Quality Quandaries: How to Set Up a Robust Shewhart Control Chart for Dispersion?

Hafiz Z. Nazir [a] , Marit Schoonhoven [b] , Muhammad Riaz [c d] & Ronald J. M. M. Does [b]

[a] Department of Statistics , University of Sargodha , Sargodha , Pakistan

[b] Institute for Business and Industrial Statistics, Department of Operations Management ,
Amsterdam Business School , Amsterdam , The Netherlands

[c] Department of Statistics , Quaid-i-Azam University , Islamabad , Pakistan

[d] Department of Mathematics and Statistics , King Fahad University of Petroleum and
Minerals , Dhahran , Saudi Arabia
Published online: 11 Dec 2013.

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

Quality Quandaries: How to Set Up a Robust Shewhart Control Chart for Dispersion?

Hafiz Z. Nazir[1],
Marit Schoonhoven[2],
Muhammad Riaz[3,4],
Ronald J. M. M. Does[2]

[1]Department of Statistics, University of Sargodha, Sargodha, Pakistan
[2]Institute for Business and Industrial Statistics, Department of Operations Management, Amsterdam Business School, Amsterdam, The Netherlands
[3]Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan
[4]Department of Mathematics and Statistics, King Fahad University of Petroleum and Minerals, Dhahran, Saudi Arabia

Edited by Ronald J. M. M. Does.

Address correspondence to Ronald J. M. M. Does, IBIS UvA, Plantage Muidergracht 12, 1018 TV, Amsterdam, The Netherlands. E-mail: r.j.m.m.does@uva.nl

INTRODUCTION

In real-world processes, performance indicators are monitored in order to control process performance. A change in the process characteristic can affect process performance and such changes should therefore be detected as soon as possible and prevented from occurring again. A control chart is a practical tool that can be used to visually monitor a process characteristic and provides objective rules to decide what action to take. Since the introduction of the control chart by Shewhart (1931), many enhancements and variants have been proposed in the literature. However, the general concept of the control chart has remained the same: the chart consists of a graph on which estimates of the relevant process parameter are plotted through time. Control limits, sometimes supplemented with additional decision rules, provide easy checks on the stability of the process parameter. When the estimated value falls outside the control limits, it is probable that the process has changed (i.e., is out of control). The operator should then investigate the cause of the change and adjust the process to the in-control state.

As noted above, the concept of the control chart was proposed by Shewhart in 1931. His charts are referred to as Shewhart control charts. The first charts are based on the assumption that the characteristic to be controlled is normally independently and identically distributed (i.i.d). When the characteristic has an i.i.d. normal distribution, it is common practice to monitor both its mean and its spread. These standard deviation and location Shewhart control charts are easy to construct and apply. After a fixed period of time, a sample is taken from the process and the statistic of interest is calculated and plotted on a chart with an upper control limit (UCL) and lower control limit (LCL), representing the extent of natural variation in the in-control statistic. The control limits are calculated as the average of the statistic plus or minus a multiplier times the standard deviation of the statistic. The multipliers are chosen such that the probability of a false signal and therefore time wasted on an unnecessary process adjustment is small, usually 0.0027. The standard deviation parameter of the process is controlled first, followed by the location parameter. Examples of such a standard deviation chart (S chart) and location chart ($\overline{X}$ chart) are given in Figures 1a and 1b, respectively.
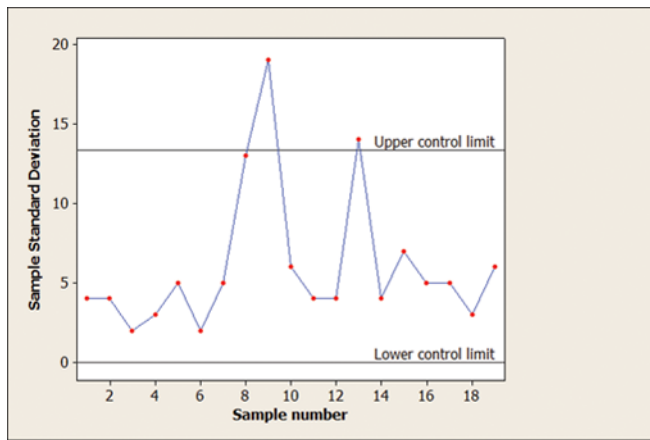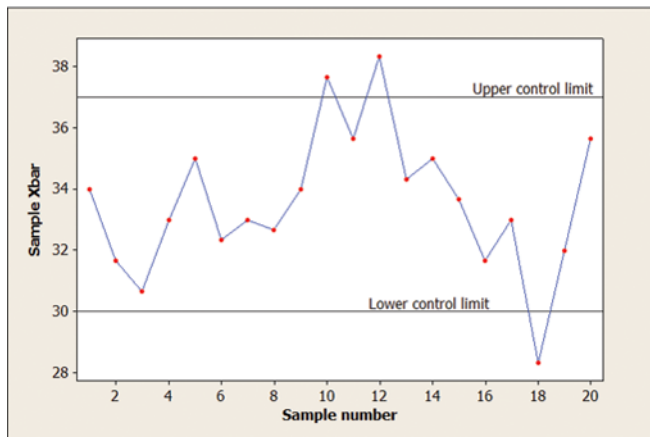
FIGURE 1 (a) Shewhart standard deviation control chart; (b) Shewhart location control chart. (Color figure available online.)

