Robut control charts (`rcc`) in the `rQCC` package

Chanseok Park* and Min Wang†

December 2022

Abstract

In this note, we provide a brief summary of variables control charts and a description of how they are constructed using the `rcc` function in the robust quality control chart (`rQCC`) R package. Using `rcc` function, one can construct the traditional Shewhart-type variables control charts. In addition, using various robust location and scale estimates provided by the `rQCC` package, one can easily obtain robust alternatives to the traditional charts.

1 Introduction

Control charts, also known as Shewhart control charts [1, 2, 3], have been widely used to monitor whether a manufacturing process is in a proper state of control or not. The traditional Shewhart-type control charts are made up of the upper control limit (UCL), the center line (CL) and the lower control limit (LCL) and they have the form of $\mathrm{CL} \pm g \cdot \mathrm{SE}$, where the American Standard is based on $g = 3$ with a target false alarm rate of 0.027% and the British Standard is based on $g = 3.09$ with a target false alarm rate of 0.020%. The UCL is given by $\mathrm{CL} + g \cdot \mathrm{SE}$ and the LCL is $\mathrm{CL} - g \cdot \mathrm{SE}$.

In what follows, we provide how to construct the traditional Shewhart-type control charts and robust alternatives to them using various robust location and scale estimates provided by the `rQCC` package. In this note, we assume that we have m samples and that each sample has the same sample size of n. Let X_{ij} be the ith sample (subgroup) from a stable manufacturing process, where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. We also assume that X_{ij} are independent and identically distributed as normal with mean μ and variance σ^2. The A, B and D notations here follow the definitions in ASTM (STP 15-C) [4] and ASTM (STP 15-D) [5].

*Applied Statistics Laboratory, Department of Industrial Engineering, Pusan National University, Busan 46241, Korea. His work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1A2C1091319).

†Department of Management Science and Statistics, The University of Texas at San Antonio, San Antonio, TX 78249, USA.

2 The $\bar{X}$ chart

In order to construct the CL $\pm g \cdot$ SE control limits, we consider the relation

$$\frac{\bar{X}_k - E(\bar{X}_k)}{\mathrm{SE}(\bar{X}_k)} = \pm g.$$

Since $E(\bar{X}_k) = \mu$ and $\mathrm{SE}(\bar{X}_k) = \sigma/\sqrt{n_k}$, we have

$$E(\bar{X}_k) \pm g \cdot \mathrm{SE}(\bar{X}_k) = \mu \pm \frac{g}{\sqrt{n_k}}\sigma.$$

Then the control limits for the $\bar{X}$ chart with the sample size n_k are given by

$$\begin{aligned} \mathrm{UCL} &= \mu + A(n_k)\sigma, \\ \mathrm{CL} &= \mu, \\ \mathrm{LCL} &= \mu - A(n_k)\sigma, \end{aligned}$$

where $A(n_k) = g/\sqrt{n_k}$. In practice, the values of the parameters, μ and σ, are not known. Thus, with the estimates $\hat{\mu}$ and $\hat{\sigma}$, we have

$$\begin{aligned} \mathrm{UCL} &= \hat{\mu} + \frac{g}{\sqrt{n_k}}\hat{\sigma}, \\ \mathrm{CL} &= \hat{\mu}, \\ \mathrm{LCL} &= \hat{\mu} - \frac{g}{\sqrt{n_k}}\hat{\sigma}. \end{aligned} \tag{1}$$

Thus, we need to estimate μ and σ by using each sample and then pooling these estimates. Using the ith sample above, the sample mean and variance are given by

$$\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \quad \text{and} \quad S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2,$$

where $i = 1, 2, \ldots, m$. Then we can estimate μ using all the samples as below:

$$\bar{\bar{X}} = \frac{\bar{X}_1 + \bar{X}_2 + \cdots + \bar{X}_m}{m} = \frac{1}{m}\sum_{i=1}^{m}\bar{X}_i.$$

