The use of statistical transformation in six sigma analysis

Mustafid, I Yulvia, D Ispriyanti, Sugito, A R Hakim

Department of Statistics, Faculty of Science and Mathematics, Diponegoro University, Jl. Prof. Soedarto, SH, Tembalang, Semarang 50275, Indonesia

Corresponding author: mustafid55@gmail.com

Abstract. The basic assumptions in six sigma analysis are use data assumptions from the quality characteristic variables in-control and normal distributed. The use of normal distribution assumptions aims to facilitate the computational and analytical processes. In fact, there are many observational data obtained from industry which are uncontrolled and not normally distributed. The aim of the research is to use exponentially weight moving average (EWMA) as statistical transformation to make data uncontrolled and not normally distributed into data in-control and normal distributed. This research uses a case study with empirical data Weibull distributed, and use the statistics of EWMA to the data can be transformed into in-control and normal distributed. Based on the results of the transformation from empirical data, the EWMA control chart were made to determine the data position of in-control and also measured six sigma values.

1. Introduction

Industry and business competition is getting wider and tighter, so that every company strives to create or produce quality products as expected by consumers. The increasing use of advanced technology and well-established resources, each company can produce better quality products or produce smaller defective products. So that the observational data on defective data are more likely to be positively-skewness, or in a statistical perspective, the observational data tends not to be normally distributed.

Statistical control charts are widely used in the industry by using observational data that are not normally distributed [1]. In six sigma analysis, many articles propose formulas for performance measurement and put forward flowcharts for calculating the performance index from normal and non-normal data. Another approach to making a normal distribution uses a transformation in such a way as to meet normality conditions [2]. The Six Sigma performance evaluation approach without assuming a normal probability distribution is carried out [1]. They evaluate the performance of Six Sigma using the exponential distribution, Gamma and Weibull. Other articles discussing six sigma analysis with the Weibull distribution, among others, are by [3] and [4].

EWMA as a statistical transformation has been widely used as an approach to data analysis that is not normally distributed [5]. The use of the t distribution and Gamma distribution are used for analysis with the robust theory on EWMA control charts for abnormal data [6]. In this study, Monte Carlo simulation was used to obtain the appropriate smoothing parameter and distance between the control limits for EWMA control charts.

One of the main analyses in statistical quality control is the six sigma analysis. The analyst uses the assumption that what is used must fulfill the assumption that the data is in an in-control position or is stable and normally distributed. Control and supervision is the application of quality control statistical methods for activities carried out, which ensure that production and operation activities are carried out in accordance with what is targeted, or specifications are in accordance with management targets. If
there is a deviation, then the deviation can be corrected to take action so that the target can be achieved. The aim of the research is to use statistical transformation EWMA to make data out-of-control and not normally distributed into data in-control and normally distributed. This research uses a case study with empirical data Weibull distributed, and uses the statistics of EWMA to the data can be transformed into in-control and normally distributed. The results of the transformation from observational data, the EWMA control chart were designed to determine the data in-control and to derive level six sigma as the quality of the product or process of production.

2. Research methods

The research method uses statistical transformation EWMA for six sigma quality control analysis for data that does not meet the in-control assumptions and is not normally distributed. The use of EWMA transformation aims to obtain observational data into a symmetrical normal distribution or normal distribution with zero-skewness. Hypothesis testing of the normal distribution is done by using the Darling Anderson test. In this case, the optimal smoothing parameter in EWMA is determined, that is, the EWMA transformation results in a symmetrical normal distribution with the largest p-value.

This research uses case study of the barecore wood industry in Central Java, Indonesia. The data used are monthly data from production data and defect data for 33 months from 2017 to 2019. Barecore is type of processed wood which is the result of combining small bundles and in the form of sticks / blocks of a certain size with a standard unit of 1 barecore sheet which is 0.038695 m³ / sheet. To find out the distribution of the observational data, a hypothesis was tested with various alternative distributions, namely normal, Weibull, Gamma, and exponential distributions. The computation of observational data and transformation data was carried out using Microsoft Excel, SPSS, and Minitab software.

