Quality Engineering, 26:246–252, 2014

ISSN: 0898-2112 print/1532-4222 online
DOI: 10.1080/08982112.2013.874562

Quality Quandaries: A Stepwise Approach for Setting Up a Robust Shewhart Location Control Chart

Hafiz Z. Nazir[1], Marit Schoonhoven[2], Muhammad Riaz[3], 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, University of Amsterdam, Amsterdam, The Netherlands
[3]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

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lqen.

INTRODUCTION

Shewhart control charts are extensively used in practice to monitor changes in process characteristics from the in-control state. There are many types of control chart applications, ranging from charts that monitor industrial process characteristics to charts that monitor the number of chronic diseases within health care. A control chart consists of a time series plot of the relevant process characteristic augmented by control chart limits (i.e., an upper control limit [UCL] and a lower control limit [LCL]). These control limits, sometimes supplemented with additional decision rules, provide easy checks on the stability of the process parameter. When the statistic falls outside the control limits, it is probable that the process has changed (i.e., is out of control). The employee or operator should then investigate the cause of the change and adjust the process to the in-control state.

A broad range of control charts is available: there are control charts for monitoring individual measurements of a process characteristic—used when the process characteristic provides no basis for rational subgrouping—as well as subgroup control charts suitable for monitoring processes from which samples are taken periodically. Within the category of subgroup control charts, there is a distinction between control charts for the location and control charts for the spread. Figure 1 provides an example of a control chart; in this case the Shewhart control chart for monitoring the location of the process characteristic ($\overline{X}$ control chart).

A further distinction is made based on the size of the shift to be detected: when the application requires that smaller shifts have to be detected (small means $\delta/(\sigma/\sqrt{n}) < 2$, with δ the shift, σ the in-control standard deviation, and n the sample size), control charts that take into account the history of observations are recommended. Examples include the cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) control charts (see Page [1954] and Roberts [1959] respectively). Here we focus on the Shewhart control chart.

Control charts were originally developed and tested assuming that the process characteristics are known. In real-world applications, however, the mean and standard deviation of the process characteristic are not known and have to be estimated before the control chart can be deployed. To derive such estimates, it is common practice to separate Phase I from Phase II. During

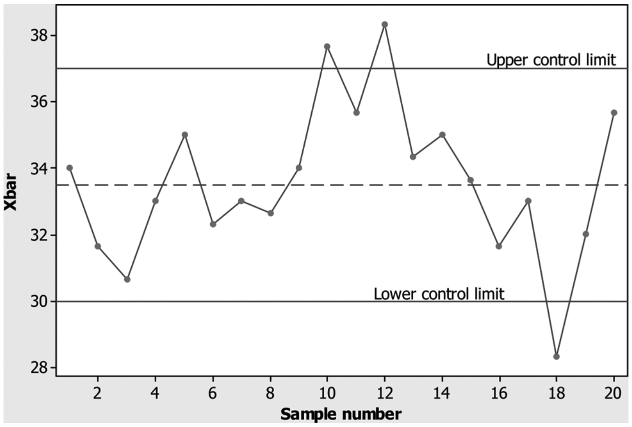

FIGURE 1 Shewhart location control chart.

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 online monitoring in Phase II. Jensen et al. (2006) conducted a literature survey of the effects of parameter estimation on control chart properties and identified the following issue for future research:

> *The effect of using robust or other alternative estimators has not been studied thoroughly. Most evaluations of performance have considered standard estimators based on the sample mean and the standard deviation and have used the same estimators for both Phase I and Phase II. However, in Phase I applications it seems more appropriate to use an estimator that will be robust to outliers, step changes and other data anomalies. Examples of robust estimation methods in Phase I control charts include Rocke (1989), Rocke (1992), Tatum (1997), Vargas (2003) and Davis and Adams (2005). The effect of using these robust estimators on Phase II performance is not clear, but it is likely to be inferior to the use of standard estimates because robust estimators are generally not as efficient* (Jensen et al. 2006, p. 360).

