CHILEAN
JOURNAL OF
STATISTICS

Edited by Víctor Leiva

Volume 10 Number 2
December 2019
ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society
SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA

Aims

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (`www.soche.cl`). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS is an international scientific forum strongly committed to gender equality, open access of publications and data, and the new era of information. The ChJS covers a broad range of topics in statistics, data science, data mining, artificial intelligence, and big data, including research, survey and teaching articles, reviews, and material for statistical discussion. In particular, the ChJS considers timely articles organized into the following sections: Theory and methods, computation, simulation, applications and case studies, education and teaching, development, evaluation, review, and validation of statistical software and algorithms, review articles, letters to the editor.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

Contents

Chilean Journal of Statistics
Vol. 10, No. 2, December 2019, 131–154

Statistical Quality Control
Research Paper

Performance of Shewhart control charts based on neoteric ranked set sampling to monitor the process mean for normal and non-normal processes

Guilherme Parreira da Silva[1], Cesar Augusto Taconeli[1,*], Walmes Marques Zeviani[1], and Isadora Aparecida Sprengoski do Nascimento

[1]Department of Statistics, Federal University of Paraná, Curitiba, Brazil

(Received: 05 April 2019 · Accepted in final form: 28 May 2019)

Abstract

In this study, we consider the design and performance of control charts using the neoteric ranked set sampling (NRSS) in monitoring industrial processes. NRSS is a recently proposed sampling design, based on the traditional ranked set sampling (RSS). NRSS differs from RSS by constituting, originally, a single set of k^2 sample units, instead of k sets of size k, where k is the final sample size. We evaluate NRSS control charts by average, median and standard deviation of run lengths, based on Monte Carlo simulation results. NRSS control charts perform the best, compared to RSS and some of its extensions, in most simulated scenarios. The impact of imperfect ranking and non normality are also evaluated. An application to concrete strength data serves as an illustration of the proposed method.

Keywords: Generalized normal distribution · Imperfect ranking · Perfect ranking · Run length · Skew-normal distribution

Mathematics Subject Classification: Primary 62D05 · Secondary 62P30

1. Introduction

Nowadays, technological resources are widely available for the real-time monitoring of many industrial processes. Even so, it must be recognized that sampling still plays a fundamental role in statistical quality control. Factors such as high costs, time of inspection and destructive tests may limit the evaluation of a large number of items. In this context, efficient sampling designs, providing more accurate results with smaller sample sizes, are highly useful. Ranked set sampling (RSS) and its extensions have been shown as efficient alternatives to more conventional methodologies (such as simple random sampling - SRS), when ranking sample units, according to their possible values, is substantially cheaper or easier than effectively measuring them. In the area of statistical quality control, RSS and its extensions can be applied, for example, to develop statistical quality control charts.

Originally proposed in 1924 by Walter A. Shewhart, statistical quality control charts (or

*Corresponding author. Email: Email: taconeli@ufpr.br

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

simply control charts) constitute a relevant tool for visualizing industrial processes and identifying assignable causes of variation (Shewhart, 1924; Montgomery, 2009). A process is said to be under statistical control when no special or assignable causes are present. Several alternatives to the original control charts were proposed, providing greater speed in detecting out-of-control situations. These alternatives include: the use of additional or alternative decision rules (Koutras et al., 2007); adaptive sampling schemes (Costa and De Magalhaes, 2007; Santore et al., 2019); nonparametric control charts (Qiu, 2018) or even the use of alternative sampling designs to the usual SRS. In this study, we consider a variety of RSS-based designs for constructing control charts.

Proposed by McIntyre (1952), the RSS is an effective sampling design when the variable of interest is expensive or difficult to measure, but it is possible ranking sample units efficiently according to some accessible and cheap criterion (Chen et al., 2003). The ranking process can be performed based, for example, on an expert's judgment or using some concomitant variable. In the first case (personal judgment), the sample units may be ordered based on visual inspection by using photos or videos, among others. In the other case, the sample units are ordered according their possible values for the variable of interest, but based only on values assessed for some correlated and accessible concomitant variable. In both cases, if the ranking criterion is not susceptible to errors, we have the perfect ranking scenario. Errors in the ranking process, however, frequently happen. In this situation, we say that the ranking process is imperfect.

RSS becomes more efficient than SRS as long as a more accurate and accessible ordering criterion is available. Several studies have shown the superiority of RSS over SRS for estimation of different population parameters (see Chen, 2007; Al-Omari and Bouza, 2014; Consulin et al., 2018). Additionally, a large number of sampling designs derived from the original RSS were proposed, such as median ranked set sampling (MRSS) by Muttlak (1997), extreme ranked set sampling (ERSS) by Samawi et al. (1996), and double ranked set sampling (DRSS) by Al-Saleh and AlKadiri (2000), among others.

RSS and its related sampling designs have been studied in the context of statistical quality control. Muttlak and Al-Sabah (2003) considered RSS and two of its modifications, ERSS and MRSS, in the design of Shewhart control charts. The authors have shown, based on an extensive simulation study, that RSS-based control charts dominate their SRS counterpart, requiring, on average, fewer samples to detect a change in the process mean. Additionally, MRSS have showed the best performance among the three sampling designs based on ranked sets. Improved control charts for DRSS schemes were also considered in the design of quality control charts. This class of sampling designs is characterized by the initial selection and ranking of k^3 (instead of k^2) sample units to draw a sample of size k after two ranking cycles. DRSS control charts outperform those based on a single ordering cycle. Recently, Mahdizadeh and Zamanzade (2019) presented a an economic variation of double RSS which reduces the number of training sample units to almost half. Furthermore, memory-based control charts using RSS, as cumulative sum or exponentially weighted moving average chart, were developed and discussed in Abid et al. (2017) and Haq et al. (2015), among others. Al-Omari and Bouza (2014) present a bibliographic review of RSS and control charts based on their related designs.

Zamanzade and Al-Omari (2016) recently proposed neoteric ranked set sampling (NRSS), another sampling design originated from RSS. Technically, its fundamental difference to RSS is the constitution and ordering of a single set of k^2 sample units, instead of k sets of size k like in RSS, MRSS and ERSS. After the ordering process, k units are chosen to compose the final sample, selected according to their specific ranks. The effect of creating a large initial set is the reduction of sample units variance, once the dispersion of order statistics decreases as the sample size increases. This reduction overcomes the covariances induced by sample units selected from the same ranked set. In this way, it was found, for different sample sizes, correlation levels between the variable of interest and an auxiliary variable and probability distributions that NRSS overcomes RSS and SRS for estimating population mean and variance. As additional studies regarding NRSS and its higher efficiency over RSS and other RSS-based designs we recommend Koyuncu (2018) and Taconeli and Cabral (2019).

NRSS was firstly considered for control charts by Koyuncu and Karagöz (2018) to monitor the mean of bivariate asymmetric distributions. The authors studied the type I error using different RSS designs under perfect ranking (that is, when there are no errors in the ranking process). They considered the Type I Marshall-Olkin bivariate Weibull and bivariate lognormal distributions. They verified that the NRSS and RSS designs have type I error closest to 0.0027, an usual type I error adopted for Shewhart control charts. Moreover, Nawaz and Han (2019) have compared NRSS, RSS, MRSS, and ERSS in the design of homogeneously weighted moving average control charts, registering that NRSS turns out to present the best performance among the considered RSS-based schemes in monitoring the process mean under bivariate normal distribution.

In this paper, we analyze the power of Shewhart-type control charts for monitoring the process mean based on NRSS. The remainder of this article is organized as follows. In Section 2, we briefly describe the RSS-based designs. The Shewhart-type control chart based on NRSS is presented in Section 3. Section 4 covers a simulation study conducted to evaluate the performance of NRSS control charts. A case study is in Section 5, while our concluding remarks are provided in Section 6.

2. Neoteric ranked set sampling and other sampling designs based on ranked sets

In this section, we briefly describe the sampling designs considered in this study. Initially, the original RSS design can be described as presented in Algorithm 1.

Algorithm *1* RSS scheme

1: Selection of k^2 units of the population using SRS, allocating them, randomly, in k sets of size k;
2: Ranking the sample units in each set according to the possible values of the variable of interest, using the pre-established ordering criterion;
3: Selection, for the final sample, of the ith judged unit in the ith set, for $i = 1, \ldots, k$.
4: Steps 1 to 3 can be replicated n times (n cycles) producing a sample of size nk.

We denote the RSS sample by $Y_{[i]j}$, for $i = 1, \ldots, k; j = 1, l \ldots, n$, where $Y_{[i]j}$ represents the observation ranked in the ith position in the jth cycle. In this case, the sample units are independent, but not identically distributed random variables, as a result of the ordering process. Furthermore, in the perfect ranking scenario $Y_{[i]\cdot}$ becomes to the ith order statistic from a SRS of size n, which is usually denoted by $Y_{(i)\cdot}$. In this work, however, we only use $Y_{[i]\cdot}$ for both perfect and imperfect ranking scenarios. When the results are specific to just one of the ranking scenarios, it will be emphasized in the text.

The usual estimator of the population mean using RSS is given by

$$\bar{Y}_{\text{RSS}} = \frac{1}{nk} \sum_{j=1}^{n} \sum_{i=1}^{k} Y_{[i]j},$$

with variance

$$\mathrm{Var}(\bar{Y}_{\mathrm{RSS}}) = \frac{\sigma^2}{nk} - \frac{1}{nk^2}\sum_{j=1}^{n}\sum_{i=1}^{k}(\mu_{[i]} - \mu)^2,$$

where μ and σ^2 are the population mean and variance and $\mu_{[i]} = \mathrm{E}[Y_{[i]j}]$.
The MRSS scheme is detailed in Algorithm 2.

Algorithm *2* MRSS scheme

1: Selection of k^2 units of the population using SRS, allocating them, randomly, into k sets of size k;
2: Ranking the sample units in each set according to the possible values of the variable of interest, using the pre-established ordering criterion;
3: For odd k, selection, for the final sample, of the $(k+1)/2$th judged unit in the each set. For even k, we must select the units judged in position $k/2$ in half of the sets and those judged in position $(k+2)/2$ in the remaining sets;
4: Steps 1 to 3 can be replicated n times (n cycles) producing a sample of size nk.

Next, we present the steps to drawn an ERSS sample in Algorithm 3.

