International Journal of Statistics and Applied Mathematics 2022; 7(2): 63-67

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2022; 7(2): 63-67

www.mathsjournal.com
Received: 11-01-2022
Accepted: 15-02-2022

Khalid Ul Islam Rather
Division of Statistics and Computer Science, Faculty of Basic Sciences, Main Campus SKUAST-J, Jammu, Jammu and Kashmir, India

SEH Rizvi
Division of Statistics and Computer Science, Faculty of Basic Sciences, Main Campus SKUAST-J, Jammu, Jammu and Kashmir, India

Manish Sharma
Division of Statistics and Computer Science, Faculty of Basic Sciences, Main Campus SKUAST-J, Jammu, Jammu and Kashmir, India

M Iqbal Jeelani
Division of Statistics and Computer Science, Faculty of Basic Sciences, Main Campus SKUAST-J, Jammu, Jammu and Kashmir, India

Corresponding Author:
Khalid Ul Islam Rather
Division of Statistics and Computer Science, Faculty of Basic Sciences, Main Campus SKUAST-J, Jammu, Jammu and Kashmir, India

Optimum stratification for equal allocation using ranked set sampling as a method of selection

Khalid Ul Islam Rather, SEH Rizvi, Manish Sharma and M Iqbal Jeelani

DOI: https://doi.org/10.22271/maths.2022.v7.i2a.809

Abstract
Present investigation deals with the problem of optimum stratification on the auxiliary variable for equal allocation when method of selection of units in individual strata is done by ranked set sampling (RSS). A $cum\sqrt{K_4(x)}$ rule of obtaining approximate optimum strata boundaries has been developed by minimizing the variance and solution of minimal equations thereof. A numerical analysis on the basis of percent relative efficiency has also been made, which indicated remarkable gain in efficiency.

Keywords: Stratified ranked set sampling, approximately optimum stratification, equal allocation

Introduction

In sample survey, the precision of an estimator depends on the homogeneity of the units of the population besides the sample size and sampling fraction, the role of stratified sampling technique comes into play as one of the possible ways to enhance the precision of the estimator. The optimum stratum boundaries (OSB) are the ones that resemble to the least amount of variance for any given allocation method. Dalenius (1950) [1] was the first to explore the objective of optimum stratification on the response variable for proportional and Neyman allocation methods, followed by Hayashi and Maruyama (1948) [7]. Singh and Sukhatme (1969) [11] considered the problem of obtaining approximate optimum strata boundaries (AOSB) using auxiliary information. Further problem of optimum stratification for equal allocation under simple random sampling (SRS) has been done by Singh and Parkash (1973) [12]. Rizvi, *et al.* (2000) [9] developed the theoretical frame work for determination of AOSB on a auxiliary variable. Rizvi, *et al.* (2002) [10] developed a methodology for obtaining AOSB using auxiliary information under compromise method of allocation. Danish, *et al.* (2017) [3] studied the problem of obtaining OSB in proportional allocation by changing cost of each unit. Danish and Rizvi (2018) [4] proposed a technique under Neyman allocation and recently, Gupt *et al.* (2022) [6] considered problem of optimum stratification of heteroscedastic populations in stratified sampling for a known allocation under SRS strategy.

McIntyre (1952) [8] introduced the concept of ranked set sample (RSS). Samawi (1996) [13] defined stratified ranked set sample (SRSS) as a sampling scheme in which a population is partitioned into L mutually exclusive and exhaustive strata, with each stratum containing a ranked set RSS of n items. The strata are sampled independently of one another. As a result, an SRSS scheme may be thought of as a collection of L independent ranked set samples.

If the population under consideration be alienated into L strata and a sample $n_{0h} = (R_h \times n_h)$ units is selected from h^{th} stratum by using RSS, where R_h be the number of cycles and n_h be the sample size of each cycle and each sample element is assessed in relation to a variable Y, then estimate of population mean under stratified ranked set sampling estimate is given by

$$\bar{y}_{SRSS} = \sum_{h=1}^{L} \frac{W_h}{n_{0h}} \left[\sum_{j=1}^{R_h} \sum_{r=1}^{n_h} \bar{y}_{ij(r)}\right] \tag{1.1}$$

where W_h is the weight of units in the h^{th} stratum and $\bar{y}_{ij(r)}$ is the sample mean based on n_{0h} units drawn from the h^{th} stratum.