Shewhart control charts are powerful for detecting large parameter shifts (large means $\delta/(\sigma/\sqrt{n}) > 2$, with δ the shift, σ the in-control standard deviation, and n the sample size) but they are less effective when shifts are small. The reason is that, for the Shewhart chart, the charting statistic is based on information within the current sample only and not on information from previous samples. To overcome this drawback, Page (1954) introduced the cumulative sum (CUSUM) chart for the location and subsequently for the standard deviation as well (Page 1963). The CUSUM chart has an elephant-like memory, paying equal attention to the most ancient observation and the most recent. Roberts (1959) proposed a general variant that, depending on its smoothing parameter, can behave like a Shewhart or a CUSUM chart. This so-called exponentially weighted moving average (EWMA) control chart gives less and less weight to data as they get older

and older. Ng and Case (1989) developed an EWMA control chart for monitoring both the process mean and standard deviation. Hunter (1986) provided an insightful analysis of the different ways in which the Shewhart, CUSUM, and EWMA control charts use the history of data to detect a change in the process mean.

The present article focuses on controlling the standard deviation of a single key process characteristic using a Shewhart standard deviation control chart with estimated parameters. There is an extensive literature available on Shewhart control charts for monitoring the process dispersion with estimated parameters, including Maravelakis et al. (2002), Mahmoud et al. (2010), and Psarakis et al. (2013). The main objective of this article is to give practitioners a stepwise procedure on how to set up such dispersion charts. In another article we shall give the complementary control chart for the location. In control chart design, it is standard practice to separate Phase I from Phase II (cf. Vining 2009). During Phase I, control charts are used retrospectively to study historical data samples. Once representative samples are established, the parameters are estimated and control limits are determined and used for monitoring in Phase II. The following two sections describe these stages for the standard deviation control chart. Next, we demonstrate the procedure for a real-life data example. The final section offers some concluding remarks.

PHASE I PROCEDURE

The UCL and LCL of the Shewhart standard deviation control chart are given by

$$\widehat{UCL} = U\hat{\sigma}, \quad \widehat{LCL} = L\hat{\sigma}, \qquad [1]$$

with $\hat{\sigma}$ the estimated in-control standard deviation and U and L the constants such that the desired in-control performance is obtained. Usually, these constants are chosen such that the false alarm probability is sufficiently small, namely, 0.0027. Note that for normally distributed random variables the expectation and standard deviation of the sample standard deviation are linear functions of σ and hence the formula [1] makes sense. Recall that formula [1] can be used for control charts in Phase I as well as in Phase II. For the sake of clarity,

we shall add subscripts I and II to $\widehat{UCL}$, $\widehat{LCL}$, U, and L to indicate the phase to which we refer.

Schoonhoven and Does (2013) analyzed the performance of the Phase I control charts. They showed that the type of estimator used to construct these charts is important. A robust estimator should be selected first because then the Phase I limits are not affected by disturbances and therefore the correct data samples from which σ is estimated are retained. However, an efficient estimator of σ should be used to obtain the final estimate in order to ensure efficiency under normality. Below, we describe a practical step-by-step approach that meets these requirements.

