Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v9n1p111

A robust dispersion control chart based on modified trimmed standard deviation

By Sindhumol, Srinivasan, Gallo

Published: 26 April 2016

This work is copyrighted by Università del Salento, and is licensed under a Creative Commons Attribuzione - Non commerciale - Non opere derivate 3.0 Italia License.
For more information see:
http://creativecommons.org/licenses/by-nc-nd/3.0/it/

Electronic Journal of Applied Statistical Analysis
Vol. 09, Issue 01, April 2016, 111-121
DOI: 10.1285/i20705948v9n1p111

A robust dispersion control chart based on modified trimmed standard deviation

Sindhumol M.R.*[a], Srinivasan M.R.[a], and Gallo M.[b]

[a] *Department of Statistics, University of Madras, Chennai-600004, Tamilnadu, India*
[b] *Department of Human and Social Sciences, University of Naples - L'Orientale, P.zza S.Giovanni 30, Napoli-80134, Italy*

Published: 26 April 2016

Control Chart is a widely used on-line process control techniques to control variability. This paper focuses on variability due to dispersion of a quality characteristic. Classical methods of estimating parameters of the distribution of quality characteristic may be affected by the presence of outliers. In order to overcome such situation, robust estimators, which are less affected by the extreme values or small departures from the model assumptions, are introduced in industrial application. This article introduced a modification to trimmed standard deviation to increase its efficiency, and is used in controlling process dispersion. Authors constructed a phase-I control chart derived from standard deviation of trimmed mean, which is robust. Simulation study is conducted to assess its performance at phase-II. This robust control chart is compared with s-chart in terms of its efficiency to detect outliers or assignable causes of variation as well as its Average Run Length.

keywords: Average Run Length, Control Limits, Outlier, Robust Control Chart, Trimmed Mean.

1 Introduction

Shewhart control chart for sample standard deviation is a widely used process control technique. Performance of a control chart is depending on designing its limits, as narrowing or widening limits can influence probability of type I and type II errors respectively. Its performance can also be affected by the presence of outliers. As sample standard

*Corresponding author: mgallo@unior.it.

©Università del Salento
ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

deviation is defined on all data points in a sample, it is influenced by the extreme values present in the subgroup. If subgroups taken in phase-I contain many outliers, 3-sigma limits of s-chart may inflate and outliers remain unnoticed. This chart is based on the assumption that the underlying distribution of quality characteristic is approximately normally distributed, even though the actual distribution of s has long right tail. If robust measures are used to estimate process dispersion and to construct limits, it can over come such problems.

A robust estimator is an estimator that is insensitive to changes in the underlying distribution and also resistant against the presence of outliers. There are many robust measures of location and scale available in literature. Wilcox (2012) says by substituting robust measures of location and scale for the usual mean and variance, it should be possible to obtain test statistics which are insensitive to the combined effects of variance heterogeneity and non-normality. When robust measures are used to estimate parameters of distribution of quality characteristics constructing limits, control chart procedures become robust. Rocke (1989) suggested that in order to identify outliers easily, limits of a control chart should frame based on robust measures while non-robust measures should plotted on it.

Iglewicz and Langenberg (1986) proposed mean and range charts with control limits determined by trimmed mean of the subgroup means and the trimmed mean of the range. Rocke (1989) proposed standard deviation control charts based on the mean or the trimmed mean of the subgroup ranges or subgroup Inter Quartile Ranges. Several authors have developed robust control charts based on various measures of scale namely Median Absolute Deviation (MAD) (Abu-Shawiesh, 2008; Adekeye, 2012; Adekeye and Azubuike, 2012), Dispersion on M-estimate (Shahriari*et al.*, 2009), the Q_n and S_n estimates (Das, 2011), estimate obtained by the mean subgroup average deviation from the median MD (Schoonhoven and Does, 2012) and standard deviation of Modified Maximum Likelihood Estimator (Sindhumol and Srinivasan, 2015).