If finite population correction is overlooked, then the minimization of the variance expression

$$\frac{1}{n}\left[\sum_{h=1}^{L} w_h^2 \sigma_{h(r)}^2 + \mu_{h\eta}\right] \text{ is corresponding to the minimization of } \left[\sum_{h=1}^{L} W_h^2 \sigma_{h(r)}^2\right] \quad (1.2)$$

where $\sigma_{h(r)}^2$represents the variance of r^{th} order statistics in h^{th} stratum of the random sample size n_h.
The variance in (1.2) is a function of the stratum boundaries for given L, n, and the variable y. Dalenius and Gurney (1951) [2] developed one of the several approaches for determining AOSB on the study variable y. In the present investigation we deal with the problem of determining AOSB for equal allocation using ranked set sampling in individual stratum as a method of selection.

Minimal equations under equal allocation
Let the regression of the estimation variable $'Y'$ on the stratification variable$'X'$, in the infinite super population is specified by

$$y = \varphi(x) + e \quad (2.1)$$

where $\varphi(x)$ is a function of auxiliary variable, e is the error term such that $E(e|x) = 0$ and $V(e|x) = \eta(x) > 0 \forall x \in (a, b)$ with$(b - a) < \infty$. Assume $f(x, y)$ be the joint density function of (x, y) and $f(x)$ be the marginal density function ofx. Then, we have the following relations:

$$W_h = \int_{x_{h-1}}^{x_h} f_i(x),$$

$$\mu_{hc} = \frac{1}{w_h}\int_{x_{h-1}}^{x_h} \varphi(x) f_i(x)$$

$$\text{and } \sigma^2{}_{hy} = \sigma^2{}_{hc} + \mu_{h\eta} \ (h = 1,2,3\ldots, L) \quad (2.2)$$

where (x_{h-1}, x_h) are lower and upper boundaries of the h^{th} stratum with $x_0 = a$ and $x_L = b$, $\mu_{h\eta}$ is the expected value of $\eta(x)$ and σ_{hc}^2 is the variance of $\varphi(x)$ in the h^{th} stratum.
By these relations, variance expression are reduced to

$$V(\bar{y}_{SRSS})_{Eq} = \left[\sum_{h=1}^{L} W_h^2 \sigma_{h(r)}^2\right] \quad (2.3)$$

$$\sigma_{h(r)}^2 = \left(\sigma^2{}_{hc} - \frac{1}{n}(\mu_i - \mu)^2\right)$$

Let $[x_h]$ represent the set of optimum stratification points on the range (a, b) for which the $V(\bar{y}_{SRSS})$ is the smallest. These points $[x_h]$ are the solutions of minimal equations', which are derived by equating to zero, the partial derivatives of $V(\bar{y}_{SRSS})$ with respect to $[x_h]$. We will now get this equal allocation minimal equation.
To get the minimal equations for this allocation method, we differentiate the variance expression given in (2.3) w.r.to$'x_h'$, and get minimal equations as

$$W_h\left\{\left(\varphi(x_h) - \mu_{hc(r)}\right)^2 - \sigma_{hc(r)}^2 + \xi(x_h) - \mu_{h\eta}\right\}$$

$$= W_i\left\{\left(\varphi(x_h) - \mu_{ic(r)}\right)^2 - \sigma_{ic(r)}^2 + \xi(x_h) - \mu_{i\eta}\right\} \quad (2.5)$$

where $i = h + 1, h = 1,2, \ldots, L - 1$

These equations are also complicated to solve and therefore could be used to find stratification points. Further, better approximation can be obtained by using some approximate iterative procedures.

