Expert Systems with Applications 38 (2011) 667–672

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Using cause selecting control charts to monitor dependent process stages with attributes data

Su-Fen Yang *, Jin-Tyan Yeh

Department of Statistics, National Chengchi University, Taipei 116, Taiwan, ROC

A R T I C L E I N F O

Keywords:
Control chart
Dependent process steps
Attribute data
Bivariate binomial control region

A B S T R A C T

In this study, we propose cause selecting control charts to monitor two dependent process stages with attributes data. The control limits on the bivariate binomial control region can be obtained. The detection ability of the cause selecting control charts is compared to those of Shewhart attributes control charts and the bivariate binomial control region by different correlation. Numerical example and simulation study show that the cause selecting control charts perform better than Shewhart attributes control charts and the bivariate binomial control region.

1. Introduction

Most of the products are produced by several different process steps these days. If the process steps are independent then using a Shewhart control chart to monitor each individual step is meaningful. However if many process steps were dependent then the Shewhart charts are difficult to interpret the process state correctly. The multivariate control charts have become a popular topic in quality control. Lowery and Montgomery (1995) reviewed Hotelling multivariate control chart, multivariate cumulative sum (MCUSUM) control procedure, and multivariate exponentially weighted moving average (MEWMA) control chart. But, little research has been done on multi-attribute processes.

Let variable X_j be the number of defects or nonconformities with respect to quality characteristic j, $j = 1,2,\ldots,n$, and $p = (p_1,p_2,\ldots,p_n)$ be the vector of fraction nonconformities. However X_j's are correlated. Hence, control chart for multivariate-attribute processes should be used. Patel (1973) proposed a Hotelling-type χ^2 chart to monitor the time-dependent observations from multinomial or multivariate Poisson populations. Because of its complexity, the scheme was not widely used in practice. Lu, Xie, Goh, and Lai (1998) established a multivariate *np* control chart to deal with the multivariate-attribute processes. The weighted sum of nonconforming counts of each quality characteristic was defined as X statistic. Control limits of the Shewhart-type charts were derived using X. The drawbacks of this work were the normality assumption and the lack of discussion on the average run length. Niaki (2006) employed the concept of simultaneous confidence intervals to derive control limits for several correlated quality characteristics in a multi-attribute data. He took advantage of the bootstrap method in designing the control charts, compare its performance to other method. Niaki and Abbasi (2007) first proposed a new transformation technique to reduce the amount of skewness of distribution of the attributes data and then use a Hotelling T^2 control chart on the transformed data. Mukhopadhyay (2008) expanded the concept of 'Mahalanobis Distance' in a multinomial distribution and thereby proposed a multivariate-attribute control chart. A drawback of this work is that when there was an out-of-control signal, it is often difficult to determine which component of the process was out of control.

In this study, we propose cause selecting control charts to monitor two dependent process stages with attributes data. The cause selecting control chart is similar to the regression control chart by Mandel (1969) in that a control chart is constructed for a variable only after the observations have been adjusted for the effect of some other random variables. Cause selecting control chart was first introduced by Zhang (1984). In the past ten years, many works have been done on this. Wade and Woodall (1993) gave excellent review on the cause selecting control chart, and discussed its relationship to the Hotelling T^2 control chart. In their opinion, the cause selecting chart outperformed Hotelling T^2 chart.

Let X be the quality variable in the first stage, and Y the quality variable in the second stage. Since the two stages are dependent, Y is influenced by X. To monitor the variation of X, attributes control chart for X should be constructed. However, we cannot construct attributes control chart for Y to monitor the second stage since the out-of-control attributes control chart for Y may be influenced by the out-of-control first stage. The correct approach is to adjust the effect of X on Y, and the simple cause selecting control chart is thus constructed to control the specific quality on the second stage.

