REVISTA INVESTIGACION OPERACIONAL VOL. 45, NO. 3, 354-368, 2024

APPROXIMATE OPTIMUM STRATA BOUNDARIES FOR NEYMAN ALLOCATION UNDER RANKED SET SAMPLING

Khalid Ul Islam Rather*, S.E.H Rizvi*, Manish Sharma*, M. Iqbal Jeelani** Sira Allende***
and Carlos N. Bouza***
*Division of Statistics and Computer Science, Main Campus SKUAST-J, Jammu, J&K, India, 180009.
**Division of Social and Basic Sciences Faculty *of Forestry, SKUAST-Kashmir, India*
****F*acultad de Matematica y Computacion. Universidad de La Habana La Habana, Cuba.

ABSTRACT
Using convenient stratification criteria such as geographical regions or other natural conditions like age, gender, etc., is not beneficial in order to maximize the precision of the estimates of variables of interest. Thus, one has to look for an efficient stratification design to divide the whole population into homogeneous strata that achieves higher precision in the estimation. In this paper the problem of optimum stratification on an auxiliary variable $'X'$ for Neyman allocation under ranked set sampling has been considered, when the form of the regression of the estimation variable $'Y'$ on the auxiliary variable $'X'$ given the variance function $V(y \mid x)$ are known. A $\text{cum}\sqrt[3]{K_2(x)}$ rule of finding approximately optimum strata boundaries has been proposed. Further, empirical study has been made and presented along with relative efficiency.

KEYWORDS: Ranked set Sampling, approximately optimum strata boundaries, auxiliary variable, optimum strata width.

MSC: 62D05

RESUMEN
No es beneficioso usar criterios de estratificación convenientes, como regiones geográficas u otras condiciones naturales como edad, género, etc., para maximizar la precisión de las estimaciones de interés. Por lo tanto, se debe buscar un diseño de estratificación eficiente para dividir a toda la población en estratos homogéneos, que logre una mayor precisión en la estimación. En este trabajo se ha abordado el problema de la estratificación óptima usando una variable auxiliar para la asignación de Neyman bajo un muestreo estratificado, considerando, que se conoce la forma de la regresión entre la variable de estimación y la variable auxiliar, dada la función de varianza. Se ha propuesto una regla para encontrar límites de estratos aproximadamente óptimos. Además, se ha realizado y presentado un estudio empírico junto con una eficiencia relativa. Se hace una comparación para bajo varias distribuciones. También se considera el uso de Programación dinámica.

PALABRAS CLAVE: Muestreo conjuntos clasificados, aproximación de límites de estratos óptimos, variable auxiliar, amplitud de estratos óptimos.

1. INTRODUCTION

When planning a survey sampling samplers focus primarily on the methods of sample selection and in teh reduction of sampling errors. The precision of an estimator of a population parameter depends on the heterogeneity of the units as well as in fixing the sample size and sampling fraction. Stratified sampling method plays a significant role for enhancing the precision of the estimator. For achieving a greater precision of the estimate to decrease the heterogeneity of the population units stratifying is a popular approach. Commonly, stratification is made considering administrative grouping, geographic regions etc., as well as on the basis of auxiliary characters. The main objective of stratification is to dealt with enhancing precision of an estimator and the construction of strata, the number of strata, the allocation of sample size to strata and stratification variable(s) should be considered when possible for increasing efficiencies.
Recent contributions on the detection of Optimal Boundaries of Strata (OBS) are Gupta and Ahamed (2021) who used a model-based allocation under a super population model.
McIntyre (1952) introduced the concept of RSS. Rather *et al.* (2021, 2022, 2023) considered situation of optimum stratification for different allocation under RSS. Samawi (1996) introduced concept of Stratified Ranked Set Sample (SRSS) We can think of a SRSS scheme as a collection of L separate Ranked Set Samples.

Let the population under consideration be divided into L strata and a sample $n_{0h} = (R_h \times n_h)$ units is selected from h^{th} stratum is drawn using RSS, where R_h is the number of cycles and n_h is sample size of each cycle. Each sample element is measured with respect to some variable Y, and estimator of the population mean is given by

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

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

If the finite correction is ignored, the variance of the estimate will be

$$\frac{1}{n}\left[\sum_{h=1}^{L} W_h \sqrt{\sigma^2_{h(r)} + \mu_{h\eta}} \right]^2$$

It is equivalent to the minimization of the expression

$$\left[\sum_{h=1}^{L} W_h \sqrt{\sigma^2_{h(r)} + \mu_{h\eta}} \right] \tag{1.2}$$

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

denotes the variance of r^{th} order statistics in h^{th} stratum of the random sample of size n_h.

For a given method of allocation, the variance is clearly a function of the strata boundaries. In most of these investigations related to optimum stratification, both the estimation and stratification variables are taken to be the same. Since the distribution of the estimation variable $'Y'$ is rarely known in practice, it is desirable to stratify on the basis of some suitably chosen concomitant variable $'X'$. An investigation in this direction has been made by Taga (1967) who has considered the general problem of optimum stratification based on auxiliary variables for the case of proportional allocation. We consider the problem of optimum stratification on the auxiliary variable $'X'$, assuming knowledge about the form of the regression of $'Y'$ on $'X'$ and the variance function $V(y \mid x)$, minimal equations giving optimum strata boundaries have been obtained for proportional allocations under ranked set sampling. Since these equations cannot be solved easily, various methods of finding approximations to the exact solutions have been given.

In this paper, the problem of construction of strata boundaries will be dealt using classical approach when the sample is selected from the strata using RSS.

We consider the fact that different stratification problems have been modeled using Mathematical Programming (MP) tools, as Allende-Alonso and Bouza-Herrera (1987) who derived optimization criteria for multivariate strata construction, .Bouza, et al. (2013, 2018). Brito et al. (2022) prepondering heuristic methods, . The OBS has been formulated as a Nonlinear Programming Problem (NLPP) in different papers. Lone *et al*.(2017) used a Branch and Bound Method foe solving it. The problem of determining the OSB can be also considered as the problem of determining Optimum Strata Width (OSW). Popularly MPP is solved by using a convenient tool of the Dynamic Programming (DP) or GD toolbox. We develop a MPP model for determining OBSS under the use of stratified RSS.

We will illustrate its behavior when dealing with RSS considering that the distribution of the auxiliary variable. the empirical studies developed sustain that increasing the number of strata increase the precision.