Step 1: Select Phase I Data

We draw k samples of size n from the process when the process is assumed to be in control and we denote these samples by X_{ij}, $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n$. The subgroups (i.e., samples) should reflect random, short-term rather than special cause variation. To ensure this, items within a subgroup should be produced under conditions in which only random effects are responsible for the observed variation. Additional variability due to potential special causes such as a change in materials or personnel will then occur only between subgroups. Furthermore, the subgroup should not be selected over an interval that is too short because measurements may then be highly correlated and not represent just short-term variation.

However, in practice the k samples of size n may contain outliers, shifts, or other contaminations. These can be filtered out by following the next few steps.

Step 2: Construct a Phase I Standard Deviation Control Chart

We start with a robust estimator of the standard deviation based on the trimmed mean of the sample interquartile ranges. Estimate the 10% trimmed mean of the sample interquartile ranges, defined by

$$\overline{IQR}_{10} = \frac{1}{k - 2(\lceil k/10 \rceil - 1)} \times \left[\sum_{v=\lceil k/10 \rceil}^{k-\lceil k/10 \rceil+1} IQR_{(v)} \right],$$

[2]

where $\lceil z \rceil$ denotes the ceiling function (i.e., the smallest integer not less than z) and $IQR_{(v)}$ is the

TABLE 1 Constants for Phase I Procedure

n	$d_{\overline{IQR}_{10}}$	U_I	L_I	d_{IQR}	$d_{\bar{S}}$
3	1.644	2.923	0.042	1.692	0.998
4	2.020	2.525	0.108	2.060	0.997
5	0.951	3.220	0.035	0.990	0.980
6	1.253	2.688	0.093	1.284	0.983
7	1.490	2.403	0.154	1.514	0.985
8	1.683	2.225	0.208	1.704	0.986
9	1.122	2.474	0.146	1.144	0.984
10	1.293	2.281	0.198	1.312	0.985

vth ordered value of the sample interquartile ranges. The interquartile range of sample i is defined by

$$IQR_i = Q_{i,3} - Q_{i,1},$$

where $Q_{i,1} = X_{i,(a)}$ and $Q_{i,3} = X_{i,(b)}$, with $a = \lceil n/4 \rceil$, $b = n - a + 1$, and $X_{i,(v)}$ the vth-order statistic in sample i.

To obtain an unbiased estimate of σ from $\overline{IQR}_{10}$ we divide this quantity by $d_{\overline{IQR}_{10}}$, which is a normalizing constant. Values for this constant for various sample sizes are given in Table 1.

The Phase I standard deviation control chart limits are derived from

$$\widehat{UCL}_I = U_I \overline{IQR}_{10}/d_{\overline{IQR}_{10}}, \quad \widehat{LCL}_I = L_I \overline{IQR}_{10}/d_{\overline{IQR}_{10}},$$

with U_I and L_I the 0.99865 and 0.00135 quantiles of the distribution of IQR/d_{IQR} (see Table 1).

Step 3: Screen for Sample Shifts

Plot the IQR/d_{IQR}s of the Phase I samples on the standard deviation control chart generated in step 2 (charting the IQR instead of the sample standard deviation or the sample range ensures that localized variance disturbances are identified and samples that contain only one single outlier are retained).

Exclude from the Phase I data set all samples whose IQR/d_{IQR} falls outside the control limits.

Step 4: Construct a Phase I Individuals Chart

Update the spread estimate according to the formula

$$\overline{IQR}' = \frac{1}{k'} \sum_{i \in K} IQR_i \times 1_{\widehat{LCL_I} \leq IQR_i/d_{IQR} \leq \widehat{UCL_I}}(IQR_i), \quad [3]$$

H. Nazir et al.

with $1_D(x)$ the indicator function, K is the set of samples that are not excluded in step 3, and k' is the number of nonexcluded samples.

