Nigerian Journal of Mathematics and Applications
Volume B, **29**, (2019), 48 − 69. *Printed by Unilorin press*

http://www.njmaman.com

On the Performance Assessment of Mean Methods in Estimating Process Capability for Skewed Distributions

J. O. Braimah*, S. O. Elakhe and N. Edike

Abstract

This study compares the performances of Gini Mean, Clements and Box-Cox transformation methods for estimating process capability Indices when the distribution of the process data is (skewed) non-normal. The use of process performance index (PPI) is implored for process capability analysis (PCA) using Weibull distribution. Data was simulated using R software with a decision interval (target point) of 1.0 and 1.5. Performance assessment was carried out using Boxplots, descriptive statistics and the root mean square deviation. It is observed that Gini mean difference based process capability indices performs best in estimating the process capability indices closest to a set target for varying distribution parameters at different sample sizes, followed by Clements and lastly, the Box-Cox transformation method.

1. Introduction

Statistical analysis has become a powerful tool for a better market valuation, taking a leading role in the development of new products and services that try to circumvent the increasing amount of risks that an investor has to take. The

Received: 12/09/2019, Accepted: 01/10/2019, Revised: 30/11/2019. * Corresponding author.
2015 *Mathematics Subject Classification.* 68M20, & 97K60.
Key words and phrases. Process Control, Capability Indices, Performance Index, Standard Error, Skewed
Department of Mathematics and Statistics, Faculty of Physical Sciences, Ambrose Alli University, Ekpoma, Edo State, Nigeria.
E-mail: ojbraimah2014@gmail.com

need and demand for the knowledge of statistics is steadily increasing in almost every sector, where young statisticians could develop a professional career (Golafshani, 2003). Statistical Process Control is the application of statistical tools and techniques in monitoring variation in a continuous process in order to detect variations that are of assignable causes, and therefore make recommendations for corrective check on the process. Control charts are used to monitor processes in order to detect assignable cause(s) that change the process parameters. Ding (2004) emphasized the importance of identification of assignable cause. When the distribution of the output quality of the process variable is continuous, the combination of two control charts such as an X-chart and an R-chart are usually required to monitor both the process mean and the process variance (Montgomery and Runger, 2009). However, recently Vännman and Albing (2007) have shown that the two combined charts are not always reliable in identifying the nature of the change. For example, if the joint X and R charts are in use, and if the X -chart signals the presence of an assignable cause, then the user should investigate the type of the assignable cause that is affecting the process parameter(s). This is because an X -chart is not only sensitive to a shift in the process mean, but is also sensitive to an increase in the process variance.

Measuring a process performance and acting upon the assessments based on the measurements are critical elements of any continuous quality improvement efforts (Spiring, 1995), however, companies make assessments of process performance based on different indicators.
Most common of these indicators can be described in terms of process yield, process expected loss and capability indices of a particular process characteristic (Chen, 2001). Among these indicators, Process Capability Indices (PCIs) have gained substantial attention both in academic community and several types of manufacturing industries since 1980s (McCormack Jr, *et al.*, 2000; Kotz and Johnson, 2002 and Wu and Swain, 2009). The first process capability index proposed in the literature more than three decades ago is the $C_p$index, which is defined as:

$$C_p = \frac{USL - LSL}{6} \quad (1)$$

where USL and LSL denote the upper and lower specification limits respectively and is the standard deviation of the process characteristic of interest. The C_p index measures the process spread in relation to its specification range. Since C_p does not take the process mean of the quality characteristic into account, it does not give any information about whether the process is centered (Bordignon and Scagliarini, 2002). In order to overcome this problem, a second generation PCI, the C_{pk} index, is introduced.

The C_{pk} is defined as:

$$C_{pk} = \min\left[\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3}\right] \quad (2)$$

where μ and σ are the mean and the standard deviation of the quality characteristic studied, respectively. The mean of the process characteristic has an influence on the C_{pk} index and therefore it is more sensitive to departures from centrality than the C_p index (Anis, 2008).

The C_p and C_{pk} indices are the most commonly used process capability indices in industry and they are called standard or basic PCIs (Kotz and Johnson, 2002 and Anis, 2008).

P_p and P_{pk} are measures of process performance from a customer perspective. P_p is similar in definition to C_p, and P_{pk} is similar in definition to C_{pk} but in each we use the overall standard deviation of the data (Lovelace and Swain, 2009).

The earlier practical applications of these basic PCIs require the fulfillment of two assumptions:

(i) The process has to be in statistical control.

(ii) The process characteristic of interest has to be normally distributed.

Utilization of the basic PCIs when these underlying assumptions are not satisfied may lead to incorrect assessments and misleading interpretations of process capability (Chen, 2000).

Non-normally distributed processes are not uncommon in practice. Combining this fact with the misleading results of applying basic PCIs to non-normal processes while treating them as normal distributions forced academicians and practitioners to investigate the characteristics of process capability indices with non-normal data (Kotz and Johnson, 2002, Spiring, *et al.*, 2003 and Yum and Kim, 2010).

There are two approaches adopted in estimating PCI for non-normal process situation include:

(i) Data Transformation Approach: Data transformation approach is aimed at transforming the non-normal process data into normal process data. Several methods have been proposed for approximating normally distributed data by using mathematical functions. Most known amongst these methods are Johnson transformation system, which is based on derivation of the moments of the distribution, and Box-Cox power transformation. The main rationale behind these methods is to first transform the non-normal data into normal data and then use standard process capability indices, which are based on the normality assumption (Kotz and Johnson, 2002 and Ding, 2004).