* Corresponding author. Tel.: +886 29387459; fax: +886 29398024.
E-mail address: yang@nccu.edu.tw (S.-F. Yang).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2010.07.018

Let X be the number of nonconforming units of a specific part of product, Y the number of nonconforming units of the product and ρ is the correlation coefficient. A random sample of n units of a product is selected. The interested quality characteristics (X,Y) is assumed to follow a bivariate binomial distribution $BB(n,p_x,p_y,\rho)$, *where* p_x, p_y are the fraction nonconforming of X and Y, respectively. The control limits of the bivariate binomial control region (BB control region) are found using the exact distribution. When any sample point lies outside the control region, we deem the entire process is out-of-control. Otherwise it is in-control. The drawback of the bivariate binomial control region is that it might indicate the entire process is out-of-control, but does not show which process part is out-of-control.

In Section 2, process description for attribute is illustrated. In Section 3, a bivariate binomial control region is constructed for different p_x, p_y, n and ρ. The effect of p_x and p_y on control region is explained. In Section 4, numerical example for Shewhart np_x–np_y chart, cause selecting np_x–e chart, and BB control region is presented and the detection ability for different method is compared. In Section 5, the ARL computation is carried out. In Section 6, concluding remarks are provided.

The detection ability of the cause selecting control charts is compared to those of Shewart attributes control charts and the bivariate binomial control region by considering three levels of correlation between X and Y – low, medium and high. Numerical example and simulation data showed that the cause selecting control charts performed better than Shewhart attributes control charts and the bivariate binomial control region. The cause selecting control charts are thus recommended to monitor the dependent process stages with attributes data.

2. Process description for attributes data

X is the input quality variable and Y is the outgoing quality variable. Here we assume that the paired data can only be collected at the end of the second stage. A random sample of n units of a product is taken. Biswas and Huang (2002) gave the joint p.d.f of X and Y as follows:

$$p(X=x, Y=y) = \binom{n}{x} p_X^x (1-p_X)^{n-x} f(y|x),$$

where

$$f(y|x) = (1+w)^{-n} \sum_{(i,j)\in S} \binom{x}{i}\binom{n-x}{j} \{p_Y + w(p_Y - p_X) + w\}^i$$

$$*\{1-p_Y - w(p_Y-p_X)\}^{x-i}\{p_Y + w(p_Y-p_X)+w\}^j$$

$$*\{1-p_Y - w(p_Y-p_X)+w\}^{n-x-j}$$

with

$$S = \{(i,j) : i+j = y,\ i = 0,1,\ldots,x,\ j = 0,1,\ldots,n-x\}$$

$$w = \frac{\rho k}{1-\rho k},\ \ k = \sqrt{\frac{p_Y(1-p_Y)}{p_X(1-p_X)}}$$

Fig. 1 shows the two-stage process.

The distribution of the bivariate quality characteristic (X,Y) is not symmetric. To solve the asymmetric problem, Freeman and Tukey (1950) proposed a better arcsin approach of normalized transformation.

Since the two stages are dependent, i.e. the second stage is influenced by the first stage, the relation of Y and X may be expressed by the arcsin transformed model, $(Y^*|X^*) = f(x^*) + \varepsilon$, where Y should be no less than X, $Y^* = Y/n$, $X^* = X/n$, and ε is a random error, $\varepsilon \sim N(0,\sigma^2)$. In order to exclude the effect from the first stage while monitoring the second stage, we let

$$e = \arcsin(Y^*|X^*) - \hat{\arcsin}(Y^*|X^*) \sim N(0,\sigma^2)$$

where $\hat{\arcsin}(Y^*|X^*)$ is fitted value of arcsin $(Y^*|X^*)$. To control the two dependent process stages effectively, the Shewhart np_x chart and e chart (cause selecting control charts) are used. When both stages are in-control

$$X \sim bin(n,p_X), \quad e \sim N(0,\sigma^2)$$

We constructed the np_x chart and e chart as follows:

$$\begin{aligned} UCL_X &= np_X + k_X\sqrt{np_X(1-p_X)} \\ CL_X &= np_X \\ LCL_X &= np_X - k_X\sqrt{np_X(1-p_X)} \end{aligned} \tag{1}$$

$$\begin{aligned} UCL_e &= k_e\sigma \\ CL_e &= 0 \\ LCL_e &= -k_e\sigma \end{aligned} \tag{2}$$

This is to say, we use np_x chart to monitor the first stage, and e chart to monitor the second stage. When p_x and σ are unknown, they are estimated from the sample.

