Construction of stratification points under optimum allocation using dynamic programming

Faizan Danish
Division of Statistics and Computer Science
Faculty of Basic Sciences
SKUAST-JAMMU,J&K,India-180009
danishstat@gmail.com

Abstract

In the present investigation, some theory has been developed for optimum stratification, when two auxiliary variables treated as the basis of stratification with one study variable under study. The problem has been formulated as mathematical programming problem and then solved by dynamic programming. Empirical studied have been made to illustrate the proposed method with the comparisons of other existing methods.

Keywords: Optimum allocation; Multistage decision problem; Cost function

Introduction

Let the target population consisting of 'N' units be stratified into L×M strata based on two auxiliary variables 'X' and 'Z' when the estimation of mean of the study variable 'Y' is of interest. In order to have the estimate ,we divide the whole population into the desired number of strata like L×M such that each strata is homogenous within itself and heterogeneous between strata with respect to the character under study 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 selected from a population 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 population unit in the (h ,k)th stratum ,then the population total can be expressed as

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

Under stratified random sampling the unbiased estimator of the population mean $\overline{Y_N}$,is

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

where ,

$\bar{y}_{hk} = \frac{1}{n_{hk}}\sum_{i} y_{hki}$ and 'W_{hk}' denotes the weight of the (h, k)th stratum as given as

$W_{hk} = \frac{N_{hk}}{N}$

The basic consideration involved in the determination of optimum strata boundaries (OSB) is that the strata should be internally as homogenous as possible, that is, the stratum variances should be as small as possible, given some sample allocation. When it is difficult to made stratification on the basis of study variable we make use of information highly related with the stud variable known as auxiliary information. The determination of OSB was pioneered by Dalenius (1950). Moreover several other authors have developed methods using study variable as stratification variable as well as auxiliary variable used as stratification variable. Dalenius and Gurney (1951) developed a technique with respect to an auxiliary variable closely related to study variable. Singh and Sukhatme (1969) considered the problem of finding approximately optimum strata boundaries on an auxiliary variable for one estimation variable. Further several authors have developed several methods at different allocation procedures like Singh, R. (1975), S.E.H. Rizvi *et al*. (2002) , Khan *et al*. (2003),Gupta et al. (2005),Rao *et al*. (2012),Khan *et al*. (2015).Danish and Rizvi (2017a) discussed made an attempt to cover all the developed methods made towards construction of strata boundaries. Danish *et al* .(2017) developed a method of obtaining optimum strata boundaries when the cost of every unit varies in the whole strata. Further development on optimum strata boundaries has been made by Danish and Rizvi (2018), Danish, F. (2018) and Danish and Rizvi (2019).
In this paper, the author is going to develop a method under optimum allocation using two stratification variables and one study variable.

Formulation of Problem under Optimum allocation

In stratified sampling, variance of the estimator depends on values of n_{hk} (h=1,2,...,L ;1,2,3,...,M) apart from values of Y. Even for a fixed n, the values of the $V\left(\bar{y}_{st}\right)$ will differ for different configurations $\left(n = n_{11},...,n_{LM}\right)$.The combination 'n' for which $V\left(\bar{y}_{st}\right)$ is minimum among all the values of variances for different possible combinations 'n' is an optimal allocation for a fixed n. Since sampling involves cost in a survey, consider the linear cost function

$$C = C_0 + \sum_h \sum_k C_{hk} n_{hk} \qquad (1)$$

where C_0 is an overhead cost (e.g cost of setting up and maintaining an office ,recruiting survey personal and other capital expanses etc.) C_{hk} is the cost of sampling unit from the $(h,k)^{th}$ stratum and C is the total cost.(1) equation is a very simple and reasonable cost function expressing the total cost of field operation. The variance under stratified random sampling is

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

(2)

To determine the value of n_{hk} , consider the function

$$\psi = V\left(\bar{y}_{st}\right) + \lambda C \qquad (3)$$

where λ is the known constant.

Using the calculus method of Langrangian multiplier, select n_{hk} and a constant λ to minimize ψ .

