OPTIMUM STRATIFICATION FOR TWO STRATIFYING VARIABLES

Faizan Danish*[1], S.E.H.Rizvi**, Manish Kumar Sharma**, M. Iqbal Jeelani**, Bunti Kumar** and Q. Farooq Dar**
*Department of Mathematics and Statistics, Govt. Degre College, Shopian-192303 (J&K), India
Department of Statistics and Computer Science, Faculty of Basic Sciences, SKUAST-Jammu
**Department of Statistics, Ramanujan School of Mathematical Science, Pondicherry University, India

ABSTRACT
In this paper, we have taken under considerations two auxiliary variables used as stratification variables and single study variable. For obtaining the optimum strata boundaries, we have solved the non linear equations using dynamic programming that resulted in more efficient results rather than obtained on using single study variable and single auxiliary variable. The proposed method is illustrated by using different distributions followed by stratification variables.

KEYWORDS: Concomitant Variables, Dynamic Programming, Multistage Decision Problem

MSC: 62D05

RESUMEN
Hemos considerado en este paper que se usan dos variables auxiliares como variables de estratificación para una sola variable de estudio. Para obtener fronteras de estratificación óptimas, hemos resuelto el problema de las ecuaciones no- lineales usando programación dinámica lo que determinó resultados más eficientes que al usar una sola variable con una sola variable auxiliar . El método propuesto es ilustrado usando diferentes distribuciones de las variables de estratificación.

PALABRAS CLAVE: Varaibles Concomitantes, Programación Dinámica, Problema de Decisión Multietápica

1. INTRODUCTION

The main aim for stratification in the design of sample survey is to reduce the sample variance of the estimates. In this type of sampling the whole population is partitioned into a number of strata and from each of the stratum a sample is selected by using the desired sampling design. The strata are made in such a way that the stratum is homogenous within itself and are as much as heterogeneous as possible between itself. The main factors that influence the reduction of variance to a very large extent are like choice of the variable on the basis of stratification will be one, the total number of strata that should be made from the whole population while partitioning, determination of the stratum boundaries is one the most important factor that influences the variation and the last but not the least is the design of sampling used in each stratum for selecting a sample from it. The effect of stratification using one stratifying variable has received considerable attention from research point of view. This issue was first considered by Dalenius (1950): He exhibited an arrangement of minimal equations for obtaining optimum boundary points. Small number of strata could only be the option for solving the equations. Subsequently endeavours have been made by many authors such as Dalenius and Gurney (1951), Mahalanobis (1952), Hansen, *et al.* (1953), Aoyama (1954), Ekman (1959), Dalenius and Hodges (1959), Sethi (1963), Hess, *et al.* (1966), Singh and Sukhatme (1969, 1973), Singh (1977), Yadava and Singh (1984), and others. Most of the authors suggested different approaches and obtained the calculus equations in terms of stratum mean and stratum variance for determining the strata boundaries. These equations are ill adapted to practical computations, since both stratum mean and stratum variance themselves depend on the boundary points. For this reason, they suggested different rules to obtain approximate optimum strata boundaries.
An iterative technique utilizing Shanno's changed Newton strategy while deciding the strata boundaries that prompts a local least of the variance for Neyman portion, once the best starting solution is choosen was recommended by Unnithan (1978): The technique is turned out to be quicker than the Dalenius and Hodges iterative methodology. Bhler and Deutler (1975) figured the issue of deciding OSB as an improvement issue

[1] corresponding author: danishstat@gmail.com

and built up a computational system to tackle the issue utilizing dynamic programming. Bouza, C. (2002) studies the ability in diagnosis of a separator variable is studied using a superpopulation model for stratified sampling. Khan, *et al*. (2002) proposed a method for constructing stratification points using dynamic programming approach. Later Khan, *et al.* (2005) formulated the problem of OSB for exponential study variable as a Mathematical programming Problem (MPP) and determined the optimum boundary points using dynamic programming. Khan *et al*. (2015) determined strata points having log normal distribution for non symmetrical population.

Danish *et al*. (2017) talked about the method for getting optimum strata boundaries when the cost of each unit fluctuates in the entire strata. Danish *et al.* (2017) brought several developments towards construction of strata boundaries together. Danish and Rizvi (2017) examined issue of discovering OSB by thinking about as the issue of optimum strata width (OSW), utilizing MPP by dynamic programming strategy, when the investigation variable is uniformly distributed. Likewise the proposed strategy demonstrates superior to other stratification technique (Singh, 1967): Danish and Rizvi (2018) proposed a method when the auxiliary variables have standard Cauchy and power distributions.

The current investigation emphasis on one study variable and two auxiliary variables used as stratification variables. The proposed method would be compared with several existing methods.

