Two New Estimators for a Normal Coefficient of Variation based on Ranked Set Sampling

Mohamed S. Abdallah[1], Mohamed S. Hamouda[2]

Quantitative Techniques, Faculty of Commerce, Aswan University, Egypt,

1.E-mail: mohamed_abdallah@com.aswu.edu.eg ; statisticsms.2010@gmail.com

2.E-mail: m.sattar@com.aswu.edu.eg ; msattar_statistic@yahoo.com

Abstract

This article addresses the estimation for the coefficient of variation (CV) of the normal distribution considering ranked set sampling. Using EM algorithm and linear interpolation technique, two new estimators for the CV are proposed. The finite sample size properties of the proposed estimators are examined by simulation studies in terms of their relative efficiency, Pitman measures of closeness and bias criteria. It turns out that the proposed estimators are substantially more efficient and less sensitive to the perfectness assumption than those recently suggested in the literature. Also it was verified the superiority of the proposed estimators using empirical data set.

Keywords: Coefficient of Variation; Concomitant variable; EM algorithm; Estimation methods ; Imperfect Models; Missing Data Approach; Ranked set sampling.

1. Introduction

Ranked set sampling (RSS) is a sampling technique rooted by McIntyre (1952) to estimate more effectively yields of pastures. This sampling technique is helpful in the settings when ranking the sampling items of the interested variable Y without referring to their actual values is much cheaper and easier rather than getting their actual values. The ranking process can be proceeded by eye estimation or using an auxiliary variable frequently called as a concomitant variable X which is expected to be relatively highly correlated with Y. Nowadays RSS is well-established as a procedure for increasing the accuracy or reducing sampling costs.

In order to obtain a RSS based on X, first draw k random samples each of size k of the bivariate variables (X, Y). Second, for each sample, X values are fully measured and sorted in ascending way. Then, measure only the Y values associated with the i^{th} smallest observation of X corresponding to the i^{th} sample ($i = 1 \ldots k$). Finally, repeat the preceding steps m times (cycles) in order to obtain the $n = km$ values of Y. If the number of the selected items across the cycles equals, RSS is hence called balanced RSS. Otherwise, it is called unbalanced RSS. It should be emphasized that we will confine ourselves that (X, Y) are continuous variables drawn through balanced RSS. To clarify the notation, let $x_{lj(i)}$ be the i^{th} smallest observation from j^{th} sample corresponding to l^{th} cycle. And also, let $y_{li[i]}$ be the interested variable's values associated with $x_{lj(i)}$ $((i,j) = 1..k, l = 1 \ldots m)$. Further, Perfect (Imperfect) ranking refers that the rankings are done without (with) errors denoted by rounded (squared) bracket.

Since RSS being introduced, it has been widely adopted in many disciplines as several statistical problems started to be revisited. For instance, Zamanzade and Vock (2015) presented a new estimator for the population variance, Zamanzade and Mohammadi (2016) used RSS for producing efficient estimators for the population mean, Ashour and Abdallah (2018a) introduced novel estimators for the population cumulative distribution function (CDF) under RSS, Ozturk (2018) provided a new ratio estimator under a linearity assumption between X and Y, Ozturk et al. (2018) constructed a Rao-Blackwellized version of maximum likelihood estimators and best linear unbiased estimators under judgment post stratified (JPS) samples. Other interesting studies based on RSS can be found in the monograph of Bouza and Al-Omari (2018) and the references cited therein.

Despite RSS has been applied on a broad class of statistical aspects, but it is still relatively scarce regarding the estimation of the coefficient of variation (CV). The CV, also termed as relative standard deviation, is considered as one of the most important statistical measure for describing the variation within the data. The reason is that CV is a unit-free, dimensionless, measure which is independent of the units in which the variable has been taken. Thereby CV is more meaningful than the standard deviation enables us to compare the volatility of different populations into different scales. This makes CV is widely used into several fields such as: engineering, physics, economy ... etc. Nevertheless, CV has a certain disadvantage as it will approach infinity and hence not reliable whenever the mean value closes to zero. Therefore, it may be advisable to firstly investigate if the mean is significantly far away from zero or not.

Albatineh et al (2014) can be considered the first study which decided to estimate the CV under RSS. The authors compared between RSS and simple random sample (SRS) technique in estimating various confidence intervals for CV. With much extended series of Monte Carlo simulations, it has been observed that the RSS uniformly performed much better in terms of coverage probability and shorter width of the confidence intervals compared with those under SRS. Finally, the authors demonstrated that the two confidence intervals proposed by Mahmoudvand and Hassani (2009) have the shortest width among all the considered intervals. On the other hand, Consulin et al. (2018) proposed four novel estimators, which will be discussed later, for CV based on RSS assuming the normality condition. The authors constructed their proposed estimators via utilizing the suggested estimators for the population mean and the population variance existing in the literature. It was merged from their comparison results that the traditional SRS estimator is outperformed by all the proposed estimators even for small sample sizes. This superiority is increasing whenever perfect ranking scenario holds and diminishes otherwise.

Most recently, Ashour and Abdallah (2019b) proposed different parameter estimation methods under RSS. This motivated us to utilize these proposed methods for introducing two new normal CV estimators and then evaluating the performance of our new estimators relative with those mentioned in the literature. The remainder of this manuscript is structured as follows. Section 2 exhibits several different estimators for CV under RSS. Section 3 presents the two proposed estimators for CV under RSS. The numerical comparisons using simulated data are summarized and organized in Section 4. In Section 5, real data set is also used to present the applicability of the proposed estimators. At the end, section 6 shows our concluding points and some possible extension works.

2. Different Estimators for CV

Throughout this study, Let $\{y_{li[i]}: i = 1 \ldots k; l = 1 \ldots m\}$ be a concomitant-based RSS of size n drawn from a continuous distribution $F\left(\frac{y-\mathcal{M}}{\delta}\right)$, where $F(\cdot)$ is the normal CDF having non-zero mean $\mathcal{M}$ and existence of the variance δ^2. The population CV can be defined as:

$$\tau = \frac{\delta}{\mathcal{M}}.$$

In the rest of this section, we will first briefly review several point estimators for τ one by one as follows.

2.1 Mahmoudvand and Hassani's CV estimator

Mahmoudvand and Hassani (2009) derived an asymptomatic unbiased estimator for τ under SRS assuming the normality distribution. His idea is to expand the naïve estimator for τ which can be expressed as:

$$\hat{\tau}_1 = \frac{\sqrt{\hat{\delta}_1^2}}{\widehat{\mathcal{M}}_1}, \tag{1}$$

where $\widehat{\mathcal{M}}_1$ and $\hat{\delta}_1^2$ are respectively the sample mean and sample variance. Applying the Taylor series expansion on (1) around $\widehat{\mathcal{M}}$, we will get:

$$\hat{\tau}_1 = \hat{\delta}_1 \sum_{i=1}^{\infty} \frac{(-1)^{i-1} \left(\widehat{\mathcal{M}}_1 - \mathcal{M}\right)^{i-1}}{\mathcal{M}^i}.$$

