IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 11, Issue 4 Ver. II (Jul - Aug. 2015), PP 05-10
www.iosrjournals.org

A Simple RobustDispersionControl Chart Based on MMLE

M.R.Sindhumol[1], M.R.Srinivasan[2]
[1, 2] *(Department of Statistics, University of Madras, Chennai, India)*

Abstract:*Shewhart S-control chart is one of the widely used charts to control process dispersion. Presence of outliers may extremely affect this control charting procedure. More- over, if the quality characteristic is not following normality assumption, one has to look for a robust control chart. A robust control chart is a better choice to overcome it. This article presented a Shewhart type robust dispersion control chart based on Modified Maximum Likelihood Estimator (MMLE) which is robust to the presence of outliers as well as the change in assumptions on distribution. Though robust properties of MMLE are available in literature, its exposure to industrial application is very less. A comparison of this dispersion chart with sample standard deviation S-chart is also given in terms of outlier detection. Moreover, a simulation study is conducted to compare charts in terms of Average Run Length (ARL) for both in-control and out-of-control situations. Numerical examples are given to check its performance.*

Keywords: *Average Run Length, Control Chart, Control Limits, Modified Maximum Likelihood Estimator, Outlier, S-chart.*

I. Introduction

Control chart is now an on-line process control technique to detect and remove variations in the process introduced by Walter.A.Shewhart (1924) in Bell Telephones Laboratory. Determination of assignable causes of variation in control chart is possible with the use of control limits. Specifying the control limits is the most important step in designing a control chart. When the limits are narrow, the risk of a point falling beyond the limits increases, and hence increase the false indication that the process is out of control. If the limits are wider, the risk increases the points falling within the limits, falsely indicates that the process is in control [1]. There are many measures of scale available in literature and many control charts are available to control process dispersion or the dispersion of quality characteristics.

Sample standard deviation (S) is the most commonly used measure of dispersion in many of the statistical applications including industrial application. It is defined as the square root of the mean of the squares of deviations from the common sample mean. It can be regarded as a true representative of the data, since all data values are taken into account in its calculation. If sample size is moderately large, S-chart is better than range (R) method. Though S is a good estimator; it is non-robust to slight deviations from non-normality. When extreme values are presented, sample standard deviation is not a good representative of process dispersion. In control chart application, in order to estimate process dispersion using preliminary samples, pooling sample variance gives a better estimate than pooling sample standard deviation [2]. Zhang [3] introduced two charts namely Improves R Chart (IRC) and Improved S Chart (ISC).

There are many robust measures of scale available in literature. A robust estimator is an estimator which is insensitive to changes in the underlying distribution and also resistant against the presence of outliers. Several authors have worked on robust control charts and several robust measures are experimented in process control. Some of them are Trimmed mean of Range [4],Inter Quartile Range (IQR)[5-6], Gini's Mean Difference [7], Median Absolute Deviation (MAD)[8-10], Dispersion on M-estimate [1],The Q_n and S_n estimates [11], and estimate obtained by the mean subgroup average deviation from the median MD [12].The present study introduces the application of another robust measure of dispersion namely Modified Maximum Likelihood Estimator (MMLE) in process control based on Shewhart type control chart.

II. Modified Maximum Likelihood Estimator

Let f(x; Θ) be a probability density function of x, Θ being an unknown parameter. TheMaximumLikelihood equation based on a sample of size is obtained from

$$\xi(\underline{x},\ n,\theta) = \frac{d\log L}{d\theta} = 0. \tag{2.1}$$

Here $\underline{x}^T = (x_1, x_2, \dots x_n)$ and L is the likelihood function. Suppose the Maximum Likelihood equation based on a sample of size n has no explicit solution, one has to seek another function

$$\xi^*(\underline{x},\ n,\theta) = \frac{d\log L^*}{d\theta} = 0. \tag{2.2}$$

$$\text{Limit } n\to\infty\ \xi(\underline{x},\ n,\theta) = \xi^*(\underline{x},\ n,\theta). \tag{2.3}$$

