

A note on the function for the attributes control charts (`acc`) in the `rQCC` package

Chanseok Park* and Min Wang†

January 2022

Abstract

In this note, we provide a brief summary of attribute control charts along with a description of how they are constructed using the `acc` function in the R package `rQCC`. In addition, various traditional attribute control charts such as p , c and u can be easily constructed using the `acc` function.

1 Introduction

In many practical applications, attribute control charts are widely used in the scenario where the quality characteristics cannot be conveniently measured and/or represented. Attribute (or count) data deal with quality characteristics, such as color, types of design, defective or non-defective, conforming or nonconforming, etc. In general, it is easier, cheaper, and faster to collect attribute data than quantitative data, whereas attribute data are less informative and precise. For example, the fraction or percentage of nonconforming units does not provide any information in which the measurements are centered and how dispersive they are.

Attribute control charts can also be constructed based on the idea of the traditional Shewhart control charts. These are the control charts for the fraction of nonconforming units (p chart), for nonconformities (c chart) and for nonconformities per unit (u chart). The traditional Shewhart-type control charts consist of the upper control limit (UCL), the center line (CL) and the lower control limit (LCL). They have the form of $CL \pm g \cdot SE$, 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 $CL + g \cdot SE$ and the LCL is $CL - g \cdot SE$.

In this note, we consider two kinds of attribute control charts for fraction nonconforming and for nonconformities (defects), which are also called the p and u charts,

*Applied Statistics Laboratory, Department of Industrial Engineering, Pusan National University, Busan 46241, Korea. His work was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2017R1A2B4004169).

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

respectively. The Shewhart-type attribute control charts are constructed under the tacit assumption that the normal approximation to the binomial distribution is acceptable. In what follows, we provide the construction of the p and u control charts based on the traditional Shewhart-type control charts and implement these charts in the R package `rQCC`.

2 The p and np charts

The p chart is usually used to monitor fraction nonconforming. Using this p chart, it is also available to construct the control chart (np chart) for the number of nonconforming instead of fraction nonconforming.

It is assumed that we have m samples and that the i th sample have either the equal number of observations n (balanced case) or different number of observations n_i (unbalanced case) for $i = 1, 2, \dots, m$. For notational simplicity, let X_i be the number of nonconforming units in the i th sample (subgroup) from a stable manufacturing process. We assume that X_i are independent and identically distributed according to the binomial distribution with the sample size n or n_i and the Bernoulli probability p .

2.1 The conventional p chart

Let $\hat{p}_k = X_k/n_k$. We have $E(\hat{p}_k) = p$ and $\text{Var}(\hat{p}_k) = p(1-p)/n_k$. To construct the control charts of the form $\text{CL} \pm g \cdot \text{SE}$, we consider the relation

$$\frac{\hat{p}_k - E(\hat{p}_k)}{\sqrt{\text{Var}(\hat{p}_k)}} = \pm g,$$

which showed that the control limits for the conventional Shewhart-type p chart with n_k are given by

$$p \pm g \sqrt{\frac{p(1-p)}{n_k}}. \quad (1)$$

It is worth noting that the value of the parameter p is unknown in practice. To tackle this issue, we combine all the subgroups and estimate p by

$$\bar{p} = \frac{\sum_{i=1}^m X_i}{\sum_{i=1}^m n_i}. \quad (2)$$

Replacing p in (1) with \bar{p} provides the control limits of the p chart

$$\bar{p} \pm g \sqrt{\frac{\bar{p}(1-\bar{p})}{n_k}}. \quad (3)$$

Since the control limits should be between zero and one, the control limits for the p chart are actually given by

$$\begin{aligned} \text{UCL} &= \min \left\{ \bar{p} + g \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_k}}, 1 \right\}, \\ \text{CL} &= \bar{p}, \\ \text{LCL} &= \max \left\{ \bar{p} - g \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_k}}, 0 \right\}. \end{aligned}$$

For the case of the known p , use the known value instead of \bar{p} in the above.