Using the fact that under the normality assumption, $\widehat{\mathcal{M}}_1$ and $\hat{\delta}_1^2$ are independent random variables. This enabled the authors to obtain analytically the expected value $\hat{\tau}_1$ given by:

$$E(\hat{\tau}_1) = c\,\tau \left(1 + \sum_{i=1}^{\infty} \frac{(2i)!}{2^i i!} \left(\frac{\tau^2}{n}\right)^i\right), \tag{2}$$

where $c = \sqrt{\frac{2}{n-1}} \frac{\left(\frac{n}{2}-1\right)!}{\left(\frac{n-1}{2}-1\right)!}$. As $n \to \infty$, the authors claimed that (2) can be rewritten as:

$$E(\hat{\tau}_1) \cong (2 - c)\tau.$$

Naturally, the authors stated that an asymptomatic unbiased estimator for τ can be formulated as follows:

$$\hat{\tau}_2 = \frac{1}{(2-c)}\hat{\tau}_1. \tag{3}$$

In addition, the authors successfully derived the variance of $\hat{\tau}_2$ and proved analytically that $\hat{\tau}_2$ is asymptotically consistent estimator to τ and also it is asymptotically more efficient than $\hat{\tau}_1$. Intuitively, $\hat{\tau}_1$ and $\hat{\tau}_2$ can be straightforwardly applied under RSS. Therefore, for simplicity in the notation, $\hat{\tau}_1$ and $\hat{\tau}_2$ are also denoted respectively to (1) and (3) under RSS.

2.2 Consulin et al.'s CV estimator

As mentioned earlier, Consulin et al. (2018) proposed four estimators for τ under RSS. The first one is the same as the naïve estimator denoted by $\hat{\tau}_1$. Whereas the second one is a modified version of $\hat{\tau}_1$ as replacing $\hat{\delta}_1^2$ in (1) with another more efficient estimator. Since it was proven that $\hat{\delta}_1^2$, see stokes (1980), is a biased estimator for δ^2 under RSS, the authors alternatively preferred to deal with an unbiased estimator provided by MacEachern et al. (2002) given by:

$$\hat{\delta}_2^2 = \frac{1}{n}\left(\sum_{l=1}^{m}\sum_{i=1}^{k}\left(y_{li[i]} - \widehat{\mathcal{M}}_1\right)^2 + \sum_{l=1}^{m}\sum_{i=1}^{k}\left(y_{li[i]} - \widehat{\mathcal{M}}_{1[i]}\right)^2 + \frac{(n-k+1)}{k(n-1)}\sum_{l=1}^{m}\sum_{i=1}^{k}\left(y_{li[i]} - \widehat{\mathcal{M}}_{1[i]}\right)^2\right),$$

where $\widehat{\mathcal{M}}_{1[i]}$ is the sample mean of the i^{th} judgment order statistics. Consequently another CV estimator expectedly enjoys with nice properties given by:

$$\hat{\tau}_3 = \frac{\sqrt{\hat{\delta}_2^2}}{\widehat{\mathcal{M}}_1}.$$

The third proposed estimator is a parametric estimator in its nature. As, the authors suggested to replace both the sample mean and sample variance in (1) with another estimators incorporate the distributional information as well as the structural information supported by RSS design. The authors preferred to use the parametric mean estimator proposed by Stokes (1995) which takes:

$$\widehat{\mathcal{M}}_2 = \frac{\sum_{i=1}^{k} \widehat{\mathcal{M}}_{1[i]} / v_i}{\sum_{i=1}^{k} 1 / v_i},$$

where v_i is the variance of the i^{th} order statistic from a sample of size k under the standard normal distribution. Whilst the authors decided to use the parametric variance estimator proposed by Yu et al. (1999) which takes:

$$\hat{\delta}_3^2 = \frac{1}{n - 1 + \frac{1}{k}\sum_{i=1}^{k} \alpha_i^2} \sum_{l=1}^{m} \sum_{i=1}^{k} \left(y_{li[i]} - \widehat{\mathcal{M}}_1\right)^2,$$

where α_i is the expected value of the i^{th} order statistic from a sample of size k under the standard normal distribution. One can easily realize that either $\widehat{\mathcal{M}}_2$ or $\hat{\delta}_3^2$ should be used with care in practice. As these two estimators are based on the assumption that there is no errors in the ranking process. This assumption is absolutely unrealistic and most likely to be violated. Accordingly, the third CV estimator can be immediately formulated as:

$$\hat{\tau}_4 = \frac{\sqrt{\hat{\delta}_3^2}}{\widehat{\mathcal{M}}_2}.$$

The final CV estimator proposed by Consulin et al. (2018) is constructed by replacing both the sample mean and sample variance in (1) with their maximum likelihood estimators assuming the normality and the perfectness of ranking. In the light of Stokes (1995), the maximum likelihood estimators for $\mathcal{M}$ and δ^2 can be obtained by either maximizing the following likelihood function:

$$L(\mathcal{M},\delta|Y)=\sum_{l=1}^{m}\sum_{i=1}^{k}\log\left(f_i\left(\frac{y_{li[i]}-\mathcal{M}}{\delta}\right)\right)\propto -n\log(\delta)+\sum_{l=1}^{m}\sum_{i=1}^{k}\log\left(f\left(\frac{y_{li[i]}-\mathcal{M}}{\delta}\right)\right)$$
$$+\sum_{l=1}^{m}\sum_{i=1}^{k}(i-1)\log\left(F\left(\frac{y_{li[i]}-\mathcal{M}}{\delta}\right)\right)$$
$$+\sum_{l=1}^{m}\sum_{i=1}^{k}(k-i)\log\left(1-F\left(\frac{y_{li[i]}-\mathcal{M}}{\delta}\right)\right),$$

or equivalently solving simultaneously the following two equations:

$$\left.\begin{aligned}\frac{\partial L(\mathcal{M},\delta|Y)}{\partial\mathcal{M}}&=\sum\nolimits_{l=1}^{m}\sum\nolimits_{i=1}^{k}\frac{f_i'^{\mathcal{M}}\left(\frac{y_{li[i]}-\mathcal{M}}{\delta}\right)}{f_i\left(\frac{y_{li[i]}-\mathcal{M}}{\delta}\right)}=0\\ \frac{\partial L(\mathcal{M},\delta|Y)}{\partial\delta}&=\sum\nolimits_{l=1}^{m}\sum\nolimits_{i=1}^{k}\frac{f_i'^{\delta}\left(\frac{y_{li[i]}-\mathcal{M}}{\delta}\right)}{f_i\left(\frac{y_{li[i]}-\mathcal{M}}{\delta}\right)}=0\end{aligned}\right\}$$

where $f_i(\cdot)$ is the probability density function (pdf) of i^{th} order statistics from a sample of size k under the normal distribution and $f(\cdot)$ is the pdf of the normal distribution. The resulting maximum likelihood estimators for $\mathcal{M}$ and δ are denoted here, respectively, by $\widehat{\mathcal{M}}_3$ and $\hat{\delta}_4^2$. Hence the authors formulated their final CV estimator:

$$\hat{\tau}_5=\frac{\sqrt{\hat{\delta}_4^2}}{\widehat{\mathcal{M}}_3}.$$

It may be easily to realize that the same criticism thrown to $\widehat{\mathcal{M}}_2$ or $\hat{\delta}_3^2$ can be also extended to $\widehat{\mathcal{M}}_3$ and $\hat{\delta}_4^2$, as the latter both also basically assume the perfect assumption. To close this section, it should be indicated that Consulin et al. (2018) made an extensive comparison study among $\hat{\tau}_1$, $\hat{\tau}_3$, $\hat{\tau}_4$ and $\hat{\tau}_5$. The authors concluded that $\hat{\tau}_3$ ($\hat{\tau}_5$) is the best estimator among the nonparametric (parametric) estimators. Accordingly, $\hat{\tau}_3$ and $\hat{\tau}_5$ are reserved for the comparison purposes.

