Clinical Epidemiology and Global Health 7 (2019) 17–21

Contents lists available at ScienceDirect

Clinical Epidemiology and Global Health

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

Variance estimation of persons infected with AIDS under ranked set sampling☆

Carlos N. Bouza[a], Jose F. García[b], Gajendra K. Vishwakarma[c,*], Sayed M. Zeeshan[c]

[a] Department of Mathematics and Computation, University of Havana, Havana, Cuba
[b] DACEA, Universidad Juárez Autónoma de Tabasco, Mexico
[c] Department of Applied Mathematics, Indian Institute of Technology Dhanbad, India

ARTICLE INFO

Article history:
Received 5 November 2017
Accepted 8 December 2017
Available online 13 December 2017

Keywords:
Distribution
Expected model variance
Order statistics
Ranked set sampling
Simple random sampling

ABSTRACT

This paper deals with the analysis of the estimation of the variance of a sensitive variable. The randomized response (RR) procedure gives confidence to the interviewee that his/her information on private (sensitive variable) is protected. In general, a simple random sampling (SRS) with replacement design is used for selecting a sample. We develop a model for the estimation of the variance of sensitive variable in randomized response procedure, when ranked set sampling (RSS) design is used under ranking criteria. The usual gain in accuracy related to the use of ranked set sampling is revealed. The performance of the proposed model is demonstrated using data provided by a study of samples of persons infected with AIDS disease.

1. Introduction

Ranked set sampling is an alternative sample design, which generally provided gains in accuracy with respect to simple random sampling with replacement (SRSWR). It was proposed for estimating pastures yield by Ref. [5] in 1952. He established that to estimate the mean pasture yield using RSS and found their inference more efficient than to select the sample using a simple random sampling (SRS) design. The units may be ranked by means of a cheap procedure and then an order statistics is selected from each of the independent samples selected using srs with replacement (SRSWR). It turned out that the use of ranked set sampling is highly beneficial and leads to estimators which are more precise than the usual sample mean per unit ones. The method is now referred to as ranked set sampling (RSS) method in the literature. Ref. [10] were the first who proved that the mean estimator from RSS is more efficient than that from SRS. This led a lot of research has been done by various authors including, Refs. [1–4,7–9].

In this paper, we propose a model using RSS, instead of SRS with replacement (SRSWR), for studies of the variance. The rest of this paper is organized as follows: Section 2 develops the study of the One-Way Analysis of Variance. Section 3 is devoted to the presentation of estimators of the variance. Section 4 is devoted to the development of numerical studies of the behavior of the analyzed models in testing hypothesis. We discuss the results obtained from the use of SRSWR and develop alternative RSS models in the next section. Samples of persons infected with the AIDS virus are analyzed and the behavior of the accuracy of the different alternative estimators is studied.

2. Estimation of the treatment effects in a one-way layout in ranked set sampling

Consider the one-way layout

$$Y_{ij} = \mu_i + å_{ij} = \mu + \alpha_i + å_{ij},\ i = 1, \ldots, k,\ j = 1, \ldots, n(i). \tag{2.1}$$

This issue is important in many applications and has been studied extensively. Let Y be the variable of interest. We select independent samples of size $n(i)$, $i = 1, \ldots, k$, using simple random sampling with replacement, (SRSWR), for estimating the parameters of interest μ, $\alpha_i = (\mu_i - \mu)$, $i = 1, \ldots, k$. We assume that for any $i = 1, \ldots, k$ and $j = 1, \ldots, n(i)$, $E(å_{ij}) = 0$, $V(å_{ij}) = \sigma^2{}_i$ and $\mathrm{Cov}(a^\circ_{ij} a^\circ_{i'j'}) = 0$, if $i \neq i'$ and/or $j \neq j'$. The usual estimation of the

☆ Mathematics Subject Classification: Primary 62D05; Secondary 62G05; Tertiary 62G08.