Suppose that the underlying distribution of the type (1/σ) f(X-μ)/σ), where μ and σ are location and scale parameters respectively. Let $F(z) = \int_{-\infty}^{z} f(z)dz$, where z = (X-μ)/σ. Then the likelihood function based on symmetric censored sample is not having an explicit solution. Tiku.et.al [13]has studied MMLE and its variance for symmetric and non-symmetric distributions and for symmetric and non-symmetric censoring. The MMLE for μ and σ, for a symmetric censored normal distribution are:

$$\hat{\mu} = \frac{\sum_{i=r+1}^{n-r} x_{(i)} + r\beta(x_{(r+1)} + x_{(n-r)})}{p} \quad (2.4)$$

$$\text{and} \quad \hat{\sigma} = \frac{B + (B^2 + 4AC)^{1/2}}{2[A(A-1)]^{1/2}}. \quad (2.5)$$

Here, n=sample size, p= n-2r+2rβ,

A= n-2r, B=rα ($x_{(n-r)} - x_{(r+1)}$) and

$C = \sum_{i=r+1}^{n-g} x_{(i)}^2 + r\beta\left(x_{(r+1)}^2 + x_{(n-r)}^2\right) - p(\bar{x}_{pr})^2$.

The parameters α and β are in such a way that g(z) = f(z) / [1-F(z)] = α + β z.

When r=0, these estimates are reduced to sample mean and sample standard deviation.Tiku.et.al [13] showed that the MMLE of mean and its estimator of standard deviation are useful pair of robust estimators of location and scale parameters. They also showed these MMLE $(\hat{\mu}, \hat{\sigma})$ are almost as efficient as MLE $(\hat{\mu}, \hat{\sigma})$ for type-2 censored normal samples and more efficient than the Best Linear Unbiased Estimator (BLUE) of $(\hat{\mu}, \hat{\sigma})$. Also they are jointly more efficient overall than the trimmed estimators and as efficient as some of the celebrated robust estimators named Huber's M estimators, including the wave estimators, the bi-square estimators [13-14] for symmetric distributions.

III. Robust Control Charts For Process Dispersion

Lax [15] studied finite sample performance of robust estimators of scale in long tailed symmetric distributions. Langenberg and Iglewicz [4] proposed mean and range charts with control limits determined trimmed mean of the subgroup means and the trimmed mean of the range. Rocke[5] proposed standard deviation control charts based on the mean or the trimmed mean of the subgroup ranges or subgroup Inter Quartile Ranges. He suggested that in order to identify outliers easily, limits of a control chart should frame based on robust measures while non-robust measures are plotted on it. Abu-Shawiesh[8] derived the control limits for the standard deviation control chart using the Median Absolute Deviation (MAD). Mahmoud. et.al.[2]showed that the square root of average of sample variance is more efficient than the mean of the subgroup standard deviations.

This paper considered two phases of control charting procedure. Preliminary samples are considered for constructing limits. After framing the limits, values of sample standard deviations (s_i) are plotted and if any $s_i$values are outside the limits, the chart is revised. Once the limits are constructed, it is used for phase II to control the process. If the subgroup size is n and number of preliminary subgroups used in phase I is *m*, usually an estimate of process standard deviation using sample standard deviation is calculated based on the pooled standard deviation from *m* subgroups. In order to estimate standard deviation in quality control applications, Mahmoud et.al [2] used the square root of average of variances from the subgroups. He showed that averaging the unbiased estimators gave much more efficient multiple sample estimators than averaging biased. In this article, process standard deviation is estimated based on this method.

The Center Line (CL), Lower Control Limits (LCL) and Upper Control Limits (UCL) of a Shewhart type dispersion chart are:

$CL = c_4\,\sigma$; $LCL = c_4\,\sigma - 3\,\sigma\sqrt{1-c_4^2} = B5\,\sigma$; $UCL = c_4\,\sigma + 3\,\sigma\sqrt{1-c_4^2} = B6\,\sigma$. (3.1)

where $c_4(n) = \left(\frac{2}{n-1}\right)^{1/2} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$.

When σ is estimated from the preliminary samples, average of subgroup standard deviation s_i is used.

$\bar{s} = (\sum_{i=1}^{m} s_i)/m$, where $s_i = \left(\frac{1}{n-1}\sum_{j=1}^{n}\left(x_{ij} - \bar{x_i}\right)^2\right)^{1/2}$.