Schoonhoven, Nazir et al. (2011) and Schoonhoven, Riaz, and Does (2011) performed an extensive study of robust estimation in the context of Shewhart location and standard deviation control charts respectively. Nazir et al. (2014) summarized these findings in a practical procedure that enables practitioners to create and implement a robust standard deviation control chart without too much difficulty. The main purpose of the present column is to give practitioners a stepwise procedure on how to set up a robust location control chart.

The following two sections describe the Phase I and II stages for the Shewhart location control chart. Next, we demonstrate the procedure for a real-life example. The final section offers some concluding remarks.

PHASE I PROCEDURE

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

$$\widehat{UCL} = \hat{\mu} + C\hat{\sigma}/\sqrt{n}, \ \ \widehat{LCL} = \hat{\mu} - C\hat{\sigma}/\sqrt{n}. \quad [1]$$

with $\hat{\mu}$ and $\hat{\sigma}$ the estimates of the in-control mean μ and standard deviation σ, respectively, and C the constant chosen such that the desired in-control performance is obtained. Usually, C is chosen such that the false alarm probability is sufficiently small, namely, 0.0027. 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 $\widehat{UCL}$, $\widehat{LCL}$, and C to indicate the phase to which we refer.

Schoonhoven, Nazir et al. (2011) analyzed the performance of the Phase I location control charts. They showed that the type of location 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 μ is estimated are retained. However, an efficient estimator should be used to obtain the final estimates in order to ensure efficiency under normality. Below, we describe a practical step-by-step approach which 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 samples should reflect random, short-term rather than special-cause variation. To ensure this, items within a sample 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 samples. Furthermore, the sample 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. In practice the k samples of size n may contain outliers, sample shifts, or other contaminations. These can be filtered out in Phase I.

Step 2: Obtain a Robust Estimate of the Standard Deviation

The second step is to estimate the standard deviation using the stepwise procedure given by Nazir et al. (2014). The resulting estimate is denoted by $\hat{\sigma}$.

Step 3: Construct a Phase I Location Control Chart

We estimate the mean with a robust estimator, namely, the 10% trimmed mean of the sample trimeans (cf. Tukey [1977]), defined by

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

where $\lceil z \rceil$ denotes the ceiling function (i.e., the smallest integer not less than z) and $TM_{(v)}$ denotes the v-th ordered value of the sample trimeans. The trimean of sample i is defined by

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

where $Q_{i,2}$ is the median and $Q_{i,1} = X_{i,(a)}$ and $Q_{i,3} = X_{i,(b)}$ the first and third quartiles with $X_{i,(\nu)}$ the ν-th order statistic in sample i and $a = \lceil n/4 \rceil$, $b = n - a + 1$.

The Phase I location control chart limits are derived from

$$\widehat{UCL}_I = \overline{TM}_{10} + 3\hat{\sigma}/\sqrt{n}, \ \widehat{LCL}_I = \overline{TM}_{10} - 3\hat{\sigma}/\sqrt{n}.$$

Step 4: Screen for Sample Shifts

We plot the TM_is of the Phase I samples (cf. [3]) on the location control chart generated in step 3 (charting the TM_is instead of the sample means ensures that localized mean disturbances are identified and samples that contain only one single outlier are retained).

We exclude from the Phase I data set all samples whose TM_i falls outside the control limits ($i = 1, 2, \ldots, k$).

Step 5: Construct a Phase I individuals chart

The mean estimate is updated according to formula

$$\overline{TM}' = \frac{1}{k'} \sum_{i \in K'} TM_i \times I_{\widehat{LCL}_I \le TM_i \le \widehat{UCL}_I}(TM_i), \quad [4]$$

with $1_D(x)$ the indicator function, K' the set of samples that are not excluded in step 4 and k' the number of non excluded 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

$$\widehat{UCL}_{ind} = \overline{TM}' + 3\hat{\sigma}, \ \widehat{LCL}_{ind} = \overline{TM}' - 3\hat{\sigma}. \quad [5]$$

Step 6: Screen for Individual Outliers

We plot the individual observations of the samples remaining from step 4 on the individuals chart derived in step 5 and remove from the Phase I data set the observations that fall outside the limits.

