RESEARCH REVIEW JOURNALS
Volume-03
Issue-08
August-2018

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com[UGC Listed Journal]

A new method of estimating the process spread using confidence interval of sample range

[*1]Boya Venkatesu;[2]R. Abbaiah &[3]V. Sai Sarada

[*1]Research Scholar, Department of Statistics, Sri Venkateswara University, Tirupati,(India)
[2]Professor, Department of Statistics, Sri Venkateswara University, Tirupati,(India)
[3]Research Scholar, Department of Statistics, Sri Venkateswara University, Tirupati, (India)

ARTICLE DETAILS

Article History
Published Online: 07August 2018

Keywords
Sample ranges, Shewhart control chart, Confidence interval, Classical estimate, New estimate

[*]Corresponding Author
Email: venkatesusvu@gmail.com

ABSTRACT

This paper deals with a novel method of estimating the process spread (σ) in the construction of Shewhart control chart for means, basing ona new estimatederived from the confidence interval (CI) of sample range. The classical estimate ($\bar{R}/d_{2,n}$)of σ for a normally distributed data is utilized to arrive at anew estimateproposed as the weighted sum of the lower, middle and upper values of the 100(1-α) %CI for the process rangebasing on sample range. The weights are defined as inversely proportional to the absolute bias from the target spread. It is shown by simulation that the new estimate is more consistent than the classical point estimate based on $\bar{R}/d_{2,n}$. It is also shown that the $\bar{X}$ chart performs better in terms of β-risk when the new estimate is used.

1. Introduction

Consider a quality characteristic represented by the random variable XandletX ~ N (μ, σ[2]). The classical Shewhart control chart is a process control tool used to detect the presence of assignable causes, if any, on the process. The $\bar{X}$ chart is used to control the process mean basing onsubgroupsof size n, drawn from the process periodically using m independentsamples. When the process meanand standard deviation (SD) are known as$\mu_0$and σ_0 respectively, thenwesay that the process is under control, if$\bar{X}_i \in (\mathrm{UCL}, \mathrm{LCL})$ fori = 1,2, … m such that there is no recognizable pattern on the control chart, whereLCL and UCL are the lower and upper control limits given as $\mu_0 \pm A\sigma_0$and$A = \frac{3}{\sqrt{n}}$. The condition for the process to be in control is therefore given as

$$\bar{X}_i \in (\mu_0 - A\sigma_0, \mu_0 + A\sigma_0) \qquad \forall\, i = 1,2 \ldots m \quad (1)$$

In general, μ and σ are unknown and estimated from sample data after monitoring the process for a period of time. This helps in establishing LCL and UCL so that at any time point it is enough to verify whether (1) is satisfied or not. Details on the basics of control charts can be found in Douglas Montgomery [1] and a mathematical treatment in Mittag and Rinne [2].

One method of constructing the control limits is based on the point estimates of μ and σ. The control limits for the $\bar{x}$ chart however depend on estimate of process spread (σ). Apoint estimate of $\hat{\sigma}$ for normally distributed data is $\frac{\bar{R}}{d_{2,n}}$ where $\bar{R}$ is the mean of subgroup ranges and$d_{2,n}$is a constant (correction factor to make the estimate unbiased). The values of $d_{2,n}$for different values of n are derived from the sampling distribution of R and available in statistical tables [2].

Even though a point estimate is a convenient way of describing the true (unknown) value of a parameter, it is a practice to provide 100(1-α)% Confidence Interval (CI) in which, the true range is expected toliegiven by $\left[\frac{\bar{R}_L}{d_{2,n}}, \frac{\bar{R}_U}{d_{2,n}}\right]$. These two limits of CI serve as two potential candidates for $\hat{\sigma}$. Infact, any value in the CI is a likely candidate for the true range. By including the value at the middle of CI, we may perceive three different estimates of the range is given by the triplet$\tau = \left[\frac{\bar{R}_L}{d_{2,n}}, \frac{\bar{R}}{d_{2,n}}, \frac{\bar{R}_U}{d_{2,n}}\right]$, instead of an interval. Each of the elements of τ is a potential candidate to estimate σ.