2. MINIMAL EQUATIONS UNDER NEYMAN ALLOCATION

Optimum stratification looks for constructing strata in such a way that is obtained the minimum variance of the estimator. The main objective of this set of techniques stratification determining a better cross-section of the population gaining in relative precision with respect to Simple Random Sampling (SRS) . The problem of determining optimum strata boundaries (OSB) is the result presented in h seminal paper of Dalenius (1950). Using MPP fo OSB is gaining in attention. See for example the series of papers of Khan et al. (2009)

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

$$y = c(x) + e \tag{2.1}$$

Where $'c(x)'$ is a function of auxiliary variable, $'e'$ is the error term such that $E(e \mid x) = 0$ and $V(e \mid x) = \eta(x) > 0\ \forall x \in (a,b)$ with $(b-a) < \infty$. Let $f(x,y)$ and $f(x)$ be the joint density function and marginal density function of (x,y) and x respectively. Then, we have

$$w_h = \int_{x_{h-1}}^{x_h} f_i(x) \ , \ \mu_{hc} = \frac{1}{w_h}\int_{x_{h-1}}^{x_h} c(x) f_i(x) \quad \text{and} \quad \sigma^2{}_{hy} = \sigma^2{}_{hc} + \mu_{h\eta}\ , \ \left(h = 1,2,3\ldots,L\right) \tag{2.2}$$

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

$$V\left(\bar{y}_{SRSS}\right)_N = \left[\sum_{h=1}^{L} W_h \sqrt{\sigma^2_{h(r)} + \mu_{h\eta}}\right] \tag{2.3}$$

Let $[x_h]$ denote the set of optimum points of stratification on the range (a,b), for which the $V\left(\bar{y}_{SRSS}\right)$ is minimum. These points $[x_h]$ are the solutions of the minimal equations which are obtained by equating to zero the partial derivatives of $V\left(\bar{y}_{SRSS}\right)$ with respect to $[x_h]$. We shall now obtain these minimal equations for proportional allocations. The minimization of this variance is equivalent to the minimization of the expression

$$\left[\sum_{h=1}^{L}\sum_{i=1}^{n_h} W_h \sqrt{\sigma^2_{hc(r)} + \mu_{h\eta}}\right].$$

On equating to zero the partial derivative of this expression with respect to $[x_h]$, we get

$$\frac{\left(\left(c(x_h) - \mu_{h(r)}\right)^2 - \sigma^2_{h(r)} + \eta(x_h) - \mu_{h\eta}\right)}{\sqrt{\sigma^2_{h(r)} + \mu_{h\eta}}} = \frac{\left(\left(c(x_h) - \mu_{i(r)}\right)^2 - \sigma^2_{i(r)} + \eta(x_h) - \mu_{i\eta}\right)}{\sqrt{\sigma^2_{i(r)} + \mu_{h\eta}}}\ i = h+1, h = 1,2,\ldots,L-1 \tag{2.4}$$

These equations are implicit functions of the strata boundaries $[x_h]$ and their exact solutions are somewhat difficult to find. Therefore, we proceed to find the method of solving these minimal equations by conducting approximations. However, we shall first obtain certain approximate expressions for the conditional mean and variance which will be necessary to obtain approximate solutions.

3. APPROXIMATE EXPRESSIONS FOR CONDITIONAL MEAN AND VARIANCE

Let the functions $f_i(x)$, $c(x)$ and $\eta(x)$ are bounded away from zero and possess first two derivatives continuous $\forall x \in (a,b)$. Then, we have the following identities due to Ekman (1959).

$$I_i(y,x) = \int_Y^X (t-y)^i f_i(t)dt = \sum_{j=0}^{3} \frac{(k)^{i+j+1}}{j!(i+j+1)} f^{(j)} + O(k^{i+5}) \tag{3.1}$$

where $f^{(j)}$ is the j^{th} derivative of $f_i(t)$ at $t = y$ and $k = x - y$

$$I_i(y,x) = \int_Y^X (t-y)^i f_i(t)dt = \sum_{j=0}^{3} \frac{(-k)^{i+j+1}}{j!(i+j+1)} f^{(j)} + O(k^{i+5}) \tag{3.2}$$

$O(k^i)$ is the higher order terms with power $\geq i$

Let $\mu_\eta(y,x)$ denote the conditional expectation of function $\eta(t)$ in the interval (y,x), so that

$$\mu_\eta(y,x)=\frac{\int_y^x \eta(t) f_i(t)dt}{\int_y^x f_i(t)dt} \tag{3.3}$$

we have from the definition of $\mu_\eta(y,x)$

$$\mu_\eta(y,x)\int_y^x f_i(t)dt=\int_y^x \eta(t) f_i(t)dt)$$

therefore, we have

$$\mu_\eta(y,x).I_0(y,x)=\left[\eta I_0(y,x)+\sum_{i=1}^{3}\eta^{(i)}I_i(y,x)/i!\right]+O(k^5)$$

Using the Taylor Series expansions for $I_i(y,x)$ and $I_i(x,y)$ from (3.2) and simplifying the result at point $t=y$, we have

$$\mu_\eta(y,x)=\eta\left[1+\frac{\eta'}{2\eta}k+\frac{(\eta' f'+2f\eta'')}{12f\eta}k^2+\frac{(ff''\eta'+ff'\eta''+f^2\eta'''-\eta' f'^2)}{24f^2\eta}k^3+O(k^4)\right] \tag{3.4}$$

Proceeding in the same fashion using Taylor series expansions about the point 'x', the expression for $\mu_\eta(y,x)$ is obtained as

$$\mu_\eta(y,x)=\eta\left[1-\frac{\eta'}{2\eta}k+\frac{(\eta' f'+2f\eta'')}{12f\eta}k^2-\frac{(ff''\eta'+ff'\eta''+f^2\eta'''-\eta' f'^2)}{24f^2\eta}k^3+O(k^4)\right] \tag{3.5}$$

Let $\sigma^2{}_\eta(y,x)$ denotes the conditional variance of the function $\eta(t)$ in the interval (y,x), we have

$$\sigma^2{}_\eta(y,x)=\mu_{\eta^2}(y,x)-(\mu_\eta(y,x))^2$$

substituting values, we get

$$\sigma^2{}_\eta(y,x)=\frac{k^2(\eta'(y))^2}{12}\left[1+(\eta''(y)/\eta'(y))k+O(k^2)\right] \tag{3.6}$$

Using the above results, several other approximations can be obtained. Multiplying the series expansions for $\mu_\eta(y,x)$ about the points $t=y$, $t=x$ and taking the square root, we obtain

$$\mu_\eta(y,x)=\sqrt{\eta(y)\eta(x)}\left[1+O(k^2)\right] \tag{3.7}$$

From (3.3), we have

$$\mu_\eta(y,x).I_0(y,x)=\int_y^x \eta(t) f_i(t)dt$$

Taking $\eta(t)=t^2$ and using (3.7), we get

$$\int_y^x f_i(t)dt=\frac{1}{xy}\int_y^x t^2 f_i(t)dt\left[1+O(k^2)\right] \tag{3.8}$$

Similarly expanding $\sqrt[2]{f_i(t)}$ about the point t=y, we have

$$\left[\int_Y^X \sqrt[\lambda]{f_i(t)}dt\right]^{\lambda} = k^{\lambda-1}\int_Y^X f_i(t)dt\left[1+O(k^2)\right] \tag{3.9}$$

4. APPROXIMATE SOLUTIONS OF THE MINIMAL EQUATIONS

To find approximate solutions to the minimal equation (2.4), we shall obtain the series expansions of system of equations about the point $[x_h]$, the common boundary of h^{th} and $(h+1)^{th}$ strata. The expansions for the two sides of the equation (2.4) are obtained by using various results proved in the preceding section. For the expansion of the right hand side about the point x_h, (y,x) is replaced by (x_{h-1},x_h) while for the left hand side we replace (y,x) by (x_{h-1},x_h).