$\hat{\sigma} = \bar{s}/c_4$. (3.2)

$CL = \bar{s}$; $LCL = \left(1 - \frac{3}{c_4}\sqrt{1-c_4^2}\right)\bar{s} = B_3\bar{s}$; $UCL = \left(1 + \frac{3}{c_4}\sqrt{1-c_4^2}\right)\bar{s} = B_4\bar{s}$. (3.3)

Mahmoud.et.al. [2] considered $\bar{s}_{var} = \left(\frac{1}{m}\sum_{i=1}^{m} s_i^2\right)^{1/2}$ and hence an unbiased estimator is given based on sample variance as

$$\widehat{\sigma *} = \frac{\bar{s}_{var}}{c_4(m(n-1)+1)} = \frac{\bar{s}_{var}}{c_4(\nu+1)}. \quad (3.4)$$

Control limits based on this estimator are

$CL = \bar{s}_{var}$; $LCL = B_3(\nu+1)\bar{s}_{var}$; $LCL = B_4(\nu+1)\,\bar{s}_{var}$ (3.5)

Here, B_3(.) and B_4(.) can be obtained from the same tables used for classical dispersion charts.

Adekeye[9] developed a modified limits based on MAD.Schoonhoven and Does [12]evaluated an adaptive trimmer where the estimate of S is obtained by the mean subgroup average deviation from the median. They also studied trimmed mean of IQR. Zhang [3] made an improved R-chart and S-chart based on its cumulative distribution function.

When the chart based on dispersion of MMLE is considered, the control limits are:

CL =$\widehat{\sigma_{MMLE}}$; LCL = $B_3(\nu^* + 1)\widehat{\sigma_{MMLE}}$; LCL = $B_4(\nu^* + 1)\widehat{\sigma_{MMLE}}$. (3.6)

where, $\widehat{s_{MMLE}} = \left(\frac{1}{m}\sum_{i=1}^{m} s^2_{MMLE\ i}\right)^{1/2}$ and $\widehat{\sigma_{MMLE}} = \widehat{s_{MMLEa}} / c_4(\nu^* + 1)$. (3.7)

Control chart constants$c_4(m(n - 2r + 2r\beta - 1) + 1) = c_4(\nu^* + 1)$

$B_3(\nu^* + 1) = (1 - \frac{3}{c_4(\nu^*+1)}\sqrt{1 - {c_4}^2(\nu^* + 1)}\,)$ and

$B_4(\nu^* + 1) = (1 + \frac{3}{c_4(\nu^*+1)}\sqrt{1 - {c_4}^2(\nu^* + 1)}\,)$.

Here, B_3(.) and B_4(.) can be obtained from the same tables used for classical dispersion charts.

IV. Robust Dispersion Charts Based On Mmle With Examples

An empirical study based on Monte Carlo simulation is conducted for 1000 runs and random samples are simulated from N(0,1). Subgroup sizes considered are n=10 and 20and number of subgroups are considered 20 for each case. Two levels of symmetric censoring for each subgroup are 10% (r=1) and 20% (r=2). Clean samples are considered to analyze type-1 error and contaminated samples are included to study the effect of detection of outliers or assignable causes of variation. Out-of-control situation is studied based on the samples taken from N(0,2) and N(0,4).

A classic way of illustrating the effect of slight departure from normality is with the contaminated or mixed normal distribution. Let X is a standard normal random variable having distribution function φ(x). Then for any constant K>0, φ(x/K) is a normal distribution with standard deviation K. Let ε be any constant, $0 \le \varepsilon \le 1$. The contaminated normal distribution is H(x) = (1- ε) φ(x) + ε φ(x/K) which has mean 0 and variance $(1- \varepsilon + \varepsilon K^2)$.

Contamination is made for 40% of the subgroup so that 10% and 20% censoring will be meaningful for a contaminated subgroup.

The mixture models are 0.60 N(0,1)+0.40 N(0,2), and 0.60 N(0,1)+0.40 N(0,4).

Among the 20 subgroups, 5% subgroups are out of control and hence outlier models are fixed as 16 N(0,1) and 4 N(0,3)mixed and 16 N(0,1) and 4 N(0,5) mixed.

Fig. 1, 2 and 3 show the in-control signal of control charts based on estimates from pooled sample variance, MMLE with 10% censoring and 20% respectively. The simulation study showed that for an in-control process, many times s-chart and s_{var}-chart are performing all most equal, s_{var}-chart is more efficient as it showed false alarm very rarely. The chart based on MMLE always performs better than s-chart and s_{var}-chart and it is the most efficient among the three in terms of false alarm. Simulation study of Average Run Length (ARL) also supports this. A simulation study is made for comparison of charts in terms of detecting out-of control signal. The Fig. 4 indicates that among the four out-of-control signals fixed, MMLE is detecting three.

