Mathematics and Statistics 9(5): 617-629, 2021 http://www.hrpub.org
DOI: 10.13189/ms.2021.090501

Methods of Stratification for a Generalised Auxiliary Variable Optimum Allocation

Md. Irphan Ahamed[1,*], Bhuwaneshwar Kumar Gupt[2], Manoshi Phukon[1]

[1]Department of Mathematics, Umshyrpi College, Shillong, Meghalaya, 793004, India
[2]Department of Statistics, North-Eastern Hill University, Shillong, Meghalaya, 793022, India

Received May 23, 2021; Revised August 2, 2021; Accepted August 22, 2021

Cite This Paper in the following Citation Styles

***(a):** [1] Md. Irphan Ahamed, Bhuwaneshwar Kumar Gupt, Manoshi Phukon, "Methods of Stratification for a Generalised Auxiliary Variable Optimum Allocation," Mathematics and Statistics, Vol. 9, No. 5, pp. 617 - 629, 2021. DOI: 10.13189/ms.2021.090501.*

***(b):** Md. Irphan Ahamed, Bhuwaneshwar Kumar Gupt, Manoshi Phukon (2021). Methods of Stratification for a Generalised Auxiliary Variable Optimum Allocation. Mathematics and Statistics, 9(5), 617 - 629. DOI: 10.13189/ms.2021.090501.*

Abstract In stratified sampling, ever since Dalenius [1] undertook the problem of optimum stratification, the research in the area has been progressing in various perspectives and dimensions till date. Amidst the multifaceted developments in the trend of the research, consideration of the topic by taking into account various aspects such as different sample selection methods and allocations, study variable based stratification, auxiliary variable based stratification, superpopulation models, extension to two study variables for a single auxiliary variable, extension to two stratification variables for a single study variable etc., are a few noteworthy ones. However, with regard to considering optimum stratification of heteroscedastic populations, as live populations are generally heteroscedastic, it was Gupt and Ahamed [2,3] who considered the problem for a few allocations under a heteroscedastic regression superpopulation (HRS) model. As a sequel to the work of the authors, in this paper, the problem of optimum stratification for an objective variable y based on a concomitant variable x under the HRS model is considered for an allocation proposed by Gupt [4,5] and termed as Generalised Auxiliary Variable Optimum Allocation (GAVOA). Methods of stratification in the form of equations and approximate solutions to the equations which stratify populations at optimum strata boundaries (OSB) and approximately optimum strata boundaries (AOSB) respectively are obtained. Mathematical analysis is used in minimizing sampling variance of the estimator of population mean and deriving all the proposed methods of stratification. The proposed equations divide heteroscedastic populations, symmetrical or moderately skewed or highly skewed, at OSB, but, the equations are implicit in nature and not easy in solving. Therefore, a few methods of finding AOSB are deduced from the equations through analytically justified steps of approximation. The methods may provide practically feasible solutions in survey planning in stratifying heteroscedastic population of any level of heteroscedasticity and the work may contribute, to some extent, theoretically in the research area. The methods are empirically examined in a few generated heteroscedastic data of varied shapes with some assumed levels of heteroscedasticity and found to perform with high efficiency. The proposed methods of stratification are restricted to the particular allocation used.

Keywords Characteristic Under Study, Heteroscedastic Regression Superpopulation Model, Generalised Auxiliary Variable Optimum Allocation, Optimum Strata Boundaries, Probability Density Function

1. Introduction

In sample survey, since the precision of an estimator of a population parameter depends on the heterogeneity of the units of the population besides the sample size and sampling fraction, the role of stratified sampling method comes into play as one possible way to enhance the precision of the estimator. In stratified sampling, a heterogeneous population is divided into a number of strata so as to increase the homogeneity among population units within strata and then a sample is drawn from each stratum by using any suitable sample selection method. The main aspects which are to be dealt with tactically for enhancing precision of an estimator of a population parameter are construction of strata, number of strata to be made, and allocation of sample size to strata. In the construction of strata, the major concerns are determination of OSB and choice of the best characteristic.

Tschuprow [6] and Neyman [7] developed method of allocation of sample size to strata for the first time in stratified sampling based on the characteristic under study. Cochran [8] showed that superpopulation model could be constructed such that finite population under study can be considered as a simple random sample from the superpopulation that provided information on auxiliary variable highly correlated with study variable is available. Among many sample size allocations to strata, it is pertinent to mention that Hanurav [9] and Rao [10] used information of the auxiliary variable for allocation of sample size to strata under the following superpopulation model.

$$\begin{aligned} &\text{(i) } \xi(y_i|x_i) = \alpha + \beta x_i \\ &\text{(ii) } \nu(y_i|x_i) = \sigma^2 x_i^g \qquad (1) \\ &\text{(iii) } \varsigma(y_i, y_j|x_i, x_j) = 0 \end{aligned}$$

where α, β, σ^2 and g are the superpopulation parameters and the script letters ξ, ν and ς denote conditional expectation, variance and covariance given $x's$ respectively.

Gupt and Rao [11] considered problem of optimum allocation of sample size to strata for probability proportional to size with replacement (PPSWR) under particular case, i.e., intercept $\alpha = 0$, of the superpopulation model (1).

It was Dalenius [1] who pioneered the work for determining OSB in stratifying population based on characteristic under study. Dalenius and Gurney [12] conjectured that by taking $W_h\sigma_{hy}$ constant, the optimum points of the study variable that divided the population into strata could be determined, where W_h and σ_{hy} are the stratum weight and standard deviation of the characteristic y in the h^{th} stratum. Mahalanobis [13], and Hansen, et al. [14] postulated that OSB of the study variable y could be determined by keeping $W_h\mu_{hy}$ constant, where μ_{hy} is the mean for y in h^{th} stratum. Dalenius and Hodges [15] endorsed the conjecture of Dalenius and Gurney [12] and again Dalenius and Hodges [16] proposed method of cumulating the values of $\sqrt{f(y)}$ for finding OSB. Ekman [17] proposed that OSB of study variable y could be obtained by ensuring $W_h(y_h - y_{h-1})$ constant. Sethi [18] demonstrated that postulates made by Hansen, et al. [14] did not give OSB for certain types of population.

It is unrealistic to assume that stratification should be done based on study variable y whose information is not available in practice and therefore some other known variable x which is highly correlated with the study variable y should be used. For a given number of strata, Dalenius [97] obtained equations giving OSB based on auxiliary variable for proportional and Tschuprow [6] and Neyman [7] optimum allocation (TNOA). Taga [20] too obtained OSB for the objective variable y based on the concomitant variable x and showed that the optimum stratification method, he proposed, reduced to the optimum decomposition of the distribution function $H(z)$for the random variable z= $\eta(x)$, where $\eta(x)$ is the regression function of y on x. It was Singh and Sukhatme [21] who first made a breakthrough in not only obtaining equations giving OSB but also methods of approximation for obtaining AOSB for TNOA and proportional allocation based on auxiliary variable x when both the form of regression of study variable y on x and variance function $V(y|x)$ are known, and since then a number of researchers such as Singh [22], Singh and Sukhatme [23], Singh [24-26], Singh and Prakash [27], and Yadava and Singh [28], to mention a few notable ones among many, furthered the work in the direction under various allocations and sample selection methods. The theory of optimum stratification based on single study variable with and without auxiliary variable was extended to more than one variate by, inter alia, Ghosh [29], Sadasivan and Aggarwal [30], Gupta and Seth [31], Rizvi et al. [32], Rizvi et al. [33,34], Verma [35], etc., under various allocations, sample selection methods and some other exceptional conditions. Rabee et al., [36] proposed a multivariate calibration estimation of the population mean of a study variable under stratified random sampling scheme using two auxiliary variables.