For example, one can obtain the control limits of the conventional p chart using the `acc` function in the R package `rQCC`.

```
> acc(x, n, type="p")
```

The conventional np chart is easily obtained by multiplying n and the control limits of the conventional p chart. It should be noted that this np chart is appropriate only when samples are balanced. Thus, we have the control limits for the np chart as follows.

$$\begin{aligned} \text{UCL} &= \min \left\{ n\bar{p} + g\sqrt{n\bar{p}(1 - \bar{p})}, n \right\}, \\ \text{CL} &= n\bar{p}, \\ \text{LCL} &= \max \left\{ n\bar{p} - g\sqrt{n\bar{p}(1 - \bar{p})}, 0 \right\}. \end{aligned}$$

One can obtain the control limits of the conventional np chart by setting `type="np"` in the `acc` function.

2.2 The p chart based on the Wilson confidence interval

In the statistics literature, Equation (3) is the most commonly-used interval for the Bernoulli parameter p and is also related to the Wald confidence interval for p ; see [1]. However, it is well-known that the Wald confidence limits are unsatisfactory for small n . See the references in [1, 2, 3]. Wilson [4] obtained an improved confidence interval by using the idea of directly inverting a test statistic.

Park [5] introduced an improvement of the p chart based on the approach of Wilson [4]. When p is known, the control limits with the sample size n_k are given by

$$\frac{p + g^2/(2N)}{1 + g^2/N} \pm \frac{g}{1 + g^2/N} \sqrt{\frac{p(1 - p)}{n_k} + \frac{g^2}{4n_k N}},$$

where $N = \sum_{i=1}^m n_k$. When p is unknown, the control limits are then given by

$$\frac{\bar{p} + g^2/(2N)}{1 + g^2/N} \pm \frac{g}{1 + g^2/N} \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_k} + \frac{g^2}{4n_k N}},$$

where \bar{p} is given in (2). Since the control limits should be between zero and one, the control limits are actually given by

$$\begin{aligned} \text{UCL} &= \min \left\{ \frac{\bar{p} + g^2/(2N)}{1 + g^2/N} + \frac{g}{1 + g^2/N} \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_k} + \frac{g^2}{4n_k N}}, 1 \right\}, \\ \text{CL} &= \frac{\bar{p} + g^2/(2N)}{1 + g^2/N}, \\ \text{LCL} &= \max \left\{ \frac{\bar{p} + g^2/(2N)}{1 + g^2/N} - \frac{g}{1 + g^2/N} \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_k} + \frac{g^2}{4n_k N}}, 0 \right\}. \end{aligned}$$

As afore-mentioned, for the case of the known p , we can use the known value instead of \bar{p} in the above.

For example, one can obtain the control limits of the p chart using the `acc` function in the R package `rQCC`.

```
> acc(x, n, type="p", pEstimator="Wilson")
```

The np chart based on the Wilson confidence interval is also obtained by multiplying n and the control limits of the p chart based on the Wilson confidence interval. This chart is also appropriate only when samples are balanced. Thus, we have the following control limits.

$$\begin{aligned} \text{UCL} &= \min \left\{ \frac{n\bar{p} + g^2/(2m)}{1 + g^2/(nm)} + \frac{g}{1 + g^2/(nm)} \sqrt{n\bar{p}(1 - \bar{p}) + g^2/(4m)}, n \right\}, \\ \text{CL} &= \frac{n\bar{p} + g^2/(2m)}{1 + g^2/(nm)}, \\ \text{LCL} &= \max \left\{ \frac{n\bar{p} + g^2/(2m)}{1 + g^2/(nm)} - \frac{g}{1 + g^2/(nm)} \sqrt{n\bar{p}(1 - \bar{p}) + g^2/(4m)}, 0 \right\}. \end{aligned}$$

One can obtain the control limits of the np chart based on the Wilson confidence interval with `type="np"` as follows.

```
> acc(x, n, type="np", pEstimator="Wilson")
```

3 The u and c charts

The u chart is the control chart for monitoring nonconformities (defects) that occur in each unit. Let X_i be the number of defects with the n_i inspection units. We assume that the number of defects in a *single* unit is distributed as the Poisson distribution with the mean λ . Then X_i are independent and identically distributed according to the Poisson distribution with the mean λn_i , such that $E(X_i) = n_i \lambda$ and $\text{Var}(X_i) = n_i \lambda$.