We consider the left-hand side of (2.4), we have

$$[\mu_{ic}-c(x_h)]^2 = \left(\frac{k_i^{\,2}}{4}\right)\left[c'^2 + \frac{\left(c'^2 f' + 2fc'c''\right)}{3f}k_i + \frac{\left(6ff''c'^2+10ff'c'c''+6f^2c'c'''-5c'^2f'^2+4f^2c''^2\right)}{36f^2}k_i^{\,2} + O(k_i^{\,3})\right]$$

and

$$\eta(x_h)+\mu_{h\eta} = \eta\left[2+\frac{\eta'}{2\eta}k_i + \frac{\left(\eta' f' + 2f\eta''\right)}{12f\eta}k_i^{\,2} + \frac{\left(ff''\eta'+ff'\eta''+f^2\eta'''-\eta' f'^2\right)}{24f^2\eta}k_i^{\,3} + O(k_i^{\,4})\right]$$

From the above results and using (3.6), we obtain

$$\left[\mu_{h(r)}-c(x_h)\right]^2+\sigma^2_{h(r)}+\eta(x_h)+\mu_{h\eta} = 2\eta\left[1+\frac{\eta'}{4\eta}k_i+\frac{\left(4fc'^2+\eta' f'+2f\eta''\right)}{24f\eta}k_i^{\,2} + \frac{\left(2ff'c'^2+6f^2c'c''+ff''\eta'+ff'\eta''+f^2\eta'''-\eta' f'^2\right)}{48f^2\eta}k_i^{\,3}+O(k_i^{\,4})\right] \tag{4.1}$$

Also, we have

$$\mu_{h\eta}+\sigma^2_{h(r)} = \eta\left[1+\frac{\eta'}{2\eta}k_i+\left(\frac{\eta' f'+2f\eta''}{12f\eta}+\frac{c'^2}{12\eta}\right)k_i^{\,2}+\left(\frac{ff''\eta'+ff'\eta''+f^2\eta'''-\eta' f'^2}{24f^2\eta}+\frac{fc'c''}{12f\eta}\right)k_i^{\,3} + O(k_i^{\,4})\right] \tag{4.2}$$

Using the above relations, we obtain on simplification

$$\frac{\left[\mu_{hc}-c(x_h)\right]^2+\sigma^2_{h(r)}+\eta(x_h)+\mu_{h\eta}}{\sqrt{\left(\mu_{h\eta}+\sigma^2{}_{h(r)}\right)}} = 2\sqrt{\eta}\left[1+(B_2)k_h^{\,2}+(B_3)k_h^{\,3}+O(k_h^{\,4})\right] \tag{4.3}$$

Where

$$(B_2)=\left(\frac{4\eta c'^2+\eta'^2}{32\eta^2}\right),\quad (B_3)=\left(\frac{8f'c'^2\eta^2+16fc'c''\eta^2+2\eta f'\eta'^2+4f\eta\eta''\eta'-4f\eta\eta'c'^2-3f\eta'^3}{192f\eta^3}\right)=\frac{1}{96f\sqrt{\eta}}\frac{d}{dx_h}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)$$

Similarly from right hand side of (2.4), we obtain

$$\frac{\left[\mu_{ic}-c(x_h)\right]^2+\sigma^2_{i(r)}+\eta(x_h)+\mu_{i\eta}}{\sqrt{\left(\mu_{i\eta}+\sigma^2{}_{i(r)}\right)}} = 2\sqrt{\eta}\left[1+(B_2)k_i^{\,2}+(B_3)k_i^{\,3}+O(k_i^{\,4})\right] \tag{4.4}$$

Therefore the right hand side of the minimal equation (2.4) can be put as

$$\frac{[\mu_{ic}-c(x_h)]^2+\sigma_{ic}^2+\eta(x_h)+\mu_{i\eta_{ic}}}{\sqrt{(\mu_{i\eta}+\sigma^2{}_{ic})}}=2\sqrt{\eta}\left[1+\left(\frac{4\eta c'^2+\eta'^2}{32\eta^2}\right)k_i{}^2+\frac{k_i{}^3}{96f\sqrt{\eta}}\frac{d}{dx_h}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+O(k_i{}^4)\right]$$

Similarly the expansion of the left hand side of the equation (2.4) can be obtained. The expansion is given by

$$\frac{[\mu_{hc}-c(x_h)]^2+\sigma_{h(r)}^2+\eta(x_h)+\mu_{h\eta}}{\sqrt{(\mu h_\eta+\sigma^2{}_{h(r)})}}=2\sqrt{\eta}\left[1+\left(\frac{4\eta c'^2+\eta'^2}{32\eta^2}\right)k_h{}^2+\frac{k_h{}^3}{96f\sqrt{\eta}}\frac{d}{dx_h}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+O(k_h{}^4)\right]$$

Whereas before the functions f , η , c and derivatives are again evaluated at x_h .

The equations (2.7), after cancelling $2\sqrt{\eta}$ on both sides and multiplying the two sides by $f(x_h)$, can be put as

$$=\frac{k_h{}^2}{16}\left[\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+\frac{k_h}{3}\frac{d}{dx_h}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+O(k_h{}^4)\right]$$

$$=\frac{k_i{}^2}{16}\left[\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+\frac{k_i}{3}\frac{d}{dx_h}\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)+O(k_i{}^4)\right] \quad (4.5)$$

As can be easily seen, this equation can be written as

$$\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)^{\frac{1}{3}}\left[k_h^2\int_{x_{h-1}}^{x_h}g_1(t)f_i(t)dt.[1+O(k_h^2)]\right]^{\frac{2}{3}}=\left(\frac{4f\eta c'^2+f\eta'^2}{\eta^{\frac{3}{2}}}\right)^{\frac{1}{3}}\left[k_i^2\int_{x_h}^{x_{h+1}}g_1(t)f_i(t)dt.[1+O(k_i^2)]\right]^{\frac{2}{3}} \quad (4.6)$$

Where $i=h+1, h=1,2,...,L$

$$g_1(t)=\frac{\eta'^2(t)+4\eta(t)c'^2(t)}{[\eta(t)]^{\frac{3}{2}}} \quad (4.7)$$

In case it is possible to find a function $Q_1(x_{h-1},x_h)$ such that

$$k_h^2\int_{x_{h-1}}^{x_h}g_1(t)f_i(t)dt = \text{constant} \quad (4.8)$$

$$=Q_1(x_{h-1},x_h)[1+O(k_i^2)] \quad (4.9)$$

the above results can be put in the form of theorem as follows.