Gupt [4,5] dealt with problem of allocation of sample size to strata by modifying the model (1) into a more general form in which element of correlation among the units within strata was taken into account, and hence he obtained a few generalised model-based allocations; Gupt and Ahamed [2,3] obtained methods of stratifying heteroscedastic populations for finding OSB and AOSB for two of the few generalised model-based allocations under simple random sample with and without replacement (SRSWR and SRSWOR) designs. Gupt et al., also [37] obtained equations giving OSB and AOSB for the auxiliary variable optimum allocation (AVOA) proposed by Hanurav [9]. One of the above mentioned generalised allocations proposed by Gupt [4,5] under the presumption of equality of co-efficient of variation of $x^{g/2}$ in all strata, is as follows:

$$n_h \alpha\, N_h\sigma_h\left(x^{g/2}\right), \qquad (2)$$

provided the ratio $\frac{\sigma_h(x)}{\sigma_h(x^{g/2})}$ are equal in all strata .

The above allocation can give $n_h \alpha N_h \sigma_{hx}$, i.e., AVOA when $g = 2$; therefore, Gupt [5] justifiably named the allocation (2) as Generalised Auxiliary Variable Optimum Allocation (GAVOA).

In this paper, we deal with the problem of finding OSB and AOSB for the allocation (2), i.e., GAVOA under SRSWR design which also holds true for SRSWOR design when finite population correction is neglected.

The paper comprises five sections. Section 2 contains the derivation of equations giving OSB and section 3 gives the derivation of the approximate solutions of the equations giving OSB to obtain a few methods of stratification giving AOSB. In section 4, empirical illustration of all the proposed methods is conducted in some generated populations and the results are discussed about. Section 5 gives conclusion.

2. Method of Obtaining OSB

The GAVOA given by (2) can be represented as follows:

$$n_h = n\frac{W_h\sigma_h(x^{g/2})}{\sum_h^L W_h\sigma_h(x^{g/2})} \tag{3}$$

where L is the number of strata into which the population of size N is divided such that $\sum_{h=1}^{L} N_h = N$, and n_h is size of the sample to be selected from the population of size N_h of the h^{th} stratum such that $\sum_{h=1}^{L} n_h = n$.

Using (1) and (3), the sampling variance of the stratified sampling can be expressed as follows:

$$nV(\bar{y}_{st}) = \sum_{h=1}^{L}\frac{W_h\left(\beta^2\sigma_{hx}^2+\sigma^2\mu_h(x^g)\right)}{\sigma_h(x^{g/2})}\sum_{h=1}^{L} W_h\sigma_h\left(x^{g/2}\right) \tag{4}$$

From (4), we can again obtain

$$n\,V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h\left[\beta^2\frac{\sigma_{hx}}{\sigma_h(x^{g/2})}\sigma_{hx} + \sigma^2\frac{1}{C_h(x^{g/2})}\{1 + C_h^2\left(x^{g/2}\right)\}\mu_h\left(x^{g/2}\right)\right]\sum_{h=1}^{L} W_h\sigma_h\left(x^{g/2}\right) \tag{5}$$

where $C_h\left(x^{g/2}\right) = \frac{\sigma_h(x^{g/2})}{\mu_h(x^{g/2})}$ is the coefficient of variation of $x^{g/2}$ in h^{th}stratum.

Since Gupt [4,5] presumed $C_h\left(x^{g/2}\right)$ is equal for all h and obtained the allocation (2) provided $\frac{\sigma_{hx}}{\sigma_h(x^{g/2})}$ is equal for all h, here we assume constants k and k' in lieu of $\frac{\sigma_{hx}}{\sigma_h(x^{g/2})}$ and $\frac{1+\ C_h^2(x^{g/2})}{C_h(x^{g/2})}$ in (5), we get the following:

$$n\,V(\bar{y}_{st}) = \beta^2 k\sum W_h\sigma_{hx}\sum W_h\sigma_h\left(x^{g/2}\right) + \sigma^2 k'\sum W_h\mu_h\left(x^{g/2}\right)\sum W_h\sigma_h\left(x^{g/2}\right). \tag{6}$$

Since β^2, σ^2, k and k' are the positive real valued quantities, minimising $n\,V(\bar{y}_{st})$ in (6) is equivalent to minimising the following expression.

$$\sum W_h\sigma_{hx}\sum W_h\sigma_h\left(x^{g/2}\right) + \sum W_h\mu_h\left(x^{g/2}\right)\sum W_h\sigma_h\left(x^{g/2}\right). \tag{7}$$

Differentiating expression (7) partially with respect to x_h and equating the result to zero, we get

$$\left\{\sum W_h\sigma_{hx} + \sum W_h\mu_h\left(x^{g/2}\right)\right\}\frac{\delta}{\delta x_h}\sum W_h\sigma_h\left(x^{g/2}\right) + \sum W_h\sigma_h\left(x^{g/2}\right)\left\{\frac{\delta}{\delta x_h}\sum W_h\sigma_{hx} + \frac{\delta}{\delta x_h}\sum W_h\mu_h\left(x^{g/2}\right)\right\} = 0\ . \tag{8}$$

Considering $f(x)$ as probability density function of the auxiliary variable x, we have the following:

$$W_h\mu_h\left(x^{g/2}\right) = \int_{x_{h-1}}^{x_h} x^{g/2} f(x)dx. \tag{9}$$

$$2W_h\sigma_h\left(x^{g/2}\right)\frac{\delta\sigma_h(x^{g/2})}{\delta x_h} = \left\{x_h^{g/2} - \mu_h\left(x^{g/2}\right)\right\}^2 f(x_h) - \sigma_h^2\left(x^{g/2}\right)f(x_h). \tag{10}$$

From (8), we have

$$\left\{\sum W_h\sigma_{hx} + \sum W_h\mu_h\left(x^{g/2}\right)\right\}\left\{\frac{\delta}{\delta x_h}W_h\sigma_h\left(x^{g/2}\right) + \frac{\delta}{\delta x_h}W_{h+1}\sigma_{h+1}\left(x^{g/2}\right)\right\} + \sum W_h\sigma_h\left(x^{g/2}\right)\left\{\frac{\delta}{\delta x_h}W_h\sigma_{hx} + \frac{\delta}{\delta x_h}W_{h+1}\sigma_{(h+1)x} + \frac{\delta}{\delta x_h}W_h\mu_h\left(x^{g/2}\right) + \frac{\delta}{\delta x_h}W_{h+1}\mu_{h+1}\left(x^{g/2}\right)\right\} = 0. \tag{11}$$

Putting $A = \sum W_h\sigma_{hx} + \sum W_h\mu_h\left(x^{g/2}\right)$ and $B = \sum W_h\sigma_h\left(x^{g/2}\right)$ in (11) and using (9) and (10), we obtain the terms of the equations (11) as follows:

$$\begin{aligned}
&A\frac{\delta}{\delta x_h}W_h\sigma_h\left(x^{g/2}\right)\\
&= A\left[\frac{\left\{x_h^{g/2}-\mu_h(x^{g/2})\right\}^2+\sigma_h^2(x^{g/2})}{2\sigma_h(x^{g/2})}\right]f(x_h).\\
&A\frac{\delta}{\delta x_h}W_{h+1}\sigma_{h+1}\left(x^{g/2}\right)\\
&= A\left[\frac{\left\{x_h^{g/2}-\mu_{h+1}(x^{g/2})\right\}^2+\sigma_{h+1}^2(x^{g/2})}{2\sigma_{h+1}(x^{g/2})}\right]\{-f(x_h)\}.\\
&B\left\{\frac{\delta}{\delta x_h}W_h\sigma_{hx} + \frac{\delta}{\delta x_h}W_h\mu_h\left(x^{g/2}\right)\right\}\\
&= B\left[\frac{(x_h-\mu_{hx})^2+\sigma_{hx}^2}{2\sigma_{hx}} + x_h^{g/2}\right]f(x_h).\\
&B\left\{\frac{\delta}{\delta x_h}W_{h+1}\sigma_{h+1x} + \frac{\delta}{\delta x_h}W_{h+1}\mu_{h+1}\left(x^{g/2}\right)\right\}\\
&= B\left[\frac{\left(x_h-\mu_{(h+1)x}\right)^2+\sigma_{(h+1)x}^2}{2\sigma_{(h+1)x}} + x_h^{g/2}\right]\{-f(x_h)\}.
\end{aligned} \tag{12}$$