Differentiating (3) w.r.to n_{hk} ,we have

$$-\frac{W_{hk}^2\sigma_{hky}^2}{n_{hk}^2}+\lambda c_{hk}=0 \qquad \begin{matrix} ,h=1,2,...,L \\ K=1,2,...,M \end{matrix}$$

$$\therefore n_{hk}=\frac{W_{hk}\sigma_{hky}}{\sqrt{\lambda c_{hk}}} \tag{4}$$

while taking summation on both sides ,we get

$$n=\sum_h\sum_k\left(\frac{W_{hk}\sigma_{hky}}{\sqrt{\lambda c_{hk}}}\right) \tag{5}$$

From (4) and (5),we can obtain

$$n_{hk}=n\frac{W_{hk}\sigma_{hky}\Big/\sqrt{c_{hk}}}{\sum_h\sum_k\left(W_{hk}\sigma_{hky}\Big/\sqrt{c_{hk}}\right)} \tag{6}$$

Thus (6) relation leads to the following important conclusion that, in a given strata, we have to take a large sample size if

i. the stratum size is large.
ii. the stratum has large variability.
iii. the cost per unit is cheaper in the stratum.

The sample size ' n_{hk} ' from $(h,k)^{th}$ stratum required for estimating the population with a specified cost 'C' is given by

$$n_{hk}=\frac{(C-c_0)W_{hk}\sigma_{hky}\Big/\sqrt{c_{hk}}}{\sum_h\sum_k\left(W_{hk}\sigma_{hky}\Big/\sqrt{c_{hk}}\right)} \tag{7}$$

If $c_{hk}'s$ are the same from stratum to stratum, relation (7) will lead to Neyman allocation. Similarly if $c_{hk}'s$ and $\sigma_{hky}'s$ doesn't vary from stratum to stratum, relation (6) leads to proportional allocation.

Now, substituting (7) in (2),we get

$$V\left(\bar{y}_{st}\right)=\frac{\left(\sum_h\sum_k W_{hk}\sigma_{hky}\sqrt{c_{hk}}\right)^2}{(C-c_0)}-\sum_h\sum_k\frac{W_{hk}\sigma_{hky}^2}{N_{hk}}$$

(8)

If the finite population correction is ignored ,then minimizing the expression on right hand side of (8) is the same as minimizing

$$V\left(\bar{y}_{st}\right)=\sum_{h}\sum_{k}W_{hk}\sigma_{hky}\sqrt{c_{hk}}$$

(9)

When the study variable 'Y' itself is not used for stratification variable, we propose a model based on bi-variate stratified sampling design. Let the regression model of study variable on auxiliary variables is of the form as

$$Y=\lambda\left(x,z\right)+\varepsilon \qquad (10)$$

where, $\lambda\left(x,z\right)$ be a linear function of 'X' and 'Z' and is equal to $\alpha+\beta x+\gamma z$ and 'ε' denotes the error term such that

$$E\left(\varepsilon/_{(x,z)}\right)=0 \qquad \text{and} \qquad V\left(\varepsilon/_{(x,z)}\right)=\phi\left(x,z\right) \text{ for all (x,z)}$$

Under model (10) the stratum mean 'μ_{hky}' and the stratum variance 'σ^2_{hky}' can be written as

$$\mu_{hky}=\mu_{hk\lambda} \qquad (11)$$

and

$$\sigma^2_{hky}=\sigma^2_{hk\lambda}+\mu_{hk\phi} \qquad (12)$$

where $\mu_{hk\lambda}$ and $\mu_{hk\phi}$ are the expected value of $\lambda\left(x,z\right)$ and $\phi\left(x,z\right)$ respectively 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 'σ^2_{hky}' can be expressed as

$$\sigma^2_{hky}=\sigma^2_{hk\lambda}+\sigma^2_{hk\varepsilon} \qquad (13)$$

where $\sigma^2_{hk\varepsilon}$ is the variance of in (h, k)th stratum. It can be verified by (12) and (13) .