Theorem : If the regression of the estimation variable $'Y'$ on the stratification variable $'X'$, in the infinite super population is given by $y=c(x)+e$, where $'c(x)'$ is a function of auxiliary variable, $'e'$ is the error term such that $E(e\mid x)=0$ and $V(e\mid x)=\eta(x)>0\ \forall x\in(a,b)$ with $(b-a)<\infty$, and further if the function $g_1(x)\ f_i(x)\ \in\Omega$: then the system of equations (2.4) given strata boundaries (x_h) which correspond to the minimum of $V(\bar{Y}_{st})_N$ can be written as

$$\left[k_h^2\int_{x_{h-1}}^{x_h}g_1(t)f_i(t)dt.[1+O(k_h^2)]\right]^{\frac{2}{3}}=\left[k_i^2\int_{x_h}^{x_{h+1}}g_1(t)f_i(t)dt.[1+O(k_i^2)]\right]^{\frac{2}{3}}$$

Neglecting the terms of order $O\left(Sup_{(a,b)}\left(k_h\right)\right)^3$ can be neglected; these equations can be replaced by the approximate system of equations

$$k_h^2 \int_{x_{h-1}}^{x_h} g_1(t) f_i(t) dt = \text{constant}$$

Or equivalently by

$$Q_1\left(x_{h-1}, x_h\right) = \text{constant} \quad , \quad k_h = \left(x_h - x_{h-1}\right)$$

$$Q_1\left(x_{h-1}, x_h\right)\left[1 + O(k_i^2)\right] = k_h^2 \int_{x_{h-1}}^{x_h} g_1(t) f_i(t) dt \quad , \ i = h+1, h = 1,2,\ldots, L$$

The similar results can also be obtained by minimizing the function

$$\sum_{h=1}^{L} k_h^2 \int_{x_{h-1}}^{x_h} g_1(t) f_i(t) dt$$

Thus we find that if the function $g_1(t) f_i(x)$ belongs to Ω, the minimum value of

$$\left[\sum_{h=1}^{L} W_h \sqrt{\sigma_{h(r)}^2 + \mu_{h\eta}}\right]$$

and therefore $\left\{nV(\overline{Y}_{STRSS})\right\}_N$, exists and the solutions of the system of equations (2.4) or equivalently of (4.7). These equations as such are very difficult to solve and therefore it is essential to find some way out of this difficulty. It is done by replacing these systems of equations by other systems of equations which are comparatively easier to solve but are only asymptotically equivalent to the exact minimal equations. The error factor is introduced because we neglect the terms of higher powers of strata widths which is of course justifiable if the number of strata is large. We have obtained these systems of equations after neglecting the terms of order $O\left(Sup_{(a,b)}\left(k_h\right)\right)^4 = O\left(m^4\right)$ where $m = Sup_{(a,b)}\left(k_h\right)$, on both sides of the equation (4.7). If the number of strata is large and therefore terms of order $O\left(m^4\right)$ are quite small, the error involved in the approximate systems of equations is expected to be quite small and the set of points $\left(x_h\right)$ obtained from them shall be quite near the optimum values.

Now we proceed to develop the approximate systems of equations given in (4.8) and (4.9). Here, in finding various forms of the function $Q_1\left(x_{h-1}, x_h\right)$, we shall keep in mind that the function $Q_1\left(x_{h-1}, x_h\right)$ is such that

$$k_h^2 \int_{x_{h-1}}^{x_h} g_1(t) f_i(t) dt = Q_1\left(x_{h-1}, x_h\right)\left[1 + O(k_i^2)\right]$$

If in (2.4) we retain only the first term on both sides of the equation and neglect the others, the two sides are equalized if

$$k_h = \text{constant}, \ \left(\frac{b-a}{L}\right)h \quad , \ \text{for} \quad h = 1,2,\ldots, L \qquad (4.10)$$

and therefore $x_h = a + \left(\frac{b-a}{L}\right)h$, with $\left(x_0 = a\right)$ and $\left(x_L = b\right)$

This set of solutions cannot be expected to yield very good results as we have neglected terms of order $O\left(m^3\right)$ on both sides of the exact minimal equations. This solution holds for all $g_1(t) f_i(x)$ provided they belong to Ω and all density functions with finite range. Due to its universality of application it can be recommended in case of less information about $g_1(t)$ and $f_i(x)$. Apart from this, it gives the strata boundaries at once without any difficulty that may arise even in solving the approximate systems of equations. This approximate method fails if

the range of $'x'$ is infinite, but one can resort to truncation of the density function to any suitable probability level before using this approximation.
We obtain next approximate systems of equations; the optimum points of stratification are such that

$$k_h^2 \int_{x_{h-1}}^{x_h} g_1(t) f_i(t) dt = c_1 , \quad h = 1,2,...,L \tag{4.11}$$

The solutions obtained from this approximation are expected to be quite close to the optimum points as only terms of $O(m^4)$ have been neglected. All the approximate systems that will now follow also give the points of stratification to the same degree of accuracy.
From (3.9) and equation (4.11) is obtained the following class of approximate equations. The approximations to optimum (x_h) are obtained from

$$\left[k_h^{(3\lambda-1)} \int_{x_{h-1}}^{x_h} g_1(t) f_i(t) dt \right]^{\frac{1}{\lambda}} = \text{constant}, \quad h = 1,2,...,L$$

For λ = 1/2 and 1/3 we have

$$\left[\sqrt{k_h^2 \int_{x_{h-1}}^{x_h} g_1(t) f_i(t) dt} \right] = c_2 = \text{constant}, \quad h = 1,2,...,L \tag{4.12}$$

For λ =1/3, we have the system of equation as

$$\left[\sqrt[3]{k_h^2 \int_{x_{h-1}}^{x_h} g_1(t) f_i(t) dt} \right] = c_3 , \quad h = 1,2,...,L \tag{4.13}$$

In all these systems of equations, $c's$ are constants to be determined and in some cases, the few equations may be meaningless such as for $h = 1$, i,e $x_{h-1} = 0$.

If the $K_2(x) = g_1(t) f_i(t) dt$ is bounded and its first two derivative exists $\forall x \in [a,b]$, then for given value of L taking equal intervals on the $cum\sqrt[3]{K_2(x)}$ rule will give AOSB.

5. A MATHEMATICAL PROGRAMMING APPROACH

The problem of determining the optimum strata boundaries, has been formulated in the framework of mathematical programming problem (MPP). It is well-known that with some mild modifications the corresponding MPP may be converted into a multistage decision problem and dynamic programming techniques provide solution to it. This approach is being used in applications, see Mostafa et al. (2008). Computing the solutions , using DP models for the optimization problems, are solved using R or Python tools. See Reddy and Khan [12] who implemented the solutions of the MPP via a R package. We used as a basis Package 'dynprog' and Scientific Python, Using SciPy for Optimization and Hands-On Linear Programming: Optimization With Python.
Optimum stratification is the method of choosing the best boundaries that make strata internally homogeneous, given some sample allocation. In order to make the strata internally homogenous, the strata should be constructed in such a way that the strata variances for the characteristic under study be as small as possible. This could be achieved effectively by having the distribution of the main study variable known and create strata by cutting the range of the distribution at suitable points. If the frequency distribution of the study variable is unknown, it may be approximated from the past experience or some prior knowledge obtained at a recent study.
Solving the OSB problem is equivalent to determining Optimum Strata Widths (OSW). The MPP under Neyman allocation uses as constraint that that sum of the widths of the strata equals the total range of the distribution. Within this framework we consider

$$d_h = y_h - y_{h-1}$$

as strata widths and determining the strata boundaries is looking for dividing the range $y_L - y_0 = d$ determining adequate intermediate points $y_1 \le y_2 \le \cdots \le y_{L-1}$. Let us assume that the probability density function, fdp, f(y) is integrable by parts, being a piece-wise continuous linear or non-linear functions

$$f(y) = \begin{cases} \gamma_1(y)\ if\ y \in [a_0, a_1] \\ \gamma_2(y)\ if\ y \in]a_1, a_2] \\ \vdots \\ \gamma_n(y)\ if\ y \in]a_{n-1}, a_n] \end{cases}$$