From (11) and (12), we get

$$\left\{\sum W_h\sigma_{hx} + \sum W_h\mu_h\left(x^{g/2}\right)\right\}\left[\frac{\left\{x_h^{g/2}-\mu_h\left(x^{g/2}\right)\right\}^2+\sigma_h^2\left(x^{g/2}\right)}{\sigma_h\left(x^{g/2}\right)}\right]$$
$$+\sum W_h\sigma_h\left(x^{g/2}\right)\left\{\frac{(x_h-\mu_{hx})^2+\sigma_{hx}^2}{\sigma_{hx}}\right\} = \left\{\sum W_h\sigma_{hx} + \sum W_h\mu_h\left(x^{g/2}\right)\right\}\left[\frac{\left\{x_h^{g/2}-\mu_{h+1}\left(x^{g/2}\right)\right\}^2+\sigma_{h+1}^2\left(x^{g/2}\right)}{\sigma_{h+1}\left(x^{g/2}\right)}\right]$$
$$+\sum W_h\sigma_h\left(x^{g/2}\right)\left[\frac{\left\{x_h-\mu_{(h+1)x}\right\}^2+\sigma_{(h+1)x}^2}{\sigma_{(h+1)x}}\right]. \quad (13)$$

Equations (13) give OSB of the objective variable y in terms of the auxiliary variable x.

3. Approximate Solutions to the Equations Giving OSB to Obtain Methods of finding AOSB

In this section, we derive approximate solutions of equations (13) that will give AOSB in stratifying heteroscedastic populations. The techniques used by Singh and Sukhatme [21], Gupt and Ahamed [2,3] etc., are followed in obtaining series expansions of the system of equations (13). The techniques adopted by Singh and Sukhatme [21] by using Ekman's [38] identity in carrying out series expansion of conditional mean and variance are also followed and extended in this section in the same way Gupt and Ahamed [2,3] explored. We assume $\psi(x) = x^{g/2}$, and it is also stipulated at this juncture to assume the existence of continuous first two partial derivatives of $f(x)$ and $\psi(x)$ with respect to x_h in $[x_{h-1}, x_{h+1}]$ for all values of h.

Considering parameters, terms and expressions of the right hand side of equations (13), evaluating all the derivatives at $z = x_h$, $\forall z \in [x_h, x_{h+1}]$ and putting $k_{h+1} = x_{h+1} - x_h$, we proceed as follows:

$$\mu_{h+1}\left(x^{g/2}\right) = \mu_{(h+1)\psi} = \psi + \frac{\psi'}{2}k_{h+1} + \frac{2\psi'' f + \psi' f'}{12f}k_{h+1}^2$$
$$+\frac{ff''\psi' + ff'\psi'' + f^2\psi''' - \psi' f'^2}{24f^2}k_{h+1}^3$$
$$+\frac{\left\{\begin{matrix}9\psi' f''' f^2 + 16\psi'' f'' f^2 + 9\psi''' f^2 f' + 6\psi^{iv} f^3 \\ -25\psi' f' f'' f - 15ff'^2\psi'' + 15\psi' f'^3\end{matrix}\right\}}{720f^3}k_{h+1}^4 + O(k_{h+1}^5) \ .$$

$$\Rightarrow \left\{x_h^{g/2} - \mu_{h+1}\left(x^{g/2}\right)\right\}^2 = \frac{\psi'^2}{4}k_{h+1}^2 + \frac{2f\psi'\psi'' + f'\psi'^2}{12f}k_{h+1}^3$$
$$+\frac{\left\{\begin{matrix}4f^2\psi''^2 + 10ff'\psi'\psi'' - 5f'^2\psi'^2 \\ +6ff''\psi'^2 + 6f^2\psi'\psi'''\end{matrix}\right\}}{144f^2}k_{h+1}^4 + O(k_{h+1}^5). \quad (14)$$

$$\sigma_{(h+1)\psi}^2 = \frac{\psi'^2}{12}k_{h+1}^2 + \frac{\psi'\psi''}{12}k_{h+1}^3$$
$$+\frac{\left\{\begin{matrix}2ff''\psi'^2 + 24ff'\psi'' + 18f^2\psi'\psi''' \\ +16f^2\psi''^2 - 5f'^2\psi'^2 - 20ff'\psi\psi''\end{matrix}\right\}}{720f^2}k_{h+1}^4 + O(k_{h+1}^5). \quad (15)$$

Adding (14) and (15)

$$\left\{x_h^{g/2} - \mu_{h+1}\left(x^{g/2}\right)\right\}^2 + \sigma_{(h+1)\psi}^2 = \frac{\psi'^2}{3}k_{h+1}^2 + \frac{3f\psi'\psi'' + f'\psi'^2}{12f}k_{h+1}^3$$
$$+\frac{\left\{\begin{matrix}36f^2\psi''^2 + 50ff'\psi'\psi'' - 30f'^2\psi'^2 + 32ff''\psi'^2 \\ +48f^2\psi'\psi''' + 24ff'\psi'' - 20ff'\psi\psi''\end{matrix}\right\}}{720f^2}k_{h+1}^4 + O(k_{h+1}^5). \quad (16)$$

Again from (15), we get

$$\sigma_{(h+1)\psi}^{-1} = \frac{2\sqrt{3}}{k_{h+1}\psi'}\left[1 - \frac{\psi''}{2\psi'}k_{h+1} - \frac{\left\{\begin{matrix}2ff''\psi'^2 + 24ff'\psi'' + 18f^2\psi'\psi''' \\ -29f^2\psi''^2 - 5f'^2\psi'^2 - 20ff'\psi\psi''\end{matrix}\right\}}{120f^2\psi'^2}k_{h+1}^4\right] + O(k_{h+1}^5). \quad (17)$$

Multiplying (16) and (17)

$$\frac{\left\{x_h^{g/2} - \mu_{h+1}\left(x^{g/2}\right)\right\}^2 + \sigma_{(h+1)\psi}^2}{\sigma_{(h+1)\psi}} = \frac{2}{\sqrt{3}}\psi' k_{h+1} + \frac{f\psi'' + f'\psi'}{2\sqrt{3}f}k_{h+1}^2$$
$$+\frac{\left\{\begin{matrix}10ff'\psi'\psi'' + 2f^2\psi''^2 - 20f'^2\psi'^2 + 14ff''\psi'^2 \\ +6f^2\psi'\psi''' - 12ff'\psi'' - 30ff'\psi\psi''\end{matrix}\right\}}{60\sqrt{3f^2\psi'}}k_{h+1}^3 + O(k_{h+1}^4). \quad (18)$$

Similarly, we can obtain

$$\frac{\left(x_h - \mu_{(h+1)x}\right)^2 + \sigma_{(h+1)x}^2}{\sigma_{(h+1)x}} = \frac{2}{\sqrt{3}}k_{h+1} + \frac{f'}{2\sqrt{3f}}k_{h+1}^2$$
$$+\frac{12ff'' - 5f'^2}{60\sqrt{3f^2}}k_{h+1}^3 - \frac{2ff'f'' - 5f'^3}{120\sqrt{3}f^3}k_{h+1}^4 + O(k_{h+1}^5). \quad (19)$$