After the observational data meets the normal distribution assumptions, then EWMA control chart is designed to determine the optimal control limit distance using the ARL value. The EWMA control chart is used to find out whether the transformed data is in the in-control position or not. After it is known that the transformed data is in an in-control position and is normally distributed, six sigma analysis are carried out, which begins with determining the Defect Per Million Opportunities (DPMO). Furthermore, the DPMO value determines the sigma level as a measure of the quality or performance of the industry or business.

3. Six sigma analysis

Six sigma consists of 2 words, namely six which means 6, and sigma which is a symbol of the standard deviation of statistics. Six sigma is defined as structured management methodology for improving processes by focusing on efforts to minimize variations that occur while reducing defects or products or services that do not comply with standards. Six sigma is also a measure of sigma level describes ability of process or process quality to produce the desired product. The integration between management and statistical aspects, Six Sigma is a vision of improving quality towards the target of 3.4 failures per 1000000 (DPMO) which means that the industry or company produces products with a customer satisfaction level of 99.9997%.

Six sigma analysis uses assumptions that the observational data are normally distributed and in stable or in control conditions. For the assumption of the process in a stable condition, it can be shown using the EWMA control chart which is designed using the optimal smoothing value λ and the appropriate control limit distance value L. The application of the six sigma analysis begins with determining the number of defect products. A defect product is a failure in the sense that it does not meet the product quality requirements as desired by the customer. Defect per opportunities is a measure of failure calculated in the production of quality improvement, which shows the number of defects in failure per one opportunity, where as for DPMO is the amount of DPO multiplied by 1000000, or the number of defects in 1000000 products. In calculating the DPMO value for attribute data and variable data has a different calculation. The DPMO calculation uses the assumption that the data used is in-control and normally distributed. DPMO calculations use formulas (1) and (2).
DPMO = \( 1 - \Phi(\frac{USL - \bar{x}}{s}) \times 1000000 \) for USL \\
= \( 1 - \Phi(\frac{\bar{x} - LSL}{s}) \times 1000000 \) for LSL \hspace{1cm} (1) \\
wheredef(\cdot) is cumulative distribution function of normal standard:

\[
\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}w^2} dw
\]

Furthermore, the sigma level for variable data is obtained with Motorola Six Sigma, which allows a shift in the value around the process mean of ± 1.5σ, so that DPMO can be determined by DPMO = \( 1 - \Phi(k - 1.5) \times 1000000 \). Based on formulas (1), (2) and standard normal inversion \( \Phi(k - 1.5) \), the value of the \( k \) sigma level can be obtained by formula (3):

\[
k = NORMSINV(1 - \frac{DPMO}{1000000}) + 1.5
\hspace{1cm} (3)
\]

The smoothing value \( \lambda \) is determined on the condition that the EWMA designed is a statistic that has a symmetrical normal distribution, that is, with the largest p-value. The p-value can be obtained from Minitab or SPSS using the Darlin Anderson test. The value of length \( L \) is determined on the condition that the in-control value of Average Run Length (ARL) is large, or the out-of-control data is large. For 3 sigma, the optimal weighting value (\( \lambda \)) and control limit distance (\( L \)) are determined using the values \( ARL_0 = \frac{1}{p} = \frac{1}{0.0027} = 370.4 \) \[7\]. \( ARL_0 \) is used to determine the optimal smoothing value (\( \lambda \)) in the form of an EWMA control chart based on the standard deviation of the data. The control limits of the EWMA control chart are obtained based on mean (5) and variance (6). So that the EWMA control charts can be defined by formulas (7), (8) and (9):

\[
UCL = \mu_Y + L \sigma_Y \\
= \mu + L \sigma \sqrt{\frac{(1-(1-\lambda)^2)}{2-\lambda} \hspace{1cm} (7)}
\]

\[
CL = \mu \hspace{1cm} (8)
\]

4. EWMA transformation
Statistical EWMA is used to transform observational data that are not in-control or are not normally distributed into in-control data and are normally distributed. Furthermore, EWMA can also be used for control chart design that are built from the results of EWMA transformations. The control chart has three lines, the upper control limit (UCL), the lower control limit (LCL) which is the limit for a deviation that can still be tolerated and the centre line is the average value of the quality characteristics related to controlled conditions. The process is said to be in control if all the observed data are between the control limits of the UCL and LCL.

