Aktüerya Derneği

İstatistikçiler Dergisi: İstatistik & Aktüerya

Journal of Statisticians: Statistics and Actuarial Sciences

IDIA 12, 2019, 1, 1-14

Geliş/Received:13.11.2018, Kabul/Accepted: 19.05.2019

www.istatistikciler.org

Araştırma Makalesi / Research Article

Ranked set sampling designs for monitoring the skew-normal process

Derya Karagöz

Hacettepe University, Department of Statistics, Beytepe, Ankara, TURKEY
deryacal@hacettepe.edu.tr
orcid.org/0000-0002-8283-8573

Nursel Koyuncu

Hacettepe University, Department of Statistics, Beytepe, Ankara, TURKEY
nkoyuncu@hacettepe.edu.tr
orcid.org/0000-0003-1065-3411

Abstract

In this study, we have proposed to use various ranked set sampling designs to construct the mean charts based on Shewhart, weighted variance and skewness correction methods that are applied to monitor the process variability under the skew-normal process. The performance of the mean charts based on various ranked set sampling designs are compared with simple random sampling by Monte Carlo simulation. Simulation results revealed that the mean charts based on various ranked set sampling perform much better than simple random sampling.

Keywords: Weighted variance method, Skew-normal distribution, Ranked set sampling designs, Skewness correction method.

Öz

Çarpık-normal sürecin izlenmesi için sıralı küme örnekleme tasarımları

Bu çalışmada, çarpık-normal süreç altında süreç değişkenliğini izlemek için uygulanan ortalama kartları oluşturmak için Shewhart, ağırlıklı varyans ve düzeltilmiş çarpıklık yöntemlerine dayanan çeşitli sıralı küme örnekleme tasarımlarını kullanmak önerilmiştir. Sıralı küme örnekleme tasarımlarına dayanan ortalama kartlarının performansı Monte Carlo benzetimi kullanılarak basit rastgele örnekleme ile karşılaştırılmıştır. Simülasyon sonuçları, çeşitli sıralı küme örneklemesine dayanan ortalama kartlarının basit rastgele örneklemeden çok daha iyi performans gösterdiğini ortaya koymuştur.

Anahtar sözcükler: *Ağırlıklı varyans yöntemi, Çarpık-normal dağılım, Sıralı küme örnekleme tasarımları, Düzeltilmiş çarpıklık yöntemi.*

1. Introduction

In recent years, the statisticians tried to take the advantage of using various sampling designs to construct control chart limits. Ranked Set Sampling (RSS) is one of the most popular sampling and effective design which is introduced by [1]. Most statisticians modified this design and proposed various ranked set sampling designs. Recently Zamanzade and Al-Omari [2] have introduced neoteric ranked set sampling (NRSS). They preferred to use these sampling designs since they are more efficient compared to simple random sampling (SRS). In the literature there are some important studies which combine sampling designs and control charts as follows:

Muttlak and Al-Sabah [3] have studied different quality control charts for the sample mean using RSS, median rank set sampling (MRSS) and extreme ranked set sampling (ERSS). Abujiya and Muttlak [4] have developed control chart using double ranked set sampling. Al-Nasser and Al-Rawwash [5] have examined control charts under robust ranked set sampling. Pongpullponsak and Sontisamran [6] have studied mean charts for multiple characteristics under ranked set sampling design. Li et.al. [7] developed two control charts based on the exact sampling distributions of sample mean and range.

Koyuncu and Karagoz [8] proposed Shewhart $\bar{X}$ and R control charts using some ranked-based sampling designs when the parent distribution follows a bivariate normal distribution. Koyuncu and Karagoz [9] studied the mean control chart limits also Karagoz and Koyuncu [10] examined range charts control limits under the bivariate skewed distributions by using different ranked set sampling designs. Silva et al. [11] considered the design and performance of control charts using NRSS in monitoring normal distributed processes.

The most commonly used procedures of statistical quality control (SQC), control charts acceptance sampling plans, are often implemented under the assumption of normal data, which rarely holds in practice. The type of data usually exhibit moderate to strong asymmetry as well as light to heavy tails. This study focus on parametric family of skew-normal distribution introduced by O'Hagan and Leonard [12] and investigated with more detail by Azzalini [13]. The aim of this paper is to construct the mean charts based on Shewhart, weighted variance (WV) and skewness correction (SC) methods are applied to monitor the process variability under the skew-normal distribution by using different ranked set sampling designs.

The rest of this paper is organized as follows: In Section 2, we have explained skew-normal distribution and also its multivariate extension is given. The sampling designs are explained in Section 3. The Shewhart, WV and SC methods are given respectively in Section 4. The performance of the mean charts based on various ranked set sampling designs are compared with SRS by Monte Carlo simulation. The simulation study is given to compare the Type I risk probabilities of these mean charts with respect to different subgroup sizes for skew-normal distribution in Section 5. Finally, Section 6 and 7 end up with the results of this study.

2. Skew-normal distribution

Definition 1.1: A random variable Y is said to have a location-scale skew-normal distribution, with location λ scale δ and shape parameter α, being denoted $Y \sim SN(\lambda, \delta^2, \alpha)$, if its probability density function (pdf) is given by

$$f(y;\lambda,\delta,\alpha) = \frac{2}{\delta}\phi(\frac{Y-\lambda}{\delta})\Phi(\alpha\frac{Y-\lambda}{\delta}), \tag{2.1}$$

where ϕ and Φ denote the standard normal probability function and it's cumulative distribution function (cdf), respectively. If $\lambda = 0$ and $\delta = 1$, we obtain the standard skew-normal distribution, denoted $SN(\alpha)$. Let a random variable $Z = \frac{X-\lambda}{\delta}$ is said to be skew-normal with parameter α, written $Z \sim SN(\alpha)$, if its density function is

$$f(z;\alpha) = 2\phi(z)\Phi(\alpha z) \tag{2.2}$$

[13]. Note that, if $Y \sim SN(\lambda, \delta^2, \alpha)$ then $Z = \frac{Y-\lambda}{\delta} \sim SN(\alpha)$.