Step 7: Obtain the final estimate of the mean

We now obtain a new estimate of the mean from the mean of the sample means

$$\overline{\overline{X}}'' = \frac{1}{k''} \sum_{i \in K''} \frac{1}{n_i'} \sum_{j \in N_i'} X_{ij} \times I_{\widehat{LCL}_{ind} \le X_{ij} \le \widehat{UCL}_{ind}}(X_{ij}), \quad [6]$$

with K'' the set of samples that are not excluded in steps 4 and 6, k'' the number of non excluded samples, N_i' the set of observations that are not excluded in sample i, and n_i' the number of non excluded observations in sample i.

PHASE II

In Phase I, we have established the reference data set. From this data set, the in-control mean is estimated. We then derive the control limits for Phase II; that is, the online monitoring stage. Recall that the formula for the Phase II control limits is given by [1], $\hat{\sigma}$ is calculated with the procedure described by Nazir et al. (2014), and $\hat{\mu}$ is calculated in step 7 of the Phase I procedure described in the previous section.

What remains is the determination of the factor C_{II} for the control limits.

Schoonhoven et al. (2009) presented a formula to calculate C_{II} for the $\overline{X}$ control chart based on the pooled mean of the sample standard deviations, $\tilde{S}$. They tested this formula for $\overline{X}$ charts derived from a broad range of standard deviation estimators and concluded that the formula is suitable when the variance of the estimator is close to the variance of $\tilde{S}$. Since the variance of the standard deviation estimator described by Nazir et al. (2014) is close to the variance of $\tilde{S}$ (see Schoonhoven and Does [2013]), the formula presented by Schoonhoven et al. (2009) 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 8: Construct a Phase II Location Chart

We obtain C_{II} for the Phase II control limits from the equation

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

where $c_4(m)$ is defined by

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

and $t_{k(n-1)}(1-\alpha/2)$ denotes the $(1-\alpha/2)$-th percentile of a t-distribution with $k(n-1)$ degrees of freedom and α the desired false alarm probability (usually 0.0027). Values for C_{II} and $c_4(m)$ with $m = k(n-1)+1$ for $n = 3, \ldots, 10$, $k = 20, 50$ and $\alpha = 0.0027$ are provided in Table 1.

The values for $\hat{\sigma}$, $\hat{\mu}$, and C_{II} are substituted into the formula for the $\overline{X}$ control limits given by [1].

Step 9: Use the Phase II Location Chart for Online Monitoring

Newly available data (Y_{ij} with $i = 1, 2, 3, \ldots$ and $j = 1, 2, \ldots, n$) are collected periodically and used to calculate $\overline{Y}_i = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$ (the mean in Phase II is used as plotting statistic because the mean is efficient under normality and sensitive to disturbances). When $\overline{Y}_i$ falls outside the control limits, the cause of this out-of-control signal should be investigated.

TABLE 1 Constants for Phase II Procedure

	$k = 20$		$k = 50$	
n	$c_4(m)$	C_{II}	$c_4(m)$	C_{II}
3	0.994	3.257	0.998	3.100
4	0.996	3.194	0.998	3.076
5	0.997	3.163	0.999	3.064
6	0.998	3.145	0.999	3.057
7	0.998	3.133	0.999	3.053
8	0.998	3.124	0.999	3.049
9	0.998	3.118	0.999	3.047
10	0.999	3.113	0.999	3.045

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

TABLE 2 Melt Index Measurements

Sample	Observations				TM
1	218	224	220	231	223.25
2	238	236	247	234	238.75
3	280	228	228	221	239.25
4	210	249	241	246	236.50
5	243	240	230	230	235.75
6	225	250	258	244	244.25
7	240	238	240	243	240.25
8	244	248	265	234	247.25
9	238	233	252	243	241.50
10	228	238	220	230	229.00
11	218	232	230	226	226.50
12	226	231	236	242	233.75
13	224	221	230	222	224.25
14	230	220	227	226	225.75
15	224	228	226	240	229.50
16	232	240	241	232	236.25
17	243	250	248	250	247.75
18	247	238	244	230	239.75
19	224	228	228	246	231.50
20	236	230	230	232	