Solutions of the minimal equations
In this section, we intend to find the series expansions of the system of equations given in (2.5) about the point $[x_h]$, the common boundary of h^{th} and $(h + 1)^{th}$ strata and obtain their approximated solutions. To find the expressions on left hand side of (2.5) we shall use the relations obtained in different lemmas by replacing (y, x) by (x_{h-1}, x_h), and for the right side corresponding relations after replacing (y, x) by (x_{h-1}, x_h) will be used Singh *et, al.* (1969) [11].
The corresponding expansion for the left hand side of the equation can be obtained from the expansion of right side by merely varying the symbols of the coefficients of odd powers of k_i where$k_i = (x_{h+1}, x_h)$, although the same result will be obtained, we have

$$\mu_\eta(y,x) = \xi\left[1 + \frac{\xi'}{2\xi}k + \frac{(\xi' f' + 2f\xi'')}{12f\xi}k^2 + \frac{(ff''\xi' + ff'\xi'' + f^2\xi''' - \xi' f'^2)}{24f^2\xi}k^3 + O(k^4)\right]$$

$O(k^i)$ is the higher order terms with power $\geq$ i
After simplification, we get

$$[\mu_{hc} - \varphi(x_h)]^2 + \sigma^2_{hc(r)} + \xi(x_h) + \mu_{i_\eta} = 2\eta\left[1 + \frac{\xi'}{4\xi}k_i + Ak_i{}^2 + Bk_i{}^3 + O(k_i{}^4)\right]$$

$$Where,\ A = \frac{(4fc'^2 + \xi' f' + 2f\xi'')}{24f\xi}$$

$$B = \frac{(2ff'c'^2 + 6f^2c'c'' + ff''\xi' + ff'\xi'' + f^2\xi''' - \xi' f'^2)}{48f^2\xi}$$

When we evaluate all of the functions and derivatives in the preceding equations with regard to x_h, we get

$$W_i[\mu_{ic} - \varphi(x_h)]^2 + \sigma^2_{ic(r)} + \xi(x_h) + \mu_{i_\eta} = 2\sqrt{\xi}\left[f\sqrt{\xi}k_i + \frac{(\xi' f + 2f'\xi)}{4\sqrt{\xi}}k_i{}^2 + O(k_i{}^3)\right] \quad (3.1)$$

$$\Rightarrow W_i[\mu_{ic} - \varphi(x_h)]^2 + \sigma^2_{ic(r)} + \eta(x_h) + \mu_{i_\eta} = \int_{x_{h-1}}^{x_h}\sqrt{g(t)}f_i(t)dt\left[1 + O(k_i{}^2)\right]$$

Similarly from (2.5), we have

$$W_h[\mu_{hc} - c(x_h)]^2 + \sigma^2_{hc(r)} + \eta(x_h) + \mu_{i_\eta}$$

$$= 2\sqrt{\eta}\left[f\sqrt{\eta}k_h + \frac{(\eta' f + 2f'\eta)}{4\sqrt{\eta}}k_h{}^2 + O\left(k_h{}^3\right)\right]$$

$$= \int_{x_{h-1}}^{x_h}\sqrt{g(t)}f_i(t)dt\left[1 + O(k_h{}^2)\right] \quad (3.2)$$

where $i = h+1, h = 1,2,\ldots,L-1$

Theorem
Let the regression of the estimation variable $'Y'$ on the stratification variable $'X'$, in the infinite super population is given by

$$y = \varphi(x) + e$$

where $\varphi(x)$ is a function of auxiliary variable, e is the error term such that $E(e|x) = 0$ and $V(e|x) = \eta(x) > 0 \forall x \in (a,b)$ with$(b-a) < \infty$, and further if the function $^{g_1(x)}f_i(x) \in \Omega$: then the system of equations (2.4) give strata boundaries (x_h) which resemble to the minimum of $V(\bar{Y}_{st})_{eq}$ can be written as

$$\left[\int_{x_{h-1}}^{x_h}\sqrt{g(t)}f_i(t)dt.\left[1 + O(k_h^2)\right]\right] = \left[\int_{x_h}^{x_{h+1}}\sqrt{g(t)}f_i(t)dt.\left[1 + O(k_i^2)\right]\right] \quad (3.3)$$

Hence, if we have adequately large number of strata, the k_h's (strata widths) are small and their higher powers in the expansion can be ignored, then the set of equations (3.3) can be approximated by

$$\int_{x_{h-1}}^{x_h}\sqrt{g(t)}f_i(t)dt = constant,\ h = 1,2,\ldots,L \quad (3.4)$$

where terms of order $O\left(Sup_{(a,b)}(k_h)\right)$ have been ignored on both sides of the equation (3.2)
and $Q_3(x_{h-1},x_h)$ is of order $O\left(Sup_{(a,b)}(k_h)\right)$, then the minimal equations (2.4), can to the similar degree of approximation as involved in (3.4), be put as