3. Determination of the bivariate binomial control region

(X,Y) follows a bivariate binomial distribution $BB(n,p_X,p_Y,\rho)$. The control limits can be calculated by using the exact probability distribution. That is

$$p((X,Y) \geqslant UCL_B) \leqslant \alpha, \quad \text{where } (X,Y) \sim BB(n,p_X,p_Y,\rho)$$

$$\text{this implies} \sum_{(x,y)=(0,0)}^{UCL_B} \binom{n}{x} p_X^x(1-p_X)^{n-x} f(y|x) > 1-\alpha$$

where

$$f(y|x) = (1+w)^{-n} \sum_{(i,j)\in S} \binom{x}{i}\binom{n-x}{j} \{p_Y + w(p_Y - p_X) + w\}^i$$

$$\bullet\{1-p_Y - w(p_Y-p_X)\}^{x-i}\{p_Y + w(p_Y-p_X)+w\}^j$$

$$\bullet\{1-p_Y - w(p_Y-p_Y)+w\}^{n-x-j}$$

with

$$S = \{(i,j) : i+j = y,\ i = 0,1,\ldots,x,\ j = 0,1,\ldots,n-x\}$$

$$w = \frac{\rho k}{1-\rho k},\ \ k = \sqrt{\frac{p_Y(1-p_Y)}{p_X(1-p_X)}}$$

When n, p_X, p_Y, ρ, α are given, the bivariate binomial control region,the triangular and upper control limit (UCL_B) can be constructed as shown in Fig. 2

When the BB control region is determined, we can plot statistics (X,Y) on the BB control region and monitor the process.

3.1. Numerical example

Table 1 lists critical points on the Upper Control Limit (UCL_B) for three different combinations of p_x, p_y, ρ = 0.1–0.9, n = 100 and α = 0.0027.

Fig. 3 shows the BB control region for three different combinations of p_x, p_y, ρ, n = 50 and α = 0.0027.

Table 2 shows critical points on the Upper Control Limit (UCL_B) for three different combinations of p_x, p_y, ρ = 0.1–0.9, n = 100 and α = 0.0027.

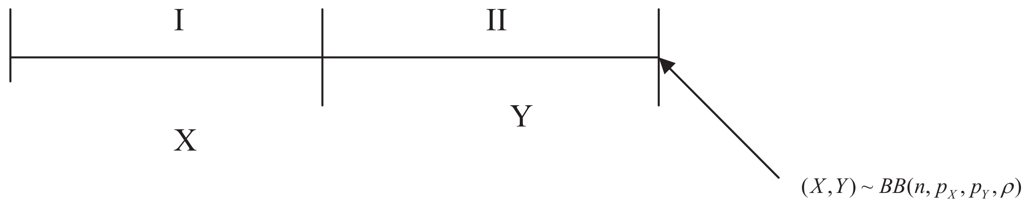

Fig. 1. Two-stage process for attribute data.

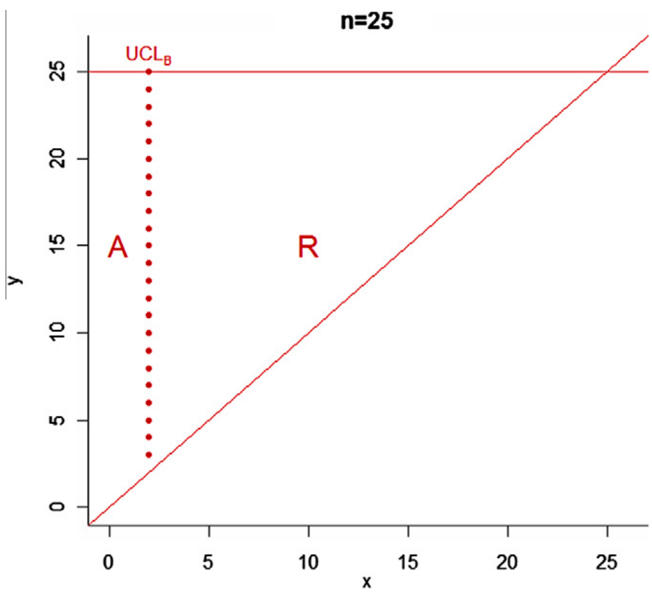

Fig. 2. BB control region for $n = 25$, $p_x = 0.01$, $p_y = 0.01$, $\rho = 0.1$, $\alpha = 0.0027$, where A: acceptance region, R: rejection region.