Trimmed mean (Tukey, 1948) and its standard error are more appealing because of its computational simplicity. Apart from that, these measures are less affected by departures from normality than the usual mean and standard deviation, as observations in the tail are removed. Standard error of trimmed mean is not sufficient to estimate process dispersion because of trimming subgroups (Dixon and Yuen, 1974). Standard estimator of variance of trimmed mean is obtained through Winsorization (Wilcox, 2012). Huber (1981) showed a jackknife estimator for its variance. Capéraà and Rivest (1985) derived an exact formula for variance of the trimmed mean as a function of order statistics, when trimming percentage is small.

As trimming makes a reduction in dispersion, estimating population dispersion based standard error of trimmed mean will not give a clear picture of actual dispersion. Variance of trimmed mean which is a function of order statistics or its variance modification based on Winsorization, are not helping in this regard. Hence trimmed standard deviation has a limited exposure to applications in literature. In this paper, authors made a modification to improve variance of the trimmed mean by multiplying it with a tuning constant to reduce the effect of loss due to trimming so that its robust qualities are not much disturbed. A robust control chart for controlling process dispersion is developed based on this modified measure.

2 Robust control chart based on trimmed standard deviation

It is interesting to see that about 1895 Mendeleev preferred to take the mean of the middle third to test for harmony of the series of observation. One problem with the mean is that the tails of a distribution can dominate its value. In order to reduce the effect of tails of a distribution, it can be simply removed. The γ-trimmed mean is

$$\mu_\gamma = \frac{1}{1-2\gamma}\int_{x_\gamma}^{x_{1-\gamma}} x dF(x)\,. \tag{1}$$

Let $x_{(1)} \leq x_{(2)} ... \leq x_{(n)}$ denote an order statistics sample of size n, from a population having symmetric distribution. The r-times symmetrically trimmed sample is obtained by dropping both r-lowest and r-highest values. Here $r = [\gamma n]$ is the greatest integer and trimming is done for $\gamma\%$ $(0 \leq \gamma \leq .5)$ of n. Trimmed mean is defined as

$$\overline{x}_T = \frac{1}{n-2r}\sum_{i=r+1}^{n-r} x_{(i)}. \tag{2}$$

Sample standard deviation of observations from trimmed mean is

$$s_T = \sqrt{\frac{1}{n-2r-1}\sum_{i=r+1}^{n-r}\left(x_{(i)} - \overline{x}_T\right)^2}. \tag{3}$$

By assuming symmetric trimming and normal distribution

$$P\left\{\left|\frac{x-\mu}{\sigma}\right| \leq \frac{s_T}{\sigma}\right\} = 1-2\gamma \tag{4}$$

$$\Phi\left(-\frac{s_T}{\sigma}\right) = \gamma = 1-\Phi\left(\frac{s_T}{\sigma}\right) \tag{5}$$

$$\widehat{\sigma} = \left[\frac{1}{\phi^{-1}(1-\gamma)}\right] s_T \tag{6}$$

$$\widehat{\sigma}_T = s_T^* = 1.4826 s_T \tag{7}$$

The function Φ is the distribution function of standard normal. The authors considered a maximum trimming of 25%, for framing this constant, beyond which loss of information from the sample is high. Hence modified estimate of dispersion using trimmed mean and the tuning constant are relevant and meaningful. It is interesting that this is the same constant introduced by Hampel (1974) for Median Absolute Deviation (MAD) and is equivalent to the trimmed mean of 25% for a symmetric trimming.

If σ^2 is the population variance, Mean Square Error (MSE) of its estimators based on sample variance s^2 and trimmed variance s_T^{2*} are

$$MSE\left(s^2\right) = \left(\frac{1}{c_4^2}\right)\left(1 - c_4^2\right)\sigma^2 \tag{8}$$

$$MSE\left(s_T^{2*}\right) = (1.4826)^2\left(\frac{1}{c_4^2}\right)\left(1 - c_4^2\right)\sigma^2. \tag{9}$$

The constant c_4 is a function of sample size n only and same tables for control chart parameters can be used to get these values.

$$c_4\left(n\right) = \frac{\Gamma\left[\frac{n}{2}\right]\sqrt{\frac{2}{n-1}}}{\Gamma\left[\frac{n-1}{2}\right]}. \tag{10}$$

