different data transformation methods are all the same, namely make data close to normal distribution by adjusting kurtosis and skewness of the data distribution, only if original data distribution obey normal distribution or approximate normal distribution, method 1 will be valid. When the distribution of the data does not satisfy the above conditions, method 1 won’t work. Furthermore, for above conversion methods, there does not exist selection criteria, we need to find the optimal way to transform data by repeating verification process, that means a large amount of work is required, and we could not make sure whether we can achieve a good result.

**Method 2**

Generalized CP calculation. Method 2 can be realized by JMP or we can calculate non-normal process capability index by JMP.

**Method 3**

Non-parameter calculation method. We could calculate CP if there are enough data. Method 3 is usually carried out by Minitab. Math principles can be programmed to macros operations ncap which can be invoked. Related operations can be referred to Minitab help.

**Question**

A production process of LG needs a kind of raw material (bottom). Quality control personnel will test bottom’s quality after suppliers change their abrasives. The QC will measure every mould’s length, width and thickness. Since the data of length and thickness usually obey normal distribution, CP of length and thickness are easily to calculate. However, under normal circumstances, the data of width does not obey normal distribution. How to calculate width’s CP is a difficult point.

<table>
<thead>
<tr>
<th>num</th>
<th>width</th>
<th>num</th>
<th>width</th>
<th>num</th>
<th>width</th>
<th>num</th>
<th>width</th>
<th>num</th>
<th>width</th>
<th>num</th>
<th>width</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4.320</td>
<td>10</td>
<td>4.325</td>
<td>18</td>
<td>4.310</td>
<td>26</td>
<td>4.296</td>
<td>34</td>
<td>4.301</td>
<td>42</td>
<td>4.286</td>
</tr>
<tr>
<td>3</td>
<td>4.316</td>
<td>11</td>
<td>4.327</td>
<td>19</td>
<td>4.320</td>
<td>27</td>
<td>4.301</td>
<td>35</td>
<td>4.299</td>
<td>43</td>
<td>4.320</td>
</tr>
</tbody>
</table>

Specifications: 4.25-4.35
Under one-dimensional normal distribution, the formula for the capacity index of the measurement data process shows as:

Statistical Model

Statistical Model 1

If the upper and lower limits are required, the process capability index is:

\[
C_p = \frac{T_U - T_L}{6\sigma} \approx \frac{T_U - T_L}{6s}
\]

(1)

Statistical Model 2

Bilateral tolerance, symmetrical distribution, center offset process capability index is:

\[
C_{pk} = \min\{C_{pu}, C_{pl}\} = \frac{\min\{T_U - \mu, \mu - T_L\}}{3\sigma} \approx \min\left\{\frac{T_U - \bar{x}}{3s}, \frac{\bar{x} - T_L}{3s}\right\}
\]

(2)

\(\bar{x}, \mu, \sigma\) respectively means sample data mean, sample variance, the corresponding random variables mean and standard deviation of the normal distribution; \(T_L, T_U\) respectively for lower and upper limits. The process capability index of the normal distribution data can be calculated by the JMP software.

Calculate process capability according to normal distribution directly:
This is an animation of the process capability. Dragging the mouse to offset block, we can show how the process capability changes with offset sample. At this time, the process capability index $C_p=0.889$, $C_{pk}=0.63$.

Histogram and normal probability chart indicate that the wide data distribution is not normal. Therefore, the process capability index is not accurate, there exists a deviation.

This paper talks about the software realization of the process capability index of abnormal data as follows. As mentioned above, there are three ways.

Method 1: Data conversion method. The implementation steps are

- **Step1.** With data conversion commonly used such as logarithmic conversion, square root transformation, reciprocal transformation, square root after reciprocal, square root after retaking arcsin, power conversion. These transformation can be realized in JMP software. As for Box-Cox transformation method or Johnson transformation method for converting abnormal data to normal data can be directly transformed in Minitab software menu.