(ii) Distribution Fitting Method for Empirical Data: Distribution fitting methods use the empirical process data, of which the distribution is unknown. These methods later fit the empirical data set with a non-normal distribution based on the parameters of the empirical distribution. Clements' Method is one of the most popular distribution approaches (Kotz and Johnson, 2002). The method employs both anew process capability index based on percentiles and a distribution fitting approach. The basic Cp index is calculated by using the 99.865th and the 0.135th percentiles of the distribution of quality characteristic instead of 6, which is meaningful only when the process is normally distributed. Therefore, the percentile-based Cp is obtained by:

$$C_p = \frac{USL - LSL}{\xi_{0.99865} - \xi_{0.00135}} \tag{3}$$

where $\xi_{0.99865}$ and $\xi_{0.00135}$ denote the upper and lower $0.135th$ percentiles of the process distribution, respectively.
Following the same logic, the C_{pk} index can be obtained using a percentile approach:

$$C_{pk} = \min\left[\frac{USL - \xi_{0.5}}{\xi_{0.99865} - \xi_{0.5}}, \frac{\xi_{0.5} - LSL}{\xi_{0.5} - \xi_{0.00135}}\right] \tag{4}$$

where $\xi_{0.5}$ is the median of the process distribution, which is used instead of the process mean, because the process mean is not indicative of the centrality of a non-normal distribution specially when skewness of the distribution is taken into account (Anis, 2008).
However, it is worthy of notice since, unlike other quantities designed for measuring the dispersion of a random variable, the mean difference is independent of any central measure of localization, which can be seen from its definition.

$$\Delta_1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x - y| dF(x) dF(y) \tag{5}$$

When the random variable X is discrete (a case more often considered) the formula has the form:

$$\Delta_1 = \sum_{i=-\infty}^{i=+\infty} \sum_{j=-\infty}^{j=+\infty} |x - y| p_i p_j \tag{6}$$

The analytic investigation of the discussed characteristic is made difficult because of the absolute value occurring in the formula. However, it facilitates the computations on numerical data, which also concerns, as is well known, the mean deviation.
Wu, PearnL and Kotz (2009) used Gini's mean difference as a measure of variability in computing process capability indices and compare the performance with other process capability indices under low, moderate and high symmetry levels

and concludes that the performance of GMD based estimator of C_p is more robust and gives higher values under high asymmetry and that it also has lower bias and MSE when C_p values equal to 2.00 or higher.

2. Methodology

2.1 Process Capability Indices

The process capability index, the Cp index, which is defined as:

$$C_p = \frac{USL - LSL}{6\sigma} \tag{7}$$

The C_{pk} can be defined as:

$$min\left[\frac{USL - \mu}{3}, \frac{\mu - LSL}{3}\right] \tag{8}$$

where USL and LSL denote the upper and lower specification limits, respectively, and σ is the standard deviation of the process characteristic of interest.
Process Capability relative to one sided specification limit are:

$$C_{pu} = \frac{USL - \mu}{3\sigma} \; Process\ Capability\ relative\ to\ upper\ specification\ limit$$

$$C_{pl} = \frac{\mu - LSL}{3\sigma} \; Process\ Capability\ relative\ to\ lower\ specification\ limit$$

$$P_{pu} = \frac{USL - \mu}{3\sigma} \; Process\ performance\ relative\ to\ upper\ specification\ limit$$

$$P_{pl} = \frac{\mu - LSL}{3\sigma} \; Process\ performance\ relative\ to\ lower\ specification\ limit$$

2.2 Clements Method (CM)

For non-normal Pearsonian distribution (which includes a wide class of "populations" with non-normal characteristics), Chang, Choi and Bai (2002) proposed a method of non-normal percentiles to calculate process capability C_p and process capability for off center process C_{pk} indices based on the mean, standard deviation, skewness and kurtosis.
Under the assumption that these four parameters determine the type of the Pearson distribution curve, Clements utilized the table of the family of Pearson curves as a function of skewness and kurtosis.

Clements replaced 6σ by $(U_P - L_P)$ in the below equation:

(9) $$\text{Cp}= \frac{USL - LSL}{U_P - L_P}$$

where, U_P is the 99.865 percentile and L_P is the 0.135 percentile, for C_{pk}, the process mean U is estimated by median M, and the two 305 are estimated by $(U_P - M)$ and $(M - L_P)$ respectively,

(10) $$\text{Cpk} = \min \left[\frac{USL- \ M}{U_P - M}, \ \frac{M - LSL}{M - L_P}\right]$$

2.2.1 Algorithm for calculating PCIs using Clements method

(i) Obtain the specification limits USL and LSL for a given quality characteristic;
(ii) Estimate sample statistics for the given sample data: sample size, mean, standard deviation, skewness and kurtosis Calculate estimated 0.135 percentile L_P;
(iii) Calculate estimated 99.865 percentile U_P;
(iv) Calculate estimated median M; and
(v) Calculate non-normal process capability indices using equations.