The next steps should be applied if individual outliers are likely. First, we construct the limits of the Phase I individuals control chart from

$$\widehat{UCL}_{ind} = 3\overline{IQR}'/d_{IQR}, \quad \widehat{LCL}_{ind} = -3\overline{IQR}'/d_{IQR}. \quad [4]$$

Step 5: Screen for Individual Outliers

Determine the residuals in each sample by subtracting the trimean from each observation in the corresponding sample (to filter out changes representing shifts in the standard deviation only): $resid_{ij} = X_{ij} - TM_i$ with

$$TM_i = (Q_{i,1} + 2Q_{i,2} + Q_{i,3})/4,$$

where $Q_{i,2}$ is the median of sample i. Note that the trimean is a robust estimate of the location (cf. Tukey 1997).

Plot the residuals on the individuals chart derived in step 4. Remove the observations from the Phase I data set corresponding to the residuals that fall outside the limits.

Step 6: Obtain the Final Estimate of the Standard Deviation

Obtain a new estimate of the standard deviation from the mean of the sample standard deviations

$$\overline{S}' = \frac{1}{k'} \sum_{i \in K} S_i'/c_4(n_i'), \quad [5]$$

with K the set of samples that are not excluded, k' is the number of nonexcluded samples, n_i' is the number of nonexcluded observations in sample i, S_i' is the standard deviation derived from the remaining n_i' observations

$$S_i' = \left(\frac{1}{n_i' - 1} \sum_{j: X_{ij} \in K_i} (X_{ij} - \overline{X}_i)^2 \right)^{1/2} \quad [6]$$

with K_i the set of retained observations in sample i and $c_4(n_i')$ defined by

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

Divide the result by $d_{\overline{S}}$ (see Table 1) to obtain an overall unbiased estimate.

PHASE II

Once the in-control reference data set is established, the standard deviation can be estimated from the Phase I data and control limits determined for use in Phase II. Recall that the Phase II control limits are given by [1] and $\hat{\sigma}$ is determined in step 6 of the Phase I procedure described in the previous section. What remains is the determination of the factors U_{II} and L_{II} in order to obtain the desirable in-control performance.

Schoonhoven et al. (2011) presented a formula for U_{II} and L_{II} of the Phase II standard deviation control chart based on the pooled mean of the sample standard deviations, $\tilde{S}$. They tested this formula for charts derived from a broad range of Phase I estimators and concluded that the formula is suitable when the variance of the estimator is close to the variance of $\tilde{S}$. Subsequently, Schoonhoven and Does (2013) showed that the variance of the estimator given in [5] is close to $\tilde{S}$. Hence, the formula presented by Schoonhoven et al. (2011) can also be applied in the procedure discussed here. This is a nice result because the formula is a plug-in so no simulations are required to obtain the constants.

In the next part, we continue with the steps in the approach applicable to Phase II.

Step 7: Construct a Phase II Standard Deviation Chart

Obtain U_{II} and L_{II} for the Phase II control limits from

$$U_{II} = \frac{\sqrt{F_{n-1,k(n-1)}(1 - \alpha/2)}c_4(k(n-1) + 1)}{c_4(n)} \quad [7]$$

and

$$L_{II} = \frac{\sqrt{F_{n-1,k(n-1)}(\alpha/2)}c_4(k(n-1) + 1)}{c_4(n)}, \quad [8]$$

TABLE 2 Constants for Phase II Procedure

n	$c_4(n)$	$k = 20$		$k = 50$	
		U_{II}	L_{II}	U_{II}	L_{II}
3	0.886	3.138	0.041	2.992	0.041
4	0.921	2.625	0.107	2.535	0.108
5	0.940	2.352	0.171	2.286	0.172
6	0.952	2.178	0.227	2.126	0.228
7	0.959	2.055	0.274	2.012	0.276
8	0.965	1.963	0.314	1.926	0.316
9	0.963	1.890	0.349	1.858	0.351
10	0.973	1.832	0.378	1.803	0.380

with $F_{v, w}$ denoting an F distribution with v numerator degrees of freedom and w denominator degrees of freedom and α the desired false alarm probability (usually 0.0027). Values for U_{II} and L_{II} when $n = 3,\ldots, 10$ and $k = 20, 50$ are provided in Table 2.

Substitute the values for $\hat{\sigma}$, U_{II}, and L_{II} into the formula for the standard deviation control limits given by [1].