- **Step2.** Normal distribution of the process capability index formula can be directly achieved
in the Minitab software or JMP software. See references [3] and [4]. The reasons have mentioned above.

Method 2: Software method and generalized process capability index method

(1) If the following abnormal distribution (showed below) is known, it can be calculated directly,

\[
\begin{align*}
0.00135 &= P_{T - T_L} \\
0.5 &= P_{T - T_U} \\
0.99865 &= P_{T - T_U}
\end{align*}
\]

(2) Calculation formula for the generalized process capability index of the measurement data

In this paper, the formula for calculating the generalized process capability index of the measurement data of a particular distribution in one dimensional abnormal distribution is given by the reference[1] and reference [2]:

\[
C_p = \frac{T_U - T_L}{P_{0.99865} - P_{0.00135}}
\]

\[
C_{pk} = \min\left\{ \frac{T_U - P_{0.05}}{P_{0.99865} - P_{0.05}}, \frac{P_{0.5} - T_L}{P_{0.99865} - P_{0.00135}} \right\}
\]

where \( P_{0.00135} \), \( P_{0.5} \), \( P_{0.99865} \) respectively means the data in the specific distribution of the percentiles 0.135%, 50%, 99.865%, \( T_L, T_U \) respectively for the lower and the upper limit.

[JMP solution]
Step 1: Find the best corresponding distribution by JMP through “analysis-distribution-continuous fitting-all”. The best fitting for Johnson Su distribution
Step 2: In the “fit Johnson Su” choose “Quantile”, and then enter up three probabilities 0.00135, 0.99865, 0.5 respectively. Then output, JMP9.02 shows: \( X_{0.00135} = 4.2749 \), \( X_{0.5} = 4.314337 \), \( X_{0.995} = 4.353767 \). By formula (3), (4): \( C_p = 1.268 \), \( C_{pk} = 0.904 \).

(3) A new formula for calculating process capacity of variables data. Reference [1] shows that under the one dimensional normal distribution, the measurement data process capability index do not obey any particular distribution. The new formula shows as follows:

\[
C_p = \frac{T_U - T_L}{X_{0.995} - X_{0.005}}
\]
\[ C_{pk} = \min \left\{ \frac{T_U - X_{0.5}}{X_{0.995} - X_{0.5}}, \frac{X_{0.5} - T_L}{X_{0.5} - X_{0.005}} \right\} \]  \tag{6}

\( X_{0.005}, X_{0.5}, X_{0.995} \) means the data in the percentile 0.5\%, 50\%, 99.5\% regardless of any specific distribution. JMP9.02 shows: \( X_{0.005}=4.276 \), \( X_{0.5}=4.32 \), \( X_{0.995}=4.341 \). By the formula (5), (6): \( C_p=1.538 \), \( C_{pk}=\min\{1.428571, 1.59\}=1.428571 \)

Method 3 Non parametric method. Assume that the impurity content quality standard allows the upper limit of 4.35, the lower limit is 4.25, process capability analysis for example.

[Minitab Solution]
1 Copy ncap.mac into the Macro directory
2 Enter “% ncap c1 4.35 4.25” in the Edit > edit command editor
3 Click submit button
Output-answer:

**Data Display**

| CNP 1.57406 | CNPK 1.44543 |
| Non parametric capability index CNP | The upper non parametric ability index CNPU |
| CNPK 1.63647 | CNPK 1.44543 |
| The lower non parametric ability index CNPL | Bilateral non parametric capability index CNPK |

**Conclusion**

After the comprehensive comparison of wide abnormal distribution capacity analysis, the best results are known as: \( C_p=1.538 \), \( C_{pk}=1.428571 \); You can also use \( C_p=1.574 \), \( C_{pk}=1.44543 \). The better software algorithms are method 2 and 3.

**Acknowledgements**

This work was financial support by Top-notch Academic Programs Project of Jiangsu Higher Education Institutions (TAPP).

**Reference**