3. Proposed Estimators for CV

In this part, the same methodology mentioned in Consulin et al. (2018) is adopted here in which replacing both the sample mean and sample variance in (1) with another more efficient estimators. In the light of the estimators introduced by Ashour and Abdallah (2019b), two novel CV estimators are proposed as shown below.

3.1 CV Estimator based on EM Algorithm

Ashour and Abdallah (2019b) proposed a new strategy for parameters estimation for location and scale family under RSS using EM algorithm. Their idea is based on incorporating the stochastic relationship between the measured items and the unmeasured items to estimate the unknown parameters. In other words, their strategy is to impute the unmeasured items in the view of the measured items, then use the complete sample for estimate the unknown parameters, finally use the parameters' estimates to upgrade these imputed unmeasured items. By repeating these steps iteratively until the convergence of the parameters' estimates occurs. In order to make the strategy more robust against the imperfect ranking error, the authors resort to participate the fraction-of-random-rankings model proposed by Frey et al. (2007) during implementing their algorithm. Accordingly, $\mathcal{M}$ and δ^2 can be estimated through applying the following steps:

1- Let $\left(\widehat{\mathcal{M}}^{(0)}, \hat{\delta}^{2(0)}\right)$ be the seed estimates for $\mathcal{M}$, δ^2.
2- Set $p = 0$.
3- Estimate the controller parameter λ of the fraction-of-random-rankings model by:

$$\hat{\lambda} = \underset{\lambda \in [0,1]}{\text{Max}} \prod_{i=1}^{k} \prod_{l=1}^{m} \left((1-\lambda)\, b_{i,k-i+1} \left(F\left(\frac{y_{[i,l]} - \widehat{\mathcal{M}}^{(p)}}{\hat{\delta}^{2(p)}} \right) \right) + \lambda \right).$$

where $b_{a,b}(t)$ is the pdf of the Beta distribution with parameters a and b at the point t.

4- Estimate $\mathcal{M}$ and δ^2 by the following equations:

$$\left.\begin{aligned}
\widehat{\mathcal{M}}^{(p+1)} &= \frac{1}{kn} \sum_{l=1}^{m} \sum_{i=1}^{k} \left((1-\hat{\lambda}) \left(y_{li[i]} + E(y \mid y > y_{li[i]}) + E(y \mid y < y_{li[i]}) \right) + \hat{\lambda}\, E(y) \right) \\
\hat{\delta}^{2(p+1)} &= \frac{1}{kn} \sum_{l=1}^{m} \sum_{i=1}^{k} \left((1-\hat{\lambda}) \left(y_{li[i]}^{2} + E(y^2 \mid y > y_{li[i]}) + E(y^2 \mid y < y_{li[i]}) \right) + \hat{\lambda}\, E(y^2) \right) - \left(\widehat{\mathcal{M}}^{(p+1)} \right)^2
\end{aligned}\right\},$$

where:

$$E(y^r \mid y > y_{li[i]}) = \int_{y_{li[i]}}^{\infty} y^r f_{j,i}^{1}(y \mid y_{li[i]})\ dy,$$

$$f_{j,i}^{1}(y, \mathcal{M}, \delta \mid \mathrm{w}) =$$

$$\frac{(k-i)!}{(j-i-1)!\,(k-j)!} \left(\frac{F\left(\frac{y-\mathcal{M}}{\delta}\right) - F\left(\frac{w-\mathcal{M}}{\delta}\right)}{1 - F\left(\frac{y-\mathcal{M}}{\delta}\right)} \right)^{j-i-1} \left(1 - \frac{F\left(\frac{y-\mathcal{M}}{\delta}\right) - F\left(\frac{w-\mathcal{M}}{\delta}\right)}{1 - F\left(\frac{w-\mathcal{M}}{\delta}\right)} \right)^{k-j} \frac{f\left(\frac{y-\mathcal{M}}{\delta}\right)}{1 - F\left(\frac{w-\mathcal{M}}{\delta}\right)},$$

$$w \le y \le \infty$$

and

$$E(y^r \mid y < y_{li[i]}) = \int_{-\infty}^{y_{li[i]}} y^r f_{j,i}^{2}(y, \mathcal{M}, \delta \mid y_{li[i]}) dy,$$

$$f_{j,i}^{2}(y, \mathcal{M}, \delta \mid \mathrm{w}) = \frac{(i-1)!}{(j-1)!\,(i-j-1)!} \left(\frac{F\left(\frac{y-\mathcal{M}}{\delta}\right)}{F\left(\frac{w-\mathcal{M}}{\delta}\right)} \right)^{j-1} \left(1 - \frac{F\left(\frac{y-\mathcal{M}}{\delta}\right)}{F\left(\frac{w-\mathcal{M}}{\delta}\right)} \right)^{i-j-1} \frac{f\left(\frac{y-\mathcal{M}}{\delta}\right)}{F\left(\frac{w-\mathcal{M}}{\delta}\right)},$$

$$-\infty \le y \le w.$$

5- Set $p = p + 1$.

6- Repeat steps $(3-5)$ until stopping rule satisfies. i.e.

$$\text{Max}\left(\left|\widehat{\mathcal{M}}^{(p)} - \widehat{\mathcal{M}}^{(p-1)}\right|, \left|\hat{\delta}^{2(p)} - \hat{\delta}^{2(p-1)}\right|\right) \leq \epsilon.$$

where ϵ is the tolerance size.

7- The final estimates are $\left(\widehat{\mathcal{M}}_4, \hat{\delta}_5^2\right) = \left(\widehat{\mathcal{M}}^{(p)}, \hat{\delta}^{2(p)}\right)$.

Now we can immediately provide our first CV estimator formulated as:

$$\hat{\tau}_{P1} = \frac{\sqrt{\hat{\delta}_5^2}}{\widehat{\mathcal{M}}_4}.$$

3.2 CV Estimator based on the Concomitant Variable

Our second CV estimator is based on the information ranking supported by X. Ashour and Abdallah (2019b) suggested also a nonparametric procedure to estimate $\mathcal{M}$ and δ^2 which can be explained through the following steps:

1- Combining $y_{l[i]}$ and their corresponding values of $x_{li(i)}$ into two new variables $(y_z^*, x_z^* \ , z = 1 \ldots n)$ respectively.

2- Sorting ascending (y_z^*, x_z^*) according to x^* values yielding $(y_{[z]}^*, x_{(z)}^*)$.

3- Estimating $\mathcal{M}$ and δ^2 by the following equations:

$$\left.\begin{aligned} \widehat{\mathcal{M}}_5 &= \frac{1}{nk}\sum\nolimits_{t=1}^{nk} h_1^1(x_{(t)}^*) \\ \hat{\delta}_6^2 &= \frac{1}{nk}\sum\nolimits_{t=1}^{nk} h_1^2(x_{(t)}^*) - \left(\widehat{\mathcal{M}}_5\right)^2 \end{aligned}\right\},$$

where

$$h_1^r(x) = \begin{cases} {y^r}^{iso}_{[1]} & x < x^*_{(1)} \\ {y^r}^{iso}_{[t]} + \dfrac{{y^r}^{iso}_{[t+1]} - {y^r}^{iso}_{[t]}}{x^*_{(t+1)} - x^*_{(t)}}\left(x - x^*_{(t)}\right) & x^*_{(t)} < x < x^*_{(t+1)} \quad t = 1 \dots n-1 \\ {y^r}^{iso}_{[n]} & x^*_{(n)} < x \\ {y^r}^{iso}_{[t]} & x = x^*_{(t)} \quad t = 1 \dots n \end{cases},$$

and ${y^r}^{iso}$is the isotonized values of y^r obtained by the Pool-Adjacent-Violators Algorithm introduced by Ozturk (2007).