* Corresponding author.
E-mail addresses: bouza@matcom.uh.cu (C.N. Bouza), jfgr55@hotmail.com (J.F. García), vishwagk@rediffmail.com (G.K. Vishwakarma), zeeshan008x52@gmail.com (S.M. Zeeshan).

https://doi.org/10.1016/j.cegh.2017.12.002
2213-3984/

effects α_i is

$$\alpha_i^* = \frac{\sum_{j=1}^{n(i)} y_{ij}}{n(i)} - \frac{\sum_{i=1}^{k}\sum_{j=1}^{n(i)} y_{ij}}{n} = (\bar{y}_i - \bar{y}); \quad \text{where } n = \sum_{i=1}^{k} n(i) \tag{2.2}$$

Its variance is given by

$$V(\alpha_i^*) = E(\bar{y}_i \pm \mu_i \pm \mu - \bar{y})^2 = \frac{\sigma_i^2}{n(i)} + \frac{\sigma^2}{n} + (\mu_i - \mu)^2. \tag{2.3}$$

Ref. [6] proposed to use RSS. As usual, the model was based on the selection of *n(i)* independent samples of size *n(i)* using SRSWR and to rank each of them. That is we have hypothetically for each i = 1, . . . ,k $s_i = \{(Y_{i11}\ldots, Y_{i1n(i)})_1, \ldots, (Y_{in(i)1}\ldots, Y_{in(i)n(i)})_{n(i)}\}$ and by ranking, we have the ranked samples$\{(Y_{i1(1)}\ldots, Y_{i(n(i))}), \ldots, (Y_{in(i)(1)}\ldots, Y_{in(i)(n(i))})\}$. Y is measured in the statistic of order (SO) t in the t$^{\text{th}}$ sample. Then our set of results for treatment "i" is

$$\begin{aligned} s(i) &= \{(Y_{i1(1)}), (Y_{i2(2)}), \ldots, (Y_{it(t)})\ldots, (Y_{in(1)n(i)})\} \\ &= \{Y_{i(1)}, Y_{i(2)}, \ldots, Y_{i(t)}\ldots, Y_{i(n(i))}\} \end{aligned} \tag{2.4}$$

We deal with the linear model
$Y_{i(j)t} = \mu_i + \varepsilon_{i(j)t} = \mu + \alpha_i + \varepsilon_{i(j)t}$, *i = 1, . . . , k, j = 1, . . . , n(i)*

$$\bar{y}_{(i)} = \frac{\sum_{j=1}^{n(i)} y_{i(j)}}{n(i)}, \quad \bar{y}_{RSS} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n(i)} y_{i(j)j}}{n}, \quad n = \sum_{i=1}^{k} n(i)$$

It is unbiased and

$$V(\bar{y}_{(i)}) = \frac{\sum_{j=1}^{n(i)} \sigma^2{}_{i(j)}}{n^2(i)}, \sigma^2{}_{i(j)} = V(y_{i(j)})$$

It was hypothesized that, $\sigma^2{}_{i(j)} = \sigma^2{}_{(i)}$, which is the counterpart of the hypothesis used in the development of the one-way layout ANOVA. $\sigma^2{}_i = \sigma^2$, in the inferences based on SRSWR. Using the relation established by Ref. [10] we can derive that for any *i = 1, . . . ,k*

$$V(\bar{y}_{(i)}) = \frac{\sigma^2{}_i}{n(i)} - \frac{\sum_{j=1}^{n(i)} \Delta_{i(j)}^2}{n^2(i)}, \Delta_{i(j)} = \mu_{i(j)} - \mu_{(i)}, \mu_{i(j)} = E(y_{i(j)}),$$

Let us look for the RSS counterpart of the results (2.2) and (2.3).

Proposition 2.1. $\alpha_{(i)}^* = \frac{\sum_{j=1}^{n(i)} y_{i(j)}}{n(i)} - \frac{\sum_{i=1}^{k}\sum_{j=1}^{n(i)} y_{i(j)}}{n} = \left(\bar{y}_{(i)} - \bar{y}_{RSS}\right)$ is unbiased and more accurate thanα_i^*.

