Statistics and Applications {ISSN 2454-7395 (online)}
Volume 20, No. 1, 2022 (New Series), pp 181-192

Methods of Stratification for Allocation Proportional to Stratum Total Under a Superpopulation Model

Bhuwaneshwar Kumar Gupt[1] **and Md. Irphan Ahamed**[2]
[1]*Department of Statistics, North-Eastern Hill University, Shillong, India*
[2]*Department of Mathematics, Umshyrpi College, Shillong, India*

Received: 08 September 2020; Revised: 23 March 2021; Accepted: 28 March 2021

Abstract

In this paper, the problem of optimum stratification in stratified sampling is considered on a concomitant variable *x* which is highly correlated with the estimation variable *y*, in the light of a-priori distributions, for the allocation proportional to stratum total of the auxiliary variable *x*. Unlike earlier techniques and approaces, available in literature, used in obtaining methods of stratification for the said allocation, the problem is dealt with in a different way and hence a distinct set of equations giving optimum points of stratification and a few distinct methods for finding approximately optimum points of stratification have been obtained. All these proposed methods of stratification are found efficient as well as easy to use when examined empirically by illustrating them in several generated data.

Keywords: Allocation; Auxiliary variable; Optimum points of stratification; Probability density functions; Simple random sampling with and without replacement.

AMS Subject Classification: 62D05, 62-08

1. Introduction

In the field of construction of strata in stratified sampling, it was Dalenius (1950) who first considered the problem of optimum stratification based on estimation variable for Tschuprow (1923) and Neyman (1934) optimum allocation. Since then, the work in this area has been being progressively extended by several researchers into various ways, dimensions and results. To mention a few pertinent ones, among many, are Dalenius and Gurney (1951), Dalenius and Hodges (1959), Taga (1967), Singh and Sukhatme (1969, 1972), Singh and Prakash (1975), Singh (1971, 1975 a, b, c) *etc*., who considered the problem of construction of strata for various allocations with different methods of sampling.

Rao (1968) obtained allocation proportional to stratum total of auxiliary variable which is highly correlated with estimation variable, auxiliary variable proportional allocation (AVPA), under the following superpopulation model:

$$\begin{aligned} &(i)\ \ \xi\left(y_i \middle| x_i\right) = \alpha + \beta x_i \\ &(ii)\ V\left(y_i \middle| x_i\right) = \sigma^2 x_i^2 \\ &(iii)\ \varsigma\left(y_i, y_j \middle| x_i, x_j\right) = 0 \end{aligned} \tag{1}$$

Corresponding Author: B.K. Gupt
Email: bhuwaneshwargupt@gmail.com

Yadava and Singh (1984) considered the problem of optimum stratification under simple random sampling scheme for the abovementioned allocation – AVPA - proposed by Rao (1968). They obtained equations giving optimum points of stratification (OPS) and a few methods of finding approximately optimum points of stratification (AOPS). Gupt and Ahamed (2020) too considered problem of optimum stratification for a generalized AVPA obtained by Gupt (2003, 2012), *i.e.*,

$$n_h \alpha Z_h \tag{2}$$

where $Z_h = \sum X_{hj}^{g/2}$, $j = 1, 2, ..., N_h$, provided $\eta_h(g) = \frac{\sigma_h(x)}{Z_h}$ are equal in all strata h=1, 2, …, L under the model (1) when $V(y_i|x_i)$ is proportional to x_i^g and g is the level of hetero-scedasticity. A particular case of the methods of stratification obtained by the authors when g =2 gives the methods of stratification for the allocation AVPA which were also obtained by Yadava and Singh (1984).

In this paper too, the problem of optimum stratification for allocation AVPA in the light of the priori distributions (1) is considered, but a different technique and procedure is used. We obtain equations giving OPS and a few methods for finding AOPS which are quite different from the ones obtained by Yadava and Singh (1984), *i.e.*, the particular case of the methods obtained by Gupt and Ahamed (2020). The methods of approximation obtained in this paper are suitable for practical applications. These proposed methods will hold good for stratified simple random sampling without replacement also when finite population correction is ignored in each stratum.