(11) $$\text{Cp}=\frac{USL - LSL}{U_P - L_P}$$

$$\frac{\text{USL}- \ \text{M}}{U_P - M}, \ \frac{\text{M} - \text{LSL}}{M - L_P}$$

(12) $$\text{Cpu} =\frac{\text{USL}- \ \text{M}}{U_P - M}$$

(13) $$\text{Cpl} = \frac{\text{M} - \text{LSL}}{M - L_P}$$

2.3 Box-Cox power Transformation (BCT)

The Box-Cox transformation was proposed by Box and Cox in 1964 and used for transforming non-normal data (Hoerl and Snee, 2010). The Box-Cox transformation uses the parameter λ. In order to transform the data as closely as possible to normality, the best possible transformation should be performed by selecting the most appropriate value of λ. In order to obtain the optimal λ value, Box-Cox transformation method requires maximization of a log-likelihood function. After the transformation, process capability can be evaluated. They proposed a useful family of power transformations on the necessarily positive response variable X.

(14) $$X^{(\lambda)} = \begin{cases} \frac{X^{\lambda}-1}{\lambda} \ , \ for \ \lambda \ \neq 0 \\ \ln X \ , \ for \ \lambda = 0 \end{cases}$$

where the variable X takes positive values. If the variable X takes negative values, then a constant value will be added in order to make the values positive. This continuous family depends on a single parameter λ that can be estimated by using maximum likelihood estimation.
Firstly, a value of λ from a pre-assigned range is collected. Then $L_{\max}$ is computed as in

$$L_{max} = -\frac{1}{2}\ln\widehat{\sigma}^2 \; + \; \ln J\left(\lambda, X\right) = -\frac{1}{2}\;\ln\widehat{\sigma}^2 \; + \left(\lambda - 1\right)\sum_{i=1}^{n} \ln X_i \tag{15}$$

For all $\lambda, J\left(\lambda, X\right)$ is evaluated as in Equation

$$J\left(\lambda, X\right) = \prod_{i-}^{n} \frac{\partial W_i}{\partial X_i} = \prod_{i=1}^{n} X_i^{\lambda-1} \tag{16}$$

$$\ln J\left(\lambda, X\right) = \left(\lambda - 1\right)\sum_{i=1}^{n} \ln X_i \tag{17}$$

For fixed λ, σ^2 is estimated by using $(S\lambda)$, which is the residual sum of squares of $X^{(\lambda)}.\sigma^2$ is estimated by the formula in the equation below:

$$\widehat{\sigma}^2 = \frac{S(\lambda)}{n} \tag{18}$$

When the optimum value is obtained, for all the quality characteristics values of X, the upper and the lower specification limits are transformed to normal variables (Spiring, Leung, Cheng and Yeung, 2003). The classic formula for the computation of process capability indices under normal assumption is used.

2.4 Gini's Mean Difference (GM)

The Gini's mean difference for a set of n ordered observations, $\{x_1, x_2, \cdots, x_n\}$, of a random variable X is defined as:

$$\text{Gn}=\frac{2}{n(n-1)}\sum_{j=1}^{n}\sum_{i=1}^{n} |x_i - x_j| \tag{19}$$

$$\text{Gn}=\frac{2}{n(n-1)}\sum_{i=1}^{n} \left[(x_i - x_1) + (x_i - x_2) + \cdots + (x_i - x_{n-1})\right] \tag{20}$$

$$\text{Gn}=\frac{2}{n(n-1)}\sum_{i=1}^{n} \left(2i_i - n - 1\right) x_{(i)} \tag{21}$$

If the random variable X follows normal distribution with mean μ and variance σ^2, then Yitzhaki (2010) suggests a possible unbiased estimator of standard deviation (σ) as:

$$\sigma = \text{c}\ \frac{\sum_{i=1}^{n}\left[(2i_i - n - 1)\, x_i\right]}{n(n-1)} \tag{22}$$

where $c = \sqrt{\pi} = 1.77245, \sigma = 0.8862$, G_n is an unbiased measure of variability. Gini's mean difference can be rewritten as:

$$\text{Gn=}\frac{2}{n(n-1)}\sum_{i=1}^{n}(2i - n - 1)\, x_{(i)} \tag{23}$$

If we write this as:

$$\text{Gn=}\ \frac{2}{n(n-1)}\sum_{i=1}^{n}[(i-1) - (n-1)]x_{(i)} \tag{24}$$

$$\text{Gn} = \frac{2}{n(n-1)}\left[\sum_{i=1}^{n}(i-1)\, x_{(i)} - \sum_{i=1}^{n}(n-1)\, x_{(i)}\right] \tag{25}$$

$$\text{Gn=}\frac{2}{n(n-1)}\,[U - V] \tag{26}$$

where $U = \sum_{i=1}^{n}(i-1)x_{(i)}$ and $V = \sum_{i=1}^{n}(n-1)x_{(i)}$ are the estimated means. The unbiased estimator of Gini Mean difference for Weibull distribution is

$$E(\text{Gn}) = \left(2 - 2^{1-\frac{1}{\beta}}\right)\frac{G\left(1+\frac{1}{\beta}\right)}{\lambda} = \sigma_{gw} \tag{27}$$

The Weibull probability density function is given as:

$$f\,(x) =\ \lambda\beta(\lambda x)^{\beta 00000-1}e^{-(\lambda x)^{\beta}} \tag{28}$$