Note that it is easily seen that $\bar{\bar{X}}$ is unbiased for μ. However, S_i is not unbiased for σ since $E(S_i) = c_4(n_i)\sigma$, where

$$c_4(n_i) = \sqrt{\frac{2}{n_i - 1}} \cdot \frac{\Gamma(n_i/2)}{\Gamma(n_i/2 - 1/2)}.$$

Thus, $S_i/c_4(n_i)$ is unbiased for σ. Then we can easily show that $\bar{S}/c_4(n_k)$ is unbiased for σ, where

$$\bar{S} = \frac{1}{m}\sum_{i=1}^{m} S_i.$$

Thus, by substituting $\hat{\mu} = \bar{\bar{X}}$ and $\hat{\sigma} = \bar{S}/c_4(n)$ into (1), we have the control limits

$$\begin{aligned}
\text{UCL} &= \bar{\bar{X}} + \frac{g}{\sqrt{n}}\frac{\bar{S}}{c_4(n)} = \bar{\bar{X}} + A_3(n)\bar{S}, \\
\text{CL} &= \bar{\bar{X}}, \\
\text{LCL} &= \bar{\bar{X}} - \frac{g}{\sqrt{n}}\frac{\bar{S}}{c_4(n)} = \bar{\bar{X}} - A_3(n)\bar{S},
\end{aligned}$$

where $A_3(n) = A(n)/c_4(n) = g/\{c_4(n)\sqrt{n}\}$.

It is also known that

$$E(R) = d_2(n)\sigma,$$

where R is the sample range from $X_i \sim N(\mu, \sigma^2)$ and

$$d_2(n) = 2\int_0^\infty \Big\{1 - \big[\Phi(z)\big]^n - \big[1 - \Phi(z)\big]^n\Big\} dz.$$

For more details on $d_2(n)$, one can refer to the vignette below.

```
> vignette("factors.cc", package="rQCC")
```

Then, with the ith sample, $R_i/d_2(n)$ is unbiased for σ, where

$$R_i = \max_{1\le j\le n}(X_{ij}) - \min_{1\le j\le n}(X_{ij}).$$

Then, with the m samples, $\bar{R}/d_2(n)$ is unbiased for σ, where

$$\bar{R} = \frac{1}{m}\sum_{i=1}^{m} R_i.$$

Substituting $\hat{\mu} = \bar{\bar{X}}$ and $\hat{\sigma} = \bar{R}/d_2(n)$ into (1), we have the control limits

$$\begin{aligned}
\text{UCL} &= \bar{\bar{X}} + \frac{g}{\sqrt{n}}\frac{\bar{R}}{d_2(n)} = \bar{\bar{X}} + A_2(n)\bar{R}, \\
\text{CL} &= \bar{\bar{X}}, \\
\text{LCL} &= \bar{\bar{X}} - \frac{g}{\sqrt{n}}\frac{\bar{R}}{d_2(n)} = \bar{\bar{X}} - A_2(n)\bar{R},
\end{aligned}$$

where $A_2(n) = A(n)/d_2(n) = g/\{d_2(n)\sqrt{n}\}$.

As alternatives to the above, we can use robust estimates of location and scale. For example, using the median, we can estimate μ

$$\hat{\mu} = \frac{M_1 + M_2 + \cdots + M_m}{m} = \frac{1}{m}\sum_{i=1}^{m} M_i,$$

where

$$M_i = \underset{1 \le j \le n}{\text{median}}(X_{ij}).$$

One can also consider estimating σ based on the conventional MAD (median absolute deviation) given by

$$\text{MAD} = \frac{\underset{1 \le i \le n}{\text{median}}|X_i - M|}{\Phi^{-1}(3/4)} \approx 1.4826 \cdot \underset{1 \le i \le n}{\text{median}}|X_i - M|,$$

where $X_i \sim N(\mu, \sigma^2)$ and $M = \text{median}(X_i)$. Here $\Phi^{-1}(3/4)$ is needed to make this estimator Fisher-consistent [6] for the standard deviation under the normal distribution. For more details, see the references [7, 8]. It should be noted that the above conventional MAD estimator is Fisher-consistent but not unbiased. The "*unbiased* MAD" (uMAD) with a finite sample is developed by Park, Kim and Wang [8] and implemented in the `rQCC` package (see `mad.unbiased` function).