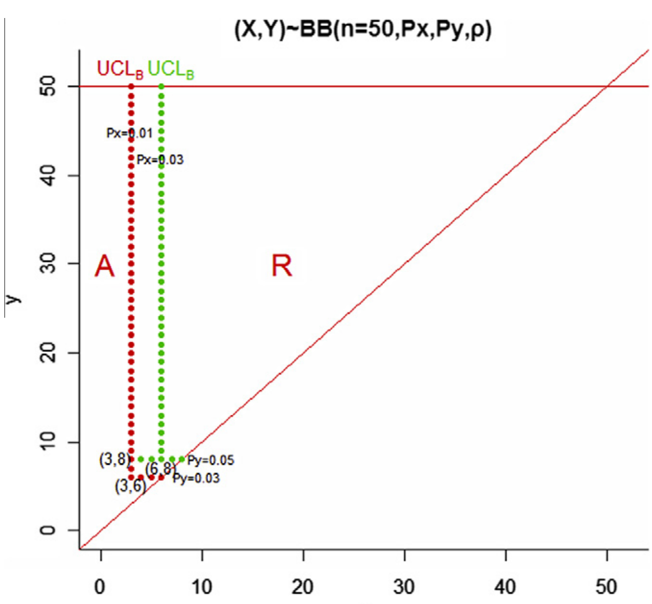

Fig. 3. BB control regions for ($n = 50$, $p_x = 0.01$, 0.03, $p_y = 0.03$, 0.05, $\rho = 0.5$, $\alpha = 0.0027$) and A: acceptance region, R: rejection region.

Fig. 4 shows the BB control region for three different combinations of p_x, p_y, $n = 50$, $\rho = 0.5$ and $\alpha = 0.0027$

3.2. The effect of p_x and p_y on BB control region

From above, we found that the area of accepted region would increase toward right-side and the area of the rejection region would shrink gradually while p_x is increasing at $\rho = 0.1$–0.9. When p_y is increased, the area of rejection region would decrease upward and the area of the acceptance region would expand gradually.

3.3. Relationship to the cause selecting chart

The control limit of BB control region is obtained by the 1-α quantile of the bivariate binomial distribution with a false rate of α. Figs. 2–4 show the typical control region. When a sample point falls in R it gives an out-of-control signal of the process, however it does not tell us which step of the process is out-of-control. The advantage of the cause selecting chart over the BB control region is that it is easier to identify which stage is out-of-control.

4. Numerical example

The paint defect data in Table 3 is taken from Mukhopadhyay (2008) but the sample size is changed to 100. This example deals with the fraction defective of two types of paint defects of a ceiling fan cover. Let X be the number of patty defect, and Y be the number of poor covering. The correlation coefficient of X and Y is 0.553, and Y is influenced by X.

Scatter plot indicates that the relationship of $\hat{\arcsin}(Y^*|X^*)$ and X^* is linear. Using the least square error method to find their relationship, the regression model is

$$\hat{\arcsin}(Y^*|X^*) = 0.0643 + 0.874X^*$$

The residual is

$$e = \arcsin(Y^*|X^*) - (0.0643 + 0.874X^*)$$

The in-control distribution of e is $N(0,(0.002)^2)$

To compare the performance among (np_x and e charts), BB control region and (Shewhart np_x and np_y charts), let the identical false alarm rate be 0.0054 and we plot the data on Figs. 5–7.

Fig. 5 shows an out-of-control point (the sample 20) on np_x chart and an out-of-control point (the sample 10) on e chart.

Fig. 6 shows that all (X, Y)'s are inside the BB control region.

Fig. 7 shows an out-of-control point (the sample 20) on np_x chart, but none on np_y chart.

The detection results of the 3-typed control charts show that using np_x–e chart outperforms others.