Proof: Due to the unbiasedness of the RSS estimators

$E(\alpha_{(i)}^*) = E(\alpha_i^*) = E(\bar{y}_{(i)}) - E(\bar{y}_{RSS}) = \mu_i - \mu$ and $V(\alpha_{(i)}^*) = E(\bar{y}_{(i)} \pm \mu_i \pm \mu - \bar{y}_{RSS})^2 = \frac{\sum_{j=1}^{n(i)} \sigma_{i(j)}^2}{n^2(i)} + \frac{\sum_{i=1}^{k}\sum_{j=1}^{n(i)} \sigma_{i(j)}^2}{n^2} + (\mu_i - \mu)^2$

We have that $\sigma_{i(j)}^2 = \sigma_i^2 - \Delta_{i(j)}^2$ then substituting in the above equation

$$V(\alpha_{(i)}^*) = \frac{\sigma_i^2}{n(i)} + \frac{\sum_{i=1}^{k} n(i)\sigma_i^2}{n^2} + (\mu_i - \mu)^2 - \psi(1)$$

where, represents the gain in accuracy due to the use of

$$\text{RSS}.\Psi(1) = \frac{\sum_{j=1}^{n(i)} \Delta_{i(j)}^2}{n^2(i)} + \frac{\sum_{i=1}^{k}\sum_{j=1}^{n(i)} \Delta_{i(j)}^2}{n^2} \geq 0$$

Remark 1. If $\forall i = 1, \ldots, k$, $\sigma_i^2 = \sigma^2 \Rightarrow \frac{\sigma^2}{n(i)} + \frac{\sigma^2}{n}$ and the usual relation is obtained.

3. The estimation of the variance in RSS

A basic relationship in RSS is

$$\sigma^2 = \frac{1}{k}\sum_{r=1}^{k} \sigma_{(r)}^2 + (\mu_{(r)} - \mu_{(r')})^2; \quad \text{if } r \neq r' \tag{3.1}$$

Ref. [9] suggested as estimator of it, for one cycle,

$$\sigma_S^2 = \frac{1}{(k-1)}\sum_{r=1}^{k} (Y_{(r)i} - \mu_{rss})^2 \quad \text{where}, \mu_{rss} = \frac{1}{k}\sum_{r=1}^{k} Y_{(r)}$$

and its expectation is $E(\sigma_S^2) = \sigma^2 + \frac{1}{k(k-1)}\sum_{r=1}^{k} (\mu_{(r)} - \mu_{rss})^2$.

Considering the structure of the One-Way ANOVA the estimator proposed by Ref. [9] is given in the next proposition.

Proposition 3.1: Ref. [9]; $\sigma_S^2 = \sigma^2 + \frac{1}{(nk-1)}\sum_{i=1}^{n}\sum_{r=1}^{k} (Y_{(r)i} - \mu_{rss})^2$

where, $\mu_{rss} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{r=1}^{k} Y_{(r)i}$ estimates the RSS variance and if $n \to \infty$then $E(\sigma_S^2) = \sigma^2$.

Stokes derived that this estimator over estimates σ^2 and its variance is

$$V(\sigma_S^2) = \frac{n}{(nk-1)^2}\left\{ (\frac{nk-1}{nk})^2 \sum_{r=1}^{k} \mu_{4(r)} + 4\sum_{r=1}^{k} \Delta_{(r)}^2 \sigma_{(r)}^2 + 4(\frac{nk-1}{nk})\sum_{r=1}^{k} \Delta_{(r)} \mu_{3(r)} \right.$$

Note that this error depends on moments of the distribution of the order statistics then the variable's distribution must be known. Hence to derive an explicit formula is very complex.

Ref. [4] proposed to use as estimator

$$\sigma_M^2 = \sigma_{M1}^2 + \sigma_{M2}^2 \tag{3.2}$$

where,

$$\sigma_{M1}^2 = \frac{1}{2n^2k^2}\sum_{r \neq r'}^{k}\sum_{i=1}^{n}\sum_{j=1}^{n} (Y_{(r)i} - Y_{(r')j})^2 \tag{3.3}$$

and

$$\sigma_{M2}^2 = \frac{1}{2n(n-1)^2k^2}\sum_{r=1}^{k}\sum_{j=1}^{n}\sum_{j=1}^{n} (Y_{(r)i} - Y_{(r')j})^2 \tag{3.4}$$

It is unbiased. Next proposition gives its properties.

