International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Volume 7 Issue 3, March-2018, Impact Factor: 3.578

Optimum Strata Boundaries for Right Triangular and Gamma Auxiliary Variables with Single Study Variable

Faizan Danish[1], Yasmeena Ismail[2], Mir Subzar[3]

[1]Division of Statistics and Computer Science, Faculty of Basic Sciences, Main Campus, SKUAST-J, Chatha, Jammu (India)
[2,3]Division of Agricultural Statistics, SKUAST-K, Shalimar (India)

ABSTRACT

While using stratified sampling one has to select an appropriate stratification technique, which corresponds to minimum variance in order to increase the efficiency of the estimators. The stratification technique which results in minimum possible variance is known as optimum stratification. Strata may be constructed either on the basis of study variable itself or by using some other variable(s) highly correlated with the variable under study. In the present investigation, we are using two auxiliary variables having single study variable. The auxiliary variables are assumed to follow right triangular and gamma distribution. The proposed method under neyman allocation has been compared with [21] by estimating percent relative efficiency which clearly indicates the proposed method more preferable.

Keywords: Optimum Stratification, Neyman Allocation, Gamma Distribution

I.INTRODUCTION

The proper choice of strata boundaries In stratified random sampling, is one of the important factors as regards to the efficiency of estimator of population characteristic under consideration. The strata boundaries should be choosen, such that the strata are homogenous within itself and heterogeneous between them. The stratification can be done on the basis of study variable but in most cases the it is either difficult or the information is unknown regarding the study variable. In such cases stratification is done o the basis of variable (s) closely related to the study variable. [1] first considered the problem of optimum stratification for one variable using study variable itself as the basis of stratification. [2] developed a technique of obtaining optimum strata boundaries (OSB) using an auxiliary variable closely related to study variable. [3] considered the problem of finding approximately optimum strata boundaries (AOSB) on an auxiliary variable for one estimation variable case. The problem of optimum stratification, when two or more characters of the population are under study, seems to be relatively of greater practical importance. They provide some approximate solutions for the strata boundaries by using an auxiliary variable as the stratification variable under Neyman and proportional allocations. [4] tackled the problem for proportional and equal allocation procedures. [5] considered the problem of finding OSB when sample sizes to different strata are allocated in proportion to strata totals of the auxiliary variable. several other authors made an attempt to solve the problems related to strata boundaries such as [6] , [2] , [7] , [8], [9] , [10],[11] ,[12] and [13] , [14], [15] etc. [16] developed the theory of optimum stratification for bivariate case, on the basis of auxiliary variable, in case of simple random sampling. [17], [18]) applied their procedure to determine OSB to the population various distributions.[19] made an attempt to present all the developed methods introduced for construction of stratification points using mathematical programming technique. Also, [20] proposed a method for determing OSB for single study variable having one auxiliary variable when the cost of every unit varies in the whole strata.

Thus, in the present paper the problem of optimum stratification has been considered for two auxiliary variables having single study variable. The problem is formulated as NLPP which is solved by dynamic programming. The results obtained are compared with [21] by estimating percent relative efficiency

II.FORMULATION OF PROBLEM AS MPP

Let there be a finite population consisting of N units, for which it is required to estimate the total or mean for the characteristic Y under study, using simple random sampling technique. In order to have this, we divide the whole population into L × M strata on the basis of two auxiliary variables say, X and Z, such that the number of units in the (h , k)th stratum is N_{hk} so that

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

A sample of size 'n' is to be drawn from the whole population and suppose that the allocation of sample size to the (h , k)th stratum is n_{hk} such that

$$\sum_{h=1}^{L}\sum_{k=1}^{M} n_{hk} = n$$

The value of population unit in the (h , k)th stratum be denoted by y_{hki} (i = 1 , 2 , 3,…, N_{hk}) and then the population total is

$$Y = \sum_{h=1}^{L}\sum_{k=1}^{M}\sum_{i=1}^{N_{hk}} y_{hki}$$

Since the study variable is denoted by 'Y'. The unbiased estimate of population mean $\overline{Y}$ is

$$\overline{y}_{st} = \sum_{h=1}^{L}\sum_{k=1}^{M} W_{hk}\overline{y}_{hk}$$

Where $W_{hk} = \frac{N_{hk}}{N}$ denotes the stratum weight for the (h , k)th and $\overline{y}_{hk} = \frac{1}{n_{hk}}\sum_{i=1}^{n_{hk}} y_{hki}$

For stratified simple random sampling, the sample estimate $\overline{y}_{st}$ is unbiased and its sampling variance is given as below:

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

where $f_{hk} = \frac{n_{hk}}{N_{hk}}$ denotes the sampling fraction in the $(h,k)^{th}$ stratum.