Likewise, our second proposed CV estimator can intuitively be formulated as:

$$\hat{\tau}_{P2} = \frac{\sqrt{\hat{\delta}_6^2}}{\widehat{\mathcal{M}}_5}.$$

Of course, $\hat{\tau}_{P2}$ has a limitation in the practice as it assumes that RSS has to be done in the light of X.

4. Efficiency Comparison using Simulated Data Set

This part exhibits the performance of the aforementioned estimation methods based on the simulated data generated through Dell and Clutter (1972) model which assuming (Y, X) has a standard bivariate normal distribution with a correlation coefficient ρ. The chosen values are taken to be $\rho = 1$ for perfect ranking, $\rho = .90$ for closely perfect ranking, $\rho = .50$ for closely imperfect ranking and $\rho = .00$ for imperfect ranking. To illustrate the effect of set and sample sizes, similar to what is done in Ashour and Abdallah (2019 a, b), we considered four different configurations: $(k, m) = (2,5), (2,10), (5,2)$ and $(5,4)$. Without losing of generality, the τ values were obtained by fixing the $\delta = 1$ and varying the

$\mathcal{M}$. The values of $\tau = .25, .5, .75$ and 1. For each combination of ρ, k, m and τ, 5000 data sets are generated. To address the performance of the considered CV estimators, $\hat{\tau}_L$ are computed for each simulated sample, where $L = \{2, 3, 5, P1$ and $P2\}$. Afterwards, we made the comparison study between these estimators through three different criteria. The first criterion is the relative efficiency (RE) which is computed as:

$$RE(\hat{\tau}_L) = \frac{MSE(\hat{\tau}_2)}{MSE(\hat{\tau}_L)},$$

where MSE refers to the mean square error. It is clear that if RE less than one, this refers to the superiority of $\hat{\tau}_2$ over $\hat{\tau}_L$ and vice versa. On the other hand, the second criterion is the Pitman measure of closeness (PC) which is commonly defined as:

$$PC(\hat{\tau}_L) = \Pr(|\hat{\tau}_2 - \tau| < |\hat{\tau}_L - \tau|).$$

One can easily observe if PC less than $.5$, this indicates that $\hat{\tau}_2$ is closer to τ compared to $\hat{\tau}_L$ and vice versa. Finally, the third criterion is the bias of the estimators computed as:

Table 1: Estimated RE and PC of the methods of estimation CV using the simulated data

			ρ = 1		ρ = .9		ρ = 0.5		ρ =0	
	$(k.m)$		RE	PC	RE	PC	RE	PC	RE	PC
$\tau = .25$	(2,5)	$\hat{\tau}_3$	1.02	0.17	1.01	0.15	0.99	0.04	0.97	0.02
		$\hat{\tau}_5$	1.09	0.67	1.06	0.58	0.77	0.26	0.36	0.07
		$\hat{\tau}_{P1}$	1.07	0.72	1.05	0.68	0.99	0.65	0.99	0.58
		$\hat{\tau}_{P2}$	1.47	0.65	1.18	0.63	1.01	0.61	0.96	0.81
	(2,10)	$\hat{\tau}_3$	1.01	0.10	1.00	0.05	0.99	0.00	0.98	0.00
		$\hat{\tau}_5$	1.11	0.66	1.03	0.48	0.58	0.11	0.26	0.01
		$\hat{\tau}_{P1}$	1.08	0.69	1.09	0.56	0.90	0.54	0.99	0.89
		$\hat{\tau}_{P2}$	1.60	0.61	1.33	0.57	0.96	0.60	0.94	0.59
	(5,2)	$\hat{\tau}_3$	1.03	0.75	1.03	0.56	1.00	0.17	0.99	0.11
		$\hat{\tau}_5$	1.15	0.70	1.06	0.44	0.27	0.04	0.07	0.00
		$\hat{\tau}_{P1}$	1.18	0.71	1.07	0.51	0.71	0.58	0.94	0.93
		$\hat{\tau}_{P2}$	1.65	0.71	1.30	0.76	1.00	0.70	0.98	0.66
	(5,4)	$\hat{\tau}_3$	1.05	0.81	1.02	0.58	1.00	0.05	0.99	0.01
		$\hat{\tau}_5$	1.20	0.58	0.98	0.28	0.16	0.00	0.04	0.00
		$\hat{\tau}_{P1}$	1.22	0.58	1.03	0.33	0.74	0.76	1.00	0.99
		$\hat{\tau}_{P2}$	2.02	0.67	1.35	0.66	0.99	0.67	0.95	0.64
$\tau = .5$	(2,5)	$\hat{\tau}_3$	0.99	0.20	0.98	0.15	0.97	0.04	0.94	0.02
		$\hat{\tau}_5$	1.09	0.68	1.03	0.58	0.71	0.26	0.42	0.07
		$\hat{\tau}_{P1}$	1.08	0.74	1.02	0.68	0.98	0.65	0.98	0.68
		$\hat{\tau}_{P2}$	1.51	0.69	1.35	0.69	1.01	0.71	1.01	0.81
	(2,10)	$\hat{\tau}_3$	1.01	0.12	1.01	0.05	0.99	0.00	0.97	0.00
		$\hat{\tau}_5$	1.14	0.66	1.10	0.48	0.65	0.11	0.31	0.01
		$\hat{\tau}_{P1}$	1.13	0.69	1.06	0.56	0.89	0.54	0.96	0.89
		$\hat{\tau}_{P2}$	1.55	0.60	1.27	0.57	0.93	0.57	0.99	0.60
	(5,2)	$\hat{\tau}_3$	1.03	0.74	1.02	0.56	0.99	0.17	0.95	0.11

