Process Capability Indices Based on Median Absolute Deviation

Kayode S. Adekeye
Redeemer's University
Department of Mathematical Sciences
Redemption Camp
Mowe, Nigeria.

Abstract

Process capability is the performance of a process under normal and in-control conditions. Its indices are to measure the inherent variability of a process and thus to reflect its performance. To calculate the capability indices for key characteristics, most industries normally assume that the distribution of their process output is normal. However, in practice, most key quality characteristics do fails the normality test and thus the accuracy of normal based process capability indices becomes doubtful and hence they cannot really reflect the performance of a process. This paper proposes the determination of the common process capability indices based on the median absolute deviation. The median absolute deviation is a robust estimate of variability when the sample data are non- normal or are skewed. Real process data and simulated process data from heavily skewed distributions are presented to demonstrate the application of the proposed indices and the results were compared with the existing indices for non-normal and skewed process data.

Key words: capability index, median absolute deviation, non-normal, percent - nonconforming, process data.

1. Introduction

Process capability analysis deals with how to assess the capability of a manufacturing process, where information about the process is used to improve the capability. One can determine how well the process will perform relative to product requirement or specifications. Process capability analysis is to predict how well the process will hold the tolerance. It is very useful for assisting product developers and process developers in selecting or designing product/process, evaluating and selecting among competing vendors and determining the bottleneck process in terms of process quality. Process capability is the performance of a process under normal and in-control conditions. Its indices are to measure the inherent variability of a process and thus to reflect its performance. Process capability studies are conducted and their associated measures determined under the assumption that the process variation is due only to random causes, and are in fact valid only when the process under investigation is free from any special or assignable cause (Spring, 1991). Pan and Wu (1997) presented the following benefits for process capability analysis: (i) continuously monitoring the process quality through the capability indices in order to assure that the products manufactured are conforming to the specifications. (ii) supplying information on product design and process quality improvement for engineers and designer, and (iii) providing the basis for reducing the cost of product failures.

To calculate and measure the quality characteristics in the industry, one usually assumes that the distribution of measurements follows a normal distribution, and then calculates its process capability indices without verifying their accuracy.

The process capability ratio, also called the Cp index, is commonly used to measure the process quality. It assumes that the distribution of the process output is normal. The Cp index is defined as:

$$C_p = \frac{USL - LSL}{6\sigma} \tag{1}$$

where *USL* is the upper specification limit, *LSL* is the lower specification limit, and σ is the standard deviation.

The C_{pk} index proposed by Sullivan (1985) is a measure of the capability of a process in relation to the process average. It is based on the distance between the process average and the closest specification limit, and is defined as

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

Tau (1997) modified the above formular for C_{pk} to become

$$C_{pk} = C_p(1 - k) \quad (3)$$

where, $k = \frac{|T-\mu|}{m}$, $m = \frac{USL - LSL}{2}$, $and\, T = \frac{USL + LSL}{2}$.

Chan et al. (1988) proposed another index, called C_{pm} which is defined as

$$C_{pm} = \frac{USL - LSL}{6\sigma\sqrt{1+\left(\frac{\mu - T}{\sigma}\right)^2}} \quad (4)$$

The preceding equations are the process capability indices when the normality assumption is not violated. Gunter (1989) identified three non-normal distributions with the same mean and standard deviation as those of a normal distribution: (1) chi-squared distribution with 4.5 d.f.; (2) t distribution with 8 d.f. and (3) uniform distribution. Although having the same C_p and C_{pk} indices, the defects falling outside plus or minus 3σ are significantly different. Thus, Gunter (1989) stated that the defects in parts per million parts (ppm) for a (i) chi-squared distribution is 14,000; (ii) for t distribution, 4000; and (iii) for uniform distribution, zero; while the ppm for normal distribution is 2700. Therefore, if the distribution of measurements of process output violates the assumption of normal distribution, its process capability indices are very questionable (Pan and Wu,1997). Clements (1989) used a Pearson distribution curve to estimate the non-normal process capability index. If the distribution of measurements of a quality characteristic belongs to the Pearson family of probability curves consisting of normal, lognormal, t, F, beta and gamma distributions, then $P(LPL < \mu < UPL) = 1 - 0:0027 = 0:9973$. If the process mean is replaced by median, UPL is the 99.865 percentile, and LPL is the 0.135 percentile of the Pearson family. The non-normal process capability indices are defined as:

$$C_p = \frac{USL - LSL}{X_{99.865} - X_{0.135}} \text{ and } C_{pk} = \min\left\{\frac{USL - X_{50}}{X_{99.865} - X_{50}}, \frac{X_{50} - LSL}{X_{50} - X_{0.135}}\right\},$$

where $X_p = p*100^{th}$ percentile.