2. FORMULATION OF PROBLEM

Suppose the target population consisting of 'N' units be subdivided LxM strata dependent on two explanatory variables 'X' and 'Z' when the estimation of mean of the study variable 'Y' is of intrigue. In order to have the estimate ,we divide the whole population into the desired number of strata like LxM such that each stratum is homogenous within itself and heterogeneous between strata regarding the character under investigation such that the number of units in the $(h,k)^{th}$ stratum is N_{hk} ,such that

$$\sum_{h=1}^{L}\sum_{k=1}^{M} N_{hk} = N$$

From each of the stratum, a sample of size n_{hk} is to drawn such that a total of 'n' units of sample would be taken from a universe consisting of size 'N'. Let the sample size selected from $(h , k)^{th}$ stratum be 'n_{hk}' ,such that

$$\sum_{h}\sum_{k} n_{hk} = n$$

Let y_{hki} ,(i=1,2,3,...,N_{hk}) denotes the populace unit in the (h , k)th stratum ,at that point the populace total can be expressed as

$$y = \sum_{h}\sum_{k}\sum_{i} y_{hki}$$

An unbiased estimator of the population mean $\overline{Y}_N$ unde above sampling ,is

$$\overline{y}_{st} = \sum_{h}\sum_{k} W_{hk}\,\overline{y}_{hk}$$

where ,

$\overline{y}_{hk} = \dfrac{1}{n_{hk}}\sum_{i} y_{hki}$ and 'W_{hk}' denotes the weight of the $(h,k)^{th}$ stratum and is equal to $\dfrac{N_{hk}}{N}$

The sampling variance of the unbiased estimator '$\overline{y}_{st}$' is

$$V\left(\overline{y}_{st}\right) = \sum_{h}\sum_{k}\left(\frac{1}{n} - \frac{1}{N}\right) W_{hk}^{2}\sigma_{hky}^{2} \qquad h=1,2,...,L\ ;\ k = 1,2,...,M$$

where σ_{hky}^{2} represents the variance for the $(h,k)^{th}$ stratum.

However, if the finite population correction (f.p.c) is ignored, we can express the above equation as

$$V\left(\overline{y}_{st}\right) = \sum_{h}\sum_{k}\frac{W_{hk}^{2}\sigma_{hky}^{2}}{n} \tag{1}$$

At the point when the study variable 'Y' itself isn't utilized for stratification variable, we propose a model based on bi-variate stratified sampling design. Let the regression model of study variable on auxiliary variables be as

$$Y = \lambda\left(x,z\right) + \varepsilon \tag{2}$$

where, $\lambda\left(x,z\right)$ is function of 'X' and 'Z' which is assumed to be linear and 'ε' denotes the blunder term with the end goal that

$$E\left(\varepsilon\middle|(x,z)\right) = 0 \quad \text{and} \quad V\left(\varepsilon\middle|(x,z)\right) = \phi\left(x,z\right) \text{ for all (x, z)}$$

Under (2), we can have

$$\mu_{hky} = \mu_{hk\lambda} \tag{3}$$

and

$$\sigma^2_{hky} = \sigma^2_{hk\lambda} + \mu_{hk\phi} \tag{4}$$

where $\mu_{hk\lambda}$ and $\mu_{hk\phi}$ are the weighted average value of $\lambda\left(x,z\right)$ and $\phi\left(x,z\right)$ individually and $\sigma^2_{hk\lambda}$ denotes the variance of $\lambda(x,z)$ in the (h,k)th stratum.

If 'λ' and 'ε' are uncorrelated, then in model (2) 'σ^2_{hky}' can be expressed as (T.Dalenius and M. Gurney 1951)

$$\sigma^2_{hky} = \sigma^2_{hk\lambda} + \sigma^2_{hk\varepsilon} \tag{5}$$

where $\sigma^2_{hk\varepsilon}$ is the fluctuation of blunder term in (h, k)th stratum. It can be verified by (4) and (5):

Let the joint density function of the super population is f(x, y, z) and joint marginal of X and Z is f(x, z) .Let f(x) and f(z) be the recurrence capacity of the auxiliary factors X and Z respectively defined in the interval [a, b] and [c, d].

On the off chance that the populace average of the variable under 'Y' is evaluated under the variance given in condition (1) ,at that point the issue of deciding the strata limits is to cut up the ranges $d_x = b - a$ and $t_z = \quad d - c$, at (L - 1) and (M-1) intermediate points as

$$a = x_0 \le x_1 \le \quad \ldots \quad \le x_{L-1} \le x_L = b$$

and

$$c = z_0 \le z_1 \le \ldots \le z_{M-1} \le z_M = d$$

respectively with the end goal that the condition (2) is least.