If the finite population correction (f.p.c) is ignored, the variance of the estimate is given by

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

σ_{hky}^{2} represents the population variance for the character Y in the $(h,k)^{th}$ stratum and is defined as

$$\sigma_{hky}^{2} = \frac{1}{N_{hk}}\sum_{i=1}^{N_{hk}}\left(y_{hki} - \overline{y}_{hk}\right)^{2}$$

When the study variable 'Y' itself is not used for stratification variable, we propose a model based on two auxiliary variables. Let the regression model of study variable on auxiliary variables is of the form as:

$$Y = \lambda\left(x,z\right) + e \qquad (2)$$

where, $\lambda\left(x,z\right)$ be a linear or non-linear function of 'X' and 'Z' and 'e' denotes the error term such that

$E\left(e \mid x,z\right) = 0$ and $V\left(e \mid x,z\right) = \phi\left(x,z\right)$ for all (x, z)

Under model (5.1.2) the stratum mean 'μ_{hky}' and the stratum variance 'σ_{hky}^{2}' can be written as

$$\mu_{hky} = \mu_{hk\lambda} \qquad (3)$$

and

$$\sigma_{hky}^{2} = \sigma_{hk\lambda}^{2} + \mu_{hk\phi} \qquad (4)$$

where $\mu_{hk\lambda}$ and $\mu_{hk\phi}$ are the expected values of $\lambda\left(x,z\right)$ and $\phi\left(x,z\right)$,respectively and $\sigma_{hk\lambda}^{2}$ denotes the variance of $\lambda(x,z)$ in the (h, k)th stratum.

If 'λ' and 'ε' are uncorrelated, then in model (1) 'σ_{hky}^{2}' can be expressed as

$$\sigma_{hky}^{2} = \sigma_{hk\lambda}^{2} + \sigma_{hk\varepsilon}^{2} \qquad (5)$$

where $\sigma^2_{hk\varepsilon}$ is the variance of error term in (h, k)th stratum.

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

If the population mean of the study variable 'Y' is estimated under the variance given in equation (1) ,then the problem of determining the strata boundaries is to cut up the ranges $d_x = b - a$ and $t_z = d - c$,at (L-1) and (M-1) intermediate points as $a = x_0 \le x_1 \le \ldots \le x_{L-1} \le x_L = b$ and $c = z_0 \le z_1 \le \ldots \le z_{M-1} \le z_M = d$, respectively, such that the equation (1) is minimum.

For a fixed size 'n', minimizing the expression of the right hand side of (1) is equivalent to minimizing

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

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

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

If $f(x,z), \lambda(x,z)$ and $\phi(x,z)$ are known and also integrable then, $W_{hk}, \sigma^2_{hk\lambda}$ and $\mu_{hk\phi}$ can be obtained as a function of boundary points $(x_{h-1}, x_h, z_{k-1}, z_k)$ by using the following expression

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

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

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,

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

and $(x_h, x_{h-1}, z_k, z_{k-1})$ are the boundary points of the (h, k)th stratum.

Thus, the objective function (6) could be expressed as the function of boundary points $(x_{h-1}, x_h, z_{k-1}, z_k)$ only.