Step 8: Use the Phase II Chart for Online Monitoring

Periodically collect newly available data (Y_{ij} with $i = 1, 2, 3,\ldots$ and $j = 1, 2,\ldots, n$) and calculate $S_i/c_4(n)$ according to [6] (use S in Phase II because this estimator is efficient under normality and sensitive to disturbances). When $S_i/c_4(n)$ falls outside the control limits, look for the cause of the out-of-control signal.

APPLICATION TO A REAL-WORLD DATA EXAMPLE

In this section we demonstrate the Phase I and II procedures detailed above.

Step 1: Select Phase I Data

Our data set was supplied by Wadsworth et al. (2001, pp. 235–237). The operation concerns the melt index of a polyethylene compound. The data consist of 20 subgroups of size 4 (Table 3). In Figure 2 the individual value plot of the observations versus sample number is given.

We use the first 19 samples to demonstrate the Phase I calculations and the last sample to illustrate Phase II monitoring.

TABLE 3 Melt Index Measurements

Sample	Observations				IQR/d_{IQR}
1	218	224	220	231	6.31
2	238	236	247	234	6.31
3	280	228	228	221	28.65
4	210	249	241	246	18.94
5	243	240	230	230	6.31
6	225	250	258	244	16.03
7	240	238	240	243	2.43
8	244	248	265	234	15.06
9	238	233	252	243	9.23
10	228	238	220	230	8.74
11	218	232	230	226	6.80
12	226	231	236	242	7.77
13	224	221	230	222	4.37
14	230	220	227	226	4.86
15	224	228	226	240	7.77
16	232	240	241	232	4.37
17	243	250	248	250	3.40
18	247	238	244	230	8.26
19	224	228	228	246	10.68
20	236	230	230	232	

Step 2: Construct a Phase I Standard Deviation Control Chart

First, the 10% trimmed mean of the sample interquartile ranges is derived from formula [2]. This leads to a value of 17.59. To obtain an unbiased estimate of σ, we divide this value by $d_{\overline{IQR}_{10}}$, or 2.02 as shown in Table 1. This yields a final value of 8.71.

The constants U_I and L_I for the Phase I standard deviation control limits are 2.53 and 0.11, respectively (see Table 1). The resulting upper and lower control limits are 22.03 and 0.94, respectively.
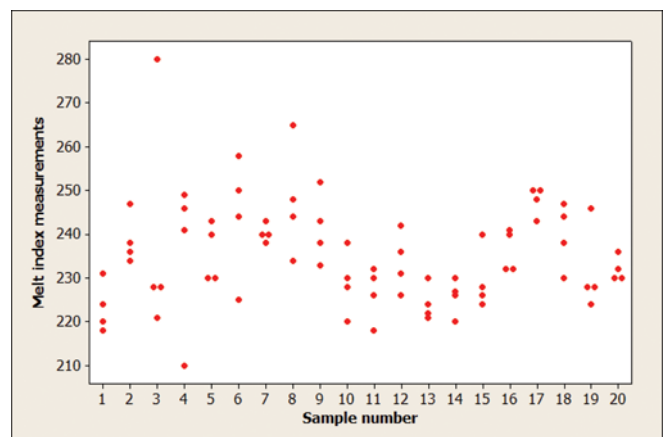
FIGURE 2 Individual value plot of 20 samples of the melt index. (Color figure available online.)
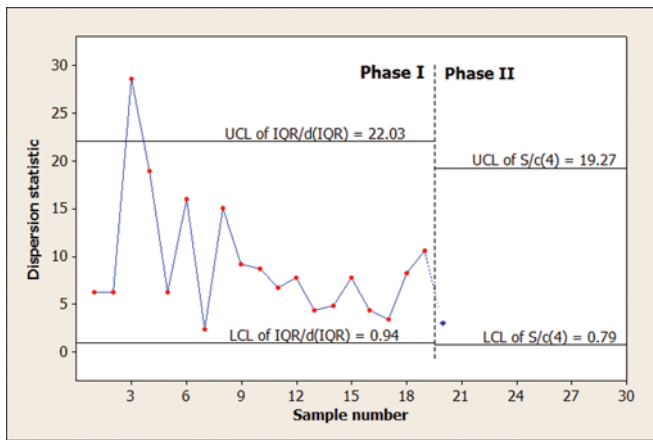
FIGURE 3 Phase I (and Phase II) standard deviation control chart for the melt index. (Color figure available online.)