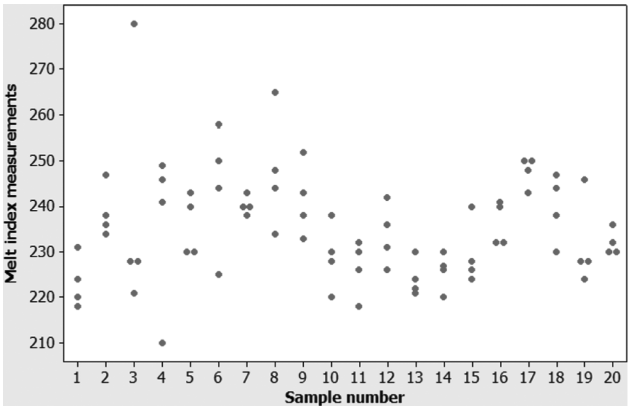

FIGURE 2 Individual value plot of 20 samples of the melt index.

of 20 samples of size 4 (Table 2). Figure 2 plots individual values for each sample.

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

Step 2: Obtain a Robust Estimate for the Standard Deviation

We use the procedure described by Nazir et al. (2014) to obtain an estimate of the standard deviation. Note that in this paper the same data set is used for illustration purposes. The final estimate is 7.32.

Step 3: Construct a Phase I Location Control Chart

First, the 10% trimmed mean of the sample trimeans is derived from formula [2]. This leads to a value of 235.22. The constant C_l for the location control chart is 3. The resulting upper and lower control limits are 246.20 and 224.24.

Step 4: Screen for Sample Shifts

The goal of this step is to filter out localized disturbances in the mean. To do this, we first determine the sample trimeans (see the TMs in Table 2). Samples 1, 8, and 17 fall outside the control limits of the Phase I location control chart and are therefore removed from the Phase I data set (see Figure 3). Note that this suggests the Phase I data are out of control: it seems that there is an underlying issue causing the observed pattern. The best step forward before applying the Shewhart location chart for monitoring is to identify the cause and prevent it from occurring again. In what follows, we continue the procedure to demonstrate how the remainder of the Phase I analysis can identify individual outliers.

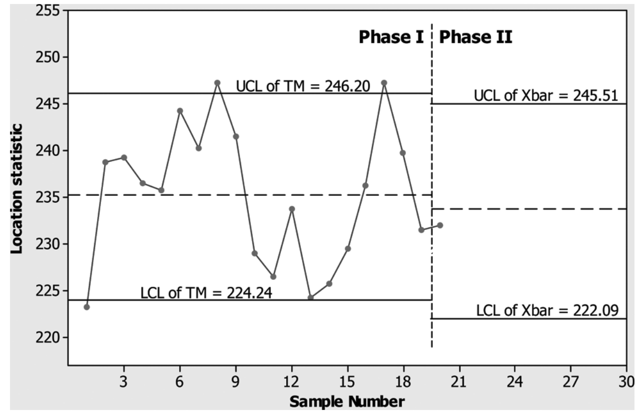

FIGURE 3 Phase I (and Phase II) location control chart for the melt index.

Step 5: Construct a Phase I Individuals Chart

We now update the location estimate according to [4], which means that we take the sum of the trimeans of all the samples excluding samples 1, 8, and 17. The mean of the non excluded trimeans is 234.53. The upper and lower control limits of the individuals chart are calculated from [5], giving 256.49 and 212.57.

Step 6: Screen for Individual Outliers

We plot the resulting individual melt index measurements (except observations from samples 1, 8, and 17) on the individuals chart. The first observations in samples 3 and 4 and the third observation in sample 6 fall outside the control chart limits and are removed from the Phase I data set (see Figure 4).

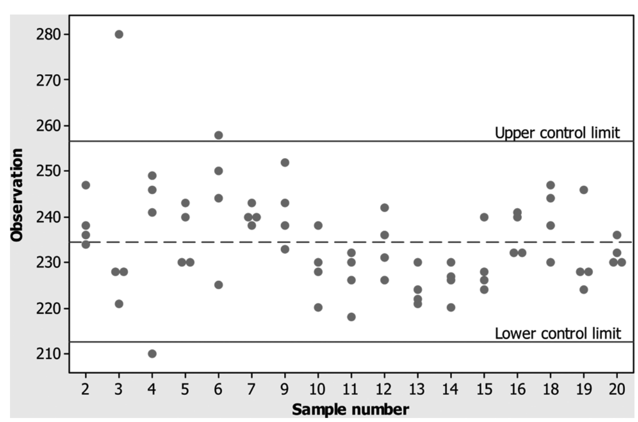

FIGURE 4 Phase I individuals control chart for the melt index.