Let

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

and the ranges as:

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

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

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 (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 functions are known then σ^2_{hky} can be expressed as a function of boundary 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 or width of the (h, k)th stratum. Then, using (12) and (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 (14) suggests that, for determination of two way stratification, a two-dimensional dynamic programming approach should be used. Employing the general concept of dynamic programming with the state and decision variables by the pairs (h, k).Then problem of two way optimum stratification can be expressed as to

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

Though the formulation of (17) seems to be difficult, it can in fact be seen with respect to the decision space $B_h(x_{h-1}) \times B_k(z_{k-1})$ .This decision space depends upon the states of both past and future stages since the variable x_{h-1} is present in the decision space of the M stages (h , 1) to (h , M) and the variable z_{k-1} in the decision spaces of the L stages (1,k) to (L , K).Hence, Bellman's principle of optimality that states, "An optimal policy has the property that, whatever the initial state and decisions are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decisions" is not applicable. It should be noted that the problem is partially due to the way the stages have been defined. In fact instead of $L \times M$ stages which can be viewed through the decision variables V_h and U_k .However, due to nature of objective function given in general by (14),its transformation to reflect the (L x M) stages does not seem to be mathematically tractable for most allocations.

We propose a simple approach which permits a solution to the problem (17) using the unidmensional dynamic programming iteratively. Before the first iteration, some trail values say x_0 and z_0, such that $a = x_0 \le x_1 \le \ldots \le x_{L-1} \le x_L = b$ and $c = z_0 \le z_1 \le \ldots \le z_{M-1} \le z_M = d$ are choosen for the initial points of the stratification. Then for the ith iteration (i=1,2,...) the points of stratification z^{i-1} are first considered as fixed. Note that the points of stratification x^{i-1}could also be chosen instead of z^{i-1}.Fixing the values of z^{i-1} has in fact the effect of reducing the problem exactly to the one of two-way optimum stratification with one categorical stratification variable. This can be seen by comparing the formulation (17) to the one which is defined on univarite auxiliary variable used as stratification variable with the values of the points of stratification Z taken as constant in (17).

Let $\phi^*_{x_h}\left(x_{h-1}, z^{i-1}\right)$ be the optimal value for the objective function (14) for the strata (h, k) to (L, k) for all k=1,2,...,M given that the lower bound for the strata (h, k) for k = 1,2,...,M is x_{h-1} .The functional equation of Bellman with respect to the first part of the ith iteration is then given by

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

$B_h(x_{h-1})$ is defined in (17).

Using this last equation, new points of stratification x^i with respect to the variable 'X' can be obtained to response the proceeding value x^{i-1} .Hence, the OSB for the first part of the ith iteration are given by $\left(x^i, z^{i-1}\right)$ .For the second part of the ith iteration ,the points of stratification x^i are in turn considered as fixed. Restating the problem of determining OSB as the problem of determining optimum points (V_h, U_k) ,adding equation (15) and (16) as a constraint ,the problem (14) can be treated as an equation problem of determining Optimum Strata Width (OSW), $V_1, V_2, \ldots, V_L \, and \, U_1, U_2, \ldots, U_M$ and is expressed as the following Mathematical Programming Problem (MPP):

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

Subject to (18)

$\sum_h V_h = d_x$

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

and

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

Initially, (x_0, z_0) the initial values of the auxiliary variables X and Z, respectively, are known. Therefore, the first term $\phi_{11}(x_1, x_0, z_1, z_0)$ in the objective function (18) is the function of (V_1, U_1) alone, once the (V_1, U_1) is known. The second term $\phi_{22}(x_2, x_1, z_2, z_1)$ will be the function of (V_2, U_2) alone, and so on. Due to special nature of function, the MPP (18) may be treated as the function of (V_h, U_k) and can be expressed as

Minimize $\sum_h \sum_k \phi_{hk}(V_h, U_k)$

Subject to (19)

$\sum_h V_h = d_x$

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

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

III.NEYMAN ALLOCATION

This allocation of the total sample size to different strata is called minimum variance allocation and given by Neyman (1934).In this case, the allocation of samples among different strata is based on a joint consideration of the stratum size and the stratum variance. In this allocation, it is assumed that the sampling cost per unit among different strata is constant and size of the sample is fixed. Sample sizes are allocated by

$$n_{hk} = \frac{nW_{hk}\sigma_{hky}}{\sum\sum W_{hk}\sigma_{hky}}$$

A formula for the minimum variance with fixed 'n' is obtained by substituting the value of n_{hk}

$$\sum_h \sum_k \frac{(1-f)}{n_{hk}} W_{hk}^2 \sigma_{hky}^2$$

then we get

$$V\left(\bar{y}_{st}\right) = \frac{\left(\sum_h \sum_k W_{hk}\sigma_{hky}\right)^2}{n} - \frac{\sum_h \sum_k W_{hk}\sigma_{hky}^2}{N} \qquad (20)$$

There may be difficulty in using this as the value of σ_{hky} will usually be unknown. However, the stratum variance may be obtained from previous surveys or from a specially planned pilot survey. The other alternative is to conduct the main survey in a phased manner and utilize the data collected in the first phase for ensuring better allocation in the second phase.