Relative Efficiency (RE) of the estimator based on variance of the trimmed mean is

$$RE = \frac{MSE\left(s\right)}{MSE\left(s_T^*\right)} = 2.19810276. \tag{11}$$

The RE is also influenced by the percentage of trimming, while estimating s_T^*. So a simulation study is conducted by generating samples of sizes $n = 5, 8, 10$ and 20 from $N(0, 1)$. Sample $s.d(s)$ and this robust $s.d\left(s_T^*\right)$ were calculated and $10,000$ runs were simulated for each sample using SAS software. Variance of s and s_T^* were calculated across the runs and RE of s_T^* were calculated. In order to calculate simulation error of RE of s_T^*, this was repeated 10 times and standard deviation of the 10 RE values were calculated. The results of this simulation study are given in Table 1. It shows that efficiency will also improve if sample size and trimming percentage are increased.

Table 1: Relative Efficiency of the modified robust estimator s_T^*.

Sample size n	**RE** (trim 10%)	**Simulation Error**	**RE** (trim 20%)	**Simulation Error**
5	0.47909	0.00522	0.59384	0.00801
8	0.49552	0.00584	0.64991	0.01054
10	0.59897	0.00642	0.72579	0.00975
20	0.60686	0.00647	0.76741	0.01223

This paper considered two phases of control charting procedure. Preliminary samples are considered for constructing limits in Phase I. 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.

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

$$\mathrm{CL} = \mathrm{c}_4\sigma; \quad \mathrm{LCL} = \mathrm{c}_4\sigma - 3\sqrt{1-\mathrm{c}_4^2} = \mathrm{B}_5\sigma; \quad \mathrm{UCL} = \mathrm{c}_4\sigma + 3\sqrt{1-\mathrm{c}_4^2} = \mathrm{B}_6\sigma. \tag{12}$$

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

$\overline{s} = \sum_{i=1}^{m} s_i/m$, where $s_i = \left(\frac{1}{n-1}\sum_{j=1}^{n}(x_{ij}-\overline{x})^2\right)^{1/2}$ and an estimate of process dispersion using it is $\widehat{\sigma} = \overline{s}/\mathrm{c}_4$.

$$\mathrm{CL} = \overline{s}; \quad \mathrm{LCL} = \left(1 - \frac{3}{\mathrm{c}_4}\sqrt{1-\mathrm{c}_4^2}\right)\overline{s} = \mathrm{B}_3\overline{s}; \quad \mathrm{UCL} = \left(1 + \frac{3}{\mathrm{c}_4}\sqrt{1-\mathrm{c}_4^2}\right)\overline{s} = \mathrm{B}_4\overline{s}. \tag{13}$$

When σ is estimated from the preliminary samples using average of the robust estimator s_T^* say, $\widehat{\sigma} = \overline{s_T^*}/\mathrm{c}_4$ control limits are

$$\mathrm{CL} = \overline{s_T^*}; \quad \mathrm{LCL} = \mathrm{B}_3\overline{s_T^*} \text{ and } \mathrm{UCL} = \mathrm{B}_4\overline{s_T^*} \tag{14}$$

As sample size is same, there is no change of control chart constants and table used for classical control chart can be used here also. If trimmed standard deviation is directly taken for framing limits, it will not cover process dispersion. The loss of information due to trimming is covered by multiplying a tuning constant with standard deviation of the trimmed data. This constant is selected in such a way that, it will compensate the loss due to maximum allowable trimming.

In order to fix this tuning constant, a simulation study is conducted and validated the effect multiplier 1.4826 on modified trimmed standard deviation to control process. The constants 0.7803041 (10% trimming), 1.1881829 (20% trimming), 1.4826022 (25% trimming) and 1.9069394 (30% trimming) are used as multipliers to develop control limits and compared its performance with classical non-robust limits. Limits are verified based on its efficiency to detect outlier with a minimum of false detection without disturbing its robustness.