Therefore, using (18) and (19) we can obtain right hand side of equations (13) as

$$\text{RHS} = A\left\{\frac{2}{\sqrt{3}}\psi' k_{h+1} + \frac{f\psi'' + f'\psi'}{2\sqrt{3f}}k_{h+1}^2 + O(k_{h+1}^3)\right\}$$
$$+B\left\{\frac{2}{\sqrt{3}}k_{h+1} + \frac{f'}{2\sqrt{3f}}k_{h+1}^2 + \frac{12ff'' - 5f'^2}{60\sqrt{3f^2}}k_{h+1}^3\right\} + O(k_{h+1}^4)$$

$$= \frac{2}{\sqrt{3}}(A\psi' + B)k_{h+1} + \frac{A(f\psi'' + f'\psi') + Bf'}{2\sqrt{3f}}k_{h+1}^2 + O(k_{h+1}^3).$$

Taking $g(t) = A\psi'(t) + B$, where $t \in [x_{h-1}, x_{h+1}]$, $\forall\, h = 1, 2, \ldots, L$, we get

$$\text{RHS} = \frac{2}{\sqrt{3}} gf k_{h+1}\left\{1 + \frac{(gf)'}{gf}\frac{k_{h+1}}{4}\right\} + O(k_{h+1}^2)$$

Similarly, the left hand side of (13) can be obtained as

$$\text{LHS} = \frac{2}{\sqrt{3}} gf k_h\left\{1 - \frac{(gf)'}{gf}\frac{k_h}{4}\right\} + O(k_h^2),$$

where $k_h = x_h - x_{h-1}$.

Thus, equations (13) can be written as

$$\frac{2}{\sqrt{3}} gf k_h\left\{1 - \frac{(gf)'}{gf}\frac{k_h}{4}\right\} + O(k_h^2) = \frac{2}{\sqrt{3}} gf k_{h+1}\left\{1 + \frac{(gf)'}{gf}\frac{k_{h+1}}{4}\right\} + O(k_{h+1}^2).$$

Squaring both sides and expanding by binomial theorem, we can get

$$k_h^2(gf)^2\left\{1 - \frac{(gf)'}{gf}\frac{k_h}{2}\right\} + O(k_h^2) = k_{h+1}^2(gf)^2\left\{1 + \frac{(gf)'}{gf}\frac{k_{h+1}}{2}\right\} + O(k_{h+1}^2)$$

$$\Rightarrow k_h^2(gf)\left\{1 - \frac{(gf)'}{gf}\frac{k_h}{2}\right\} + O(k_h^2)$$

$$= k_{h+1}^2(gf)\left\{1 + \frac{(gf)'}{gf}\frac{k_{h+1}}{2}\right\} + O(k_{h+1}^2) \quad (20)$$

The identity obtained by Singh and Sukhatme [21] and used by Yadava and Singh [28], Gupt and Ahamed [2,3], Gupt et al. [37] etc., is as follows:

$$\left[\int_{x_h}^{x_{h+1}} \sqrt[\gamma]{f(z)}\, dz\right]^{\gamma}$$

$$= k_{h+1}^{\gamma} f(x_h)\left[1 + \frac{k_{h+1}}{2}\frac{f'(x_h)}{f(x_h)} + O(k_{h+1}^2)\right]$$

$$= k_{h+1}^{\gamma-1}\int_{x_h}^{x_{h+1}} f(z)dz\,\{1 + O(k_{h+1}^2)\} \quad (21)$$

By using (20) and (21), we can proceed as follows:

$$k_h\int_{x_{h-1}}^{x_h} g(t)f(t)dt\big(1 + O(k_h^2)\big)$$

$$= k_{h+1}\int_{x_h}^{x_{h+1}} g(t)f(t)dt\{1 + O(k_{h+1}^2)\}$$

$$\Rightarrow k_h\int_{x_{h-1}}^{x_h} g(t)f(t)dt = \text{constant} = C_1 \quad (22)$$

$$\Rightarrow \left[\int_{x_{h-1}}^{x_h}\sqrt{g(t)f(t)}dt\right]^2 = C_1$$

$$\Rightarrow \int_{x_{h-1}}^{x_h}\sqrt{g(t)f(t)}dt = C_2 \quad (23)$$

Thus, the AOSB are given by (22) and equivalently by (23), and the values of constants C_1 and C_2 can be approximately evaluated as,

$$C_1 = \frac{1}{L}(u - v)\int_u^v g(t)f(t)dt,\ C_2 = \frac{1}{L}\int_u^v\sqrt{g(t)f(t)}dt$$

respectively, where we assume u and v are upper and lower bounds of the points of stratification $x_h's$, i.e., $u \le x_h \le v$. The approximate solutions of OSB, i.e., $x_h's$ can be calculated by any of the methods (22) and (23) given lower boundaries x_{h-1}.

From the above analytical work involved in finding AOSB, the following theorem has been inferred.

Theorem 3.1: For a given number of strata, constructing equal intervals on the cumulative of $g(x)f(x)$ or $\sqrt{g(x)f(x)}$ yields AOSB provided the function $F(x) = g(x)f(x)$ is bounded and its first two derivatives exist $\forall x \in (u, v)$.

However, in the application of methods of finding AOSB (22) and (23), it is required to determine the values A and B under the conditions of AOSB for which the following lemma is used.

Lemma 3.1: Under the conditions of AOSB, the approximate values of A and B in the above assumption, i.e., $g(t) = A\psi'(t) + B, \forall t \in [x_{h-1}, x_{h+1}]$ and h =1, 2, ..., L are obtained as,

$$A = \int_u^v \psi(t)\,f(t)dt + \frac{1}{2\sqrt{3}L}\left\{\int_u^v\sqrt{f(t)}dt\right\}^2,$$

$$B = \frac{1}{2\sqrt{3}L}\left\{\int_u^v\sqrt{g_1(t)f(t)}\right\}^2, \text{ where } g_1(t) = \psi'(t).$$

Proof: From (15), we get,

$$\sigma_{(h+1)x}^2 = \frac{k_{h+1}^2}{12} + \frac{2ff' - 5f'^2}{720f^2}k_{h+1}^4 + O(k_{h+1}^5). \quad (24)$$

As obtained and used by Singh and Sukhatme [21] and followed by Yadava and Singh [28], Gupt and Ahamed [2,3], Gupt et al. [37], etc., we have

$$W_{h+1} = k_{h+1}f + \frac{k_{h+1}^2}{2}f' + \frac{k_{h+1}^3}{6}f'' + O(k_{h+1}^4). \quad (25)$$

Again by taking square root on both sides of (24)

$$\sigma_{(h+1)x} = \frac{k_{h+1}}{2\sqrt{3}} + \frac{2ff'' - 5f'^2}{240\sqrt{3}f^2}k_{h+1}^3 + O(k_{h+1}^4). \quad (26)$$

Adding the series expansion of $\mu_{(h+1)\psi}$, obtained above, to (26), we get,

$$\sigma_{(h+1)x} + \mu_{(h+1)\psi} = \psi + \frac{1 + \sqrt{3}\psi'}{2\sqrt{3}}k_{h+1} + \frac{\psi'f' + 2f\psi''}{12f}k_{h+1}^2$$

$$+ \frac{\left\{\begin{matrix}2ff'' - 5f'^2 + 10\sqrt{3}ff''\psi' + 10\sqrt{3}ff'\psi'' \\ +10\sqrt{3}f^2\psi''' - 10\sqrt{3}\psi'f'^2\end{matrix}\right\}}{240\sqrt{3}\,f^2}k_{h+1}^3 + O(k_{h+1}^4). \quad (27)$$