If the finite population correction is ignored, equation (20) can be written as

$$\frac{1}{n}\left(\sum_h \sum_k W_{hk}\sigma_{hky}\right)^2$$

However, minimising this is equivalent to minimize (since 'n' is fixed constant)

$$\sum_h \sum_k W_{hk}\sigma_{hky}$$

Let us assume the regression model defined in equation (2) be linear as:

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

then $\sigma^2_{hky} = \beta^2 \sigma^2_{hkx} + \gamma^2 \sigma^2_{hkz}$ and the weight and variance of the (h, k)th stratum having auxiliary variables as 'X' and 'Z'.

$$W_{hk} = \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} f(x,z)\,\partial x \partial z \tag{21}$$

$$\sigma^2_{hkx} = \frac{1}{W_{hk}} \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} x^2 f(x)\,\partial x \partial z - \mu^2_{hkx} \tag{22}$$

$$\sigma^2_{hkz} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} z^2 f(z)\,\partial z \partial x - \mu^2_{hkz} \tag{23}$$

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

Then

$$\sum_h \sum_k W_{hk} \sqrt{\beta^2 \sigma^2_{hkx} + \gamma^2 \sigma^2_{hkz}} \tag{24}$$

Thus, under Neyman allocation for obtaining OSB we need to minimize the objective function (21) for which MPP is written as:

Minimize $\sum_h \sum_k W_{hk} \sqrt{\beta^2 \sigma^2_{hkx} + \gamma^2 \sigma^2_{hkz}}$

Subject to

$$\sum_h V_h = d_x$$

$$\sum_k U_k = t_z \qquad \forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{matrix} h = 1,2,\ldots,L \\ k = 1,2,\ldots,M \end{matrix} \tag{25}$$

IV.MPP FOR RIGHT TRIANGULAR AND GAMMA DISTRIBUTION WITH SINGLE STUDY VARIABLE

If the X variable has right triangular distribution with the pdf as

$$f(x) = \begin{cases} 2(2-x) & ; 0 \leq x \leq 1 \\ 0 & ; otherwise \end{cases} \tag{26}$$

and the variable Z follows Gamma Distribution with pdf as

$$f(z) = f(z,s,\theta) = \begin{cases} \frac{1}{\theta^s \Gamma(s)} z^{s-1} e^{-\frac{z}{\theta}} & ; \quad z \geq 0, s, \theta > 0 \\ 0 & , otherwise \end{cases} \tag{27}$$

where's' is the slope and 'θ' is the scale parameter and Γs is a gamma distribution function defined as

$$\Gamma s = \int_0^\infty e^{-z} z^{s-1} \partial z \qquad ; s > 0$$

this function is also defined by the upper incomplete gamma function $\Gamma s, z$ and a lower incomplete gamma function γ(s,z) ,respectively as

$$\Gamma(s,z) = \int_z^\infty u^{s-1} e^{-u} \partial u$$

and

$$\gamma(s,z) = \int_0^z u^{s-1} e^{-u} \partial u$$

There also exists normalized incomplete gamma function which gives a value restricted between '0' and '1' that can be stated as

$$Q(s,z) = \frac{1}{\Gamma(s)} \int_z^{\infty} u^{s-1} e^{-u} \partial u \qquad , \quad s, z > 0$$

$$P(s,z) = \frac{1}{\Gamma(s)} \int_0^{z} u^{s-1} e^{-u} \partial u \qquad , \quad s, z > 0, \Gamma(s) \neq 0$$

Where Q(s, z) denotes the upper regularized incomplete gamma function with P(s,z) denotes regularized lower incomplete gamma.