$Q_3(x_{h-1},x_h)$= constant, $h = 1,2,\ldots,L$ (3.5)

Or equivalently bys

$$Q_3(x_{h-1}, x_h)[1 + O(k_i^2)] = \int_{x_{h-1}}^{x_h} \sqrt{g(t)} f_i(t) dt$$

$$i = h + 1, h = 1,2, \ldots, L$$

Since $\int_{x_{h-1}}^{x_h} \sqrt{g(t)} f_i(t) dt = O(m)$ when the function $\sqrt{g(t)} f_i(t)$ is bounded$\forall x \in (a, b)$. Thus we get the following $cum\sqrt{K_4(x)}$ rule for finding AOSB for equal allocation method.

$cum\sqrt{K_4(x)}$

Rule

Let the function $\sqrt{K_4(x)} = g(t) f_i(t)$ is restricted and its first two derivative exists$\forall x \in (a, b)$, then for specified value of L taking identical intervals on the $cum\sqrt{K_4(x)}$ rule will give us AOSB.

Empirical Study

We will study the following density functions for 'x' to demonstrate the utility of approximate solutions to the minimal equations yielding optimum points of stratification.

i. Rectangular $f(x) = \frac{1}{b-a}, a \leq x \leq b$

ii. Standard normal $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} - \infty \leq x \leq \infty$

For clarity, the simple regression line of $'Y'$on $'X'$ of the form $y = \alpha + \beta x + e$, assuming the value of $\beta = 0.65$. For the conditional variance function, it is assumed to have two different forms like the first form could be a constant and the second could be function of auxiliary variable. i,e. $\eta(x) = \kappa$ and $\eta(x) = \lambda x$, where κ and λ are constants.

For the empirical studies under equal allocation let us assume small values of $= 0.022$, $\lambda = 0.004$, such that there may be very small effect of these constants over the estimation.

If the stratification variable follows the uniform distribution with pdf $(x) = \frac{1}{b-a}$,$x \in [1,2]$, utilizing the $cum\sqrt{K_4(x)}$ rule, we get the stratification points as given in Table I.

Table 1: AOSB and % R.E. for uniformly distributed variable $\eta(x) = \kappa$

L	AOSB	Total variance $\left\{nV(\bar{y}_{SRSS})_{Eq}\right\}$	%R.E
2	1.4854	0.75522	102.067
3	1.3385, 1.6479	0.75233	102.459
4	1.2565, 1.4934, 1.7593	0.75132	102.597
5	1.1951, 1.3866, 1.5803, 1.7931	0.75083	102.664
6	1.1632, 1.3333, 1.5056, 1.7183,1.8283	0.75063	102.691

If the stratification variable follows the standard normal distribution with pdf $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ truncated at $x \in [0,1]$, utilizing the $cum\sqrt{K_4(x)}$ rule, we get the stratification points as given in Table II.

Table 2: AOSB and % R.E. for standard normal distributed variable $\eta(x) = \kappa$

L	AOSB	Total Variance $\left\{nV(\bar{y}_{SRSS})_{Eq}\right\}$	%R.E
2	0.4901	0.08576	113.307
3	0.3086, 0.6313	0.07920	122.688
4	0.2296, 0.4689, 0.7326	0.07893	123.114
5	0.1857, 0.3706, 0.5768, 0.7812	0.07877	123.360
6	0.1560, 0.3058, 0.4671, 0.6381, 0.8083	0.07868	123.496

If the stratification variable follows the uniform distribution with pdf $(x) = \frac{1}{b-a}$,$x \in [1,2]$, utilizing the $cum\sqrt{K_4(x)}$ rule, we get the stratification points as given in Table III. Table III: AOSB and Percent relative efficiency when the auxiliary variable is uniformly distributed.

Table 3: AOSB and % R.E. for uniformly distributed variable $\boldsymbol{\eta}(\boldsymbol{x}) = \boldsymbol{\lambda x}$

L	AOSB	Total Variance $\left\{nV\left(\bar{y}_{SRSS}\right)_{eq}\right\}$	%R.E
2	1.5069	0.75521	102.069
3	1.3601, 1.6833	0.75221	102.476
4	1.2673, 1.9914, 1.7621	0.75289	102.384
5	1.2618, 1.4585, 1.6247, 1.7959	0.75091	102.654
6	1.1785, 1.3848, 1.5199, 1.6805, 1.8476	0.75061	102.695

If the stratification variable follows the standard normal distribution with pdf $(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ $x \in [0,1]$, utilizing $cum\sqrt{K_4(x)}$ rule, we get the stratification points as given in Table IV.