Step 3: Screen for Sample Shifts

The goal of this step is to filter out localized disturbances. To do this, we first determine the sample IQR/d_{IQR}s (see Table 3). Sample 3 does not meet the upper limit of the Phase I standard deviation control chart and is therefore removed from the Phase I data set (see Figure 3).

Step 4: Construct a Phase I Individuals Chart

We now update the spread estimate according to [3], which means that we take the sum of the interquartile ranges of all of the samples excluding sample 3. The mean of the nonexcluded interquartile ranges is 16.89. The upper and lower control limits of the individuals chart are calculated from [4], giving 24.59 and −24.59.

Step 5: Screen for Individual Outliers

We determine the residuals by subtracting the trimean from each observation in the corresponding sample. The resulting residuals of the 18 nonexcluded Phase I samples are presented in Table 4. The first residual of the fourth sample (−26.50) falls outside the control limits and is removed from the Phase I data set (see Figure 4).

Step 6: Obtain the Final Estimate of the Standard Deviation

In this step, we obtain the final estimate of σ from [5]. This means that we determine $S_i/c_4(n)$ for every

TABLE 4 Residuals of Melt Index Measurements

Sample	Residuals			
1	−5.25	0.75	−3.25	7.75
2	−0.75	−2.75	8.25	−4.75
4	−26.50	12.50	4.50	9.50
5	7.25	4.25	−5.75	5.75
6	−19.25	5.75	13.75	−0.25
7	−0.25	−2.25	−0.25	2.75
8	−3.75	0.25	17.25	−13.75
9	−3.50	−8.50	10.50	1.50
10	−1.00	9.00	−9.00	1.00
11	−8.50	5.50	3.50	−0.50
12	−7.75	−2.75	2.25	8.25
13	−0.25	−3.25	5.75	−2.25
14	4.25	−5.75	1.25	0.25
15	−5.50	−1.50	−3.50	10.50
16	−4.25	3.75	4.75	−4.25
17	−4.75	2.25	0.25	2.25
18	7.25	−1.75	4.25	−9.75
19	−7.50	−3.50	−3.50	14.50

sample except sample 3, which has been eliminated. Note that sample 4 contains three instead of four observations, because the first observation was excluded when the sample was screened with the individuals chart. The mean of these sample standard deviations is divided by the unbiasing constant (0.997 as shown in Table 1), leading to a final value of 7.32.

Step 7: Construct a Phase II Standard Deviation Chart

The values for the constants U_{II} and L_{II} used to calculate the limits of the Phase II standard deviation
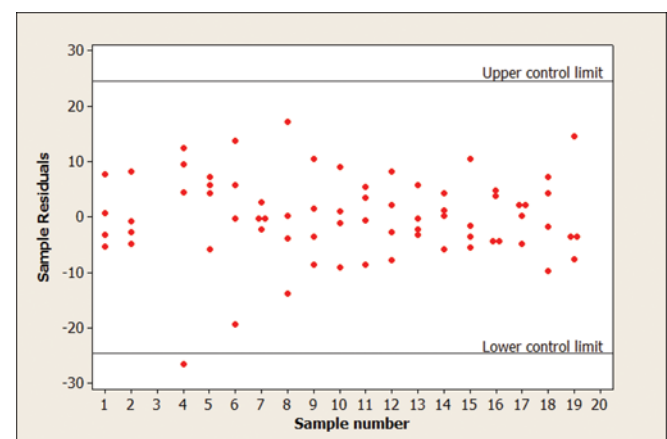

FIGURE 4 Phase I individuals control chart for the residuals. (Color figure available online.)

control are chart given by formulas [7] and [8]. Given these values, the Phase II upper and lower control limits are 19.27 and 0.79, respectively.

Step 8: Use the Phase II Chart for Online Monitoring

We use sample 20 to demonstrate Phase II monitoring of the standard deviation. First, we calculate $S/c_4(4)$ from this sample, giving a value of 3.07. This value falls between the upper and lower control limits (19.27 and 0.79), so no action is required and the process can continue (see Figure 3).