		$\hat{t}_5$	1.26	0.78	1.05	0.44	0.27	0.04	0.09	0.00
		$\hat{t}_{P1}$	1.27	0.68	1.06	0.51	0.85	0.58	0.90	0.93
		$\hat{t}_{P2}$	1.67	0.70	1.35	0.72	0.98	0.65	0.92	0.66
	(5,4)	$\hat{t}_3$	1.01	0.85	1.01	0.58	0.98	0.05	0.98	0.01
		$\hat{t}_5$	1.20	0.65	0.98	0.28	0.14	0.00	0.05	0.00
		$\hat{t}_{P1}$	1.21	0.66	1.01	0.33	0.67	0.76	1.00	0.99
		$\hat{t}_{P2}$	1.97	0.68	1.38	0.67	0.95	0.67	1.00	0.64
$\tau = .75$	(2,5)	$\hat{t}_3$	0.99	0.22	0.99	0.16	0.97	0.03	0.94	0.01
		$\hat{t}_5$	1.19	0.70	1.13	0.63	0.77	0.28	0.46	0.07
		$\hat{t}_{P1}$	1.18	0.74	1.14	0.71	1.01	0.64	1.00	0.80
		$\hat{t}_{P2}$	1.87	0.66	1.53	0.66	1.03	0.63	0.91	0.58
	(2,10)	$\hat{t}_3$	0.99	0.09	0.99	0.05	0.99	0.01	0.98	0.00
		$\hat{t}_5$	1.14	0.59	1.10	0.51	0.71	0.11	0.35	0.00
		$\hat{t}_{P1}$	1.12	0.63	1.03	0.57	0.95	0.53	1.00	0.91
		$\hat{t}_{P2}$	1.60	0.60	1.30	0.61	0.96	0.64	1.00	0.59
	(5,2)	$\hat{t}_3$	1.03	0.72	1.02	0.52	0.98	0.17	0.94	0.11
		$\hat{t}_5$	1.25	0.68	0.98	0.40	0.31	0.05	0.09	0.01
		$\hat{t}_{P1}$	1.28	0.68	1.01	0.43	0.87	0.58	0.94	0.89
		$\hat{t}_{P2}$	1.95	0.70	1.46	0.68	1.00	0.74	1.05	0.69
	(5,4)	$\hat{t}_3$	1.01	0.83	1.01	0.57	0.99	0.05	0.97	0.00
		$\hat{t}_5$	1.17	0.63	0.90	0.27	0.21	0.01	0.07	0.00
		$\hat{t}_{P1}$	1.17	0.62	0.97	0.44	0.87	0.70	1.01	0.96
		$\hat{t}_{P2}$	1.99	0.60	1.47	0.64	0.98	0.64	0.98	0.61
$\tau = 1$	(2,5)	$\hat{t}_3$	0.99	0.20	0.96	0.16	0.93	0.03	0.95	0.02
		$\hat{t}_5$	1.18	0.69	0.77	0.61	0.51	0.29	0.22	0.07
		$\hat{t}_{P1}$	1.18	0.74	0.87	0.69	0.92	0.62	0.91	0.80
		$\hat{t}_{P2}$	1.86	0.64	1.58	0.63	1.03	0.64	0.97	0.64
	(2,10)	$\hat{t}_3$	0.99	0.13	0.99	0.07	0.98	0.00	0.97	0.00
		$\hat{t}_5$	1.15	0.65	1.02	0.48	0.72	0.12	0.43	0.01
		$\hat{t}_{P1}$	1.16	0.68	1.04	0.56	0.92	0.52	1.03	0.87
		$\hat{t}_{P2}$	1.50	0.60	1.37	0.58	0.92	0.62	1.02	0.60
	(5,2)	$\hat{t}_3$	1.02	0.70	1.01	0.55	0.95	0.20	0.99	0.10
		$\hat{t}_5$	1.23	0.62	1.06	0.44	0.37	0.09	0.30	0.00
		$\hat{t}_{P1}$	1.23	0.63	1.09	0.50	0.88	0.60	1.02	0.91
		$\hat{t}_{P2}$	2.27	0.63	1.28	0.70	1.00	0.69	1.00	0.65
	(5,4)	$\hat{t}_3$	1.02	0.87	1.01	0.52	0.98	0.02	0.98	0.02
		$\hat{t}_5$	1.21	0.65	0.92	0.28	0.23	0.01	0.08	0.00
		$\hat{t}_{P1}$	1.23	0.62	0.97	0.33	0.82	0.72	1.02	0.90
		$\hat{t}_{P2}$	2.14	0.66	1.38	0.63	0.96	0.63	0.95	0.57

Table 2: Estimated bias of the methods of estimation CV using the simulated data

	$(k.m)$		$\hat{t}_2$	$\hat{t}_3$	$\hat{t}_5$	$\hat{t}_{P1}$	$\hat{t}_{P2}$
$\tau = .25$	(2,5)	ρ =1	0.010	0.008	0.018	0.024	0.020
		ρ =0.9	0.007	0.004	0.009	0.019	0.015
		ρ =0.5	0.007	0.001	0.016	0.017	0.007
		ρ =0	0.009	0.003	0.046	0.015	0.010
	(2,10)	ρ =1	0.002	0.009	0.006	0.011	0.007
		ρ =0.9	0.005	0.003	0.003	0.011	0.006
		ρ =0.5	0.006	0.003	0.056	0.011	0.001
		ρ =0	0.005	0.002	0.200	0.008	0.005
	(5,2)	ρ =1	0.002	0.003	0.010	0.020	0.010
		ρ =0.9	0.007	0.007	0.002	0.024	0.004
		ρ =0.5	0.011	0.001	0.068	0.023	0.001
		ρ =0	0.013	0.001	0.184	0.020	0.013
	(5,4)	ρ =1	0.000	0.001	0.000	0.010	0.004
		ρ =0.9	0.003	0.003	0.010	0.012	0.008
		ρ =0.5	0.003	0.001	0.083	0.009	0.005
		ρ =0	0.006	0.003	0.200	0.009	0.008
$\tau = .5$	(2,5)	ρ =1	0.013	0.008	0.028	0.041	0.033
		ρ =0.9	0.006	0.008	0.007	0.034	0.017
		ρ =0.5	0.010	0.001	0.037	0.025	0.001
		ρ =0	0.016	0.003	0.094	0.029	0.016
	(2,10)	ρ =1	0.004	0.002	0.011	0.020	0.013
		ρ =0.9	0.010	0.007	0.005	0.023	0.011
		ρ =0.5	0.003	0.001	0.051	0.014	0.008
		ρ =0	0.001	0.004	0.120	0.001	0.002
	(5,2)	ρ =1	0.003	0.004	0.013	0.038	0.014

		ρ =0.9	0.006	0.004	0.015	0.034	0.008
		ρ =0.5	0.030	0.021	0.152	0.048	0.011
		ρ =0	0.017	0.004	0.388	0.033	0.021
	(5,4)	ρ =1	0.003	0.005	0.011	0.026	0.012
		ρ =0.9	0.003	0.003	0.021	0.024	0.014
		ρ =0.5	0.004	0.008	0.190	0.009	0.017
		ρ =0	0.009	0.004	0.400	0.015	0.014
$\tau = .75$	(2,5)	ρ =1	0.013	0.021	0.011	0.037	0.015
		ρ =0.9	0.019	0.009	0.019	0.054	0.037
		ρ =0.5	0.018	0.035	0.085	0.009	0.018
		ρ =0	0.020	0.040	0.210	0.001	0.016
	(2,10)	ρ =1	0.008	0.003	0.013	0.048	0.014
		ρ =0.9	0.004	0.004	0.022	0.050	0.012
		ρ =0.5	0.014	0.029	0.290	0.017	0.050
		ρ =0	0.007	0.028	0.650	0.008	0.005
	(5,2)	ρ =1	0.007	0.002	0.017	0.051	0.019
		ρ =0.9	0.007	0.007	0.034	0.054	0.017
		ρ =0.5	0.010	0.003	0.250	0.033	0.020
		ρ =0	0.014	0.034	0.660	0.002	0.015
	(5,4)	ρ =1	0.004	0.007	0.015	0.039	0.015
		ρ =0.9	0.001	0.001	0.041	0.025	0.031
		ρ =0.5	0.009	0.015	0.286	0.009	0.031
		ρ =0	0.005	0.004	0.620	0.010	0.012
$\tau = 1$	(2,5)	ρ =1	0.049	0.060	0.014	0.029	0.005
		ρ =0.9	0.092	0.100	0.083	0.009	0.060
		ρ =0.5	0.110	0.140	0.200	0.080	0.100
		ρ =0	0.050	0.087	0.300	0.041	0.053
	(2,10)	ρ =1	0.022	0.027	0.006	0.020	0.005
		ρ =0.9	0.029	0.036	0.034	0.004	0.023
		ρ =0.5	0.031	0.045	0.140	0.006	0.053
		ρ =0	0.044	0.057	0.290	0.035	0.044
	(5,2)	ρ =1	0.038	0.032	0.056	0.047	0.002
		ρ =0.9	0.018	0.017	0.040	0.047	0.024
		ρ =0.5	0.077	0.090	0.450	0.032	0.126
		ρ =0	0.049	0.078	0.980	0.030	0.039
	(5,4)	ρ =1	0.027	0.022	0.006	0.027	0.005
		ρ =0.9	0.011	0.011	0.065	0.026	0.045
		ρ =0.5	0.045	0.054	0.430	0.020	0.060
		ρ =0	0.013	0.024	0.890	0.003	0.002