Algorithm *3* ERSS scheme

1: Selection of k^2 units of the population using SRS, allocating them, randomly, into k sets of size k;
2: Ranking the sample units in each set according to the possible values of the variable of interest, using the pre-established ordering criterion;
3: For even k, selection, we must select, for the final sample, the units judged as the minimum in half of the sets and those judged as the maximum in the others. However, if k is odd we must select the units judged as the minimum in $(k-1)/2$ sets; those judged as the maximum in other $(k-1)/2$ sets, and the unit judges as the median (position $(k+1)/2$) in the final set;
4: Steps 1 to 3 can be replicated n times (n cycles) producing a sample of size nk.

Additionally, Zamanzade and Mahdizadeh (2019) proposed the RSS with extreme ranks, which is a more general sampling design including ERSS as a special case. Finally, NRSS scheme (Zamanzade and Al-Omari, 2016) consists of the steps described in Algorithm 4.

Algorithm *4* NRSS scheme

1: Selection of k^2 units of the population using SRS;
2: Ranking the k^2 sample units based on the pre-established ordering criterion;
3: Selection of the $[(i-1)k+l]$-th sample unit for the final sample, for $i = 1, \ldots, k$. If k is odd, then $l = (k+1)/2$; if k is even, then $l = (k+2)/2$ when i is odd and $l = k/2$ when i is even;
4: Again, steps 1-3 can be repeated n times, setting up n cycles and producing a final sample of size nk.

As previously stated, in NRSS the k^2 original sample units must compose (and must be ordered in) a single set, which induces dependence between the observations (differently from the RSS design). The variances of these variables, however, are reduced due to the greater set size, which justifies its higher efficiency. For the sake of illustration, to select a NRSS sample of size $k = 3$, we must select the sample units ranked in positions 2, 5 and

8 from a original ordered sample of size $k^2 = 9$; for a sample of size $k = 4$, we must select those ranked in positions 3, 6, 11 and 14 from a ordered sample of size $k^2 = 16$; and for a sample of size $k = 5$, the sample units ranked in positions 3, 8, 13, 18 and 23 must be selected from a ordered sample of size $k^2 = 25$. These are the sample sizes considered in this study. It is possible to observe that the positions of the selected sample units are, in general, regularly spaced.

The NRSS sample is denoted by $\{Y_{[(i-1)k+l]j}, i = 1, \ldots, k, j = 1, \ldots, n\}$, in which $Y_{[(i-1)k+l]j}$ refers to the unit ranked in position $[(i-1)k + l]$ (of an initial sample of size k^2), in the jth cycle. Under perfect ranking, particularly, $Y_{[(i-1)k+l]j}$ corresponds to the $((i-1)k + l)$th order statistics from a SRS sample of size k^2.

According to Zamanzade and Al-Omari (2016), the NRSS sample mean is an unbiased estimator for the population mean for symmetric distributions, which can be written by:

$$\bar{Y}_{\text{NRSS}} = \frac{1}{nk} \sum_{j=1}^{n} \sum_{i=1}^{k} Y_{[(i-1)k+l]j}, \tag{1}$$

and its variance is given by:

$$\text{Var}(\bar{Y}_{\text{NRSS}}) = \frac{1}{nk^2} \sum_{i=1}^{k} \text{Var}(Y_{[(i-1)k+l]}) + \frac{2}{nk^2} \sum_{1 \le i < i' \le k}^{k} \text{Cov}(Y_{[(i-1)k+l]}, Y_{[(i'-1)k+l]}). \tag{2}$$

3. Statistical quality control charts using NRSS

In this section, the Shewhart-type control chart based on NRSS is presented. Control charts for the process mean based on simple random samples of size k are defined by a central line (CL) and a pair of control limits (LCL and UCL) given by

$$\text{LCL} = \mu_0 - A\sqrt{\text{Var}(\bar{Y}_{\text{SRS}})} = \mu_0 - A\frac{\sigma_0}{\sqrt{k}},$$

$$\text{CL} = \mu_0,$$

$$\text{UCL} = \mu_0 + A\sqrt{\text{Var}(\bar{Y}_{\text{SRS}})} = \mu_0 + A\frac{\sigma_0}{\sqrt{k}},$$

where μ_0 and σ_0 are the in-control process mean and standard deviation, $\bar{Y}_{\text{SRS}}$ the mean of a simple random sample of k units and A the amplitude parameter of the control chart. An observed sample mean beyond the control limits is an indicator of an out-of-control process. It is usual to consider $A = 3$, which, under normal distribution, is associated to a probability of a false alarm (a point outside the control limits for an in-control process) of approximately 0.0027.

We consider control charts for the process mean using NRSS, based on the structure

$$\mathrm{LCL} = \mu_0 - \mathrm{A}\sqrt{\mathrm{Var}(\bar{\mathrm{Y}}_{\mathrm{NRSS}})},$$

$$\mathrm{CL} = \mu_0, \tag{3}$$

$$\mathrm{UCL} = \mu_0 + A\sqrt{\mathrm{Var}(\bar{Y}_{\mathrm{NRSS}})},$$

where $\bar{Y}_{\mathrm{NRSS}}$ and $\mathrm{Var}(\bar{Y}_{\mathrm{NRSS}})$ are defined in (1) and (2), respectively.

Our proposal constitutes an extension of the conventional SRS control charts, in such a way that the samples are periodically selected using NRSS and the control limits are based on (3). Alternatively, extensions of control charts were previously proposed for some other designs based on RSS. The performance of these control charts are used here as reference to NRSS control charts results.

In our study, to set the values for NRSS control limits, as described in (3), it was firstly necessary to get the values for $\mathrm{Var}(\bar{Y}_{\mathrm{NRSS}})$, for a process under statistical control, for each simulated scenario. Under perfect ranking, $Y_{[(i-1)k+l]}$ is equivalent to the $(i-1)k+l$ order statistic from a SRS sample of size k^2, for $i = 1, \ldots, k$. Thus, in this case we calculated $\mathrm{Var}(\bar{Y}_{\mathrm{NRSS}})$ as presented in (2), by using the properties of order statistics from the normal distribution, presented, for example, in Balakrishnan and Rao (1998).

Under imperfect ranking, due to the ranking errors, the sampling units no longer match to order statistics. In this case, we obtained the values for $\mathrm{Var}(\bar{Y}_{\mathrm{NRSS}})$ by means of a preliminary simulation study. So we simulated $B = 10^6$ NRSS samples from a bivariate normal distribution for different combinations of k and ρ (the correlation between the variable of interest and an auxiliary variable). Bivariate normal distribution is very usual in several industrial applications (Montgomery, 2009). Also, it is largely considered to evaluate the performance of control charts for RSS-based designs. Then, $\mathrm{Var}(Y_{[(i-1)k+l]})$ and $\mathrm{Cov}(Y_{[(i-1)k+l]}, Y_{[(i'-1)k+l]})$ are estimated, respectively, by

$$\mathrm{Var}(Y_{[(i-1)k+l]}) = \frac{\sum_{h=1}^{B}\left(Y_{[(i-1)k+l],h} - \bar{Y}_{[(i-1)k+l]}\right)^2}{B-1}, \quad i = 1, \ldots, k, \tag{4}$$

where

$$\bar{Y}_{[(i-1)k+l]} = \frac{\sum_{h=1}^{B} Y_{[(i-1)k+l],h}}{B},$$

and

$$\begin{aligned}\mathrm{Cov}(Y_{[(i-1)k+l]}, Y_{[(i'-1)k+l]}) = \frac{1}{B-1}\sum_{h=1}^{B}&\left(Y_{[(i-1)k+l],h} - \bar{Y}_{[(i-1)k+l]}\right)\\ &\times\left(Y_{[(i'-1)k+l],h} - \bar{Y}_{[(i'-1)k+l]}\right),\end{aligned} \tag{5}$$

for $1 \leq i < i' \leq k$. Then, we replace (4) and (5) in (2) to obtain the variances, and we used them to set the NRSS control limits under imperfect ranking.

In practice, the true process parameters are rarely (if ever) known. When they are unknown, it is usual to perform the statistical process control in two distinct stages: phase I and phase II (Chakraborti et al., 2008). Phase I consists in selecting a number of samples when the process operates in-control. Their sample units should then be used for estimating the process parameters and calculating the control limits. It is usually recommended the selection of 20-25 samples in phase I, aiming to accurately define the control limits; see Montgomery (2009). Once the control limits were calculated, in phase II the obtained control chart must be used to monitor the process, based on new samples selected over time.

When the process parameters are unknown, we propose the estimation of μ_0 and $\mathrm{Var}(\bar{Y}_{\mathrm{NRSS}})$ based on the results of m independent samples of size k selected from the process in the absence of assignable causes of variation (in-control process), according to

$$\bar{\bar{Y}}_{\mathrm{NRSS}} = \frac{1}{m}\sum_{p=1}^{m} \bar{Y}_{\mathrm{NRSSp}}$$

and

$$\widehat{\mathrm{Var}}\left(\bar{Y}_{\mathrm{NRSS}}\right) = \frac{1}{k^2}\sum_{i=1}^{k} \widehat{\mathrm{Var}}\left(Y_{[(i-1)k+l]}\right) + \frac{2}{k^2}\sum_{i<i'} \widehat{\mathrm{Cov}}\left(Y_{[(i-1)k+l]}, Y_{[(i'-1)k+l]}\right), \tag{6}$$

where

$$\widehat{\mathrm{Var}}\left(Y_{[(i-1)k+l]}\right) = \frac{1}{m-1}\sum_{p=1}^{m}\left(Y_{[(i-1)k+l]p} - \bar{Y}_{[(i-1)k+l]}\right)^2,$$

where $\bar{Y}_{[(i-1)k+l]} = (\sum_{p=1}^{m} Y_{[(i-1)k+l]p})/m$ and

$$\begin{aligned}\widehat{\mathrm{Cov}}(Y_{[(i-1)k+l]}, Y_{[(i'-1)k+l]}) = \frac{1}{m-1}\sum_{p=1}^{m}&[(Y_{[(i-1)k+l]p} - \bar{Y}_{[(i-1)k+l]}) \\ &\times(Y_{[(i'-1)k+l]p} - \bar{Y}_{[(i'-1)k+l]})], 1 \leq i < i' \leq k.\end{aligned}$$

Thus, in practice the NRSS control charts for the process mean with estimated control limits are defined by substituting, in (3), μ_0 by $\bar{\bar{Y}}_{\mathrm{NRSS}}$ and $\mathrm{Var}(\bar{Y}_{\mathrm{NRSS}})$ by $\widehat{\mathrm{Var}}\left(\bar{Y}_{\mathrm{NRSS}}\right)$

$$\begin{aligned}\mathrm{LCL} &= \bar{\bar{Y}}_{NRSS} - A\sqrt{\widehat{\mathrm{Var}}\left(\bar{Y}_{\mathrm{NRSS}}\right)}, \\ \mathrm{CL} &= \bar{\bar{Y}}_{\mathrm{NRSS}}, \\ \mathrm{UCL} &= \bar{\bar{Y}}_{\mathrm{NRSS}} + A\sqrt{\widehat{\mathrm{Var}}\left(\bar{Y}_{\mathrm{NRSS}}\right)}.\end{aligned}$$

In order to investigate the bias of (6) in estimating (2), an additional simulation study was carried out, considering $k = 3, 4, 5$. For each value of k, we simulated 5×10^4 replications of m samples, using NRSS, from a normal standard distribution. For m, values between 5 and 25 were set. At each step, the m simulated samples were considered to estimate $\mathrm{Var}(\bar{Y}_{\mathrm{NRSS}})$. We found that the bias of this estimator is negligible (a relative bias lower than 0.001 was verified for all sample sizes for $m \geq 20$).