Rodriguez (1992) suggested using the goodness of fit test instead of finding its Pearson distribution curve to determine a specific distribution. The maximum likelihood estimation (MLE) method can also be used to estimate its parameters and percentiles.

Vännman (1995) proposes the superstructure which unifies the four basic process capability indices (PCIs), namely C_p, C_{pk}, C_{pm} and C_{pmk} as follows

$$C_p(u,v) = \frac{d - u|\mu - m|}{3\sqrt{\sigma^2 + v(\mu - T)^2}} \quad (5)$$

where u and v are non-negative constants that takes value 0 or 1. Here, $d=\frac{USL-LSL}{2}, m=\frac{USL+LSL}{2}$ and T is the target value. It should be noted that if T is unknown, then T = m is often used (Pearn and Kotz, 2006). The process mean is μ and the process standard deviation is σ.

From Equation (5), the four capability indices can be deduced to become

$$C_p(u,v)=\begin{cases} C_p, & u=0,\ v=0 \\ C_{pk}, & u=1,\ v=0 \\ C_{pm}, & u=0,\ v=1 \\ C_{pmk} & u=1.\ \ v=1 \end{cases} \tag{6}$$

Chen and Pearn (1997) modified Vännman (1995) $C_p(u, v)$ and proposed a quantile based PCI superstructure without implicitly assuming normality of the underlying process for a two sided specifications defined by

$$C_p(u,v) = \frac{d-u|q_{50}-m|}{3\sqrt{(\frac{q_{99.865}-q_{0.135}}{6})^2+v(q_{50}-T)^2}} \tag{7}$$

Vännman and Albing (2007) proposed family of quantile based process capability indices (qPCI) and $C_{MA}(\tau,\eta)$. Peng (2010) extends the work of Vännman and Albing (2007) by developing both asymptotic parametric and non-parametric confidence limits and testing procedures of $C_{MA}(\tau,\eta)$. Chao and Lin (2005) proposed a very general process yield - based PCI as follows

$$C_y = \frac{1}{3}\Phi^{-1}(\frac{1}{2}(F_\theta(USL)-F_\theta(LSL)+1)) \tag{8}$$

where Φ is the CDF of the standard normal distribution, F(.) and θ are the CDF and the vector of parameters of the underlying process distribution. This paper proposes the process capability indices under non-normal distribution, using the median absolute deviation as an estimate of the process variation.

2. *Median Absolute Deviation Based Process Capability Indices*

Let X_{ij} represent a random sample of size n taken over m subgroup, $i = 1, 2,\ldots, n$ and $j = 1,2,\ldots,m$. The sample are assumed to be independent and taken from a continuous identical distribution functions. If σ^2 is unknown, then an unbiased estimate of σ^2 is the sample variance (S^2). In practice, the normality assumption is often violated by real life data, therefore, using S^2 as an estimate of σ^2 will affect the process capability indices and thus this might leads to wrong signal and invalid inference. The median absolute deviation (MAD) has been claimed in the literature to be the best estimate of sigma (σ) when the data under consideration is non-normal (see e.g. Abu-Shawiesh (2008) and Shahriari, Maddahi, and Shokouhi (2009)). Therefore, if MAD is used as an estimate of variability (i.e. σ), then we have

$$\hat{\sigma} = b_n\overline{MAD} \tag{9}$$

where $\overline{MAD} = \frac{\sum_{j=1}^{m} MAD_j}{m}$, b_n is a function of the sample size n and

$$MAD_j = \frac{1}{n}\sum_{i=1}^{n} median\left|X_{ij} - MD_j\right| \qquad (10)$$

where MD_j is the median of the j^{th} rational subgroup or sample. It should be noted that the values of b_n can be obtained in Adekeye (2012) and Abu-Shawiesh (2008). Using MAD as an estimate of σ, then the Chen and Pearn (1997) modification of Vännman (1995) can be further modified by substituting Equation (9) into Equation (5). Thus, we have

$$C_p(u,v) = \frac{d - u\left|M - m\right|}{3\sqrt{(b_n\overline{MAD})^2 + v(M - T)^2}} \qquad (11)$$

where M= Median(X_{ij}), d and m are as earlier defined.