Proposition 3.2: Ref. [4]; $\sigma_M{}^2$ is unbiased and its variance, if $\mu_{(r)} < \infty$, is $V(\sigma_M^2) = A + B + C + D + F$

where, $A = \frac{1}{nk^2}\sum_{r=1}^{k} \mu_{4(r)}$, $B = \frac{4}{nk^2}\sum_{r=1}^{k} \mu_{3(r)} \Delta_{(r)}$, $C = \frac{4}{nk^2}\sum_{r=1}^{k} \sigma_{(r)}^2 \Delta_{(r)}^2$,

$D = \frac{4}{n^2k^4}\sum_{r<r'} \sigma_{(r)}^2 \sigma_{r'}^2$, $F = \frac{k^2(n-1)-2}{n(n-1)k^4}\sum_{r=1}^{k} \sigma_{(r)}^2$

Using the mean square errors (MSE) and One Way ANOVA decomposition ideas we have that

$MST = MS1 - MS2$

taking $\mu_{rss(r)} = \frac{1}{n}\sum_{i=1}^{n} Y_{(r)i}$, then

$$MS1 = \frac{1}{k-1}\sum_{r=1}^{k}\sum_{i=1}^{n}(Y_{(r)i} - \mu_{rss})^2$$

$$MS2 = \frac{1}{k-1}\sum_{r=1}^{k}\sum_{i=1}^{n}\left(Y_{(r)i} - \mu_{rss(r)}\right)^2$$

Then the rank-residual *MSE* is:

$$MSR = \frac{1}{k(n-1)}\sum_{r=1}^{k}\sum_{i=1}^{n}\left(Y_{(r)i} - \mu_{rss(r)}\right)^2$$

The expectations of the *MSE*'s are

$$E(MST) = \frac{1}{k}\sum_{r=1}^{k}\sigma^2_{(r)} + \frac{1}{n(k-1)}\sum_{r=1}^{k}\left(\mu_{(r)} - \mu\right)^2$$

$$E(MSR) = \frac{1}{k}\sum_{r=1}^{k}\sigma^2_{(r)}.$$

Then the variance of σ^2_M as

$$\sigma^2_M = \frac{(k-1)MST + (nk-k+1)MSR}{nk}$$

and its expectations given by

$$E(\sigma^2_M) = \frac{n+2}{nk}\sum_{r=1}^{k}\sigma^2_{(r)} + \frac{1}{n^2k}\sum_{r=1}^{k}(\mu_{(r)} - \mu)^2$$

The ordering made using an auxiliary variable *X* is equivalent to the use of SRS in the inferior of the scenarios. It is well known that RSS is equivalent to it, in terms of accuracy, in such cases. That is, for any $r\mu_{(r)} = -\mu$. Hence the statistic

$$V(n) = \frac{MST}{MSR} \tag{3.5}$$

under the hypothesis of random ranking must be close to 1. Therefore, we can evaluate the usefulness of the ranking of *Y*, produced by *X*, by analysing *V(n)*, as in regression analysis, through the coefficient of determination. In this case, large values of *V(n)* implies that the ranking is more different than the ranking produced by pure randomness. That is reasoning similar to the non-parametric evaluation of the goodness of regression fitting. Under the hypothesis of normality *V(n)* is distributed *F(k, k(n − 1))* and inferences can be developed using the classic parametric theory using *F* tests.

4. Monte Carlo evaluation

4.1. Normality based tests

1000 runs (samples) were generated using the Uniform *(0,2)*, Normal *(0,1)*, Exponential *(1)*, Gamma with density function $f(x) = x^4 \exp(-x)/\Gamma(5)$; $x > 0$, U-shaped *with density function* $f(x) = 3\times^2/2$; $x\in[0,1]$ and the Lognormal *(0,1)* distribution. The sample size parameters were $n\in\{2, 3, 4, 5\}$ and $k\in\{2, 3, 4, 5\}$. The ANOVA was performed using the normal approximation and $\alpha=0.05$. An estimation of the percentage of samples in which we accepted the true hypothesis H_0 was computed for SRS and RSS and the results presented in Table 1.

Tables 2–7 present the results of *1000* runs (samples) generated, using different bivariate distributions, with $\rho\in\{0, 0.05, 0.75, 0.90, 0.95\}$, of the mean of *V(n)* then we computed the values of *V(mean)* using the following formula

$$V(mean) = \frac{1}{1000}\sum_{1\le h\le 100} V(n)_h \tag{4.1}$$

Table 1
Percentage of acceptance of the true hypothesis H_0 using One Way ANOVA for 1000 runs (samples) generated and $\alpha = 0.05$.

