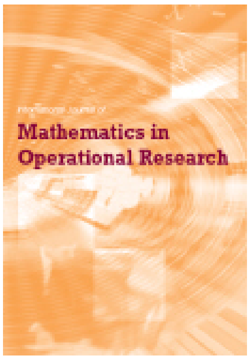

International Journal of Mathematics in Operational Research

ISSN online: 1757-5869 - ISSN print: 1757-5850
https://www.inderscience.com/ijmor

Approximately optimum strata boundaries under super population model

Faizan Danish, S.E.H. Rizvi

DOI: 10.1504/IJMOR.2022.10052805

Article History:

Received:	24 August 2022
Accepted:	06 November 2022
Published online:	23 February 2024

Approximately optimum strata boundaries under super population model

Faizan Danish* and S.E.H. Rizvi

Department of Mathematics,
School of Advanced Sciences,
VIT-AP University,
Inavolu, Beside AP Secretariat, Amaravati AP, India
and
Division of Statistics and Computer Science,
Faculty of Basic Sciences,
Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu,
J&K, India
Email: danishstat@gmail.com
Email: sehrizvistat@gmail.com
*Corresponding author

Abstract: In the present investigation, a methodology has been developed under super population model for obtaining approximately optimum strata boundaries (AOSB) on highly related variable with the variable under consideration which is applicable for the situation having more than one study variables with product estimator. Minimal equations have been obtained under certain assumptions by minimising the generalised variance expressions. Under this complex nature we developed $\mathrm{Cum}\sqrt[3]{R_5(x)}$ rule for obtaining the AOSB. Empirical studies have also been made on certain density functions, which showed remarkable gain in precision.

Keywords: product estimator; super population; optimum strata boundaries; optimum stratification.

Reference to this paper should be made as follows: Danish, F. and Rizvi, S.E.H. (2024) 'Approximately optimum strata boundaries under super population model', *Int. J. Mathematics in Operational Research*, Vol. 27, No. 1, pp.63–74.

Biographical notes: Faizan Danish is currently working as an Assistant Professor (Statistics) in VIT-AP University, Vijayawada Andhra Pradesh, India and has done his PhD from Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, J&K, India. He has worked as a Postdoctoral Fellow in New York University, New York, USA. He has around three years of teaching experience and published around 25 research papers in reputed journals. Further, he has worked as a Biostatiscian under Research Consultation Services, Doha Qatar. He has research expertise in sampling theory, mathematical programming, applied statistics and biostatistics. He is well versed with Statistical Software's: R, STATA, SPSS, Fortran 77, O.P STAT, WINDOSTAT, and Mathematica.

S.E.H. Rizvi started his career as a Lecturer at IAAS, Tribhuvan University, Nepal in 1987. Presently, he is working as a Professor (Statistics)/Dean, Faculty of Basic Sciences at SKUAST-Jammu, India where he joined in 1989 as an Assistant Professor. He has more than 34 years of teaching experience with active involvement in management and planning of various academic activities. He has vast research experience in the field of sampling theory and applied statistics and supervised MSc/PhD students of Statistics. He has a good number of research publications in journals of national and international repute and associated in research projects in different capacities.

1 Introduction

Dalenius (1950) pioneered the concept of obtaining stratification points using study variable as stratification variable. A significant contribution has been made by Dalenius and Hodges (1957), Serfling (1968), Ekman (1969), and Lavall'ee and Hidiroglou (1988) for obtaining the stratification points, and Neyman (1934) on the number of observations to be selected from each stratum. Ghosh (1963) has worked in proportional allocation under two characters. Due to complex formulation of equations, an iterative procedure has been developed by Unnithan (1978) for obtaining initial feasible solution. Study has been made on highly skewed population in business and agricultural surveys by Rivest (2002). Rizvi et al. (2002) used the compromised method and Verma (2008) ratio and regression method for obtaining AOSB and Cochran (1977) considered the distribution of the data to be approximately uniform. Danish and Rizvi (2018) proposed a method for obtaining stratification points using two auxiliary variables.