In section 2 of this paper, the equations giving OPS and a few methods giving AOPS are obtained. In section 3, numerical illustrations of all the proposed methods in this paper by using generated data and comparison of the methods with respect to the methods proposed by Gupt and Ahamed (2020) for particular case g=2 are carried out. In section 4, conclusion is given.

2. Equations Giving OPS and Methods of Finding AOPS

The allocation taken is

$$n_h \propto X_h , \tag{3}$$

where, X_h is total of h^{th} stratum. The allocation (3) can be written as

$$\eta_h = n \frac{W_h \overline{X}_h}{\overline{X}} , \tag{4}$$

where W_h is the proportion of population units in the h^{th} stratum, $\overline{X}_h$ is the mean for x in h^{th} stratum and $\overline{X}$ is the population mean. Using (4) in the sampling variance for stratified sampling in simple random sampling with replacement (SRSWR), *i.e.*, $V(\overline{y}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{\sigma_{hy}^2}{n_h}$, we get,

$$V(\overline{y}_{st}) = \frac{\overline{X}}{n} \sum_{h=1}^{L} W_h \frac{\sigma_{hy}^2}{\overline{X}_h} \tag{5}$$

where, L is total number of strata.

Considering the superpopulation model (1), Gupt (2003, 2012) obtained conditional expectation of σ_{hy}^2 given x as

$$E\left(\sigma_{hy}^2 \mid \underline{\mathbf{X}}_h\right) = \beta^2\sigma_{hx}^2 + \sigma^2\,\frac{N_h - 1}{N_h}\left(\frac{1}{N_h}\sum_{i=1}^{N_h} X_{hi}^2\right), \text{ where } \underline{\mathbf{X}}_h' = \left(\mathrm{X}_{h1}\ \mathrm{X}_{h2}\ \ldots\ \mathrm{X}_{hN_h}\right).$$

For strata of large sizes, we can consider $\frac{N_h - 1}{N_h} \approx 1$, and therefore, we obtain the following expression

$$E\left(\sigma_{hy}^2 \mid \underline{\mathbf{X}}_h\right) = \left(\beta^2 + \sigma^2\right)\sigma_{hx}^2 + \sigma^2\,\overline{X}_h^2$$

$$\Rightarrow E\left(V\left(\bar{y}_{st} \mid \widehat{\mathbf{X}}\right)\right) = \frac{\overline{X}}{n}\sum_h \frac{W_h}{\overline{X}_h}\left\{\left(\beta^2 + \sigma^2\right)\sigma_{hx}^2 + \sigma^2\,\overline{X}_h^2\right\}, \tag{6}$$

where $\widehat{\mathbf{X}} = (\underline{\mathbf{X}}_1, \underline{\mathbf{X}}_2, \ldots, \underline{\mathbf{X}}_L)$.

The conditional expectation of variance in (6) is partially differentiated with respect to x_h (h=1, 2, …, L-1), and the derivative is equated to zero to obtain equations that minimize the expected conditional variance. Thus, we proceed as follows:

$$\frac{\delta}{\delta x_h} E\left(V(\bar{y}_{st}) \mid \widehat{\mathbf{X}}\right) = 0$$

$$\Rightarrow \frac{\delta}{\delta x_h}\left[\frac{W_h}{\overline{X}_h}\left\{\left(\beta^2 + \sigma^2\right)\sigma_{hx}^2 + \sigma^2\,\overline{X}_h^2\right\}\right] + \frac{\delta}{\delta x_h}\left[\frac{W_{h+1}}{\overline{X}_{h+1}}\left\{\left(\beta^2 + \sigma^2\right)\sigma_{(h+1)x}^2 + \sigma^2\,\overline{X}_{h+1}^2\right\}\right] = 0 \tag{7}$$

Taking the first term, we get

$$\frac{\delta}{\delta x_h}\left[\frac{W_h}{\overline{X}_h}\left\{\left(\beta^2 + \sigma^2\right)\sigma_{hx}^2 + \sigma^2\,\overline{X}_h^2\right\}\right] = \left(\beta^2 + \sigma^2\right)\frac{\delta}{\delta x_h}\left(\frac{W_h\sigma_{hx}^2}{\overline{X}_h}\right) + \sigma^2\,\frac{\delta}{\delta x_h}\left(W_h\overline{X}_h\right) \tag{8}$$

and considering the expression

$$\frac{\delta}{\delta x_h}\left(\frac{W_h\sigma_{hx}^2}{\overline{X}_h}\right) = \frac{\overline{X}_h\,\frac{\delta}{\delta x_h}\left(W_h\sigma_{hx}^2\right) - W_h\sigma_{hx}^2\,\frac{\delta\overline{X}_h}{\delta x_h}}{\overline{X}_h^2} \tag{9}$$