If 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 (2.9) ,then the problem of determining the strata boundaries is to cut up the ranges V=b-a and U = 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 (9) is minimum.

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

$$W_{hk}=\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k}f\left(x,z\right)\partial x\partial z \qquad (14)$$

$$\sigma^2_{hk\lambda}=\frac{1}{W_{hk}}\int_{x_{h-1}}^{x_h}\int_{z_{k-1}}^{z_k}\lambda^2\left(x,z\right)f\left(x,z\right)\partial x\partial z-\mu^2_{hk\lambda} \qquad (15)$$

and $$\mu_{hk\phi} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \phi(x,z) f(x,z) \partial x \partial z \qquad (16)$$

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

(17) where $(x_h - x_{h-1})$ and $(z_k - z_{k-1})$ are the boundary points of the (h,k)th stratum. If a simple random sample of size 'n_{hk}' drawn 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 $\bar{Y}_{hk}$.

However for 'X' and 'Z' independent of error term 'ε'

then

$$\sigma_{hky}^2 = \beta^2 \sigma_{hkx}^2 + \gamma^2 \sigma_{hkz}^2 \qquad (18)$$

Thus the variance that we need to minimize is

$$V(\bar{y}_{st}) = \sum_h \sum_k W_{hk} \left(\beta^2 \sigma_{hkx}^2 + \gamma^2 \sigma_{hkz}^2 \right) \sqrt{c_{hk}}$$

the weight and variance of the (h,k)th stratum having auxiliary variables as 'X' and 'Z'.

$$W_{hk} = \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} f(x,z) \partial x \partial z \qquad (19)$$

$$\sigma_{hkx}^2 = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} x^2 f(x) \partial x \int_{z_{k-1}}^{z_k} \partial z - \mu_{hkx}^2 \qquad (20)$$

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

(21)

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

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

Let $$\phi(x_h, x_{h-1}, z_k, z_{k-1}) = W_{hk} \sigma_{hky} \sqrt{c_{hk}} \qquad (22)$$

we have already let the range

$$d_x = b - a = x_L - x_0 \qquad (23)$$

$$t_z = d - c = z_M - z_0 \qquad (24)$$

Then, in the bivariate stratification a problem of determining the strata boundaries (x_h, z_k) is to break up the ranges of (23) and (24) 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 (25)

$$a = x_0 \le x_1 \le ... \le x_{L-1} \le x_L = b$$
$$c = z_0 \le z_1 \le ... \le z_{M-1} \le z_M = d$$

and

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

When, the marginal frequency function are known and σ^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 and width of the (h, k)th stratum. Then, using (23) and (24) ,the ranges can be expressed as

$$\sum_h V_h = \sum_h (x_h - x_{h-1}) = b - a = d_x \qquad (26)$$

$$\sum_k U_k = \sum_k (z_k - z_{k-1}) = d - c = t_z \qquad (27)$$

Let $\phi^*_{x_h}(x_{h-1}, z^{i-1})$ be the optimal value for the objective function (25) 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

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

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 (x^i, z^{i-1}) .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 (26) and (27) as a constraint ,the problem (25) can be treated as an equation problem of determining Optimum Strata Width (OSW) , $V_1, V_2, ..., V_L$ *and* $U_1, U_2, ..., U_M$ and is expressed as the following MPP:

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

Subject to (28)

$$\sum_h V_h = d_x$$

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

and

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

Initially, (x_0, z_0) are 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 (28) 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 (28) 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 (29)

$$\sum_h V_h = d_x$$

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

and

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

The solution Procedure

The problem (29) 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 equations link te different stages of the problem.