In order to obtain OSB having distribution function of two auxiliary variables defined in (23) and (24) equations we have to find the value of W_{hk}, σ^2_{hkx} and σ^2_{hkz} for that substitute above distribution functions in (21)-(23),we have

$$W_{hk} = Q_1 V_h (4 - V_h - 2x_{h-1}) \tag{28}$$

$$\sigma^2_{hkx} = \frac{q_1 - U_k \left[2V_h (V_h + 2x_{h-1}) - \frac{2}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) \right]}{Q_1^2 V_h^2 (4 - V_h - 2x_{h-1})^2} \tag{29}$$

$$\sigma^2_{hkz} = \frac{\theta^2 s Q_1 (s+1)[Q_2] - \theta^2 s^2 (4 - V_h - 2x_{h-1})[Q_3]^2}{Q_1^2 (4 - V_h - 2x_{h-1})} \tag{30}$$

where

$$q_1 = V_h U_k Q_1 (4 - V_h - 2x_{h-1})$$
$$\left\{ \frac{4}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) - \frac{1}{2}\left[(V_h + 2x_{h-1})\left(V_h^2 + 2x_{h-1}^2 + 2V_h x_{h-1}\right) \right] \right\}$$

$$Q_1 = Q\left(s, \frac{z_{k-1}}{\theta}\right) - Q\left(s, \frac{z_k}{\theta}\right) \quad , \quad Q_2 = Q\left(s+2, \frac{z_{k-1}}{\theta}\right) - Q\left(s+2, \frac{z_k}{\theta}\right)$$

and $Q_3 = Q\left(s+1, \frac{z_{k-1}}{\theta}\right) - Q\left(s+1, \frac{z_{k-1} + U_k}{\theta}\right)$

By substituting values obtained in equations (28) to (30) in (25) ,we have

$$\text{Minimize} \sum_h \sum_k Q_1 V_h (4 - V_h - 2x_{h-1}) \sqrt{ \beta^2 \frac{q_1 - U_k \left[2V_h (V_h + 2x_{h-1}) - \frac{2}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) \right]}{Q_1^2 V_h^2 (4 - V_h - 2x_{h-1})^2} + \gamma^2 \frac{\theta^2 s Q_1 (s+1)[Q_2] - \theta^2 s^2 (4 - V_h - 2x_{h-1})[Q_3]^2}{Q_1^2 (4 - V_h - 2x_{h-1})} }$$

Subject to

$$\sum_h V_h = d_x$$
$$\sum_k U_k = t_z \tag{31}$$
$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad h = 1, 2, \dots, L; \quad k = 1, 2, \dots, M$$

V.THE SOLUTION PROCEDURE

The problem (19) is a problem of multistage decision in which the objective function and the constraints are separable functions of (V_h, U_k), which allows us to use a dynamic programming technique. Dynamic programming determines optimal solution of a multi-variable problem by decomposing into stages, each stage compromising a single variable sub problem. A dynamic programming model is generally a recursive equation. These recursive equation links to different stages of the problem.

Consider the following sub problem of equation (19) for first $(L_1 \times M_1)$ strata, where $(L_1 \times M_1) \le (L \times M)$, i, e $L_1 < L, M_1 < M$

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

Subject to

$$\sum_{h=1}^{L-1} V_h = d_{L_1} \tag{32}$$

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

and $V_h \ge 0 \quad and \quad U_k \ge 0$

where $d_{L_1} < d_x, t_{M_1} < t_z$

Note: If $d_{L_1} = d_x$ and $t_{M_1} = t_z$ then $(L_1 \times M_1) = (L \times M)$

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 $\phi_{L_1 \times M_1}\left(V_{L_1} \times U_{M_1}\right)$ denotes the minimum value of the objective function of the equation (32), that is,

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

$$and \quad V_h \ge 0, U_k \ge 0; h = 1, 2, 3, \ldots, L_1 \quad ; \quad k = 1, 2, 3, \ldots, M_1$$

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 $L_1 = 1, 2, \ldots, L$ and $M_1 = 1, 2, \ldots, M$; $0 \le d_{L_1} \le V, 0 \le t_{M_1} \le U$.

