IJASRM

International Journal of Advanced Scientific Research and Management, Volume 4 Issue 4, April 2019

www.ijasrm.com

ISSN 2455-6378

On the Estimating the Operating Characteristic of Shewart control chart for Means using Interval estimates of Process mean and Spread

Boya Venkatesu[1], R. Abbaiah[2], and V. Sai Sarada[3]

[1,3] Research Scholar, Department of Statistics, Sri Venkateswara University, Tirupati, Andhra Pradesh, India.

[2] Professor, Department of Statistics, Sri Venkateswara University, Tirupati, Andhra Pradesh, India.

Abstract

The design of Shewart control chart is based on the point estimates $\hat{\mu}$ and $\hat{\sigma}$ of process mean (μ) and standard deviation (σ), each of which is a single value and the control limits are constructed around $\hat{\mu}$ and $\hat{\sigma}$. An interval estimate takes into account the 'within sample variation' and provides a range of values in the form of a confidence interval. It is possible to utilize the information from the confidence intervals and reexamine the performance of the control chart in terms of the Operating Characteristic (OC). In this paper we discuss a method of converting interval estimates of μ and σ into 'improvised point estimates' and there by achieve a better estimate of the OC. The new estimate will be brought out as a convex combination of 3^2 possible point estimates of the OC. Since the interval estimate reflects sampling variation in addition to location, we claim that 'proper summary of interval estimate' serves as a better point estimate and we call it Improvised Point Estimator (IPE). It is shown by simulated experiments that the new estimator is more consistent than the process OC obtained by using the classical method of point estimates.

Keywords: *Control chart, Interval estimate, point estimate, Operating Characteristic, Process mean, New estimate.*

1. Introduction

Let **X** be a quality characteristic following N (μ, σ^2) where μ and σ represents the mean and Standard deviation of **X** respectively. Shewart (1931) proposed the concept of control chart for Statistical Process Control. A control chart describes the pattern of variation and detects the unusual patterns, if any, in the process parameter. The $\bar{x}$ chart is used to control the process mean basing on m independent samples each of size n, drawn from the process periodically. Let LCL and UCL be the lower and upper control limits that holds a type-I risk (α) of 0.0027. These limits are called 3-sigma limits and the sample means $\bar{x}_i$ are plotted against the sample number (i) for i = 1,2…, m. where m is the number of samples each of size n. If $\bar{x}_i \in (\text{UCL}, \text{LCL})$ for all i, then the process is said to be *under control*, provided the values are randomly spread around the central line (CL) which is set at the mean of the sample means $\bar{\bar{x}}$. When μ and σ are known as μ_0 and σ_0 then the control limits are given by $\mu_0 \pm A\,\sigma_0$ where $A = \frac{3}{\sqrt{n}}$ but in general μand σ are unknown. For the ith sample, μ is estimated by $\bar{x}_i = \frac{\sum_{j=1}^{n} x_{ij}}{n}$ and R_i = {Max_i (x_{ij})-Min_i (x_{ij})} is the process range which will be used to estimate σ.

The classical estimate of σ is to use $\hat{\sigma} = \frac{\bar{R}}{d_{2,n}}$ where, $\bar{R} = \frac{\sum_{i=1}^{m} R_i}{m}$ and $d_{2,n}$ is a constant whose values are tabulated at different n. The control limits for the R-chart are given as $UCL = D_4 \bar{R}$, CL $= \bar{R}$ and LCL $=$ D3 $\bar{R}$

The constants D_3, D_4 are also tabulated for various values of n and the control limits of $\bar{X}$ chart will be UCL $= \bar{\bar{x}} + A_2\bar{R}$, CL = $\bar{\bar{x}}$ and LCL = $\bar{\bar{x}} - A_2\bar{R}$ where $A_2 = \frac{3}{d2\sqrt{n}}$.

In order to control the process mean, it is enough to use $\bar{X}$ chart. When σ is known we need to control the process variation also with the help of R chart (or s-chart). Hence, we set a pair of charts known as ($\bar{X}$,R) charts.

The performance of Shewhart control chart is measured by the type–II risk β, which gives the probability of "accepting the process to be in control" even though there is a shift in the process (false alarm). This measure is treated as the OC of the process. Shifts of larger magnitude will have lower false alarm rate when compared to shifts of smaller magnitude because such shifts are rarely missed. Another measure of assessing the performance of the control chart is the Average Run Length (ARL) which is the number of samples that failed to detect a shift before giving the true alarm. Higher value of ARL indicates poor ability to detect a shift in the process, when it really occurs. Several basic details of control charts can be found in Montgomery (2008), Mittag and Rinne (1993) and E.L. Grant (1964).