From (25) & (27), we get

$$W_{h+1}\left(\sigma_{(h+1)x}+\mu_{(h+1)\psi}\right)$$
$$=\psi f k_{h+1}+\frac{f+\sqrt{3}f\psi'+\sqrt{3}f'\psi}{2\sqrt{3}}k_{h+1}^2$$
$$+\frac{\left\{\begin{matrix}\sqrt{3}\psi'f'+2\sqrt{3}f\psi''+3f'\\+3\sqrt{3}f'\psi'+2\sqrt{3}f''\psi\end{matrix}\right\}}{12\sqrt{3}}k_{h+1}^3+O(k_{h+1}^4). \quad (28)$$

By using Taylor series expansion, we get the following expression.

$$\int_{x_h}^{x_{h+1}}\psi(t)f(t)dt=\psi f k_{h+1}+\frac{\psi' f+f'\psi}{2}k_{h+1}^2$$
$$+\frac{\psi'' f+2f'\psi'+f''\psi}{6}k_{h+1}^3+O(k_{h+1}^4). \quad (29)$$

Subtracting (29) from (28), we get,

$$W_{h+1}\left(\sigma_{(h+1)x}+\mu_{(h+1)\psi}\right)-\int_{x_h}^{x_{h+1}}\psi(t)f(t)dt$$
$$=\frac{f}{2\sqrt{3}}k_{h+1}^2+\frac{3f'}{12\sqrt{3}}k_{h+1}^3+O(k_{h+1}^4)$$
$$=\frac{k_{h+1}^2}{2\sqrt{3}}f\left\{1+\left(\frac{f'}{f}\right)\frac{k_{h+1}}{2}\right\}+O(k_{h+1}^4)\ .$$

Using identity (21), we get,

$$W_{h+1}\left(\sigma_{(h+1)x}+\mu_{(h+1)\psi}\right)-\int_{x_h}^{x_{h+1}}\psi(t)f(t)dt$$
$$=\frac{k_{h+1}}{2\sqrt{3}}\int_{x_h}^{x_{h+1}}f(t)dt\{1+O(k_{h+1}^2)\}$$
$$=\frac{1}{2\sqrt{3}}\left\{\int_{x_h}^{x_{h+1}}\sqrt{f(t)}dt\right\}^2.$$

Assuming stratification into L number of equal intervals under AOSB, we can get

$$A=\sum_{h=1}^{L}W_h\left(\sigma_{hx}+\mu_{h\psi}\right)$$
$$=\int_u^v\psi(t)f(t)dt+\frac{1}{2\sqrt{3}L}\left\{\int_u^v\sqrt{f(t)}dt\right\}^2.$$

Similarly from (15), we get,

$$\sigma_{(h+1)\psi}=\frac{\psi'}{2\sqrt{3}}k_{h+1}+\frac{\psi''}{4\sqrt{3}}k_{h+1}^2$$
$$+\frac{\left\{\begin{matrix}2ff''\psi'^2+24ff'\psi''+18f^2\psi'\psi'''\\+f^2\psi''^2-5f'^2\psi'^2-20ff'\psi\psi''\end{matrix}\right\}}{240\sqrt{3}f^2\psi'}k_{h+1}^3+O(k_{h+1}^4).$$

Then, multiplying the above by (25) and using the identity (21), we proceed as follows:

$$W_{h+1}\sigma_{(h+1)\psi}=\frac{1}{2\sqrt{3}}k_{h+1}\int_{x_h}^{x_{h+1}}g_1(t)f(t)dt$$
$$\{1+O(k_{h+1}^2)\}.$$
$$\Rightarrow W_{h+1}\sigma_{(h+1)\psi}=\frac{1}{2\sqrt{3}}\left\{\int_{x_h}^{x_{h+1}}\sqrt{g_1(t)f(t)}dt\right\}^2.$$

As before, assuming L number of equal intervals stratification under AOSB, we can get

$$\sum_{h=0}^{L}W_{h+1}\sigma_{(h+1)\psi}=\frac{1}{2\sqrt{3}}\sum_{h=1}^{L}\left\{\int_{x_h}^{x_{h+1}}\sqrt{g_1(t)f(t)}dt\right\}^2.$$
$$\Rightarrow\sum_{h=0}^{L}W_{h+1}\sigma_{(h+1)\psi}=\frac{L}{2\sqrt{3}L^2}\left\{\int_u^v\sqrt{g_1(t)f(t)}dt\right\}^2,$$

Hence we get

$$B=\sum_{h=1}^{L}W_h\sigma_{h\psi}=\frac{1}{2\sqrt{3}L}\left\{\int_u^v\sqrt{g_1(t)f(t)}dt\right\}^2$$

The proof of the Lemma is completed.

4. Empirical Illustrations of the Proposed Methods of Stratification and Discussion on Their Results

In the empirical illustration, as was done by Singh and Sukhatme [21], Gupt and Ahamed [2,3], Gupt et al. [37] etc., the data generated by the following probability density functions (pdf) in the given ranges are used.

a) Uniform distribution: $f(x)=1,\ 1\le x\le 2$

b) Right triangular distribution: $f(x)=2(2-x),\ 1\le x\le 2$

c) Exponential distribution: $f(x)=e^{-x+1},\ 1\le x<\infty$

d) Chi-square distribution: $f(x)=\frac{1}{2}e^{\frac{-x+1}{2}},\ \ 1\le x<\infty$

We calculate OSB by using the proposed equations (13) and, in the calculation of AOSB, among the proposed two equivalent methods of approximation (22) and (23), we conveniently use (23). Population generated by each of the pdfs for every assumed level of heteroscedasticity – g =1, g =1.5 and g =2 – is divided into number of strata L = 2, 3, 4, 5, 6. For any number of strata while using the proposed methods of stratification in stratifying the populations to get OSB and AOSB, successive iterations are executed till optimum points converge and then the resulting sampling variances are calculated. At the same time, for each population, equal interval stratification is done for each considered number of strata and corresponding sampling variances are calculated. Efficiency of each method of stratification is found by comparing its variance at optimum points with respect to that of equal interval stratification in each population for each considered number of strata. The comparisons are shown in Tables (1-12). In the tables, equal interval stratification is denoted by the abbreviation EIS and relative efficiencies by RE.

In the generation of populations, $\xi(y|x)=\alpha+\beta x$ is taken to be linear with slope at 45^o and $\alpha=0$ is also assumed. The constant σ^2 in the conditional variance $V(y|x)=\sigma^2x^g$ is determined, given g =1, g =1.5 and g =2, so that 90% of the total variation is accounted for by the regression. We truncate right triangular distribution such that area under the curve to the right of the truncation point is 0.05. In the case of exponential and chi-square distributions, truncation is done in such a way that area under the curve to the left as well as right is 0.05. Numerical differentiation and integration methods are used in solving (23).

Table 1. Population generated by Uniform Distribution, g=1

No. of strata (L)	OSB and $nV(\bar{y}_{st})$ due to equations (13)		Boundaries and $nV(\bar{y}_{st})$ due to EIS		RE	AOSB and $nV(\bar{y}_{st})$ due to method (23)		RE
	OSB	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		AOSB	$nV(\bar{y}_{st})$	
2	1.5236	0.0280	1.5	0.0280	100	1.4996	0.0279	100.36
3	1.3234, 1.7095	0.0168	1.334, 1.667	0.0169	100.59	1.3316, 1.6634	0.0166	101.81
4	1.2856, 1.5240, 1.7316	0.0086	1.25, 1.50, 1.75	0.0088	102.33	1.2489, 1.4981, 1.7473	0.0089	98.88
5	1.2144, 1.3721, 1.5344, 1.7321	0.0128	1.20, 1.40, 1.60, 1.80	0.0131	102.34	1.1993, 1.3987, 1.5982, 1.7977	0.0126	103.97
6	1.1698, 1.3041, 1.5217, 1.7249, 1.8728	0.0114	1.1667, 1.334, 1.499, 1.6667, 1.8334	0.0112	98.25	1.1668, 1.3338, 1.5807, 1.6677, 1.8346	0.0114	98.25