If we consider *f(x)* is the probability density function for stratification variable x, we have

$$W_h = \int_{x_{h-1}}^{x_h} f(t)\,dt$$

$$W_h\,\frac{\delta\overline{X}_h}{\delta x_h} = \left(x_h - \overline{X}_h\right) f(x_h) \tag{10}$$

Therefore, we can get

$$\frac{\delta}{\delta x_h}\left(W_h\sigma_{hx}^2\right) = \left(x_h - \overline{X}_h\right)^2 f(x_h) \tag{11}$$

Using (10) and (11) in equation (9), we get

$$\frac{\delta}{\delta x_h}\left(\frac{W_h\sigma_{hx}^2}{\overline{X}_h}\right) = \frac{\left(x_h - \overline{X}_h\right)}{\overline{X}_h}\left[\overline{X}_h\left(x_h - \overline{X}_h\right) - \sigma_{hx}^2\right] f\left(x_h\right) \tag{12}$$

Again using (10) and (12) in equations (8), we get

$$\frac{\delta}{\delta x_h}\left[\frac{W_h}{\overline{X}_h}\left\{\left(\beta^2+\sigma^2\right)\sigma^2_{hx}+\sigma^2\overline{X}^2_h\right\}\right]$$
$$=\left[\left(\beta^2+\sigma^2\right)\frac{\left(x_h-\overline{X}_h\right)}{\overline{X}^2_h}\left\{\overline{X}_h\left(x_h-\overline{X}_h\right)-\sigma^2_{hx}\right\}+\sigma^2 x_h\right]f\left(x_h\right) \tag{13}$$

Similarly,

$$\frac{\delta}{\delta x_h}\left[\frac{W_{h+1}}{\overline{X}_h}\left\{\left(\beta^2+\sigma^2\right)\sigma^2_{(h+1)x}+\sigma^2\overline{X}^2_{h+1}\right\}\right]$$
$$=-\left[\left(\beta^2+\sigma^2\right)\frac{\left(x_h-\overline{X}_{h+1}\right)}{\overline{X}^2_{h+1}}\left\{\overline{X}_{h+1}\left(x_h-\overline{X}_{h+1}\right)-\sigma^2_{(h+1)x}\right\}+\sigma^2 x_h\right]f\left(x_h\right) \tag{14}$$

Substituting (13) and (14) in (7), we get

$$\left(\beta^2+\sigma^2\right)\frac{\left(x_h-\overline{X}_h\right)}{\overline{X}^2_h}\left\{\overline{X}_h\left(x_h-\overline{X}_h\right)-\sigma^2_{hx}\right\}+\sigma^2 x_h$$
$$=\left(\beta^2+\sigma^2\right)\frac{\left(x_h-\overline{X}_{h+1}\right)}{\overline{X}^2_{h+1}}\left\{\overline{X}_{h+1}\left(x_h-\overline{X}_{h+1}\right)-\sigma^2_{(h+1)x}\right\}+\sigma^2 x_h$$

$$\Rightarrow\left(x_h-\overline{X}_h\right)\left\{\frac{x_h}{\overline{X}_h}-(1+C^2_{hx})\right\}=\left(x_h-\overline{X}_{h+1}\right)\left\{\frac{x_h}{\overline{X}_{h+1}}-(1+C^2_{(h+1)\,x})\right\} \tag{15}$$

where, $C_{hx}=\dfrac{\sigma_{hx}}{\overline{X}_h}$ is the coefficient of variation of x variable in the h^{th} stratum.

Thus, we have got equations (15) which will give OPS.

For finding methods of approximation, we neglect square of coefficients of variation C^2_{hx} in (15) as square of coefficients of variation are expected to be very small quantities relatively with unity, then we can get

$$x_h=\sqrt{\overline{X}_h\,\overline{X}_{h+1}} \tag{16}$$

which will give AOPS.