Minimizing (1) is equivalent to minimzing

$$\sum_h \sum_k W^2_{hk} \sigma^2_{hky}$$

as the value of 'n' is known in advance. Thus while using (4), we have

$$\sum_h \sum_k W^2_{hk} \left(\sigma^2_{hk\lambda} + \mu_{hk\phi}\right) \tag{6}$$

On the off chance that $f\left(x,z\right), \lambda\left(x,z\right)$ and $\phi\left(x,z\right)$ are known and furthermore imtegrable at that point, $W_{hk}, \sigma^2_{hk\lambda}$ and $\mu_{hk\phi}$ can be acquired as a component of boundary points $\left(x_h, x_{h-1}, z_k, z_{k-1}\right)$ utilizing the accompanying articulation

$$W_{hk} = \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} f\left(x,z\right) \partial x \partial z \tag{7}$$

$$\sigma^2_{hk\lambda} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \lambda^2\left(x,z\right) f\left(x,z\right) \partial x \partial z - \mu^2_{hk\lambda} \tag{8}$$

and

$$\mu_{hk\phi} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \phi(x,z) f(x,z)\, \partial x \partial z \tag{9}$$

Where,

$$\mu_{hk\lambda} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \lambda(x,z) f(x,z)\, \partial x \partial z \tag{10}$$

where $(x_h - x_{h-1})$ and $(z_k - z_{k-1})$ are the boundary points of the (h,k)th stratum. If a sample is taken randomly of size 'n_{hk}' from (h, k)th stratum unbiased estimator of $\bar{x}$ and $\bar{y}$ can be obtained as

$$\bar{x}_{st} = \sum_h \sum_k W_{hk} \bar{x}_{hk} \ , \ \bar{y}_{st} = \sum_h \sum_k W_{hk} \bar{y}_{hk}$$

where $\bar{x}_{st}$ and $\bar{y}_{hk}$ are the unbiased sample estimator of $\bar{X}_{hk}$ and $\bar{Y}_{hk}$ .Thus the objective function (6) could be expressed as the function of boundary points $(x_h, x_{h-1}, z_k, z_{k-1})$ only.

Let

$$\phi(x_h, x_{h-1}, z_k, z_{k-1}) = W_{hk}^2 \left(\sigma_{hk\lambda}^2 + \mu_{hk\phi}\right) \tag{11}$$

we have already let the range

$$d_x = b - a = x_L - x_0 \tag{12}$$

$$t_z = d - c = z_M - z_0 \tag{13}$$

Then, in the bivariate stratification a problem of determining the strata boundaries (x_h, z_k) is to break up the ranges of (12) and (13) at intermediate points in order to estimate $x_1 \le x_{2\le} \ldots \le x_{L-2} \le x_{L-1}$ and $z_1 \le z_{2\le} \ldots \le z_{M-2} \le z_{M-1}$.Then, the reasonable criterion for determining optimum strata boundaries(OSB) $(x_h, z_k)$is to minimize

Minimize $\sum_h \sum_k \phi_{hk}(x_h, x_{h-1}, z_k, z_{k-1})$

Subject to constraint (14)

$$a = x_0 \le x_1 \le \ldots \le x_{L-1} \le x_L = b$$

$$c = z_0 \le z_1 \le \ldots \le z_{M-1} \le z_M = d$$

and

$$\sum_h \sum_k n_{hk} = n$$

When, the marginal frequency function are known and σ_{hky}^2 communicated as a component of stratification points (x_h, z_k).For the rectangular stratification let $V_h = x_h - x_{h-1}$ and $U_k = z_k - z_{k-1}$ denotes the total length of the (h, k) th stratum. Then, using (12) & (13) ,the ranges can be expressed as

$$\sum_h V_h = \sum_h (x_h - x_{h-1}) = b - a = d_x \tag{15}$$

$$\sum_k U_k = \sum_k (z_k - z_{k-1}) = d - c = t_z \tag{16}$$

The objective function in (4) proposes that, for determining of two sided subdivision, a two sided dynamic programming approach ought to be utilized. Utilizing the general concept of dynamic programming with the state and decision variables b y the sets (h, k):Then issue of two way ideal stratification can be communicated as

$$\begin{aligned}
&Minimize \quad \sum_h \sum_k \phi_{hk}\left(x_h, x_{h-1}, z_k, z_{k-1}\right) \\
&\quad Subject \quad to \\
&\qquad \left(x_h, z_k\right) = \left(x_{h-1} + V_h, z_{k-1} + U_k\right) \\
&\qquad\quad \left(x_h, z_k\right) \in \left[a, d\right] \times \left[c, d\right] \\
&\qquad\quad \left(V_h, U_k\right) \in B_h\left(x_{h-1}\right) \times B_k\left(z_{k-1}\right) \\
&\qquad\qquad = \left[0, b - x_{h-1}\right] \times \left[0, d - z_{k-1}\right] \\
&\qquad\quad \left(x_0, z_0\right) = \left[a, c\right] \qquad h = 1,2,...,L \quad and \quad k = 1,2,...,M
\end{aligned} \tag{17}$$