Note that $\min(y) = y_0 = a_0$; $\max(y) = y_L = a_n$. Considering that the i-th fdp determines L_i strata and $d_h = y_h - y_{h-1}$ $then\ L = \sum_{i=1}^{n} L_i\ and\ d = \sum_{h=1}^{L} d_h$

Each stratification point may be represented by

$$y_k = y_{k-1} + d_h$$

Take

$$\phi_h(y_h, y_{h-1}) = w_h Var(\bar{y}_{h(r)})$$

The optimization problem to be solved by using DP is

$$minimize \left\{ \sum_{h=1}^{L} \phi_h(d_h, y_{h-1}) | \sum_{h=1}^{L} d_h = d;\ \forall h = 1,..,L;\ d_h \ge 0 \right\}$$

As y_0 is known $\phi_1(d_1, y_0)$ is only function of d_1.

Hence, the MPP to be solved is a multistage decision one and, at each stage the value of the OSW , and so of the OSB, for a stratum is calculated using a Dynamic Programming technique with a forward recursive equation. In the numerical studies will be considered continuous fdp`s.

6. EMPIRICAL STUDY:

Comparisons of methods of obtaining approximate optimum strata boundaries (AOSB) have been developed for various allocation procedure viz. optimum with unstratified RSS has been done empirically. For this purpose the definite integrals involved in calculations have been solved using Mathematica software in order to obtain approximate optimum boundaries. Thereafter calculation of variance and ultimately percent relative efficiencies have been obtained for various values of number of strata i.e, L=2,3,4,5,6 through softwares.

With the purpose of illustrating the usefulness of the approximate solutions to the minimal equations giving optimum points of stratification an empirical study was developed. The effectiveness of the methods of finding approximation to the optimum points of stratification was evaluated. We have considered the system of minimal equations, obtained for the case of optimum, and the MPP approach based on DP. In this illustration we shall consider approximation intervals and the system of approximations given in approximate system of equations. We used the following distributions viz, uniform, right triangular, exponential and standard normal. For the proposed methods the Relative Efficiency in percent (%RE) of them were computed.

For clarity, the linear regression line of $'Y'$ on $'X'$ of the form $y = \alpha + \beta x + e$, assuming the value of $\beta = 0.5$. For the conditional variance function, it is assumed to have two different forms like the first form could be a constant and the second could be function of auxiliary variable. i,e. $\eta(x) = \delta$ and $\eta(x) = \lambda x$, where δ and λ are constants.

For the empirical studies under Neyman allocation let us assume small values of $\delta = 0.0214$, $\lambda = 0.00437$, such that there may be very small effect of these constants over the estimation.

If the stratification variable follows the uniform distribution with pdf $f(x) = \dfrac{1}{b-a}$, $x \in [1,2]$, utilizing the $cum\sqrt[3]{K_2(x)}$ rule, we get the stratification points as given in Table 1.

Table 1: AOSB and Variance when the auxiliary variable is uniformly distributed

$\eta(x) = \delta$							
L	**AOSB**	$n(V(\bar{y}_{SRSS})_{opt})$	%R.E	$n(V(\bar{y}_{SRSS})_{MPP-R})$	%R.E	$n(V(\bar{y}_{SRSS})_{MPP-Ph})$	%R.E
2	1.5000	0.74968	102.821	0.67838	113.668	0.67848	113.521

3	1.3299, 1.6599	0.74607	103.319	0.64129	119.619	0.64329	119.263
4	1.2499, 1.4999, 1.7500	0.74470	103.509	0.63893	120.606	0.638903	124.380
5	1.1972, 1.3945, 1.5933, 1.8027	0.74406	103.599	0. 63337	121.631	0. 63335	121.638
6	1.1650, 1.3300, 1.4950, 1.6600,1.8250	0.74371	103.647	0.62945	122.438	0.62925	122.476

Table 1 depicts the AOSB for different values of the number of strata L, when the stratification variable X follows uniform distribution. Also, Table 1 shows the percent relative efficiency of optimum stratification for single stratification variable as compared to unstratified RSS having variance 0.77083. It reveals that the percent relative efficiency ranges from 102.821-103.647 for L=2,3,4,5,6. The use of MPP moved within the interval 113.668-122-438, when an R-code is used and in 113.521-122.476 if Phyton algorithm was used. As expected, the gain in efficiency is much more satisfactory as compared to unstratified. MPP provides better AOSB´s and they do not differ significatively between them. From the above table, it is obvious that the relative efficiency has an increasing trend with respect to increase in number of strata.

If the stratification variable follows the Right triangular distribution with pdf $f(x) = \frac{2(2-x)}{(b-a)^2}$ using $a = 1$ and $b = 2$, utilizing the $cum\sqrt[3]{K_2(x)}$ rule, we get the stratification points as given in Table 2.

Table 2: AOSB and Variance when the auxiliary variable is Right-triangular distributed

$\eta(x) = \delta$							
L	**AOSB**	$(nV(\bar{y}_{SRSS})_{opt})$	**%R.E**	$(nV(\bar{y}_{SRSS})_{MPP-R})$	**%R.E**	$n(V(\bar{y}_{SRSS})_{MPP-Ph})$	**%R.E**
2	1.4179	0.66702	115.563	0.66806	115.439	0.68006	113.393
3	1.2648, 1.5674	0.66408	116.075	0.65870	117.021	0.69 870	111.537
4	1.1973, 1.4142, 1.6574	0.66304	116.258	0.60182	128.465	0.60912	100.268
5	1.1573, 1.3224, 1.5103, 1.7188	0.66254	116.345	0.60006	128.472	0.60 006	128.472
6	1.1220, 1.2516, 1.3911, 1.5443, 1.7253	0.66224	116.397	0.60070	128.424	0.600681.	128.222

Table 2 reveals that the AOSB for different values of the number of strata L, when the stratification variable X follows Right- triangular distribution. Also Table 2 displays the percent relative efficiency of optimum stratification for single stratification variable as compared to unstratified RSS having variance 0.68056. It depicts that the percent relative efficiency varies from 115.563-116.397 for L=2,3,4,5,6. The use of MPP provided %RE´s in 1135.439-128.424, when an R-code is used and in 113.39-128.22 if Phyton algorithm was used. Note that the gain in efficiency is satisfactory larger, when the methods are compared with unstratified. From the above table, it is obvious that the relative efficiency has an increasing trend with respect to increase in number of strata.