To compute C_p and C_{pk} using Gini's mean difference as a measure of variability when the data follow a Weibull distribution

$$C_{npg} = \frac{USL - LSL}{5.3172\sigma_{gw}} \tag{29}$$

$$C_{npkg} = \frac{\min(USL - m, m - LSL)}{2.6586\sigma_{gw}}$$

$$C_{npug} =\ \frac{USL - m}{2.6586\sigma_{gw}} \tag{30}$$

$$C_{nplg} = \frac{m - LSL}{2.6586\sigma_{gw}} \tag{31}$$

2.5 Performance Comparison of the Clements, Box-Cox transformation and the Gini Methods

The performance comparison is carried out by generating Weibull data through simulation and for this reason, process performance indices (PPIs) are executed for computing process capability rather than process capability indices (PCIs). Weibull distribution is used or modeling most industrial processes especially in reliability field which is concerned with the failure of a product or the time to failure of the product. Only one sided (USL) process performance index P_{pu} is considered. The USL is computed from the equation below using a targeted P_{pu} of 1.0 and 1.5. The targeted P_{pu} of 1.0 is indicating the process is marginally capable of meeting the specifications and the P_{pu} of 1.5 is indicating the process is good and very capable of meeting the specification limits (Montgomery, 2009). Box plots, descriptive statistics, the root-mean-square deviation (RMSD), which is used as a measure of error, are utilized for evaluating the performances of the methods. In addition, the bias of the estimated values is important as the efficiency measured by the mean square error.

2.6 The Root-Mean-Square Deviation (RMSD)

The root-mean-square deviation (RMSD) is used to measure the differences between the targeted P_{pu} values and the estimates obtained by BCT, Clements and Gini mean difference based methods.

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{r}\left(Estimated\ Ppu_i - Targeted\ Ppu_i\right)^2}{r}} \tag{32}$$

where r is the number of data sets generated randomly for each Weibull distribution with specified parameters. The RMSD serves to aggregate the magnitudes of the errors in the predictions for various times into a single measure of predictive power and a measure of accuracy (Hoerl and Snee, 2010).

3. Result and Data Analysis

3.1 The Descriptive Statistics

The tables below show the corresponding quantiles, mean, median along with skewness and kurtosis based on the specified parameter values of Weibull distribution. Kurtosis gives information about the relative concentration of values in the center of the distribution as compared to the tails. Datasets with high kurtosis tend to have prominent peak and heavy tails. Skewness gives information about whether the distribution of the data is symmetrical. The skewness for a normal distribution is zero. The positive skewness values indicate that the distribution is positively skewed, which corresponds that right tail is longer than the left tail, and for negative skewness values it is vice versa. Therefore, it can be stated that kurtosis and skewness give information about tail behavior of a distribution.

Table 1: Summary statistics of Weibull distribution at $\alpha = 1$ and $\beta = 1$ for different sample sizes:

	Weibull(,)	$\mathbf{X}_{0.99865}$	**Median =** $\mathbf{X}_{0.50}$	**Mean**	**Skewness**	**Kurtosis**
n=25	Weibull(1,1)	3.8939	0.7469	1.0296	1.5138	2.7650
n=50	Weibull(1,1)	4.3501	0.6915	0.9989	1.6654	3.2510
n=75	Weibull(1,1)	4.8491	0.6803	0.9733	1.8448	4.7940
n=100	Weibull(1,1)	5.1722	0.7167	1.0130	1.8384	4.6222

Table 2: Summary statistics of Weibull distribution at $\alpha = 1$ and $\beta = 2$ for different sample sizes:

	Weibull(,)	$\mathbf{X}_{0.99865}$	**Median =** $\mathbf{X}_{0.50}$	**Mean**	**Skewness**	**Kurtosis**
n=25	Weibull(1,2)	8.4542	1.4737	2.0537	1.7316	3.9214
n=50	Weibull(1,2)	8.9368	1.4228	2.0194	1.6610	3.7344
n=75	Weibull(1,2)	9.5834	1.3896	2.0026	1.7548	4.1068
n=100	Weibull(1,2)	10.2353	1.4067	2.0281	1.7980	4.2492

Table 3: Summary statistics of Weibull distribution at $\alpha = 2$ and $\beta = 1$ for different sample sizes:

	Weibull(,)	$\mathbf{X}_{0.99865}$	**Median =** $\mathbf{X}_{0.50}$	**Mean**	**Skewness**	**Kurtosis**
n=25	Weibull(2,1)	1.9352	0.8284	0.8743	0.5722	0.3388
n=50	Weibull(2,1)	2.0577	0.8495	0.8916	0.5426	0.0124
n=75	Weibull(2,1)	2.0978	0.8295	0.8774	0.5650	0.0292
n=100	Weibull(2,1)	2.2503	0.8158	0.8719	0.6686	0.3788

Table 4: Summary statistics of Weibull distribution at $\alpha = 2$ and $\beta = 2$ for different sample sizes:

	Weibull(,)	$\mathbf{X}_{0.99865}$	**Median =** $\mathbf{X}_{0.50}$	**Mean**	**Skewness**	**Kurtosis**
n=25	Weibull(2,2)	1.6916	1.7779	1.7770	0.5108	0.1892
n=50	Weibull(2,2)	1.7177	1.5458	1.7537	0.5820	0.1896
n=75	Weibull(2,2)	4.2828	1.6516	1.7703	0.5638	-0.0554
n=100	Weibull(2,2)	4.4739	1.6662	1.7733	0.5898	0.1090

