Pak J. Statist.
2003 Vol. 19(3) pp 291 – 300

NONPARAMETRIC CONTROL CHART BASED ON SUM OF RANKS

Fahad S. Altukife[1]

Institute of Public Administration, King Faisal Specialist Hospital and Research Center / Biostatistics, Epidemiology and Scientific Computing Department, Riyadh, Saudi Arabia

ABSTRACT

In this paper, new nonparametric control chart based on the sum of ranks called the RS control chart has been proposed. The new control chart uses the sum of ranks. The RS control chart has the ability to detect small shifts from the process mean when the heavy tailed distribution function is considered.

KEY WORDS

Average run length (ARL); Control Charts; Sum of Ranks

1. INTRODUCTION

Quality control is defined as an aggregate of functions designed to insure adequate quality in process. This is accomplished by an "initial critical study of design, materials, and processes, followed by periodic inspection of the products”. Most of the procedures involving the study of the quality use the statistical methods where they have been widely used in business and industry. While the quality control was used at the first time in the industry, quality control extends beyond manufacturing to all business disciplines, including marketing, finance, operations and human resources management. Quality control techniques apply as well to management within service industries, such as health care, transportation, and information systems. Quality control has implications for organizational development and business policy. Moreover, since the improvement of quality provides many benefits, the methodologies of quality control are recommended to many fields.

The Shewhart control chart is the most widely used statistical process control (SPC) technique and was proposed by Shewhart in 1931. In the case of monitoring the mean of such a process, the observations $x_{i1}, x_{i2}, \ldots, x_{in}$ are taken over time i, where x_{ik} is the k element taken at time i, i= 1, 2, …. Let $\bar{x}_i = \sum_{j=1}^{n} (x_{ij} / n)$ be the mean of the sample taken at time i. The mean of the process can be estimated as the mean of the m preliminary

[1] Assistance professor at the Institute of Public Administration, Riyadh, Saudi Arabia

samples averages and given by $\bar{\bar{x}} = \sum \bar{x}_i / m$. The standard deviation of the process can be estimated by $\frac{\bar{S}}{c_4\sqrt{n}}$ where $S_i = \sqrt{\frac{\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2}{n-1}}$ is the standard deviation of the i^{th} sample, $\bar{S} = \frac{\sum_{i=1}^{m} S_i}{m}$, and the values of c_4 are constants given for various sample sizes. Thus, the control limits for the Shewhart control chart are set at

$$\bar{\bar{x}} \pm L\left(\frac{\bar{S}}{c_4\sqrt{n}}\right), \qquad (1)$$

where L is arbitrary chosen to be 3 so that ARL, Average Run length, is approximately equal to 370. A value of L=3 implies that there is a probability of only 0.0026 of sample statistics falling outside the control limits if the process is in control.

A fundamental assumption of the Shewhart control chart is that the underlying distribution of the quality characteristics is normal where most data occurring in engineering manufacturing, and business is considered to be normally distributed. Due to the fact that so many processes, occur in practice, do not follow the normal distribution, the need for alternatives to the traditional control chart comes to play. The development and applications of control charts, which do not depend on any parametric distributional assumption, were discussed in few papers in the literature. Alloway and Raghavachari (1991) and Pappanastos and Adams (1996) studied control charts based on the Hodges-Lehmann estimator. Where this control chart based on the weighted variance was proposed by Bai and Chai (1995). Ferrell (1953) constructed a control chart using the medians of the sample ranges. Amin, et-al (1995) developed a control chart based on a sign statistic. Bakir (1995,1997), Arnold (1985,1986) developed other alternatives to the Shewhart control chart.

This paper proposes a new control chart based on the sum of ranks of the observations when the assumption of normality of the Shewhart control chart is not met. Mean and variance of sum of ranks are presented first. General guidelines are given for constructing the RS control chart. A comparison between the RC and the traditional Shewhart control chart is discussed by considering four distributions.

2. MEAN AND VARIANCE OF RANKS WITHIN SAMPLES

Let $x_1, x_2, \ldots, x_N$ be a random sample of size N from a population with cumulative distribution function *F(x)*. Let R_i be the rank of the i^{th} unit drawn from the sample. Suppose R is the sum of the n ranks selected at random , without replacement, from the N ranks.

Theorem 1: $E(R) = n\frac{N+1}{2}$.

Theorem 2: $\mathrm{Var}(R) = \frac{n(N+1)(N-n)}{12}$

Proofs are simple.