The moment generating function and the first moments of Z are given in Azzalini [13]. In particular,

$$E(Z) = \frac{2^{\frac{1}{2}}}{\pi}\delta$$

$$Var(Z) = 1 - \frac{2}{\pi}\delta^2$$

where δ is related to α via the relationships $\delta = \frac{\alpha}{1+\alpha^2}^{\frac{1}{2}}$, $\alpha = \frac{\delta}{1-\delta^2}^{\frac{1}{2}}$. Also,

$$k_3 = \frac{(4-\pi)E(Z)^3}{2Var(Z)^{\frac{3}{2}}} \quad (2.3)$$

where k_3 is the skewness, varies in the interval $(-0.995, 0.995)$. Let $Z = (Z_1, \dots, Z_k)^T$, the multivariate extension of its density function is

$$f_k(z) = 2\phi_k(z; \Omega)\Phi(\alpha^T z), z\epsilon R^k \quad (2.4)$$

where

$$\alpha^T = \frac{\lambda^T \Psi^{-1}\Delta^{-1}}{(1+\lambda^T\Psi^{-1}\lambda)^{\frac{1}{2}}} \quad (2.5)$$

$$\Delta = diag((1-\delta_1^2), \dots, (1-\delta_k^2)), \quad (2.6)$$

$$\lambda = (\lambda(\delta_1), \dots, \lambda(\delta_k))^T, \quad (2.7)$$

$$\Omega = \Delta(\Psi + \lambda\lambda^T)\Delta \quad (2.8)$$

and $\phi_k(z; \Omega)$ denotes the density function of the k dimensional multivariate normal distribution with the standardized marginals and correlation matrix Ω (detailed information see at [13], p.717).

3. Sampling designs

In this study one of the most important point is to generate correlated skewed finite population. In ranked set sampling designs we assumed that there is an auxiliary variable which is ranked easily. To use these sampling designs, we need to rank quality data (Y) with correlated auxiliary variable (X). Therefore we have assumed that our two finite population have bivariate skew-normal distribution with the parameters as explained in Section 2. We consider bivariate skew-normal distribution so lets say that a random variable Z with density function given in Eq.(2.4) is a 2-dimensional skew normal variable, with vector λ of shape parameters and dependence parameter Ψ. For brevity, we shall write $Z \sim SN_2(\lambda, \Psi)$.

3.1. Simple random sampling design

Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be a bivariate random sample from a population of size N with pdf $f(x, y)$, means μ_x, μ_y, variances σ_x^2, σ_y^2 and correlation coefficient ρ_{xy}. Assume that the auxiliary variable X is correlated with the variable of interest Y. The unbiased estimators of the population mean of study and auxiliary variables based on SRS are as follows: $\bar{Y}_{(SRS)} = \frac{1}{n}\sum_{i=1}^{n} Y_i$; $\bar{X}_{(SRS)} = \frac{1}{n}\sum_{i=1}^{n} X_i$. The range of sample for study and auxiliary variables can be calculated using following equations respectively:

$$R_{Y(SRS)} = max(Y_i) - min(Y_i); \quad (3.1)$$

$$R_{X(SRS)} = max(X_i) - min(X_i) \tag{3.2}$$

where $i = 1,2, \dots, n$.

3.2. Ranked Set Sampling Design

RSS design can be described as follows:

1. Select a simple random sample of size n^2 units from the target finite population and divide them into n samples each of size n.

2. Rank the units within each sample in increasing magnitude by using personal judgement, eye inspection or based on a concomitant variable.

3. Select the ith ranked unit from the ith sample

4. Repeat steps 1 through 3, m times if needed to obtain a RSS of size $N = nm$.

Let

$$(X_{11h}, Y_{11h}), (X_{12h}, Y_{12h}), \dots, (X_{1nh}, Y_{1nh});$$

$$(X_{21h}, Y_{21h}), (X_{22h}, Y_{22h}), \dots, (X_{2nh}, Y_{2nh});$$

$$\vdots;$$

$$(X_{n1h}, Y_{n1h}), (X_{n2h}, Y_{n2h}), \dots, (X_{nnh}, Y_{nnh})$$

be n independent bivariate random samples with pdf $f(x,y)$, each of size n in the hth cycle, $(h = 1,2, \dots, m)$. Let $(X_{i(1:n)h}, Y_{i[1:n]h}), (X_{i(2:n)h}, Y_{i[2:n]h}), \dots, (X_{i(n:n)h}, Y_{i[n:n]h})$ be the order statistics of $X_{i1h}, X_{i2h}, \dots, X_{inh}$ and the judgement order of $Y_{i1h}, Y_{i2h}, \dots, Y_{inh}$ $(i = 1,2, \dots, n)$, where $()$ and $[]$ indicate that the ranking of X is perfect and ranking of Y has errors. Assume that the measured units using RSS are $(X_{1(1:n)h}, Y_{1[1:n]h}), (X_{2(2:n)h}, Y_{2[2:n]h}), \dots, (X_{n(n:n)h}, Y_{n[n:n]h})$.

Then the RSS estimators of population mean for study and auxiliary variables can be written as

$$\bar{Y}_{(RSS)} = \frac{1}{nm} \sum_{h=1}^{m} \sum_{i=1}^{n} Y_{i[i:n]h}; \tag{3.3}$$

$$\bar{X}_{(RSS)} = \frac{1}{nm} \sum_{h=1}^{m} \sum_{i=1}^{n} X_{i(i:n)h}. \tag{3.4}$$

The RSS estimators of population range for study and auxiliary variables can be written as

$$R_{Y(RSS)} = max(Y_{i[i:n]h}) - min(Y_{i[i:n]h}); \tag{3.5}$$

$$R_{X(RSS)} = max(X_{i(i:n)h}) - min(X_{i(i:n)h}) \tag{3.6}$$

where $(i = 1,2, \dots, n)$.

3.3. Median ranked set sampling design

For the sake of brevity we follow Al-Omari [14]' sampling design and notations. MRSS design can be described as in the following steps:

1. Select n random samples each of size n bivariate units from the population of interest.

2. The units within each sample are ranked by visual inspection or any other cost free method with respect to a variable of interest.