CONCLUDING REMARKS

In this article, we have outlined procedures to construct Phase I and Phase II standard deviation control charts. The initial estimate of σ for the Phase I standard deviation control chart is based on the 10% trimmed mean of the sample interquartile ranges $\overline{IQR}_{10}$. This estimator is robust against both diffuse and localized disturbances so that the Phase I limits are not affected by these disturbances. The Phase I data are then screened for localized and diffuse disturbances by means of both a Phase I sample standard deviation control chart and a Phase I individuals chart. At the end of Phase I, σ is estimated from the screened data using the mean of the sample standard deviations S, ensuring efficiency under normality. The procedures are described step by step and give constants that can be plugged into the formulas directly, thus making them easy for practitioners to implement and apply.

ABOUT THE AUTHORS

Hafiz Zafar Nazir obtained his M.Sc. in statistics in 2006 and his M.Phil. in statistics in 2008 from the Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan. Currently, he is working on his Ph.D. focusing on robust control charting techniques at the University of Amsterdam, The Netherlands. He is a lecturer in the Department of Statistics, University of Sargodha, Sargodha, Pakistan.

Marit Schoonhoven obtained her Ph.D. in statistics in 2011 at the University of Amsterdam, The Netherlands. She is senior consultant at the Institute for Business and Industrial Statistics of the University of Amsterdam, The Netherlands. Her current research

interests include control charting techniques and operational management methods.

Muhammad Riaz obtained his Ph.D. in statistics in 2008 at the University of Amsterdam, The Netherlands. He is assistant professor in the Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan, and Associate Professor in the Department of Mathematics and Statistics, King Fahad University of Petroleum and Minerals, Dhahran, Saudi Arabia. His current research interests include statistical process control, nonparametric techniques, and experimental design.

Ronald J. M. M. Does is Professor of Industrial Statistics at the University of Amsterdam; Managing Director of the Institute for Business and Industrial Statistics, which operates as an independent consultancy firm within the University of Amsterdam; and Director of the Institute of Executive Programmes at the Amsterdam Business School. He is a Fellow of the ASQ and Academician of the International Academy for Quality. His current research activities include the design of control charts for nonstandard situations, healthcare engineering and operational management methods.

REFERENCES

Hunter, J. S. (1986). The exponentially weighted moving average. *Journal of Quality Technology*, 18:203–210.

Mahmoud, M. A., Henderson, G. R., Epprecht, E. K., Woodall, W. H. (2010). Estimating the standard deviation in quality-control applications. *Journal of Quality Technology*, 42:348–357.

Maravelakis, P. E., Panaretos, J., Psarakis, S. (2002). Effect of estimation of the process parameters on the control limits of the univariate control charts for process dispersion. *Communications in Statistics - Simulation and Computation*, 31:443–461.

Ng, C. H., Case, K. E. (1989). Development and evaluation of control charts using exponentially weighted moving averages. *Journal of Quality Technology*, 21:242–250.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 42:243–254.

Page, E. S. (1963). Controlling the standard deviation by CUSUM and warning lines. *Technometrics*, 5:307–315.

Psarakis, S., Vyniou, A. K., Castagliola, P. (2013). Some recent developments on the effects of parameter estimation of control charts. *Quality and Reliability Engineering International*, 29(7):951–970.

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, 1:239–250.

Schoonhoven, M., Does, R. J. M. M. (2013). A robust $\overline{X}$ control chart. *Quality and Reliability Engineering International*. doi:10.1002/qre.1447

Schoonhoven, M., Riaz, M., Does, R. J. M. M. (2011). Design and analyis of control charts for standard deviation with estimated parameters. *Journal of Quality Technology*, 43:307–333.

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product.* Princeton, NJ: Van Nostrand.

Tukey, J. W. (1997). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Vining, G. (2009). Technical advice: Phase I and phase II control charts. *Quality Engineering*, 21:478–479.

Wadsworth, H. M., Stephens, K. S., Godfrey, A. B. (2001). *Modern Methods for Quality Control and Improvement*, 2nd ed. New York, NY: Wiley.

H. Nazir et al.