4. Monte Carlo evaluation of NRSS-based control charts

In this section we present the run length properties for NRSS-based control charts, and for other RSS-based designs, obtained through a Monte Carlo simulation study. First, we evaluate their performance when the process follows the normal distribution. Thereafter, we analyze how NRSS-based control charts, and its competitors, were affected by different departures from normal distribution, considering models with different levels of skewness and kurtosis. For this purpose, we developed computational routines using `R` language. All simulations were performed using R software (R Core Team, 2019). The packages `MASS` (Venables and Ripley, 2002), `sn` (Azzalini, 2019), and `normalp` (Mineo, 2018) were used to generate samples from normal and non-normal distributions.

To evaluate the performance of NRSS control charts under normal distribution, we simulated samples from a bivariate normal distribution, according to

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ \mu_Y \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

where Y corresponds to the variable of interest and X was the concomitant variable. We assume $\mu_Y = \mu_0 = 0$ as the in-control process mean. The efficiency in ranking the sample units into each set was specified thorough ρ, such that higher levels of imperfect ranking was introduced by decreasing ρ. For the out-of-control scenarios, we consider

$$\mu_Y = \mu_0 + \frac{\delta \sigma_0}{\sqrt{k}},$$

such that δ determines the shift in the process mean:

$$\delta = |\mu_Y - \mu_0| \frac{\sqrt{k}}{\sigma_0}, \tag{7}$$

and $\delta = 0$ implies to an in-control process.

As parameters settings for the simulation study we had $k = 3, 4$ and 5; $\delta = 0, 0.1, 0.2, 0.3, 0.4, 0.8, 1.2, 1.6, 2, 2.4$ and 3.2 and $\rho = 0, 0.25, 0.50, 0.75, 0.9$ and 1. To evaluate the performance of control charts we consider the average run length (ARL), defined as the average number of points in a control chart until one exceeds the control limits. Particularly, if we have an in-control process, ARL_0 is the reciprocal of the false alarm error rate; for an out-of-control process, ARL_1 is inversely proportional to the detection probability, representing the average number of samples until the out-of-control state is detected. For each combination of k, and δ we simulated 10^6 independent NRSS samples under perfect ranking and 10^7 under imperfect ranking, for each considered correlation level (ρ value). At this stage, we have to increase the number of simulations, due to some numerical instability in estimating the run length properties. Based on results provided by a previous convergence study (results were not showed), we noticed some additional instability when considering imperfect ranking. This study pointed out that the adopted simulation sizes were satisfactory to achieve satisfactory convergence. The ARL values were calculated as the inverse of the proportion of points (sample means) beyond the control limits. In addition, the simulation results were also summarized by means of standard deviation of the run length (SDRL) and median run length (MRL), since the run length distribution is quite skewed.

The parameters for the simulation study were chosen in such a way to allow the comparison of the ARL values with those presented in other publications, referring to control charts for other sampling designs based on RSS. Moreover, it becomes evident that the considered scenarios (198 in total) comprises a great variety of processes. The sample size was limited to $k = 5$ given the context for application of sampling designs based on RSS (restrictions related to draw big samples, initial selection and ranking of k^2 - or even k^3 or more - sample units, among others). Moreover, the amplitude parameter (A) for the control limits were set, under perfect ranking, so that $\mathrm{ARL}_0 = 370.51$. This is the ARL_0 corresponding to SRS control charts when we set $A = 3$. In this way, we could fairly compare the ARL_1 values for NRSS control charts with those provided by the other sampling designs. The DRSS designs control charts, particularly, produce low values for ARL_0 and, consequently, high false alarms rates when $A = 3$.

Table 1 presents the simulated run length results for RSS-based control charts. Besides NRSS, results obtained by SRS, RSS, ERSS and MRSS are also presented. In this first part of the analysis, we consider perfect ranking ($\rho = 1$), allowing to assess the maximum power provided by each design. Some conclusions drawn from Table 1 are the following:

- The efficiency of NRSS control charts for detecting shifts in process mean increases, as expected, for higher values of δ and k. As an illustration, for $k = 3$ and $\delta = 0.40$ we have ARL = 120.60 compared to ARL = 6.41 for $\delta = 1.20$, while for $k = 5$ and $\delta = 0.40$ we have ARL = 102.60 for $k = 3$ against ARL = 60.14 for $k = 5$;
- The NRSS control charts perform better than SRS control charts in all simulated scenarios. For example, for $k = 3$ and $\delta = 0.80$ we have ARL = 21.25 for NRSS control charts compared to ARL = 71.55 for SRS, while for $k = 5$ and $\delta = 1.60$ we have ARL = 1.46 for NRSS against ARL = 12.38 for SRS;
- The NRSS control charts dominates RSS and ERSS designs in all the simulated scenarios. For example, when compared to RSS, for $k = 3$ and $\delta = 0.80$ we have ARL = 21.25 for NRSS control charts against ARL = 35.43 for RSS, while for $k = 5$ and $\delta = 1.60$ we have ARL = 1.46 for NRSS against ARL = 2.83 for RSS;
- The NRSS control charts overcome the MRSS competitor in all simulated scenarios. This is remarkable, once MRSS is well known by its higher efficiency in estimating the mean, compared to RSS, for symmetric distributions. Additionally, MRSS performs best under both single and DRSS strategies for control charts for the process mean (Mehmood et al., 2013). When $k = 3$ and $\delta = 0.80$ it was verified ARL = 21.25 for NRSS control charts compared to ARL = 29.52 for MRSS, while when $k = 5$ and $\delta = 1.60$ we have ARL = 1.46 for NRSS against ARL = 2.04 for MRSS.

In order to summarize the performance of the different control charts designs, Figure 1 presents the geometric means of the ratios of ARL values for SRS control charts relative to the ones obtained by each of the other sampling designs, for each sample size. The ARL values for SRS control charts were, on average, 2.39 times larger than the corresponding NRSS when $k = 3$; 3 times for $k = 4$ and 3.59 times for $k = 5$. The best performance of NRSS control charts over the RSS, ERSS and MRSS counterparts becomes evident. For MRSS, for example, we have, on average, ARL 1.22 times higher than NRSS for $k = 3$; 1.25 times for $k = 4$ and 1.28 times for $k = 5$.

Table 1. ARL, MRL and SDRL for control charts constructed by SRS and designs based on RSS under perfect ranking

		SRS			RSS			ERSS			MRSS			NRSS		
k	δ	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL
	0.00	370.40	257	369.90	370.51	257	370.01	370.51	257	370.01	370.51	257	370.01	370.51	257	370.01
	0.10	352.93	245	352.43	333.89	232	333.39	340.25	236	339.75	339.67	236	339.17	325.63	226	325.13
	0.20	308.43	214	307.93	266.03	185	265.53	272.18	189	271.68	265.11	184	264.61	234.03	162	233.53
	0.30	253.14	176	252.64	196.93	137	196.43	197.20	137	196.70	186.22	129	185.72	157.23	109	156.73
	0.40	200.08	139	199.58	139.43	97	138.93	137.99	96	137.49	128.12	89	127.62	102.60	71	102.10
3	0.80	71.55	50	71.05	35.43	25	34.93	35.35	25	34.85	29.52	21	29.02	21.25	15	20.74
	1.20	27.82	19	27.32	11.54	8	11.03	11.43	8	10.92	9.22	7	8.71	6.41	5	5.89
	1.60	12.38	9	11.87	4.76	3	4.23	4.75	3	4.22	3.80	3	3.26	2.76	2	2.20
	2.00	6.30	5	5.78	2.50	2	1.94	2.49	2	1.93	2.06	2	1.48	1.61	1	0.99
	2.40	3.65	3	3.11	1.61	1	0.99	1.61	1	0.99	1.40	1	0.75	1.20	1	0.49
	3.20	1.73	1	1.12	1.09	1	0.31	1.09	1	0.31	1.04	1	0.20	1.01	1	0.10
	0.00	370.40	257	369.90	370.51	257	370.01	370.51	257	370.01	370.51	257	370.01	370.51	257	370.01
	0.10	352.93	245	352.43	328.08	228	327.58	341.30	237	340.80	318.07	221	317.57	310.56	215	310.06
	0.20	308.43	214	307.93	249.81	173	249.31	266.81	185	266.31	232.45	161	231.95	210.30	146	209.80
	0.30	253.14	176	252.64	174.89	121	174.39	192.64	134	192.14	156.42	109	155.92	126.90	88	126.40
	0.40	200.08	139	199.58	119.36	83	118.86	135.85	94	135.35	100.29	70	99.79	77.86	54	77.36
4	0.80	71.55	50	71.05	27.78	19	27.28	33.69	24	33.19	21.42	15	20.91	13.89	10	13.38
	1.20	27.82	19	27.32	8.54	6	8.02	10.70	8	10.19	6.38	5	5.86	4.09	3	3.56
	1.60	12.38	9	11.87	3.55	3	3.01	4.41	3	3.88	2.73	2	2.17	1.89	1	1.30
	2.00	6.30	5	5.78	1.94	1	1.35	2.33	2	1.76	1.59	1	0.97	1.25	1	0.56
	2.40	3.65	3	3.11	1.35	1	0.69	1.53	1	0.90	1.19	1	0.48	1.06	1	0.25
	3.20	1.73	1	1.12	1.03	1	0.18	1.07	1	0.27	1.01	1	0.10	1.00	1	0.00
	0.00	370.40	257	369.90	370.51	257	370.01	370.51	257	370.01	370.51	257	370.01	370.51	257	370.01
	0.10	352.93	245	352.43	331.68	230	331.18	333.00	231	332.50	329.60	229	329.10	299.58	208	299.08
	0.20	308.43	214	307.93	244.98	170	244.48	254.77	177	254.27	223.41	155	222.91	181.06	126	180.56
	0.30	253.14	176	252.64	165.54	115	165.04	173.73	121	173.23	136.91	95	136.41	104.59	73	104.09
	0.40	200.08	139	199.58	107.88	75	107.38	117.44	82	116.94	85.20	59	84.70	60.14	42	59.64
5	0.80	71.55	50	71.05	22.53	16	22.02	26.59	19	26.09	15.56	11	15.05	9.55	7	9.04
	1.20	27.82	19	27.32	6.73	5	6.21	8.13	6	7.61	4.55	3	4.02	2.86	2	2.31
	1.60	12.38	9	11.87	2.83	2	2.28	3.38	2	2.84	2.04	2	1.46	1.46	1	0.82
	2.00	6.30	5	5.78	1.63	1	1.01	1.87	1	1.28	1.31	1	0.64	1.10	1	0.33
	2.40	3.65	3	3.11	1.21	1	0.50	1.32	1	0.65	1.08	1	0.29	1.01	1	0.10
	3.20	1.73	1	1.12	1.01	1	0.10	1.03	1	0.18	1.00	1	0.00	1.00	1	0.00