In all these charts determination of limits depends on how well the parameters are estimated from sample data. The calculation of OC is a function of the point estimates of μ and σ. An estimate of the OC is obtained by using the moment-estimates in the place of μ and σ.

In this paper we discuss a method of converting interval estimates of μ and σ into 'improvised point estimates' and there by achieve a better estimate of the process OC. Since the interval estimate reflects sampling variation in addition to location, we claim that a *proper summary of interval estimate* serves as a better point estimates and we call it *Improvised Point Estimator* (IPE).

In the following section we review some basics of Interval estimates.

2. A brief review of interval estimation

In statistical inference we come across the problem of estimating a population mean/ percentage/ ratio by using sample data. One way of estimating the unknown parameter θ is to provide a single value t_n which is a function of sample $(x_1, x_2, ..., x_n)$. A major portion of statistical methodology is based on this type of estimates called *point estimate.* Sample mean, standard deviation and percentiles are all point estimates of the corresponding population parameters. Estimates are broadly classified as point estimates and interval estimates. Koch and Link (1970) observed that point estimation often gives incorrect information and interval estimates capture reliable information of the parameter.

However, in real life situations point estimates are easy to understood and apply in a context but the accuracy of the point estimate is a matter of concern. For instance, in geological applications, a point estimate of a hilly area can be given in terms of coordinates. If a parachute has to land in that place, the accuracy matters a lot. For this reason, practitioners prefer a range of values instead of a single value as an estimate since such a range gives more information than the single value.

An interval estimate of a parameter θ contains all possible values between two end points $[a_\mathbf{x}, b_\mathbf{x}]$ which are functions of sample $\{\boldsymbol{x} = (x_1, x_2, ..., x_n)\}$. Let $0 \leq \alpha \leq 1$ and (1-α) be the desired confidence with which the interval estimate $[a_\mathbf{x}, b_\mathbf{x}]$ contain the true parameter θ. This interval is called Confidence Interval (CI) or random interval and $P[\boldsymbol{x} \epsilon [a_x, b_x]]$ depends on the sample *x.* If a number of samples are generated and for each sample the CI of the parameter is worked out then the proportion of samples that contain θ is an empirical estimate of (1-α), also known as the *coverage probability.* The method of estimation and the quality of the data show influence on the coverage probability.

3. Interval estimates of process mean and SD

In the case of normal distribution the100(1-α) % CI of the process mean μ is given by

$$\left[\bar{x} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{x} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right] \qquad (1)$$

where $t_{\alpha/2}$ is the (1-α) % quantile of the Student's t-distribution. If we write L ($\bar{x}$,n) and U($\bar{x}$, n) as the limits of the CI then P$[L(\bar{x},n) \leq \mu \leq U(\bar{x},n)]$ =(1-α) (2)

The CI of μ_1 is an Interval estimate and the coefficient $t_{\alpha/2}$ is called confidence coefficient which depends on the sampling distribution of $\bar{x}$ and the sample size n. If n is large, we may use normal quantile ($z_{\alpha/2}$) in the place of $t_{\alpha/2}$.

On similar lines an interval estimate of the process spread σ can be obtained from the sample ranges as

$$\left[\frac{\bar{R}}{d_{2,n}}G_1, \frac{\bar{R}}{d_{2,n}}G_2\right] \quad (3)$$

where G_1, G_2 are constants and R denotes sample range and $d_{2,n}$ is a scaling constant.(Mittag and Rinne(1993))

In the section 5 we propose a new method of evaluating the OC of $\bar{x}$ chart by utilizing the interval estimates in the place of point estimates. The method of point estimates is discussed below.

4. OC function of $\bar{x}$ chart with point estimates

Let μ_0 be the mean of the process, when the process is in control. Suppose the process mean has undergone a shift to $\mu_1 = \mu_0 + k\sigma$, where k is a constant. The quantity $\left[\frac{\mu_1-\mu_0}{\sigma}\right]$ is called standardized shift and the OC of $\bar{x}$ chart is given by

$$\beta = P\{LCL \leq \bar{x} \leq UCL/\mu = \mu_0 + k\,\sigma\} \quad (4)$$

$$\Rightarrow \Phi\left[\frac{UCL-(\mu_0+k\,\sigma)}{\sigma/\sqrt{n}}\right] - \left[\frac{LCL-(\mu_0+k\,\sigma)}{\sigma/\sqrt{n}}\right]$$

$$\Rightarrow \Phi(3-k\sqrt{n}) - \Phi(-3-k\sqrt{n}) \quad (5)$$

The evaluation of β depends on k, which is estimated as $\hat{k} = \frac{\hat{\mu}_1-\hat{\mu}_0}{\hat{\sigma}}$ by using point estimates. For different values of k, the graph of β gives OC curve. As expected, the shape of the OC curve depends on k as well as n.

In the following section we propose a new estimate of β by utilizing the interval estimates of μ and σ.