$$bias(\hat{\tau}_L) = E(|\hat{\tau}_L - \tau|).$$

The RE, PC and bias for $\hat{\tau}_2$, $\hat{\tau}_3$, $\hat{\tau}_5$, $\hat{\tau}_{P1}$, and $\hat{\tau}_{P2}$ were obtained and reported in Table $[1-2]$. In the context of the simulation results, the following points can be concluded:

1- At the perfect ranking, increasing the sample size has a positive effect on the REs and the bias of the CV estimators for a fixed set size with some exceptions. However, when the quality of the ranking tends to the randomization, increasing the sample size has a negative effect on the REs of the CV estimators for a fixed set size particularly when the set size is large.

2. In almost cases, increasing the sample size has a negative effect on the PCs of the CV estimators for a fixed set size particularly corresponding to $\hat{\tau}_3$ and $\hat{\tau}_5$.

3- For a fixed sample size, increasing the set size rather than the number of cycles has strongly a positive effect on the on the REs, PCs and biases provided that the quality of the ranking tends to the perfectness.

4- As expected, there is a touchable effect of the values of ρ on the REs and PCs of the considered estimators, as increasing the values of ρ raising the values of REs and PCs corresponding to all estimators and vice versa. Yet concerning to the bias of the estimators, this effect is slightly week expect for $\hat{\tau}_5$ which the latter is no longer unbiased estimator as $\rho \to 0$.

5- It seems that the true value of the CV has a weak effect on the performance of the estimators. In some cases, higher RE and bias for the largest values for τ.

6- Comparing the behavior of the four estimators with τ_2, one can easily observe the superiority of all the four estimators with respect to both the RE (greater than 1) and the PC (greater than $.5$) criteria with a few exceptions related to τ_3 provided that the ranking are perfectly done. However when the quality of the rankings tends to the randomization, the RE and PC corresponding to τ_3 and τ_5 destroy particularly when the rankings are completely random. Concerning to τ_{P1}, as $\rho \to 0$, the RE (PC) also reduces but this reduction never less than 70% (40%). Likewise concerning to τ_{P2}, the efficiencies loss are not so much as the correspondence RE and PC never less than 90% and 50% respectively.

7- Concerning to the bias criterion, τ_2 and τ_3 are the best estimators in almost considered cases.

8- It is interesting to note that τ_{P2} is the best estimator in view of RE as long as the relation between Y and X is fairly good. However, as this relation becomes weak, the difference of the performance between τ_2 and τ_{P2} is negligible yet recall that the latter is uniformly the better according to PC criterion even at $\rho = 0$

9- Not surprising, it also evident that despite τ_{P1} is slighty outperformed by τ_5 at the perfect scenario, the latter is substantially outperformed by τ_{P1} when the rankings tends to the randomization specially when $\rho = 0$. The reason for this phenomenon is that τ_{P1} uses the fraction-of-random-rankings model which reduces the effect of violating the perfectness assumption.

5. Efficiency Comparison using Empirical Data Set

To help illustrate the CV estimators presented in sections $(2-3)$, in what follows, we assess the performances of $\hat{\tau}_2$, $\hat{\tau}_3$, $\hat{\tau}_5$, $\hat{\tau}_{P1}$ and $\hat{\tau}_{P2}$ estimators using the empirical data set known as body fat data set. This dataset of size 252 observations is provided by Carnegie

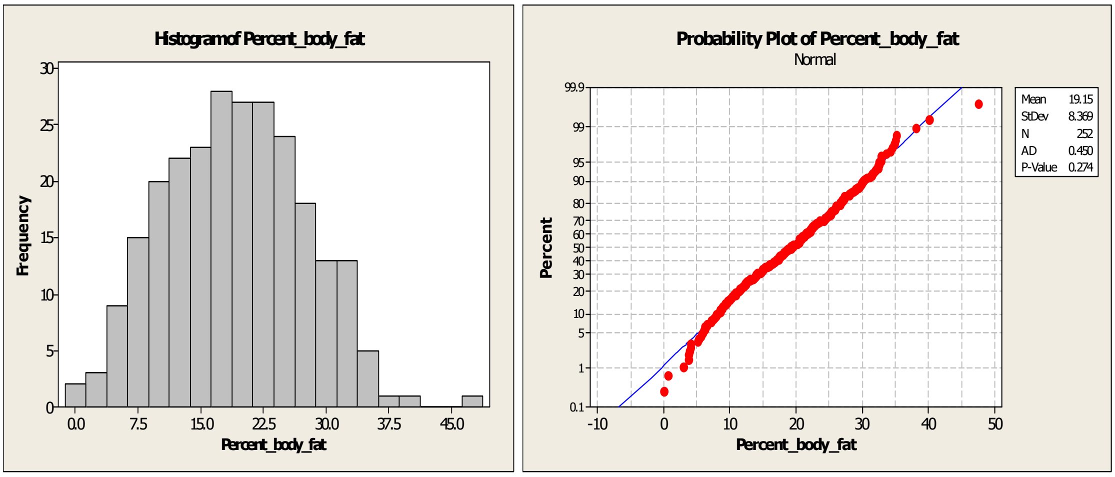
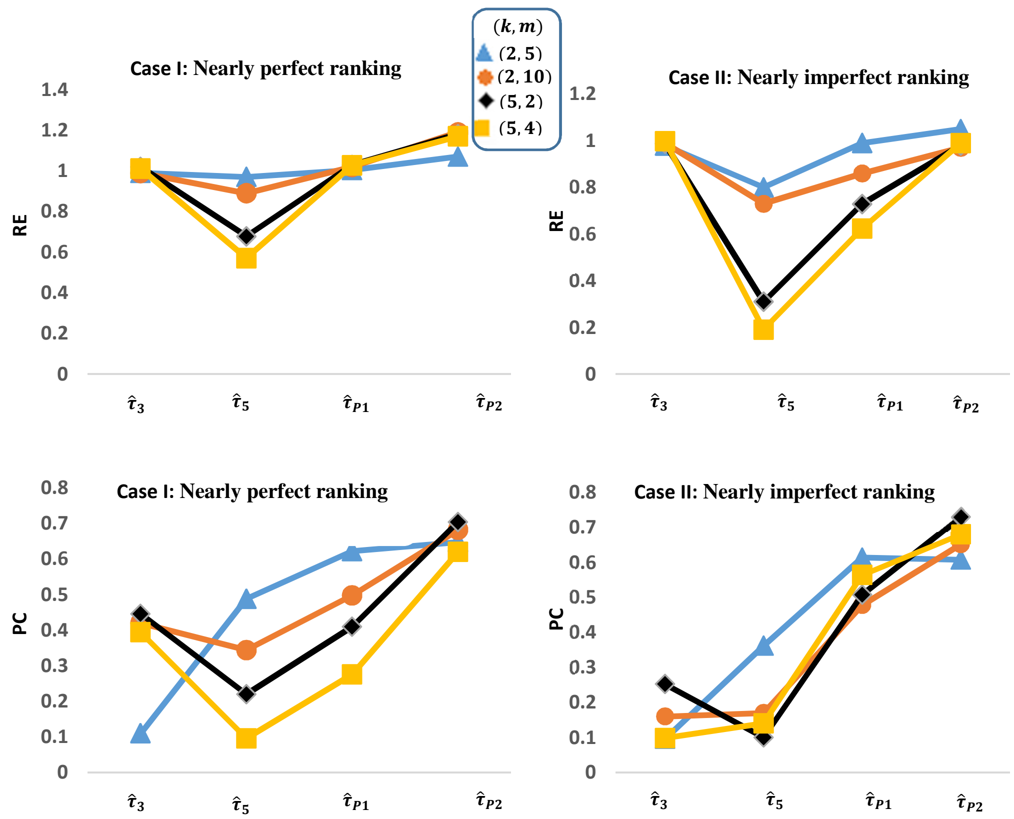