Consider the following sub problem of equation (29) for first $(L_1 \times M_1)$ strata, where $(L_1 \times M_1) \leq (L \times M)$,i,e $L_1 < L, M_1 < M$

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 (30)

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

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

and

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

where

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

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

The transformation functions are given by

$$d_{L_1} = V_1 + V_2 + ... + V_{L_1}$$
$$d_{L_1-1} = V_1 + V_2 + ... + V_{L_1-1} = d_{L_1} - V_{L_1}$$
$$d_{L_1-2} = V_1 + V_2 + ... + V_{L_1-2} = d_{L_1-1} - V_{L_1-1}$$
$$\vdots$$
$$d_2 = V_1 + V_2 = d_3 - V_3$$
$$d_1 = V_1 = d_2 - V_2$$

Similarly, we have

$$t_{M_1} = U_1 + U_2 + ... + U_{M_1}$$
$$t_{M_1-1} = U_1 + U_2 + ... + U_{M_1-1} = t_{M_1} - U_{M_1}$$
$$t_{M_1-2} = U_1 + U_2 + ... + U_{M_1-2} = t_{M_1-1} - U_{M_1-1}$$
$$\vdots$$
$$t_2 = U_1 + U_2 = t_3 - U_3$$
$$t_1 = U_1 = t_2 - U_2$$

Let

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

and $V_h \geq 0, U_k \geq 0; h = 1,2,3,...,L_1$ *and* $k = 1,2,3,...,M_1$

and $1 \leq L_1 \leq L$, $1 \leq M_1 \leq M$

Let $\phi_{L_1\times M_1}\left(d_{L_1}, t_{M_1}\right)$ denotes the minimum value of the objective function of the equation (30), that is,

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

and $V_h \geq 0, U_k \geq 0; h = 1,2,3,...,L_1$ *and* $k = 1,2,3,...,M_1$

with the above definition of $\phi_{L_1\times M_1}\left(V_{L_1}, U_{M_1}\right)$,the MPP (22) 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 $L_1 = 1,2,...,L$ *and* $M_1 = 1,2,...,M$; $0 \leq d_{L_1} \leq V$, $0 \leq t_{M_1} \leq U$.

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

and $V_h\geq 0, U_k\geq 0; h=1,2,3,...,L_1$ *and* $k=1,2,3,...,M_1$

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

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

and

$V_h\geq 0$, $h=1,2,...,L_1$

$U_k\geq 0$, $k=1,2,...,M_1-1$

and

$1\leq L_1\leq L$, $1\leq M_1\leq M$

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

Empirical Studies

I: If the variable is having right triangular distribution with pdf as

$$f\left(x\right)=\begin{cases}2\left(2-x\right) & ;\quad 0\leq x\leq 1\\ 0 & ;otherwise\end{cases} \tag{31}$$

and the other variable follows an exponential distribution as

$$f\left(z\right)=\begin{cases}\lambda e^{-\lambda z} & ;\quad z\geq 0,\lambda>0\\ 0 & ;otherwise\end{cases} \tag{32}$$

In order to obtain OSB when the auxiliary variables have as given above we need to estimate the values of W_{hk},σ^2_{hkx} and σ^2_{hky} .For estimating this use above pdf's in (19)-(21),we get

$$W_{hk}=V_h g_1 \tag{33}$$

$$\sigma^2_{hkx}=\frac{U_k g_1\left\{\frac{4}{3}\left(V_h^2+3x_{h-1}^2+3V_h x_{h-1}\right)-\frac{1}{2}\left[\left(V_h+2x_{h-1}\right)\left(V_h^2+2x_{h-1}^2+2V_h x_{h-1}\right)\right]\right\}-4U_k^2\left[\begin{array}{l}V_h\left(1-\frac{1}{3}V_h\right)\\ +x_{h-1}\left(2-x_{h-1}-3V_h\right)\end{array}\right]^2}{g_1^2} \tag{34}$$

$$\sigma_{hkz}^2=\frac{z_{k-1}^2-\left(U_k+z_{k-1}\right)^2e^{-\lambda U_k}+2\lambda\left[z_{k-1}+\left(U_k+z_{k-1}\right)e^{-\lambda U_k}\right]+2\left(1-e^{-\lambda U_k}\right)}{\left(1-e^{-\lambda U_k}\right)\left(4-V_h-2x_{h-1}\right)}-\frac{\left[\left(z_{k-1}+\frac{1}{\lambda}\right)\left(1-e^{-\lambda U_k}\right)-U_ke^{-\lambda U_k}\right]^2}{\left(1-e^{-\lambda U_k}\right)^2\left(4-V_h-2x_{h-1}\right)^2} \tag{35}$$

where

$$g_1=e^{-\lambda z_{k-1}}\left(1-e^{-\lambda U_k}\right)\left(4-V_h-2x_{h-1}\right)$$