3. If n is odd, select the $((n+1)/2)$ th-smallest ranked unit X together with the associated Y from each set, i.e., the median of each set. If n is even, from the first $n/2$ sets select the $(n/2)$th ranked unit X together with the associated Y and from the other sets select the $((n+2)/2)$ th ranked unit X together with the associated Y .

4. The whole process can be repeated m times if needed to obtain a sample of size nm units.

Let $(X_{i(1)h}, Y_{i[1]h}), (X_{i(2)h}, Y_{i[2]h}), \dots, (X_{i(n)h}, Y_{i[n]h})$ be the order statistics of $X_{i1h}, X_{i2h}, \dots, X_{inh}$ and the judgement order of $Y_{i1h}, Y_{i2h}, \dots, Y_{inh}$ $(i = 1,2, \dots, n), (h = 1,2, \dots, m)$ where () and [] indicate that the ranking of X is perfect and ranking of Y has errors. For odd and even sample sizes the units measured using MRSS are denoted by MRSSO and MRSSE, respectively. For odd sample size let $(X_{1(\frac{n+1}{2})h}, Y_{1[\frac{n+1}{2}]h}), (X_{2(\frac{n+1}{2})h}, Y_{2[\frac{n+1}{2}]h}), \dots, (X_{n(\frac{n+1}{2})h}, Y_{n[\frac{n+1}{2}]h})$ denote the observed units by MRSSO. The sample mean of X and Y are given as following respectively,

$$\bar{Y}_{(MRSSO)} = \frac{1}{nm} \sum_{h=1}^{m} \sum_{i=1}^{n} Y_{i[\frac{n+1}{2}]h} \tag{3.7}$$

$$\bar{X}_{(MRSSO)} = \frac{1}{nm} \sum_{h=1}^{m} \sum_{i=1}^{n} X_{i(\frac{n+1}{2})h} \tag{3.8}$$

For even sample size let $\left(X_{1(\frac{n}{2})h}, Y_{1[\frac{n}{2}]h}\right), \left(X_{2(\frac{n}{2})h}, Y_{2[\frac{n}{2}]h}\right), \dots, \left(X_{\frac{n}{2}(\frac{n}{2})h}, Y_{\frac{n}{2}[\frac{n}{2}]h}\right),$

$\left(X_{\frac{n+2}{2}(\frac{n+2}{2})h}, Y_{\frac{n+2}{2}[\frac{n+2}{2}]h}\right), \left(X_{\frac{n+4}{2}(\frac{n+2}{2})h}, Y_{\frac{n+4}{2}[\frac{n+2}{2}]h}\right), \dots, \left(X_{n(\frac{n}{2})h}, Y_{n[\frac{n}{2}]h}\right)$ denote the observed units by MRSSE.

The sample mean of X and Y are given respectively

$$\bar{Y}_{(MRSSE)} = \frac{1}{nm} \sum_{h=1}^{m} (\sum_{i=1}^{\frac{n}{2}} Y_{i[\frac{n}{2}]h} + \sum_{i=\frac{n+2}{2}}^{n} Y_{i[\frac{n+2}{2}]h} \tag{3.9}$$

$$\bar{X}_{(MRSSE)} = \frac{1}{nm} \sum_{h=1}^{m} (\sum_{i=1}^{\frac{n}{2}} X_{i(\frac{n}{2})h} + \sum_{i=\frac{n+2}{2}}^{n} X_{i(\frac{n+2}{2})h}). \tag{3.10}$$

The ranges using MRSS can be calculated if the sample size odd

$$R_{Y(MRSSO)} = max(Y_{i[\frac{n+1}{2}]h}) - min(Y_{i[\frac{n+1}{2}]h}); \tag{3.11}$$

$$R_{X(MRSSO)} = max(X_{i(\frac{n+1}{2})h}) - min(X_{i(\frac{n+1}{2})h}) \tag{3.12}$$

for $(i = 1,2, \dots, n)$; if the sample size even

$$R_{Y(MRSSE)} = max(Y_{i[\frac{n}{2}]h}, Y_{i[\frac{n+2}{2}]h}) - min(Y_{i[\frac{n}{2}]h}, Y_{i[\frac{n+2}{2}]h}); \tag{3.13}$$

$$R_{X(MRSSE)} = max\left(X_{i(\frac{n}{2})h}, X_{i(\frac{n+2}{2})h}\right) - min\left(X_{i(\frac{n}{2})h}, X_{i(\frac{n+2}{2})h}\right). \tag{3.14}$$

RSS and MRSS designs can be seen in Figure1.

RSS (n=5)

Cycle	Rank 1	2	3	4	5
	■	◊	◊	◊	◊
	◊	■	◊	◊	◊
1	◊	◊	■	◊	◊
	◊	◊	◊	■	◊
	◊	◊	◊	◊	■
	■	◊	◊	◊	◊
	◊	■	◊	◊	◊
2	◊	◊	■	◊	◊
	◊	◊	◊	■	◊
	◊	◊	◊	◊	■
	■	◊	◊	◊	◊
.	◊	■	◊	◊	◊
.	◊	◊	■	◊	◊
.	◊	◊	◊	■	◊
	◊	◊	◊	◊	■
	■	◊	◊	◊	◊
	◊	■	◊	◊	◊
m	◊	◊	■	◊	◊
	◊	◊	◊	■	◊
	◊	◊	◊	◊	■

MRSS n=5 (n is odd)

Cycle	Rank 1	2	3	4	5
	◊	◊	■	◊	◊
	◊	◊	■	◊	◊
1	◊	◊	■	◊	◊
	◊	◊	■	◊	◊
	◊	◊	■	◊	◊
	◊	◊	■	◊	◊
	◊	◊	■	◊	◊
2	◊	◊	■	◊	◊
	◊	◊	■	◊	◊
	◊	◊	■	◊	◊
	◊	◊	■	◊	◊
.	◊	◊	■	◊	◊
.	◊	◊	■	◊	◊
.	◊	◊	■	◊	◊
	◊	◊	■	◊	◊
	◊	◊	■	◊	◊
	◊	◊	■	◊	◊
m	◊	◊	■	◊	◊
	◊	◊	■	◊	◊
	◊	◊	■	◊	◊