Table 1
Critical points on the UCL_B for $n = 50$, p_x, p_y, $\alpha = 0.0027$, $\rho = 0.1$–0.9.

(p_x, p_y)	**(0.01,0.03)**	**(0.01,0.05)**	**(0.03,0.05)**
Critical points	{(3,6),(3,7)···(3,50),(4,6),(5,6),(6,6)}	{(3,8),(3,9)···(3,50),(4,8),(5,8)···(8,8)}	{(6,8),(6,9)···(6,50),(7,8),(8,8)}

Table 2
Critical points on the UCL_B for $n = 100$, p_x, p_y, $\alpha = 0.0027$, $\rho = 0.1$–0.9.

(p_x, p_y)	**(0.01,0.03)**	**(0.01,0.05)**	**(0.03,0.05)**
Critical points	{(5,9),(5,10)···(5,100),(6,9),(7,9),(8,9),(9,9)}	{(5,12),(5,13)···(5,100),(6,12),(7,12)···(12,12)}	{(9,12),(9,13)···(9,100),(10,12),(11,12)···(12,12)}

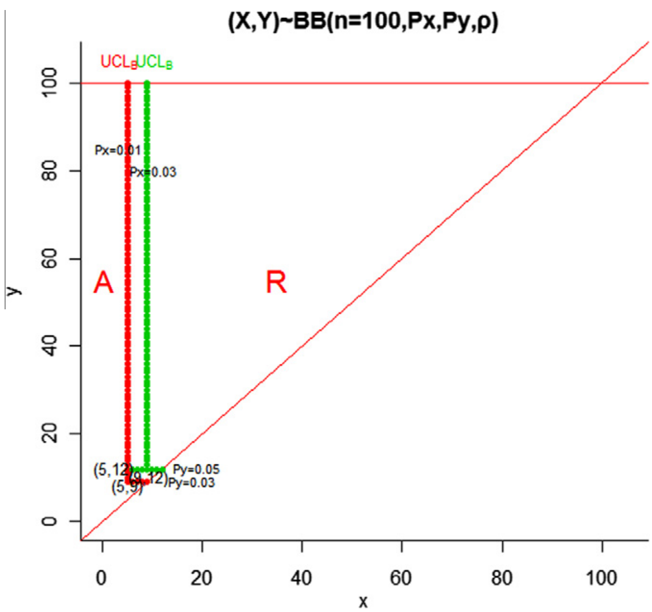

Fig. 4. BB control region for ($n = 100$, $p_x = 0.01$, 0.03, $p_y = 0.03$, 0.05, $\rho = 0.5$, $\alpha = 0.0027$).

Table 3
Paint defect data and monitoring results of the 3-typed charts.

Sample number	n	X	Y	np_x–e chart		BB control region	np_x–np_y chart	
1	100	1	9	in	in	in	in	in
2	100	3	8	in	in	in	in	in
3	100	3	10	in	in	in	in	in
4	100	3	7	in	in	in	in	in
5	100	1	3	in	in	in	in	in
6	100	1	6	in	in	in	in	in
7	100	2	7	in	in	in	in	in
8	100	4	8	in	in	in	in	in
9	100	3	8	in	in	in	in	in
10	100	3	14	in	out	in	in	in
11	100	2	15	in	in	in	in	in
12	100	1	5	in	in	in	in	in
13	100	3	8	in	in	in	in	in
14	100	2	11	in	in	in	in	in
15	100	2	7	in	in	in	in	in
16	100	2	6	in	in	in	in	in
17	100	5	10	in	in	in	in	in
18	100	1	7	in	in	in	in	in
19	100	2	8	in	in	in	in	in
20	100	10	15	out	in	in	out	in
21	100	3	10	in	in	in	in	in
22	100	2	10	in	in	in	in	in
23	100	3	8	in	In	in	in	in
24	100	2	10	in	in	in		

5. The ARL computation and comparison

The average run length (ARL) provides a measure of the sensitivity of the control chart. With the assumption of the process in control, the in-control ARL (ARL_0) for a control chart is the average number of samples before a signal is given. The out-of-control ARL (ARL_1) is the average number of samples that must be taken to detect the fraction nonconforming shift when the process is out of control. In this section, we compute ARL_1 for np_x–e chart, BB control region, Shewhart np_x–np_y chart and compare their detection ability.