Figure 1 shows a comparison on positions of various UCL based on above four multipliers named as UCL_1, UCL_2 ,UCL_3 and UCL_4 respectively, along with classical UCL, for a particular trimming and for a fixed subgroup size, say $n = 10$. Limits are calculated for 20 random data sets, each contains 20 subgroups of size 10. Position of UCL_3, based on the multiplier 1.4826 provide the lowest UCL one can think about that covers process dispersion without the loss of robustness, for all levels of trimming except for 10% trimming. When trimming is small the limits are coinciding with that of s-chart, which is non-robust.

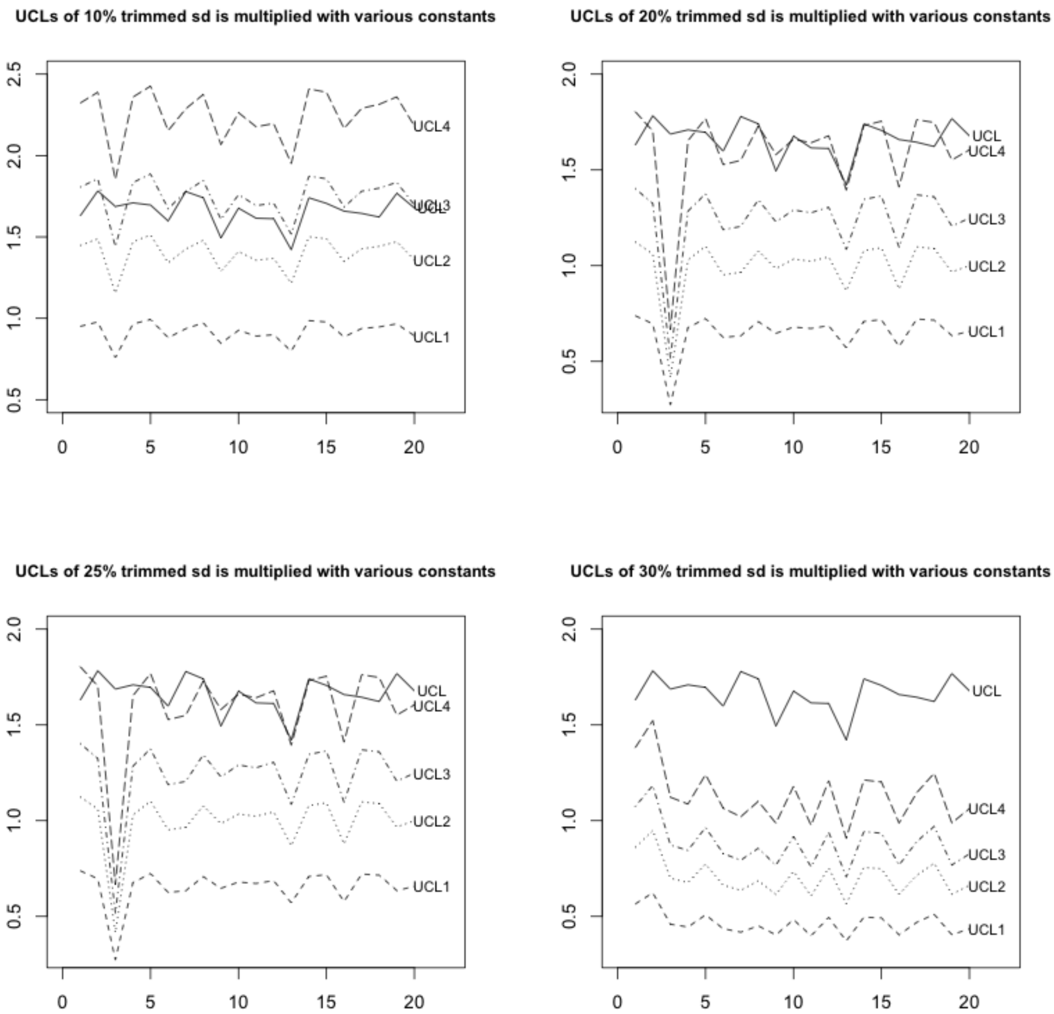

Figure 1: Comparison of UCLs based on different multipliers for trimmed s.d along with UCL of s-chart for 20 random data sets each contains 20 subgroups of size 10 each.