If the stratification variable follows the exponential distribution with pdf $f(x) = e^{-x+1}$ $x \in [1,5]$, utilizing the $cum\sqrt[3]{K_2(x)}$ rule, we get the stratification points as given in Table 3.

Table 3: AOSB and Variance when the auxiliary variable is exponentially distributed

$\eta(x) = \delta$							
L	**AOSB**	$(nV(\bar{y}_{SRSS})_{opt})$	**%R.E**	$(nV(\bar{y}_{SRSS})_{MPP-R})$	**%R.E**	$n(V(\bar{y}_{SRSS})_{MPP-Ph})$	**%R.E**
2	2.2562	0.98779	114.65	0.42614	252.65	0.42925	250.93
3	1.6329, 3.5718	0.95144	119.03	0.42746	254.91	0.42941	253.71

4	2.3676, 3.0136, 3.8711	0.93050	121.71	0.42747	265.05	0.42970	263.28
5	1.4904, 2.0771, 2.7868, 3.7301w	0.90027	125.79	0.42850	264.51	0.43054	263.58
6	1.4071, 1.8624, 2.5062, 3.1199, 3.8408	0.89740	126.19	0.42851	264.98	0.43129	263.21

Table 3 presents that the AOSB for different values of the number of strata L, when the stratification variable X follows Exponential distribution. Also Table 3 displays the percent relative efficiency of optimum stratification for single stratification variable as compared to unstratified RSS having variance 1.13251. It depicts that the percent relative efficiency ranges from 114.65-126.19 for L=2,3,4,5,6. MPP stratification produced %RE`s which moved within the interval 252.657.264.98, when an R-code is used and in 250.93-263.21 if Phyton algorithm was used. The gain in efficiency due to the use of MPP very large when compared with the other models. From the above table, it is obvious that the relative efficiency has an increasing trend with respect to increase in number of strata.

Let us suppose that the auxiliary variable X follows Standard normal distribution with pdf as

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad -\infty \le x \le \infty$$

In order to obtain the OSB when the variable is Standard normal distributed, using the proposed $cum\sqrt[3]{K_2(x)}$ rule the distribution is truncated at $x \in [0,1]$. By solving it in Mathematica software we get the stratification points as below:

Table 4: AOSB and Variance when the auxiliary variable is Standard normally distributed

$\eta(x) = \delta$							
L	**AOSB**	$(nV(\bar{y}_{SRSS})_{opt})$	**%R.E**	$(nV(\bar{y}_{SRSS})_{MPP-R})$	**%R.E**	$n(V(\bar{y}_{SRSS})_{MPP-Ph})$	**%R.E**
2	0.5528	0.02553	380.571	0.02236	434.588	0.02409	423.000
3	0.4685, 0.7255	0.02459	395.156	0.02213	433.010	0.02429	395.337
4	0.3398, 0.5705, 0.7720	0.02425	400.780	0.02209	439.198	0.02343	414.079
5	0.2981, 0.4929, 0.6676, 0.8350	0.02403	404.383	0.02116	4599.104	0.02116	4599.104
6	0.2636, 0.4316, 0.5836, 0.7251, 0.8584	0.02391	406.317	0.021080	460.861	0.021082	460.861

Table 4 reveals that the AOSB for different values of the number of strata L, when the stratification variable X follows Standard normal distribution. Also Table 4 displays the percent relative efficiency of optimum stratification for single stratification variable as compared to unstratified RSS having variance 0.09717. It depicts that the percent relative efficiency ranges from 380.571-406.317 for L=2,3,4,5,6. In this case the MPP overcome significatively the gain in efficiency due to the proposed OBS procedure. From the above table, it is obvious that the relative efficiency has an increasing trend with respect to increase in number of strata.

If the stratification variable follows the uniform distribution with pdf $f(x) = \frac{1}{b-a}$, $x \in [1,2]$, utilizing the $cum\sqrt[3]{K_2(x)}$ rule, we get the stratification points as given in Table 5.

Table 5: AOSB and Variance when the auxiliary variable is uniformly distributed

$\eta(x) = \lambda x$							
L	**AOSB**	$n\{V(\bar{y}_{SRSS})_{opt}\}$	**%R.E**	$(nV(\bar{y}_{SRSS})_{MPP-R})$	**%R.E**	$n(V(\bar{y}_{SRSS})_{MPP-Ph})$	**%R.E**
2	1.4798	0.74996	102.783	0.6881	111.938	0.6599	120.261
3	1.3241, 1.6504	0.74606	103.321	0.6760	114.011	0.6353	121.371

4	1.2391, 1.4891, 1.7411	0.74469	103.511	0.65732	117.205	0.6316	121.300
5	1.1914, 1.3707, 1.5902, 1.7970	0.74405	103.599	0.61904	124.520	0.61728	124.866
6	1.1589, 1.3167, 2.2109, 1.6470, 1.8196	0.74399	103.609	0.60542	127.312	0.60549	127.312

Table 5 depicts the AOSB for different values of the number of strata L, when the stratification variable X follows uniform distribution. Also Table 5 shows the percent relative efficiency of optimum stratification for single stratification variable as compared to unstratified RSS having variance 0.77083. It reveals that the percent relative efficiency ranges from 102.783-103.609 for L=2,3,4,5,6. The MPP produces OBS that are more efficient than the use of the use of the usual approximation . The computation of the DP solutions provided by R-code (111.938-127.312)behaves worse than the Phyton´s results (120.261-127.312). Note that $n(\mathrm{V}(\bar{y}_{\mathrm{SRSS}})_{\mathrm{MPP-Ph}})$ is less variable than $\mathrm{nV}(\bar{y}_{\mathrm{SRSS}})_{\mathrm{MPP-R}}$.

If the stratification variable follows the Right triangular distribution with pdf $f(x)=\frac{2(2-x)}{(b-a)^2}$ using $a=1$ and $b=2$, utilizing the $cum\sqrt[3]{K_2(x)}$ rule, we get the stratification points as given in Table 6.

Table 6: AOSB and Variance when the auxiliary variable is Right-triangular distributed

$\eta(x)=\lambda x$							
L	**AOSB**	$n\{V(\bar{y}_{SRSS})_{opt}\}$	**%R.E**	$(nV(\bar{y}_{SRSS})_{MPP-R})$	**%R.E**	$n(V(\bar{y}_{SRSS})_{MPP-Ph})$	**%R.E**
2	1.4089	0.66703	115.562	0.66045	116.826	0.66886	115.460
3	1.2590, 1.5604	0.66387	116.113	0.66248	116.356	0.66208	116.356
4	1.1904, 1.4031, 1.6480	0.66308	116.251	0.65770	117.201	0.66256	116,342
5	1.1494, 1.3115, 1.4957, 1.7086	0.66276	116.306	0.65718	117.308	0.65219	118.208
6	1.1236, 1.2538, 1.3938, 1.5504, 1.7302	0.66225	116.397	0.64882	118,851	0.64246	120.087

Table 6 reveals that the AOSB for different values of the number of strata L, when the stratification variable X follows Right- triangular distribution. Also Table 6 displays the percent relative efficiency of optimum stratification for single stratification variable as compared to unstratified RSS having variance 0.68056. It depicts that the percent relative efficiency ranges from 115.562-116.397 for L=2,3,4,5,6. The gain in efficiency associated with the three methods are very similar.