Table 4: AOSB and % R.E. for Standard normally distributed variable $\boldsymbol{\eta}(\boldsymbol{x}) = \boldsymbol{\lambda x}$

L	AOSB	Total Variance $\left\{\boldsymbol{n}\,\boldsymbol{V}(\bar{\boldsymbol{y}}_{\boldsymbol{SRSS}})_{\boldsymbol{Eq}}\right\}$	%R.E
2	0.7018	0.08125	119.60
3	0.3577, 0.6815	0.07929	122.55
4	0.2675, 0.5191, 0.7547	0.07894	123.10
5	0.2296, 0.4259, 0.6235, 0.8088	0.07878	123.34
6	0.1982, 0.3652, 0.5252, 0.6806, 0.8389	0.07869	123.48

The above results in Table I-IV are more promising in the case of uniform and normal distributions as far long as the problem of constructing strata boundaries for ranked set sampling is concerned.

Conclusion

For obtaining the optimum stratification points under SRSS, we have assumed different distributions for the auxiliary variable used as stratification variable. The AOSB obtained for uniform and standard normal distributions are presented in table I-II and table III-IV for $\eta(x) = \delta$ and $\eta(x) = \lambda x$, respectively. The standard normal distribution shows highest % R.E for $\eta(x) = \delta$ and $\eta(x) = \lambda x$, respectively. The upsurge in the number of strata is directly proportional to the decrease in total variance. These figures show a considerable gain in the efficiency of estimators when the proposed method of determining AOSB is used for all $L = 2,3, \dots ,6$. Further, with increase in number of strata the efficiency shows increasing trend. Thus, the proposed method of $cum\sqrt{K_4(x)}$ shows an increase in percent gains in precision while selecting samples using RSS.

References

1. Dalenius T. The problem of optimum stratification, Skand. Aktuarietidskr. 1950;150(3-4):203.
2. Dalenius T, Gurney M. The problem of optimum stratification II, Skand. Aktuarietidskr. 1951;34(1-2):133-148.
3. Danish F, Rizvi SEH, Jeelani MI, Reashi JA. Obtaining strata boundaries under proportional allocation with varying cost of every unit. Pakistan Journal of Statistics and Operation Research. 2017;13(3):567-74.
4. Danish F. A Mathematical Programming Approach for obtaining Optimum Strata Boundaries using two Auxiliary variables under Proportional allocation, Statistics In Transition New Series, ISSN 2450-0291, Exeley, New York, NY. 2018;19(3):507-526.
5. Gupt KB, Ahamed IM. Construction of Strata for a Model-Based Allocation under a Super population Model. Journal of Statistical Theory and Applications. 2021;20(1):46-60.
6. Gupt KB, Swer M, Ahamed IM, Singh BK, Singh KH. Stratification Methods for an Auxiliary Variable Model-Based Allocation under a Super-population Model. Mathematics and Statistics. 2022;10(1):15-24.
7. Hayashi C, Maruyama F. On some method for stratification, (in Japanese), Research Memoir, Inst. Statist. Math. 1948;4:399-411.
8. McIntyre GA. A method for unbiased selective sampling, using ranked set sampling and stratified simple random sampling. Journal of Applied Statistics. 1952;23:231-255.
9. Rizvi SEH, Gupta JP, Singh R. Approximately optimum Stratification for two study variables using auxiliary information. Journal of the Indian Society of the Agricultural Statistics. 2000;53(3):287-298.
10. Rizvi SEH, Gupta JP, Bhargava M. Optimum stratification based on auxiliary variable for compromise allocation. Metron. 2002;28(1):201-215.
11. Singh R, Sukhatme BV. Optimum stratification, Annals of the Institute of Statistical Mathematics. 1969;21(1):515-528.
12. Singh R, Prakash D. Optimum stratification for equal allocation, Annals of the Institute of Statistical Mathematics. 1975;27(1):273-280.
13. Samawi HM. Stratified ranked set sample. Pakistan Journal of Statistics. 1996;12:9-16.