The distribution plot of Weibull distribution for various shape and scale parameter is as shown below:

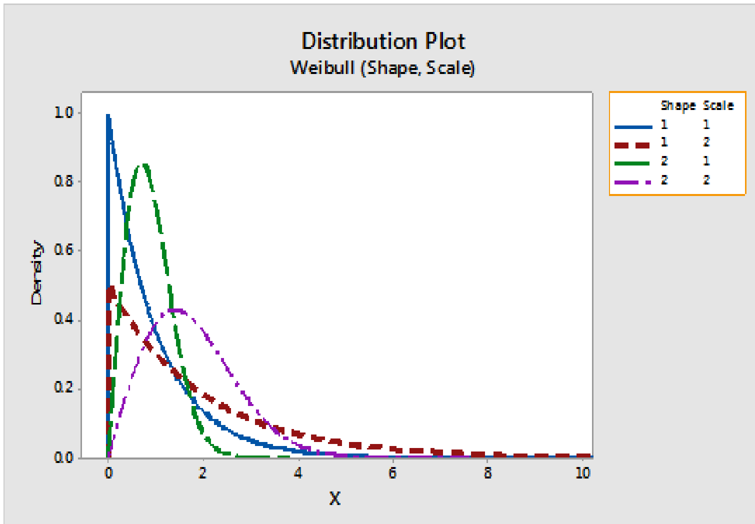


Figure 1: Distribution plot of Weibull distribution using different shape and scale parameters

From the distribution plot in Figure 1, the distribution plots are positively skewed (non-normal) for the combinations of the shape and scale parameters with Weibull $(1, 1)$ the most peaked.

3.2 Process Capability Analysis

3.2.1 Gini Mean Difference based Process Capability Analysis

The table of parameter estimation is given below using the generated data from Weibull distribution with varying shape and scale parameters of (1,1), (1,2) (2,1) and (2,2) at different sample sizes of $n = 25, 50, 75$ and 100.

Table 5: Gini's Estimated USL obtained from the data:

GMD	C_{npug}	**USL FOR GINI**			
		n=25	**n=50**	**n=75**	**n=100**
Weibull(1,1)	1.0	3.4055	3.3501	3.3389	3.3753
	1.5	4.7348	4.6794	4.6682	4.7046
Weibull(1,2)	1.0	3.8298	3.7789	3.7457	3.7629
	1.5	5.0079	4.9570	4.9238	4.9409
Weibull (2,1)	1.0	2.2086	2.2297	2.2096	2.1960
	1.5	2.8986	2.9198	2.8997	2.8861
Weibull (2,2)	1.0	3.1118	3.0668	3.0633	3.0778
	1.5	3.8176	3.7726	3.7691	3.7837

Table 6: Clements's Mean Difference based Process Capability Analysis:

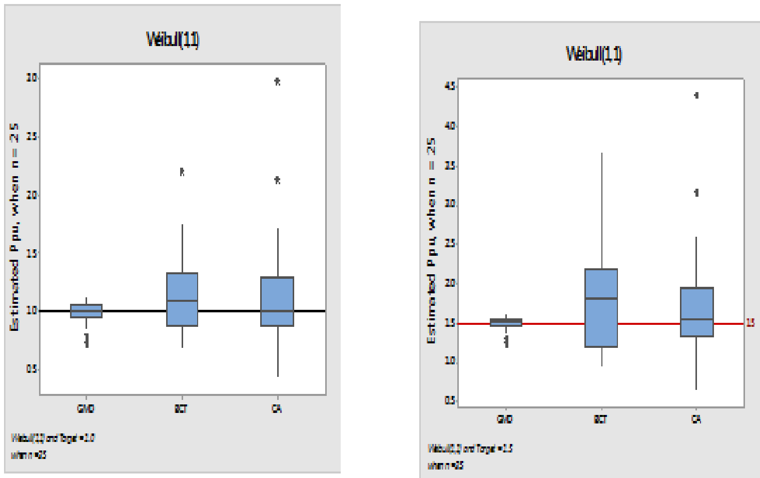
CA	C_{npug}	**USL FOR CLEMENTS ANALYSIS**			
		n=25	**n=50**	**n=75**	**n=100**
Weibull (1,1)	1.0	3.8939	4.3501	4.8491	5.1722
	1.5	5.4674	6.1794	6.9335	7.3999
Weibull (1,2)	1.0	8.4542	8.9368	9.5834	10.2353
	1.5	11.9445	12.6937	13.6802	14.6495
Weibull (2,1)	1.0	1.9352	2.0577	2.0978	2.2503
	1.5	2.4886	2.6618	2.7319	2.9675
Weibull (2,2)	1.0	1.6916	1.7177	4.2828	4.4739
	1.5	1.6484	1.8036	5.5983	5.8777

Table 7: Box-Cox's Mean Difference based Process Capability Analysis:

BCT	C_{npug}	**USL FOR BCT**			
		n=25	**n=50**	**n=75**	**n=100**
Weibull (1,1)	1	2.6189	2.3984	2.0027	1.8314
	1.5	3.7938	3.4562	2.7452	2.3774
Weibull (1,2)	1	3.4239	3.0202	2.5266	2.1994
	1.5	4.7835	4.1252	3.3355	2.8110
Weibull (2,1)	1	1.6913	1.6490	1.6384	1.6528
	1.5	2.2016	2.1041	2.0553	2.0790
Weibull (2,2)	1	2.6310	2.3786	2.3879	2.3661
	1.5	3.3389	2.9853	2.9895	2.9084