MRSS n=6 (n is even)

Cycle	Rank 1	2	3	4	5	6
	◊	◊	■	◊	◊	◊
	◊	◊	■	◊	◊	◊
1	◊	◊	■	◊	◊	◊
	◊	◊	◊	■	◊	◊
	◊	◊	◊	■	◊	◊
	◊	◊	◊	■	◊	◊
	◊	◊	■	◊	◊	◊
	◊	◊	■	◊	◊	◊
2	◊	◊	■	◊	◊	◊
	◊	◊	◊	■	◊	◊
	◊	◊	◊	■	◊	◊
	◊	◊	◊	■	◊	◊
	◊	◊	■	◊	◊	◊
.	◊	◊	■	◊	◊	◊
.	◊	◊	■	◊	◊	◊
.	◊	◊	◊	■	◊	◊
	◊	◊	◊	■	◊	◊
	◊	◊	◊	■	◊	◊
	◊	◊	■	◊	◊	◊
	◊	◊	■	◊	◊	◊
m	◊	◊	■	◊	◊	◊
	◊	◊	◊	■	◊	◊
	◊	◊	◊	■	◊	◊
	◊	◊	◊	■	◊	◊

Figure 1: The representation of RSS and MRSS designs

3.4. Neoteric ranked set sampling design

Zamanzade and Al-Omari [2] have defined a new neoteric ranked set sampling. The NRSS sceheme can be described as follows:

1. Select a simple random sample of size n^2 units from the target finite population.

2. Ranked the n^2 selected units in an increasing magnitude based on a concomitant variable, personel judgment or any inexpensive method.

3. If n is an odd, then select the $[\frac{n+1}{2}+(i-1)n]$ th ranked unit for $(i=1,2,\dots,n)$. If n is an even, then select the $[l+(i-1)n]$ th ranked unit, where $[l=\frac{n}{2}]$ if i is an even and $[l=\frac{n+2}{2}]$ if i is an odd for $(i=1,2,\dots,n)$.

4. Repeat steps 1 through 3 m times if needed to obtain a NRSS of size $N=nm$.

Let $(X_{1h},Y_{1h}),(X_{2h},Y_{2h}),\dots,(X_{n^2h},Y_{n^2h})$ be n^2 simple random units selected from the population with the pdf $f(x,y)$ and let $(X_{(1)h},Y_{[1]h}),(X_{(2)h},Y_{[2]h}),\dots,(X_{(n^2)h},Y_{[n^2]h})$ be order statistics of measured units by NRSS $(h=1,2,\dots,m)$. The sample means of study and auxiliary variables under NRSS scheme are given as following

$$\bar{Y}_{(NRSS)} = \frac{1}{nm}\sum_{h=1}^{m}\sum_{i=1}^{n} Y_{[(i-1)n+i]h} \quad (3.15)$$

$$\bar{X}_{(NRSS)} = \frac{1}{nm}\sum_{h=1}^{m}\sum_{i=1}^{n} X_{[(i-1)n+i]h}. \quad (3.16)$$

The the ranges of study and auxiliary variables under NRSS scheme are given as following

$$R_{Y(NRSS)} = max(Y_{[i]h}) - min(Y_{[i]h} \quad (3.17)$$

$$R_{X(NRSS)} = max(X_{(i)h}) - min(X_{(i)h}). \quad (3.18)$$

4. Methods for mean control charts

We aim to monitor the process variability under the skew-normal distribution by using the mean control charts based on Shewhart, Weighted Variance and Skewness Correction methods.

4.1. Shewhart Method

Control charts, also known as Shewhart charts (after Walter A. Shewhart) or process-behavior charts, are a statistical process control tool used to determine if a manufacturing or business process is in a state of control. The control chart was invented by Walter A. Shewhart in the 1920s [15] to construct the control limits. The control limits of mean chart for Shewhart method are given as follows:

$$UCL_{Shewhart} = \bar{\bar{Y}} + \frac{3}{d_2\sqrt{n}}\bar{R}_y \quad (4.1)$$

$$LCL_{Shewhart} = \bar{\bar{Y}} - \frac{3}{d_2\sqrt{n}}\bar{R}_y \quad (4.2)$$

where $\bar{\bar{Y}}$ is the mean of the subgroups means and d_2 is constant that depends on the subgroup size n, and this constant is calculated when the distribution is normal [13]. In Eq.(4.1) and Eq.(4.2) the location and scale estimators considered are the mean of the sample means and sample ranges, respectively

$$\bar{\bar{Y}} = \frac{1}{k}\sum_{i=1}^{k}\bar{Y}_i = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{1}{n}\sum_{i=1}^{n} Y_{ij}\right) \quad (4.3)$$

$$\bar{R}_y = \frac{1}{k}\sum_{i=1}^{k} R_{yi} \quad (4.4)$$

where k is the number of samples, R_{yi} is the range of the ith sample where $i = 1,2,\ldots,k$ and