Table 2 contains average width of s-chart along with average width of control charts calculated with various multipliers against four levels of trimming, calculated for 20 random data sets, each containing 20 subgroups of size 10. The trimmed s.d chart with multiplier 1.90694 is close to the non-robust s-chart. It is clear that except for small percentage of trimming, s_T^*-chart with multiplier 1.4826 gives the maximum width that can represents process dispersion without the loss of robustness. For small sample size, maximum trimming one can think about is 10% and for large sample it is 25%. In such case, the tuning constant introduced by the authors, provide the most narrowed interval one can think about that covers process dispersion as well as detection of outliers beyond which reduces its robust nature.

Table 2: Comparison of average width of control charts with s-chart.

	Average width of control charts				
Multipliers	**1**	**0.78034**	**1.18818**	**1.4826**	**1.90694**
10%	1.387746	0.767212	1.168178	1.457642	1.874799
20%	1.387746	0.544989	0.829816	1.035436	1.331764
25%	1.387746	0.544989	0.829816	1.035436	1.331764
30%	1.387746	0.386639	0.588707	0.734583	0.944810

3 Performance of robust control chart

An empirical study based on Monte Carlo simulation is conducted for 1,000 runs using SAS software and random samples are simulated from $N(0, 1)$. The study considers subgroup sizes of $n = 10$ and 20 and number of subgroups considered are 20 in each case. Two levels of symmetric trimming are considered for each subgroup at 10% ($r = 1$) and 20% ($r = 2$). Random subgroups generated from $N(0, 1)$ are considered as data from in-control state and sample standard deviations are calculated from it to test against these control limits. Any value of the sample statistic lie outside the limits is considered as a false-alarm. In order to study the effect of detection of outliers or assignable causes of variation, two out-of-control situations are created, one is based on random samples taken from $N(0, 2)$ and the other is from random samples from $N(0, 4)$.

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 variable having distribution function $\varphi(x)$. Then for any constant $K > 0$, $\varphi(x/K)$ is a normal distribution with standard deviation K. Let ϵ be any constant, $0 \leq \epsilon \leq 1$. The contaminated normal distribution is $\mathrm{H}(x)) = (1 - \epsilon)\, \varphi(x) + \epsilon\varphi(x/K)$ which has mean 0 and variance $(1 - \epsilon + \epsilon K)$. Contamination is made for 40% of the subgroup so that 10% and 20% censoring will be meaningful for a contaminated subgroup. The mixture models considered are $0.60N(0, 1) + 0.40N(0, 2)$, and $0.60N(0, 1) + 0.40N(0, 4)$. Also 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.

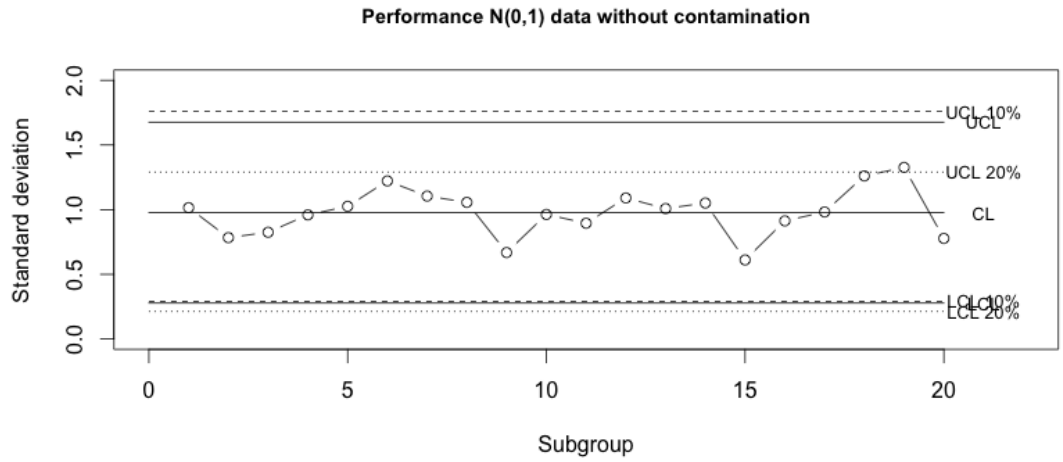

Figure 2: Comparison of s-chart, s_T^*-charts for 10% and 20% trimming for data from $N(0,1)$.