3.4 Graphical Comparison of the computed Process Capabilities

In order to compare the process capability methods graphically at each targeted P_{pu} (1.0 and 1.5), box plot or whisker plot is used to show the shape of the distribution, its central value (0.50), variability $(0.75 - 0.25)$ and outliers by star symbol if it exists. The position of the median line in a box plot indicates the location of the values. The figures below shows the comparison:



a.) Weibull (1,1), target Ppu 1.0 and n= 25 b.) Weibull (1,1), target Ppu 1.5 and n=25

Figure 2: Distribution plot of Weibull distribution using different shape and scale parameters

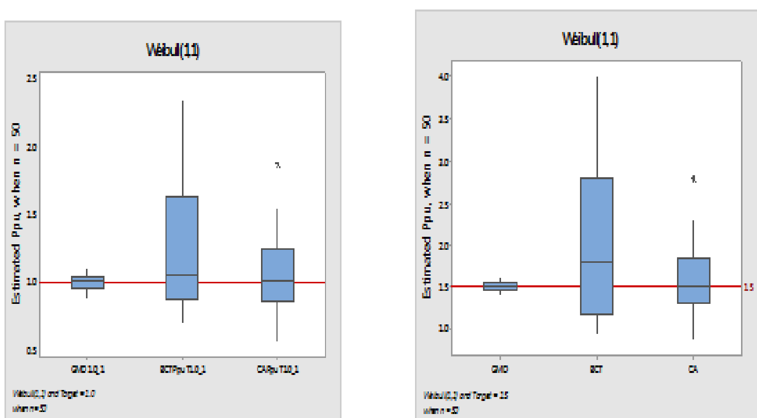


Figure 2: Weibull (1,1), target Ppu 1.0 and n= 50 **Figure 3:** Weibull (1,1), target Ppu 1.5 and n= 50

Figure 3: Distribution plot of Weibull distribution using different shape and scale parameters

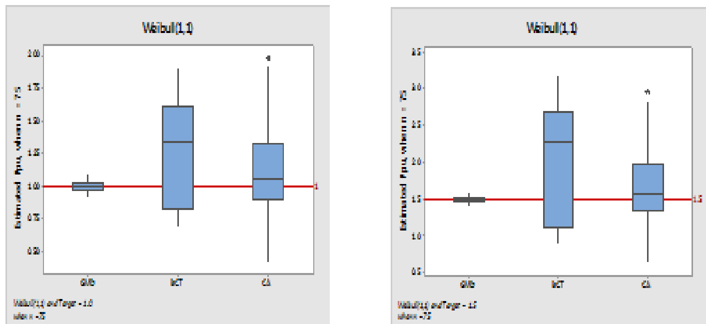


Figure 4: Weibull(1,1), target Ppu 1.0 and n= 75

Figure 5: Weibull(1,1), target Ppu 1.5 and n=

Figure 4: Distribution plot of Weibull distribution using different shape and scale parameters

From the boxplots, the results show that for different distribution parameters at different sample sizes, GMD methods is the best of the three methods for computing process capability for when the process is non-normal.

3.5 Mean and Standard deviation of Computed Capability Indices

To confirm the result shown from the boxplots above, the mean values and the standard deviation (which shows how concentrated the data are around the mean) of the computed process capabilities are computed in the tables below:

Table 8: Descriptive statistics for CA, BCT, and GMD methods when $n = 25$

$n = 25$						
Target Ppu	**Statistics**	**Method**	**Weibull(1,1)**	**Weibull(1,2)**	**Weibull(2,1)**	**Weibull(2,2)**
1.0	Mean	CA	1.1230	1.1930	1.0702	1.1163
		BCT	1.1314	1.0663	1.0125	1.1055
		GMD	1.0000	1.0000	0.9999	1.0000
	Standard Deviation	CA	0.4235	0.5112	0.3542	0.4386
		BCT	0.3097	0.2478	0.2126	0.3483
		GMD	0.0758	0.1706	0.0832	0.1696
1.5	Mean	CA	1.6849	1.7897	1.6191	1.6774
		BCT	1.7820	1.6351	1.5815	1.6583
		GMD	1.5000	1.5000	1.5000	1.5000
	Standard Deviation	CA	0.6275	0.7665	0.5182	0.6465
		BCT	0.6087	0.4404	0.3048	0.4828
		GMD	0.0758	0.1706	0.0832	0.1696

Table 9: Descriptive statistics for CA, BCT, and GMD methods when $n = 50$

$n = 50$						
Target Ppu	**Statistics**	**Method**	**Weibull(1,1)**	**Weibull(1,2)**	**Weibull(2,1)**	**Weibull(2,2)**
1.0	Mean	CA	1.0577	1.1384	1.0262	1.0616
		BCT	1.2337	1.1833	0.9929	1.0356
		GMD	1.0000	1.0000	1.0000	1.0000
	Standard Deviation	CA	0.2644	0.3955	0.1658	0.2691
		BCT	0.4849	0.5377	0.1333	0.1778
		GMD	0.0525	0.1204	0.0603	0.1172
1.5	Mean	CA	1.5856	1.7068	1.5418	1.5930
		BCT	1.9822	1.8171	1.5461	1.5546
		GMD	1.5000	1.5000	1.5000	1.5000
	Standard Deviation	CA	0.3907	1.5000	0.2517	0.4002
		BCT	0.9365	0.8660	0.1953	0.2396
		GMD	0.0525	0.1204	0.0603	0.1172