2. Sample range and confidence intervals

Mittag and Rinne [2] has summarized different estimates of process spread basing on sample range as well as standard deviation. Other methods of estimating σ include Boyles and Burr [3]. For the i[th]sample,

Let R_i= **(**Max – Min)$_i$ for i = 1, 2, …, m. If ρ denotes therange for the entire process, then $R_i$is an estimate of ρ. It can be shown that R_i based on a subgroup size n has the property

$$\mathrm{E}\left(\frac{R_i}{d_{2,n}}\right) = \sigma_0 \ and\ \mathrm{V}\left(\frac{R_i}{d_{2,n}}\right) = \left[\frac{e_n}{d_{2,n}}\right]^2 \left[\frac{R_i}{d_{2,n}}\right]^2 \qquad (2)$$

Therefore $\widehat{\sigma_0} = \frac{R_i}{d_{2,n}}$ is an unbiased estimate of σ with a SD of$\frac{\mathrm{e_n} R_i}{\mathrm{d}_{2,n}^2}$. The constants $d_{2,n}$and e_n are compiled and available in statistical tables [2]. Since R_i is only a point estimate of ρ it is possible to construct 100(1-α)% CI for ρ as $\frac{R_i}{d_{2,n}} \pm \mathrm{g}\sqrt{\mathrm{V}\left(\frac{R_i}{d_{2,n}}\right)}$ where g is a constant which is a function of n and α.Using (2) these limits reduce to$\left[\frac{R_i}{d_{2,n}} \pm \mathrm{g}\left(\frac{\mathrm{e_n} R_i}{\mathrm{d}_{2,n}^2}\right)\right]$.

Since the limits are not symmetric, define constants $g_1 \geq 0$and$g_2 \geq 0$ such that the CI will be

$$\left[\frac{R_i}{d_{2,n}} - g\left(\frac{e_n R_i}{d_{2,n}^2}\right), \frac{R_i}{d_{2,n}} + g\left(\frac{e_n R_i}{d_{2,n}^2}\right)\right] \quad (3)$$

Rearrangement of the terms in (3) leads to the lower and upper confidence limits as

$$L = \frac{R_i}{d_{2,n}} G_1 \text{and } U = \frac{R_i}{d_{2,n}} G_2$$

where $G_1 = \left[1 - g\left(\frac{e_n}{d_{2,n}}\right)\right]$ and $G_2 = \left[1 + g\left(\frac{e_n}{d_{2,n}}\right)\right]$(4)

This leads to the triplet $\tau = [\theta_1, \theta_2, \theta_3]$ where

$$\theta_1 = \frac{R_i}{d_{2,n}} G_1,\ \theta_2 = \frac{R_i}{d_{2,n}} \text{ and } \theta_3 = \frac{R_i}{d_{2,n}} G_2 \quad (5)$$

In the following section, we develop a new estimate of σ using the triplet elements and study the properties of the new estimate.

3. Values of $G_1$and G_2

One way of constructing the CI is based on the percentiles of the standardized sample range from normal population. Let r_n denote a sample range with sub group size. Given $0 \leq \alpha \leq 1$, the upper confidence limit is based on the cumulative distribution function of $R_n$for which

$$F(r_{n,1-\alpha}) = P(R_n \leq r_{n,1-\alpha}) = 1-\alpha$$

where $r_{n,1-\alpha}$ is the percentileon the sampling distribution of$R_n$corresponding to (1-α). Similarly, the lower confidence limit satisfies the percentile

$$F(r_{n,\alpha}) = P(R_n \leq r_{n,\alpha}) = \alpha \quad (7)$$

The values of $r_{n,\alpha}$for different n and α are given in [2].For instance, withα = 0.05, we get$r_{n,\alpha}$= 4.47 for n = 10. Similarly, with α = 0.05 the lower percentile for n = 10 will be 1.86. For some selected values of n, the coefficients are shown in Table1.

Table 1
Coefficients for the standardized sample range.

N	$d_{2,n}$	e_n	$r_{n,0.05}$	$r_{n,0.95}$
10	3.078	0.797	1.86	4.47
15	3.472	0.755	2.32	4.80
20	3.735	0.729	2.63	5.01

Forn = 10we get e_n = 0.797, d_n = 3.078, G_1 = 0.51838 and G_2 = 2. 15744.So, the 95% confidence limits for the true range of the process will be 1.86 to 4.47.

In the following section we propose a new estimator of σ using the triplet given in (5).

4. The new estimator of σusing CI ofr_n

We propose a linear combination of these three values given in (5) as a new estimator of σ is given by

$$\hat{\sigma}_{R,CI} = \ell_1\theta_1 + \ell_2\theta_2 + \ell_3\theta_3 \quad (8)$$

such that $\ell_i \geq 0$ for i = 1,2,3. and $\sum_{i=1}^{3} \ell_i$

Suppose the process is designed in such a way that it should have a target spread σ'. This can be known from historical data on the process or from the engineering specifications. There are different ways of proposing the weightsℓ_1, ℓ_2, ℓ_3. One way is an *ad hoc* method proposed by Vishnu Vardhan and Sharma [4], according to which $\ell_2 = 0.5$ and$\ell_1 = \ell_3 = 0.25$. Hence the weights can be taken as

$$w_i = |\theta_i - \sigma'|^{-1} \quad (9)$$

such that $\ell_i = \frac{w_i}{\sum_{i=1}^{3} w_i}$for i = 1,2,3.(10)

As a result, the new estimator given in (8) becomes a convex combination of the middle and extreme values of the confidence interval. It can be seen that the weights decrease as the absolute error increases.