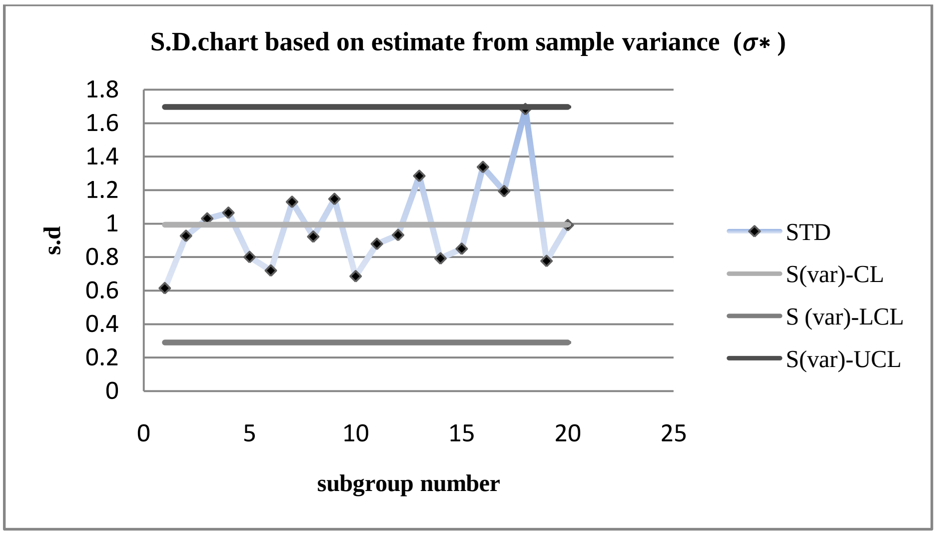

Figure-1: Dispersion chart based on the estimate $\widehat{\sigma^*}$when process is in-control

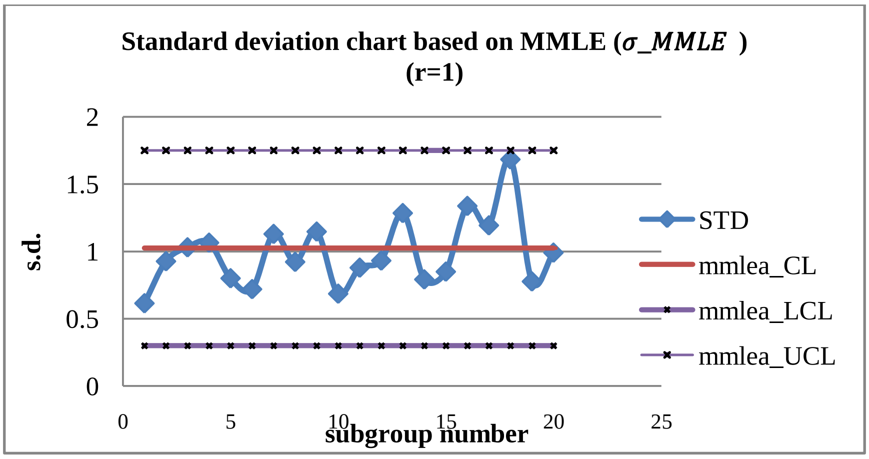

Figure-2: Dispersion chart based on the estimate $\widehat{\sigma_{MMLE}}$ for censoring r=1 when process is in-control