Table 10: Descriptive statistics for CA, BCT, and GMD methods when $n = 75$

$n = 75$						
Target Ppu	**Statistics**	**Method**	**Weibull(1,1)**	**Weibull(1,2)**	**Weibull(2,1)**	**Weibull(2,2)**
1.0	Mean	CA	1.1044	1.0783	1.0396	1.0223
		BCT	1.2484	1.3056	0.9957	1.0203
		GMD	1.0000	1.0762	1.0000	1.0000
	Standard Deviation	CA	0.3418	0.3200	0.2112	0.1540
		BCT	0.3859	0.5790	0.0945	0.1513
		GMD	0.0459	0.3183	0.0523	0.1025
1.5	Mean	CA	1.6550	1.6168	1.5589	1.5345
		BCT	2.0129	2.0084	1.5295	1.5439
		GMD	1.5000	1.5000	1.5000	1.5055
	Standard Deviation	CA	0.5053	0.4764	0.3071	0.2295
		BCT	0.7514	0.9211	0.1434	0.2088
		GMD	0.0459	0.0965	0.0523	0.1025

Table 11: Descriptive statistics for CA, BCT, and GMD methods when $n = 100$

$n = 100$						
Target Ppu	**Statistics**	**Method**	**Weibull(1,1)**	**Weibull(1,2)**	**Weibull(2,1)**	**Weibull(2,2)**
1.0	Mean	CA	1.0627	1.0552	1.0311	1.0287
		BCT	1.1825	1.2099	0.9890	1.0056
		GMD	1.0000	1.0000	1.0000	1.0000
	Standard Deviation	CA	0.2580	0.2305	0.1834	0.1856
		BCT	0.3095	0.3763	0.0844	0.0865
		GMD	0.0394	0.0707	0.0436	0.0859
1.5	Mean	CA	1.5936	1.5829	1.5476	1.5428
		BCT	1.8554	1.8531	1.5156	1.5081
		GMD	1.5000	1.5000	1.5000	1.5000
	Standard Deviation	CA	0.3850	0.3449	0.2776	0.2700
		BCT	0.5458	0.5991	0.1000	0.1215
		GMD	0.0394	0.0707	0.0436	0.0859

At Weibull (1, 1) and Weibull (1, 2) at sample size of 25, 50, 75 and 100, the Gini Mean Difference based process capability estimates approximately the the target P_{pu} of 1.0 and 1.5, the Clements method estimates is also close to the target P_{pu} while the Box-Cox transformation method is at deviance from the target (overestimated) the Ppu of 1.0 and1.5 as the sample size increases.
At Weibull (2,1) and Weibull (2,2) which indicate low symmetry and at sample size of 25, 50, 75 and 100, the three method estimates are all approximately target P_{pu} of 1.0 and 1.5 with the Gini Mean Difference based process capability estimates the best (closest).

3.6 The Root-Mean-Square Deviation (RMSD)

The root-mean-square deviation (RMSD) is used to measure the differences between the targeted Ppu values and the estimates obtained by Box-Cox Transformation, Clements and Gini mean difference based methods.
The tables below summaries the result obtained for each of the distribution parameter at different sample sizes:

Table 12: The root-mean-square deviations for CA, BCT, and GMD methods when $n = 25$

$n = 25$					
Target Ppu	**Method**	**Weibull(1,1)**	**Weibull(1,2)**	**Weibull(2,1)**	**Weibull(2,2)**
1.0	CA	0.4369	0.5416	0.3576	0.4495
	BCT	0.4794	0.2541	0.2108	0.3606
	GMD	0.0750	0.1688	0.0824	0.1679
1.5	CA	0.6482	0.8122	0.5266	0.6641
	BCT	0.6653	0.4564	0.3125	0.5035
	GMD	0.0750	0.1688	0.0824	0.1679

Table 13: The root-mean-square deviations for CA, BCT, and GMD methods when $n = 50$

$n = 50$					
Target Ppu	**Method**	**Weibull(1,1)**	**Weibull(1,2)**	**Weibull(2,1)**	**Weibull(2,2)**
1.0	CA	0.3749	0.4153	0.1662	0.2734
	BCT	0.5339	0.5629	0.1321	0.1795
	GMD	0.0525	0.1192	0.0597	0.1160
1.5	CA	0.3961	0.6173	0.2526	0.4069
	BCT	1.0048	0.9141	0.1988	0.2434
	GMD	0.0525	0.1192	0.0597	0.1160

Table 14: The root-mean-square deviations for CA, BCT, and GMD methods when $n = 75$

$n = 75$					
Target Ppu	**Method**	**Weibull(1,1)**	**Weibull(1,2)**	**Weibull(2,1)**	**Weibull(2,2)**
1.0	CA	0.3541	0.3263	0.2128	0.1540
	BCT	0.4557	0.6496	0.0937	0.1512
	GMD	0.0455	0.0956	0.0518	0.1015
1.5	CA	0.5237	0.4859	0.3097	0.2298
	BCT	0.9036	1.0440	0.1450	0.2113
	GMD	0.0455	0.0956	0.0518	0.1016

Table 15: The root-mean-square deviations for CA, BCT, and GMD methods when $n = 100$

$n = 100$					
Target Ppu	**Method**	**Weibull(1,1)**	**Weibull(1,2)**	**Weibull(2,1)**	**Weibull(2,2)**
1.0	CA	0.2630	0.2348	0.1841	0.1860
	BCT	0.3566	0.4276	0.0843	0.0858
	GMD	0.0390	0.0700	0.0432	0.0850
1.5	CA	0.3925	0.3514	0.2789	0.2707
	BCT	0.6467	0.6902	0.1002	0.1205
	GMD	0.0390	0.0700	0.0432	0.0850

Results from the root-mean-square deviation (RMSD) in Table 12 to Table 15 shows that the GMD methods have the lowest RMSDs across all the different distribution parameters and sample sizes.