Using values obtained in equations (33)-(35) in MPP (29),we have

Minimize

$$\sum_h\sum_k V_hg_1\left(\begin{array}{l}\beta^2\dfrac{U_kg_1\left\{\begin{array}{l}\frac{4}{3}\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)\\-\frac{1}{2}\left[\left(V_h+2x_{h-1}\right)\left(V_h^2+2x_{h-1}^2+2V_hx_{h-1}\right)\right]\end{array}\right\}-4U_k^2\left[\begin{array}{l}V_h\left(1-\frac{1}{3}V_h\right)\\+x_{h-1}\left(2-x_{h-1}-3V_h\right)\end{array}\right]^2}{g_1^2}\\+\gamma^2\left[\begin{array}{l}\dfrac{z_{k-1}^2-\left(U_k+z_{k-1}\right)^2e^{-\lambda U_k}}{+2\lambda\left[z_{k-1}+\left(U_k+z_{k-1}\right)e^{-\lambda U_k}\right]+2\left(1-e^{-\lambda U_k}\right)}\\\left(1-e^{-\lambda U_k}\right)\left(4-V_h-2x_{h-1}\right)\end{array}\right]-\dfrac{\left[\left(z_{k-1}+\frac{1}{\lambda}\right)\left(1-e^{-\lambda U_k}\right)-U_ke^{-\lambda U_k}\right]^2}{\left(1-e^{-\lambda U_k}\right)^2\left(4-V_h-2x_{h-1}\right)^2}\end{array}\right)\sqrt{c_{hk}}$$

Subject to

$$\begin{aligned}&\sum_h V_h=d_x\\&\sum_k U_k=t_z\\&\forall V_h\geq 0,U_k\geq 0\quad,\quad\begin{array}{l}h=1,2,...,L\\k=1,2,...,M\end{array}\end{aligned} \tag{36}$$

By executing a programme in R-software relating to generating random numbers for estimating the parameters we get β=1.52 and γ=0.42.

Now let us assume that the auxiliary variable X is defined in $x\in[0,1]$ and Z in $z\in[0,6]$ i,e.

$x_0=0,x_L=1,z_0=0,z_M=6$ and λ = 1.we can write MPP (36) as

Minimize

$$\sum_h\sum_k V_hg_1{}'\left(\begin{array}{l}(2.31)\dfrac{U_kg_1{}'\left\{\frac{4}{3}\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-\frac{1}{2}\left[\left(V_h+2x_{h-1}\right)\left(V_h^2+2x_{h-1}^2+2V_hx_{h-1}\right)\right]\right\}-4U_k^2\left[\begin{array}{l}V_h\left(1-\frac{1}{3}V_h\right)\\+x_{h-1}\left(2-x_{h-1}-3V_h\right)\end{array}\right]^2}{g_1^2}\\+(0.17)\dfrac{z_{k-1}^2-\left(U_k+z_{k-1}\right)^2e^{-U_k}+2\left[z_{k-1}+\left(U_k+z_{k-1}\right)e^{-U_k}\right]+2\left(1-e^{-U_k}\right)}{\left(1-e^{-U_k}\right)\left(4-V_h-2x_{h-1}\right)}-\dfrac{\left[\left(z_{k-1}+1\right)\left(1-e^{-U_k}\right)-U_ke^{-U_k}\right]^2}{\left(1-e^{-U_k}\right)^2\left(4-V_h-2x_{h-1}\right)^2}\end{array}\right)\sqrt{c_{hk}}$$

Subject to

$$\sum_h V_h = 1$$
$$\sum_k U_k = 6 \qquad (37)$$
$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{matrix} h = 1,2,...,L \\ k = 1,2,...,M \end{matrix}$$