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

$$and \quad V_h \ge 0, U_k \ge 0; h = 1, 2, 3, \ldots, L_1 \quad and \quad k = 1, 2, 3, \ldots, M_1 \tag{33}$$

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

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

$V_h \ge 0$, $h = 1, 2, \ldots, L_1$, $U_k \ge 0$, $k = 1, 2, \ldots, M_1$, $1 \le L_1 \le L$, $1 \le M_1 \le M$

Using the same procedure to write the forward recursive equation of the dynamic programming technique and could obtain OSB.

VI.NUMERICAL ILLUSTRATION

For computation details let us define the interval of both variables as $x \in [0, 1]$ and $z \in [0, 6]$ by generating the data in R-software using the pdf's (26) and (27) to estimate β = 0.81 and
γ = 0.72.Also the minimum likelihood estimates of the parameters of the gamma distribution while compiling generated data of distribution are s = 4.3829 and θ = 3.0124.By substituting these values in MPP (31), we get
Minimize

$$\sum_h \sum_k Q_1' V_h \left(4 - V_h - 2x_{h-1}\right) \sqrt{ \begin{array}{l} (0.6561) \dfrac{q_1' - U_k \left[2V_h\left(V_h + 2x_{h-1}\right) - \dfrac{2}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) \right]}{Q_1'^2 V_h^2 \left(4 - V_h - 2x_{h-1}\right)^2} \\ +(161.8539) \dfrac{Q_1'\left[Q_2'\right] - \theta^2 s^2 \left(4 - V_h - 2x_{h-1}\right)\left[Q_3'\right]^2}{Q_1'^2 \left(4 - V_h - 2x_{h-1}\right)} \end{array} }$$

Subject to

$$\sum_h V_h = 1$$

$$\sum_k U_k = 6 \tag{34}$$

$$\forall V_h \ge 0, U_k \ge 0 \quad , \quad \begin{array}{l} h = 1, 2, \ldots, L \\ k = 1, 2, \ldots, M \end{array}$$

where

$$q_1' = V_h U_k Q_1' \left(4 - V_h - 2x_{h-1}\right) \left\{ \frac{4}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) - \frac{1}{2}\left[\left(V_h + 2x_{h-1}\right)\left(V_h^2 + 2x_{h-1}^2 + 2V_h x_{h-1}\right) \right] \right\}$$

$Q_1^{'} = Q\left(4.3829, \frac{z_{k-1}}{3.0124}\right) - Q\left(4.3829, \frac{z_k}{3.0124}\right), Q_2^{'} = Q\left(6.3829, \frac{z_{k-1}}{3.0124}\right) - Q\left(6.3829, \frac{z_k}{3.0124}\right)$, and

$Q_3^{'} = Q\left(5.3829, \frac{z_{k-1}}{3.0124}\right) - Q\left(5.3829, \frac{z_{k-1}+U_k}{3.0124}\right)$

The MPP (34) is executed in the LINGO with total target of 6 (2×3) strata in which 2 are to be made along X variable and 3 along Z variable. The OSB so obtained is shown in the following Table

Table 5.6.4: OSB and Variance of proposed method when the auxiliary variables X and Z have right-triangular and gamma distribution respectively

OSB (x_h, z_k)	Variance (Proposed method)	Variance (Reddy et al. 2016)	% R.E.
(0.4231,1.9237) (1.0000,1.9237) (0.4231,3.6829) (1.0000,3.6829) (0.4231,6.0000) (1.0000,6.0000)	0.004869	0.0079251	162.76

The above tables show us the OSB under Neyman allocation when one of the auxiliary variables is having Right triangular distribution and the other is having Gamma distribution.

CONCLUSION

The problem of determining OSB is discussed by many authors mostly either separately or they determined the OSB under a particular allocation. The OSB so obtained may be infeasible or sub-optimum, especially for small and skewed populations. In this paper, an attempt has been made to solve the problem and the problem so obtained is formulated as MPP, which is solved by developing a coded in a user friendly software LINGO. Furthermore, the variance obtained through proposed method when comparing with the [21] results in the percent of relative efficiency as 162.76 which is clearly indicates that the proposed method is more preferable.