Sai Sarada *et.al* [5] have used a method in which the weights are inversely proportional to absolute difference between θ_i and σ' $\forall$ i = 1,2,3. Reddi Rani *et.al* [6] have used the CI based estimate of process mean and proposed a new method of evaluating the OC curve of a single sampling plan by variable type of inspection.

In essence we propose that the CI based estimator of σ contains information on the probable variation in the parameter and hence we claim that this estimator will be more useful than the point estimator obtained at the middle of theCI.
In the following section a stepwise procedure is discussed followed by simulated experiments in support of the new estimate**.**

5. Step wise procedure

The following steps can be followed to implement the new method.

- Generate 'm' random samples each of size 'n' from N (μ, σ^2).
- Evaluate the sample mean and range for each subgroup.
- Calculate the CI using (3)

- Read $e_n$andd_n, from tables and calculates G_1, G_2 using (4)
- Calculateθ_1, $\theta_2$andθ_3 of each subgroup, using (5)
- Calculate w_1, w_2, $w_3$using (9)
- Calculate ℓ_1, ℓ_2, ℓ_3 using (10)
- Find new estimator$\hat{\sigma}_{R,CI}$ using (8)

This procedure leads to $\hat{\sigma}_{R,CI}$ for each of the m samples and this data can be used to study the empirical properties of the new estimate.

For each value of n, the new estimate can be found either on real data or on a simulated dataset. Repeating this exercise over 'm' samples like m = 500 or 1000, it is possible to study the statistical properties of $\theta_1, \theta_2, \theta_3$ and new estimate$\hat{\sigma}_{R,CI}$. Since $\theta_2$is an estimate mostly followed in practice, the simulated expected value, bias and standard error of the new estimate can be compared with those ofθ_2.

In the following section a simulated experiment is reported to demonstrate the working of the new method.

6. Illustration by simulation

Let us take X~N (10.5, 2.25) so that the process has σ = 1.5 and the data be expressed in consistent units. We have generated 500 samples each of size n = 10,15 and 20 from the above distribution using the Random Number Generation tool of the Data Analysis Pak in MS-Excel.

Taking n = 10, we get e_n = 0.797, d_n = 3.078,G_1= 0.51838, G_2 = 2.15744. The confidence interval for each sample is found using an MS-Excel template developed for this purpose. The summary of results is shown in Table2.

Table 2

Estimate of σ using sample Ranges under the two methods.

n	M	Estimate ± S. E	
		Classical Point Estimate	New Estimate
10	500	1.4572 ± 0.0161	1.4767 ± 0.0070
15	500	1.3103 ± 0.0151	1.4185 ± 0.0065
20	500	1.2472 ± 0.1490	1.3909 ± 0.0076

From Table2. It can be seen that irrespective of subgroup size n, the new estimate using confidence intervals of R is found to be closer to the hypotheticalprocess spread of σ = 1.5. Further the standard error is much lower than that of classical estimate. Hence, we propose that the new estimate of σ basing on the CI of sample ranges provides a consistent estimate with less bias when compared to $\frac{\bar{R}}{d_{2,n}}$.

In the following section we study the impact of the new estimate on the Operating Characteristic (OC) of the$\bar{\mathbf{x}}$Chart.

7. Effect of new estimate on the OC function of the $\bar{x}$ chart

The performance of the$\bar{X}$ chart depends on the control limits which are based on the estimate of σ. One of the performance measures of the chart is the Operating Characteristic (OC), which measures the type-2 risk (β). The probability of accepting the process when there is a true shift in the mean is given by the OC function. It can be shown that the OC is a function of the magnitude of shift in the mean given by$k = \frac{(\mu_1 - \mu_0)}{\sigma}$ where μ_0 is the current mean and μ_1 is the shifted mean, so that $\mu_1 = \mu_0 + k\sigma$.

In a practical situation when the suspected shift in mean isμ_1 we can determine k if σ is known.FollowingMontgomery andRunger [7], the OC function can be calculated in terms of n and k as

$$\beta = \Phi(3 - k\sqrt{n}) - \Phi(-3 - k\sqrt{n}) \quad (11)$$

where Φ denotes the cumulative standard normal distribution.

Since n is fixed in this study, the pattern of the OC Curve depends on k only, which also depends on σ. We consider two estimates of σ given by a) Classical estimate $\frac{\bar{R}}{d_{2,n}}$ and b) New estimate $\hat{\sigma}_{R,CI}$ given in (8) and compare the OC values with the true OC obtainable when σ is known.

For different values of shiftedmean, we have calculated k along with the OC values of $\bar{x}$ Chart.

Suppose the true σ specified by the hypothesis is used instead of the estimated value. Then the resulting k and β will be the true values with which the estimated values will be compared. The results are shown in Table3.