5. The triplet estimate method of the OC function

The estimation of OC depends on the statistic k = $\left[\frac{\mu_1-\mu_0}{\sigma}\right]$ which takes the form as $\frac{\{\mu_1-\bar{x}\}}{\bar{R}/d_2}$.

Given the CI of a parameter the true value can be anywhere in the CI, by taking the two end points and the middle value of the CI we can obtain a as 3-point estimates of the parameter and call it the *triplet estimate.*

Define $\theta_1 = \frac{\bar{R}}{d_{2,n}}$ G1, $\theta_2 = \frac{\bar{R}}{d_{2,n}}$ and $\theta_3 = \frac{\bar{R}}{d_{2,n}}$ G2.

Then $\tilde{R} = \{\theta_1, \theta_2, \theta_3\}$ is a triplet estimate of R.

Similarly, the triplet estimate for the mean will be $\tilde{\mu} = \{m_1, m_2, m_3\}$

where $m_1 = \bar{x} - t\frac{\bar{R}/d_{2,n}}{\sqrt{n}}$, $m_2 = \bar{x}$, $m_3 = \bar{x} + t\frac{\bar{R}/d_{2,n}}{\sqrt{n}}$

Each of the three components in a triplet estimate is a potential candidate for the true parameter with 100(1-α) % confidence.

Propositon-1: Let O_{ij} be the OC value evaluated at mean m_i and standard deviation θ_j obtained from the triplet estimates defined in (2). Listing these 3^2 values as $O_1, O_2, \ldots, O_9$. A new point estimate will be of the form $O_{new} = \sum_{j=1}^{9} w_j O_j$ where $w_j \geq 0$, $\sum_{j=1}^{9} w_j = 1$ are real numbers.

Proof:

The value of k in relation to the given shift μ_1 can be found at each of the 9 combinations of m_i and θ_i as follows.

$k_1 = \frac{\mu_1 - m_1}{\theta_1}$	$k_4 = \frac{\mu_1 - m_1}{\theta_2}$	$k_7 = \frac{\mu_1 - m_1}{\theta_3}$
$k_2 = \frac{\mu_1 - m_2}{\theta_1}$	$k_5 = \frac{\mu_1 - m_2}{\theta_2}$	$k_8 = \frac{\mu_1 - m_2}{\theta_3}$
$k_3 = \frac{\mu_1 - m_3}{\theta_1}$	$k_6 = \frac{\mu_1 - m_3}{\theta_2}$	$k_9 = \frac{\mu_1 - m_3}{\theta_3}$

Hence the 9 possible OC values using k_i are given by

$$O_i = \Phi(3 - k_i\sqrt{n}) - \Phi(-3 - k_i\sqrt{n}) \quad (6)$$

The linear combination $\sum_{j=1}^{9} w_j O_j$ is the new estimate. The weights can be either fixed (*apriori*) or derived from sample data.

Hence the proposition.

6. Selection of Weights

There are different ways of proposing the weights. One way is an adhoc method proposed by Vishnu Vardhan and Sarma (2010), according to which w_i= 0.5 and the remaining weights are $w_j = 0.25\ (i \neq j)$. Another way is to give equal weights $w_i = \frac{1}{9}$ for all i. Both these methods are *apriori* in the sense that they are fixed before collecting data. Sai Sarada et.al (2018) have used weights that are inversely proportional to the squared deviation from the target.

Suppose the process has a true mean μ′ and true SD is σ′. Then the true value of k (standardized shift) will be $k_0 = \frac{\mu_1 - \mu'}{\sigma'}$ where μ_1 is the shifted mean. Then the true OC, denoted by O_{true} will be

$$O_{true} = \Phi(3 - k_0\sqrt{n}) - \Phi(-3 - k_0\sqrt{n}) \quad (7)$$

Now instead of squared error we use absolute error and define the weights as

$w_i = (O_i - O_T)^{-1}$ ∀ i = 1, 2,…,9. Venkatesu Boya et al (2018) used this method to estimate the process spread using confidence interval of sample range.

Hence the new estimate of the OC will be of the form

$$O_{new} = \frac{\sum_{i=1}^{9} w_i\ O_i}{\sum_{i=1}^{9} w_i} \quad (8)$$

We call this *improvised point estimator* derived from interval estimates. It follows that O_{new} is a convex combination of O_i , i = 1, 2,…,9.

Remark: The method depends on the knowledge about the shifted mean μ_1 from which k_i's will be calculated. For running simulated trails, we have to input the hypothetical shift in the mean μ_1 and workout the resulting O_{new}.

In the following section we report the results of simulated experiments with different sample sizes and compare the resulting OC values.