$j = 1,2,\ldots,n.$

The control limits of the mean chart for Shewhart method using different sampling designs are defined as follows:

$$UCL_{Shewhart(j)} = \bar{\bar{Y}}_{(j)} + \frac{3}{d_2\sqrt{n}}\bar{R}_{y(j)} \quad (4.5)$$

$$LCL_{Shewhart(j)} = \bar{\bar{Y}}_{(j)} - \frac{3}{d_2\sqrt{n}}\bar{R}_{y(j)}. \quad (4.6)$$

where (j) implies the sampling methods which are SRS, RSS, MRSS and NRSS sampling designs, respectively. In Equation 4.5 and 4.6, $\bar{Y}_j$ can be rewriten

$$\bar{\bar{Y}}_{(SRS)} = \frac{1}{k}\sum_{i=1}^{k} \bar{Y}_{(SRS)}, \quad \bar{\bar{Y}}_{(RSS)} = \frac{1}{k}\sum_{i=1}^{k} \bar{Y}_{(RSS)}$$

$$\bar{\bar{Y}}_{(MRSS)} = \frac{1}{k}\sum_{i=1}^{k} \bar{Y}_{(MRSS)}, \quad \bar{\bar{Y}}_{(NRSS)} = \frac{1}{k}\sum_{i=1}^{k} \bar{Y}_{(NRSS)} \tag{4.7}$$

and $\bar{R}_j$ can be rewritten

$$\bar{R}_{y(SRS)} = \frac{1}{k}\sum_{i=1}^{k} R_{y(SRS)}, \quad \bar{R}_{y(RSS)} = \frac{1}{k}\sum_{i=1}^{k} R_{y(RSS)},$$

$$\bar{R}_{y(MRSS)} = \frac{1}{k}\sum_{i=1}^{k} R_{y(MRSS)}, \quad \bar{R}_{y(NRSS)} = \frac{1}{k}\sum_{i=1}^{k} R_{y(NRSS)} \tag{4.8}$$

for the mean charts limits under different ranked set sampling. Here k is the number of samples.

4.2. Weighted variance method

The WV method proposed by Choobineh and Ballard [16] decomposes the skewed distribution into two parts at its mean and both parts are considered symmetric distributions which have the same mean and different standard deviation. In this method, μ_y and μ_R are normally estimated using the grand mean of the subgroup means $\bar{\bar{Y}}$ and the mean of the subgroup ranges $\bar{R}$, respectively.

The control limits of mean chart for WV method are defined by Bai and Choi [17] as follows:

$$UCL_{WV} = \bar{\bar{Y}} + 3\frac{\bar{R}}{d_2^*\sqrt{n}}\sqrt{2\hat{P}_Y} \tag{4.9}$$

$$LCL_{WV} = \bar{\bar{Y}} - 3\frac{\bar{R}}{d_2^*\sqrt{n}}\sqrt{2(1-\hat{P}_Y)} \tag{4.10}$$

where d_2^* is the control chart constant for mean chart based on the WV method and $P_Y = P(Y \le \bar{Y})$ is the probability that the quality variable Y will be less than or equal to its mean $\bar{Y}$.

The control limits of mean chart for WV method using different sampling designs are defined as follows:

$$UCL_{WV(j)} = \bar{\bar{Y}}_{(j)} + 3\frac{\bar{R}_{y(j)}}{d_2^*\sqrt{n}}\sqrt{2\hat{P}_{Y_{(j)}}} \tag{4.11}$$

$$LCL_{WV(j)} = \bar{\bar{Y}}_{(j)} - 3\frac{\bar{R}_{y(j)}}{d_2^*\sqrt{n}}\sqrt{2(1-\hat{P}_{Y_{(j)}})}. \tag{4.12}$$

In Equation (4.11) and (4.12), $\hat{P}_{Y_{(j)}}$ is the probability to be estimated from

$$\hat{P}_{Y_{(j)}} = \frac{\sum_{i=1}^{k}\sum_{m=1}^{n}\delta(\bar{\bar{Y}}_{(j)} - Y_{im(j)})}{nk} \tag{4.13}$$

where k and n are the number of samples and the number of observations in a subgroup, respectively.

4.3. Skewness correction method

The skewness correction method was proposed by [6] for constructing the $\bar{X}$ and R control charts under skewed distributions. It's asymmetric control limits are obtained by taking into consideration the degree of skewness estimated from subgroups, and making no assumptions about distributions. When the distribution is symmetric, $\bar{X}$ chart is closer to the Shewhart chart.

The control limits of the mean chart for SC method are defined by Chan and Cui[6] as follows:

$$UCL_{SC} = \bar{\bar{Y}} + (3 + c_4^*)\frac{\bar{R}}{d_2^*\sqrt{n}}$$
$$LCL_{SC} = \bar{\bar{Y}} + (-3 + c_4^*)\frac{\bar{R}}{d_2^*\sqrt{n}} \quad (4.14)$$

where c_4^* and d_2^* are the control chart constants for the SC method. The constant c_4^* is obtained as follows:

$$c_4^* = \frac{\frac{4}{3}k_3(\bar{Y})}{1+0.2k_3^2(\bar{Y})} \quad (4.15)$$

where $k_3(\bar{Y})$ is the skewness of the subgroup mean $\bar{Y}$ [6].

The control limits of the $\bar{Y}$ chart for SC method using different sampling designs are defined as follows:

$$UCL_{SC(j)} = \bar{\bar{Y}}_{(j)} + (3 + c_4^*)\frac{\bar{R}_{y(j)}}{d_2^*\sqrt{n}} \quad (4.16)$$

$$LCL_{SC(j)} = \bar{\bar{Y}}_{(j)} + (-3 + c_4^*)\frac{\bar{R}_{y(j)}}{d_2^*\sqrt{n}} \quad (4.17)$$

where j implies the sampling methods which are SRS, RSS, MRSS and NRSS designs.

5. The simulation study

When the parameters of the process are unknown, control charts can be applied in a two-phase procedure. In Phase I, control charts are used to define the in-control state of the process and to assess process stability for ensuring that the reference sample is representative of the process. The parameters of the process are estimated from Phase I sample and control limits are estimated for using in Phase II. In Phase II, samples from the process are prospectively monitored for departures from the in-control state. The Type I risk indicates the probability of a subgroup $\bar{Y}$ falling outside the ± 3 sigma control limits. When the process is in-control, the Type I risks are 0.27%. However, due to the control limits, about 0.0027 of all control points will be false alarms and have no assignable cause of variation [10]. We have used RStudio 1.2 version to run simulation studies.