Distribution	*n*	*k*	SRS	RSS	Distribution	*n*	*k*	SRS	RSS
Uniform (0,2)	2	2	0.76	0.72	*Normal (0,1)*	2	2	0.88	0.87
	2	3	0.74	0.71		2	3	0.87	0.87
	2	4	0.78	0.71		2	4	0.88	0.87
	2	5	0.78	0.77		2	5	0.91	0.88
	5	2	0.79	0.77		5	2	0.93	0.89
	5	3	0.79	0.76		5	3	0.93	0.92
	5	4	0.82	0.78		5	4	0.93	0.93
	5	5	0.81	0.78		5	5	0.94	0.92
Exponential(1)	2	2	0.54	0.67	*Gamma* $x^4\exp(-x)/\Gamma(5)$ $x > 0$	2	2	0.66	0.69
	2	3	0.54	0.67		2	3	0.66	0.69
	2	4	0.66	0.72		2	4	0.66	0.73
	2	5	0.71	0.73		2	5	0.68	0.76
	5	2	0.70	0.72		5	2	0.75	0.79
	5	3	0.71	0.72		5	3	0.76	0.77
	5	4	0.72	0.74		5	4	0.77	0.84
	5	5	0.74	0.79		5	5	0.77	0.88
U-shaped $f(x)=3x^2/2x\in[0, 1]$	2	2	0.54	0.68	*Lognormal(0,1)*	2	2	0.86	0.83
	2	3	0.56	0.69		2	3	0.86	0.82
	2	4	0.56	0.69		2	4	0.88	0.85
	2	5	0.59	0.69		2	5	0.88	0.85
	5	2	0.61	0.74		5	2	0.89	0.87
	5	3	0.69	0.74		5	3	0.88	0.87
	5	4	0.69	0.82		5	4	0.89	0.90
	5	5	0.71	0.84		5	5	0.89	0.90

Remark: When data are not Normally distributed, the RSS-ANOVA had a better performance than the classic SRS procedure. It can be due to the convergence of linear rank statistics to normality.

Table 2
Values of *V(mean)* under different values of ρ and a joint Uniform distribution.

Distribution	n	k	$\rho=0$	$\rho=0.05$	$\rho=0.50$	$\rho=0.75$	$\rho=0.90$	$\rho=0.95$
Uniform (0,2)	2	2	1.07	1.59	3.43	4.85	5.55	6.77
	2	3	1.06	1.59	3.43	4.88	5.59	6.77
	2	4	1.07	1.59	3.42	4.89	5.80	7.08
	2	5	1.07	1.59	3.44	4.94	5.85	7.28
	5	2	1.08	1.59	3.46	4.94	5.84	7.39
	5	3	1.08	1.59	3.45	4.94	5.85	7.42
	5	4	1.06	1.65	3.47	4.96	5.87	7.48
	5	5	1.08	1.65	3.48	4.98	5.89	7.55

Table 3
Values of *V(mean)* under different values of ρ and a joint Normal distribution.

Distribution	n	k	$\rho=0$	$\rho=0.05$	$\rho=0.50$	$\rho=0.75$	$\rho=0.90$	$\rho=0.95$
Normal (0,1)	2	2	1.07	1.15	8.63	8.91*	8.88*	9.09*
	2	3	1.06	1.15	8.73	8.98*	8.98*	9.17*
	2	4	1.07	1.14	8.72*	8.18*	8.78*	9.28*
	2	5	1.07	1.11	8.72*	8.78*	8.79*	9.28*
	5	2	1.02	1.11	8.74*	8.78*	8.84*	9.39*
	5	3	1.02	1.11	8.65*	9.07*	8.95*	9.47*
	5	4	1.06	1.13	8.79*	9.16*	9.07*	9.58*
	5	5	1.02	1.13	8.81*	9.28*	9.59*	9.75*

Table 4
Values of *V(mean)* under different values of ρ and a joint Exponential distribution.

