

## A Simple Robust Control Chart Based on MAD

Moustafa Omar Ahmed Abu-Shawiesh  
Mathematics Department, Faculty of Science, Hashemite University,  
Al-Zarqa 13115, Jordan

---

**Abstract:** Control charts are one of the most powerful tools used to detect aberrant behavior in industrial processes. The Shewhart S-control chart is one of the most widely used statistical process control techniques developed to control the process variability based on the fundamental assumption that the underlying distribution of the quality characteristic is normal. When the underlying normality assumption is not met, the robust methods are one of the most commonly used statistical methods preferred in such situations. We presented a simple approach to robust estimation of the process standard deviation  $\sigma$  based on a very robust scale estimator, namely, the median absolute deviation from the sample median (MAD). The proposed method provides an alternative to the Shewhart S-control chart. A numerical example was given and a Monte Carlo simulation study was conducted to illustrate the performance of the proposed method and compare it with that of the Shewhart method. The proposed robust method was found to lead to a better performance than the Shewhart method and has good properties for the heavy tailed distributions and moderate sample sizes.

**Keywords:** Average run length, breakdown point, influence function, median absolute deviation

---

### INTRODUCTION

Scale estimators are very important in many statistical applications. The most common of scale estimators is the sample standard deviation,  $S$ . Many statistical quality control textbooks recommend the use of it for estimating the process standard deviation  $\sigma$  for a normally distributed random variable. Unfortunately,  $S$  is not necessarily the most efficient or meaningful estimator of scale in skewed and leptokurtic distributions and it is notably that  $S$  is non-robust to slight deviations from normality<sup>[1]</sup>. The most appropriate measure of scale often depends on estimator efficiency, performance of inferential methods in realistic settings and effectiveness in describing the most interesting aspect of distribution variation. The sample standard deviation,  $S$ , has good efficiency in platykurtic and moderately leptokurtic distributions but the classic inferential methods for  $S$  may perform poorly in realistically nonnormal distributions<sup>[2]</sup>. Therefore, we are looking for a scale estimator which is robust, explicit and easily computable as an alternative to  $S$ . The median absolute deviation from the sample median (MAD) is a more meaningful measure of variation and may be preferred to the sample standard deviation,  $S$ , in certain nonnormal distributions.

**Shewhart S-control chart:** The Shewhart S-control chart which is one of the most widely used statistical

process control technique developed to monitor the standard deviation of a production process  $\sigma$  in order to control the process variability. It is used as a standard practice to estimate  $\sigma$  the average of the subgroup standard deviation,  $\hat{\sigma} = \bar{S}/c_4$ , where the constant  $c_4$  is needed to make the estimator an unbiased estimator for  $\sigma$ . The fundamental assumption of the S-control chart is that the underlying distribution of the quality characteristic is normal, but unfortunately many processes, occur in practice, do not follow the normal distribution and due to the fact that the sample standard deviation,  $S$ , is non-robust to slight deviations from normality assumption, the need for alternatives to the Shewhart S-control chart comes to play.

**Robust Methods:** The robust methods are one of the most commonly used statistical methods when the underlying normality assumption is violated. These methods offer useful and viable alternative to the traditional statistical methods and can provide more accurate results, often yielding greater statistical power and increased sensitivity and yet still be efficient if the normal assumption is correct. By a robust estimator, we mean an estimator which is insensitive to changes in the underlying distribution and also resistant against the presence of outliers. The robust estimator is considered to be good if it has high efficiency, high breakdown point which is the maximum fraction of outliers an estimator can cope, redescending influence function

which measures how an estimator reacts to a small fraction of outliers and has low gross-error sensitivity which measures the worst influence a small amount of contamination of fixed size can have on the value of the estimator<sup>[3,4]</sup>. In this paper, we will restrict attention to estimator that have an explicit formula, always yield a unique estimate, have a 50% breakdown point, a bounded influence function and being easily computable and needs little computation time.

