Journal of Scientific Research & Reports
8(6): 1-7, 2015; Article no.JSRR.19675
ISSN: 2320-0227

SCIENCEDOMAIN *international*
www.sciencedomain.org

SCIENCEDOMAIN

An Alternative Mean Estimator for Ranked Set Sample

Özlem Ege Oruç[1*]

[1]Department of Statistics, Dokuz Eylül University, Faculty of Sciences, Turkey.

Author's contribution

The sole author designed, analyzed and interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/JSRR/2015/19675
Editor(s):
(1) Shunpu Zhang, Department of Statistics, University of Nebraska – Lincoln, USA.
Reviewers:
(1) Anonymous, Prague, Czech Republic.
(2) Anonymous, Slippery Rock University of PA, USA.
Complete Peer review History: http://sciencedomain.org/review-history/10618

Original Research Article

Received 23rd June 2015
Accepted 31st July 2015
Published 19th August 2015

ABSTRACT

Aims: We introduce a new estimator for population mean by using coefficient of variation as prior information in ranked set sampling (RSS). Then we compare it with the estimator of the mean in RSS, the estimator of the mean in simple random sampling (SRS) in the sense of mean square error (MSE). We conclude that the proposed RSS mean estimator is more efficient than the aforementioned estimators.
Study Design: This was rank set sampling, improved estimation study.
Place and Duration of Study: Dokuz Eylul University Department of Statistics between December 2014 and June 2015.
Methodology: In this study, We introduce a new estimator for population mean by using coefficient of variation as prior information in Ranked Set Sampling (RSS). The performance of this estimator is compared in the sense of mean square error (MSE).
Results: When we compared the improved RSS mean estimator, SRS mean estimator and traditional RSS mean estimator in the sense of mean square error. We conclude that the proposed RSS mean estimator is more efficient than the aforementioned estimators.
Conclusion: We have shown that a biased estimator with a smaller MSE can be obtained by using a priori information which is the coefficient of variation. To compare the efficiencies of the mean estimators with each other, we evaluate the relative efficiencies of each estimator using MSE. It is shown that the proposed mean estimator for RSS is more efficient than the conventional estimators. In particular the better efficiencies are obtained for small sample sizes.

**Corresponding author: Email: ozlem.ege@deu.edu.tr;*

Keywords: Improved estimator; ranked set sample; coefficient of variation; mean squared error.

1. INTRODUCTION

Ranked set sampling is an alternative to simple random sampling that can sometimes offer large improvements in precision. In many applications it is very difficult or expensive to collect the sampling data, but the data may be ranked without any cost. In such cases the use of RSS, which originally proposed in connection with estimating herbage yield in McIntyre [1] gives better estimate of the population mean compared to the simple random sampling(SRS). It was re-discovered by Takahasi & Wakimoto [2], who developed the underlying theory of the method. The RSS is a method of collecting data that improves estimation by utilizing the samplers judgment or auxiliary information about the relative sizes of the sampling units. In practice, ranking is bound to be performed with some error. Dell & Clutter [3] showed that the RSS mean estimator remains unbiased in the presence of ranking errors, and that when ranking is completely random, the RSS estimator has the same precision as the SRS estimator. RSShas been successfully applied in medicine, census data and environmental research, and many other applications, Chen et al. [4], Muttlak and McDonald [5],Johnson et al. [6] and Patil and Taillie [7]. Barabesi and Fattorini [8], Sengupta and Mukhuti [9] and Kadilar et al. [10].

Improved estimation is a very important concept in statistical inference. The main aim of this estimation method is to examine the conditions under which biased estimators can lead to an improvement over the conventional unbiased procedures. Given additional information such as coefficient of variation, kurtosis or skewness, the problem has been studied extensively. The population coefficient of variation is usually fairly stable over time and characteristics of similar nature and its value may be known see Murthy [11]. Govindarajulu and Sahai [12] pointed out that in many life sciences and biological experiments, the observations have normal distribution with known coefficient of variation. For illustration, in clinical laboratory experiments the routine procedures are repeated often enough so that coefficient of variation is known for practical purposes. Searls[13], Khan [14] and Arnholt and Hebert [15], Unsal and Ege Oruc [16], Srisodaphol and Tongmol [17] employed the known coefficient of variation for the population mean related with the improved estimator. They were also of the opinion that a simple type of priori information usually available to the experimenter is the coefficient of variation, particularly those in the biological fields through long association with their experimental material. Similarly, we assume that the value of the coefficient of variation is quite accurately known.