Table 2. Population generated by Uniform Distribution, g=1.5

No. of strata (L)	OSB and $nV(\bar{y}_{st})$ due to equations (13)		Boundaries and $nV(\bar{y}_{st})$ due to EIS		RE	AOSB and $nV(\bar{y}_{st})$ due to method (23)		RE
	OSB	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		AOSB	$nV(\bar{y}_{st})$	
2	1.5169	0.0279	1.5	0.0277	99.28	1.4995	0.0277	100
3	1.3228, 1.7123	0.0170	1.334, 1.667	0.0168	98.82	1.3299, 1.6603	0.0165	101.82
4	1.2862, 1.5199, 1.7327	0.0085	1.25, 1.50, 1.75	0.0086	101.18	1.2480, 1.4962, 1.7446	0.0083	103.61
5	1.2139, 1.3699, 1.5354, 1.7347	0.0126	1.20, 1.40, 1.60, 1.80	0.0129	102.38	1.1987, 1.3975, 1.5964, 1.7953	0.0125	103.20
6	1.1698, 1.3041, 1.5216, 1.7257, 1.8767	0.0104	1.1667, 1.334, 1.499, 1.6667, 1.8334	0.0110	105.77	1.1664, 1.3329, 1.4995, 1.6660, 1.8326	0.0112	98.21

Table 3. Population generated by Uniform Distribution, g=2

No. of strata (L)	OSB and $nV(\bar{y}_{st})$ due to equations (13)		Boundaries and $nV(\bar{y}_{st})$ due to EIS		RE	AOSB and $nV(\bar{y}_{st})$ due to method (23)		RE
	OSB	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		AOSB	$nV(\bar{y}_{st})$	
2	1.5149	0.0277	1.5	0.0276	99.64	1.5000	0.0276	100
3	1.3285, 1.7372	0.0169	1.334, 1.667	0.0166	98.22	1.3286, 1.6573	0.0163	101.84
4	1.2609, 1.5031, 1.7438	0.0083	1.25, 1.50, 1.75	0.0085	102.41	1.2473, 1.4946, 1.7418	0.0081	104.94
5	1.2138, 1.3689, 1.5332, 1.7351	0.0124	1.20, 1.40, 1.60, 1.80	0.0128	103.23	1.1982, 1.3965, 1.5947, 1.7929	0.0122	104.92
6	1.1701, 1.3047, 1.5266, 1.7270, 1.8766	0.0094	1.1667, 1.334, 1.500, 1.6667, 1.8334	0.0108	114.89	1.1662, 1.3324, 1.4986, 1.6648, 1.8309	0.0111	97.30

For uniform populations, it is logical to consider that equal interval stratification is efficient stratification method. The proposed methods of stratification (13) and (23) are stratifying the populations of all the three considered levels of heteroscedasticity for each number of strata, mostly, with almost same efficiencies or slightly higher efficiencies than that of equal interval stratification. Therefore, the proposed stratification methods can be considered efficient stratification methods.

Table 4. Population generated by Right Triangular Distribution, g=1

No. of strata (L)	OSB and $nV(\bar{y}_{st})$ due to equations (13)		Boundaries and $nV(\bar{y}_{st})$ due to EIS		RE	AOSB and $nV(\bar{y}_{st})$ due to method (23)		RE
	OSB	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		AOSB	$nV(\bar{y}_{st})$	
2	1.3674	0.0226	1.5	0.0278	123.01	1.3631	0.0226	123.01
3	1.2726, 1.6217	0.0125	1.334, 1.667	0.0146	116.80	1.2323, 1.5082	0.0127	114.96
4	1.2345, 1.4318, 1.6381	0.0076	1.25, 1.50, 1.75	0.0102	134.21	1.1711, 1.3627, 1.5898	0.0077	132.47
5	1.1804, 1.3812, 1.5800, 1.7367	0.0106	1.20, 1.40, 1.60, 1.80	0.0126	118.87	1.1358, 1.2832, 1.4479, 1.6431	0.0087	144.83
6	1.1488, 1.2964, 1.4593, 1.6395 1.8336	0.0664	1.1667, 1.334, 1.500, 1.6667, 1.8334	0.0716	107.83	1.1125, 1.2324, 1.3628, 1.5083, 1.6805	0.0587	121.98

Table 5. Population generated by Right Triangular Distribution, g=1.5

No. of strata (L)	OSB and $nV(\bar{y}_{st})$ due to equations (13)		Boundaries and $nV(\bar{y}_{st})$ due to EIS		RE	AOSB and $nV(\bar{y}_{st})$ due to method (23)		RE
	OSB	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		AOSB	$nV(\bar{y}_{st})$	
2	1.3729	0.0225	1.5	0.0277	123.11	1.3630	0.0225	123.11
3	1.2734, 1.6064	0.0109	1.334, 1.667	0.0146	133.94	1.2321, 1.5069	0.0134	108.96
4	1.2386, 1.4339, 1.6384	0.0089	1.25, 1.50, 1.75	0.0134	150.56	1.1711, 1.3627, 1.5898	0.0076	132.89
5	1.1834, 1.3987, 1.5905, 1.7551	0.0106	1.20, 1.40, 1.60, 1.80	0.0126	118.87	1.1358, 1.2831, 1.4479, 1.6430	0.0087	144.83
6	1.1491, 1.2970, 1.4594, 1.6426, 1.8337	0.0697	1.1667, 1.334, 1.500, 1.6667, 1.8334	0.0750	107.60	1.1124, 1.2324, 1.3628, 1.5083, 1.6804	0.0686	109.33

Table 6. Population generated by Right Triangular Distribution, g=2

No. of strata (L)	OSB and $nV(\bar{y}_{st})$ due to equations (13)		Boundaries and $nV(\bar{y}_{st})$ due to EIS		RE	AOSB and $nV(\bar{y}_{st})$ due to method (23)		RE
	OSB	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		AOSB	$nV(\bar{y}_{st})$	
2	1.3732	0.0224	1.5	0.0275	122.77	1.3634	0.0224	122.77
3	1.2787, 1.5979	0.0131	1.334, 1.667	0.0145	110.67	1.2318, 1.5047	0.0132	109.84
4	1.2377, 1.4347, 1.6426	0.0072	1.25, 1.50, 1.75	0.0099	137.50	1.1715, 1.3630, 1.5902	0.0075	132.00
5	1.1839, 1.3999, 1.5904, 1.7541	0.0082	1.20, 1.40, 1.60, 1.80	0.0125	152.44	1.1358, 1.2833, 1.4480, 1.6430	0.0085	147.06
6	1.1493, 1.2983, 1.4596, 1.6453, 1.8337	0.0722	1.1667, 1.334, 1.500, 1.6667, 1.8334	0.0776	107.48	1.1125, 1.2325, 1.3629, 1.5084, 1.6805	0.0724	107.18

Table 7. Population generated by Exponential Distribution, g=1

No. of strata (L)	OSB and $nV(\bar{y}_{st})$ due to equations (13)		Boundaries and $nV(\bar{y}_{st})$ due to EIS		RE	AOSB and $nV(\bar{y}_{st})$ due to method (23)		RE
	OSB	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		AOSB	$nV(\bar{y}_{st})$	
2	2.3327	0.1736	2.50	0.2021	116.42	1.9732	0.1895	106.65
3	1.7864, 2.9432	0.1194	2.0, 3.0	0.1271	106.45	1.5930, 2.4435	0.1056	120.36
4	1.7027, 2.4636, 3.2728	0.0791	1.75, 2.50, 3.25	0.0852	107.71	1.4273, 1.9726, 2.7273	0.0629	135.45
5	1.4051, 1.9235, 2.5231, 3.2728	0.0761	1.6, 2.2, 2.8, 3.4	0.0984	129.30	1.3342, 1.7363, 2.2407, 2.9192	0.0707	139.18
6	1.3818, 1.7889, 2.3125, 2.7695, 3.2776	0.0616	1.5, 2.0, 2.3, 2.8, 3.5	0.0761	123.54	1.2745, 1.5930, 1.9725, 2.4425, 3.0580	0.0608	125.16

In the skewed populations generated by Right Triangular Distribution, it is seen that the proposed methods of stratification (13) and (23) are performing with much higher efficiencies than that of equal interval stratification for all the number of strata, although when number of strata is 6, the efficiency is relatively less compared to efficiencies in lower number of strata in each of the three populations. Therefore, the proposed methods perform efficiently in stratifying populations of all the considered levels of heteroscedasticity. Both the proposed methods work with more or less same efficiencies.