**Sample Standard Deviation:** The sample standard deviation,  $S$ , is the most commonly used measure of scale. It is defined as the square root of the mean of the squares of deviations from the common sample mean,  $\bar{X}$ . The main advantage of the sample standard deviation, is that, it can be regarded as truly representative of the data, since all data values are taken into account in its calculation, while the main disadvantage, is that, it is non-robust to slight deviations from normality and can be easily influenced by outliers. The breakdown point of the sample standard deviation for a sample of size  $n$  is merely  $1/n$ , that is, it can be destroyed by even a single outlier<sup>[5]</sup>.

**Median Absolute Deviation:** The median absolute deviation from the sample median (MAD) is considered one of the good robust estimators for scale which satisfies the above combination of requirements. Due to the good properties of this estimator, it will be used as an alternative to the sample standard deviation,  $S$ , in estimating the process standard deviation,  $\sigma$ , to construct a simple univariate robust control chart when the assumption of normality for the Shewhart S-control chart is not met.

**MATERIALS AND METHODS**

The median absolute deviation from the sample median (MAD) is a very robust scale estimator than the sample standard deviation. It measures the deviation of the data from the sample median. It promoted first by Hampel<sup>[6]</sup> who attributed it to Gauss. This estimator is simple and easy to compute. The MAD is often used as an initial value for the computation of more efficient robust estimators. The MAD for a random sample of size  $n$  observations  $x_1, x_2, \dots, x_n$  is defined as follows:

$$MAD = 1.4826MD\{|X_i - MD|\}, \quad i = 1, 2, \dots, n \quad (1)$$

where  $MD$  is the sample median. The statistic  $b_nMAD$  is an unbiased estimator of  $\sigma$  if the random sample  $x_1, x_2, \dots, x_n$  are normally distributed<sup>[7]</sup>. The correction factor

$b_n$  is given in Table 1 for different values of  $n$ . The main properties of the MAD, is that, it has a maximal 50% breakdown point which is twice as much as the interquartile range, IQR. The influence function of the MAD is bounded by the sharpest possible bound among all scale estimators. The gross-error sensitivity of the MAD is equal to 1.167 which is the smallest value that one can obtain for any scale estimator in the case of the normal distribution. The MAD efficiency at normal distributions is equal to 37%. Let  $x_{i1}, x_{i2}, \dots, x_{in}$  be a random sample of size  $n$  of independent observations taken at period  $i$ , where  $i = 1, 2, \dots$ . The samples are assumed to be equal, independent and taken from continuous identical distribution functions. The proposed robust control chart based on the MAD estimator is a chart of subgroup standard deviations in which the control limits for the sake of robustness are set using the median absolute deviation from the sample median (MAD). Thus we may set the control limits and central line for the Shewhart S-control chart based on

$$\overline{MAD} = \sum_{i=1}^m MAD_i / m \text{ as follows:}$$

$$\begin{aligned} LCL &= c_4 \hat{\sigma} - 3\hat{\sigma} \sqrt{1 - c_4^2} \\ &= c_4 b_n \overline{MAD} - 3b_n \overline{MAD} \sqrt{1 - c_4^2} \\ &= B_3 b_n \overline{MAD} = B_3^* \overline{MAD} \end{aligned} \quad (2)$$

$$CL = c_4 \hat{\sigma} = c_4 b_n \overline{MAD} = c_4^* \overline{MAD} \quad (3)$$

$$\begin{aligned} UCL &= c_4 \hat{\sigma} + 3\hat{\sigma} \sqrt{1 - c_4^2} \\ &= c_4 b_n \overline{MAD} + 3b_n \overline{MAD} \sqrt{1 - c_4^2} \\ &= B_6 b_n \overline{MAD} = B_6^* \overline{MAD} \end{aligned} \quad (4)$$

where,

$$c_4 = \frac{E(\bar{S})}{\sigma} \approx \frac{4(n-1)}{4n-3}, c_4^* = b_n c_4 \quad (5)$$

$$\begin{aligned} B_3^* &= b_n B_3 = b_n (c_4 - 3\sqrt{1 - c_4^2}) \\ B_6^* &= b_n B_6 = b_n (c_4 + 3\sqrt{1 - c_4^2}) \end{aligned} \quad (6)$$

The values of the control limit factors  $c_4^*$ ,  $B_3^*$  and  $B_6^*$  are calculated and given in Table 1. Now, after the LCL and UCL and the central line CL, are calculated, the values of  $S_i$ ,  $i = 1, 2, \dots, m$ , are plotted on the chart. If any of the plotted  $S_i$ 's is falling outside the control limits, the process is considered to be out of control.