In recent years there has been an incredible interest of researchers towards the area of stratification points. Danish and Rizvi (2018) proposed a method for obtaining stratification points using two highly related variables. Reddy et al. (2018) studied stratification points and size for health related populations and Danish et al. (2017) utilised single auxiliary variable for stratification purpose using mathematical programming approach. Two stratification variables have been used by Danish et al. (2020), and Danish and Rizvi (2021, 2019). Abo-El Hassan et al. (2021) proposed goal programming for obtaining the stratification points. In recent years there has been dramatically increase in the literature regarding obtained the strata boundaries some of the latest work has been done by Brito et al. (2021), Reddy and Khan (2020), Danish (2019), Reddy and Khan (2020). Alshqaq et al. (2022) discussed about linear approximation of the multivariate stratified sampling problem with examples. Hamid et al. (2021) suggested that mathematical goal programming model to determine the optimum strata boundary by bi-variate variables in multi-objective problem with minimum variance.

Thus, in the present investigation we have used product estimator using to study variables and one auxiliary variable and minimised the general variance for obtaining AOSB and a method has been proposed for the same. The paper concludes with an empirical study.

2 Variance and covariance for product estimators under super population set up

Let us make L strata from the given population of size N and assume that in each stratum the regression lines of the two interested variables on the highly related variable are linear and pass through origin.

Let the regression of Y on X be given by the linear model

$$Y_j = C_j(X) + e_j \qquad (j = 1, 2) \tag{1}$$

where $C_j(X)$ is a real function of X, and e_j is disturbance so that $E(e_j/X) = 0$, $E(e_j e_j'/X, X') = 0$, for $x \neq x'$ $V(e_j/X) = \eta_j(x_i) > 0, j = 1, 2, x \in (a, b), (b - a) < \infty$. $E(e_j c_j) = 0$ but $E(c_1 c_2) \neq 0$. If $f_s(x, y_1, y_2)$ denoted joint density function of (X, Y_1, Y_2) and $f(x)$ marginal of X in the super-population model, then we have

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

$$\mu_{hy_j} = \mu_{hc_j} = \frac{1}{W_h}\int_{x_{h-1}}^{x_h} C_j(x) f(x)dx$$

$$\sigma_{hc_j}^2 = \frac{1}{W_h}\int_{x_{h-1}}^{x_h} C_j^2(x) f(x)dx - \left(\mu_{hc_j}\right)^2$$

$$\sigma_{hy_j}^2 = \sigma_{hc_j}^2 + \mu_{h\eta_j}$$

where (x_{h-1}, x_h) stratification points, $\mu_{h\eta_j}$ is the average value, $\eta_j(x_i)$ the conditional variance of the h^{th} subpopulation.

Further, we assume that for character Y, $R_{11} = R_{12} = \ldots = R_{1L} = R_1$ and that for character Z, $R_{21} = R_{22} = \ldots = R_{2L} = R_2$ so that we can use combined product estimators. The combined estimators in case of stratified sampling are given by

$$\bar{Y}_{st.p} = \frac{\left(\sum_{h=1}^{L} W_h \bar{y}_h\right)\left(\sum_{h=1}^{L} W_h \bar{x}_h\right)}{\bar{X}} \tag{2}$$

and

$$\bar{Z}_{st.P} = \frac{\left(\sum_{h=1}^{L} W_h \bar{z}_h\right)\left(\sum_{h=1}^{L} W_h \bar{x}_h\right)}{\bar{X}} \tag{3}$$

where $W_h = \dfrac{N_h}{N}$ is the stratum weight and $\bar{x}_h, \bar{X}, \bar{Z}_h, \bar{Y}_h$, and $\bar{X}_h$ sample mean, population mean, sample mean, sample mean and population mean of X, X, Z, Y and X respectively.

For large population size, the approximate variance expressions are given by

$$V\left(\bar{Y}_{st.p}\right) = \frac{1}{n}\sum_{h=1}^{L} \frac{W_h^2}{n_h}\left(\sigma_{hy}^2 + R_1^2\sigma_{hx}^2 - 2R_1\sigma_{hxy}\right) \tag{4}$$

and

$$V\left(\bar{Z}_{st.P}\right)=\frac{1}{n}\sum_{h=1}^{L}\frac{W_h^2}{n_h}\left(\sigma_{hz}^2+R_1^2\sigma_{hx}^2-2R_1\sigma_{hxz}\right) \tag{5}$$

The covariance expression has been given in the following Lemma.