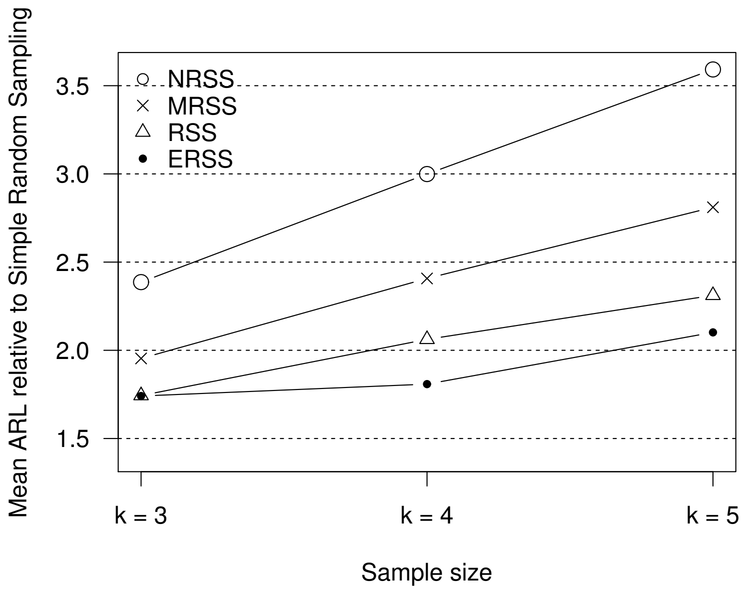

Figure 1. Average relative efficiency from control charts of designs based on RSS compared to SRS under perfect ranking. ARL from RSS, MRSS and ERSS were taken from Al-Omari and Haq (2012).

Table 2 presents the simulation results under imperfect ranking by setting $A = 3$ (3-sigma limits). This is a traditional choice for Shewhart control charts. The ARL values for $\rho = 0$ are identical to the corresponding ones from SRS, once NRSS and SRS are equivalent if the ordering is done completely at random. Based on these results, it is possible to assess the impact of ranking errors in the performance of control charts. Some conclusions from Table 2 are highlighted next:

- Control charts for all RSS based designs lose performance when the correlation between the variables decreases. For example, for NRSS control charts, $k = 3$ and $\delta = 0.8$, ARL = 21.34 when $\rho = 1$, ARL = 31.23 when $\rho = 0.90$; 44.02 when rho $\rho = 0.75$ and 59.55 when $\rho = 0.50$;
- The ARL values for NRSS control charts are smaller compared to the ones provided by SRS in almost all simulated scenarios with $\delta \neq 0$. NRSS only loses in a few scenarios described by low shifts in process mean and low values for ρ;
- The ARL_0 values from NRSS control charts are around 370.4, as intended. The individual ARL_0 values range from 365.62 when $k = 3$ and $\rho = 0.75$, to 372.04, when $k = 5$ and $\rho = 1$.

Figure 2 shows the geometric means of the ratios of ARL values for SRS control charts relative to the ones obtained by each of the RSS based designs. These results are presented for each sample size and considering the different correlation levels between the auxiliary and the variable of interest. We can notice that NRSS control charts are, in general, more efficient than all other considered sampling designs. Moreover, the superiority of NRSS control charts becomes higher when the correlation between the variables increases. For $\rho = 0.9$ and $\rho = 1$, we have, on average, higher efficiency for the NRSS control charts with $k = 4$ than for the other sampling designs taking $k = 5$, which can reflect in resource savings and lower operational effort.

Table 2. ARL, MRL and SDRL for control charts constructed by NRSS under imperfect ranking

		$\rho = 0$			$\rho = 0.25$			$\rho = 0.50$			$\rho = 0.75$			$\rho = 0.90$			$\rho = 1.00$		
k	δ	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL
	0.00	370.40	257	369.90	369.11	256	368.61	371.24	257	370.74	365.62	254	365.12	369.37	256	368.87	369.15	256	368.65
	0.10	352.93	245	352.43	349.32	242	348.82	349.84	243	349.34	343.41	238	342.91	335.38	233	334.88	322.07	223	321.57
	0.20	308.43	214	307.93	307.94	214	307.44	297.73	207	297.23	279.09	194	278.59	259.64	180	259.14	233.58	162	233.08
	0.30	253.14	176	252.64	248.44	172	247.94	237.54	165	237.04	211.75	147	211.25	183.29	127	182.79	155.70	108	155.20
	0.40	200.08	139	199.58	195.30	136	194.80	182.48	127	181.98	156.13	108	155.63	127.46	89	126.96	101.62	71	101.12
3	0.80	71.55	50	71.05	68.42	48	67.92	59.55	41	59.05	44.02	31	43.52	31.23	22	30.73	21.34	15	20.83
	1.20	27.82	19	27.32	26.29	18	25.79	22.04	15	21.53	15.07	11	14.56	9.92	7	9.41	6.44	5	5.92
	1.60	12.83	9	12.32	11.67	8	11.16	9.55	7	9.04	6.34	5	5.82	4.15	3	3.62	2.76	2	2.20
	2.00	6.30	5	5.78	5.92	4	5.40	4.84	3	4.31	3.26	2	2.71	2.23	2	1.66	1.61	1	0.99
	2.40	3.65	3	3.11	3.43	3	2.89	2.85	2	2.30	2.01	2	1.42	1.49	1	0.85	1.20	1	0.49
	3.20	1.73	1	1.12	1.65	1	1.04	1.45	1	0.81	1.19	1	0.48	1.06	1	0.25	1.01	1	0.10
	0.00	370.40	257	369.90	369.89	257	369.39	370.89	257	370.39	371.00	257	370.50	368.05	255	367.55	371.43	258	370.93
	0.10	352.93	245	352.43	355.25	246	354.75	345.83	240	345.33	341.05	237	340.55	333.70	231	333.20	310.75	216	310.25
	0.20	308.43	214	307.93	310.98	216	310.48	296.94	206	296.44	274.31	190	273.81	244.20	169	243.70	208.73	145	208.23
	0.30	253.14	176	252.64	246.69	171	246.19	237.34	165	236.84	206.25	143	205.75	167.96	117	167.46	127.24	88	126.74
	0.40	200.08	139	199.58	193.12	134	192.62	178.81	124	178.31	147.41	102	146.91	112.07	78	111.57	76.67	53	76.17
4	0.80	71.55	50	71.05	68.61	48	68.11	58.03	40	57.53	39.93	28	39.43	24.92	17	24.41	13.91	10	13.40
	1.20	27.82	19	27.32	26.15	18	25.65	21.15	15	20.64	13.28	9	12.77	7.66	5	7.14	4.10	3	3.57
	1.60	12.83	9	12.32	11.53	8	11.02	9.16	6	8.65	5.57	4	5.05	3.23	2	2.68	1.89	1	1.30
	2.00	6.30	5	5.78	5.86	4	5.34	4.64	3	4.11	2.89	2	2.34	1.82	1	1.22	1.25	1	0.56
	2.40	3.65	3	3.11	3.41	2	2.87	2.74	2	2.18	1.82	1	1.22	1.29	1	0.61	1.06	1	0.25
	3.20	1.73	1	1.12	1.64	1	1.02	1.42	1	0.77	1.14	1	0.40	1.02	1	0.14	1.00	1	0.00
	0.00	370.40	257	369.90	367.97	255	367.47	368.85	256	368.35	369.72	256	369.22	369.33	256	368.83	372.04	258	371.54
	0.10	352.93	245	352.43	354.70	246	354.20	345.42	240	344.92	341.92	237	341.42	328.10	228	327.60	296.14	205	295.64
	0.20	308.43	214	307.93	308.61	214	308.11	298.38	207	297.88	274.64	191	274.14	235.35	163	234.85	182.99	127	182.49
	0.30	253.14	176	252.64	249.19	173	248.69	234.05	162	233.55	197.91	137	197.41	155.22	108	154.72	104.91	73	104.41
	0.40	200.08	139	199.58	193.77	134	193.27	177.52	123	177.02	141.57	98	141.07	100.78	70	100.28	60.11	42	59.61
5	0.80	71.55	50	71.05	67.86	47	67.36	56.81	40	56.31	37.11	26	36.61	21.01	15	20.50	9.65	7	9.14
	1.20	27.82	19	27.32	26.02	18	25.52	20.68	14	20.17	12.20	9	11.69	6.34	5	5.82	2.88	2	2.33
	1.60	12.83	9	12.32	11.48	8	10.97	8.88	6	8.37	5.10	4	4.57	2.72	2	2.16	1.46	1	0.82
	2.00	6.30	5	5.78	5.83	4	5.31	4.52	3	3.99	2.67	2	2.11	1.59	1	0.97	1.10	1	0.33
	2.40	3.65	3	3.11	3.39	2	2.85	2.67	2	2.11	1.71	1	1.10	1.19	1	0.48	1.01	1	0.10
	3.20	1.73	1	1.12	1.64	1	1.02	1.40	1	0.75	1.11	1	0.35	1.01	1	0.10	1.00	1	0.00