Let \( X_i, i = 1, 2, 3, \cdots \) are independent and identically distributed observational data with mean \( \mu \) and standard deviation \( \sigma \). In this case \( \mu \) and \( \sigma \) are known.

The EWMA for \( Y_i \) is defined by the formula (4):

\[
Y_i = \lambda X_i + (1 - \lambda)Y_{i-1} \hspace{1cm} for \ i = 1, 2, 3, \cdots, \hspace{1cm} Y_0 = \mu \hspace{1cm} (4)
\]

where \( \lambda \) is smoothing parameter pembobot with value \( 0 < \lambda \leq 1 \). The exact mean and variance of EWMA \( Y_i \) are given in formulas (5) and (6):

\[
\mu_Y = E[Y_i] = Y_0 = \mu \hspace{1cm} (5)
\]

\[
\sigma_Y^2 = \text{var}(Y_i) = \lambda \sigma^2 \frac{(1-(1-\lambda)^2)}{(2-\lambda)} \hspace{1cm} (6)
\]

The smoothing value \( \lambda \) is determined on the condition that the EWMA designed is a statistic that has a symmetrical normal distribution, that is, with the largest p-value. The p-value can be obtained from Minitab or SPSS using the Darlin Anderson test. The value of length \( L \) is determined on the condition that the in-control value of Average Run Length (ARL) is large, or the out-of-control data is large. For 3 sigma, the optimal weighting value (\( \lambda \)) and control limit distance (\( L \)) are determined using the values \( ARL_0 = \frac{1}{p} = \frac{1}{0.0027} = 370.4 \) \[7\]. \( ARL_0 \) is used to determine the optimal smoothing value (\( \lambda \)) in the form of an EWMA control chart based on the standard deviation of the data. The control limits of the EWMA control chart are obtained based on mean (5) and variance (6). So that the EWMA control charts can be defined by formulas (7), (8) and (9):

\[
UCL = \mu_Y + L \sigma_Y \\
= \mu + L \sigma \sqrt{\frac{(1-(1-\lambda)^2)}{2-\lambda} \hspace{1cm} (7)}
\]

\[
CL = \mu \hspace{1cm} (8)
\]
\[ LCL = \mu_{y_j} - L\sigma_{y_j} \]
\[ = \mu - L \sigma \sqrt{\lambda \frac{(1-(1-\lambda)^2)}{2-\lambda}} \]  

(9)

Based on formula (7), (8), and (9), EWMA control charts can be derived by on the parameters of observation data, that is parameter mean \( \mu \) and standard deviation \( \sigma \) form the variables X.

5. Six sigma analysis for non normal distribution

This six sigma analysis uses case study of the barecore wood industry with monthly observation data from production data and defect data for 33 months from 2017 to 2019. The target or quality specification set by management is the number of defects in the amount of 0.35% of the total production. Based on the total production of observational data is 618175, the specification value is 0.35% \( \times \) 618175 = 2164. The specification value of 2164 is carried out by the EWMA transformation, the value of the specification of transformation is 1800. The profile for data of the observation and the results of the EWMA transformation is given in Figure 1. Figure 1 shows that the transformed data is smoother than the original observed data, and the two data remain at approximately the original mean.

![Figure 1. The profile of the observational data and EWMA transformed](image)

![Figure 2. Probability plot for observational data.](image)
The value of Six Sigma Analysis begins with performing a hypothesis test on the observational data using Minitab. Data processing is performed to determine the type of distribution of defect data for normal, exponential, Weibull, and Gamma distributions. The result of Goodness of Fit Test is given in the Probability plot (Figure 2) that the corresponding observation data distribution is the Weibull distribution with p-value > 0.25 with AD = 0.346. So that the test results have provided sufficient evidence that the defect data of wood production follows the Weibull distribution with the shape parameter \(\beta = 1.96988\) and scale parameter \(\theta = 1554.26992\). The mean and variance of the Weibull distribution are derived as the formulas [1]:

\[
\mu = E(X) = \theta r\left(1 + \frac{1}{\beta}\right)
\]
\[
\sigma^2 = E(X - \mu)^2 = \theta^2 r\left(1 + \frac{2}{\beta}\right) - r^2 \left(1 + \frac{1}{\beta}\right)
\]