This work is concerned with the improved estimator of the population mean by using coefficient of variation as prior information in RSS. It is known that one important way of improving estimation of distribution parameters is by applying biased estimation procedures. The mean square error (MSE) is the most popular and widely used criterion to show when a biased estimator is better than an unbiased estimator. Under the assumption that the coefficient of variation is known a priori, we derive improved estimators for the mean in RSS with minimum MSE.

The paper is organized as follows. First, we provide an overview of ranked set sampling and improved estimator for mean. Then we propose an alternative mean estimator for RSS and evaluate its efficiency. Finally, we demonstrate the performance of mean estimators $\overline{X}_{RSS}$, $\overline{X}^*_{RSS}$ and $\overline{X}_{SRS}$ with an application on a set of real data in biological sciences.

2. MATERIALS AND METHODS

2.1 An Alternative Mean Estimator for RSS

The RSS procedure involves randomly drawing *k* sets each of size *k* from the target population. Larger values of *k* convey more information about the population but they also more likely contain ranking errors. Because of this reason, in classical RSS practice, *k* usually takes values such as 2, 3, 4 or 5. The members of each random set can be ranked with respect to the characteristic of the study variable or auxiliary variable. From the first set of *k* units, the smallest ranked is measured. From the second set of *k* units, the second lowest ranked unit is measured, so on. This procedure describes one cycle of the RSS process. We can repeat the whole procedure *m* times to get a RSS of size *n=km*. After ranking all of the study variables, we set $X_{(i)j}$,$i = 1,... k; j = 1,...,m$ denote the i^{th} order statistics in the j^{th} cycle. Note that the rank order statistics, $X_{(i)j}$ for a fixed i, are independent identically distributed (i.i.d) with mean μ_i and variance σ_i^2. For RSS, we can write the estimator of mean $\overline{X}_{RSS}$ as

$$\overline{X}_{RSS} = \frac{1}{k}\sum_{i=1}^{k}\frac{1}{m}\sum_{j=1}^{m} x_{(i)j} = \frac{1}{m}\sum_{j=1}^{m}\overline{x}_{jRSS} \tag{1}$$

where $\overline{X}_{jRSS} = \frac{1}{k}\sum_{i=1}^{k} x_{(i)j}$, $j=1,...,m$ are i.i.d. with mean μ and variance $\tau^2 = \frac{1}{k^2}\sum_{i=1}^{k}\sigma_{(i)}^2$.

The variance of $\overline{X}_{RSS}$ is given by Chen [3] is

$$Var\left[\overline{X}_{RSS}\right] = \frac{1}{mk}\left[\sigma^2 - \frac{1}{k}\sum_{i=1}^{k}\left(\mu_{(i)} - \mu\right)^2\right] = Var\left[\overline{X}\right] - \frac{1}{mk^2}\sum_{i=1}^{k}\left(\mu_{(i)} - \mu\right)^2 \tag{2}$$

where σ^2 is population variance.

Dell and Cutter [5] established that RSS method yields an unbiased estimate provided that ranking errors are not related to the process of selecting elements for quantification. This work assumes a similar condition. Searls [13] defined an improved estimator of the population mean using SRS as

$$\overline{X}_{SRS}^{*} = \frac{1}{\beta}\sum_{i=1}^{n} x_i \text{ , where } \beta = \frac{1}{n+v^2} \tag{3}$$

which can be computed using the coefficient of variation $v = \frac{\sigma}{\mu}$. It can be readily seen that $MSE\left(\overline{X}_{SRS}^{*}\right)$ is always less than $MSE\left(\overline{X}_{SRS}\right)$.