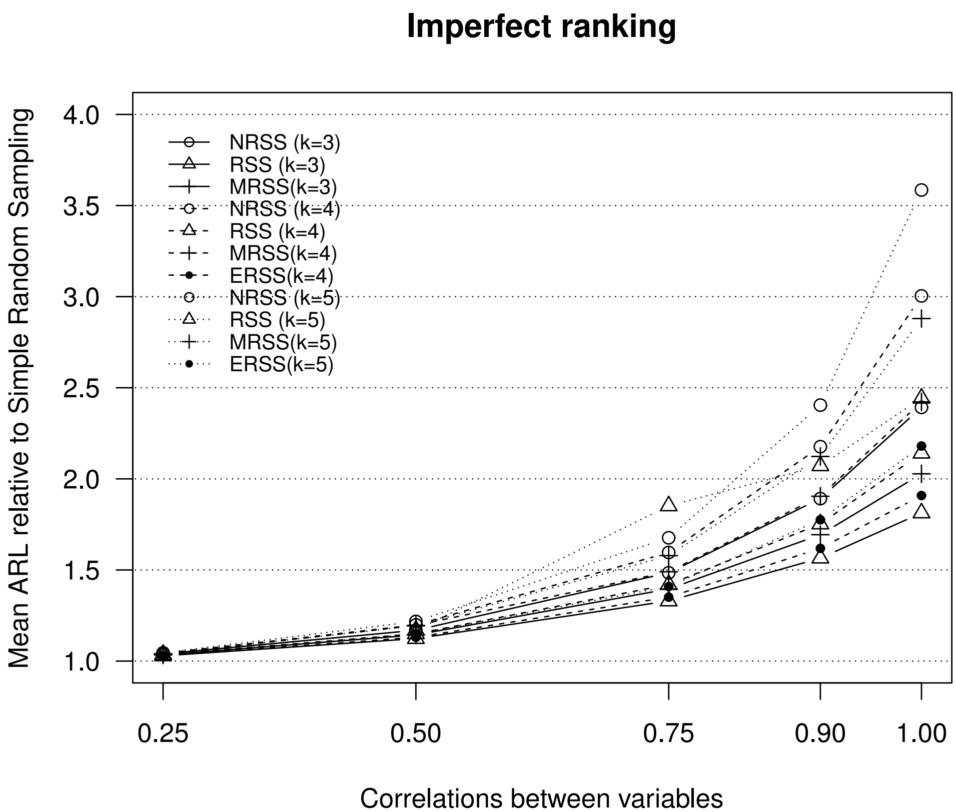

Figure 2. Average relative efficiency from control charts of designs based on RSS compared to SRS under imperfect ranking. ARL from RSS, MRSS and ERSS were taken from Al-Omari and Haq (2012). For $k = 3$ RSS and ERSS provides the same sampling design.

In order to evaluate the effect of non normality on the performance of NRSS control charts, a new simulation study was conducted. Two probability distributions are considered at this point: the skew normal and the generalized normal distributions (Azzalini, 1985; Nadarajah, 2005). Through these models, we were able to evaluate the impact of different levels of skewness and kurtosis on the run length results. The skew normal and the generalized normal models are briefly described in the following paragraphs.

The probability density function of a random variable with skew normal distribution is given by

$$f(y; \epsilon, \omega, \alpha) = \frac{2}{\omega}\phi\left(\frac{y-\epsilon}{\omega}\right)\Phi\left(\alpha\left(\frac{y-\epsilon}{\omega}\right)\right),$$

where $y \in (-\infty, \infty)$ and $\epsilon \in (-\infty, \infty)$, $\omega > 0$ and $\alpha \in (-\infty, \infty)$ are location, scale and shape parameters, respectively. Additionally, ϕ and Φ represent the probability density function and the cumulative distribution function of the standard normal distribution. The skew normal distribution becomes more asymmetric as $|\alpha|$ increases. When $\alpha > 0$, the distribution is right skewed; left skewed if $\alpha < 0$ and for $\alpha = 0$ we have the normal distribution.

A random variable has generalized normal distribution if its probability density function is given by

$$f(y;\mu,\beta,\alpha)=\frac{1}{2\,\alpha^{1/\alpha}\,\Gamma(1+1/\alpha)\,\beta}\,\mathrm{e}^{-\frac{|y-\mu|^{\alpha}}{\alpha\beta^{\alpha}}},$$

where $y\in(-\infty,\infty)$ and $\mu\in(-\infty,\infty)$, $\beta>0$ and $\alpha>0$ are location, scale and shape parameters, respectively. The generalized normal distribution is symmetric around μ and becomes the normal distribution when $\alpha=2$. In addition, for $\alpha<2$ it produces leptokurtic (fatter tails) distributions, and platykutics (thinner tails) distributions when $\alpha>2$. As particular cases of the generalized normal distribution we have, for example, the Laplace ($\alpha=1$) and uniform ($\alpha\to\infty$) distributions.

We consider four different parameter combinations for each one of the two distributions. For the skew normal model, an increasing sequence of values for α was defined ($\alpha=1,2,3$ and 5), providing distributions with different levels of skewness. Additionally, we set $\omega=1$ and, for ϵ, we have assigned appropriate values such that the process mean was equal to zero. For the generalized normal model, four different values for α were selected, producing two distributions with heavy tails (for $\alpha=1$ and 1.5) and two with light tails (for $\alpha=3$ and 4). Furthermore, for the other model parameters we set $\mu=0$ and $\beta=1$. In all cases, the process mean was set at zero since the objective here is to evaluate the robustness of the control charts in maintaining the average (and median) run length for an in-control process ($\mathrm{ARL}_0=370.4$). Also, for the sake of brevity we are only considering, at this point, the perfect ranking scenario.

For each one of the eight distributions obtained by combining the two distributions and four specific parameter settings, we have simulated 10^7 samples of sizes $k=3,4,$ and 5. Five different sampling designs are considered: NRSS, RSS, MRSS, ERSS and SRS. 3-sigma control limits were properly calculated as described in (3), for the NRSS control charts, and based on the expressions presented in Muttlak and Al-Sabah (2003), for the others. Based on the simulated results, we calculated the corresponding values for ARL, MRL and SDRL, as we can verify in Table 3.

Note in Table 3 that, although all considered sample designs have their respective ARL$'s$ affected by the distribution skewness, NRSS and MRSS provided, in general, the closest values to the nominal $\mathrm{ARL}_0=370.4$ for the skew normal distribution. This indicates that these sampling designs are more conservative than their competitors. Table 3 points higher influence in ARL and MRL for the generalized normal if compared with the skew normal distribution in the considered simulated scenarios. This is particularly evident for $\beta=1$ (Laplace distribution). However, we can also see that the NRSS control charts still dominates all its competitors, producing, in general, ARL_0 values closer to 370.4. Our results are in agreement with those found by Koyuncu and Karagöz (2018), who verified that NRSS control charts present lower type I error when applied to two asymmetric distributions: Type I Marshall-Olkin bivariate Weibull and bivariate lognormal.

Table 3. ARL, MRL and SDRL for control charts constructed by SRS and designs based on RSS under perfect ranking and Non-normal data

			SRS			RSS			MRSS			ERSS			NRSS		
Distribution	k	α	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL	ARL	MRL	SDRL
SN	3	1	351.19	244	350.69	325.09	225	324.59	351.62	244	351.12	325.09	225	324.59	361.73	251	361.23
		2	281.61	195	281.11	285.26	198	284.76	346.92	241	346.42	285.26	198	284.76	348.44	242	347.94
		3	237.08	164	236.58	254.53	177	254.03	352.77	245	352.27	254.53	177	254.03	336.65	234	336.15
		5	203.99	142	203.49	232.87	162	232.37	377.92	262	377.42	232.87	162	232.37	330.02	229	329.52
	4	1	352.94	245	352.44	335.09	232	334.59	356.46	247	355.96	321.09	223	320.59	366.05	254	365.55
		2	298.32	207	297.82	300.11	208	299.61	352.44	244	351.94	237.39	165	236.89	364.65	253	364.15
		3	259.91	180	259.41	277.96	193	277.46	380.36	264	379.86	186.96	130	186.46	378.98	263	378.48
		5	227.08	158	226.58	258.42	179	257.92	423.91	294	423.41	149.55	104	149.05	391.14	271	390.64
	5	1	357.67	248	357.17	339.87	236	339.37	357.17	248	356.67	320.18	222	319.68	368.16	255	367.66
		2	310.86	216	310.36	306.96	213	306.46	340.49	236	339.99	241.84	168	241.34	364.61	253	364.11
		3	274.15	190	273.65	292.93	203	292.43	325.44	226	324.94	191.77	133	191.27	371.25	257	370.75
		5	246.36	171	245.86	274.54	190	274.04	314.66	218	314.16	152.81	106	152.31	383.72	266	383.22
GN	3	1	126.21	88	125.71	118.37	82	117.87	150.16	104	149.66	118.37	82	117.87	189.11	131	188.61
		1.5	237.37	165	236.87	232.05	161	231.55	254.48	177	253.98	232.05	161	231.55	297.96	207	297.46
		3	730.63	507	730.13	466.32	323	465.82	558.22	387	557.72	466.32	323	465.82	429.00	298	428.50
		4	1207.28	837	1206.78	520.61	361	520.11	743.17	515	742.67	520.61	361	520.11	457.01	317	456.51
	4	1	147.39	102	146.89	133.05	92	132.55	176.31	122	175.81	143.60	100	143.10	247.53	172	247.03
		1.5	257.63	179	257.13	251.53	175	251.03	280.31	194	279.81	251.46	174	250.96	328.49	228	327.99
		3	582.92	404	582.42	419.56	291	419.06	465.33	323	464.83	385.33	267	384.83	392.97	273	392.47
		4	798.23	553	797.73	437.79	304	437.29	527.04	365	526.54	371.91	258	371.41	396.34	275	395.84
	5	1	163.59	114	163.09	146.81	102	146.31	211.19	147	210.69	155.41	108	154.91	275.21	191	274.71
		1.5	273.91	190	273.41	266.41	185	265.91	298.77	207	298.27	262.38	182	261.88	345.10	239	344.60
		3	521.44	362	520.94	399.11	277	398.61	441.48	306	440.98	375.56	260	375.06	385.42	267	384.92
		4	635.37	441	634.87	413.40	287	412.90	489.72	340	489.22	377.68	262	377.18	393.67	273	393.17