Although the formulation (17) seems to be difficult, can be truth be told be seen regarding the choice space $B_h\left(x_{h-1}\right) \times B_k\left(z_{k-1}\right)$.We propose a basic methodology which allows an answer for the issue (17) utilizing the one-dimensional powerful programming iteratively. Prior to the primary emphasis, some trail esteems say x_0 and z_0 ,with the end goal that $a = \quad x_0 \le x_1 \le ... \le x_{L-1} \le x_L \quad = b$ and $c = \quad z_0 \le z_1 \le ... \le z_{M-1} \le z_M \quad = d$ are choosen for the underlying purposes of the stratification. At that point the ith emphasis (i = 1,2,...) the purposes of subdividing the population z^{i-1} are initially supposed as settled. Keeping in mind the aim of subdividing the population x^{i-1}could likewise be choosen rather than z^{i-1}.Locking the results of z^{i-1} is having certainty the impact of diminishing problem precisely to the any of the ideal subdivision of population with one all variables taken as the basis variable. This can be seen by contrasting the plan (17) to the one which is characterized on univarite auxiliary variable utilized as stratification variable, estimations of the purposes of subdividing Z choosen as constant in (17):

Suppose $\phi^*_{x_h} \left(x_{h-1}, z^{i-1}\right)$ denotes ideal incentive for the main problem (14) for the subpopulation (h,k) to (L,k) having k = 1,2,...,M that the bottom destined for subpopulation (h,k) having k = 1,2,...,M be x_{h-1}.The initial segment of the ith emphasis of the function given by Bellman is then given by

$$\phi^*_{x_h} \left(x_{h-1}, z^{i-1}\right) = \underset{V_h \in B_h\left(x_{h-1}\right)}{Minimize} \left\{ \sum_{k=1}^{M} \phi\left(x_{h-1}, x_h, z^{i-1}_{k-1}, z^{i-1}_k\right) + \phi^*_{x_{h+1}} \left(x_h, z^{i-1}\right) \middle| x_h = x_{h-1} + V_h \right\}$$

where $B_h\left(x_{h-1}\right)$ is defined in (17):

Using this condition, new purposes behind subdividing in reference to 'X' can be gotten to response the technique regard .Hence the OSB for the underlying portion of the ith cycle as$\left(x^i, z^{i-1}\right)$ .For next piece of the ith accentuation ,the purpose of subdividing x^i are along these lines $\left(x^i, z^{i-1}\right)$supposed as settled.

Restating, issue of constructing of boundaries as the issue of stratification points $\left(V_h, U_k\right)$,including condition (15) and (16) as a requirement ,the equation (14) can be taken as a equation problem of deciding OSW,$V_1, V_2, ..., V_L \quad and \quad U_1, U_2, ..., U_M$ and is communicated as the accompanying MPP:

Minimize $\sum_h \sum_k \phi_{hk}\left(x_h, x_{h-1}, z_k, z_{k-1}\right)$

Subject to

$$\sum_h V_h = d_x \tag{18}$$

$\sum_k U_k = t_z$, h = 1,2,...,L and k =1,2,...,M

and

$V_h \geq 0 \quad and \quad U_k \geq 0$

Initially, $\left(x_0, z_0\right)$ are the initial values of the auxiliary variables X and Z respectively which are known. Therefore, the first term $\phi_{11}\left(x_1, x_0, z_1, z_0\right)$ in the objective function (18) is the function of $\left(V_1, U_1\right)$ alone, once the $\left(V_1, U_1\right)$ is known, the second term $\phi_{22}\left(x_2, x_1, z_2, z_1\right)$ will be the function of $\left(V_2, U_2\right)$ alone etc. Because of uncommon nature of function MPP (18) might be treated as the function of $\left(V_h, U_k\right)$ and can be communicated as

$$\text{Minimize} \quad \sum_h \sum_k \phi_{hk}\left(V_h, U_k\right)$$

$$\text{Subject to} \qquad \qquad (19)$$

$$\sum_h V_h = d_x$$

$$\sum_k U_k = t_z \qquad ,h=1,2,...,L \text{ and } k=1,2,...,M$$

and

$$V_h \geq 0 \quad and \quad U_k \geq 0$$

3. THE SOLUTION PROCEDURE

The problem (19) is a problem of multistage decision in which the objective function and the constraints are divisible elements of $\left(V_h, U_k\right)$, which enables us to utilize a dynamic programming system. Dynamic programming decides ideal arrangement of a multi-variable issue by disintegrating into stages, each stage bargaining a single variable sub issue. A dynamic programming model is commonly a recursive condition and these recursive equation links to different stages of the problem.