In case of ties in ranks:

$$V(R) = \frac{n(N+1)(N-n)}{12} - \frac{n(N-n)(\sum v^3 - \sum v)}{12N(N-1)},$$

Where v is the total number of ties

3. RC CONTROL CHART

Let x_{i1}, x_{i2}, . . ., x_{in} be a sample of independent observations of size n taken sequentially at time i, where i = 1, 2, ... and x_{ij} is the j^{th} element taken at time i. In order to estimate the control limits of the process based on ranks, m preliminary samples of size n are drawn assuming that the process stays in-control. The m samples of size n are assumed to be independent and taken from m identical distribution functions $F_1(x), F_2(x), \ldots, F_m(x)$. Let $N = mn$ be the total number of observations in m samples. For the purpose of constructing the control limits of the RS control chart, assign rank 1 to the smallest value of the total N observations, rank 2 to the second smallest observation among the N observations, and so on to the largest observation which will take the rank N. If several observations are equal to each other, assign the average ranks to each of the tied observations. Suppose R_{ij} is the rank assigned to the j^{th} observation in the i^{th} sample and define R_i to be the sum of ranks assigned to the i^{th} sample. That is,

$$R_i = \sum_{j=1}^{n} R_{ij}, \qquad i=1, 2, \ldots, m.$$

The sum of ranks $R_1, R_2, \ldots, R_m$ are used in Kruskal-Wallis test to test the null hypothesis that the distribution of a sequence of independent observations are obtained from m identical population distribution functions against the alternative that at least one of the populations tends to yield larger observations than the other populations. If the null hypothesis is true, it is expected that the pooled data can be treated as a single random sample from the common population of size nm. Consequently, when the observations are ranked according to the magnitude in the combined arrangement from 1 to N, it is expected that the ranks would be distributed well among the m samples. The RS control chart can help testing whether or not the m samples come from a single population.

3.1. Center Line and the Control Limits

From the mean and the variance of R_i, i=1,2, . . ., m, the center line of the RS control chart is obtained as follows:

$$E(\overline{R}) = E\left(\sum_{i=1}^{m} R_i \Big/ m\right) = \frac{n(N+1)}{2},$$

The control limits may be derived as follows:

$$\left.\begin{aligned} LCL &= \frac{n(N+1)}{2} - D\sqrt{\frac{n(N+1)(N-n)}{12}} \\ CL &= \frac{n(N+1)}{2} \\ UCL &= \frac{n(N+1)}{2} + D\sqrt{\frac{n(N+1)(N-n)}{12}}, \end{aligned}\right\} \qquad (2)$$

where D is selected to satisfy the ARL of the RS control chart. The control limits depend on the *n*, *m* and D and are constant for the same values of *n* and *m*.

The process will be in control if all R_i's are plotted inside the control limits and no systematic behavior is presented.

4. EVALUATING THE RS CONTROL CHART

The properties of the RS control chart were evaluated using a simulation based on *10,000* runs for the different combination of the sample distributions; normal, uniform, double exponential, and two parameter exponential, and sample sizes of *n*= 4, 6, 8, 10, and 12 when a total of 25 samples are drawn. The distribution of observations from the process is considered to have mean zero and a variance one for all considered distribution functions. These distribution functions are given as follows:

1. The normal distribution function:

$$f(x) = \frac{1}{\sqrt{2ps^2}} e^{-(x-m)^2/2s^2}, \text{ where } -\infty < x < \infty .$$

The values of m and s^2 are selected to be 0 and 1 respectively.

2. The uniform distribution function:

$$f(x) = \frac{1}{2b}, \quad \text{where } -b < x < b .$$

The value of b is chosen to be $\sqrt{3}$.

3. The double exponential distribution function

$$f(x) = \frac{1}{a\sqrt{2}} e^{-|(x-a)/b|}, \text{ where } -\infty < x < \infty .$$

where a, b are chosen to be 0 and $1/\sqrt{2}$ respectively

1. The two-parameter exponential distribution function:

$$f(x) = 1/a \, e^{-(x-b)/a} \text{ , where } b < x < \infty \, .$$

The values of a, b are chosen to be 1 and -1 respectively

The Shewhart control chart, was also simulated considering equation (1). The control limits of the two control charts were set such that the ARL_0 , in control average run length, approximately equaled 370 for all possible sets of sample sizes.