REFERENCES

[1] Dalenius, T., The problem of optimum stratification, Skandinavisk Aktuarietidskrift, 33, 1950, 203-213.
[2] Dalenius, T. and Gurney, M., The problem of optimum stratification II, Skandinavisk Aktuarietidskrift, 34, 1951, 133-148
[3] Singh, R. and Sukhatme, B.V., Optimum stratification. Annals of Institute of Statistical Mathematics, 21, 1969, 515-528.
[4] Singh, R., An alternative method of stratification on the auxiliary variable, Sankhya C , 37, 1975,100-108
[5] Yadava, S.S. and Singh, R., Optimum stratification for allocation proportional to strata totals for simple random sampling scheme, Communications in Statistics –Theory and Methods, 13, 1984, 2793-2806.
[6] Mahalanobis, P.C., Some aspect of design of sample surveys. Sankhya , 12 , 1952,1-17.
[7] Singh, R. and Sukhatme, B.V., Optimum stratification in sampling with varying probabilities, Annals of Institute of Statistical Mathematics, 24, 1972 ,485-494.
[8] EA Bhardwaj, RK Sharma, EA Bhadoria, A Case Study of Various Constraints Affecting Unit Commitment in Power System Planning, International Journal of Enhanced Research in Science Technology & Engineering, 2013.
[9] Singh, R. and Sukhatme, B.V., Optimum stratification with ratio and regression methods of estimation, Annals of Institute of Statistical Mathematics, 25,1973, 627-633.
[10] Singh, R. and Prakash, D., Optimum stratification for equal allocation, Annals of Institute of Statistical Mathematics, 27, 1975, 273-280.
[11] Singh, R., Approximately optimum stratification on auxiliary variable, Journal of American Statistical Association, 66, 1971, 829-833.
[12] Singh, R., An alternative method of stratification on the auxiliary variable, Sankhya C , 37, 1975a, 100-108
[13] Singh, R. 1975b. A note on optimum stratification in sampling with varying probabilities. Australian Journal of Statistics, 17: 12-21.
[14] Serfling, R.J., Approximately optimum stratification, Journal of American Statistical Association, 63, 1968, 1298-1309.
[15] Rizvi , S.E.H. Gupta, J.P. and Bhargava, M. , Optimum stratification based on auxiliary variable for compromise allocation, Metron, 28(1), 2002, 201–215.
[16] Khan, M. G. M., Najmussehar and Ahsan, M. J. Optimum stratification for exponential study variable under Neyman allocation. Journal of the Indian Society of Agricultural Statistics, 59(2), 2005, 146-150.
[17] Rizvi, S.E.H., Gupta, J.P. and Singh, R. , Approximately optimum stratification for two study variables using auxiliary information, Journal of the Indian Society of the Agricultural Statistics, 53 (3), 2000, 287-298.

[18] Khan, E. A., Khan, M. G. M. and Ahsan, M. J., Optimum stratification: A mathematical programming approach, Calcutta Statistical Association bulletin, 52, 2002,323-333.

[19] Khan, M. G. M., Nand, N. and Ahmad, N., Determining the optimum strata boundary points using dynamic programming, Survey Methodology, 34(2), 2008,205-214.

[20] Er Amit Bhardwaj, Amardeep Singh Virdi, RK Sharma, Installation of Automatically Controlled Compensation Banks, International Journal of Enhanced Research in Science Technology & Engineering, 2013.

[21] Danish, F., Rizvi, S.E.H. Jeelani, M. I and Reashi J.A., Obtaining Strata Boundaries under Proportional Allocation with Varying Cost of Every Unit, Pakistan Journal of Statistics and Operations Research, 13(3), 2017a , 567-574

[22] Danish, F., Rizvi, S.E.H., Sharma, M.K. and Jeelani, M. I.,Optimum Stratification Using Mathematical Programming Approach: A Review, Journal of Statistics Applications & Probability Letters, 4(3), 2017b ,123-129

[23] [Reddy, K.G. , Khan, M.G.M. and Rao, D.K., A procedure for computing optimal strata boundaries and sample sizes for multivariate surveys, Journal of Software, 11(8), 2016, 816-832.