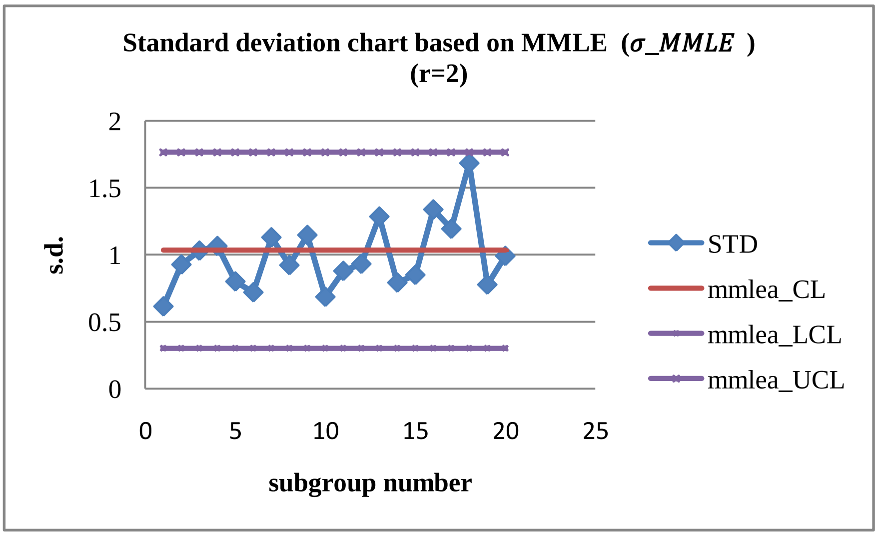

Figure-3: Dispersion chart based on the estimate $\widehat{\sigma_{MMLE}}$ for censoring r=2 when process is in-control

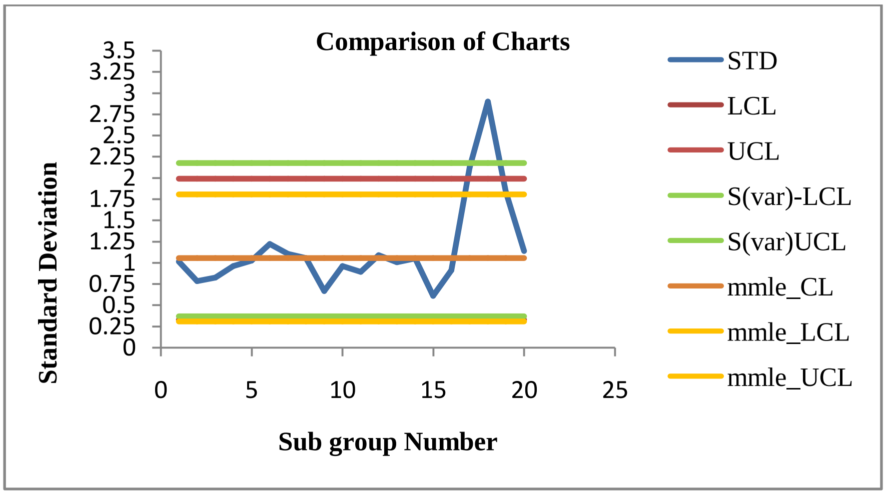

Figure-4: Comparison of dispersion charts s-chart, $s_{var}$and $s_{MMLE}$for out-of-control process based on N(0,1)Contaminated with N(0,3) for censoring r=2

V. Charts Performance Using Average Run Length

The Average Run Length (ARL) is evaluated as the number of points that must be plotted before a point indicated out-of-control signal. It is defined as the reciprocal of the probability that any point exceeds the control limits. The ARL of Shewhart control chart in the in-control process state can be varied considerably as the standard deviation of the distribution of run length is high. When the process is in-control, ARL of Shewhart control chart is expected to be close to 370. If the process is in-control, the in-control Average Run Length, ARL_0, should be large. A simulation study is done to compare ARL for both in-control and out-of-control status.

Let N (μ, σ^2) be a normal distribution, then sets of m = 20 subgroups consisting of n = 10 observations were generated from N (0, 1) distribution. The control limits for the s- chart and the proposed s_{MMLE} control chart were constructed. After determining the control limits, random N (0, 1) subgroups of size n were generated. The *s* statistic was computed for each subgroup and compared to the control limits of both control charts. The number of subgroups required for the value of the *s* estimator to exceed the control limits was recorded as a run length observation, RLi. For runs not signaling by the 1000^{th} subgroup, the run length was recorded as 1000. This process is repeated 10,000 times and the results of this simulation study are given in Table-I. The ARL_0 was calculated as:

$$ARL_0 = \frac{\sum_{i=1}^{10000} AR_i}{10000} \qquad (5.1)$$

ARL is obtained in the in-control state and out-of-control state.The same procedure is used to compare for the out-of-control ARL. If the process is out-of-control, it's ARL_1, to be small. The control limits for the ARL_1 are based on N (0, 1) distribution. The observations used to calculate the statistics *s* are from a normal distribution with mean μ = 0 and standard deviation λσ, where λ representing the size of the shift in the standard deviation. Example shows for λ=2 and 4. Number of samples required for the value of *s* estimator to exceed the control limits was recorded as a run length observation, RLi. Process is simulated 10,000 times and took average as ARL_1. Similarly, the study is conducted for the slight variations from normality, which is for contaminated normal.