Where $g_1' = e^{-z_{k-1}}\left(1-e^{-U_k}\right)\left(4-V_h-2x_{h-1}\right)$

while executing a computer programme of the MPP (37) for obtaining OSB of 6 (2×3) strata assuming the cost values as $c_{11}=2, c_{12}=3, c_{13}=4, c_{21}=5, c_{22}=6, c_{23}=7$,we get

Table 1: Displays the OSB when the auxiliary variables X and Z have pdf's as right triangular and exponential respectively

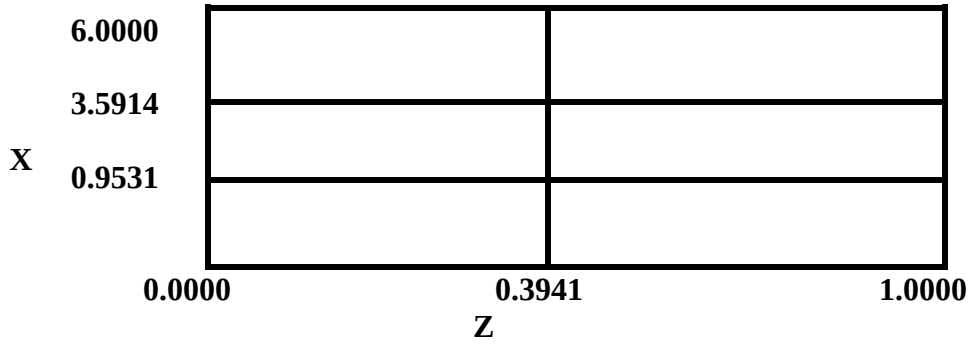

Table 2: Displays OSB and Variance of proposed method and others

OSB (x_h, z_k)	**Total Variance (Proposed method)**	**Total Variance (Fonolahi and Khan 2014)**	**% R.E.**
(0.3941,0.9531) (2.0000,0.9531) (0.3941,3.5914) (2.0000,3.5914) (0.3941,6.0000) (2.0000,6.0000)	0.03519	0.08348	237.226

Table 1 shows the stratification points when the auxiliary variables are having right triangular and exponential distribution. In order to obtain 6 strata in total from which two along the x-axis and 3 along the z-axis, the OSB obtained in case of optimum allocation are presented in Table 2 along with the variance obtained by using proposed method as well variance obtained by Fonolahi and Khan (2014).Thus it reveals that the variance obtained through the proposed method is lesser than the method proposed by Fonolahi and Khan (2014).The percentage of relative efficiency comes out to be 237.226. Hence the proposed method is preferable than the existing method.

II: Let us assume that the auxiliary variable X is having a uniform distribution with pdf as

$$f(x)=\begin{cases}\dfrac{1}{b-a} & ,a\le x\le b\\ o & ,otherwise\end{cases} \tag{38}$$

and assumed that the other variable Z is following standard normal distribution with pdf as

$$f(z)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} & ;-\infty<z<\infty\\ 0 & ;otherwise\end{cases} \tag{39}$$

To obtain OSB for the study variable which is having two auxiliary variables following uniform and standard normal distributions, we need to find the values of W_{hk},σ^2_{hkx} and σ^2_{hky} .For estimating this use above pdf's in (19)-(21),we get

$$W_{hk}=\frac{V_h}{2(b-a)}(E_1) \tag{40}$$

$$\sigma^2_{hkx}=\frac{2U_kE_1\left(V_h^2+3x_{h-1}^2+3V_hx_{h-1}\right)-12U_k^2\left(V_h+2x_{h-1}\right)^2}{3(E_1)^2} \tag{41}$$

$$\sigma^2_{hkz}=2\sqrt{2}(b-a)\begin{bmatrix} z_{k-1}\exp\left(-\dfrac{z_{k-1}^2}{2}\right)erf\left(\dfrac{U_k+z_{k-1}}{\sqrt{2}}\right)-(U_k+z_{k-1})(E_2)\\ -z_{k-1}\exp\left(-\dfrac{z_{k-1}^2}{2}\right)erf\left(\dfrac{z_{k-1}}{\sqrt{2}}\right)+(U_k+z_{k-1})\end{bmatrix}+\pi(E_1)^2 \tag{42}$$