Conclusion

In manufacturing environment, Weibull-distributed quality characteristics are encountered a lot, especially when controlling the process components in terms of times-to-failure. Weibull distributions are known to have significantly different tail behaviour, which greatly affects the process capability. In order to examine the impact of non-normal data, the parameter values of Weibull distribution are specified as (1, 1), (1, 2), (2, 1), and (2, 2) corresponding to (shape, scale) at different sample sizes of 25, 50, 75 and 100. These parameters of Weibull distributions are specified such that the effects of the tail behaviour on process capability could be examined. Principally, when the Weibull shape parameter is equal to 1, Weibull distribution reduces to Exponential distribution. Hence, this study covers all the Exponential family distributions as well.
Conclusively, from our results and findings, the Gini Mean difference based approach is the best among three methods in estimating process capability in skewed (non-normal) situations. In general, methods involving transformation seem more cumbersome in terms of calculation, though it provide estimates of PCIs that truly reflect the capability of the process when there is low symmetry as in Weibull (2, 2).

References

[1] Anis M. Z. (2008). Basic Process Capability Indices: An Expository Review. International Statistical Review. **76** (3), 347-367.
[2] Bordignon S. and Scagliarini M. (2002). Statistical Analysis of Process Capability Indices with
Measurement Errors. Quality and Reliability Engineering International. **18** (4), 321-332.

[3] Chang Y. S., Choi I. S. and Bai D. S. (2002). Process capability indices for skewed populations. Quality and Reliability Engineering International. **18** (5), 383-393.

[4] Chen J. P. (2000). Re-evaluating the Process Capability Indices for Non-normal Distributions. International Journal of Production Research. **38** (6), 1311-1324.

[5] Chen K. S., Huang M. L. and Li R. K. (2001). Process Capability Analysis for an Entire Product. International Journal of Production Research. **39** (17), 4077-4087.

[6] Ding J. (2004). A Method of Estimating the Process Capability Index from the First Moments of Non-normal Data. Quality and Reliability Engineering International. **20** (8), 787-805.

[7] Golafshani N. (2003). Understanding Reliability and Validity in Qualitative Research. The Qualitative Report. **8** (4), 597-607.

[8] Hoerl R. W. and Snee R. D. (2010). Statistical Thinking and Methods in Quality Improvement: A Look to the Futu re. Quality Engineering. **22**(3), 119-129.

[9] Hoerl R. W. and Snee R. D. (2010). Flexible Process Capability Indices. The American Statistician.. **64** (1), 10-14.

[10] Kotz S. and Johnson N. L. (1993). Process Capability Indices. 2nd Ed. Chapman & Hall, Suffolk.

[11] Kotz S. and Johnson N. L. (2002). Process Capability Indices: A review. Journal of Quality Technology. **34** (1), 2-53.

[12] Lovelace C. R. and Swain J. J. (2009). Process Capability Analysis Methodologies for Zero-Bound and Non-normal Process Data. Quality Engineering. **21** (2), 190-202.

[13] McCormack Jr D. W., Harris I. R., Hurwitz A. M. and Spagon P. D. (2000). Capability Indices for Non-Normal Data. Quality Engineering. **12** (4), 489-495.

[14] Montgomery D. C. and Runger G. C. (2009). Applied Statistics and Probability for Engineers. 3Ed. John Wiley & Sons, New York.

[15] Spiring F. A., Leung B., Cheng S. and Yeung A. (2003). A Bibliography of Process Capability Papers. Quality and Reliability Engineering International. **19** (5), 445-460.

[16] Spiring F. A. (1995). Process Capability: A Total Quality Management Tool. Total Quality Management & Business Excellence. **6** (1), 21-34.

[17] Vännman K. and Albing M. (2007). Process Capability Indices for One-Sided Specification Intervals and Skewed Distributions. Quality and Reliability Engineering International. **23** (6), 755-765.

[18] Wu C. W., Pearn W. L. and Kotz S. (2009). An Overview of Theory and Practice on Process Capability Indices for Quality Assurance. International Journal of Production Economics. **117** (2), 338-359.

[19] Wu H. H. and Swain J. J. (2009). A Monte Carlo Comparison of Capability Indices when Processes are Non-normally Distributed. Quality and Reliability Engineering International. **17** (3), 219-231.

[20] Yum B. J. and Kim K. W. (2010). A Bibliography of the Literature on Process Capability Indices: 2000-2009. Quality and Reliability Engineering International. **27** (3), 251-268.

[21] Yitzhaki S. (2010). Gini's mean difference: A superior measure of variability for non-normal distributions. Metron. **61** (2), 285-316.