Lemma 1:

The covariance term between the product estimators $\bar{Y}_{st.p}$ and $\bar{Z}_{st.P}$ can be given as

$$Cov\left(\bar{Y}_{st.p},\bar{Z}_{st.P}\right)=?\sum_{h=1}^{L}\frac{W_h^2}{n_h}\left(\sigma_{hyz}+R_1R_2\sigma_{hx}^2+R_2\sigma_{hxy}-R_1\sigma_{hxz}\right) \tag{6}$$

Proof:

If we take $\bar{y}_h=\bar{Y}_h+\in_{1h}$, $\bar{z}_h=\bar{Z}_h+\in_{2h}$, and $\bar{x}_h=\bar{X}_h+\xi_h$ so that $E(\in_{1h})=E(\in_{2h})=E(\xi_h)$ then the product estimator $\bar{Y}_{st.p}$ as given by (2) may be written in the following manner

$$\bar{Y}_{st.p}=\frac{1}{X}(\bar{X}\bar{Y}+\bar{X}\sum_h W_h\epsilon_{1h}+\bar{Y}\sum_h W_h\xi_h+\sum_h W_h\epsilon_{1h}\sum_h W_h\xi_h \tag{7}$$

and

$$\bar{Z}_{st.P}=\frac{1}{X}(\bar{X}\bar{Y}+\bar{X}\sum_h W_h\epsilon_{2h}+\bar{Z}\sum_h W_h\xi_h+\sum_h W_h\epsilon_{2h}\sum_h W_h\xi_h \tag{8}$$

Now ignoring the terms of second order we have

$$\begin{aligned}Cov\left(\bar{Y}_{st.p},\bar{Z}_{st.P}\right)&=\frac{1}{X^2}Cov\Big\{\Big(\bar{X}\bar{Y}+\bar{X}\sum W_h\epsilon_{1h}+\bar{Y}\sum W_h\xi_h\Big),\\&\quad\Big(\bar{X}\bar{Z}+\bar{X}\sum W_h\epsilon_{2h}+\bar{Z}\sum W_h\xi_h\Big)\Big\}\\&=\frac{1}{X^2}\sum_{h=1}^{L}W_h^2Cov\left\{\left(\bar{X}\epsilon_{1h}+\bar{Y}\xi_h\right),\left(\bar{X}\epsilon_{2h}+\bar{Z}\xi_h\right)\right\}\end{aligned} \tag{9}$$

The expression (9) may be easily written as follows

$$\begin{aligned}Cov\left(\bar{Y}_{st.p},\bar{Z}_{st.P}\right)&=\sum_{h=1}^{L}W_h^2\Big[Cov\left(\epsilon_{1h},\epsilon_{2h}\right)+R_2Cov\left(\epsilon_{1h},\xi_h\right)\\&\quad+R_1Cov\left(\xi_h,\epsilon_{2h}\right)+R_1R_2Cov\left(\xi_h,\xi_h\right)\Big]\end{aligned} \tag{10}$$

Which can further be expressed as

$$Cov\left(\bar{Y}_{st.p},\bar{Z}_{st.P}\right)=\sum_{h=1}^{L}\frac{W_h^2}{n_h}\left(\sigma_{hyz}+R_1R_2\sigma_{hx}^2+R_2\sigma_{hxy}+R_1\sigma_{hxz}\right)$$

Hence the lemma.

When samples from different strata are drawn using proportional method of allocation, the expressions for variances and covariance become

$$V\left(\bar{Y}_{st.P}\right)_P = \frac{1}{n}\sum_{h=1}^{L} W_h\left(\sigma_{hy}^2 + R_1^2\sigma_{hx}^2 + 2R_1\sigma_{hxy}\right) \tag{11}$$

$$V\left(\bar{Z}_{st.P}\right)_P = \frac{1}{n}\sum_{h=1}^{L} W_h\left(\sigma_{hz}^2 + R_2^2\sigma_{hx}^2 + 2R_2\sigma_{hxz}\right) \tag{12}$$

and

$$Cov\left(\bar{Y}st.p, \bar{Z}_{st.P}\right) = \frac{1}{n}\sum_{h=1}^{L} W_h^2\left(\sigma_{hyz} + R_1R_2\sigma_{hx}^2 + R_2\sigma_{hxy} + R_1\sigma_{hxz}\right) \tag{13}$$