Table 8. Population generated by Exponential Distribution, g=1.5

No. of strata (L)	OSB and $nV(\bar{y}_{st})$ due to equations (13)		Boundaries and $nV(\bar{y}_{st})$ due to EIS		RE	AOSB and $nV(\bar{y}_{st})$ due to method (23)		RE
	OSB	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		AOSB	$nV(\bar{y}_{st})$	
2	2.0053	0.1697	2.50	0.1996	117.62	1.9715	0.1704	117.14
3	1.7140, 2.5015	0.1039	2.0, 3.0	0.1248	120.12	1.5925, 2.4426	0.1041	119.88
4	1.7059, 2.4699, 3.2727	0.0749	1.75, 2.50, 3.25	0.0821	109.61	1.4270, 1.9723, 2.7265	0.0762	107.74
5	1.4060, 1.9269, 2.5241, 3.2728	0.0741	1.6, 2.2, 2.8, 3.4	0.0964	130.09	1.3339, 1.7357, 2.2401, 2.9182	0.0645	149.46
6	1.3841, 1.7893, 2.3121, 2.7694, 3.2776	0.0695	1.5, 2.0, 2.30, 2.8, 3.5	0.0816	117.41	1.2743, 1.5927, 1.9722, 2.4420, 3.0576	0.0540	151.11

Table 9. Population generated by Exponential Distribution, g=2

No. of strata (L)	OSB and $nV(\bar{y}_{st})$ due to equations (13)		Boundaries and $nV(\bar{y}_{st})$ due to EIS		RE	AOSB and $nV(\bar{y}_{st})$ due to method (23)		RE
	OSB	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		AOSB	$nV(\bar{y}_{st})$	
2	2.0026	0.1672	2.50	0.1976	118.18	1.9765	0.1778	111.14
3	1.7152, 2.6369	0.1009	2.0, 3.0	0.1227	121.61	1.5946, 2.4456	0.1065	115.21
4	1.6672, 2.3493, 3.2672	0.0548	1.75, 2.50, 3.25	0.0788	143.80	1.4283, 1.9744, 2.7287	0.0586	134.47
5	1.3927, 1.8300, 2.4927, 3.2728	0.0707	1.6, 2.2, 2.8, 3.4	0.0941	133.10	1.3348, 1.7372, 2.2423, 2.9202	0.0655	143.66
6	1.3923, 1.7842, 2.3169, 2.7695, 3.2886	0.0671	1.5, 2.0, 2.30, 2.8, 3.5	0.0786	117.14	1.2749, 1.5937, 1.9735, 2.4430, 3.0584	0.0666	118.02

In the case of populations generated by exponential distribution, the proposed methods of stratification (13) and (23) perform with much higher efficiencies than that of equal interval stratification. For g =1, except for number of strata 2, method of approximation (23) performs with slightly higher efficiency than equations (13) giving OSB. For g =1.5 and g=2, in the numbers of strata 2, 3, 4, equations (13) perform slightly better than (23), whereas in numbers of strata 5 and 6, method (23) perform slightly better than (13).

Table 10. Population generated by Chi-square Distribution, g=1

No. of strata (L)	OSB and $nV(\bar{y}_{st})$ due to equations (13)		Boundaries and $nV(\bar{y}_{st})$ due to EIS		RE	AOSB and $nV(\bar{y}_{st})$ due to method (23)		RE
	OSB	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		AOSB	$nV(\bar{y}_{st})$	
2	2.7181	0.6929	4.0	0.8326	120.16	2.9562	0.6917	120.37
3	2.4378, 4.3118	0.4361	3.0, 5.0	0.5380	123.37	2.1928, 3.9111	0.4043	133.07
4	2.1163, 3.5872, 4.9513	0.3367	2.5, 4.0, 5.5	0.3478	103.30	1.8568, 2.9561, 4.4845	0.3211	108.31
5	2.0345, 3.1770, 4.3164, 5.1908	0.2691	2.2, 3.4, 4.6, 5.8	0.3186	118.39	1.6729, 2.4844, 3.5055, 4.8818	0.2610	122.07
6	1.7398, 2.6599, 3.6814, 4.9554, 5.7889	0.2621	2.0, 3.0, 4.0, 5.0, 5.68	0.3115	118.85	1.5528, 2.1957, 2.9630, 3.9144, 5.1663	0.2637	118.13

Table 11. Population generated by Chi-square Distribution, g=1.5

No. of strata (L)	OSB and $nV(\bar{y}_{st})$ due to equations (13)		Boundaries and $nV(\bar{y}_{st})$ due to EIS		RE	AOSB and $nV(\bar{y}_{st})$ due to method (23)		RE
	OSB	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		AOSB	$nV(\bar{y}_{st})$	
2	2.7531	0.6772	4.0	0.8113	119.80	2.9562	0.6917	117.29
3	2.3719, 4.2975	0.4081	3.0, 5.0	0.5162	126.49	2.1928, 3.9111	0.4043	127.68
4	2.1602, 3.6133, 4.9461	0.1854	2.5, 4.0, 5.5	0.3276	176.70	1.8568, 2.9561, 4.4845	0.2311	141.76
5	2.0493, 3.1797, 4.3162, 5.1969	0.1670	2.2, 3.4, 4.6, 5.8	0.3011	180.30	1.6729, 2.4845, 3.5055, 4.8818	0.2010	149.80
6	1.7527, 2.6720, 3.6789, 4.9541, 5.9814	0.2585	2.0, 3.0, 4.0, 5.0, 5.68	0.2935	113.54	1.5528, 2.1957, 2.9630, 3.9144, 5.1663	0.2638	111.26

Table 12. Population generated by Chi-square Distribution, g=2

No. of strata (L)	OSB and $nV(\bar{y}_{st})$ due to equations (13)		Boundaries and $nV(\bar{y}_{st})$ due to EIS		RE	AOSB and $nV(\bar{y}_{st})$ due to method (23)		RE
	OSB	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		AOSB	$nV(\bar{y}_{st})$	
2	3.0298	0.6705	4.0	0.7949	118.55	2.9524	0.6634	119.82
3	2.1514, 3.7036	0.3487	3.0, 5.0	0.4957	142.16	2.1908, 3.9074	0.3793	130.69
4	2.1823, 3.6686, 4.9505	0.2140	2.5, 4.0, 5.5	0.3088	144.30	1.8588, 2.9589, 4.4871	0.2149	143.69
5	1.7472, 2.6881, 3.5730, 4.9283	0.2344	2.2, 3.4, 4.6, 5.8	0.2831	120.78	1.6721, 2.4829, 3.5034, 4.8799	0.2443	115.88
6	1.7907, 2.6773, 3.5923, 4.3549, 5.1565	0.2516	2.0, 3.0, 4.0, 5.0, 5.68	0.2739	108.86	1.5522, 2.1945, 2.9615, 3.9126, 5.1647	0.2463	111.21

In the populations generated by chi-square distribution too, the proposed methods stratify populations with much higher efficiencies than equal interval stratification. In the heteroscedastic population for g =1, method of approximation (23) perform slightly better than equations (13). For g=1.5 and g=2, in all considered number of strata, except strata no., 6 for g =2, equations (13) perform slightly better than method of approximation (23).