5. IN-CONTROL PERFORMANCE

In this section, the comparison of the performance between the RS control chart and the traditional parametric control chart is presented based on four distribution functions described in section (4). Those distributional forms are selected for two reasons:(1) they represent a wide range of distributions extending from the short-tailed (uniform) distribution and the heavy-tailed (double exponential) distribution to the skewed (two-parameter exponential) distribution (2). They are typical for previous literature regarding nonparametric control charts.

In the comparisons of ARL properties of the RS to the Shewhart control chart, it seems natural to match these two control charts in the sense that they both have the same ARL_0. This procedure was repeated *10,000* times using Re-sampling Stat program with different trials of L and D so that the two control charts have ARL_0 approximately 370 in all considered cases.

Table I provides the values of D and L for the RS and Shewhart control charts respectively when 25 samples consisting of n(n=4, 6, 8, 10, and 12) observations are selected. These values are used for two goals:(1) for comparative purposes of the proposed control chart to the Shewhart control chart and (2) to provide the users of the appropriate value of D such that the ARL_0 of the OS control chart is nearly equals the one from the Shewhart control chart.

Table-I

Values of D and L for the RS and Shewhart control charts assuming the in-control status.

n	Distribution Function							
	Normal		Uniform		Double Exp.		Two Para. Exp.	
	D	L	D	L	D	L	D	L
4	2.680	2.987	2.678	2.680	2.805	3.590	2.670	4.08
6	2.790	2.988	2.780	2.763	2.829	3.380	2.795	3.765
8	2.850	2.988	2.850	2.846	2.847	3.310	2.845	3.564
10	2.875	2.989	2.873	2.881	2.872	3.258	2.874	3.479
12	2.892	3.00	2.891	2.910	2.892	3.228	2.890	3.417

Based on the previous knowledge of the shape of the distribution and the sample size, the user might select the appropriate value of D. For example, if a heavy tailed distribution is considered and samples of size 8 observations are used to construct the t control limits of the RS control chart, the practitioners would use D=2.847 where other values of D can be obtained in the same manner. It seems natural, due to the central limit theorem, for the uniform distribution function that L increases as the sample size increases. Where as, for the double exponential and two-parameter exponential distribution functions, L decreases as the sample size increases. It is expected for any distribution function that L will converge to 3 as the sample size gets large due the fact that L=3 is used in the Shewhart control chart with the probability of type I error equals to 0.0027. The same findings were also concluded by Burr (1967), Schilling and Nelson (1976), and Yourstone and Zimmer (1992).

To determine whether or not the process is in control at the first stage, several criteria can be applied. The interpretation of the pattern of the plotted points on the control chart plays an important role to judge the process status. Breyfogle (1999) gave some interpretations of patterns on the traditional control chart. These interpretations may be applied to the RC control chart with some modifications. To monitor the i^{m+1} sample once the control limits are set from the initial m samples, the m+1 samples are pooled together to get the values of R_i, for i=1, 2, . . ., m+1. The value of R_{m+1} is compared to the control limits at (2) by replacing N by Ñ when Ñ = n(m + 1).

6. RS CONTROL CHART FOR UNEQUAL SAMPLE SIZES

It is easy to extend the case of unequal sample sizes. Now, let R_i and n_i be the sum of ranks and sample size of the i^{th} sample respectively. Define $\overline{R}_i$ to be the average of the ranks assigned to the i^{th} sample. The average of the ranks for each sample ($\overline{R}_i$, $i = 1, 2, \ldots, m$) is plotted against the sample number. The control limits are modified so that the center line is the weighted average of R_i, for $i = 1, 2, \ldots, m$. Thus

$$\overline{R}_c = \frac{\sum_{i=1}^{m} n_i R_i}{\sum_{i=1}^{m} n_i}$$, where the pooled variance of the process is given by

$$V_{pooled}(\overline{R}_c) = \left[\frac{\sum_{i=1}^{m} (n_i - 1) V(\overline{R}_i)}{\sum_{i=1}^{m} (n_i - 1)} \right], \text{ such that } \quad V(\overline{R}_i) = \frac{(N+1)(N-n_i)}{12 n_i}.$$

Thus, the control limits in equation (2) can be modified as follows:

$$\left.\begin{aligned} LCL &= \overline{R}_c - L \cdot \sqrt{V_{pooled}\ (\overline{R}_c)} \\ CL &= \overline{R}_c \\ UCL &= \overline{R}_c + L \cdot \sqrt{V_{pooled}\ (\overline{R}_c)} \end{aligned}\right\} \quad (3)$$

There are two approaches in applying the control limits in equation (3). The first one is to select the value of D and calculate the control limits for each individual sample. In this case, different control limits are considered for each sample. Another approach is to base the control limit calculations on the average sample size $\overline{n}$. The interpretation of the status of the chart is similar to the one with equal sample sizes.