Taking into consideration the accompanying nested issue of condition (19) for initial subpopulation $\left(L_1 \times M_1\right)$ strata, where $\left(L_1 \times M_1\right) \leq \left(L \times M\right)$, i, e $L_1 < L, M_1 < M$

$$\text{Minimize} \quad \sum_{h=1}^{L_1} \sum_{k=1}^{M_1} \phi_{hk}\left(x_{h-1}, x_h, z_{k-1}, z_k\right)$$

$$\text{Subject to} \qquad \qquad (20)$$

$$\sum_{h=1}^{L-1} V_h = d_{L_1}$$

$$\sum_{k=1}^{M-1} U_k = t_{M_1} \qquad ,h=1,2,...,L_1 \text{ and } k=1,2,...,M_1$$

and

$$V_h \geq 0 and U_k \geq 0$$

where

$$d_{L_1} < V, t_{M_1} < M$$

Note that if $d_{L_1} = V$ and $t_{M_1} = U$ then $\left(L_1 \times M_1\right) = \left(L \times M\right)$

The transformation functions are given by

$$d_{L_1} = V_1 + V_2 + \ldots + V_{L_1}$$
$$d_{L_1-1} = V_1 + V_2 + \ldots + V_{L_1-1} = d_{L_1} - V_{L_1}$$
$$d_{L_1-2} = V_1 + V_2 + \ldots + V_{L_1-2} = d_{L_1-1} - V_{L_1-1}$$

.

.

.

$$d_2 = V_1 + V_2 = d_3 - V_3$$
$$d_1 = V_1 = d_2 - V_2$$

Similarly, we have

$$t_{M_1} = U_1 + U_2 + \ldots + U_{M_1}$$
$$t_{M_1-1} = U_1 + U_2 + \ldots + U_{M_1-1} = t_{M_1} - U_{M_1}$$
$$t_{M_1-2} = U_1 + U_2 + \ldots + U_{M_1-2} = t_{M_1-1} - U_{M_1-1}$$

.

.

.

$$t_2 = U_1 + U_2 = t_3 - U_3$$
$$t_1 = U_1 = t_2 - U_2$$

Let

$$\phi_{L_1 \times M_1}\left(V_{L_1} \times U_{M_1}\right) = Min\left[\sum_{h=1}^{L_1}\sum_{k=1}^{M_1}\phi_{hk}\left(V_h, U_k\right)\middle|\sum_{h=1}^{L_1} V_h = d_{L_1}, \sum_{k=1}^{M_1} U_k = t_{M_1}\right]$$
$$and \quad V_h \geq 0, U_k \geq 0; h = 1,2,3,\ldots,L_1 \quad ; \quad k = 1,2,3,\ldots,M_1$$
$$and \quad 1 \leq L_1 \leq L \quad , \quad 1 \leq M_1 \leq M$$

means the minimum estimation of the objective function of the condition (20), that is,

$$\phi_{L_1 \times M_1}\left(d_{L_1}, t_{M_1}\right) = Min\left[\sum_{h=1}^{L_1-1}\sum_{k=1}^{M_1-1}\phi_{hk}\left(V_h, U_k\right)\middle|\sum_{h=1}^{L_1-1} V_h = d_{L_1-1}, \sum_{k=1}^{M_1-1} U_k = t_{M_1-1}\right]$$
$$and \quad V_h \geq 0, U_k \geq 0; h = 1,2,3,\ldots,L_1 \quad and \quad k = 1,2,3,\ldots,M_1$$

with the above definition of $\phi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right)$,the MPP (19) is equivalent to finding $\phi_{L \times M}\left(d_x, t_z\right)$ recursively by defining $\phi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right)$ for $L_1 = 1,2,\ldots,L \quad and \quad M_1 = 1,2,\ldots,M$; $0 \leq d_{L_1} \leq V \quad , \quad 0 \leq t_{M_1} \leq U$.

$$\phi_{L_1 \times M_1}\left(d_{L_1}, t_{M_1}\right) = Min \begin{bmatrix} \phi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right) + \\ \left[\sum_{h=1}^{L_1-1}\sum_{k=1}^{M_1-1}\phi_{hk}\left(V_h, U_k\right)\middle|\sum_{h=1}^{L_1-1} V_h = d_{L_1} - V_{L_1}, \sum_{k=1}^{M_1-1} U_k = t_{M_1} - U_{M_1}\right] \end{bmatrix} \tag{21}$$
$$and \quad V_h \geq 0, U_k \geq 0; h = 1,2,3,\ldots,L_1 \quad and \quad k = 1,2,3,\ldots,M_1$$

For fixed value of $\left(V_{L_1}, U_{M_1}\right), 0 \le d_{L_1} \le V \quad , \quad 0 \le t_{M_1} \le U$.

$$\phi_{L_1 \times M_1}\left(d_{L_1}, t_{M_1}\right) = \phi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right) + Min \quad \left[\sum_{h=1}^{L_1-1} \sum_{k=1}^{M_1-1} \phi_{hk}\left(V_h, U_k\right) \middle| \sum_{h=1}^{L_1-1} V_h = d_{L_1} - V_{L_1}, \sum_{k=1}^{M_1-1} U_k = t_{M_1} - U_{M_1}\right]$$

and

$$\begin{aligned} V_h &\ge 0 \quad , h = 1, 2, \ldots, L_1 \\ U_k &\ge 0 \quad , k = 1, 2, \ldots, M_1 - 1 \\ 1 \le L_1 &\le L \quad , 1 \le M_1 \le M \end{aligned}$$