7. Performance of the new estimate – A Simulation study

The experiment consists of 1000 random samples of size n for n = 5,8,10 generated from N (μ, σ^2) with μ_0 = 10.5 and σ_0= 1.0. For each sample, the mean $(\bar{x})$, range (R) and the 95% CI are calculated. Let the shifted mean be μ_1= 10.0 corresponding to which k can be found and this give one OC value. If we assume that the process truely operate at μ_0 and σ_0, we get k_0 = -0.5 so that the true OC denoted by O_{True} can be found. For n = 5,8 and 10 we get O_T = 0.970061, 0.943601 and 0.922028 respectively. The new OC for each of the 1000 samples is found after deriving k_i and w_i. The resulting values are combined using (8) to get the new OC. The mean and standard error of the new OC are then computed and compared with the classical method which occurs at O_5. The comparative results are shown in Table-1, where the performance is expressed interms of median absolute deviation

Table-1: Comparison of Estimates (1000 trials)

N	True OC	Estimate ± S. E		Median ± MAD	
		Classical OC	New OC	Classical OC	New OC
5	0.970061	0.6493 ±0.0093	0.9101 ±0.0023	0.7634 ± 0.2001	0.9448 ± 0.0506
8	0.943601	0.8926 ±0.0053	0.9256 ±0.0027	0.9567 ± 0.1141	0.9654 ± 0.0580
10	0.922028	0.4029 ±0.0095	0.5332 ±0.0052	0.4003 ± 0.2041	0.5581 ± 0.1127

It can be seen that the new estimate has lower standard error irrespective of the sample (sub group) size. If we use robust estimate median for location and MAD for scale then also the new estimate provides better performance than the classical one.

Figure-1: Comparison of Classical and New estimates

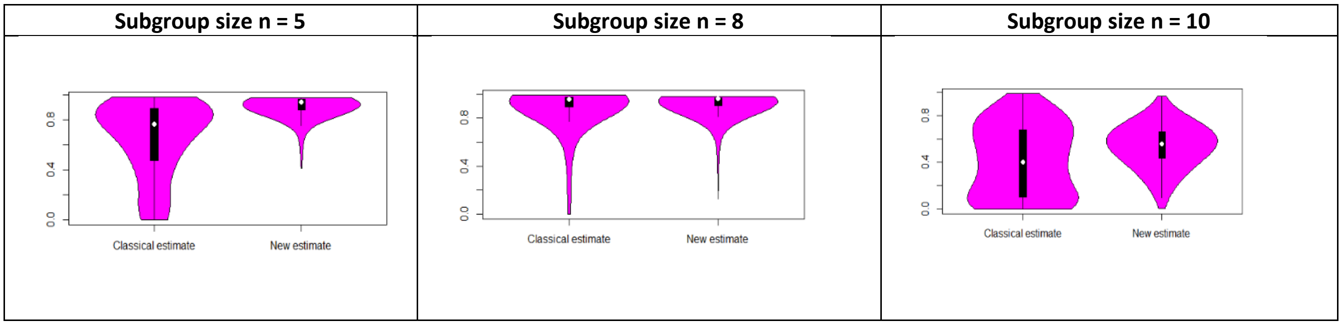

Figure-1 shows the violin plots for the new and classical estimates for different values of n.

8. Conclusions

Hence the triple estimate method offers a new system of consistent performance and provide reliable estimates of the process operating characteristic.

Acknowledgements

The first author wishes to thank Prof. K.V.S. Sarma, for his valuable suggestions and directions on a draft version of this paper.

References

[1] Venkatesu Boya, Abbaih R and Sai Sarada V, A new method of estimating the process spread using confidence interval of sample range, RESEARCH REVIEW – International journal of Multidisciplinary; Volume-03(Issue-08): 204-207p, (2018).

[2] Douglas C. Montgomery, Introduction to Statistical Quality Control, 4th Edn. John Wiley & Sons, Inc; 2008.

[3] Koch George S, Jr. Richard F. Link, Statistical analysis of geological data, Dover publications, Inc; vol (1): (1970).

[4] Mittag H.J, Rinne H, Statistical Methods of Quality Assurance, Chapman and Hall/CRC; 1993.

[5] Eugne L. Grant, Statistical Quality Control, 3rd Edn, McGraw – Hill Book Company (1964)

[6] Vishnu Vardhan R and Sarma K.V.S, Estimation of the Area under ROC Curve using confidence interval of mean, ANU Journal of Physical Sciences; 2 (1): 29-39p, (2010).

[7] Sai Sarada V, Subbarayudu M and Sarma K.V.S, Estimation of process capability index using confidence intervals of process parameters Research & Reviews: Journal of Statistics; Volume-07(Issue-03):49-57p, (2018).

[8] Shewart W.A, Economic control of quality of manufactured product, Princeton: Van Nostrand Reinhold, (1931).