Distribution	n	k	$\rho=0$	$\rho=0.05$	$\rho=0.50$	$\rho=0.75$	$\rho=0.90$	$\rho=0.95$
Exponential(1)	2	2	1.07	2.25	3.73	5.05	6.18	6.66
	2	3	1.06	2.25	3.73	5.18	6.18	6.67
	2	4	1.07	2.24	3.72	5.15	6.28	6.68
	2	5	1.07	2.22	3.77	5.35	6.39	6.68
	5	2	1.02	2.22	3.76	5.55	6.44	6.69
	5	3	1.02	2.22	3.75	5.57	6.55	6.77
	5	4	1.06	2.23	3.77	5.56	6.57	6.88
	5	5	1.02	2.23	3.78	5.58	6.59	6.85

Table 2 exhibits that for values of $\rho=0$, the value of *V(mean)* are the smallest and then, its values are increased seriously for $\rho \geq 0.50$.

The results in Table 3 give a better idea of the effect of the correlation in detecting the non-random ordering of *Y*. The values of *V(mean)* which are significant are marked with a '*'. For ρ=*0.50* the significance is accepted for the pairs {(2,4), (2,5), (5,2), (5,3), (5,4), (5,5)}. The non-randomness of the ranking is accepted in all the cases for $\rho \geq 0.75$.

Table 4 gives an idea that for the Exponential distribution for highly correlated variables, the value of *V(mean)* is expected to be larger than 3.

For the Gamma with density function $x^4 exp(-x)/\Gamma(5)$ and Lognormal distributions, see Tables 5 and 7, the values of *V(mean)* are expected to be close to 2 for $\rho < 0.50$.

Table 6 sustains a similar result for $\rho > 0.75$ in the case U-shaped distribution with density function $f(x)=3x^2/2$ $x \in [0,1]$.

4.2. Analysis of the time to death of HIV infected persons

We have considered a database of the lifetime of a set of 231 persons infected with HIV clustered by the risk-group, It constituted a population.

G1- Drug users
G2- Bisexual-homo men
G3- Bisexual-lesbian women
G4- Hetero men
G5- Hetero women
G6- Contaminated by blood transfusions
G7- Sons of VIH infected women
G8- Unknown

We selected 1000 independent (runs) samples from the data set to estimate treatment effects and compared them with the effect calculated with the population data. The estimated variance of treatment effects for model using SRS and RSS for each group (i = 1, . . . ,8) computed as

$$\hat{V}(\alpha_i^*) = \frac{1}{1000}\sum_{t=1}^{1000}\left(\alpha_{ij}^* - \alpha_i\right)^2 = E(i;\, srs) \quad (4.2)$$

Table 5
Values of *V(mean)* under different values of ρ and a joint Gamma distribution.

Distribution	n	k	$\rho=0$	$\rho=0.05$	$\rho=0.50$	$\rho=0.75$	$\rho=0.90$	$\rho=0.95$
Gamma $x^4 exp(-x)/\Gamma(5)$ $x>0$	2	2	1.17	2.69	6.06	6.67	6.61	7.44
	2	3	1.16	2.69	6.03	6.66	6.67	7.76
	2	4	1.17	2.69	6.06	6.69	6.71	7.79
	2	5	1.17	2.69	6.00	6.19	6.76	7.73
	5	2	1.19	2.76	6.06	6.19	6.77	7.78
	5	3	1.19	2.75	6.00	6.61	6.75	7.87
	5	4	1.20	2.82	6.00	6.66	6.77	7.86
	5	5	1.21	2.79	6.08	6.33	6.77	7.87

Table 6
Values of *V(mean)* under different values of ρ and a joint U-shaped distribution with densities function $f(x)=3\times^2/2$ $x \in [0,1]$.

Distribution	n	k	$\rho=0$	$\rho=0.05$	$\rho=0.50$	$\rho=0.75$	$\rho=0.90$	$\rho=0.95$
U-shaped $f(x)=3x^2/2$ $x \in [0, 1]$	2	2	1.15	2.27	3.43	3.85	4.44	5.55
	2	3	1.16	227	3.43	3.88	4.49	5.57
	2	4	1.15	2.27	3.42	3.89	4.30	5.58
	2	5	1.15	2.27	3.44	3.93	4.34	5.58
	5	2	1.18	2.27	3.46	3.93	4.34	5.59
	5	3	1.18	2.27	3.45	3.93	4.35	5.42
	5	4	1.16	2.30	3.47	3.96	4.37	5.48
	5	5	1.08	2.30	3.48	3.98	4.39	5.55

Table 7
Values of *V(mean)* under different values of ρ and a joint Lognormal distribution.