Let $U_k = X_k/n_k$. We have $E(U_k) = \lambda$ and $\text{Var}(U_k) = \lambda/n_k$. To construct the $\text{CL} \pm g \cdot \text{SE}$ control limits, we consider the relation

$$\frac{U_k - E(U_k)}{\sqrt{\text{Var}(U_k)}} = \pm g.$$

The control limits for the Shewhart-type u chart with the n_k inspection units are given by

$$\lambda \pm g \sqrt{\frac{\lambda}{n_k}}.$$

It is worth noting that if $n_k = 1$, this chart can be regarded as the c chart.

When λ is unknown, it can be estimated by

$$\bar{\lambda} = \frac{\sum_{i=1}^m X_i}{\sum_{i=1}^m n_i}.$$

Since the control limits are nonnegative, we consider the following control limits

$$\begin{aligned} \text{UCL} &= \bar{\lambda} + g \sqrt{\frac{\bar{\lambda}}{n_k}}, \\ \text{CL} &= \bar{\lambda}, \\ \text{LCL} &= \max \left\{ \bar{\lambda} - g \sqrt{\frac{\bar{\lambda}}{n_k}}, 0 \right\}. \end{aligned}$$

Again, for the case of the known λ , we just use the value of λ instead of $\bar{\lambda}$ in the above.

For example, one can obtain the control limits of the u chart using the `acc` function as follows.

```
> acc(x, n, type="u")
```

The control limits of the c chart is also easily obtained as follows.

```
> acc(x, type="c")
```

4 The g and h charts

The geometric distribution can also be used for counting event data. It is widely used for monitoring the number of conforming cases between the two consecutive appearances of nonconformities such as defects and infections.

Based on this, Kaminsky et al. [6] proposed g and h charts. The probability mass function for the geometric distribution with location a is given by

$$f(y) = P(Y = y) = p(1 - p)^{y-a}, \quad (4)$$

where $y = a, a + 1, \dots$ and a is the known minimum possible number of events (usually $a = 0, 1$). The mean and variance of Y are given by

$$\mu = E(Y) = \frac{1-p}{p} + a \quad \text{and} \quad \sigma^2 = \text{Var}(Y) = \frac{1-p}{p^2}.$$

We construct the control chart for the total number of events (g chart) and the average number of events (the h chart) per subgroup with $\text{CL} \pm g \cdot \text{SE}$ control limits.

The g chart (total number of events) with the sample size n_k has the following control limits which are actually the function of μ and σ^2 . For more details, see Park and Wang [7] and Kaminsky et al. [6].

$$\begin{aligned} \text{UCL} &= n_k \mu + g \sqrt{n_k \sigma^2} = n_k \left(\frac{1-p}{p} + a \right) + g \sqrt{\frac{n_k(1-p)}{p^2}}, \\ \text{CL} &= n_k \mu = n_k \left(\frac{1-p}{p} + a \right), \\ \text{LCL} &= n_k \mu - g \sqrt{n_k \sigma^2} = n_k \left(\frac{1-p}{p} + a \right) - g \sqrt{\frac{n_k(1-p)}{p^2}}. \end{aligned} \tag{5}$$

The smallest possible value of the total number of events is $n_k a$. Thus, the LCL can not be smaller than $n_k a$. If $\text{LCL} < n_k a$ in the above limit, we set up $\text{LCL} = n_k a$.

The h chart (average number of events) with the sample size n_k has the following control limits.

$$\begin{aligned} \text{UCL} &= \mu + g \sqrt{\frac{\sigma^2}{n_k}} = \frac{1-p}{p} + a + g \sqrt{\frac{1-p}{n_k p^2}}, \\ \text{CL} &= \mu = \frac{1-p}{p} + a, \\ \text{LCL} &= \mu - g \sqrt{\frac{\sigma^2}{n_k}} = \frac{1-p}{p} + a - g \sqrt{\frac{1-p}{n_k p^2}}. \end{aligned} \tag{6}$$

The smallest possible value of the average number of events is a . Thus, we set up $\text{LCL} = a$ if $\text{LCL} < a$ in the above limit.