Therefore, from Equation (10), using the definition in Equation (6), The Cp index based on MAD will be

$$C_p = \frac{USL - LSL}{6(b_n\overline{MAD})} \qquad (12)$$

Similarly, the C_{pk} index based on MAD will be defined by

$$C_{pk} = \frac{(USL - LSL) - \left|M - m\right|}{6(b_n\overline{MAD})} \qquad (13)$$

The C_{pm} index will be estimated by

$$C_{pm} = \frac{USL - LSL}{6\sqrt{(b_n\overline{MAD})^2 + (M - T)^2}} \qquad (14)$$

The C_{pmk} index will be estimated by

$$C_{pmk} = \frac{(USL - LSL) - \left|M - m\right|}{6\sqrt{(b_n\overline{MAD})^2 + (M - T)^2}} \qquad (15)$$

3. Results

3.1. Gain of Amplifier Data

A sample of 120 amplifiers is taken to estimate the capability of the manufacturing process which produces them. The quality characteristic of interest is the gain of an amplifier. It has a lower specification limit of 7.75 decibels, an upper specification limit of 12.2 decibels and the target value is placed at 10.0 decibels. The data for this process is contained in Osanaiye et al. (2001). The data was tested for normality using the Kolmogrov-Smrinov goodness of test and Shapiro-Wilk test. To test for the stability of the process, the data was arranged into 24 subgroups with five sample per subgroup. Thus, $\overline{X}$ and $\overline{R}$ was used to monitor the process, and the results reveals that the process that produces the data was a stable process. Using Equation (9) the MAD_j for the rational subgroups were determined to obtain the $\overline{MAD}$. The results of the preliminary analysis using S-plus 4.5 and SPSS 18 are presented in Table 1.

Table 1. Summary of Preliminary Results for the Steel Rod Data

n	$\overline{X}$	$\overline{R}$	$q_{0.135}$	$q_{99.865}$	SK	Kurtosis	K-S Test	p-value	$\overline{MAD}$	M
120	9.041	2.083	7.87	11.7	0.712	0.205	0.089	0.02	0.785	8.9

M = q_{50}, SK = Coefficient of Skewness and K-S Test = Kolmogrov - Sminorv test.

It is clear from the results in Table 1 and as confirmed by the histogram and CDF plot of the data, that the data is from a non- normal positively skewed distribution. The process capability indices based on the MAD are determined using Equations (12) through (15) using the preliminary results in Table 1. The obtained process capability indices are presented in Table 3a.

3.2. Thickness of Tablet Data

300 sample of the measured thickness of tablets were collected from the record book of a company in Nigeria for the period of 30 months with the sample size of 10 tablets per month. The specification limits for thickness of the tablet is 3:57 - 3:97 (mm). Using Equation (9) the MAD_j for the rational subgroups were determined and used to obtain the $\overline{MAD}$. The results of the preliminary analysis (see Table 2) indicate that the sampled data is not from a normal distribution but from a negatively skewed distribution. The process capability indices based on the MAD are determined and summarized in Table 3b.

Table 2. Summary of Preliminary Results

n	$\bar{X}$	$\bar{R}$	$q_{0.135}$	$q_{99.865}$	SK	Kurtosis	K-S Test	p-value	$\overline{MAD}$	M
300	3.799	0.1260	3.57	3.98	- 0.391	-1.28	0.181	0.0	0.1048	3.84

3.3. Simulated Data

Two data sets were generated from exponential distribution and Weibull distribution respectively. These two distributions are heavily positively skewed and are widely used in engineering, particularly in engineering reliability modeling. The generated process data is used to confirm the behaviour of the proposed capability indices and to compare its behaviour with the quantile based capability indices. For the exponential process data, we generated a random sample of size 300 from Exponential population with the mean rate $\lambda = 0.5$. The sampled data was arranged into 30 subgroups each with sample size 10. The data was tested for stability using control charts which confirmed the stability of the process. The information of the control limits was used to fix the specification limits to USL = 4.0 and LSL = 0.20. The computed capability indices are presented in Table 3c.

For the Weibull process data, a random sample of size 300 was generated from Weibull population with scale parameter $\theta = 2.2$ and shape parameter $\beta = 1.5$. The sample data was arranged into 30 subgroups each with a sample of size 10. The generated data was confirmed to be stable using an appropriate control chart. The information of the control limits was used to fix the specification limits as USL = 5.0 and LSL = 0.50. The computed capability indices are presented in Table 3d.

Table 3. Summary of Process Capability Indices Results

(a)Gain of Amplifier				(b)Thickness of Tablet				(c)Exponential data				(d)Weibull data			
C_p	C_{pk}	C_{pm}	C_{pmk}	C_p	C_{pk}	C_{pm}	C_{pmk}	C_p	C_{pk}	C_{pm}	C_{pmk}	C_p	C_{pk}	C_{pm}	C_{pmk}
0.78	0.59	0.51	0.39	0.59	0.48	0.50	0.41	1.03	0.84	0.90	0.73	0.70	0.35	0.48	0.24

The results in Table 3 are compared with the process capability indices based on the quantile (see Equation (7)). Table 4 presents the summary of the results using the two real (manufacturing) process data and the generated highly skewed process data.