Then, with the m samples, we have the robust unbiased estimate of σ as follows

$$\hat{\sigma} = \frac{\text{uMAD}_1 + \text{uMAD}_2 + \cdots + \text{uMAD}_m}{m} = \frac{1}{m}\sum_{i=1}^{m} \text{uMAD}_i,$$

where

$$\text{uMAD}_i = \underset{1 \le j \le n}{\text{uMAD}}(X_{ij}).$$

The `rcc` function constructs the control charts based on various *unbiased* estimates. For example, with the median and uMAD estimates, one can obtain the control limits using the following

```
> rcc(data, loc="median", scale="mad")
```

Another way of constructing the control limits is to use the Hodges-Lehmann [9] for location and Shamos [10] for scale which are respectively given by

$$\text{HL} = \text{median}\left(\frac{X_i + X_j}{2}\right)$$

and

$$\text{Shamos} = \frac{\underset{i<j}{\text{median}}\big(|X_i - X_j|\big)}{\sqrt{2}\,\Phi^{-1}(3/4)} \approx 1.048358 \cdot \underset{i<j}{\text{median}}\big(|X_i - X_j|\big),$$

where $\sqrt{2}\,\Phi^{-1}(3/4)$ is needed to make Shamos estimator Fisher-consistent for the standard deviation under the normal distribution [11]. For the Hodges-Lehmann estimate, the median is obtained by three ways: (i) the pairwise averages with $i < j$ (denoted by HL1), (ii) the pairwise averages with $i \le j$ (HL2), and (iii) all the pairwise averages (HL3). For more details, refer to [8]. It should be noted that the above Shamos is Fisher-consistent but not unbiased. The Hodges-Lehmann and "*unbiased* Shamos" are also developed by [8] and implemented in R (see `HL` and `shamos.unbiased`). For example, with the HL2 and unbiased Shamos estimates, one can obtain the control limits as below.

```
> rcc(data, loc="HL2", scale="shamos")
```

As shown above, by choosing the options for `loc` and `scale`, one can construct various control charts.

3 The S chart

In order to construct the $\mathrm{CL} \pm g \cdot \mathrm{SE}$ control limits, we can consider the relation

$$\frac{S_k - E(S_k)}{\mathrm{SE}(S_k)} = \pm g.$$

Since $E(S_k) = c_4(n)\sigma$ and $\mathrm{SE}(S_k) = \sqrt{1 - c_4(n)^2} \cdot \sigma$, we have

$$E(S_k) \pm g \cdot \mathrm{SE}(S_k) = \left\{ c_4(n) \pm g\sqrt{1 - c_4(n)^2} \right\} \sigma.$$

The control limits for the S chart are given by

$$\begin{aligned} \mathrm{UCL} &= B_6(n)\sigma, \\ \mathrm{CL} &= c_4(n)\sigma, \\ \mathrm{LCL} &= B_5(n)\sigma, \end{aligned}$$

where

$$\begin{aligned} B_5(n) &= \max\left\{ c_4(n) - g \cdot \sqrt{1 - c_4(n)^2},\ 0 \right\}, \\ B_6(n) &= c_4(n) + g \cdot \sqrt{1 - c_4(n)^2}. \end{aligned}$$

Since σ is unknown in practice, we need to choose an appropriate unbiased estimate for σ. One can consider $\hat{\sigma} = \bar{S}/c_4(n)$. Then we have

$$\begin{aligned} \mathrm{UCL} &= B_4(n)\bar{S}, \\ \mathrm{CL} &= \bar{S}, \\ \mathrm{LCL} &= B_3(n)\bar{S}, \end{aligned}$$

where $B_3(n) = B_5(n)/c_4(n)$ and $B_4(n) = B_6(n)/c_4(n)$.