The appropriate regression equations, under the assumption that function form of Y and X, and Z and X in the superpopulation are linear with regression lines passing through origin, are given as

$$Y_{ij} = R_1X_1 + e_{1ij} \tag{14}$$

$$Z_{ij} = R_2X_1 + e_{2ij} \tag{15}$$

where $R_1 = \mu_y/\mu_x$, $R_2 = \mu_z/\mu_x$ and e_{ij}'s are the deviation of Y_{ij} and Z_{ij} from the regression line, and for any given x_i we have

$$E\left(e_{1ij}/x_i\right) = 0, E\left(e_{2ij}/x_i\right) = 0, V\left(e_{1ij}/x_i\right) = \eta_1\left(x_i\right), V\left(e_{2ij}/x_i\right) = \eta_2\left(x_i\right)$$

$\eta_1(x_i)$, $\eta_2(x_i) > 0$ for every x in (a, b).

Under these regression models we have

$$\sigma_{hy}^2 + R_1^2\sigma_{hx}^2 + 2R_1\sigma_{hxy} = 4R_1^2\sigma_{hx}^2 + \mu_{h\eta_1}$$

$$\sigma_{hz}^2 + R_2^2\sigma_{hx}^2 + 2R_2\sigma_{hxz} = 4R_2^2\sigma_{hx}^2 + \mu_{h\eta_2}$$

and

$$\sigma_{hyz} + R_1R_2\sigma_{hx}^2 + R_2\sigma_{hxy} + R_1\sigma_{hxz} = 4R_1R_2\sigma_{hx}^2$$

Thus the expressions for the variances and covariance under proportional method of allocation, as given as (11) to (12), reduce to

$$V\left(\bar{Y}_{st.P}\right)_P = \frac{4R_1^2}{n}\sum_{h=1}^{L} W_h\sigma_{hx}^2 + \frac{\mu_{\eta_1}}{n} \tag{16}$$

$$V\left(\bar{Z}_{st.P}\right)_P = \frac{4R_2^2}{n}\sum_{h=1}^{L} W_h\sigma_{hx}^2 + \frac{\mu_{\eta_2}}{n} \tag{17}$$

$$Cov\left(\bar{Y}_{st.p}, \bar{Z}_{st.P}\right) = \frac{4R_1R_2}{n}\sum_{h=1}^{L} W_h\sigma_{hx}^3 \tag{18}$$

3 Minimal equations

Let $\{x_h\}$ denotes stratification define in (a, b) of the stratification variable X, then the generalised variance G_5 as given by (19), as below

$$G_5 = \begin{vmatrix} \sigma_y^2 & \sigma_{yz} \\ \sigma_{zy} & \sigma_z^2 \end{vmatrix} = \sigma_y^2 \sigma_z^2 - \left(\sigma_{yz}\right)^2 \tag{19}$$

where σ_y^2, σ_z^2 and σ_{yz} denote $V(\overline{Y}_{st.R_1})_P$, $V(\overline{Z}_{st.P})_P$, and $Cov(\overline{Y}_{st.R_1}, \overline{Z}_{st.P})_P$ respectively.

Differentiating G_5 partially with respect to $\{x_h\}$ and then equating its derivative to zero we obtain

$$\frac{\partial G_5}{\partial x_h} = \sigma_y^2 \frac{\partial s_z^2}{\partial x_h} + \sigma_z^2 \frac{\partial \sigma_y^2}{\partial x_h} - 2\sigma_{yz} \frac{\partial \sigma_{yz}}{\partial x_h} = 0,\ h = 1, 2, 3, \ldots, L-1 \tag{20}$$

Inserting the values of σ_y^2, σ_z^2 and σ_{yz} from (16) to (18) in (20), on simplification we obtain

$$W_h \frac{\partial \sigma_{hx}^2}{\partial x_h} + \sigma_{hx}^2 \frac{\partial W_h}{\partial x_h} + W_i \frac{\partial \sigma_{ix}^2}{\partial x_h} + \sigma_{ix}^2 \frac{\partial W_i}{\partial x_h} = 0,\ i = h+1,\ h = 1, 2, \quad n, L-1 \tag{21}$$