On the other hand, we may consider C^2_{hx} are not negligible in all strata, but they are approximately equal in two successive strata, $C^2_{1x}\approx C^2_{2x}$, $C^2_{2x}\approx C^2_{3x}$, ..., $C^2_{L-1\,x}\approx C^2_{L\,x}$. If each of C^2_{hx} and $C^2_{(h+1)x}$ is approximately replaced by their geometric mean in (15), we can get

$$x_h=\sqrt{1+C_{hx}\,C_{(h+1)x}}\;\sqrt{\overline{X}_h\,\overline{X}_{h+1}} \tag{17}$$

which will also give AOPS.

If we replace C^2_{hx} and $C^2_{(h+1)x}$ by their arithmetic mean such as

$$C^2_{hx}\approx C^2_{(h+1)x}=\frac{C^2_{hx}+C^2_{(h+1)x}}{2}\equiv\overline{C}^2_{h+\frac{1}{2},x}.$$

Then, we can again obtain,

$$x_h=\sqrt{1+\overline{C}^2_{h+\frac{1}{2},x}}\;\sqrt{\overline{X}_h\,\overline{X}_{h+1}} \tag{18}$$

which will also give AOPS.

It is clearly seen that in terms of analytical justification, our methods obtained in this paper do have equally strong ground with that of proposed methods by Yadava and Singh (1984) - special case of the Gupt and Ahamed's (2020) methods. However, the different techniques and procedure used in this paper have yielded different methods and results. It is also observed that our methods are much easier to use and provide more options for practical applications, whereas their methods are complicated and provide less option for practical applications, particularly in the practical applications of their approximation methods. Since their approximation methods of stratification are presented in the form of definite integrals involving known probability density function of the auxiliary variable based on which stratification is to be done, if the stratification variable in population does not follow a known probability density function, their methods cannot be used. Our methods are free from such restriction.

3. Numerical Illustrations

In order to examine the efficiencies of proposed methods - equations (15) giving OPS and methods of approximation (16), (17) and (18) giving AOPS - we use all these methods in the populations generated by the following probability density functions.

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

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

c) Normal distribution, X ~ N (3, 1): $f(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-(x-3)^2}{2}}$, $-\infty<x<\infty$

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

e) Standard Gamma distribution (Erlang distribution): $f(x)=\frac{1}{2}x e^{-x+1}$, $1\le x<\infty$

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

In the case of generated populations by Chi-square, Exponential and Standard Gamma distribution functions, we truncate the distributions such that area under the curve to the right of the truncation point is 0.05 whereas normal distribution is truncated at two points such that each area under the curve to left and right of truncation points is 0.05. We use equations (15) and methods of approximation (16), (17) and (18) in stratifying each of the generated populations into numbers of strata *L*=2, 3, 4, 5, 6 in which OPS and AOPS are found out for each number of strata, and hence sampling variances of stratified sampling at optimum points and approximately optimum points for each number of strata are calculated. OPS and AOPS are found by successive iterations. We also calculate sampling variances of stratified sampling for equal interval stratification in each of the populations for each of the considered numbers of strata. The relative efficiencies of equations (15) and methods of approximation (16), (17) and (18) with respect to equal interval stratification are calculated and shown in Tables 1, 2, 3, 4, 5 and 6.

The regression function $C(x)=\alpha+\beta x$ is taken to be linear with the slope at 45^0. The constant σ^2 in $V(y\mid x)=\sigma^2 x^2$ is determined in each case in such a way that 90% of the total variation is accounted for by the regression. Secondly, we compare the efficiencies of equations (15) and methods of approximations (16), (17) and (18) with the equations giving OPS and methods of approximation proposed by Gupt and Ahamed (2020) for particular case *g*=2, in the three populations of Uniform, Right-triangular and Exponential density functions in which the authors illustrated their proposed methods. These comparisons are given in Tables 7, 8 and 9.