For the np_x–e chart, the $ARL_1 = \frac{1}{1-\beta_1}$, β_1 is calculated from,

$$\beta_1 = P\big(LCL_x < X < UCL_x,\ LCL_e < e < UCL_e|(n, p_{x1}, p_{y1}, \rho)\big)$$

where UCL_x, LCL_x, UCL_e and LCL_e are in (1) and (2), p_{x1} and p_{y1} are out-of-control nonconforming rates.

For the BB control region, the $ARL_1 = \frac{1}{1-\beta_2}$. β_2 is calculated from

$$\beta_2 = P\big((X, Y) < UCL_B|(n, p_{x1}, p_{y1}, \rho)\big)$$

where UCL_B see (3)

For the Shewhart np_x–np_y chart, the $ARL_1 = \frac{1}{1-\beta_3}$, β_3 is computed from

$$\beta_3 = P\big(LCL_x < X < UCL_x,\ LCL_y < Y < UCL_y|(n, p_{x1}, p_{y1}, \rho)\big)$$

where $LCL_y = np_y - 3\sqrt{np_y(1-p_y)}$ and $UCL_y = np_y + 3\sqrt{np_y(1-p_y)}$.

To compare their detection ability, the ARL_1 for various combinations of ρ, p_{x1} and p_{y1} are calculated. We adopt $n = 100$, $\rho = 0.1(0.2)0.9$, $p_x = (0.001{:}0.1)$, $p_y = 0.0011$, $p_{y1} = (0.0015{:}0.30)$, and 50,000 data sets are generated.

The ARL_1 for np_x–e chart, BB control region and np_x–np_y chart are listed in Table 4.

We found that regardless of the value of ρ is, ARL_1 of np_x–e chart is always smaller than that of np_x–np_y chart and BB control region. The ARL_1 of BB control region is always smaller than that of np_x–np_y chart for small shift of p_y and small values of $\rho(0.1 \leqslant \rho \leqslant 0.5)$ and larger for large values of $\rho(0.5 \leqslant \rho \leqslant 0.9)$. The information demonstrates that the np_x–e chart detects shift in p_x and p_y faster than BB control region and np_x–np_y chart.

6. Conclusion

In this paper, we proposed a cause selecting control chart (np_x–e chart) to monitor dependent process stages with attributes data. The numerical example shows that the cause selecting control charts provide more information on the current process than the np_x–np_y chart and BB control region.

From the ARL_1 of np_x–e chart, BB control region and Shewhart np_x–np_y chart, the detection ability of np_x–e chart performs better than the other method for most cases regardless of the value of ρ.

When the correlation coefficient between Y and X, ρ, shifts from small to high, the detection ability of np_x–e chart performs better than Shewhart np_x–np_y chart and BB control region except $p_x = 0.01$. For small shift of p_y, The BB control region always has smaller ARL_1 than that of np_x–np_y chart when ρ is small $(0.1 \leqslant \rho \leqslant 0.5)$.

An advantage of the cause selecting chart over the BB control region is that it is easier to determine which process is out-of-control. The BB control region may indicate the process is out-of-control but it does not identify which step is out-of-control.

If we misused the Shewhart np_x–np_y chart or BB control region to control the second step, it might generate a false alarm. Hence the cause selecting control charts are thus recommended to monitor the process stages with attributes data.