For solution, we need partial derivatives, as can be easily obtained, which are given by

$$\frac{\partial \sigma_{hx}^2}{\partial x_h} = \frac{f(x_h)}{W_h}\left[\left(x_h - \mu_{hx}\right)^2 - \sigma_{hx}^2\right] \tag{22}$$

and

$$\frac{\partial \sigma_{ix}^2}{\partial x_h} = \frac{f(x_h)}{W_i}\left[\left(x_h - \mu_{ix}\right)^2 - \sigma_{ix}^2\right] \tag{23}$$

Further we know that

$$\frac{\partial W_h}{\partial x_h} = f(x_h) \text{ and } \frac{\partial W_i}{\partial x_h} = -f(x_h)$$

Thus the minimal equations (20) can further be expressed as

$$\left(x_h - \mu_{hx}\right)^2 = \left(x_h - \mu_{ix}\right)^2$$

$$\text{Or } x_h - \mu_{hx} = \mu_{ix} - x_h \tag{24}$$

And, finally the required minimal equations are obtained as

$$x_h = \frac{\mu_{hx} + \mu_{ix}}{2},\ i = h+1 \tag{25}$$

$$h = 1, 2, \ldots, L\text{-}1$$

This gives the points corresponding to the minimum of generalised variance G_5. This also states that the best x_h is the average of two starta means, based on auxiliary variable, which it separates. Hence, all the points x_h (h = 1, 2, 3, …, L – 1) can be obtained and desired strata formed.

Remarks:

1. It can be easily verified that by putting $C_1(x_h) = 2R_1x_h$ and $C_2(x_h) = 2R_2x_h$ in the minimal equations Rizvi et al. (2000), as obtained in case of stratified simple random sampling under proportional method of allocation, we will get exactly the same result as obtained in (25).
2. This may also be pointed out that the minimal equations, as obtained here for two characters' case, are similar to the case when we take only one study variable together with one auxiliary variable as given by Singh and Sukhatme (1969).

4 Approximate solutions of the minimal equations

The preceding section, we have developed minimal equations, the solution to which will result in stratification points $\{x_h\}$. It is difficult to get exact solution, therefore, we shall obtain approximate solutions to the minimal equations (24) which may then be utilised to get approximately optimum strata boundaries (AOSB).

As pointed earlier, the minimal equations for the present case can be obtained directly from (25) by merely putting $C_1(x_h) = 2R_1x_h$ and $C_2(x_h) = 2R_2x_h$. Also the terms of variances and covariance as given by (16) to (18) can be achieved through the same substitutions. Hence, minimal equations as obtained in product estimation case become a particular case of stratified simple random sampling estimate under proportional allocation. Thus solutions for (24) can also be obtained in the same way.

The solution can be obtained as below:

$$K_h^2 \int_{x_{h-1}}^{x_h} I_1(t) f(t) dt = \text{constant} \tag{26}$$

where

$$I_1(t) = \mu_{\eta 1} C_2'^2(t) + \mu_{\eta 2} C_1'^2(t) \tag{27}$$

Now, if we take

$$C_1(x_h) = 2R_1 x_h$$

and

$$C_2(x_h) = 2R_2 x_h$$

Then

$$C_1'(x_h) = 2R_1$$

and

$$C_2'(x_h) = 2R_2$$

Putting these values in (27) we have

$$I_3(t) = 4R_2^2 \mu_{\eta 1}) + 4R_1^2 \mu_{\eta 2} \tag{28}$$

Therefore, we have

$$K_h^2 \int_{x_{h-1}}^{x_h} I_3(t) f(t) dt = \text{constant}, \, h = 1, 2, 3, \ldots, L \tag{29}$$

where order $O(m^4)$, $m = (Sup(K_h))(a, b)$ is ignored on both side. Thus, we have following theorem:

Theorem 1:

If the estimation variables Y and Z are negatively correlated with highly related variable X so that the regression lines are presented by

$$Y = R_1 X + e_1$$

$$Y = R_1 X + e_2$$

where e_1 and e_2 are the error terms such that $E(e_j/X) = 0$, $E(e_j, e'_j)/(X, X') = 0$ for $X \neq X'$ and $V(e_j/X) > 0$ ($j = 1, 2$) for all $X \in (a, b)$ with $(b - a) < \infty$, and further if the function $I_3(X) f(x)$ belong to Ω, then the equation can be obtained as

$$K_h^2 \int_{x_{h-1}}^{x_h} I_3(t) f(t) dt = \text{constant}, \, h = 1, 2, \ldots, L$$