Step 7: Obtain the Final Estimate of the Mean

In this step, we obtain the final estimate of μ from [6]. This means that we determine the mean of all observations, except those in samples 1, 8, and 17, the first observations in samples 3 and 4, and the third observation in sample 6, leading to a final value of 233.80.

Step 8: Construct a Phase II Location Chart

The values for the constant C_{II} used to calculate the limits of the Phase II location control chart are given by formula [7], leading to a value of 3.20. Given this value and the values of $\hat{\sigma}$ (7.32) and $\hat{\mu}$ (233.80), the Phase II upper and lower control limits are 245.51 and 222.09, respectively.

Step 9: Use the Phase II Chart for Online Monitoring

We use sample 20 to demonstrate Phase II monitoring of the location. First, we calculate $\overline{Y}$ from this sample, giving a value of 232.00. This value falls between the upper and lower control limits (245.51 and 222.09), so no action is required and the process can continue (see Figure 3).

CONCLUDING REMARKS

In this article, we have outlined procedures to construct a robust location control chart. A practical procedure is given to obtain a robust estimate for the mean, based on a Phase I analysis. In Phase I, the initial estimate of μ is based on the 10% trimmed mean of the sample trimeans. This estimator is robust against both localized and diffuse disturbances so that the limits of the Phase I control chart are not affected by potential disturbances. The Phase I data are then screened for localized and diffuse disturbances by means of Phase I sample location and individuals control charts. At the end of Phase I, μ is estimated from the screened data using the grand sample mean, thus ensuring efficiency. The stepwise procedure, together with the procedure for estimating the standard deviation σ presented by Nazir et al. (2014), delivers an easy-to-implement procedure for setting up a robust monitoring device.

ABOUT THE AUTHORS

Hafiz Z. 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 operations management methods.

Muhammad Riaz obtained his Ph.D in statistics in 2008 at the University of Amsterdam, The Netherlands. He is assistant professor at the Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan, and associate professor at 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 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, health care engineering and operations management methods.

REFERENCES

Davis, C. M., Adams, B. M. (2005). Robust monitoring of contaminated data. *Journal of Quality Technology*, 37:163–174.

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: A literature review. *Journal of Quality Technology*, 38:349–364.

Nazir, H. Z., Schoonhoven, M.; Riaz, M., and Does, R. J. M. M. (2014). How to set up a robust shewhart control chart for dispersion?. *Quality Engineering*, 26:130–136.

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

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

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

Rocke, D.M. (1992). $\bar{X}_Q$ and R_Q Charts: robust control charts. *The Statistician* 41, pp. 97-104.

Schoonhoven, M., Does, R. J. M. M. (2013). A Robust $\overline{X}$ control chart. *Quality and Reliability Engineering International* 29:951–970.

Schoonhoven, M., Nazir, H. Z., Riaz, M., Does, R. J. M. M. (2011a). Robust location estimators for the $\overline{X}$ Control Chart. *Journal of Quality Technology* 44:363–379.

Schoonhoven, M., Riaz, M., Does, R. J. M. M. (2009). Design schemes for the $\overline{X}$ control chart. *Quality and Reliability Engineering International* 25:581–594.

Schoonhoven, M.; Riaz, M.; and Does, R. J. M. M. (2011b). Design and Analysis of control charts for standard deviation with estimated parameters. *Journal of Quality Technology* 44, pp. 307–333.

Tatum, L. G. (1997). Robust estimation of the process standard deviation for control charts. *Technometrics* 39:127–141.

Tukey, J. W. (1977) *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Vargas, J.A. (2003). Robust estimation in multivariate control charts for individual observations. *Journal of Quality Technology* 35: 367–376.

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