5. Conclusions

It has been demonstrated in this paper that the proposed methods of stratification for GAVOA are found to stratify heteroscedastic populations efficiently. Both the methods - equations (13) giving OSB and method of approximation (23) giving AOSB - are not only obtained analytically but also found to stratify populations of varied nature, symmetrical or moderately skewed or highly skewed with different levels of heteroscedasticity, with higher efficiencies. Although equations (13) are found arduous in practical applications, the methods of giving AOSB (22) and (23) are easy to use. It is also demonstrated that equations (13) giving OSB and method of giving AOSB (23) work with more or less same efficiencies in a number of populations. The methods of approximations (22) and (23) are proven to be equivalent. Therefore, any of the two proposed methods of stratification can be easily used in stratified sampling.

REFERENCES

[1] T. Dalenius. The problem of optimum stratification-I. Skandinavisk Aktuaritidskrift, Vol. 33, 203-213, 1950.

[2] B.K. Gupt, M.I. Ahamed. Optimum stratification for a generalized auxiliary variable proportional allocation under a superpopulation model. Communications in Statistics-Theory and Methods, Published online: 25 July 2020. https://doi.org/10.1080/03610926.2020.1793203

[3] B.K. Gupt, M.I. Ahamed. Construction of strata for a model-based allocation under a superpopulation model, Journal of Statistical Theory and Applications, Vol. 20, No. 1, 46-60, 2021.

[4] B.K. Gupt. Sample size allocation for stratified sampling under a correlated superpopulation model. METRON-International Journal of Statistics, Vol. LXI, No. 1, 35-52, 2003.

[5] B.K. Gupt. Allocation of Sample Size in Stratified Sampling Under Superpopulation Models. Saarbrucken, Deutschland/Germany: LAP LAMBERT Academic Publishing AV Akademikerverlag GmbH & Co.KG, 2012.

[6] A.A. Tschuprow. On mathematical expectation of the moments of frequency distributions in the case of correlated observations. Metron, Vol.2, 461-493, 1923.

[7] J. Neyman. On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society, Vol. 97, No.4, 558-625, 1934.

[8] W.G. Cochran. Relative accuracy of systematic and stratified random samples for a certain class of populations. Annals of Mathematical Statistics, Vol. 17, 164-177, 1946.

[9] T.V. Hanurav. Optimum Sampling Strategies and Some Related Problems, unpublished Ph.D thesis submitted to the Indian Statistical Institute, 1965.

[10] T.J. Rao. On the allocation of sample size in stratified sampling. Ann. Inst .Stat. Math, Vol. 20, 159-166, 1968.

[11] B.K. Gupt, T.J. Rao. Stratified PPS sampling and allocation of sample size. Jour. Ind. Soc. Ag. Statistics, Vol. 50, No. 2, 199-208, 1997.

[12] T. Dalenius, M. Gurney. The problem of optimum stratification-II. Skandinavisk Aktuaritidskrift, Vol. 34, 133-148, 1951.

[13] P.C. Mahalanobis. Some aspects of the design of sample surveys, Sankhya, Vol. 12, 1-7, 1952.

[14] M.H. Hansen, W.N. Hurwitz, W.G. Madow. Sample survey methods and theory, Vol. I and II, New York: John Wiley and Sons, Inc, 1953.

[15] T. Dalenius, J.L. Hodges. The Choice of stratification points, Skandinavisk Aktuaritidskrift, Vol. 34, 198-203, 1957.

[16] T. Dalenius, J.L. Hodges. Minimum variance stratification. Jour. Amer. Stat. Assoc., Vol. 54, 88-101, 1959.

[17] G. Ekman. An approximation useful in univariate stratification. The Annals of Mathematical Statistics, Vol. 30, 219-229, 1959.

[18] V.K. Sethi. A note on optimum stratification of populations for estimating the population means, Aust. J. Stat., Vol.5, 20-23, 1963.

[19] T. Dalenius. Sampling in Sweden. Contributions to the methods and theories of sample survey practice. Almqvist and Wicksell, Stockholm, 1957.

[20] Y. Taga. On optimum stratification for the objective variable based on concomittant variable. Annals of the Institute of Statistical Mathematics, Vol. 19, 101-130, 1967.

[21] R. Singh, B.V. Sukhatme. Optimum stratification. Annals of the Institute of Statistical Mathematics, Vol. 21, 515-528, 1969.

[22] R. Singh. Approximately optimum stratification on the auxiliary variable. Journal of the American Statistical Association, Vol. 66, 829-833, 1971.

[23] R. Singh, B.V. Sukhatme. Optimum stratification in sampling with varying Probabilities. Annals of the Institute of Statistical Mathematics, Vol. 24, 485-494, 1972.

[24] R. Singh. An alternate method of stratification on the auxiliary variable. Sankhya, Vol. 37, 100-108, 1975.

[25] R. Singh. On optimum stratification for proportional allocation. Sankhya, Vol. 37, Pt. I, 109-115, 1975.

[26] R. Singh. A note on optimum stratification in sampling with varying probabilities. Australian Journal of Statistics, Vol. 27, No. 1, 12–21, 1975.

[27] R. Singh, D. Prakash. Optimum stratification for equal allocation. Annals of the Institute of Statistical Mathematics, Vol. 27, 273-280, 1975.

[28] S.S. Yadava, R. Singh. Optimum stratification for allocation proportional to strata totals for simple random sampling scheme. Communications in Statistics: Theory and Methods, Vol. 13, No. 22, 2793-2806, 1984.

[29] S.P. Gosh. Optimum stratification with two characters, Ann. Maths. Statist., Vol. 34, 866-872, 1963.

[30] G. Sadasivan, R. Aggarwal. Optimum points of stratification in bi-variate populations, Sankhya, C40, 84-97, 1978.

[31] P.C. Gupta, G.R. Seth. On stratification in sampling investigation involving more than one character. J. Ind. Soc. Agril. Statist., Vol. 31, No. 2, 1-15, 1979.

[32] S.E.H. Rizvi, J.P. Gupta, R. Singh. Approximately optimum stratification for two study variables using auxiliary information, J. Ind. Soc. Agril. Statist., Vol. 53, No. 3, 287-298, 2000.

[33] S.E.H. Rizvi, J.P. Gupta, M. Bhargava. Optimum stratification based on auxiliary variable for compromise allocation, Metron, LX (3-4), 201-215, 2002.

[34] S.E.H. Rizvi, J.P. Gupta, M. Bhargava. Effect of optimum stratification on sampling with varying probabilities under proportional allocation. Statistica, Vol. 64, No. 4, 721-733, 2004.

[35] M.R. Verma. A method of optimum stratification for two study variables, J. Ind. Soc. Agril. Statist., Vol. 62, No. 3, 266-275, 2008.

[36] S. Rabee, R. Hamed, R. Kassem, M. Rashwaan, "A Goal Programming Approach for Multivariate Calibration Weights Estimation in Stratified Random Sampling," Mathematics and Statistics, Vol. 9, No. 3, pp. 326 - 334, 2021. DOI: 10.13189/ms.2021.090314.

[37] B.K. Gupt, M.I. Ahamed, M. Phukon. Optimum stratification for an auxiliary variable optimum allocation under a superpopulation model, Advances and Applications in Statistics, Vol. 67, No.1, 1-20, 2021.

[38] G. Ekman. Approximate expressions for the conditional mean and variance over small intervals of a continuous distribution. The Annals of Mathematical Statistics, Vol. 30, No. 4, 1131-1134, 1959.