For large number of strata $O(m^4)$, $m = (Sup(K_h))$ terms are ignored and the function $I_3(t)$ is given by

$$I_3(t) = 4R_2^2 \mu_{\eta 1} + 4R_1^2 \mu_{\eta 2}$$

5 Approximate system of equations

The results can be stated as follows:

1 If the expression of the obtained equations, we retain only the first term thereby neglecting the rest ones, then the two sides are equalised if

$$K_h = \text{Constant} = (b - a)/L, \quad \forall h = \ldots, L \tag{30}$$

and stratification points are

$$x_h = a + h\frac{b - a}{L} \quad \text{with } x_0 = a \text{ and } x_L = b$$

Now using Lemma-I in the system of minimal equations (29) and putting $\lambda = 1$, 1/2 and 1/3, we have following approximate system of equations.

2 $$K_h^2 \int_{x_{h-1}}^{x_h} I_3(t) f(t) dt = C_1, \, h = 1, 2, \ldots, L \tag{31}$$

3 $$K_h [\int_{x_{h-1}}^{x_h} \left[\sqrt{I_3(t) f(t)} dt\right]^2 = C_2, \quad h = 1, 2, \ldots, L \tag{32}$$

4 $$\left[\int_{x_{h-1}}^{x_h} \sqrt[3]{I_3(t) f(t)} dt\right]^3 = C_3, \, h = 1, 2, \ldots, L \tag{33}$$

In single study variable, several forms of $Q(x_{h-1}, x_h)$ have been developed by Singh and Sukhatme (1969). In all the above system of equations, C'_is (i = 1,2,3) are the constraints to be determined. This may be pointed here that the approximate system of equations (33) is more approximate from practical point of view for which the exact value of the constant C_3 is given by

$$C_3 = \frac{1}{L^3}\left[\int_{x_{h-1}}^{x_h} \sqrt[3]{I_3(t)f(t)}dt\right]^3 \tag{34}$$

Based upon the above findings, we shall give a rule to find out the AOSB for two variables of interest when product method of estimation is used.

6 Cum $\sqrt[3]{R_5(x)}$ rule

If the function $R_5(x) = I_5(X)f(x)$, where $I_5(t) = 4R_2^2\mu_{\eta_1} + 4R_1^2\mu_{\eta_2}$, is closed and initial derivatives exists of x in (a, b), then for any vale of L keeping same distance of intervals on the cumulative cube root of $R_5(x)$ will result in $\{x_h\}$.

Remark:

Since the function $I_5(x)$ is itself a constant, proposed method reduced to Cum $\sqrt[3]{f(x)}$ rule. And, in case of any distribution with fixed subpopulation size, the stratification pints will remain stagnant irrespective of conditional variance expression viz. $\eta_1(x)$ and $\eta_2(x)$.

7 Empirical study

For the empirical study, we consider the following distributions:

- uniform distribution: $f(x) = 1$ $1 \le x \le 2$
- right triangular distribution: $f(x) = 2(2 - x)$ $1 \le x \le 2$
- exponential distribution: $f(x) = e^{(x+1)}$ $1 \le x \le \infty$

This section is devoted to illustrate the usefulness of the developed rule of obtaining the set of AOSB as compared to no stratification. For this purpose the values of the AOSB, and then the other parametric values for each stratum, have been obtained by making the use of methodology. Each value of AOSB for different values of L, the number of strata, were obtained through iteration procedure by a margin of ± 0.0005, for each distribution. The numerical values of AOSB, KG_5 and % R.E. (percentage relative efficiency), where K is given by $K = n^2/(4R_2^2\mu_{\eta_1} + 4R_1^2\mu_{\eta_2})$, have been presented in Tables 1, 2 and 3 for uniform, right triangular and exponential distributions respectively. All these values depend only on the form of the distribution considered, and not on the values of g_1 or g_2.

The exponential distribution is truncated at x = 6. From Tables 1–3, it is obtained that the % R.E. enhances trend proportional to the enhance in the number of subpopulations, however, the R.E. is higher for the present case of product estimator than the usual estimator of stratified random sampling. It may be seen that the % R.E. ranges from

403.17 to 3,552.37, 356.03 to 2,961.47, and 312.61 to 2,515.47 for uniform, right triangular and exponential distributions, respectively.