Using the same procedure to write the forward recursive equation of the dynamic programming technique and could obtain OSB.
Let us assume the regression model defined in equation (2) be linear as:

$$Y = \alpha + \beta x + \gamma z + \varepsilon$$

then

$$\sigma_{hky}^2 = \beta^2 \sigma_{hkx}^2 + \gamma^2 \sigma_{hkz}^2 \tag{22}$$

when 'X' and 'Z' are independent of error term 'ε',the weight and variance of the (h,k)th stratum having auxiliary variables as 'X' and 'Z'.

$$W_{hk} = \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} f\left(x, z\right) \partial x \partial z \tag{23}$$

$$\sigma_{hkx}^2 = \frac{1}{W_{hk}} \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} x^2 f\left(x\right) \partial x \partial z - \mu_{hkx}^2 \tag{24}$$

$$\sigma_{hkz}^2 = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} z^2 f\left(z\right) \partial z \partial x - \mu_{hkz}^2 \tag{25}$$

where $\quad \mu_{hkx} = \frac{1}{W_{hk}} \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} x f\left(x\right) \partial x \partial z \quad , \quad \mu_{hkz} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} z f\left(z\right) \partial z \partial x$

4. EMPIRICAL STUDY

To outline the calculation subtleties of the proposed design, we consider the population of size 2000.
Let the 'X' auxiliary variable is exponentially distributed having distribution finction as below:

$$f\left(x, \lambda\right) = \begin{cases} \lambda e^{-\lambda x} & : x > 0 \\ 0 & : \quad otherwise \end{cases} \tag{26}$$

and Z auxiliary variable follows a right-triangular distribution with pdf as:

$$f\left(z, a, b\right) = \begin{cases} \frac{2\left(b - z\right)}{\left(b - a\right)^2} ; a \le z \le b \\ 0 \qquad ; otherwise \end{cases} \tag{27}$$

If the two stratification variables are independent then substitute the values of f(x) and f(z) taken from (26) and (27) equations respectively in (23) , (24) and (25), we have

$$W_{hk} = e^{-\lambda x_h}\left(e^{\lambda V_h}-1\right)\frac{2bU_k - 2z_{k-1}U_k - U_k^2}{\left(b-a\right)^2} \tag{28}$$

$$\sigma_{hkx}^2 = \frac{\left(\lambda^2 x_{h-1} + 2\lambda x_{h-1} + 2\right)e^{\lambda V_h} - \lambda\left(V_h^2 + 2x_{h-1}V_h + x_{h-1}^2\right) - 2\left(x_{h-1}+V_h\right)}{\lambda^2\left(e^{\lambda x_h}-1\right)} - \left[\frac{\left(\lambda x_{h-1}+1\right) - \lambda\left(x_{h-1}+V_h\right) - 1}{\lambda\left(e^{\lambda x_h}-1\right)}\right]^2 \tag{29}$$

$$\sigma_{hkz}^2 = \frac{\left(2bU_k - U_k^2\right)z_{k-1}^2 - 2z_{k-1}^3 U_k}{2bU_k - 2z_{k-1}U_k - U_{k-1}^2} - \left[\frac{\left(2bU_k - U_k\right)^2 - 2z_{k-1}^2 U_k}{2bU_k - 2z_{k-1}U_k - U_k^2}\right]^2 \tag{30}$$

Using the variance formula given in (1) and substitute obtained in (22),we get MPP as

Minimize $\sum_h \sum_k W_{hk}^2 \left(\beta^2 \sigma_{hkx}^2 + \gamma^2 \sigma_{hkz}^2\right)$

Subject to

$$\begin{aligned} &\sum_h V_h = d_x \\ &\sum_k U_k = t_z \\ &\forall V_h \ge 0, U_k \ge 0 \quad , \quad \begin{matrix} h = 1,2,\ldots,L \\ k = 1,2,\ldots,M \end{matrix} \end{aligned} \tag{31}$$

By substituting values obtained in equations (28)-(30) in MPP (31),we get

Minimize

$$\sum_h \sum_k \left(e^{-\lambda x_h}\left(e^{\lambda V_h}-1\right)\frac{2bU_k - 2z_{k-1}U_k - U_k^2}{\left(b-a\right)^2}\right)^2$$

$$\left(\begin{matrix} \beta^2 \dfrac{\left(\lambda^2 x_{h-1} + 2\lambda x_{h-1} + 2\right)e^{\lambda V_h} - \lambda\left(V_h^2 + 2x_{h-1}V_h + x_{h-1}^2\right) - 2\left(x_{h-1}+V_h\right)}{\lambda^2\left(e^{\lambda x_h}-1\right)} \\ -\left[\dfrac{\left(\lambda x_{h-1}+1\right) - \lambda\left(x_{h-1}+V_h\right) - 1}{\lambda\left(e^{\lambda x_h}-1\right)}\right]^2 + \gamma^2 \dfrac{\left(2bU_k - U_k^2\right)z_{k-1}^2 - 2z_{k-1}^3 U_k}{2bU_k - 2z_{k-1}U_k - U_{k-1}^2} - \left[\dfrac{\left(2bU_k - U_k\right)^2 - 2z_{k-1}^2 U_k}{2bU_k - 2z_{k-1}U_k - U_k^2}\right]^2 \end{matrix}\right)$$