Table 1: The control limit factors for the proposed robust univariate variable control chart

n	$b_n$	$c_4^*$	$B_5^*$	$B_6^*$
2	1.196	0.954	0.000	3.117
3	1.495	1.325	0.000	3.403
4	1.363	1.256	0.000	2.846
5	1.206	1.134	0.000	2.369
6	1.200	1.142	0.035	2.249
7	1.140	1.094	0.129	2.059
8	1.129	1.089	0.202	1.977
9	1.107	1.073	0.257	1.890
10	1.087	1.057	0.300	1.814
11	1.078	1.051	0.337	1.765
12	1.071	1.047	0.371	1.724
13	1.066	1.044	0.399	1.690
14	1.061	1.041	0.423	1.658
15	1.056	1.037	0.445	1.630
16	1.053	1.036	0.463	1.607
17	1.049	1.033	0.480	1.585
18	1.047	1.032	0.497	1.566
19	1.044	1.030	0.512	1.548
20	1.042	1.028	0.525	1.532
21	1.040	1.027	0.537	1.517
22	1.038	1.026	0.548	1.503
23	1.036	1.024	0.558	1.490
24	1.034	1.023	0.568	1.478
25	1.033	1.022	0.577	1.467

Table 2: The control chart comparison

Control Chart	Shewhart Using $S/c_4$	Robust Using $b_nMAD$
LCL	0.27221	0.27928
CL	0.95847	0.98400
UCL	1.64473	1.68871
Length	1.37252	1.40943

**RESULTS AND DISCUSSION**

**Numerical example:** A randomly generated data from the normal distribution  $N(0,1)$  is used to illustrate the two methods. The data consisting of  $m = 30$  subgroups of size  $n = 10$  observations each. The summary information regarding the control limits, central line and length for the  $S$  and the proposed robust  $MAD$  control charts are calculated and given in Table 2 where we can note that both methods lead to the same state of control and approximately same length with wider length in the case of the proposed robust method. Only minor differences in the calculated control limits are observed and this is to be expected since the data used are normally distributed.

Figure 1 indicates that the process for the both types of the control charts is out-of-control.

**Monte Carlo Simulation Study:** As indicated by Langenberg and Iglewicz<sup>[8]</sup>, the processes generated by distributions with heavier tails than the normal are of special interest. Such processes tend to have more than the expected number of points falling outside the control limits. It would thus be instructive to study the consequences of using the proposed robust method for nonnormal observations. Therefore, to evaluate the performance of the proposed robust control chart a computer programs written in FORTRAN are used. The performance of the proposed robust control chart is compared to that of the Shewhart  $S$ -control chart under