To obtain the robustness property, one can consider a robust estimate of σ. For example, the unbiased MAD or unbiased Shamos estimates of σ can be used as seen before. The limits for the S chart are calculated using the `rcc` function with `type="S"` as below.

```
> rcc(data, scale="mad", type="S")
> rcc(data, scale="shamos", type="S")
```

4 The R chart

We consider the relation

$$\frac{R_k - E(R_k)}{\mathrm{SE}(R_k)} = \pm g.$$

Since $E(R_k) = d_2(n)\sigma$ and $\mathrm{Var}(R_k) = d_3(n)^2\sigma^2$, we have

$$E(R_k) \pm g \cdot \mathrm{SE}(R_k) = \big\{d_2(n) \pm g d_3(n)\big\}\sigma.$$

The control limits for the R chart are given by

$$\begin{aligned} \mathrm{UCL} &= D_2(n)\sigma, \\ \mathrm{CL} &= d_2(n)\sigma, \\ \mathrm{LCL} &= D_1(n)\sigma, \end{aligned}$$

where

$$\begin{aligned} D_1(n) &= \max\left\{d_2(n) - g \cdot d_3(n),\ 0\right\}, \\ D_2(n) &= d_2(n) + g \cdot d_3(n), \end{aligned}$$

Since σ is unknown in practice, we need to choose an appropriate unbiased estimate for σ. One can consider $\hat{\sigma} = \bar{R}/d_2(n)$. Then we have

$$\begin{aligned} \mathrm{UCL} &= D_4(n)\bar{R}, \\ \mathrm{CL} &= \bar{R}, \\ \mathrm{LCL} &= D_3(n)\bar{R}, \end{aligned}$$

where $D_3(n) = D_1(n)/d_2(n)$ and $D_4(n) = D_2(n)/d_2(n)$. These limits are easily calculated using the `rcc` function as below.

```
> rcc(data, scale="range", type="R")
```

As afore-mentioned, we can consider a robust estimate of σ. For example, the control limits with the unbiased Shamos are calculated as below.

```
> rcc(data, scale="shamos", type="R")
```

References

[1] W. A. Shewhart. Quality control charts. *Bell Systems Technical Journal*, pages 593–603, 1926.

[2] W. A. Shewhart. Quality control. *Bell Systems Technical Journal*, pages 722–735, 1927.

[3] W. A. Shewhart. *Economic Control of Quality of Manufactured Product*. Van Nostrand Reinhold, Princeton, NJ, 1931. Republished in 1981 by the American Society for Quality Control, Milwaukee, WI.

[4] ASTM Committee E-11. *ASTM Manual on Quality Control of Materials (STP 15-C)*. American Society for Testing and Materials, Philadelphia, PA, 1951.

[5] ASTM Committee E-11. *Manual on Presentation of Data and Control Chart Analysis (STP 15-D)*. American Society for Testing and Materials, Philadelphia, PA, 4th edition, 1976.

[6] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222:309–368, 1922.

[7] F. R. Hampel, E. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, New York, 1986.

[8] C. Park, H. Kim, and M. Wang. Investigation of finite-sample properties of robust location and scale estimators. *Communication in Statistics – Simulation and Computation*, 51(5):2619–2645, 2022.

[9] J. L. Hodges and E. L. Lehmann. Estimates of location based on rank tests. *Annals of Mathematical Statistics*, 34:598–611, 1963.

[10] M. I. Shamos. Geometry and statistics: Problems at the interface. In J. F. Traub, editor, *Algorithms and Complexity: New Directions and Recent Results*, pages 251–280. Academic Press, New York, 1976.

[11] C. Lèvy-Leduc, H. Boistard, E. Moulines, M. S. Taqqu, and V. A. Reisen. Large sample behaviour of some well-known robust estimators under long-range dependence. *Statistics*, 45:59–71, 2011.