Fig. 1: Histogram and normality test of body fat dataset. This figure appears in color in the electronic version of this article.

Mellon's statistics library and can be found at http://lib.stat.cmu.edu/datasets/bodyfat. We will consider this dataset as a hypothetical population, and we consider the percentage of body fat as the interested variable (Y) and hence the CV in the target population is $\tau =$

43.7%. According to Fig. 1, the percentage of body fat appears has a symmetric shape, besides that the p value of Anderson goodness of fit test equals 27% leading to accepting the normality of the percent body fat data at a significance level 1%. Accordingly, 5,000 RSS with replacement were selected from the body fat data set for the same values of (k, m) determined in section 4. Two scenarios of the selection process were considered. First, the nearly perfect ranking setup based on "Abdomen circumference" variable as a concomitant variable, $\rho = 81.3\%$. Second, the nearly imperfect ranking setup using "Weight" variable as a concomitant variable, $\rho = 61.3\%$. For each sample, all the aforementioned CV estimators were estimated and their RE, PC and bias were also obtained and plotted as shown by Fig. 2.



Case I: Nearly perfect ranking

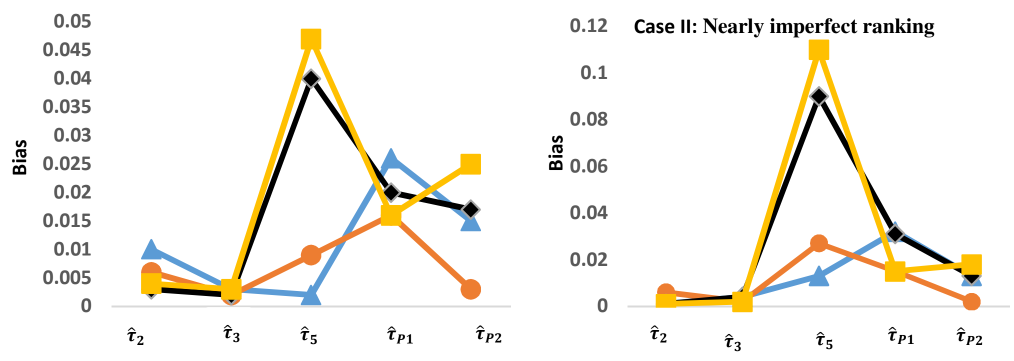


Fig. 2: Estimated RE, PC and bias of the methods of estimation CV using the empirical data. This figure appears in color in the electronic version of this article.

One can easily deduce that there is a high agreement between the results shown by Fig.2 and those reported by Table 1. As $\hat{\tau}_{P2}$ is the best estimator either in terms of RE criterion or PC criterion even at the nearly imperfect ranking scenario. While, $\hat{\tau}_2$ and $\hat{\tau}_3$ are the best estimators with respect to bias criterion. Further $\hat{\tau}_{P1}$ has uniformly better performance than $\hat{\tau}_5$ particularly at the nearly imperfect ranking setting. It is worth remarking that since neither the first scenario nor the second scenario are perfect ranking, increasing the cycle size has better effect on the behavior of the CV estimators rather than increasing the set size for a fixed sample size particularly for $\hat{\tau}_5$ and $\hat{\tau}_{P1}$. Finally it should be informed that all simulation studies in this work are programmed using R statistical software and available at the appendix of this article.

6. Conclusion

In this study, we provide two novel CV estimators under RSS for the situations where data were distributed normally. The first proposed estimator is based on EM algorithm, while the second one is derived under existing of the concomitant based-information. In the view of RE, PC and bias criteria, the numerical findings recommended that the second proposed CV estimator is at least as efficient as all competitors in almost cases and it can be reliable given the ranking process is performed in the light of a concomitant variable. Otherwise, one can adopt the first proposed CV estimator provided that the ranking process is fairly good. However, in the case of presence of ranking errors, the CV estimator proposed by Mahmoudv and Hassani (2009) may be advisable.

We would like to mention that although the proposed CV estimators can be performed under other sampling distributions, we reached finally that the efficiency of the proposed CV estimators can be ignorable except under the normality assumption. Future work may construct confidence interval for the proposed CV estimators.

Acknowledgments

The authors are grateful to anonymous referees and an editor for their valuable comments and suggestions which lead to this improved version.

References

[1.] Albatineh AN, Kibria BMG, Wilcox ML, et al. (2014). Confidence interval estimation for the population coefficient of variation using ranked set sampling: a simulation study, Journal Applied Statistics 41(4):733–751.

[2.] Ashour, S. and Abdalluh, M. (2018a). Estimation of Distribution Function based on Ranked Set Sampling: Missing Data Approach. To appear in Thailand Statistician.

[3.] Ashour, S. and Abdalluh, M. (2018b). New Parametric Estimation Methods based on Ranked Set Sampling. To appear in Gazi University: Journal of Science.

[4.] Ashour, S. and Abdalluh, M. (2018c). New Distribution Function Estimation and Perfectness Testing under Ranked Set Sampling. To appear in Journal Modern Applied Statistics Methods.

[5.] Bouza, C.N. and AL-Omari, A.I. (2018). Ranked Set Sampling: 65 Years Improving the Accuracy in Data Gathering. Elsevier inc.

[6.] Consulin, CM. Ferreira, D. Rodrigues, IA. et al. (2018). Performance of coefficient of variation estimators in ranked set sampling. Journal of Statistical Computation and Simulation. 88(2): 221–234.

[7.] Dell, TR. and Clutter, JL. (1972). Ranked set sampling theory with order statistics background. Biometrics, 28(2): 545-555.

[8.] Frey, J. (2007). New imperfect ranking models for ranked set sampling. Journal of Statistical Planning and Inference, 137: 1433–1445.

[9.] Maceachern, SN. Ozturk, O. Wolfe, DA. (2002). A new ranked set sample estimator of variance. Journal of the Royal Statistical Society: Series B, 64(2): 177-188.