Distribution	n	k	$r=0$	$r=0.05$	$r=0.5$	$r=0.75$	$r=0.90$	$r=0.95$
Lognormal(0,1)	2	2	1.17	2.22	3.23	4.55	5.55	6.11
	2	3	1.11	2.22	3.23	4.55	5.59	6.10
	2	4	1.17	2.22	3.23	4.54	5.58	6.18
	2	5	1.17	2.22	3.23	4.44	5.75	6.18
	5	2	1.18	2.22	3.22	4.44	5.77	6.26
	5	3	1.18	2.24	3.28	4.44	5.75	6.22
	5	4	1.17	2.32	3.27	4.46	5.77	6.36
	5	5	1.18	2.33	3.28	4.45	5.50	6.40

Table 8
Efficiencies of the estimates of the treatment effects in 8 groups of persons infected with HIV positive (variable time to death in years) using SRS and RSS.

Group	$\hat{V}(\alpha_i^*)$	$\hat{V}(\alpha_{(i)}^*)$
1	1.97	1.24
2	1.52	2.78
3	1.20	2.19
4	1.05	2.11
6	1.85	1.72
6	1.45	1.11
7	1.44	1.79
8	1.94	1.30

$$\hat{V}(\alpha_{(i)}^*) = \frac{1}{1000}\sum_{t=1}^{1000}\left(\alpha_{(i)j}^* - \alpha_i\right)^2 = E(i;\, rss) \tag{4.3}$$

We have computed the ratio of the efficiencies and the results are given in Table 8. Note that both of them are considerable larger for the use of SRS. These results illustrate the behaviour of RSS as alternative for estimating the treatment effects and the variability of them. Due to the nature of the data non-normality was present, hence the use of ANOVA for fixing the existence of significance of the observed differences did not make sense.

5. Conclusion

The use of RSS in ANOVA is at least as good as the SRS methodology. This result supports that RSS designed experiments can be analysed using One Way ANOVA. The estimation of the variance using RSS allows establishing the closeness of the ranking to the perfect ranking, assumed in the modeling. *V(n)* is a non-parametric statistics that can be used for analyzing the quality of the ranking. A further study of it is needed to establish rules for evaluating the relative precision of RSS as a function of the quality of the ranking.

Acknowledgements

Authors are thankful to Editor-in-Chief Prof. Shally Awasthi and anonymous referees for their constructive comments towards the improvement of this paper. Authors are also thankful to Program MÊS-CUBA EDITAL 46/2013 project 209/13 for providing necessary support to carry out the present research work.

References

1. Chen Z, Bai Z, Sinha B. *Ranked set sampling: theory and applications*, vol. 76Springer Science & Business Media; 2003.
2. Dell TR, Clutter JL. Ranked set sampling theory with order statistics background. *Biometrics*. 1972;28(2):545–555.
3. Frey J. A note on ranked-set sampling using a covariate. *J Stat Plan Inference*. 2011;141(2):809–816.
4. MacEachern SN, Öztürk Ö, Wolfe DA, Stark GV. A new ranked set sample estimator of variance. *J R Stat Soc Ser B (Stat Methodol)*. 2002;64(2):177–188.
5. McIntyre GA. A method for unbiased selective sampling, using ranked sets. *Aust J Agric Res*. 1952;3(4):385–390.
6. Muttlak HA. Median ranked set sampling with concomitant variables and a comparison with ranked set sampling and regression estimators. *Environmetrics*. 1998;9(3):255–267.
7. Patil GP, Sinha AK, Taillie C. Finite population corrections for ranked set sampling. *Ann Inst Stat Math*. 1995;47(4):621–636.
8. Perron F, Sinha BK. Estimation of variance based on a ranked set sample. *J Stat Plan Inference*. 2004;120(1–2):21–28.
9. Stokes SL. Estimation of variance using judgment ordered ranked set sample. *Biometrics*. 1980;36:35–42.
10. Takahasi K, Wakimoto K. On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann Inst Stat Math*. 1968;20(1):1–31.