5. An application to real data

In order to illustrate the application of the NRSS control charts, we used a data set with 1030 observations about the concrete strength to compression (MPa) and the amount of cement (kg) used in the production of concrete blocks (Yeh, 1998). This data set is available in the R package `AppliedPredictiveModeling` (Kuhn and Johnson, 2018). Although this data was not recorded as a case of a quality control process, it serves us, under some assumptions, as a reference population, from which samples were drawn and control charts were constructed. We assumed the concrete strength as the variable of interest and the amount of cement as an auxiliary variable, such that the sample units may be ordered with errors, producing an imperfect ranking scenario. Also, we consider an additional scenario based on perfect ranking. In this case, the sample units were ordered directly from the concrete strength values, and the ranking process did not present any error. Moreover, we assumed the concrete blocks strength distribution in this sample as the natural variability of an industrial process. A square root transformation of the concrete strength was used in order to obtain a better approximation to normal distribution.

In this application, we consider three sampling designs: SRS, RSS and NRSS; two sample sizes: $k = 3$ and $k = 5$, and processes in two different scenarios: in-control ($\delta = 0$) and out-of-control, considering $\delta = 1.2$, as described in (7). Under each sampling design and for each sample size, we selected, with replacement, 25 samples from the original data. These samples are considered for estimating the control limits with A = 3, which corresponds to a probability of a type I error of $\alpha = 0.0027$ (phase 1). Afterwards, 75 new samples were selected for monitoring the process mean (phase 2). For $\delta = 0$, these 75 samples were selected with replacement from the original data; for $\delta = 1.2$, we added to the transformed strength values a normal random variable with mean $1.2\sigma_0/\sqrt{k}$ and standard deviation equals to 0.17 (corresponding to 11.74% of the standard deviation of the transformed concrete strength). This standard deviation value is small enough to characterize the lack of control, predominantly, due to the shift in the process mean, instead of its dispersion (variance).

Figure 3 presents (on the left) the histogram for the distribution of concrete strength, with the estimated normal distribution and kernel density curves. The dispersion plot, on the right, indicates moderate positive linear relationship between the variables. The linear correlation coefficient is $\rho = 0.49$, which points to a moderately favourable scenario for RSS based designs.

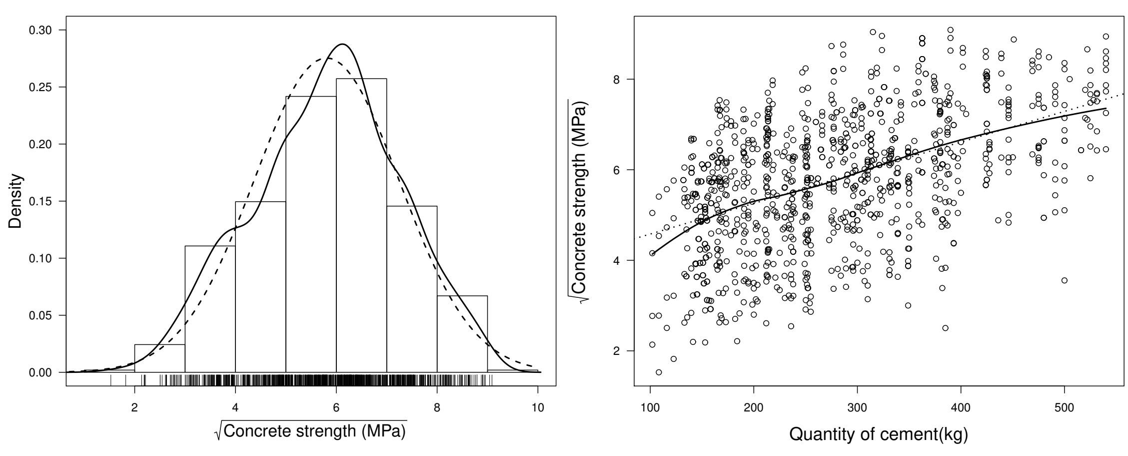

Figure 3. Histogram and scatter plot for concrete strength.

Following, Figures 4 and 5 present the SRS, RSS and NRSS control charts for the process mean considering $k = 3$. The RSS and NRSS control charts were obtained under perfect and imperfect ranking, as previously described. In Figure 4 we have the charts when $\delta = 0$ (in-control process). In all cases, it is possible to notice points randomly distributed around the central line, without any point outside the control limits. This behaviour characterizes an in-control process, as expected. In addition, Figure 5 presents the control charts for $\delta = 1.2$ (out-of-control process). It is possible to observe that the NRSS control chart showed the highest number of points exceeding the control limits (11 points under perfect ranking and 7 under imperfect ranking), followed by RSS (with 9 and 5 points exceeding the control limits, respectively) and SRS control charts (only 2 points outside the limits).

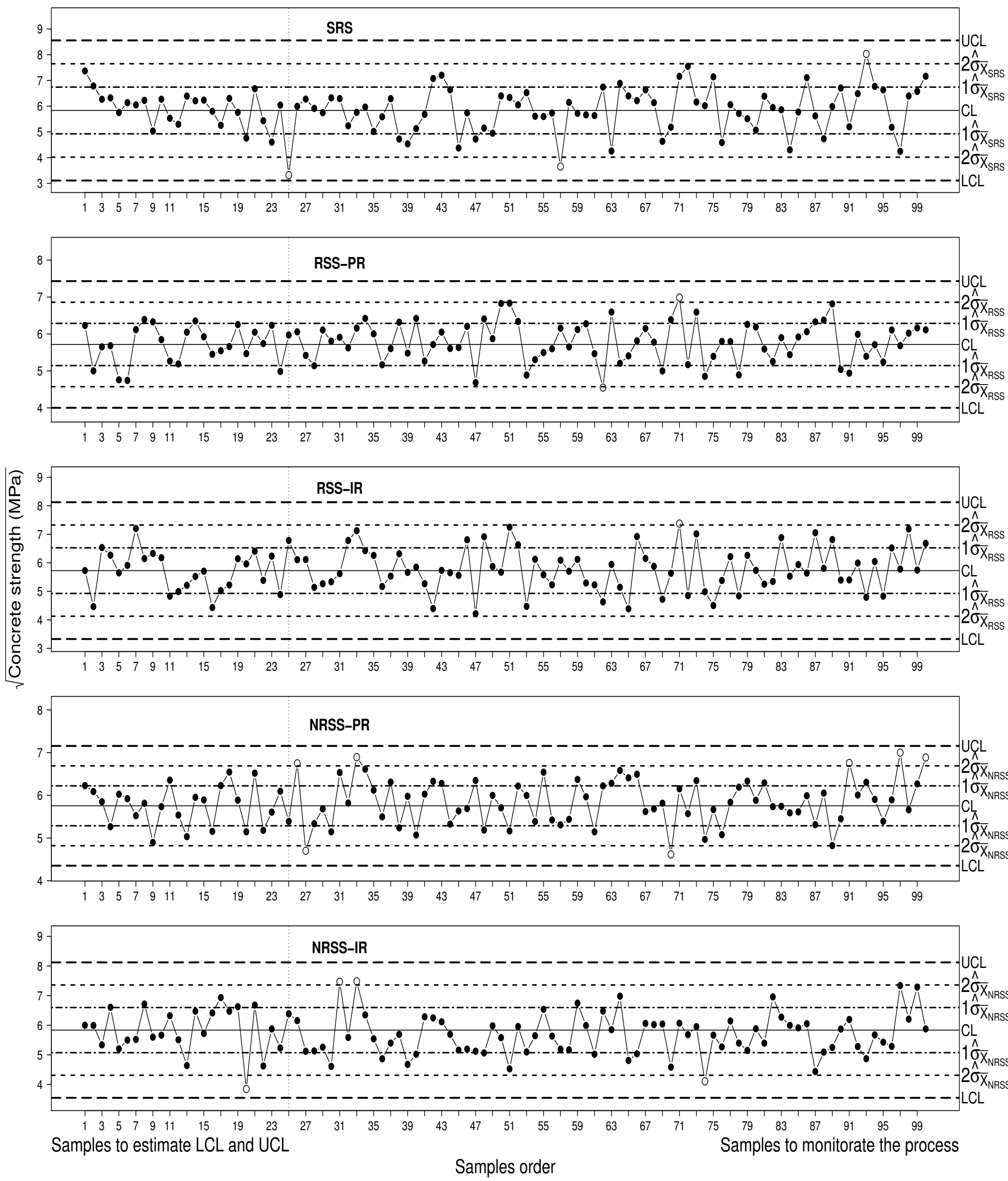

Figure 4. Control charts for concrete strength considering $k = 3$ and an in-control process ($\delta = 0$). Perfect ranking is denoted as PR, and imperfect ranking as IR.