Table 1 AOSB and % R.E. with respect to no stratification for uniform distribution

L	*AOSB*				KG_5	*% R.E.*
1	1.0000	2.0000			0.0834593	-
2	1.0000	1.49953	2.0000		0.0207009	403.17
3	1.0000	1.33280	1.66561	2.0000	0.0092094	906.24
4	1.0000	1.24949	1.49898	1.74847	0.0052410	1,592.43
	2.0000					
5	1.0000	1.19948	1.39897	1.59845	0.0033688	2,477.42
	1.79793	2.0000				
6	1.0000	1.16618	1.33235	1.49853	0.0023494	3,552.37
	1.66471	1.83089	2.0000			

Table 2 AOSB and % R.E. with respect to no stratification for right triangular distribution

L	*AOSB*				KG_5	*% R.E.*
1	1.0000	2.0000			0.0555483	-
2	1.0000	1.40492	2.0000		0.0156022	356.03
3	1.0000	1.26179	1.56029	2.0000	0.0072113	770.29
4	1.0000	1.19368	1.40447	1.64476	0.0041390	1,342.07
	2.0000					
5	1.0000	1.15368	1.31735	1.49553	0.0026796	2,073.01
	1.69852	2.0000				
6	1.0000	1.12737	1.26134	1.40402	0.0018757	2,961.47
	1.55929	1.73617	2.0000			

Table 3 AOSB and % R.E. with respect to no stratification for exponential distribution

L	*AOSB*				KG_5	*% R.E.*
1	1.0000	6.0000			0.8239026	-
2	1.0000	2.56028	6.0000		0.2635587	312.61
3	1.0000	1.94569	3.33292	6.0000	0.1249641	659.31
4	1.0000	1.67971	2.55935	3.80872	0.0721812	1,141.44
	6.0000					
5	1.0000	1.53078	2.17593	2.99829	0.0468111	1,760.06
	4.13486	6.0000				
6	1.0000	1.43549	1.94501	2.55853	0.0327534	2,515.47
	3.33067	4.37304	6.0000			

8 Concluding remarks

In this study, we have developed a methodology for estimating the stratification points on highly related variables with variables under consideration, feasible for the surveys have more than one variable under consideration for product estimators. A Cum $\sqrt[3]{R_5(x)}$ rule has been developed for estimating the stratification points. Numerical illustrations show high gain in precision in the proposed method. It shows the % R.E. enhances with the enhancement in size of subpopulations, however, the R.E. is higher for the present case of product estimator than the usual estimator of stratified random sampling. It may be seen that the % R.E. ranges from 403.17 to 3,552.37, 356.03 to 2,961.47, and 312.61 to 2,515.47 for uniform, right triangular and exponential distributions, respectively.

Data availability

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Acknowledgements

We are thankful to the editor Dr. Angappa Gunasekaran of the journal and reviewers for their suggestions for the improvement of the article.

References

Abo-El Hassan, F.S., Hamid, R.A., Ismail, E. and Ezzat, S.M. (2021) 'Behavior of mathematical goal programming for determining optimum stratum boundaries', المجلة العلمية لقطاع كليات التجارة, Vol. 26, No. 1, pp.130–143.

Alshqaq, S.S.A., Ahmadini, A.A.H., Ali, I. and Tao, J. (2022) 'Nonlinear stochastic multiobjective optimization problem in multivariate stratified sampling design', *Mathematical Problems in Engineering*, Vol. 2022, pp.1–16.

Brito, J.A., de Lima, L., González, P.H., Oliveira, B. and Maculan, N. (2021) 'Heuristic approach applied to the optimum stratification problem', *RAIRO-Operations Research*, Vol. 55, No. 2, pp.979–996.

Cochran, W.G. (1977) *Sampling Techniques*, 3rd ed., John Wiley & Sons, New York, USA.

Dalenius, T. (1950) 'The problem of optimum stratification', *Scandinavian Actuarial Journal*, Vol. 1950, Nos. 3–4, pp.203–213.

Dalenius, T. and Hodges, J.L. (1957) 'The choice of stratification points', *Skandinavisk Aktuarietidskrift*, Vol. 40, pp.198–203.

