Method of using of the Box-Cox transformation at the application of the xbar and s chart

Nicolae Eftimie1,*

1Transilvania University of Brasov, Department of Manufacturing Engineering, Mihai Viteazu No.5, Brasov, Romania

Abstract. The application of the most statistical process control techniques is based on the assumption that the distribution of the measurements is normal. However, there are many situations in practice when the process data distribution is not normal. In certain cases, the Box-Cox transformation can be used for converting the process data distribution into a normal distribution. Considering these aspects, the paper presents a method of application for the xbar and s chart that can be used in the case when the measurements distribution is not normal. The proposed method consists of the following stages: the testing of normality of the process data, the application of the Box-Cox transformation and the testing of normality of the transformed data. In the case when the distribution of the transformed data is normal, they are used at the application of the xbar and s control chart.

1 Introduction

The most of control charts that are used in practice are based on the assumption that the measurements distribution is normal. A tool often used in practice for testing this assumption is the Anderson-Darling normality test. A description of the procedure of application for this test is given in [1].

However in practice there are situations when the measurements are not normally distributed. One approach that can be applied in these cases is the transformation of the process data in order to obtain a set of normally distributed values [2].

Box and Cox [3] developed a method that can be used for converting the process data distribution into a normal distribution. According to this method, each measurement $x_i$ will be transformed by means of the following system of equations:

$$y_i = \begin{cases} x_i^\lambda & \lambda \neq 0 \\ \ln(x_i) & \lambda = 0 \end{cases}$$

The $\lambda$ parameter is determined as follows:

- The standardized values of the transformed data are calculated using the system of equations (2):

* Corresponding author: eftimien@unitbv.ro

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).
\[ z_i = \begin{cases} x_i^\lambda - 1 & \lambda \neq 0 \\ \lambda \cdot G^{\lambda-1} & \lambda = 0 \end{cases} \]  

where \( G \) represents the geometric mean of the process data.

- It is calculated the standard deviation for the distribution of the \( z_i \) variable, by means of the equation (3):

\[ \hat{\sigma}_z = \frac{\bar{x}}{c_4} \]  

where the \( c_4 \) coefficient is computed using the relation (4) [2]:

\[ c_4 = \sqrt{\frac{2}{n-1} \frac{\Gamma \left( \frac{n}{2} \right)}{\Gamma \left( \frac{n-1}{2} \right)}} \]  

where \( n \) is the size of the sample.

- It is determined the value of \( \lambda \) as being that for which the standard deviation of the \( z_i \) variable distribution is minimal.

\section*{2 Problem formulation}

In the case when the measurements distribution is not normal, the procedure of application for the \( \bar{x} \) bar and \( s \) chart becomes difficult.

The objective of the paper is to develop a method of application for this type of control chart in the cases when the distribution of the process data can be converted into a normal distribution by means of the Box-Cox transformation.

The method consists of the following stages:

- The testing of normality of the process data. If the measurements distribution is normal, they can be used at the control chart application, otherwise the data must be transformed.
- The application of the Box-Cox transformation.
- The testing of normality of the transformed data.

In the case when the distribution of the transformed data is normal, they are used at the application of the \( \bar{x} \) bar and \( s \) control chart, otherwise another method must be applied for the control chart application.

\section*{3 Problem solution}

The proposed method of application of the \( \bar{x} \) bar and \( s \) chart using the Box-Cox transformation is based on a computer program, which has been developed in Excel 2013 and Visual Basic for Applications for Excel 2013.

Each stage of the proposed method is performed on a separate worksheet.

The paper presents the operation of the program by means of an example.

The input data for the presented example were generated by computer according to a Weibull distribution, with the shape parameter \( \beta = 2.2 \), scale parameter \( \eta = 2.5 \) and threshold parameter \( \gamma = 150 \).

The first worksheet of the application, which is labelled “Data”, contains a table that is
used by the program user for recording the measurements (Figure 1).

Each sample may contain up to 10 measurements. In the presented example, the size of
the sample is equal to 5.

![Fig. 1. The Input data area.](image1)

The second worksheet of the program, which is labelled “AD”, is used by the program for
applying the Anderson-Darling normality test for the process data according to the theoretical
notions presented in [1]. The values for the number of measurements N, the average xbar and
the standard deviation s are calculated for the measurements recorded in the worksheet
labelled “Data” (Figure 2).

![Fig. 2. The Anderson-Darling test for the process data.](image2)

The user of the program must click the control button labelled “Test of normality for
Process data”. As a result of this action, the procedure A_D_x() will be run. The listing for
the first part of this procedure is presented by Figure 3.

```vba
Sub A_D_x()
    Dim N As Integer, I As Integer
    Dim xbar As Double, s As Double
    Dim A2 As Double, A2M As Double, P As Double
    Dim x() As Double, F() As Double
    N = Range("B1").Value
    xbar = Range("B2").Value
    s = Range("B3").Value
    For I = 1 To N
        ReDim Preserve x(I)
        x(I) = Application.Small(Sheets(1).Range("$A:$J"), I)
        ReDim Preserve F(I)
        F(I) = Application.Norm_Dist(x(I), xbar, s, 1)
    Next I
```

![Fig. 3. The procedure A_D_x() – Part 1.](image3)

After the declaration of the variables, the procedure collects from the worksheet the
values for the number of measurements $N$, the average $xbar$ and the standard deviation $s$. By means of a `FOR` structure, the measurements are ordered and for each of them, the cumulative distribution function is calculated. The measurements and the corresponding values of the cumulative distribution function are recorded into two arrays.

The listing for the second part of the procedure $A\_D\_x()$ is presented in Figure 4.