Table 4. Comparison of Process Capability Indices Results

PCI	Gain of Amplifier		Thickness of Tablet		Exponential data		Weibull data	
	Quantile	MAD	Quantile	MAD	Quantile	MAD	Quantile	MAD
C_p	1.14	0.78	0.96	0.59	0.76	1.03	0.28	0.70
C_{pk}	0.59	0.59	0.63	0.48	0.61	0.84	0.14	0.35
C_{pm}	0.58	0.51	0.84	0.50	0.70	0.90	0.24	0.48
C_{pmk}	0.30	0.39	0.55	0.41	0.57	0.73	0.12	0.24

4. Discussion

The results in Table 4 reflect that the C_p values for both the proposed and the quantile based are higher than the other capability indices. Furthermore, the quantile based indices are higher than the indices based on the MAD for the two real data considered in this study. However, for the positively skew data (Gain of Amplifier), it can be seen that the indices C_{pk}, C_{pm}, and C_{pmk} are either equal or almost equal. This is an indication that for a positively skewed data, the proposed MAD based capability indices and quantile based capability indices will give the same percentage of non-conforming defects and thus, the same parts per million (ppm) defects. For a negatively skewed data, the quantile based capability indices are higher than the MAD based capability indices. Therefore, the parts per million defects that will be reported using the quantile based indices will be higher than the MAD based indices. However, for the heavily skewed process data (exponential and Weibull data), it is clear that the MAD based indices are higher than the quantile based and thus gives a better indices than the quantile based indices. It should be noted that the obtained indices from the proposed approach reflect that the process that produces the data considered in this study are not capable.

5. Conclusion

This study has proposed the use of median absolute deviation (MAD) for the determination of process capability indices for a non-normal data. From the obtained results and comparison made, it can be concluded that the proposed method did not give any significant improvement over the results that are based on the quantile as a measure of variability for a slightly skewed data while there exist significant improvement in cases of heavily skewed process data. Furthermore, the method seems to be simple and does not need any rounding off values which is the experience in the determination of quantile values needed to compute quantile based indices. Therefore, engineers and users of process capability are encourage to use the propose capability indices when the process data is heavily skewed to the right.

References

Abu-Shawiesh, M.O.A. (2008). A Simple Robust Control Chart Based on MAD. *Journal of Mathematics and Statistics*, 8(1), 37 -41

Adekeye, K.S. (2012). Modified Simple Robust Control Chart Based on Median Absolute Deviation. *International Journal of Statistics and Probability* 1(2), 91 - 95

Chan, L. K., Cheng, S. W., and Spiring,F. A. (1988). A new Measure of Process Capability: C_{pm}. *Journal of Quality Technology* 20,162 - 175.

Chao, M. T. and Lin, D.K.J. (2005). Another Look at the Process Capability Index. *Quality and Reliability Engineering International* 22, 153 - 163.

Chen, K. S. and Pearn, W. L. (1997). An Application of Non-normal Process Capability Indices. *Quality and Reliability Engineering International* 13, 355 - 360.

Clements, J. A. (1989). Process Capability Calculations For Non-normal Distributions. *Quality Progress* 22, 95 - 100.

Gunter, B. (1989). The Use and Abuse of C_{pk}: Part 2 *Quality Progress*, 108 - 109.

Kane, V. E. (1986a). Process Capability Indices. *J. of Quality Technology* 18, 41-52.

Kane, V.E.(1986b). Corrigenda. *Journal of Quality Technology* 18, 265.

Osanaiye, P.A., Adekeye, K.S., and Olugbara, O.O. (2001). Effect of Normalizing Statistical Process Control Data using a Developed Johnson Type Algorithm. *Journal of the Nigerian Statistical Association* 14, 34 - 51.

Pan, J.N. and Wu, S.L. (1997). Process Capability Analysis For Non-Normal Relay Test Data. *Microelectron Reliability* 37(3), 421 - 428.

Pearn,W.L. and Kotz, S. (2006). *Encyclopedia and Handbook of Process Capability Indices: A comprehensive exposition of quality control measures*, World Scientific Publishing Co. Ltd, Singapore.

Peng, C. (2010). Estimating and Testing Quantile -based Process Capability Indices for Processes with Skewed Distributions. *Journal of Data Science* 8, 253 - 268.

Rodriguez, R.N. (1992). Recent Development in Recent Process Capability. *Journal of Quality Technology* 24, 176 - 187.

Shahriari, H. A., Maddahi, A. and Shokouhi, A. H. (2009). A Robust Control Chart Based on M-Estimate. *Journal of Industrial and System* 2(4), 297 - 307.

Spiring, F.A. (1991). Assessing Process Capability in the Presence of Systematic Assignable Cause. *Journal of Quality Technology* 23(2), 125 - 134.

Sullivan, L.P. (1984). Reducing Variability: A New Approach to Quality. *Quality Progress* 17, 15 - 21.

Sullivan, L.P. (1985). Letters. *Quality Progress* 18, 7 - 8.

Vännman, K. (1995). A Unified Approach to Process Capability Indices. *Statistical Sinica* 5, 805 - 820.