The value of the process parameter p is unknown in general so that we estimate p . We assume that we obtain m samples and denote the size of each sample by n_i for $i = 1, 2, \dots, m$. Suppose that X_{ij} be the number of independent Bernoulli trials (cases) until the first nonconforming case in the i th sample for $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n_i$. Then X_{ij} 's are independent and identically-distributed geometric random variables with location shift a and p . We let $\bar{X} = \sum_{i=1}^m \sum_{j=1}^{n_i} X_{ij} / N$ with $N = \sum_{i=1}^m n_i$. We can estimate p using the maximum likelihood (ML) and minimum variance unbiased (MVU) estimators [7] which are given by

$$\hat{p}_{\text{ML}} = \frac{1}{\bar{X} - a + 1} \quad \text{and} \quad \hat{p}_{\text{MVU}} = \frac{1 - 1/N}{\bar{X} - a + 1 - 1/N},$$

respectively. It should be noted that Benneyan [8] proposed $\hat{p}_B = (1 - 1/N)/(\bar{\bar{X}} - a + 1)$ as the MVU estimator, but Park and Wang [7] showed that \hat{p}_B is not a correct MVU estimator.

The ML estimator has the invariance property. Thus, by substituting \hat{p}_{ML} into (5) and (6), we can construct the g and h charts based on the ML method as follows.

- g chart

$$\begin{aligned} \text{UCL} &= n_k \bar{\bar{X}} + g \sqrt{n_k (\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1)}, \\ \text{CL} &= n_k \bar{\bar{X}}, \\ \text{LCL} &= n_k \bar{\bar{X}} - g \sqrt{n_k (\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1)}. \end{aligned}$$

- h chart:

$$\begin{aligned} \text{UCL} &= \bar{\bar{X}} + g \sqrt{\frac{(\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1)}{n_k}}, \\ \text{CL} &= \bar{\bar{X}}, \\ \text{LCL} &= \bar{\bar{X}} - g \sqrt{\frac{(\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1)}{n_k}}. \end{aligned}$$

When we construct the charts based on the MVU estimator, \hat{p}_{MVU} , a care should be taken because the MVU estimator has no invariance property like the ML estimator. Park and Wang [7] showed that the MVU estimators of μ and σ^2 are given by

$$\hat{\mu}_{mvu} = \bar{\bar{X}} \quad \text{and} \quad \hat{\sigma}_{mvu}^2 = \frac{N}{N+1} (\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1),$$

and they construct the g and h charts based on $\hat{\mu}_{mvu}$ and $\hat{\sigma}_{mvu}^2$ as follows.

- g chart:

$$\begin{aligned} \text{UCL} &= n_k \bar{\bar{X}} + g \sqrt{\frac{n_k N}{N+1} (\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1)}, \\ \text{CL} &= n_k \bar{\bar{X}}, \\ \text{LCL} &= n_k \bar{\bar{X}} - g \sqrt{\frac{n_k N}{N+1} (\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1)}. \end{aligned}$$

- h chart:

$$\begin{aligned} \text{UCL} &= \bar{\bar{X}} + g \sqrt{\frac{N}{N+1} \frac{(\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1)}{n_k}}, \\ \text{CL} &= \bar{\bar{X}}, \\ \text{LCL} &= \bar{\bar{X}} - g \sqrt{\frac{N}{N+1} \frac{(\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1)}{n_k}}. \end{aligned}$$

The control limits of the g or h chart is also easily obtained as follows.

```
> acc(x, type="g")  
> acc(x, type="h")
```

References

- [1] Alan Agresti and Brent A. Coull. Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*, 52:119–126, 1998.
- [2] C. R. Blyth and H. A. Still. Binomial confidence intervals. *Journal of the American Statistical Association*, 78:108–116, 1983.
- [3] B. K. Ghosh. A comparison of some approximate confidence intervals for the binomial parameter. *Journal of the American Statistical Association*, 74:894–900, 1979.
- [4] E. B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22:209–212, 1927.
- [5] C. Park. An improved p chart based on the Wilson interval. *Journal of Statistics and Management Systems*, 16:201–221, 2013.
- [6] F. C. Kaminsky, J. C. Benneyan, and R. D. Davis. Statistical control charts based on a geometric distribution. *Journal of Quality Technology*, 24:63–69, 1992.
- [7] C. Park and M. Wang. A note on the g and h control charts. <https://arxiv.org/abs/2101.07575>, 2021. ArXiv e-prints.
- [8] J. C. Benneyan. Performance of number-between g -type statistical control charts for monitoring adverse events. *Health Care Management Science*, 4:319–336, 2001.