Subject to

$$\begin{aligned} &\sum_h V_h = d_x \\ &\sum_k U_k = t_z \\ &\forall V_h \ge 0, U_k \ge 0 \quad , \quad \begin{matrix} h = 1,2,\ldots,L \\ k = 1,2,\ldots,M \end{matrix} \end{aligned} \tag{32}$$

Now let the population divided into $L \times M$ =12 strata i,e L=4 and M=3, and suppose estimates of $x_0 = 0, d_x = 10, a = z_0 = 0, z_k = 2, t_z = 2.$ Also by simulation of the linear regression model using the pdf's (26) and (27) we get the estimates of coefficients $\beta = 0.5 \quad and \quad \gamma = 1.5$.

By executing the programme coded in LINGO of NLPP given in (32) for determing the OSB gives the required result were obtained as in Table I

Table I: OSB, W_{hk} and Variance when X and Z auxiliary variables are exponentially and Right- Triangularly distributed

OSB (x_h, z_k)	Stratum Weight (W_{hk})	Variance
(0.3917, 0.3088) (0.8938, 0.3088) (1.6092, 0.3088) (10.0000, 0.3088) (0.3917, 0.7834) (0.8938, 0.7834) (1.6092, 0.7834) (10.0000, 0.7834 (0.3917, 2.0000) (0.8938, 2.0000) (1.6092, 2.0000) (10.0000, 2.0000)	0.924 0.0760 0.0596 0.0570 0.1118 0.0920 0.0721 0.0696 0.1199 0.0987 0.0774 0.0740	0.00003691

It can be large even for small values of L and M. It was observed that the variance goes on increasing after number of strata goes behind strata. Thus the sufficient strata boundaries for each of the auxiliary variables would be 4 or 5.

Now in order to make the comparison of the proposed methods and the existing methods a simulation consider is completed to discuss the efficiency of the proposed strategy. The existing strategies that are to be considered for comparison are:

1) Dalenius and Hodges (1959) cum $\sqrt{f}$ method
2) Gunning and Horgan (2004) method
3) Hidiroglou's method (1988) using Kozak's (2004) method
4) Khan *et al*. (2008) mathematical programming approach
5) Proposed method

In order to have the above comparison the population of size N = 5000 with the stratification variables 'X' and 'Z' that follow an exponential and Right-triangular distributions respectively was haphazardly created utilizing the R-programming.

For auxiliary variable X :

$$x_0 = 0.00006, x_L = 11.27211, d_x = 11.27205$$

Similarly for the auxiliary variable z:

$$z_0 = 0.00043, z_M = 2.96877, t_z = 2096834$$

If L=M=4, then the OSB's are determined by using R-package 1-3 and LINGO for 4 and 5.

The results obtained after executing the programme in both the softwares are given in Table II

Table II: The variance of variables and their total variances

Method of Stratification	$v(\bar{x}_{st})$ **(in e -07)**	$v(\bar{z}_{st})$ **(in e -07)**	**Total variance (in e-07)**
Dalenius and Hodges (1959)cum $\sqrt{f}$ method	583.1884	275.1031	585.2915
Gunning and Horgan (2004) method	3068.8590	1783.4450	4852.304
Hidiroglou's method (1988) using Kozak's (2004) method	559.6009	439.4406	999.0415
Khan *et al.* (2008) mathematical programming approach	531.1695	340.9203	872.0898
Proposed approach	499.7879	0.2451	500.0330

Thus, it reveals that the proposed method of stratification shows maximum reduction in variance of both the estimates as compared to other method. Thus, the proposed bi-variate stratification method is a better option for obtaining the OSB.

5. CONCLUSION

In this paper we developed a strategy for deciding the two dimensional Optimum Strata Boundaries (OSB) when more than one variable is available for stratification. The problem is converted into Mathematical Programming Problem (MPP) and the solution is cleared by executing a programme coded in LINGO. The results obtained through the proposed method has been compared with the methods developed by Dalenius and Hodges (1959) cum $\sqrt{f}$, Gunning and Horgan (2004), Lavallee - Hidiroglou (1988) using Kozak's (2004) and with Khan *et al.* (2008) mathematical programming approach from which the proposed method gives the best solution with highest level of precision. Thus, the study also indicates its superiority over both classical methods as well as Mathematical Programming methods. It indicates that using of two auxiliary variables as stratifying variables leads towards the increasing trend efficiency. The future work could be making comparisons for different allocation methods using two auxiliary variables as the basis of stratification. However, from perspective more than two auxiliary variables can be used for stratification.