5.1. Determination of the control charts constants

An assumption of non-normality is incorporated into the constants d_2 and c_4 to correct the control chart limits as d_2^* and c_4^* based on the distribution. Therefore, the constants are corrected under this conditions. The corrected constants are determined such that the expected value of the statistic divided by the constant is equal to the true value of σ. The WV method constant d_2^* is calculated by taking the mean of range $\left(\frac{R}{\sigma}\right)$. The SC method constant c_4^* is calculated by using Eq: (4.15). We obtain $E(\bar{R})$ by simulation: we generate 100.000 times k samples of size n, compute

R for each instance and take the average of the values. The results for all constants for $k = 30$ are presented in Table 2 for $n = 3,5,7,10$.

Table 1: The parameters of skew-normal distribution

k_3	Ω	α	δ^2
0.11	0.1	3	0.047
0.42	0.5	10	0.99
0.81	0.8	20	0.79
0.99	0.9	100	0.89

Table 2: The values of constants for the skew-normal distribution

	n =3		**n=5**		**n=7**		**n=10**	
k_3	d_2^*	c_4^*	d_2^*	c_4^*	d_2^*	c_4^*	d_2^*	c_4^*
0.11	1.9098	0.0833	2.4740	0.0656	2.8179	0.0554	3.1622	0.0456
0.42	1.9091	0.3168	2.4676	0.2460	2.8033	0.2080	3.1342	0.1748
0.81	1.9020	0.5931	2.4300	0.4665	2.7300	0.3971	3.0113	0.3340
0.99	1.8883	0.7170	2.3810	0.5636	2.6520	0.4811	2.9002	0.4060

5.2. Monte Carlo simulation study

Let E_i denote the event that the ith sample mean is beyond the limits. Further, denote by $P(Ei|\bar{Y},\hat{\sigma})$ the conditional probability that for given $\bar{\bar{Y}}$ and $\hat{\sigma}$, the sample mean $\bar{Y}_i$ is beyond the control limits

$$P(Ei|\bar{Y},\hat{\sigma}) = P(\bar{Y}_i < LCL \quad or \quad \bar{Y}_i > UCL) \tag{5.1}$$

Given $\bar{\bar{Y}}$ and $\hat{\sigma}$, the events E_s and E_t $(s \neq t)$ are independent. Therefore, the run length has a geometric distribution with parameter $P(Ei|\bar{Y},\hat{\sigma})$. When we take the expectation over the estimation data Y_{ij} we get the unconditional probability of one sample showing a Type I risk as following

$$P(Ei) = E(P(Ei|\bar{Y},\hat{\sigma})). \tag{5.2}$$

These expectations are simulated by generating 10 000 times k data samples of size n, computing for each data set the conditional value and averaging the conditional values over the data sets.

In this study, we work on control charts for skew-normal distribution in the statistical process control. Table 1 gives the parameters of skew-normal distribution based on the skewness. We consider the skewness between 0.11-0.99. The distribution parameters give this table are selected based on these skewness values.

Table 3: The results of p for mean charts based on different sampling designs

	$k_3 = 0.11$											
	Shewhart				WV				SC			
n	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**
3	0.01191	0.00825	0.00830	0.01164	0.01209	0.00884	0.00875	0.01244	0.01182	0.00781	0.00795	0.01140
5	0.00694	0.00468	0.00399	0.00746	0.00712	0.00489	0.00404	0.00845	0.00710	0.00451	0.00375	0.00713
7	0.00552	0.00375	0.00307	0.00615	0.00552	0.00395	0.00325	0.00701	0.00565	0.00349	0.00300	0.00602

10	0.00463	0.00292	0.00278	0.00642	0.00482	0.00295	0.00285	0.00672	0.00477	0.00275	0.00266	0.00632
	k_3 = 0.42											
	Shewhart				WV				SC			
n	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**
3	0.01296	0.01065	0.01066	0.01196	0.01403	0.01164	0.01201	0.01347	0.01287	0.01127	0.01085	0.01133
5	0.00715	0.00609	0.00592	0.00753	0.00786	0.00693	0.00666	0.00876	0.00723	0.00589	0.00573	0.00686
7	0.00554	0.00478	0.00450	0.00665	0.00608	0.00512	0.00501	0.00745	0.00554	0.00460	0.00437	0.00594
10	0.00452	0.00368	0.00370	0.00563	0.00487	0.00417	0.00418	0.00610	0.00438	0.00330	0.00352	0.00570
	k_3 = 0.81											
	Shewhart				WV				SC			
n	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**
3	0.01351	0.00612	0.00482	0.01115	0.01668	0.00806	0.00590	0.01246	0.01448	0.00842	0.00686	0.01537
5	0.00755	0.00227	0.00166	0.00661	0.00972	0.00316	0.00222	0.00752	0.00708	0.00294	0.00238	0.00796
7	0.00506	0.00127	0.00089	0.00491	0.00682	0.00176	0.00137	0.00544	0.00500	0.00156	0.00136	0.00606
10	0.00421	0.00070	0.00057	0.00375	0.00521	0.00101	0.00073	0.00460	0.00344	0.00092	0.00085	0.00500
	k_3 = 0.99											
	Shewhart				WV				SC			
n	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**
3	0.01410	0.00284	0.00155	0.01086	0.01866	0.00425	0.00248	0.01275	0.01425	0.00429	0.00315	0.01799
5	0.00739	0.00061	0.00015	0.00535	0.01043	0.00099	0.00032	0.00592	0.00632	0.00092	0.00036	0.00884
7	0.00448	0.00024	0.00007	0.00401	0.00685	0.00043	0.00014	0.00419	0.00399	0.00033	0.00013	0.00608
10	0.00348	0.00004	0.00002	0.00324	0.00506	0.00008	0.00006	0.00372	0.00252	0.00009	0.00002	0.00499