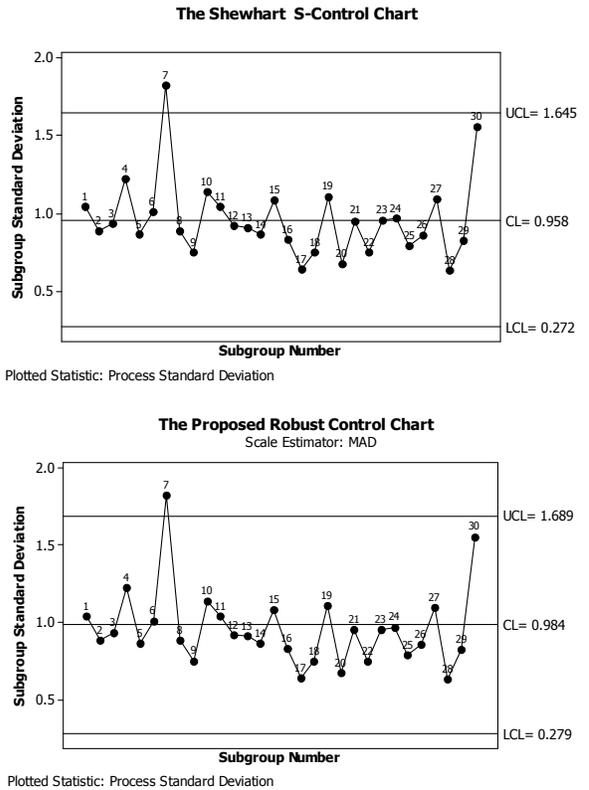


Fig. 1: The shewhart  $S$  and the proposed robust  $MAD$  control charts

four different distributions using equal sample size. In addition to the normal distribution, the logistic, double exponential and the Cauchy distributions are considered in our simulation. These distributions are typical of previous literature regarding robust control charts and represent a wide range of distributions extending to the extremely heavy-tailed (Cauchy distribution). Without loss of generality, special cases are considered where each distribution is centered at 0 with variance 1. For each distribution, 1000 trials of 20 subgroups of size 5 and 10 were run. The mean values of dispersion measures used are given in Table 3, while the variance values of them are given in Table 4. These values are used in determining the control limit values given later. It is not proper to compare the measures directly, since each measure something different. From the results, the sample standard deviation,  $S$ , has always the largest

Table 3: The mean values for dispersion

Distribution	S		MAD	
	n = 5	n = 10	n = 5	n = 10
Normal	0.944	0.975	0.557	0.616
Logistic	0.927	0.964	0.527	0.568
DE	0.903	0.949	0.471	0.487
Cauchy	16.08	20.88	1.202	1.101

Table 4: The variance of statistics

Distribution	S		MAD	
	n = 5	n = 10	n = 5	n = 10
Normal	0.006	0.003	0.005	0.003
Logistic	0.008	0.004	0.005	0.002
DE	0.010	0.006	0.005	0.002
Cauchy	2986.7	5454.0	0.073	0.019

Table 5: Control interval width for the control charts

Distribution	S/c <sub>4</sub>		b <sub>n</sub> MAD	
	n = 5	n = 10	n = 5	n = 10
Normal	1.971	1.396	1.955	1.383
Logistic	1.937	1.381	1.850	1.276
DE	1.886	1.358	1.654	1.092
Cauchy	33.59	29.90	4.220	2.471

Table 6: Number of points outside control limits

Distribution	S/c <sub>4</sub>		b <sub>n</sub> MAD	
	n = 5	n = 10	n = 5	n = 10
Normal	4	4	4	4
Logistic	14	17	21	36
DE	28	38	53	128
Cauchy	44	728	338	816

variance for all distributions and all sample sizes considered. Even that, the difference is not so big.

Table 5 presents the confidence interval width and Table 6 presents the number of points falling outside the control limits for each method. From the results, we note that for the case of the normal distribution, both methods leads approximately to the same length and also all have the same number of points falling outside the control limits for different sample sizes. As the tail weight of the distribution increases, the proposed method leads to shorter control limits and more number of points falling outside the control limits.

**The Average Run Length Simulation Study:** The criterion used to evaluate the performance of a control chart is how quick it can detect the change in the process and this can be obtained by calculating the Average Run Length (ARL) which is the number of points plotted within the control limits before one gets out. Under the normality assumption and for shewhart control charts like S, it is expected that 370.4 points would be plotted on the chart within the 3σ control limits before one gets out.

Table 7: The ARL<sub>0</sub> for S and the proposed robust MAD control charts

Control chart	n	Distribution function			
		Normal	Logistic	DE	Cauchy
Shewhart based on S/c <sub>4</sub>	5	370.2	375.4	377.3	369.6
	10	370.7	373.7	376.4	364.7
Robust based on b <sub>n</sub> MAD	5	370.2	370.3	370.5	370.3
	10	370.8	370.9	370.8	370.3