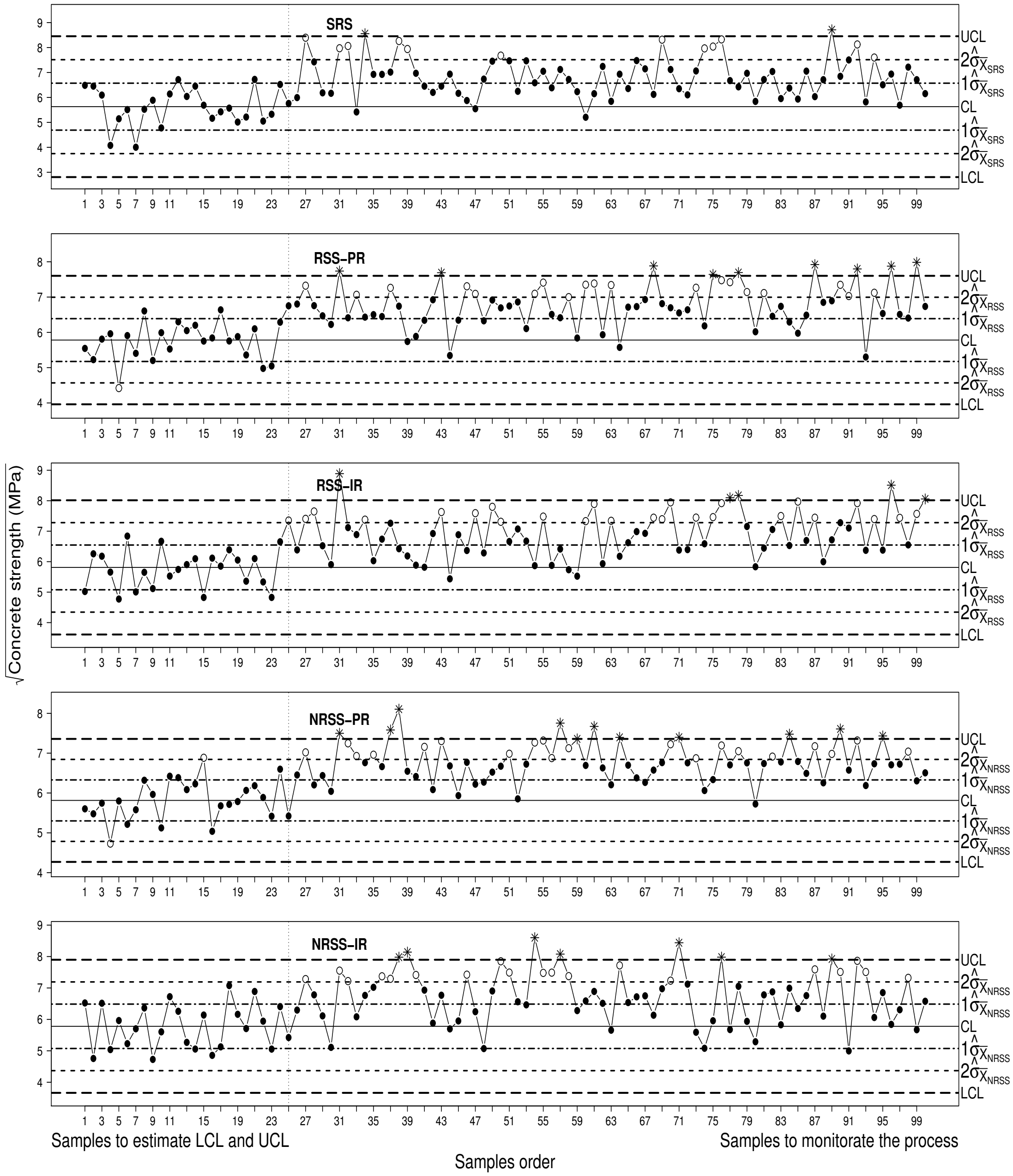

Figure 5. Control charts for concrete strength considering $k = 3$ and an out-of-control process ($\delta = 1.2$). Perfect ranking is denoted as PR, and imperfect ranking as IR.

Figures 6 and 7 present the control charts for $k = 5$, under the same three sampling designs (and five scenarios, when considering perfect and imperfect ranking), simulated, respectively, with $\delta = 0$ and $\delta = 1.2$. It is possible to notice again that NRSS control charts present satisfactory performance, showing randomness and without any point outside the control limits for an in-control process, and also presenting a large number of points exceeding the control limits in the out-of-control scenario (28 under perfect and 9 under imperfect ranking) than RSS (14 and 6 points, respectively) and SRS (with only 3 points outside the limits).

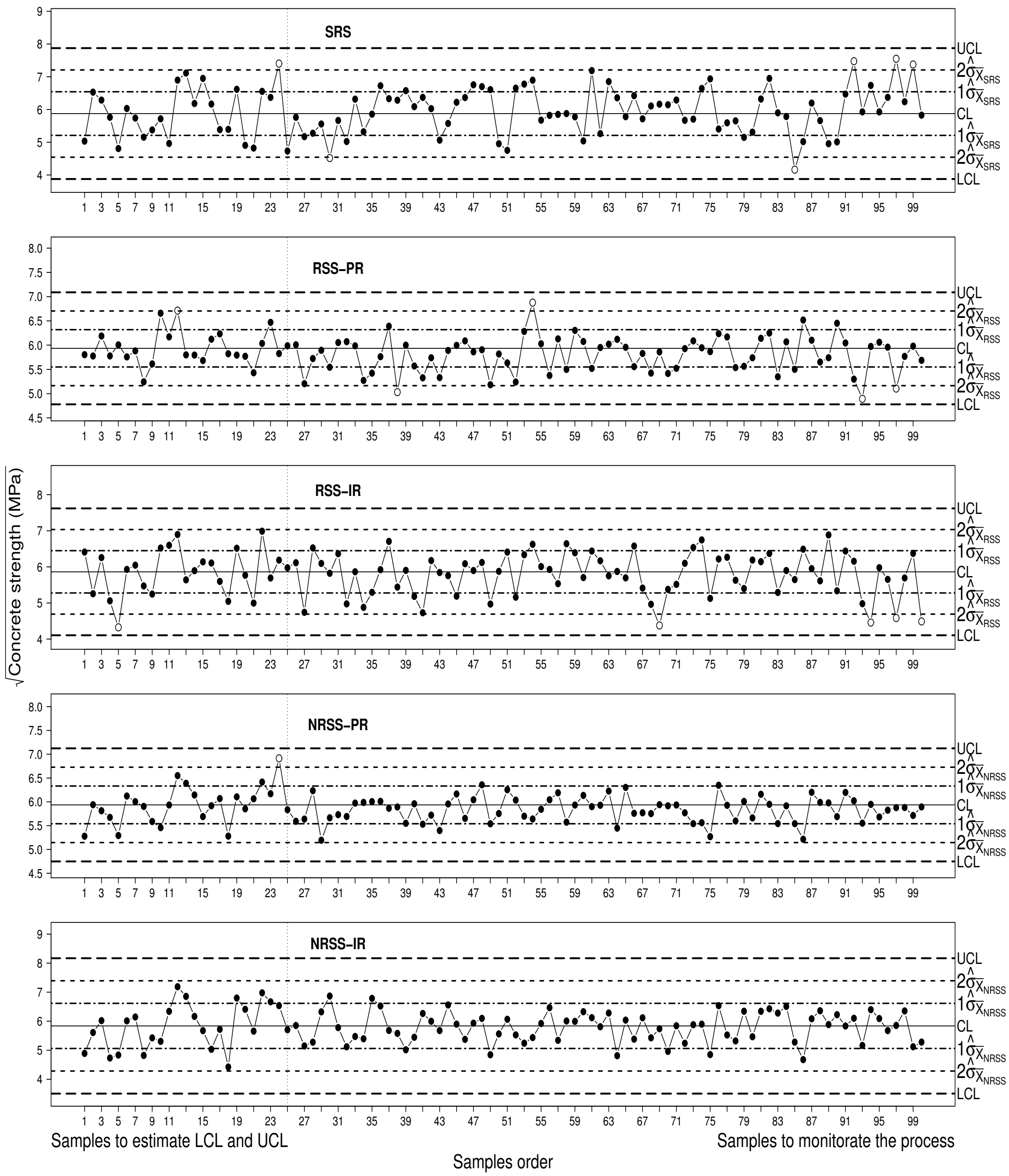

Figure 6. Control charts for concrete strength considering $k = 5$ and an in-control process ($\delta = 0$). Perfect ranking is denoted as PR, and imperfect ranking as IR.

6. Concluding remarks

In this paper, we considered control charts for the mean of a normal distributed process based on NRSS design. These charts were compared to their SRS and RSS based counterparts by means of a simulation study. Under perfect ranking, NRSS control charts overcome all their competitors, providing smaller ARL values for out-of-control process in all simulated scenarios. In addition, the NRSS control charts showed to be competitive when compared to those based on DRSS designs. However, such sampling designs require the initial selection of k^3 sample units for, after two ordering cycles, selecting a final sample of k units. For example, the ARL for NRSS control charts were smaller in all simulated