Table 4: The results of ARL for mean charts based on different sampling designs

	k_3 = 0.11											
	Shewhart				WV				SC			
n	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**
3	83.96306	121.2121	120.4819	85.91065	82.71299	113.1222	114.2857	80.38585	84.60237	128.0410	125.7862	87.7193
5	144.09222	213.6752	250.6266	134.04826	140.44944	204.4990	247.5248	118.34320	140.84507	221.7295	266.6667	140.2525
7	181.15942	266.6667	325.7329	162.60163	181.15942	253.1646	307.6923	142.65335	176.99115	286.5330	333.3333	166.1130
10	215.98272	342.4658	359.7122	155.76324	207.46888	338.9831	350.8772	148.80952	209.64361	363.6364	375.9398	158.2278
	k_3 = 0.42											
	Shewhart				WV				SC			
n	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**
3	77.16049	93.89671	93.80863	83.61204	71.27584	85.91065	83.26395	74.23905	77.70008	88.73114	92.1659	88.26125
5	139.86014	164.20361	168.91892	132.80212	127.22646	144.30014	150.15015	114.15525	138.31259	169.77929	174.5201	145.77259
7	180.50542	209.20502	222.22222	150.37594	164.47368	195.31250	199.60080	134.22819	180.50542	217.39130	228.8330	168.35017
10	221.23894	271.73913	270.27027	177.61989	205.33881	239.80815	239.23445	163.93443	228.31050	303.03030	284.0909	175.43860
	k_3 = 0.81											
	Shewhart				WV				SC			
n	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**
3	74.01925	163.3987	207.4689	89.6861	59.95204	124.0695	169.4915	80.25682	69.06077	118.7648	145.7726	65.06181
5	132.45033	440.5286	602.4096	151.2859	102.88066	316.4557	450.4505	132.97872	141.24294	340.1361	420.1681	125.62814
7	197.62846	787.4016	1123.5955	203.6660	146.62757	568.1818	729.9270	183.82353	200.00000	641.0256	735.2941	165.01650
10	266.66667	1428.5714	1754.3860	237.5297	191.93858	990.0990	1369.8630	217.39130	290.69767	1086.9565	1176.4706	200.00000
	k_3 = 0.99											
	Shewhart				WV				SC			
n	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**	**SRS**	**RSS**	**NRSS**	**MRSS**
3	70.92199	352.1127	645.1613	92.08103	53.59057	235.2941	403.2258	78.43137	70.17544	233.1002	317.4603	55.58644
5	135.31800	1639.3443	6666.6667	186.91589	95.87728	1010.1010	3125.0000	168.91892	158.22785	1086.9565	2777.7778	113.12217
7	223.21429	4166.6667	14285.7143	249.37656	145.98540	2325.5814	7142.8571	238.66348	250.62657	3030.3030	7692.3077	164.47368
10	308.64198	25000.0000	50000.0000	287.35632	197.62846	12500.0000	16666.6667	268.81720	396.82540	11111.1111	50000.0000	200.40080

Phase 1:

1. Generate correlated finite population using bivariate skew-normal distribution with the parameters as given in Table 1 of size $N = 10000$.

2. Select samples size n from bivariate skew-normal distribution for $n = 3,5,7,10$ using different sampling schemes with SRS, RSS, MRSS and NRSS respectively.

3. Repeat step 1.b 30 times $(k = 30)$ for the reference process.

4. Compute the control limits of mean chart by using different sampling designs.

Phase 2:

1. Generate correlated bivariate skew-normal distribution with the parameters as described in step 1.a.

2. Select samples size n from bivariate skew-normal distribution varieties for $n = 3,5,7,10$ using different sampling schemes with SRS, RSS, MRSS and NRSS respectively as in step 1.b.

3. Repeat step 2.b 100 times ($k = 100$) for the control process.

4. Compute the sample statistics for mean charts for different sampling designs.

5. Record whether the sample statistics calculated in step 2.d are within the control limits of step 1.d. or not for all methods.

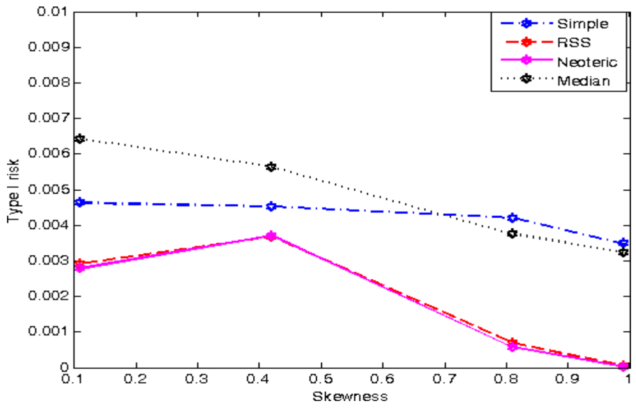
6. Repeat steps 1.a through 2.d, 10000 times and obtain an average Type I risk for each method.

6. Results

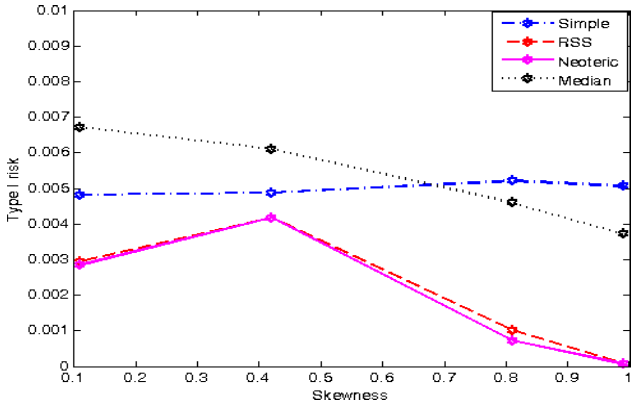
In this section, we obtain the Type I risks to compare the performance of the mean charts for bivariate skew-normal process by using different ranked set sampling designs for $n = 3,5,7,10$. When the process in control, it is expected that p is to be as low as possible. The desired p value of the control limits is to provide a p of 0.0027. Under non-normality, when skewness increases, the p increases. Tables 3 and 4, present the p and ARL values for the mean chart by using ranked set sampling designs under skew-normal distribution, respectively. Figure 2 represents Type I risks of $\bar{X}$ charts for Type I skew-normal distribution, n=10. We can sum up the results as following:

•When the distribution is approximately symmetric ($k_3 = 0.11, k_3 = 0.42$): The Shewhart mean chart under the NRSS design has the best performance. WV and SC mean charts under the NRSS designs can be also used as an alternative. As seen in Table 3, the p values of the under the RSS design Shewhart, WV and SC mean charts are more or less similar for $k_3 = 0.42$.