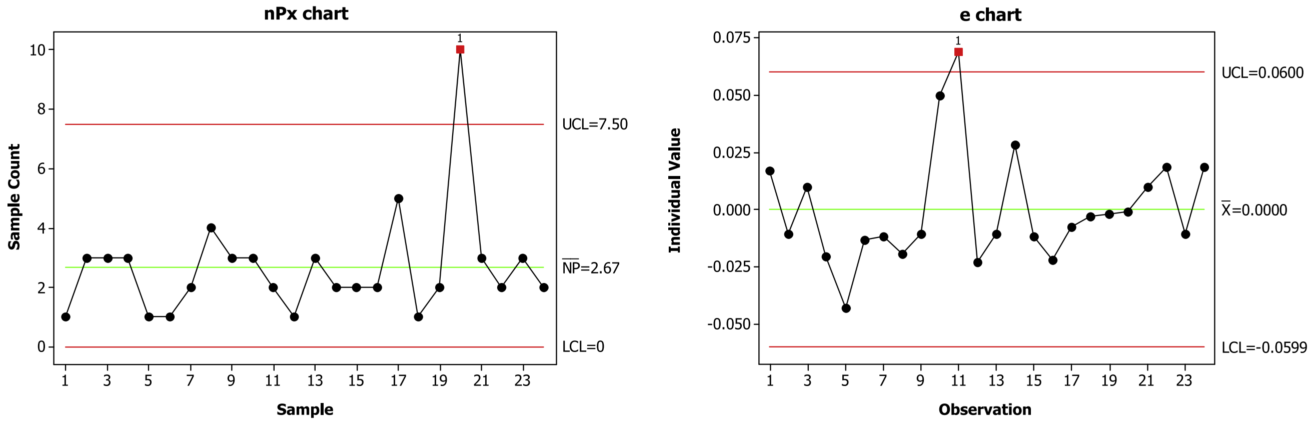

Fig. 5. Monitoring results of np_x and e charts.

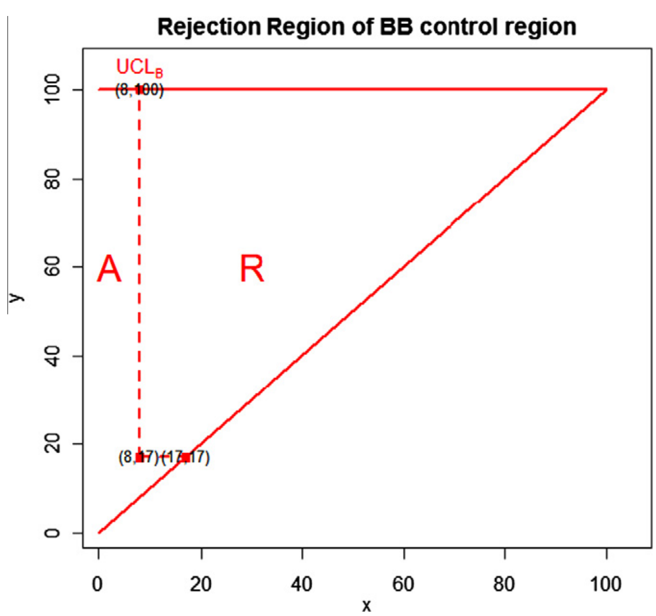

Fig. 6. Monitoring results of BB control region.

Table 4
ARL_1 for np_x–e chart, BB control region and Shewhart np_x–np_y chart.

ρ	p_x	p_{y1}	np_x–e chart	BB control region	Shewhart np_x–np_y chart
0.1	0.001	0.0015	4.69	13.16	68.03
		0.0030	3.04	14.86	24.33
	0.01	0.015	12.55	4.19	12.28
		0.030	2.77	11.82	2.74
	0.1	0.15	9.46	9.28	9.23
		0.30	1.01	100	1.01
0.3	0.001	0.0015	9.48	16.78	68.03
		0.002	3.91	18.08	45.66
	0.01	0.015	12.48	12.25	12.29
		0.020	6.56	40.82	6.38
	0.1	0.15	9.0	12.25	9.23
		0.18	4.18	40.82	2.93
0.5	0.001	0.0015	4.59	24.10	85.48
		0.0030	3.31	33.78	24.27
	0.01	0.015	12.38	6.36	12.29
		0.030	5.06	33.67	2.74
	0.1	0.15	4.73	17.45	9.23
		0.30	1.01	100	1.01
0.7	0.001	0.0015	4.52	46.51	68.03
		0.0030	8.97	71.43	45.66
	0.01	0.015	17.54	12.77	12.29
		0.030	16.40	69.44	6.38
	0.1	0.15	9.26	50.76	9.23
		0.30	1.85	100	2.93
0.9	0.001	0.0015	5.21	91.74	98.04
		0.0030	6.52	129.87	89.29
	0.01	0.011	39.06	16.37	23.31
		0.012	44.25	35.71	19.76
	0.1	0.105	93.46	4.61	181.82
		0.12	25.84	80	59.88