where

$$E_1=erf\left(\frac{U_k+z_{k-1}}{\sqrt{2}}\right)-erf\left(\frac{z_{k-1}}{\sqrt{2}}\right)\ E_2=\exp\left(-\frac{(U_k+z_{k-1})^2}{2}\right)erf\left(\frac{U_k+z_{k-1}}{\sqrt{2}}\right),$$

$$E_3=\exp\left(-\frac{(U_k+z_{k-1})^2}{2}\right)erf\left(\frac{z_{k-1}}{\sqrt{2}}\right)$$

and 'erf ' is known as the error function and is given by

$$erf(z_k)-erf(z_{k-1})=\left(\frac{2}{\sqrt{\pi}}\right)\int_{z_{k-1}}^{z_k}e^{-u^2}\partial u$$

using values obtained in (40)- (42) in equation (36),we have

Minimize

$$\sum_h \sum_k \frac{V_h}{2(b-a)}(E_1)\left\{\begin{array}{l}\beta^2 \dfrac{2U_k E_1\left(V_h^2+3x_{h-1}^2+3V_h x_{h-1}\right)-12U_k^2\left(V_h+2x_{h-1}\right)^2}{3(E_1)^2} \\ +\gamma^2\sigma_{hkz}^2 = 2\sqrt{2}(b-a)\left[\begin{array}{l} z_{k-1}\exp\left(-\dfrac{z_{k-1}^2}{2}\right) erf\left(\dfrac{U_k+z_{k-1}}{\sqrt{2}}\right)-\left(U_k+z_{k-1}\right)E_2 \\ -z_{k-1}\exp\left(-\dfrac{z_{k-1}^2}{2}\right) erf\left(\dfrac{z_{k-1}}{\sqrt{2}}\right)+\left(U_k+z_{k-1}\right)E_3 \end{array}\right] \\ +\pi(E_1)^2 \end{array}\right\}\sqrt{c_{hk}}$$

Subject to

$$\begin{aligned} &\sum_h V_h = d_x \\ &\sum_k U_k = t_z \end{aligned} \qquad (43)$$

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

Let us assume that the interval of X is defined as $x \in [0,1]$ and the variable Z be truncated at 4 i,e $z \in [0,4]$ and by simulation in R-software estimates β=1.07 and γ=0.73.Thus ,we have (43) as

Minimize

$$\sum_h \sum_k \frac{V_h}{2}(E_1)\left\{\begin{array}{l}(2.28)\dfrac{U_k E_1\left(V_h^2+3x_{h-1}^2+3V_h x_{h-1}\right)-6U_k^2\left(V_h+2x_{h-1}\right)^2}{3(E_1)^2} \\ +(1.5)\left[\begin{array}{l} z_{k-1}\exp\left(-\dfrac{z_{k-1}^2}{2}\right) erf\left(\dfrac{U_k+z_{k-1}}{\sqrt{2}}\right)-\left(U_k+z_{k-1}\right)E_2 \\ -z_{k-1}\exp\left(-\dfrac{z_{k-1}^2}{2}\right) erf\left(\dfrac{z_{k-1}}{\sqrt{2}}\right)+\left(U_k+z_{k-1}\right)E_3 \end{array}\right] \\ +\pi(E_1)^2 \end{array}\right\}\sqrt{c_{hk}}$$

Subject to

$$\begin{aligned} &\sum_h V_h = 1 \\ &\sum_k U_k = 4 \end{aligned} \qquad (44)$$

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

For obtaining OSB we assume that the values of $c_{11}=2, c_{12}=3, c_{13}=4, c_{21}=5, c_{22}=6, c_{23}=7$

for total 6 (2×3) strata, execute a computer programme in LINGO by assuming all these conditions given above we have

Table 3: Displays OSB and Variance of proposed method and others when having uniform and standard normal distribution