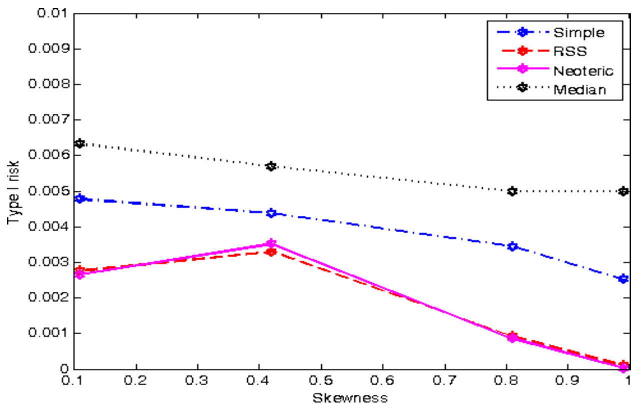
• When the skewness of distribution increases ($k_3 = 0.81, k_3 = 0.99$), the p values of the SC mean charts gives desirable results for small sample size, while the p values of the SC mean charts are deteriorated. For large sample sizes, Shewhart and WV mean charts under the MRSS design and SC mean chart under the SRS design perform better than other sampling designs for $k_3 = 0.99$.



(a) Shewhart $\overline{X}$ chart



(b) WV $\overline{X}$ chart



(c) W V R chart

Figure 2: Type I risks of $\overline{X}$ charts for skew-normal distribution, n=10

7. Conclusion

In this paper, the construction of the mean charts limits based on Shewhart, WV and SC methods under the SRS, RSS, MRSS and NRSS designs are considered for skew-normal process. We have studied the effect of the skewness on mean charts under skew-normal distributed data for small and large sample sizes by using these four sampling designs. A Monte-Carlo simulation study is run to compare the mean charts under different sampling designs. In the simulation study, the p and ARL values of these mean charts are obtained to compare the methods and designs. The results can be summed up as follows: When the distribution is approximately symmetric ($k_3 = 0.11, k_3 = 0.42$): The Shewhart mean chart under the NRSS design has the best performance. WV and SC mean charts under the NRSS designs can be also used as an alternative. When the skewness of distribution increases ($k_3 = 0.81, k_3 = 0.99$), the SC mean charts gives desirable results for small sample size, however the SC mean charts are deteriorated for large sample size. For large sample size, Shewhart and WV mean charts under the MRSS design and SC mean chart under the SRS design perform better than other sampling designs for $k_3 = 0.99$ Finally, the proposed methods using different sampling designs for the mean chart can be used to monitor the skew-normal process.

References

[1] G.A. McIntyre, 1952, A method for unbiased selective sampling using ranked sets, *Australian Journal of Agricultural Research*, 3, 385-390.

[2] E. Zamanzade, A.I. Al-Omari, 2016, New ranked set sampling for estimating the population mean and variance, *Hacettepe Journal of Mathematics and Statistics*, 45, 6, 1891-1905.

[3] H.A. Muttlak, W.S. Al-Sabah, 2003, Statistical quality control based on ranked set sampling, *Journal of Applied Statistics*, 30, 9, 1055-78.

[4] M.R. Abujiya, M.R., Muttlak H. A., 2004, Quality control chart for the mean using double ranked set sampling, *Journal of Applied Statistics*, 31, 10, 1185-1201.

[5] A.D. Al-Nasser, M. Al-Rawwash, 2007, A control chart based on ranked data, *Journal of Applied Sciences*, 7, 14, 1936-1941.

[6] A. Pongpullponsak, P. Sontisamran, 2013, Statistical quality control based on ranked set sampling for multiple characteristics, *Chiang Mai J Sci* , 40, 3, 485-498.

[7] C. Li, N.C. Su, P.F. Su, Y. Shyr, 2014, The Design $\bar{X}$ and R charts for Skew Normal Distributed Data, *Communications in Statistics- Theory and Methods*, 43:4908-4924.

[8] N. Koyuncu, D. Karagöz, 2017, Shewhart X-bar ($\bar{X}$) and R control charts using new ranked set sampling schemes, *Istatistik, Journal of the Turkish Statistical Association (IJTSA)*, 10, 3, 63-71.

[9] N. Koyuncu, D. Karagöz, 2018, New mean charts for bivariate asymmetric distributions using different ranked set sampling designs, *Quality Technology and Quantitative Management*, 15, 5, 602-621.

[10] D. Karagöz, N. Koyuncu, 2019, New ranked set sampling schemes for range charts limits under bivariate skewed distributions, *Soft Computing*, 23:5, 1573-1587.

[11] G.P. Silva, C. A. Taconeli, W.M. Zeviani, I.S. Guimares, 2017, An improved quality control chart to monitor the mean based on ranked sets, *arXiv preprint arXiv:1709.04979*.

[12] O'Hagan A., Leonard T., 1967, Bayes estimation subject to uncertainty about parameter constraints, *Biometrika*, 63, 201-203 (1976).

[13] A. Azzalini,A. Dalla Valle, 1996, The multivariate skew-normal distribution, *Biometrika*, 83, 4, 715-726.

[14] A.I. Al-Omari, 2012, Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling,. *Statistics and Probability Letters*, 82, 1883-1890.

[15] D.C. Montgomery, 1997, Introduction to Statistical Quality Control, *John Wiley&Sons. Inc*., USA .

[16] F. Choobineh, J.L. Ballard, 1987, Control-limits of QC Charts for skewed distributions using weighted variance, *IEEE Transactions on Reliability*, 36, 4, 473-477.

[17] D.S. Bai, I.S. Choi, 1995, $\bar{X}$ and R Control charts for skewed populations, *Journal Of Quality Technology*, 27, 120-131.

[18] L.K. Chan, H.J. Cui, 2003, Skewness correction $\bar{X}$ and R charts for skewed distributions, *Naval Research Logistics*, 50, 1-19.