**The In-Control Average Run Length:** If the process is in-control, we want the in-control average run length, ARL<sub>0</sub>, to be large. Let N (μ, σ<sup>2</sup>) be a normal distribution, then sets of m = 20 subgroups consisting of n = 5 and n = 10 observations were generated from N (0, 1) distribution. The control limits for the S-control chart and the proposed MAD 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, RL<sub>i</sub>. For runs not signaling by the 25,000<sup>th</sup> subgroup, the run length recorded as 25,000. The same procedure is used to compare the Shewhart S-control chart and the proposed MAD control chart for the logistic, double exponential and the Cauchy distributions. The control charts robust to the assumption of normality should exhibit relatively stable ARL<sub>0</sub> for the four different nonnormal distributions. This process is repeated 1000 times and the results for this simulation study are given in Table 7. The ARL<sub>0</sub> was calculated as  $ARL_0 = \sum_{i=1}^{1000} RL_i / 1000$ . From the results

obtained in Table 7, we noticed that the ARL<sub>0</sub> of the proposed robust control charts seems to be stable for all studied distribution functions where the values of ARL<sub>0</sub> are very closed. The performance of the ARL<sub>0</sub> of the Shewhart S-control chart is affected by the type of the distribution function of the data, where there is a difference between the different values. In general the performance of the proposed robust method based on MAD estimator is better and very close to the nominal value 370.4.

**The out-of-control average run length:** If the process is out-of-control, we want the out-of-control average run length, ARL<sub>1</sub>, to be small. The control limits for the ARL<sub>1</sub> 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. The shifts in the standard deviation are from σ to λσ. Without loss of generality,

Table 8: The values of  $ARL_1$  for S and the robust MAD control charts from the normal distribution

Shift size $\lambda$	Sample size			
	5		10	
	Robust	Shewhart	Robust	Shewhart
1.0	370.16	370.15	370.82	370.68
1.2	191.35	158.4	104.3	94.7
1.5	53.25	39.15	16.2	13.6
2.0	17.55	13.45	2.55	2.15
2.5	1.25	1.15	1.1	1.05

Table 9: The Values of  $ARL_1$  for S and the robust MAD control charts from the logistic distribution

Shift size $\lambda$	Sample size			
	5		10	
	Robust	Shewhart	Robust	Shewhart
1.0	370.30	375.38	370.92	373.72
1.2	42.5	132.1	27.33	93.57
1.5	3.85	13.75	2.1	5.13
2.0	1.2	2.5	1.03	1.4
2.5	1.1	1.6	1	1.13

the values of  $\lambda = 1.0, 1.2, 1.5, 2.0$  and  $2.5$ . The number of samples required for the value of S estimator to exceed the control limits was recorded as a run length observation,  $RL_i$ . For runs not signaling by the 25,000<sup>th</sup> sample, the run length was reported as 25,000. This process was repeated 1,000 times. The  $ARL_1$  was calculated as  $ARL_1 = \frac{\sum_{i=1}^{1000} RL_i}{1000}$ . Random samples of size  $n = 5$  and  $10$  are drawn from the four different distributions considering the different values of  $\lambda$  and the results of simulation for the  $ARL_1$  are given in Table 8-1.

From the results in Table 8-11 which show the values of  $ARL_1$  for the proposed robust MAD and Shewhart S control charts, we conclude that when the data is normal, the robust control chart is only slightly less efficient than the corresponding Shewhart S control chart to detect shifts for all considered values of  $\lambda$ . However, this loss in efficiency is small when the sample size gets large. In such cases, it is recommended to use the Shewhart S control chart in the presence of normality. When the data is selected from a heavy tailed distribution as in the case of the double exponential and logistic distributions, the  $ARL_1$  of the proposed robust control chart is generally smaller than that for the Shewhart S control chart. The proposed robust control chart has the ability to detect shifts more quickly than the S control chart for all values of  $\lambda$  and sample sizes considered. Finally, when the data is selected from extremely heavy tailed distribution as in

Table 10: The values of  $ARL_1$  for S and the robust MAD control charts from the DE distribution

Shift size $\lambda$	Sample size			
	5		10	
	Robust	Shewhart	Robust	Shewhart
1.0	370.51	377.31	370.80	376.43
1.2	179.30	282.65	75.93	132.87
1.5	33.1	72.85	2.07	7.5
2.0	3.05	8.45	1.1	1.47
2.5	1.3	3	1	1.07