Figure 2 shows a comparison of performance of s-chart with two other charts based on s.d from trimmed data for two levels of trimming. Subgroup size $n = 10$ and number subgroups taken is $m = 20$. When trimming is small, limits of s-chart and trimmed s.d chart and are nearer and have lost its robust nature. When trimming is increased to 20%, width of the limits are lesser than the classical s-chart and more robust in nature. Figure 3 shows charts performance in terms of outlier detection. The robust chart based on standard deviation of 20% trimming is detecting more points compared to s-chart.

The Average Run Length (ARL) is evaluated as the average number of points that must be plotted before a point indicated out-of-control signal. It is the reciprocal of the probability that any point exceeds the control limits. When the process is in-control, ARL of Shewharts-control chart is expected to be close to 370. A simulation study is done to compare ARL for both in-control and out-of-control status. A 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_T^*-control chart were constructed. After determining the control limits, random $N(0,1)$ subgroups of size n were generated. The s statistic is 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 is recorded as a run length observation, RL_i. For runs not signaling by the $1,000^{th}$ subgroup, the run length was recorded as $1,000$. This process is repeated $10,000$ times and the results of this simulation study are given in Table2. The ARL_0 is calculated as

$$\mathrm{ARL}_0 = \frac{1}{10000} \sum_{i=1}^{10000} \mathrm{AR}_i. \tag{15}$$

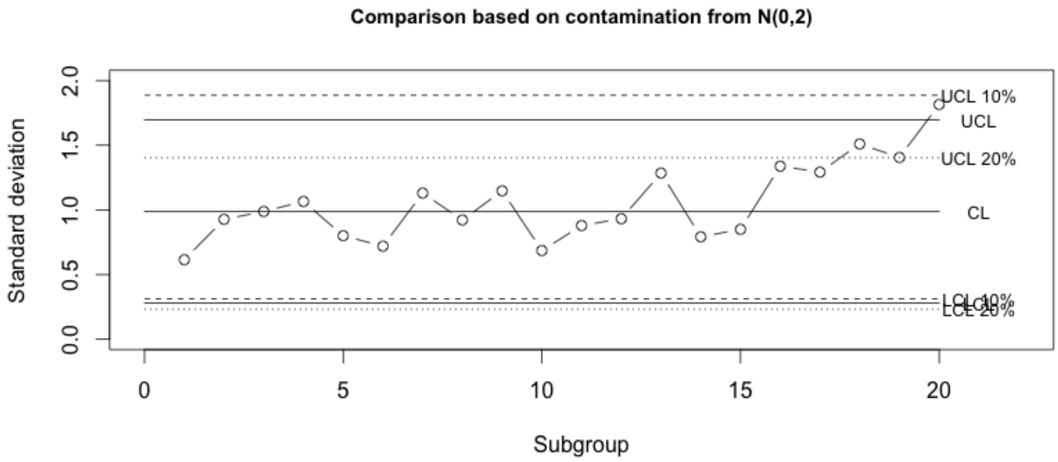

Figure 3: Performance of s-chart and s_T^*-charts for 10% and 20% trimming for $N(0,1)$ data with outlier from $N(0,2)$.

The same procedure is used to compare for the out-of-control ARL. If the process is out-of-control, its ARL_1, to be small. The control limits for the ARL_0 are based on $N(0,1)$ distribution. The observations used to calculate the statistics s are from a normal distribution with mean $\mu = 0$ and standard deviation $\lambda\sigma$, where λ representing the size of the shift in the standard deviation. Example shows for $\lambda = 2$ and 4. Number of samples required for the value of s estimator to exceed the control limits is recorded as a run length observation, RL_i. 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.