The following table shows a comparison of ARL based on simulation study. In the out-of control situation, ARL should be as low as possible.

Table-I: ARL for sample size n=10 and m=20

Distributions	S chart	S_{var} chart	S_{MMLE}	
			r=1	r=2
N(0,1)	370	438	565	614
N(0,1.5)	259	234	190	204
N(0,2)	0.42	0.47	0.54	0.59
N(0,4)	0.41	0.47	0.55	0.58
N(0,1)+40% N(0,2)	3.02	3.44	4.02	4.23
N(0,1)+40% N(0,4)	3.01	3.45	4.02	4.33

Table-I shows ARL_0 and shows that the s_{var}–chart and robust chart s_{MMLE}-chart are more efficient in in-control status of process and will rarely show false alarms than standard deviation chart. This will prevent the loss due to false alarm in any good working process. In case of ARL_1, all charts are almost equal and even if s_{var} and robust chart are less efficient, the difference is small and legible only.

VI. Conclusion

Tiku.et.al.[13] has studied MMLE and has showed that it has good properties and is better than many of the well-known robust estimators. The present study explored standard deviation based on MMLE to control process dispersion. An empirical study is done to compare its performance using software SAS. This robust chart is showing better performance in detecting outliers if compared with classical s-chart. Also a control chart is constructed for another efficient estimator, which is the square root of the average of sample variance, introduced by Mahmoud.et.al.[2],and is compared with this robust chart. In literature, both of these estimates are proved as efficient and have less Mean Square Error. When applying both of them in control charting procedure, robust chart based on MMLE is found as better one in detecting out-of-control signal as well as not showing false alarm. Charts performance is also studied based on the ARL and simulation study favors MMLE based dispersion chart in both of these studies.

References

[1]. H. Shahriari1, A. Maddahi, and A. H. Shokouhi, A Robust Dispersion Control Chart Based on M-estimate**,** Journal of Industrial and Systems Engineering, 2(4), 2009, 297-307.
[2]. M.A.Mahmoud, G.R.Henderson, E. K. Epprecht and W. H.Woodall, Estimating the standard deviation in Quality Control Applications, Journal of Quality Technology, 42(4), 2010, 348-357.
[3]. Zhang Guoyi, Improved R and s control charts formonitoring the process variance, Journal of Applied Statistics**,** 41(6), 2014, 1260–1273.
[4]. P.Langenberg, and B. Iglewicz,Trimmed Mean X bar and R chart, Journal of Quality Technology, 18,1986, 152-161.
[5]. D.M.Rocke, Robust Control charts, Technometrics, 31(2), 1989, 173-184.
[6]. A.Shabbir, Zhengyan Lin, S. A.Abbasi and M.Riaz, On Efficient Monitoring of Process Dispersionusing Interquartile Range,Open Journal of Applied Sciences, Supplement,World Congress on Engineering and Technology, 2012, 39-43.
[7]. M. Riaz and Saghirr, Monitoring process variability using Gini's mean difference, Quality Technology and Quantitative Management, Volume 4(4),2007, 439-454.
[8]. M. O. A. Abu-Shawiesh, A Simple Robust Control Chart Based on MAD**,** Journal of Mathematics andStatistics Volume4 (2),2008,102-107
[9]. K.S. Adekeye, Modified Simple Robust Control Chart Based on Median Absolute Deviation, International Journal of Statistics and Probability, 1(2), 2012, pp.91-95
[10]. K. S. Adekeye, andP. I. Azubuike, P. I., Derivation of the Limits For Control Chart Using The Median AbsoluteDeviation for Monitoring Non Normal Process,Journal of Mathematics and Statistics, 8(1), 2012, 37-41
[11]. Nandini Das, Control charts for controlling variability of Non-Normal process, Economic Quality Control, 26, 2011, 121-131.
[12]. M. Schoonhoven and R.Does, A Robust Standard Deviation Control Char**t**, Technometrics,54(1), 2012, 73-82
[13]. M. L. Tiku, W. Y. Tan and N. Balakrishnan, Robust Inference (Marcel Dekker, NY, 1986).
[14]. R. R. Wilcox, Introduction to Robust Estimation and Hypothesis Testing (2^{nd}Edn. Elsevier, USA, 2005)
[15]. D. A.Lax, Robust Estimators of Scale: Finite-sample Performance in Long-Tailed Symmetric Distributions, Journal of AmericanStatistical Association,80(391), 1985, 736-741