Table 11: The values of  $ARL_1$  for S and the robust MAD control charts from the Cauchy distribution

Shift size $\lambda$	Sample size			
	5		10	
	Robust	Shewhart	Robust	Shewhart
1.0	370.26	369.58	370.34	364.66
1.2	4.29	37.42	2.75	38.03
1.5	4.03	34.45	2.47	36.74
2.0	3.65	32.86	2.10	41.78
2.5	1.08	1.86	1.17	1.72

the case of the Cauchy distribution, the  $ARL_1$  of the proposed robust control chart is still in general smaller than that for the Shewhart S control chart for all considered cases. Therefore in the case of a heavy tailed distribution where a nonnormal distribution is present, it is recommended to use the proposed robust control chart as an alternative to the Shewhart S control chart.

## CONCLUSION

This article has presented a simple alternative robust univariate variable control chart for monitoring the process variability and the necessary table of factors for computing the control limits and the central line. The proposed robust control chart is based on the median absolute deviation from the sample median (MAD) which used for estimating the process standard deviation,  $\sigma$ , in the case of the Shewhart S control chart. The results of the numerical example show that the proposed robust method leads approximately to the same performance as the Shewhart method in the case of the normal data. The proposed method leads to wider control limits than the Shewhart method. True robust limits should be slightly larger in the case of the normal distribution<sup>[9]</sup>. The Simulation studies show that the proposed robust method has good properties for heavy tailed distributions and moderate sample sizes, where it leads to better performance than the Shewhart method, especially for the heavy tailed distributions, which seems to support the fact that these control charts are preferred in such situations. Finally, while we give an

alternative to the Shewhart S control chart based on MAD, other robust estimators alternatives to the MAD such as  $S_n$  and  $Q_n$  estimators proposed by Rousseeuw and Croux<sup>[7]</sup> can be applied to monitor the process variability and compare its performance to the method proposed in this study and to the Shewhart S control chart.

### ACKNOWLEDGEMENTS

The author would like to express his thanks to Mathematics Department in Hashemite University for the cooperation during the preparation of this study.

### REFERENCES

1. Tukey, J.W., 1960. A Survey of Sampling from Contaminated Distributions. In: Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling, Olkin, I., *et al.*, (Eds.). Stanford: Stanford University Press. pp: 448-485.
2. Bonett, D.G., 2006. Confidence interval for a coefficient of quartile variation. *Comput. Stat. Data Anal.*, 50: 2953-2957.  
DOI: 10.1016/j.csda.2005.05.007
3. Simonoff, J.S., 1984. A comparison of robust methods and detection of outliers techniques when estimating a location parameters. *Commun. Stat. Theory Methods*, 13: 813-842.  
DOI: 10.1080/03610928408828722
4. Lax, D.A., 1985. Robust estimators of scale: Finite-Sample performance in long-tailed symmetric distributions. *J. Am. Stat. Assoc., Theory Methods*, 80: 736-741.  
<http://www.amstat.org/publications/jasa/index.cfm?fuseaction=main>
5. Rousseeuw, P.J., 1991. Tutorial to robust statistics. *J. Chemometrics*, 5: 1-20.  
DOI: 10.1002/cem.1180050103
6. Hampel, F.R., 1974. The influence curve and its role in robust estimation. *J. Am. Stat. Assoc.*, 69: 383-393.  
<http://www.amstat.org/publications/jasa/index.cfm?fuseaction=main>
7. Rousseeuw, P.J. and C. Croux, 1993. Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.*, 80: 1273-1283.  
<http://www.amstat.org/publications/jasa/index.cfm?fuseaction=main>
8. Langenberg, P. and Iglewicz, B. 1986. Trimmed mean  $\bar{X}$  and R charts. *J. Qual. Technol.*, 18: 152-161. <http://www.asq.org-ASQ: Trimmed Mean X-Bar and R Charts>. QICID: 5537
9. Goodall, C., 1983. Examining Residuals. In: *Understanding Robust and Exploratory Data Analysis*, Hoaglin, D.C., F. Mosteller and J.W. Tukey (Eds.). John Wiley and Sons, New York. pp: 211-246.