7. THE AFFECT OF THE RS CONTROL CHART

Let us now study the effectiveness of the RS control chart in detecting shifts from the process mean assuming the sample sizes are equal. After matching the two control charts in the sense that they both have the same ARL_0, the process mean is shifted by the amount of δ and the ARL_1 (out of control average run length) values are obtained for all considered cases. The size of δ is measured in terms of the process mean and standard deviation and given by:

$$\delta = (\mu_1 - \mu_0)/\sigma ,$$

where μ_1 is the mean of the process after the shift has occurred, μ_0 is the mean of the process before the shift, and σ is the process standard deviation.

The control limits are set as if the process are in control considering the values of D and L from Table I. Random samples of size n are drawn from the four different distributions considering different values of δ and ARL_1. Without loss of generality, the values of δ range from 0 to 2 in the process mean.

7.1. Normal Distribution Function

The results in Table II show the values of ARL_1 for the RS and Shewhart control charts when the data is normal. The RS control chart is only slightly less efficient than the corresponding Shewhart control chart to detect shifts for all considered sample sizes and the values of δ. However, this loss in efficiency is small when the sample size gets large. In such cases, it is recommended to use the Shewhart control chart in the presence of normality.

Table-II. Values of the ARL_1 for the RS and Shewhart control charts assuming normality.

δ	Sample Size									
	4		6		8		10		12	
	RS	Shewhart	RS	Shewhart	RS	Shewhart	RS	Shewhart	RS	Shewhart
0.00	370	370	370	370	370	370	370	370	370	370
0.25	213.1	180.5	169.6	136.3	123.4	118.1	103.4	97.4	91.3	89.9
0.50	68.7	51.1	37.8	27.2	34.4	20.2	16.2	13.6	12.6	11.4
0.75	21.66	16.1	11.5	8.9	5.9	5.8	3.9	3.1	2.6	2.2
1.00	9.2	6.6	4.6	3.5	1.9	1.8	1.2	1.2	1.7	1.5
1.50	2.7	2.1	1.6	1.3	1.2	1.1279	1.1	1.0499	1.1	1.1
2.00	1.4	1.2	1.1	1.1	1.1	1	1	1	1	1

7.2 Uniform Distribution Function

Table III displays the values of ARL_1 for the RS and Shewhart control charts when the data comes from the uniform distribution function. The Shewhart control chart is more powerful in detecting shifts than the proposed control chart for all considered sample sizes.

Table- III. Values of the ARL_1 for the RS and Shewhart control charts assuming uniform distribution function.

δ	Sample Size									
	4		6		8		10		12	
	RS	Shewhart	RS	Shewhart	RS	Shewhart	RS	Shewhart	RS	Shewhart
0.00	370	370	370	370	370	370	370	370	370	370
0.25	135.3	125.7	115.7	105.7	101.9	89.3	86.7	80.1	71.5	65.1
0.50	44.1	26.5	31.9	23.2	22.9	16.5	16.4	12.6	12.7	9.6
0.75	18.	8.9	11.9	5.7	7.1	4.9	4.8	3.3	3.2	2.8
1.00	9.8	4.3	5.6	2.9	2.6	2.1	1.6	1.5	1.1	1.1
1.50	3.5	1.7	1.9	1.3	1.3	1.1	1.1	1	1	1
2.00	1.6	1.1	1.1	1	1	1	1	1	1	1

7.3. Double Exponential Distribution Function

Table IV gives the values of ARL_1 for the RS and Shewhart control charts in the presence of a heavy tailed distribution function. The RS control chat is more efficient to detect shifts than the Shewhart control chart for all values of δ when $6 \geq n$. The same finding was obtained by Amin et al.(1995) when considering the sign test.

Table- IV. Values of the ARL_1 for the RS and Shewhart control charts assuming the double exponential function.