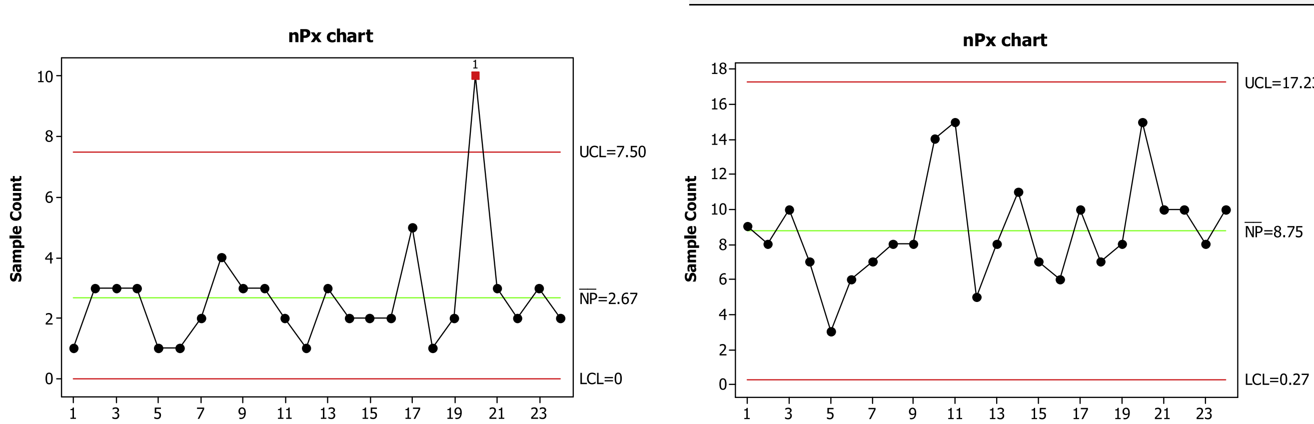

Fig. 7. Monitoring results of Shewhart np_x–np_y chart.

Acknowledgements

Support for this research was provided in part by the National Science Council of the Republic of China, Grant No. 98-2118-M-004-005-MY2 and Center for Service Innovation at National Chengchi University, Taiwan.

References

Biswas, A., & Huang, J. S. (2002). A new bivariate binomial distribution. *Statistics and Probability, 60*, 231–240.

Freeman, M. F., & Tukey, J. W. (1950). Transformations related to the angular and the square root. *Annals of Mathematical Statistics, 21*, 607–611.

Lowery, C. A., & Montgomery, D. C. (1995). A review of multivariate control charts. *IIE Transaction, 27*, 800–810.

Lu, X. S., Xie, M., Goh, T. N., & Lai, C. D. (1998). Control chart for multivariate attribute process. *International Journal of Product Research, 36*, 3477–3489.

Mandel, B. J. (1969). The regression control chart. *Journal of Quality Technology, 1*, 1–9.

Mukhopadhyay, A. R. (2008). Multivariate attribute control chart using Mahalanobis D^2 statistics. *Journal of Applied Statistics, 35*, 421–429.

Niaki, S. T. A. (2006). Bootstrap method approach in designing multi-attribute control charts. *International Journal of Advanced Manufacturing Technology, 35*, 434–442.

Niaki, S. T., & Abbasi, B. (2007). Skewness reduction approach in multi-attribute process monitoring. *Communications in Statistics-Theory and Methods, 36*, 2313–2325.

Patel, H. I. (1973). Quality control methods for multivariate binomial and Poisson distribution. *Technometrics, 15*, 103–112.

Wade, M., & Woodall, W. (1993). A review and analysis of cause-selecting control. *Journal of Quality Technology, 25*(3), 161–169.

Zhang, G. X. (1984). *A new type of control charts and a theory of diagnosis with control charts*. World quality congress transactions. American Society for Quality Control [pp. 175–185].