[10.] Mahmoudvand R, and Hassani H. (2009). Two new confidence intervals for the coefficient of variation in a normal distribution. Journal Applied Statistics, 36(4):429–442.

[11.] Mcintyre, GA. (1952). A method for unbiased selective sampling using ranked set sampling. Australian Journal of Agricultural Research, 3: 385-390.

[12.] Modarres, R. Terrence, PH. and Zheng, G. (2006). Resampling methods for ranked set samples. Computational Statistics and Data Analysis, 51: 1039-1050.

[13.] Ozturk, O. (2007). Statistical inference under a stochastic ordering constraint in ranked set sampling. Nonparametric Statistics, 19: 131–144.

[14.] Ozturk O. (2018). Ratio Estimators Based on a Ranked Set Sample in a Finite Population Setting. Journal Korean Statistics Society, 47(2): 226-238.

[15.] Ozturk, O. Sultan, K. S. and Moshref, M. E. (2018). Parametric inference based on judgment post stratified samples. Journal of the Korean Statistical Society, 47(1), 24-31.

[16.] Stokes, SL. (1980). Estimation of variance using judgment ordered ranked set samples, Biometrics: 36 (1), 35–42.

[17.] Stokes, SL. (1995). Parametric ranked set sampling. Annals of the Institute of Statistical Mathematics, 47: 465–482.

[18.] Yu PLH. Lam, K, Sinha, BMK. (1999). Estimation of normal variance based on balanced and unbalanced ranked set samples. Environmental Ecological Statistics, 6(1): 23–46.

[19.] Zamanzade, E. and Vock, M. (2015). Variance estimation in ranked set sampling using a concomitant variable. Statistics & Probability Letters, 105: 1-5.

[20.] Zamanzade, E. and Mohammadi, M. (2016). Some modified mean estimator in ranked set sampling using a covariate. Journal of Statistical Theory and Applications, 15(2): 142-152.

Appendix

R script for the studied CV estimators

```
## Required Libraries
library(MASS);library(stats4)
CV_ESTIMATORS=function (k,m,r,t) { y=matrix(0,k,m); k=nrow(y); m=ncol(y);
x=array(0,c(k,k,m));z=y;n=k*m;
for (j in 1 :k) for (i in 1:m)
{q=mvrnorm(k,c(0,0),matrix(c(1,r,r,1),2))
x[,j,i]=sort(q[,1])
z[j,i]=sort(q[,1])[j]
y[j,i]=q[order(q[,1]),2][j]}
y=qnorm(pnorm(y),1/t,1);
## 1- CV estimator Proposed by Mahmoudvand and Hassani (2009)
C=sqrt(2/(n-1))*gamma(n/2)/gamma((n-1)/2)
T1=(sd(y)/mean(y))/(2-C)
## 2- CV estimator based on the the variance estimator of MacEachern et al. (2002)
SE=sum((y-mean(y))^2);
STT=0; for (i in 1:k) STT[i]= sum((y[i,]-mean(y[i,]))^2);STT=sum(STT);MST=SE/(k-1)-
STT/(k-1)
MSE=STT/(k*(m-1))
SIG= ((k-1)*MST+(m*k-k+1)*MSE)/n;SIG=SIG^.5
T2=(SIG/mean(y))
## 3- CV estimator based on the MLE
S= function (a,b) {t=y; for (i in 1:nrow(y))  for (j in 1:ncol(y)) {t[i,j]=
pnorm(y[i,j],a,b)^(i-1)*(1-pnorm(y[i,j],a,b))^(k-i)*(dnorm(y[i,j],a,b))};
-log(prod(t))}
ST=mle(S,start=list(a=mean(y),b=sd(y)))
T3=(coef(ST)[2]/coef(ST)[1])
## 4- CV estimator based on the EM algorithm
OZE=function(a,b,F,j) {YY=matrix(0,k,k);
for (i in 1:k) YY[i,i]=y[i,j]
f1=function(z,q,i)
{
  f= function (x) {x*dnorm(x,a,b)/(1-pnorm(y[i,j],a,b))}
  c(integrate(f,y[i,j],50))$value};
f2=function(z,q,i)
{
  f= function (x) {x*dnorm(x,a,b)/pnorm(y[i,j],a,b)};
    c(integrate(f,-50,y[i,j]))$value};
f= function (x) { x*dnorm(x,a,b)}
t=integrate(f,-50,50)$value
for (p in 1:(k-1)) for (q in (p+1):k) YY[q,p]=F*f1(p,q,p)+(1-F)*t
for (p in 2:k) for (q in 1:(p-1)) YY[q,p]=F*f2(p,q,p)+(1-F)*t
YY};
OZQ=function(a,b,F,j) {YY=matrix(0,k,k);
for (i in 1:k) YY[i,i]=y[i,j]^2
f1=function(z,q,i)
{
f= function (x) {x^2*dnorm(x,a,b)/(1-pnorm(y[i,j],a,b))}
c(integrate(f,y[i,j],50))$value};

f2=function(z,q,i)
{
f= function (x) {x^2*dnorm(x,a,b)/pnorm(y[i,j],a,b)};
c(integrate(f,-50,y[i,j]))$value};
f= function (x) { (x^2)*dnorm(x,a,b)}
t=integrate(f,-50,50)$value
for (p in 1:(k-1)) for (q in (p+1):k) YY[q,p]=F*f1(p,q,p)+(1-F)*t
for (p in 2:k) for (q in 1:(p-1)) YY[q,p]=F*f2(p,q,p)+(1-F)*t
YY};
fE= function (a,b,F){
  Q=OZE(a,b,F,1); for (j in 2:ncol(y)) Q= cbind(Q,OZE(a,b,F,j));
  Q=c(Q);Q}
fQ= function (a,b,F){
  Q=OZQ(a,b,F,1); for (j in 2:ncol(y)) Q= cbind(Q,OZQ(a,b,F,j));
  Q=c(Q);Q}
L=function (w) {a=mean(y);b=sd(y);t=y;
for (i in 1:w) {Lik= function(F) {for (i in 1:nrow(y))  for (j in 1:ncol(y)) {t[i,j]=
F*dbeta(pnorm(y[i,j],a,b),i,nrow(y)+1-i) +(1-F)};-log(prod(t)) };
F=nlminb(.5,Lik,lower=0,upper=1)$par;
a1=a;b1=b;
TE=fE(a,b,F);TQ=fQ(a,b,F)
a=mean(TE);b=sqrt(mean(TQ)-(a)^2)
```

```
a2=a;b2=b
if ((abs(b2-b1))<.001) break};c(a,b)};
EM=L(20);
T4=(EM[2]/EM[1])
## 5- CV estimator based on the Concomitant Variable
Y=isoreg(y[order(z)])$yf
m1=approx(sort(c(z)),Y,x,yleft=approx(sort(c(z)),Y,min(z))$y,yright=approx(sort(c(z)),Y,m
ax(z))$y)$y;
Y=y^2;
Y=isoreg(Y[order(z)])$yf
m2=approx(sort(c(z)),Y,x,yleft=approx(sort(c(z)),Y,min(z))$y,yright=approx(sort(c(z)),Y,m
ax(z))$y)$y;
sig=sqrt(mean(m2)-mean(m1)^2);
T5=(sig/mean(m1))
c(T1,T2,T3,T4,T5)
```