Table 3

Comparison of OC values under different estimates of σ

μ_1	Classical estimate		New estimate		True value	
	K_{cla}	β_{cla}	k_{new}	β_{new}	k_{true}	β_{true}
10.5	0.00000	0.99730	0.00000	0.99730	0.00000	0.99865
11.0	0.44190	0.94548	0.37446	0.96529	0.33333	0.97417
11.5	0.88381	0.58128	0.74891	0.73622	0.66667	0.81375
12.0	1.32571	0.11658	1.12337	0.29033	1.00000	0.43554
12.5	1.76761	0.00480	1.49783	0.04123	1.33333	0.11192
13.0	2.20951	0.00003	1.87229	0.00175	1.66667	0.01159

Itfollows from Table3 that the true OC value, which is expected by the hypothesis is *under estimated* by both the methods. For instance, when μ_1=11.0, the true OC is β = 0.9741 which means 97.4% of the time the process fails to detect the shift. When$\frac{\bar{R}}{d_{2,n}}$ is used we getβ = 0.9454 only.The new estimator however is β_{new} = 0.9652, which is higher than β_{cla} with classical method.

The OC curves are drawn with n = 10, 15 and 20 as shown in Fig. 1.

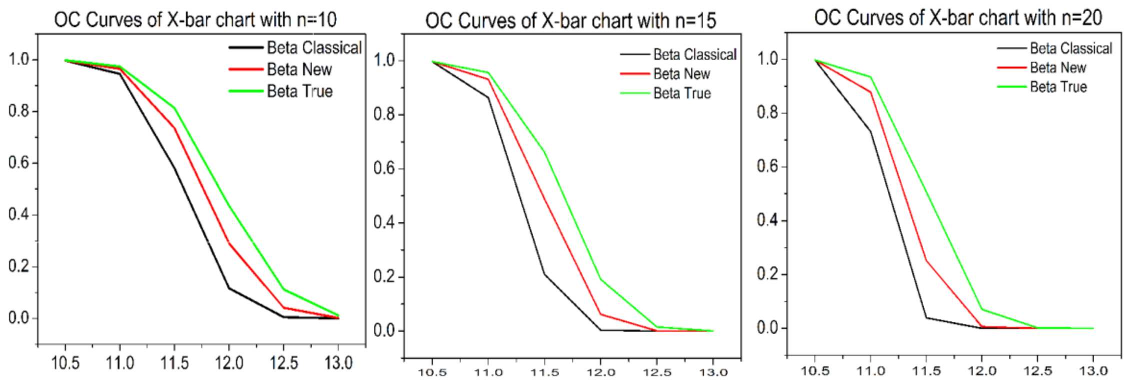

Figure1. OC Curves of X-bar Chart with different methods of estimating σ

From Fig.1, It follows that as n increases the OC curve becomes steeper, which is an expected result. In each case the OC curve based on the new estimate of σ is closer to the true OC curve, when compared to the curve based on the classical point estimate$\frac{\bar{R}}{d_{2,n}}$.

8. Conclusions

In this paper we have proposed a new method of estimating σ of a normally distributed process using sample range. We have drifted from a point estimate to an interval estimate of σ and combined the lower, middle and upper values of the CI through a convex combination. The weights are allowed to vary inversely with the distance of each value from the hypothetical σ. It is shown by simulated experiments that the new estimate performs better in terms of standard error and provides better OC curve.

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. Douglas C. (2008): Montgomery, Introduction to Statistical Quality Control, 4thEdn. John Wiley & Sons, Inc.
2. Hans-Joachim. Mittag, Horst Rinne, (1993): Statistical Methods of Quality Assurance, Chapman and Hall/CRC.
3. Boyles R.A, (1979): Estimating Common- cause σ in the presence of special causes, Journal of Quality Technology*;* 29: pp 381-395.
4. Vishnu Vardhan R and K.V.S Sarma, (2010): Estimation of the Area under ROC Curve using confidence interval of mean, ANU Journal of Physical Sciences; 2010: 2(1): pp 29-39.
5. Sai Sarada V, M. Subbarayudu and K.V.S. Sarma, Estimation of process capability index using confidence intervals of process parameters, paper presented at: International Conference on Changing Paradigms and Emerging Challenges in Statistical Sciences (IPECS-2018). Pondicherry University, Puducherry, India*.* 2018 January 29 – 31.
6. Reddi Rani K, Sai Sarada V, Sarma KVS, (2017): A New Method of implementing single sampling plan by variables using interval estimates of process mean and variance, International Journal of Scientific and Engineering Research; 8(7): pp1702-1705.
7. Douglas C. Montgomery and George C. Runger, (2002): Applied Statistics and Probability for Engineers, 3rdEdn, John Wiley & Sons, Inc.