It is the aim of this paper to investigate a similar concept using the rank set sampling. First we write the following estimator for the population mean:

$$\overline{X}_{RSS}^{*} = \alpha\sum_{j=1}^{m}\overline{x}_{jRSS}$$

where $\overline{X}_{jRSS}$, $j=1,...,m$ are i.i.d. with mean μ and variance $\tau^2 = \frac{1}{k^2}\sum_{i=1}^{k}\sigma_{(i)}^2$.

where the constant α is chosen to minimize $MSE\left(\overline{X}_{RSS}^{*}\right)$. Then it follows that

$$MSE\left(\overline{X}_{RSS}^{*}\right) = Var\left(\overline{X}_{RSS}^{*}\right) + \left[E\left(\overline{X}_{RSS}^{*}\right) - \mu\right]^2$$

where

$$Var\left(\overline{X}_{RSS}^{*}\right) = Var\left(\alpha\sum_{j=1}^{m}\overline{x}_{jRSS}\right) = \alpha^2 \mathrm{m}\tau^2$$

Then we have

$$MSE\left(\overline{X}^{*}_{RSS}\right)=\alpha^{2}m\tau^{2}\ +\mu^{2}\left(\alpha m\text{-}1\right)^{2} \tag{4}$$

We require the first and second derivative of the last equation:

$$\frac{d\left[MSE\left(\overline{X}^{*}_{RSS}\right)\right]}{d\alpha}=2m\left(\alpha\tau^{2}+\alpha\mu^{2}m-\mu^{2}\right) \tag{5}$$

and

$$\frac{d^{2}\left[MSE\left(\overline{X}^{*}_{RSS}\right)\right]}{d\alpha^{2}}=2m\left(\tau^{2}+\mu^{2}m\right) \tag{6}$$

Since the second derivative has to be non-negative, the minimizing value of α is obtained from the first derivative (5)

$$\alpha=\frac{\mu^{2}}{\tau^{2}+m\mu^{2}}$$

The fact that $v^{2}=\dfrac{\tau^{2}n(1+m)}{\mu^{2}m}$ yields $\alpha=\dfrac{a}{m\left(v^{2}+a\right)}$ where $a=n(1+m)$. Using this value we are obtained an alternative mean estimator for RSS

$$\overline{X}^{*}_{RSS}=\frac{a}{m\left(v^{2}+a\right)}\sum_{j=1}^{m}\overline{x}_{jRSS} \tag{7}$$

We can easily find the variance and MSE of this estimator

$$Var\left(\overline{X}^{*}_{RSS}\right)=\frac{a^{2}\tau^{2}}{m\left(v^{2}+a\right)^{2}} \quad \text{and} \quad MSE\left(\overline{X}^{*}_{RSS}\right)=\frac{2a^{2}\tau^{2}}{m\left(v^{2}+a\right)^{2}} \tag{8}$$

3. RESULTS AND DISCUSSION

3.1 Mean Square Error Comparisons

The mean square error (MSE) is the most widely used criterion to determine that a biased estimator is better than an unbiased estimator. In this section, using the MSE we compare the estimation (7) with the traditional mean estimator for RSS, with that of SRS. Since traditional RSS mean estimator is an unbiased estimator of population mean then the $MSE\left(\overline{X}_{RSS}\right)=Var\left(\overline{X}_{RSS}\right)$ .We now rewrite the $MSE\left(\overline{X}_{RSS}\right)$ as $MSE\left(\overline{X}_{RSS}\right)=\dfrac{\tau^{2}}{m}$ where $\tau^{2}=\dfrac{1}{k^{2}}\sum_{i=1}^{k}\left(\mu_{(i)}-\mu\right)^{2}$.

The relative efficiency (RE) with respect to the traditional RSS mean, $\overline{X}_{RSS}$ may be expressed as

$$RE\left(\overline{X}_{RSS},\overline{X}^{*}_{RSS}\right)=\frac{m\left(v^{2}+a\right)^{2}}{2a^{2}}$$

Next we examine the above formula for several values of m,k and v^{2} in Table 1.