![Fig. 4. The procedure A\_D\_x() – Part 2.](image)

First, it is calculated the value of $A^2$, which is used for computing the value of $A^{*2}$, denoted into the procedure by $A2M$. The next stage is the calculation of the $P$-value, which depends on the value of $A^{*2}$. Using the $P$-value, the procedure makes the decision about the normality of the process data distribution.

The third worksheet of the program, which is labelled “BoxCox”, is used by the program for transforming the process data. The first columns, from $A$ to $J$, are used for collecting the measurements from the worksheet “Data”. By means of the formula (2), there are calculated the standardized values of the transformed data $z_i$, for different values of the $\lambda$ parameter.

For each sample, the value of the standard deviation, $s_2$, is computed (Figure 5).

![Fig. 5. The worksheet BoxCox – Part 1.](image)
The standard deviation \( \sigma_z \) for the distribution of the variable \( z_i \) is calculated according to formula (3), by means of the values of \( \bar{s}_z \) and \( c_4 \) (Figure 6).

\[
=\sqrt{2/(X2-1))^{\text{GAMMA}(X2/2)}/\text{GAMMA}((X2-1)/2)}
\]

\[
=\text{AVERAGE(U:U)}
\]

\[
=\text{COUNT(Data!A2:J2)}
\]

\[
=W2/Y2
\]

Fig. 6. The worksheet BoxCox – Part 2.

The value of the \( \lambda \) parameter is calculated by means of the procedure \texttt{Calculate\_Lambda()}. The listing of this procedure is presented by Figure 7.

```
Sub Calculate_Lambda()
    Dim Lambda As Double, temp As Double
    Dim sigma_min As Double
    Lambda = -5
    Range("X5").Value = Lambda
    sigma_min = Range("W5").Value
    temp = -4.99
    Do While temp <= 5
        Range("X5").Value = temp
        If Range("W5").Value < sigma_min Then
            sigma_min = Range("W5").Value
            Lambda = temp
        End If
        temp = temp + 0.01
    Loop
    Range("W9").Value = sigma_min
    Range("X9").Value = Lambda
End Sub
```

Fig. 7. The procedure \texttt{Calculate\_Lambda()}. This procedure will repeat all the calculations on the worksheet “BoxCox” for the \( \lambda \) parameter taking values into the interval \([-5, 5]\) with a step value of 0.01. In this manner, the value of \( \lambda \) that correspond to the minimal value of \( \sigma_z \) will be found. By means of this value, the transformed data are calculated using the formula (1) (Figure 8).

\[
=\text{IF(ISNUMBER(Data!A2)},\text{IF(S1S2=0,LN(Data!A2),Data!A2^S1S2),""})
\]

Fig. 8. The transformed data.
The distribution of transformed data will not always result as being normal. Therefore, the next stage that must be performed, consists of application of normality test for the transformed data. This operation will be achieved in the same manner as the test of normality for the process data (Figure 9).

![Table](image)

**Fig. 9.** The test of normality for the transformed data.

For the presented example, the distribution of the transformed data has resulted as being normal.

Therefore, in this case, the classic procedure of application of the xbar and s chart can be applied for the transformed data. This operation is performed on the worksheet labelled “Chart” (Figure 10).

![Equations](image)

**Fig. 10.** The calculations that are needed for the representation of the control chart.

The columns from B to K contain the transformed data, which are used for calculating the points of the chart, as well as the central lines and the control limits. The control chart is presented by Figure 11.

![Graphs](image)

**Fig. 11.** The xbar and s chart for the transformed data.
It can be observed that for the presented example the process is under control. The process capability analysis is presented by Figure 12.

![Process Capability Analysis Diagram]

**Fig. 12.** The process capability analysis.

The program uses the value of the $\lambda$ parameter and the specification limits for calculating the transformed specification limits: $LSL_{TR}$ and $USL_{TR}$. For the presented example, the value of the $\lambda$ parameter is less than 0, therefore the value of $LSL_{TR}$ is greater than the value of $USL_{TR}$.

In the next stage, the program collects from the worksheet labelled “Chart” the values of the center lines: $ydlbar$ and $sbar$, as well as the size of the sample, $n$. The value of $c_4$ coefficient is calculated according to the formula (4).

After the calculation of the standard deviation, $\sigma$ the program computes the values of the capability indices for the transformed data: $Cp_{TR}$, $CPL_{TR}$, $CPU_{TR}$ and $CPK_{TR}$.

The program also calculates the proportions of nonconforming items that correspond to the transformed data: $p_{LTR}$, $p_{UTR}$ and $p_{notTR}$.

### 4 Conclusions

The presented method can be used in plant practice for controlling the processes in the case when the process data distribution is not normal, but can be converted into a normal distribution by means of the Box-Cox transformation.

The method is applied by means of a computer program which was developed so that the users do not have to possess computer programming knowledge.

Each stage of application of the proposed method is implemented on a separate worksheet.

The user of the program must record the measurements of the samples and all the calculations are performed by the program.

In order to assess the normality of process data distribution, the program uses a procedure developed on the base of the Anderson-Darling test. If the measurements distribution is normal, they can be used at the control chart application, otherwise the data will be transformed by means of a procedure developed on the base of the Box-Cox transformation.
The Anderson-Darling normality test is applied again, this time for the transformed data. In the case when the distribution of the transformed data is normal, they are used at the application of the \textit{xbar} and \textit{s} control chart, otherwise another method must be applied for the control chart application.


c\textbf{References}

3. G.E.P. Box, D.R. Cox, J ROY STAT SOC B MET, 26, 211 (1964)