If the stratification variable follows the exponential distribution with pdf $f(x)=e^{-x+1}$ $x\in[1,5]$, utilizing the $cum\sqrt[3]{K_2(x)}$ rule, we get the stratification points as given in Table 7.

Table 7: AOSB and Variance when the auxiliary variable is exponentially distributed

$\eta(x)=\lambda x$							
L	**AOSB**	$n\{V(\bar{y}_{SRSS})_{opt}\}$	**%R.E**	$(nV(\bar{y}_{SRSS})_{MPP-R})$	**%R.E**	$n(V(\bar{y}_{SRSS})_{MPP-Ph})$	**%R.E**
2	2.3229	0.94775	119.494	0.86595	130.914	0.86437	131.253
3	1.8078, 2.9558	0.91677	123.533	0.86185	131.322	0.86442	130.929
4	1.5695, 2.3089, 3.3625	0.90532	125.095	0.81389	139.083	0.84455	135.777
5	1.4307, 1.9592, 2.6458, 3.5701	0.89994	125.843	0.80375	140.884	0.83439	135.700
6	1.6345, 2.3289, 2.8298, 3.5498, 4.2189	0.90126	125.659	0.79463	142.486	0.80452	140.824

Table 7 presents that the AOSB for different values of the number of strata L, when the stratification variable X follows Exponential distribution. Also Table 7 displays the percent relative efficiency of optimum stratification for single stratification variable as compared to unstratified RSS having variance 1.13251. It depicts that the

percent relative efficiency ranges from 119.494-125.843 for L=2,3,4,5, and for L=6 efficiency decreases. The gain in efficiency of MPP methods are more 10%-17% larger than the obtained by the standard approach.
Let us suppose that the auxiliary variable X follows Standard normal distribution with pdf as

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad -\infty \le x \le \infty$$

In order to obtain the OSB when the variable is Standard normal distributed, using the proposed $cum\sqrt[3]{K_2(x)}$ rule the distribution is truncated at $x \in [0,1]$. By solving it in Mathematica software we get the stratification points as below:

Table 8: AOSB and Variance when the auxiliary variable is Standard normally distribute

$\eta(x) = \lambda x$							
L	**AOSB**	$n\{V(\bar{y}_{SRSS})_{opt}\}$	**%R.E**	$(nV(\bar{y}_{SRSS})_{MPP-R})$	**%R.E**	$n(V(\bar{y}_{SRSS})_{MPP-Ph})$	**%R.E**
2	0.5386	0.02570	378.108	0.02375	427.309	0.02516	424.488
3	0.3863, 0.6898	0.02480	391.755	0.02159	449.277	0.02456	394.966
4	0.3150, 0.5461, 0.7732	0.02444	397.619	0.02076	468.692	0.02335	416.702
5	0.2651, 0.4563, 0.6372, 0.8167	0.02422	401.206	0.01977	494.861	0.02258	433.376
6	0.2328, 0.3956, 0.5476, 0.6975, 0.8391	0.02409	403.388	0.01851	525.494	0.02051	474.228

Table 8 reveals that the AOSB for different values of the number of strata L, when the stratification variable X follows Standard normal distribution. Also Table 8 displays the percent relative efficiency of optimum stratification for single stratification variable as compared to unstratified RSS having variance 0.09717. It depicts that the percent relative efficiency ranges from 378.108-403.388 for L=2,3,4,5,6. The R-solutions have gains in the interval 427.309-525.494 %and for Phyton soliton they are in the interval 424.488-477.228% By perusal of tables 1, 2, 3 and 4 the distributions perform well for $\eta(x) = \delta$. It is observed that the gain in efficiency is high in case of standard normal distribution and little more in case of uniform distribution.
From tables 5, 6, 7 and 8, for $\eta(x) = \lambda x$ the distributions perform well. It is observed that the gain in efficiency is high in case of standard normal distribution and little more in case of Uniform distribution.

7. CONCLUSION

For obtaining the stratification points under RSS, we have assumed different distributions for the auxiliary variable used as stratification variable. The AOSB obtained for uniform, right triangular, exponential and standard normal distributions are presented in table 1-4 and table 5-8 for $\eta(x) = \delta$ and $\eta(x) = \lambda x$ respectively. The right-triangular distribution shows highest % R.E for $\eta(x) = \delta$ and $\eta(x) = \lambda x$. The increase in the number of strata is directly proportional to the decrease in total variance. These figures show a considerable gain in efficiency of estimators when the proposed method of determining AOSB is used for all $L = 1,2,\dots,6$. Thus, the proposed method of $cum\sqrt[3]{K_2(x)}$ shows an increase in gain % in precision while selecting samples using RSS. This method works on a single auxiliary variable but in reality, surveys involve multiple auxiliary variables.
Note that:

1. MPP approaches overcome the use of the standard procedure for determining AOSB.
2. Always the gain in efficiency was much more satisfactory as compared to unstratified.
3. The relative efficiencies has an increasing trend with respect to increase in the number of strata.

Developing methods for multiple auxiliary variables as well as for other skewed distributions are possibilities for future work.

CONFLICT OF INTERESTS. The authors declare that there is no conflict of interests regarding the publication of this paper.

ACKNOWLEDGEMENTS: The authors thanks two anonymous referees for their remarks and suggestions , which enhanced to improve the quality of the paper. The paper was benefited by a PNCB-Grant.

RECEIVED: APRIL, 2023.
REVISED: JUNE, 2023.

REFERENCES

[1] ABO_EL.HASSAN FS, HAMED R, ISMAIL EA, EZZAT SM (2022) : Optimum stratum boundaries and sample sizes for Covid-19 data in Egypt. **PLoS ONE 17,** e0271220. https://doi.org/10.1371/journal.pone.0271220

[2] ALLENDE-ALONSO AND BOUZA-HERRERA (1987): Optimization Criteria For Multivariate Strata Construction. In "**Approximation En Optimization**", 120-138, Lecture Notes In Mathematics 1353, Springer, N. York.

[3] BOUZA, C.N., S. M. ALLENDE AND POONAM SINGH (2013): Optimal Post-stratification for the study of the sustainability: an application to the monitoring of diversity in Sierra de Guerrero, **Conference: Approximation and Optimization in the Caribbean**, Puebla, Mexico, DOI : http://dx.doi.org/10.13005/ojcst/10.01.02

[4] BOUZA, C.N., S. M. ALLENDE AND POONAM SINGH (2018): Optimal sample size for stratified sampling: a numerical study, **4th Conference on Optimization Methods and Software D**ecember 16-20, 2017, Havana, Cuba.