δ	Sample Size									
	4		6		8		10		12	
	RS	Shewhart	RS	Shewhart	RS	Shewhart	RS	Shewhart	RS	Shewhart
0.00	370	370	370	370	370	370	370	370	370	370
0.25	191.7	297.0	166.9	268.3	154.2	98.9	131.5	79.8	112.9	49.1
0.50	80.8	193.6	52.8	117.6	38.6	19.9	26.7	7.8	20.1	4.8
0.75	33.1	96.6	16.9	49.1	10.9	2.8	6.9	1.9	4.7	1.5
1.00	14.1	47.8	6.2	19.9	3.8	1.4	2.4	1.1	1.1	1.1
1.50	3.5	12.2	2.5	4.7	1.8	1	1.4	1	1.2	1
2.00	1.1	4.2	1.5	1.8	1.2	1	1	1	1	1

7.4 Two Parameter Exponential Distribution Function

The values of ARL_1 for the RS and Shewhart control charts when the data is markedly skewed are displayed in table V. In all considered cases of sample sizes and values of δ, the RS control chart is only slightly less efficient than the corresponding Shewhart control chart to detect shifts.

Table- V. Values of the ARL_1 for the RS and Shewhart control charts assuming the two-parameter exponential distribution function.

δ	Sample Size									
	4		6		8		10		12	
	RS	Shewhart	RS	Shewhart	RS	Shewhart	RS	Shewhart	RS	Shewhart
0.00	370	370	370	370	370	370	370	370	370	370
0.25	140.5	47.7	123.7	37.3	112.3	31.2	89.8	27.4	78.6	23.4
0.50	41.9	11.7	31.6	8.6	23.3	5.9	17.1	5.2	13.6	4.2
0.75	15.9	4.6	11.6	3.1	6.7	2.5	5.2	2.1	3.5	1.7
1.00	7.8	2.4	5.1	1.8	2.7	1.4	1.8	1.3	1.3	1.2
1.50	2.9	1.3	2.1	1.1	1.6	1.1	1.4	1	1.2	1
2.00	1.8	1.1	1.4	1.1	1.2	1	1.1	1	1.1	1

8. CLOSING REMARKS

There is one important note that should be considered about the RS control chart. The control limits of the RC control chart are very stable and not effected by the drawn samples and capable of being deigned to achieve specific values of ARL_0. While we give an alternative to the Shewhart control chart, other SPC techniques such as CUSUM and

EWMA control charts can be applied to monitor R_i's and compare its performance to the parametric procedure.

REFERENCES

1. Alloway, J. A. Jr., and Raghavachari, M. (1991). Control chart based on the Hodges-Lehmann estimator. *Journal of Quality Technology. 23(4)* . 336-347.
2. Amin, R. W.; Reynolds, M. R., Jr.; and Bakir, S. (1995). Nonparametric quality control charts based on the sign test. *Communications in Statistics-Theory and Methods, 24*. 1597-1624.
3. Arnold, B. (1985). The sign test in current control. *Statistische Hefte, 26,* 253-262.
4. Arnold, B. (1986). Comparison of the approximate and exact optimum economic design of control charts based on the Sign Statistics. *Statistische Heft, 27*, 239-241.
5. Bai, D. S., and Chai, I. S. (1995). $\overline{X}$ and R control charts for skewed populations. *Journal of Quality Technology. 27(2)*, 120-126.
6. Bakir, S. T. (1995). A nonparametric Shewhart control chart in the presence of a blocking effect. *Proceedings of the section on quality and productivity, American Statistical Association*.110-114.
7. Bakir, S. T.(1997). Quality control charts for detecting a general change in a process. *Proceedings of the Quality and Productivity Section; American Statistical Association*. 53-56.
8. Breyfogle, F. W. (1999). *Implementing Six Sigma*: *Smarter Solutions Using Statistical Methods, 3rd edition*, Wiley and Sons Inter-science, New York.
9. Burr, I. W. (1967). The effect of non-Normality on constants for $\overline{X}$ and R charts. *Industrial Quality Control. 23*, 563-568.
10. Ferrell, E. B. (1953). Control charts using midranges and medians. *Industrial Quality Control. 9,* 30 – 34
11. Pappanastos, E. A. and Adams, B. M. (1996). Alternative Designs of the Hodges-Lehmann Control Chart. *Journal of Quality Technology. 28(2),* 213-223.
12. Schilling, E. G., and Nelson, P. R. (1976). The Effect Of non-normality on the control limits of charts. *Journal of Quality Technology, 8(4),* 183-187.
13. Yourstone, S. A. and Zimmer, W. J. (1992). Non-normality and the design of control charts for averages. *Decision Sciences, 23,*1099-1113.