Table 1. Relative efficiencies of $\overline{X}_{RSS}$ and $\overline{X}^{*}_{RSS}$ for different values of m, k and v^2

$RE\left(\overline{X}_{RSS}, \overline{X}^{*}_{RSS}\right)$	**Set size(k) and cycle size(m)**				
v^2	k=3, m=3	k=3, m=5	k=3, m=7	k=5, m=5	k=5, m=7
0.75	1,547240	2,531340	3,523470	2,518785	3,514077
1	1,584491	2,555864	3,541791	2,533444	3,525045
5	4,306713	4,081790	4,619172	3,402778	4,152902
10	21,40741	11,14198	8,906746	6,944444	6,446429

We observe that when the coefficient of variation is increasing the efficiency of the new estimator increases. Moreover the new estimator become more effective than $\overline{X}_{RSS}$ for small set size and cycle size. The largest relative efficiency gains are obtained for small sample sizes. This is the case in practice for example when observations are expensive, such sample sizes may be all that are available. When we investigate the relative efficiency of $\overline{X}_{SRS}$ with $\overline{X}^{*}_{RSS}$ we have

$$RE\left(\overline{X}_{SRS}, \overline{X}^{*}_{RSS}\right) = \frac{\left(v^2 + a\right)^2}{2na}$$

The last equation may be regarded as an alternative to RSS mean estimator, it is seen that the MSE of $\overline{X}^{*}_{RSS}$ is always smaller than the that of $\overline{X}_{SRS}$ unbiased mean estimator. The relative efficiencies of $\overline{X}_{SRS}$ and $\overline{X}^{*}_{RSS}$ for several values values of m, k and v^2 are given in Table 2.

Table 2. Relative efficiencies of $\overline{X}_{SRS}$ and $\overline{X}^{*}_{RSS}$ for different values of m, k and v^2

$RE\left(\overline{X}_{SRS}, \overline{X}^{*}_{RSS}\right)$	**Set size(k) and cycle size(m)**				
v^2	k=3, m=3	k=3, m=5	k=3, m=7	k=5, m=5	k=5, m=7
0.75	2,062982	3,037617	4,026831	3,022542	4,016088
1	2,112654	3,067037	4,047761	3,040133	4,028622
5	5,742284	4,898148	5,279053	4,083333	4,746173
10	28,54321	13,37037	10,17914	8,333333	7,367347

The Table 2 demonstrates an interesting fact. If the the coefficient of variation v^2 is increasing then the proposed RSS mean estimator $\overline{X}^{*}_{RSS}$ is more efficient than $\overline{X}_{SRS}$. Moreover the proposed estimator becomes more efficient than $\overline{X}_{SRS}$ for small *k*.

We use the data from the work of Murray et.al. [18]. In the data, leaves were sampled haphazardly, that is without any intentional positional bias but equally not according to a formal randomization scheme. Then single observer ranked the leaves within each set based on the visual appearance of the deposits on the the upper leaf surface when view under ultraviolet light. In this application we use the population which contains 125 leaves from the data. The descriptive statistics of data are given in Table 3. We then evaluate the performance of the estimators $\overline{X}_{RSS}$, $\overline{X}^{*}_{RSS}$ and $\overline{X}_{SRS}$ in estimating mean spray deposits on leaves.

To illustrate how to apply RSS to data, we select a sample whose size is *125* from the population by the simple random sampling without replacement. These data are grouped into sets each of size *k* = *5* and we repeat it *m* = *5* times. According to the RSS methodology, we use ascending order in the *x* values and assume that there is no judgement error in this ordering. Then, the smallest unit is selected from the first ordered set, the second smallest unit is selected from the second ordered set, and so on. By this way, we select *25* observations *n=mk* = *25* and we estimate the population mean by applying the estimator to these observations using the equations (1), (7) and ordinary mean estimator for SRS.

Table 3. Descriptive statistics of data

N	125
Mean	0,1768
Standard deviation	0,1544
Coefficient of variation	0,8731
Minimum	0,009
Maximum	0,729

It can be seen from the Table 4 that we obtain the closest population mean estimation using the proposed estimator. Moreover the MSE value of the proposed estimator is smaller than the other estimators. Therefore the application from real data supports the theoretical results. With this proposed estimator we have a considerable benefit in efficiency over the conventional estimators.