RECEIVED: SEPTEMBER, 2018.
REVISED: MARCH, 2019

REFERENCES

[1] AOYAMA, H. (1954): A study of stratified random sampling. **Ann. Inst. Stat. Math**., **6**, 1-36.
[2] BELLMAN, R.E. (1957): **Dynamic Programming**. Princetown University Press, New Jersey.
[3] BHLER, W., AND DEUTLER, T. (1975): Optimal stratification and grouping by dynamic programming. **Metrika**, 22, 161-175.
[4] BOUZA,C (2002): Superpopulations models and stratification in diagnostic ability studies. **Brazilian J. of Prob. and Statistics**. 16, 47-58.
[5] DALENIUS, T. (1950): The problem of optimum stratification-II. **Skand. Aktuartidskr**, 33, 203-213.
[6] DALENIUS, T. AND GURNEY, M. (1951): The problem of optimum stratification. **Skand. Akt**., 34, 133-148.
[7] DALENIUS, T., and Hodges, J.L. (1959): Minimum variance stratification. **J. Amer. Statist. Assoc**.,54, 88-101.
[8] DANISH, F. AND RIZVI, S.E.H. (2018) Optimum stratification in Bivariate auxiliary variables under Neyman allocation, **Journal of Modern Applied Statistical Methods,** 17(1) DOI: 10.22237/jmasm/1529418671.
[9] DANISH, F., RIZVI, S.E.H. JEELANI, M. I REASHI J.A. (2017) Obtaining Strata Boundaries under Proportional Allocation with Varying Cost of Every Unit **Pak.J.Stat.Oper.Res**. 13, 567-574

[10] DANISH, F., RIZVI, S.E.H. JEELANI, SHARMA, M.K. AND M. I.J (2017) Optimum Stratification Using Mathematical Programming Approach: A Review. **Stat. Appl. Pro. Lett**. 4, 123-129
[11] DANISH, F. AND RIZVI, S.E.H. (2017):On optimum stratification using mathematical programming approach. **International Research Journal of Agricultural Economics and Statistics**. 8,435-439.
[12] EKMAN, G. (1959): Approximate expression for conditional mean and variance over small intervals of a continuous distribution. **Ann. Inst. Stat. Math**., 30, 1131-1134.
[13] GUNNING, P. AND HORGAN, J. M. 2004. A new algorithm for the construction of stratum boundaries in skewed populations. **Survey Methodology,** 30, 159-166.
[14] HESS, I., SETHI, V.K. , AND BALAKRISHNAN, T.R. (1966): Stratification: A practical Investigation. **J. Amer. Statist. Assoc**., 61, 71-90.
[15] KHAN, E.A., KHAN, M.G.M., AND AHSAN, M.J. (2002): Optimum stratification: A mathematical programming approach. **Calcutta Statistical Association Bulletin**, 52 (special), 205-208.
[16] KHAN, M.G.M., NAJMUSSEHAR, AND AHSAN, M.J. (2005): Optimum Stratification for Exponential Study Variable under Neyman Allocation. **Journal of the Indian Society of Agricultural Statistics,** 59, 146-150.
[17] KHAN, M. G. M., NAND, N. AND AHMAD, N. (2008): Determining the optimum strata boundary points using dynamic programming. **Survey Methodology**, 34, 205-214.

[18] KHAN, M.G.M. ,DINESH, R. ,ANSARI, A.H. AND AHSAN, M.J. (2015): Determining optimum strata boundaries and sample sizes for skewed population with log-normal distribution. **Communication in Statistics-Simulation and Computation,** 44,1364-1387.
[19] KOZAK, M. , VERMA, M.R. AND ZIELINSKI , A. (2004):Modern approach to optimum stratification: Review and perspectives. **Statistics in Transition**, 8, 223-250.
[20] MAHALANOBIS, P.C. (1952): Some Aspects of the Design of Sample Surveys. **Sankhya,** 12, 1-7.
[21] SETHI, V.K. (1963): A note on optimum stratification of population for estimating the population mean. **Aust. J. Statist.,** 5, 20-23.
[22] SINGH, R. (1977): A note on optimum stratification for equal allocation with ratio and regression methods of estimation. **Aust. J. Statist**., 19, 96-104.
[23] SINGH, R. AND SUKHATME, B.V. (1969): Optimum stratification. **Ann. Inst .Stat. Math**., 21, 515-528.
[24] SINGH, R. AND SUKHATME, B.V. (1973): Optimum stratification with ratio and regression methods of estimation. **Ann. Inst .Stat. Math**., 25, 627-633.
[25] UNNITHAN, V.K.G. (1978): The minimum variance boundary points of stratification. **Sankhya**, 40, C, 60-72.
[26] YADAVA, S.S. AND SINGH, R. (1984): Optimum stratification for allocation proportional to strata totals for simple random sampling scheme. **Commun. Statist. Theor. Math**., 13, 2793-2806.