OSB (x_h, z_k)	Total Variance (Proposed method)	Total Variance (Danish *et al.* 2017)	% R.E.
(0.50000,1.2539) (2.0000,1.2539) (0.5000,2.9827) (2.0000,2.9827) (0.5000,4.0000) (2.0000,4.0000)	0.0008473	0.003586	423.257

Table 3 presents the OSB and variances when the auxiliary variables X and Z follow uniform and standard normal distributions respectively. The percentage of relative efficiency between proposed method and Danish *et al*. (2017) method comes out to be 423.257.Thus it reveals that the variance obtained through the proposed method is lesser than the method proposed by Danish *et al*.(2017) .Hence the proposed method is preferable than the existing method.

Conclusion

In this investigation I have developed a method for construction of optimum strata boundaries under optimum allocation when we are having single study variable consists of two auxiliary variables. The problem happens to be multistage problem which was solved by dynamic programming by executing the programme in LINGO. It is found that the construction of strata using auxiliary variable of the populations having above mentioned distribution functions, leads to substantial gains in the precision of the estimates while using the proposed technique. Empirical studies showed that the proposed method is more precise than the methods developed by Fanolahi and Khan (2014) and Danish *et al.* (2017) that concludes the proposed method more preferable.

Acknowledgment

I am highly thankful to the reviewers for suggesting comments that improves the quality of the paper. Further, the incredible efforts made by the Journal team especially Dr.Rehan Ahmad Khan and Kanwal Saleem towards the paper, deserves to be acknowledged.

References

1. Dalenius, T. (1950). The problem of optimum stratification-II. Skand. Aktuartidskr, 33, 203-213.
2. Dalenius, T. and Gurney, M. (1951). The problem of optimum stratification-II, Skand. Aktuartidskr, 34: 133-148.

3. 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(3), 123-129
4. Danish, F., Rizvi, S.E.H. Jeelani, M. I and Reashi J.A. (2017).Obtaining Strata Boundaries under Proportional Allocation with Varying Cost of Every Unit. Pak.j.stat.oper.res. 13 (3): 567-574
5. 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).2580. DOI: 10.22237/jmasm/1529418671.
6. Danish, F. (2018): A Mathematical Programming approach for obtaining optimum strata boundaries using two auxiliary variables under proportional allocation. Transition in Statistics. 19(3).507–526, DOI 10.21307/stattrans-2018-028.
7. Danish, F. and Rizvi, S.E.H. (2019): Optimum Stratification by two Stratifying Variables using Mathematical Programming. Pakistan Journal Of Statistics. 35(1),11-24.
8. Fanolahi, A.V. and Khan, M.G.M. (2014). Determing the optimum strata boundaries with constant cost factor. Conference: IEEE Asia Pacific World Congress on Computer Science and Engineering (APWC), At Plantation Island, Fiji.
9. Gupta, R. K., Singh, R. and Mahajan, P. K. (2005). Approximate optimum strata boundaries for ratio and regression estimators. Aligarh J. Statist., 25, 49-55.
10. Khan, M.G.M. , Khan, E.A. and Ahsan, M.J. 2003. An optimal multivariate stratified sampling design using dynamic programming. Aust. N. Z. J. Stat. 45(1): 107–113.
11. Khan, M. G. M. , Rao, D., 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.
12. Rao, D., Khan, M.G.M. and Khan, S. (2012).Mathematical programming on multivariate calibration estimation in stratified sampling. World academy of science, engineering and technology.6:12-27.
13. Rizvi, S. E. H., Gupta, J. P. and Bhargava, M. (2002). Optimum stratification based on auxiliary variable for compromise allocation. Metron, 28(1): 201-215.
14. Singh, R. (1971). Approximately optimum stratification on the auxiliary variable. J. Amer. Statist. Assoc., 66, 829-833.
15. Singh, R. (1975). An alternate method of stratification on the auxiliary variable. Sankhya, 37: 100-108.
16. Singh, R. and Sukhatme, B. V. (1969). Optimum stratification for equal allocation. Ann. Inst. Statist. Math., 27: 273-280.