Table 4. Population mean estimation

Estimators	Estimation	MSE values
$\overline{X}_{SRS}$	0,1949	0,000954
$\overline{X}_{RSS}$	0,1948	0,000658
$\overline{X}^{*}_{RSS}$	0,1854	0,000397

4. CONCLUSION

This paper is concerned with a new improved estimator of the population mean for RSS. We have shown that a biased estimator with a smaller MSE can be obtained by using a priori information which is the coefficient of variation. To compare the efficiencies of the mean estimators with each other, we evaluate the relative efficiencies of each estimator using MSE. It is shown in this paper that the proposed mean estimator for RSS is more efficient than the conventional estimators. In particular the better efficiencies are obtained for small sample sizes. Ranked set sampling (RSS) and the assumption of a known coefficient of variation are actually common in many biological, environmental and agricultural applications. In these areas, the measurements of the units according to variables of interest can be quite difficult in some cases, in terms of cost, time and other factors. In such conditions, by using the ranked set sampling, the sample selection process is done by less cost and in less time, than the simple random sampling (SRS) technique. Our proposed estimator is very suitable for such conditions.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. McIntyre GA. A method of unbiased selective sampling using ranked sets. Austr. J. Agri. Res.1952;3:385-390.
2. Takahasi K, Wakimoto K. On unbiased estimates of the population mean based on the sample strati_ed by means of ordering, Annals of the Institute of Statistical Mathematics. 1968;20:1-31.
3. Dell TR, Clutter JL. Ranked set sampling theory with order statistics background, Biometrics. 1972;28:545-555.
4. Chen Z, Bou Z, Sinha B. Ranked Set Sampling: Theory and Methods, Springer, New York; 2004.
5. Muttlak HA, McDonald LL. Ranked set sampling with respect to concomitant variables and with size biased probability of selection. Communications in Statistics Theory and Methods A. 1993;19:205-219.
6. Johnson GD, Patil GP, Sinha AK. Ranked set sampling for vegetation research. Abstracta Botanica. 1993;17:87-102.
7. Patil GP, Taillie C. Environmental sampling, observational economy and statistical inference with emphasis on ranked set sampling, encounter sampling and composite sampling, Bulletin ISI Proceedings of 49 Session, Firenze. 1993; 295-312.
8. Barabesi L, Fattorini L. Kernel estimators of probability density functions by ranked-set sampling, Communications in Statistics Theory and Methods A. 2002;31(4):597-610.
9. Sengupta S, Mukhuti S. Unbiased estimation of Pr(X > Y) using ranked set sample data, Statistics. 2008;42(3):223-230.
10. Kadilar C, Unyaz CY, Cingi H. Ratio estimator for the population mean using ranked set sampling, Statistical Papers. 2009;50(2):301-309.
11. Murthy MN. Sampling theory and methods, Statistical Publishing Society, Calcutta, India; 1967.
12. Govindarajulu Z, Sahat H. Estimation of the parameters of a normal distribution

with known coe_cient of variation. Stat. Appl. Res., JUSE, Section A. 1995;19(3): 85-98.
13. Searls DT. The utilization of coe_cient of variation in the estimation procedure. Journal of the American Statistician Association. 1965;59:1225-1226.
14. Khan RA. A note on estimating the mean of a normal distribution with known coe_cient of variation, Journal of the American Statistical Association.1968;63: 1039-1041.
15. Arnholt AT, Hebert JL. Estimating the mean with known coefficient of variation, The American Statistician. 1995;49: 367-369.
16. Ünsal A, Ege Oruç Ö. A note on the shrinkage techniques for estimating the mean, Scientific Research and Essays. 2011;6(12):2501-2503.
17. Srisodaphol W, Tongmol N. Improved Estimators of the Mean of a Normal Distribution with a Known Coe_cient of Variation, Journal of Probability and Statistics. 2012;1-5 .
18. Murray RA, Ridout MS, Cross JV. The use of ranked set sampling in spray deposit assessment, Aspect of Applied Biology. 2000;57:141-146.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/10618