Table 3 shows that among ARL_1 of dispersion charts, s_T^*-chart with 10% trimming is larger than that of s-chart. Hence small percentage of trimming can be advisable even for a controlled process. For out-of-control situations, ARL_1 is less for s_T^*-chart with 20% trimming than s-chart and hence this robust s_T^*-chart is preferable.

4 Conclusion

This study gives a modification to standard deviation of trimmed data in a realistic and more meaningful way by multiplying it with a suitable constant. This will improve relative efficiency of the estimate of dispersion, and control its loss of information due to trimming, while holding its robust nature. A simulation study shows that efficiency will also improve if sample size and trimming percentage are increased. Moreover, the strength of this constant is already proved by Hampel by multiplying it with MAD.

This paper also gives an application of this robust estimator of dispersion in statistical

Table 3: ARL for sample size $n = 10$ and $m = 20$, with $r = 1$ trimming 10% and $r = 2$ trimming 20%.

Distributions	s**-chart**	s_T^*	
		$r = 1$	$r = 2$
$N(0, 1)$	370	410	205
$N(0, 1.5)$	259	190.6	105
$N(0, 2)$	0.42	0.58	0.47
$N(0, 4)$	0.41	0.004	0.005
$0.60N(0, 1) + 0.40N(0, 2)$	3.02	4.26	0.62
$0.60N(0, 1) + 0.40N(0, 4)$	3.01	4.29	0.61

process control technique. Control limits are constructed and tested based on simulation study. Charts performance in both in-control and out-of-control situations in terms of detection of assignable causes or outlier is also carried out. A comparison of this chart for two levels of trimming with classical s-chart is done. Moreover, performance is evaluated based on both in-control and out-of-control ARL also. In all these situations, this robust control chart performance is remarkable.

Acknowledgement

This work was financially supported by 2013 funds of the University of Naples - L'Orientale (I).

References

Abu-Shawiesh, M. O. (2008). A simple robust control chart based on MAD. Journal of Mathematics and Statistics, 4(2), 102.

Adekeye, K. S. (2012). Modified Simple Robust Control Chart Based on Median Absolute Deviation. International Journal of Statistics and Probability, 1(2), p91.

Adekeye, K. S., & Azubuike, P. I. (2012). Derivation of the limits for control chart using the median absolute deviation for monitoring non-normal process. Journal of Mathematics and Statistics, 8(1), 37-41.

Capéraà, P., & Rivest, L. P. (1995). On the variance of the trimmed mean. Statistics & probability letters, 22(1), 79-85.

Das, N. (2011). Control charts for controlling variability of non-normal processes. Economic Quality Control, 26(2), 121-131.

Dixon, W. J., & Yuen, K. K. (1974). Trimming and winsorization: A review. Statistische Hefte, 15(2-3), 157-170.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. Journal of the American Statistical Association, 69(346), 383-393.

Huber, P. J. (1981). Robust Statistics, 2 Ed. John Wiley.

Iglewicz, B., & Langenberg, P. (1986). Trimmed mean X-bar and R charts. Journal of Quality Technology, 18(3).

Riaz, M., & Saghir, A. (2007). Monitoring process variability using Ginis mean difference. Quality Technology and Quantitative Management, 4(4), 439-454.

Rocke, D. M. (1989). Robust control charts. Technometrics, 31(2), 173-184.

Schoonhoven, M., & Does, R. J. (2012). A robust standard deviation control chart. Technometrics, 54(1), 73-82.

Shahriari, H., Maddahi, A., & Shokouhi, A. H. (2009). A robust dispersion control chart based on M-estimate. Journal of Industrial and Systems Engineering, 2(4), 297-307.

Sindhumol, M.R., & Srinivasan, M.R. (2015). A simple robust dispersion chart based on MMLE, Journal of Mathematics, 11(4), 5-10.

Wilcox, R. R. (2012). Introduction to robust estimation and hypothesis testing. Academic Press.

Tukey, J. W. (1948). Some elementary problems of importance to small sample practice. Human Biology, 205-214.