[5] BRITO, J., GUSTAVO SEMAAN, AUGUSTO FADEL, LEONARDO DE LIMA & NELSON MACULAN (2022): Mathematical programming formulations for the optimal stratification problem, **Communications in Statistics - Simulation and Computation**, DOI: 10.1080/03610918.2022.2092138

[6] BRITO J.A.M. , G.S. SEMAAN , L. DE LIMA CAND A.C.FADEL (2022): Heuristic Algorithm for Univariate Stratification Problem **arXiv:2211.10809v1**

[7] COCHRAN, W. G. (1961): Comparison of methods of determining strata boundaries, **Bull. bit. Statist. Inst**., 38, 345-358.

[8] DALENIUS, T. (1950): The problem of optimum stratification, Skand. Aktuarietidskr., 33, 203-213.

[9] Dalenius, T. And Gurney, M. (1951): The problem of optimum stratification II, **Skand. Aktuarietidskr**., 34, 133-148.

[10] DALENIUS, T. AND HODGES, J. L. (1957): The choice of stratification points, **Skand. Akluarietidskr**., 3-4, 198-203.

[11] DALENIUS, T. AND HODGES, J. L. (1959): Minimum variance stratification, J**. Amer. Statist. Assoc**., 54, 88—101.

[12] 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/

[13] .RIZVI, S.E.H., MANISH KUMAR SHARMA, M. IQBAL JEELANI, BUNTI KUMAR AND Q. FAROOQ DAR (2019): . Optimum Stratification For Two Stratifying Variables, **Revista Investigación Operacional**, 40 , 562-573.

[14] GUPTA. K, B, AND AHAMED. I, M, (2021): Construction of Strata for a Model-Based Allocation Under a Super population Model. **Journal of Statistical Theory and Applications**. 20(1): 46-60.

[15] FANOLAHI, A.V. AND KHAN, M.G.M. (2014): Determining the optimum strata boundaries with constant cost factor. Conference: **IEEE Asia-Pacific World Congress on Computer Science and Engineering (APWC),** At Plantaion Island, Fij.

[16] HIDIROGLOU, M. A., AND KOZAK, M. (2017): Stratification of skewed Populations: A Comparison of Optimisation based versus approximate Methods. **International Statistical Review**. DOI:10.1111/insr.12230.

[17] KHAN, M.G.M., AHMAD, N. AND KHAN, S. (2009): Determination the optimum stratum boundaries using. mathematical programming. **J. Mathemat. Modelling & Algorithms**, 8, 409-423.

[18] KHAN, E. A. KHAN, M. G. M. AND AHSAN, M. J. (2002): Optimum stratification: A mathematical programming approach. **Calcutta Statist. Assoc. Bull**., 52 : 323-333.

[19] LONE, M. A., S. A. MIR , IMRAN KHAN AND M. S. WANI (2017): Optimal Allocation of Stratified Sampling Design Using Gradient Projection Method. Oriental Journal of Computer Science and Technology

An International research journal of Computer Science and Technology , **DOI :** http://dx.doi.org/10.13005/ojcst/10.01.02
[20] MCINTYRE, G. A. (1952): A method for unbiased selective sampling, using ranked set sampling and stratified simple random sampling. **Journal of Applied Statistics** 23, 231-255.
[21] MAHAJAN, P.K. AND VERMA, M.R. (2004): An Alternative Method of Optimum Stratification for Scrambled Response. **Journal of the Indian Society of the Agricultural Statistics**. 58, 269-278.
[22] MEHTA, S. AND MAQNDOWARA, V.L.(2012): An optimum stratification for stratified cluster sampling design when clusters are of varying sizes. **International Journal of Scientific and Technology Research**, 1, :74-79.
[23] KHAN, M., M. G., NIRAJ NAND AND NESAR AHMAD (2008): Determining the optimum strata boundary points using dynamic programming. **Survey Methodology**, 34,. 205-214 .
[24] RATHER, K.U.I. AND KADILAR, C. (2021): Exponential Type Estimator for the population Mean under Ranked Set Sampling. **Journal of Statistics: Advances in Theory and Applications**, 25, 1–12.
[25] RATHER, K.U.I., EDA, K. G. UNAL, C. & JEELANI, M. I. (2022): New exponential ratio estimator in Ranked set sampling. **Pakistan Journal of Statistics and Operation Research**, 18, 403–409.
[26] RATHER, K.U.I., RIZVI, S.E.H., SHARMA, M. AND JEELANI, M. I. (2022): Optimum Stratification for Equal Allocation using Ranked Set Sampling as Method of Selection. **International Journal of Statistics and Applied Mathematics.** 7,: 63-67.
[27] RATHER, K.U.I. AND RIZVI, S.E.H. (2022): Approximate Optimum Strata Boundaries for Equal Allocation under Ranked Set Sampling. **Journal of Scientific Research**, 66,: 302-307.
[28] RATHER, K.U.I., RIZVI, S.E.H., SHARMA, M., JEELANI, M. I. AND DANISH, F. (2023): Approximate Optimum Strata Boundaries for Proportional Allocation using Ranked Set Sampling. **Reliability: Theory & Applications,** 72, 161-171,.
[29] REDDY, K. G. AND M. G. M. KHAN (2020): StratifyR: An R Package for optimal stratification and sample allocation for univariate populations, HTTPS://DOI.ORG/10.1111/ANZS.12301 last consult May, 2023..
[30] RICHTER-WALSH S. (2017): Implementing the Gradient Descent Algorithm in R. R bloggers, https://www.r-bloggers.com/2017/02/implementing-the-gradient-descent-algorithm-in-r/. Last consulted 15-May 2023.
[31] RIZVI, S.E.H., GUPTA, J.P. AND SINGH, R. (2000): Approximately optimum Stratification for two study variables using auxiliary information. **Journal of the Indian Society of the Agricultural Statistics**. 53, 287-298.
[32] RIZVI, S. E. H., GUPTA, J. P. AND BHARGAVA, M. (2002): Optimum stratification based on auxiliary variable for compromise allocation. **Metron**, 28, 201-215.
[33] ROSEN, J.B. (1960): The gradient projection method For Nonlinear Programming, Part i, linear constraints. **SIAM journal of Applied Mathematics** 8, 181-219.
[34] SERIFING, R. J. (1968): Approximately optimum stratification, **J. .4mer. Statist. Assoc**., 63, 1298-1309
[35] SINGH, R. AND SUKHATME, B. V. (1969): Optimum stratification, **Ann. Inst. Statist. Math**., 21, 515-528.
[36] SINGH, R. (1971): Determination of optimum strata boundaries, J**. Indian Soc. Agric. Statist.,** 23 (Silver Jubilee Number), 115—123.
[37] SAMAWI, H.M. (1996): Stratified ranked set sample. **Pakistan Journal of Statistics**, 12:9-16.
[38] TAKAHASI, K. AND WAKIMOTO, K. (1968): On unbiased estimates of the population mean based on the stratified sampling by means of ordering. **Ann. Inst. Statist. Math**., 20, 1-31.