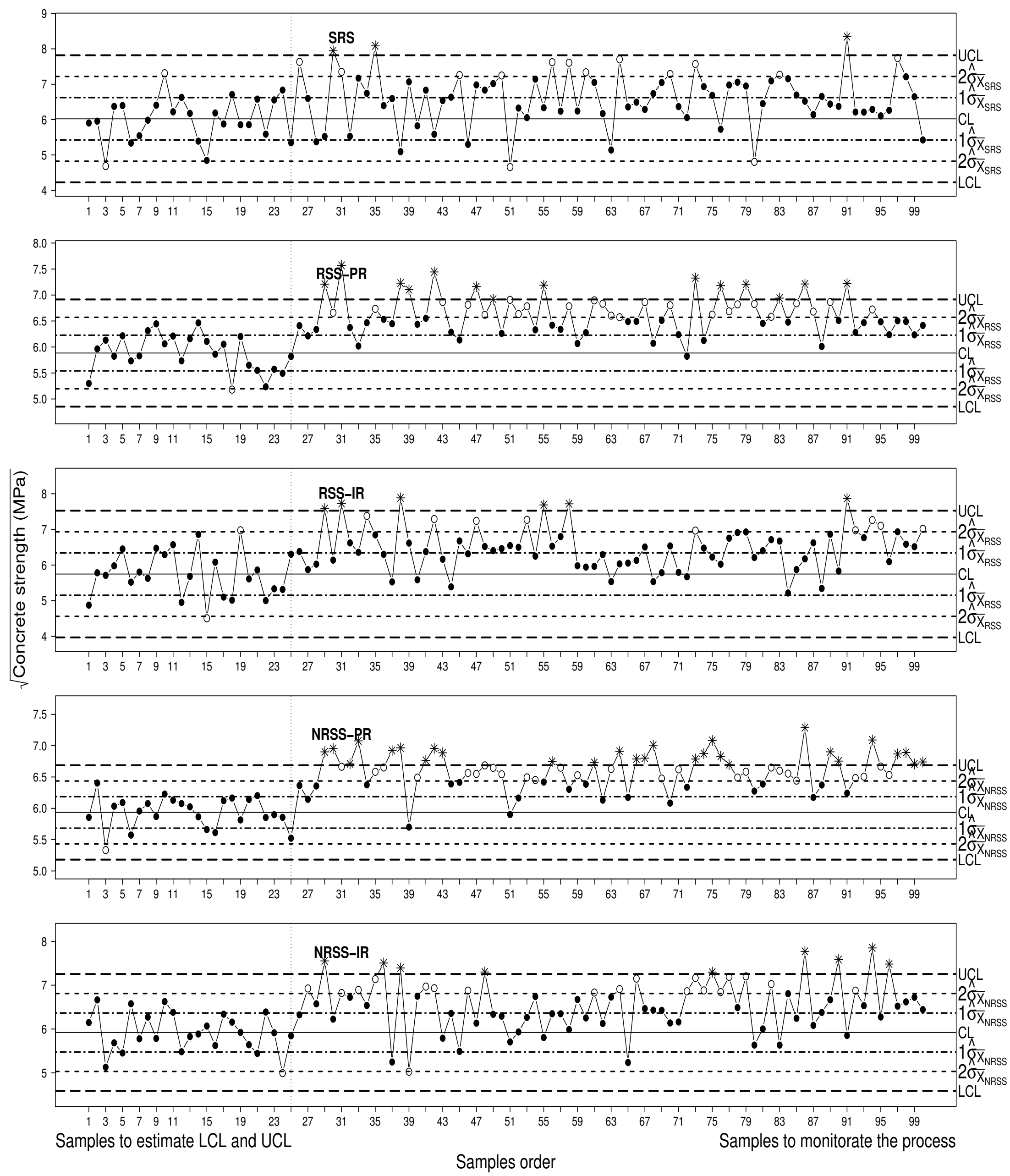

Figure 7. Control charts for concrete strength considering $k = 5$ and an out-of-control process ($\delta = 1.2$). Perfect ranking is denoted as PR, and imperfect ranking as IR.

scenarios when compared to those provided by extreme double ranked set sampling and double extreme ranked set sampling, and surpassed by those provided by double quartile ranked set sampling and quartile double ranked set sampling when $k = 5$ (Abujiya and Muttlak, 2004; Al-Omari and Haq, 2012). Moreover, this superiority is also verified against DRSS control charts for all considered sample sizes. In addition, when considering the double median ranked set sampling and median double ranked set sampling control charts, as can be seen in Abujiya and Muttlak (2004), these designs dominate NRSS, providing lower ARL values. However, it should be considered that double ranked set designs could be expensive, and sometimes infeasible, due to a high operational effort.

Under imperfect ranking, we have shown that the efficiency of NRSS control charts becomes smaller as the correlation between the variables decreases. This is a common fact to other designs based on RSS. Even so, the simulated ARL values for NRSS control charts are predominantly smaller (for out-of-control processes) than the corresponding ones reached by SRS. Additionally, it was possible to verify the superiority of the NRSS control charts with the ones provided by RSS, MRSS and ERSS in most of the simulated scenarios. Also, NRSS was the most robust method for non-normally distributed processes.

In an illustration with real data regarding concrete strength, the SRS, RSS and NRSS control charts presented points randomly distributed around the central line, without any points outside the control limits, when we simulated from a process under statistical control. However, for the out-of-control scenarios, the NRSS control charts performed better when compared to the RSS and the usual control charts based on SRS.

Therefore, based on these results, we recommend NRSS control charts for monitoring the process mean as an efficient alternative to SRS and to other RSS based designs. Under the operational point of view, the ranking of k^2 samples units in a single set (instead of ranking k sets of k units, as it occurs in RSS, MRSS and ERSS designs) may, eventually, become a complicating issue, if the ordering criterion is based, for example, on a visual judgment. However, this will usually not make great difference if the ordering criterion is based, for example, on an auxiliary variable. Finally, the impact of ties in the ranking process should be investigated; see Frey (2012) and Zamanzade and Wang (2018) for some alternatives to overcome the problem of ties in RSS.

Acknowledgement

The authors thank the Editor, an Associate Editor and the reviewers for their valuable comments on an earlier version of this manuscript.

References

Abid, M., Nazir, H.Z., Riaz, M., and Lin, Z., 2017. Investigating the impact of ranked set sampling in nonparametric cusum control charts. Quality and Reliability Engineering International, 33, 203-214.

Abujiya, M. and Muttlak, H., 2004. Quality control chart for the mean using double ranked set sampling. Journal of Applied Statistics, 31, 185-1201.

Al-Omari, A.I. and Bouza, C.N., 2014. Review of ranked set sampling: modifications and applications. Revista Investigación Operacional, 3, 215-240.

Al-Omari, A.I. and Haq, A., 2012. Improved quality control charts for monitoring the process mean, using double-ranked set sampling methods. Journal of Applied Statistics, 39, 745-763.

Al-Saleh, M.F. and AlKadiri, M.A., 2000. Double-ranked set sampling. Statistics and Probability Letters, 48, 205-212.

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics, 12, 171-178.

Azzalini, A., 2019. The R package `sn`: The Skew-Normal and Related Distributions such as the Skew-*t*. R package version 1.5-4.

Balakrishnan, N. and Rao, C.R., 1998. Order Statistics: Theory and Methods. Elsevier, Amsterdam.

Chakraborti, S., Human, S., and Graham, M., 2008. Phase i statistical process control charts: an overview and some results. Quality Engineering, 21, 52-62.

Chen, Z., 2007. Ranked set sampling: its essence and some new applications. Environmental and Ecological Statistics, 14, 355-363.

Chen, Z., Bai, Z., and Sinha, B., 2003. Ranked Set Sampling: Theory and Applications. Springer, New York.

Consulin, C.M., Ferreira, D., Rodrigues de Lara, I.A., De Lorenzo, A., di Renzo, L., and Taconeli, C.A., 2018. Performance of coefficient of variation estimators in ranked set sampling. Journal of Statistical Computation and Simulation, 88, 221-234.

Costa, A.F. and De Magalhaes, M.S., 2007. An adaptive chart for monitoring the process mean and variance. Quality and Reliability Engineering International, 23, 821-831.

Frey, J., 2012. Nonparametric mean estimation using partially ordered sets. Environmental and Ecological Statistics, 19, 309-326.

Haq, A., Brown, J., Moltchanova, E., and Al-Omari, A.I., 2015. Effect of measurement error on exponentially weighted moving average control charts under ranked set sampling schemes. Journal of Statistical Computation and Simulation, 85, 1224-1246.

Koutras, M., Bersimis, S., and Maravelakis, P., 2007. Statistical process control using Shewhart control charts with supplementary runs rules. Methodology and Computing in Applied Probability, 9, 207-224.

Koyuncu, N., 2018. Regression estimators in ranked set, median ranked set and neoteri ranked set sampling. Pakistan Journal of Statistics and Operation Research, 14, 89-94.

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

Kuhn, M. and Johnson, K., 2018. AppliedPredictiveModeling: Functions and Data Sets for 'Applied Predictive Modeling'. R package version 1.1-7.

Mahdizadeh, M. and Zamanzade, E., 2019. Efficient body fat estimation using multistage pair ranked set sampling. Statistical Methods in Medical Research, 28, 223-234.

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

Mehmood, R., Riaz, M., and Does, R.J., 2013. Control charts for location based on different sampling schemes. Journal of Applied Statistics, 40, 483-494.

Mineo, A.M., 2018. normalp: Routines for Exponential Power Distribution. R package version 0.7.0.1.

Montgomery, D.C., 2009. Statistical Quality Control Wiley, New York.

Muttlak, H., 1997. Median ranked set sampling. Journal of Applied Statistical Sciences, 6, 245-255.

Muttlak, H. and Al-Sabah, W., 2003. Statistical quality control based on ranked set sampling. Journal of Applied Statistics, 30, 1055-1078.

Nadarajah, S., 2005. A generalized normal distribution. Journal of Applied Statistics, 32, 685-694.

Nawaz, T. and Han, D., 2019. Monitoring the process location by using new ranked set sampling-based memory control charts. Quality Technology and Quantitative Management, pages in press. Available online at https://doi.org/10.1080/16843703.2019.1572288.

Qiu, P., 2018. Some perspectives on nonparametric statistical process control. Journal of Quality Technology, 50, 49-65.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Samawi, H.M., Ahmed, M.S., and Abu-Dayyeh, W., 1996. Estimating the population mean using extreme ranked set sampling. Biometrical Journal, 38, 577-586.

Santore, F., Taconeli, C.A., and Rodrigues de Lara, I.A., 2019. An adaptive control chart for the process location based on ranked set sampling. Communications in Statistics: Simulation and Computation, pages in press. Available online at https://doi.org/10.1080/03610918.2019.1622722.

Shewhart, W.A., 1924. Some applications of statistical methods to the analysis of physical and engineering data. Bell Labs Technical Journal, 3, 43-87.

Taconeli, C.A. and Cabral, A.D.S., 2019. New two-stage sampling designs based on neoteric ranked set sampling. Journal of Statistical Computation and Simulation, 89, 232-248.

Venables, W.N. and Ripley, B.D., 2002. Modern Applied Statistics with S. Springer, New York.

Yeh, I.C., 1998. Modeling of strength of high-performance concrete using artificial neural networks. Cement and Concrete Research, 28, 1797-1808.

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

Zamanzade, E. and Mahdizadeh, M., 2019. Using ranked set sampling with extreme ranks in estimating the population proportion. Statistical Methods in Medical Research, pages in press. Available online at https://doi.org/10.1177/0962280218823793.

Zamanzade, E. and Wang, X., 2018. Proportion estimation in ranked set sampling in the presence of tie information. Computational Statistics, 33, 1349-1366.

Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to Víctor Leiva, Editor-in-Chief of the ChJS (E-mail: `chilean.journal.of.statistics@gmail.com`). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a"cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at `http://chjs.mat.utfsm.cl/files/ChJS.zip`. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at `http://www.ams.org/mathscinet/msc/`. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

Copyright