Danish, F. (2019) 'Construction of Stratification points under optimum allocation using dynamic programming', *Pakistan Journal of Statistics and Operation Research*, Vol. 15, No. 2, pp.341–355.

Danish, F. and Rizvi, S.E.H. (2019) 'Optimum stratification by two stratifying variables using mathematical programming', *Pakistan Journal of Statistics*, Vol. 35, No. 1, pp.11–24.

Danish, F. and Rizvi, S.E.H. (2021) 'Approximately optimum strata boundaries for two concomitant stratification variables under proportional allocation', *Statistics in Transition New Series*, Vol. 22, No. 4, pp.19–40, DOI 10.21307/stattrans-2021-036.

Danish, F., and Rizvi, S.E.H. (2018) 'Optimum stratification in Bivariate auxiliary variables under Neyman allocation', *Journal of Modern Applied Statistical Methods*, Vol. 17, No. 1, p.13.

Danish, F., Rizvi, S., Jeelani, M.I. and Reashi, J.A. (2017) 'Obtaining strata boundaries under proportional allocation with varying cost of every unit', *Pakistan Journal of Statistics and Operation Research*, Vol. 13, No. 3, pp.567–574, https://doi.org/10.18187/pjsor.v13i3.1719.

Danish, F., Rizvi, S.E.H. and Bouza, C. (2020) 'On approximately optimum strata boundaries using two auxiliary variables', *Revista Investigacion Operacional*, Vol. 41, No. 3, pp.445–460.

Ekman, G. (1969) 'A graphical solution for the determination of an optimal stratification or grouping, with an example concerning a problem of maximizing sales revenues', *Review of the International Statistical Institute*, Vol. 37, No. 2, pp.186–194.

Ghosh, S.P. (1963) 'Optimum stratification with two characters', *Ann. Math. Statist.*, Vol. 34, No. 3, pp.866–872.

Hamid, R., Ismail, A.E. and Ezzat, S.M. (2021) 'Multi-objective mathematical programming model for optimum stratification in multivariate stratified sampling', التجارة والتمويل, Vol. 41, No. 4, pp.189–208.

Lavall'ee, P. and Hidiroglou, M. (1988) 'On the stratification of skewed populations', *Survey Methodology*, Vol. 14, No. 1, pp.33–43.

Neyman, J. (1934) 'On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection', *Journal of the Royal Statistical Society*, Vol. 97, No. 4, pp.558–625.

Reddy, K.G. and Khan, M.G.M. (2020) 'stratifyR: an R Package for optimal stratification and sample allocation for univariate populations', *Australian & New Zealand Journal of Statistics*, Vol. 62, No. 3, pp.383–405.

Reddy, K.G., Khan, M.G.M. and Khan, S. (2018) 'Optimum strata boundaries and sample sizes in health surveys using auxiliary variables', *PLoS ONE*, Vol. 13, No. 4, p.e0194787, https://doi.org/10.1371/journal.pone.0194787.

Rivest, L.P. (2002) 'A generalization of the Lavall'ee and Hidiroglou algorithm for stratification in business surveys', *Survey Methodology*, Vol. 28, No. 2, pp.191–198.

Rizvi, S.E.H., Gupta, J.P. and Bhargava, M. (2002) 'Optimum stratification based on auxiliary variable for compromise allocation', *Metron.*, Vol. 60, Nos. 3–4, pp.201–215.

Rizvi, S.E.H., Gupta, J.P. and Singh, R. (2000) 'Approximately optimum stratification for two study variables using auxiliary information', *Journal of Indian Society of Agricultural Statistics*, Vol. 53, No. 3, pp.287–298.

Serfling, R.J. (1968) 'Approximately optimal stratification', *Journal of the American Statistical Association*, Vol. 63, No. 324, pp.1298–1309.

Singh, R. and Sukhatme, B.V. (1969) 'Optimum stratification', *Ann. Inst. Statist. Math.*, Vol. 21, No. 1, pp.515–528.

Unnithan, V.K.G. (1978) 'The minimum variance boundary points of stratification', *Sankhya*, Vol. 40, No. C, pp.60–72.

Verma, M.R. (2008) 'Approximately optimum stratification for ratio and regression methods of estimation', *Applied Mathematics Letters*, Vol. 21, No. 2, pp.200–207.