Based on those formulas, the mean and standard deviation of the observed data are:

\[
\mu = 1554.26992 r\left(1 + \frac{1}{1.96988}\right) = 1377.858
\]
\[
\sigma^2 = 1554.26992^2 r\left(1 + \frac{2}{1.96988}\right) - r^2 \left(1 + \frac{1}{1.96988}\right)
\]
\[
\sigma = \sqrt{533113} = 730.146
\]

Based on observational data that are not normally distributed, the EWMA transformation was carried out using formula (4). Obtained smoothing parameter \(\lambda = 0.2\) as the optimal parameter. The data obtained from the transformation is normally distributed with p-value = 0.50. The EWMA control chart is formed using the EWMA transformation with the optimal smoothing parameter \(\lambda\), namely \(\lambda = 0.2\). Based on the smoothing parameter \(\lambda\) optimal \(\lambda = 0.20\), then determined the value of the distance control limit \(L = \) with the value of ARL = 370.4, for p-value = 0.0027 or at level 3\(\sigma\). Using smoothing parameter \(\lambda = 0.20\) and the distance control limit \(L = 2.856\) [6], and formulas (8), (9), and (10), the EWMA control chart is given in Figure 3.

Figure 3 shows that all data are between USL and LSL, the EWMA transformation data is in-control position. Thus the data assumptions for in-control and normal distribution are fulfilled. The sigma level is calculated using the optimal smoothing \(\lambda\) value and the optimal distance control limit value \(L\). DPMO is calculated using formula (1):

\[
DPMO = (1 - \phi\left(\frac{UCL - \mu}{\sigma}\right)) \times 1000000
\]
\[
= (1 - \phi\left(\frac{2300 - 1378}{730.146}\right)) \times 1000000
\]
Furthermore, with formula (3), the level of sigma is obtained:

\[ k = \text{NORMSINV} \left(1 - \frac{DPMO}{1000000}\right) + 1.5 \]

\[ = \text{NORMSINV} \left(1 - \frac{51634}{1000000}\right) + 1.5 \]

\[ = 3.13 \]

The final result of the case study shows that the level of sigma for the wood production process is \( k = 3.13 \) which indicates that the process is quite good for 3 sigma. The achievement of this sigma level still needs to be improved, by reducing the number of defective products, so that the sigma level \( k = 4 \) can be obtained, which indicates that the quality and process of wood production is getting very good.

6. Conclusions

EWMA statistical transformation can be used as a transformation method in the six sigma analysis for data that is not normally distributed or not in the control position. The data resulting from the EWMA transformation uses the mean from the observational data, and the standard deviation of the transformed data is generated from the standard deviation function of the observed data. Likewise, the EWMA control chart for USL, LSL, and central line values only uses the mean and standard deviation of the observed data. Especially for observational data with a Weibull distribution, the mean and standard deviation values were obtained directly from the Weibull distribution function.

References

[1] Aldowaisan T, Nourelfath M and Hassan J 2015 Eur. J. Oper. Res. 246 968-977
[2] Amiri A A, Bashiri M, Mogouie H and Doroudyan M H 2012 Sci. Iran. 19 6 1894–1905
[3] Hassan J, Aldowaisan T and Nourelfath M 2018 Int. J. Qual. Reliab. Manag. 35 3 1639-1652
[4] Chen F I and Yeh C H 2011 J. Oper. Res. Soc. 62 4 750-759
[5] Chou Y M, Polansky A M and Mason R L 1998 J. Qual. Technol. 30 2 133–141
[6] Alkahtani S S 2013 J. Mod. Appl. Stat. Methods 12 1 148-163
[7] Montgomery D C 2009 Introduction to Statistical Quality Control 6th edition (New York: Wiley)