Table 1: Comparison with equal interval stratification, Chi-square distribution

No. of Strata (L)	Equal Interval Stratification		Stratification due to equations (15)			Stratification due to approx. method (16)			Stratification due to approx. method (17)			Stratification due to approx. method (18)		
	Points of stratification	$nV(\overline{y_{st}})$	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency
2	4.00	0.841	2.747	0.646	130	2.661	0.646	130	2.747	0.646	130	2.747	0.646	130
3	3.00 5.00	0.795	2.112 3.759	0.349	228	2.055 3.689	0.350	227	2.111 3.758	0.349	227	2.112 3.758	0.349	227
4	2.50 4.00 5.50	0.359	2.015 3.189 4.622	0.275	131	1.735 2.725 4.300	0.277	129	2.015 3.188 4.623	0.275	131	2.016 3.189 4.623	0.275	131
5	2.20 3.40 4.60 5.80	0.279	2.015 3.172 4.504 5.860	0.262	106	1.735 2.725 4.240 5.809	0.259	108	2.015 3.172 4.504 5.860	0.262	106	2.016 3.172 4.505 5.860	0.262	106
6	2.00 3.00 4.00 5.00 6.00	0.249	1.751 2.613 3.631 4.797 6.177	0.219	114	1.729 2.593 3.622 4.789 6.171	0.219	114	1.751 2.612 3.652 4.848 6.202	0.219	114	1.751 2.613 3.652 4.849 6.207	0.219	114

It is seen that in the population of Chi-square distribution, the proposed equations (15) and methods of approximation are found to be more efficient than that of equal interval stratification for all numbers of strata and relative efficiency is remarkably high for stratum 3. Moreover, methods of approximations (16), (17) and (18) are having same efficiencies with that of the proposed equations (15) in almost all the numbers of strata.

Table 2: Comparison with equal interval stratification, Exponential distribution

No. of Strata (*L*)	Equal Interval Stratification		Stratification due to equations (15)			Stratification due to approx. method (16)			Stratification due to approx. method (17)			Stratification due to approx. method (18)		
	Points of stratification	$nV(\overline{y_{st}})$	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency
2	2.50	0.202	2.054	0.167	121	1.986	0.168	120	2.054	0.167	121	2.054	0.167	121
3	2.00 3.00	0.123	1.689 2.572	0.100	123	1.649 2.534	0.101	122	1.688 2.572	0.100	123	1.689 2.572	0.100	123
4	1.75 2.50 3.25	0.089	1.645 2.317 3.151	0.085	105	1.385 1.935 2.765	0.075	119	1.645 2.317 3.148	0.086	103	1.646 2.317 3.150	0.086	103
5	1.60 2.20 2.80 3.40	0.081	1.376 1.859 2.465 3.215	0.066	123	1.333 1.787 2.355 3.145	0.062	131	1.376 1.859 2.465 3.215	0.066	123	1.376 1.859 2.465 3.215	0.066	123
6	1.50 2.00 2.30 2.80 3.50	0.068	1.324 1.690 2.046 2.527 3.215	0.054	127	1.319 1.688 2.043 2.524 3.213	0.052	130	1.324 1.690 2.046 2.527 3.215	0.053	128	1.324 1.690 2.046 2.527 3.215	0.053	128

In the population of Exponential distribution, all the proposed methods are performing with more efficiencies than that of equal interval stratification for all the numbers of strata. Moreover, the proposed approximation methods are found in most cases to be performing with same efficiencies as or higher than that of proposed equations (15) giving OPS.

Table 3: Comparison with equal interval stratification, Normal distribution

No. of Strata (L)	Equal Interval Stratification		Stratification due to equations (15)			Stratification due to approx. method (16)			Stratification due to approx. method (17)			Stratification due to approx. method (18)		
	Points of stratification	$nV(\overline{y_{st}})$	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency
2	3.00	0.291	2.966	0.291	100	2.928	0.291	100	2.964	0.291	100	2.967	0.291	100
3	2.366 3.633	0.227	2.348 3.492	0.222	102	2.199 3.276	0.197	115	2.346 3.491	0.222	102	2.349 3.442	0.222	102
4	2.05 3.00 3.95	0.139	2.072 2.939 3.783	0.121	115	2.055 2.927 3.759	0.122	114	2.068 2.939 3.783	0.121	115	2.073 2.939 3.783	0.121	115
5	1.86 2.62 3.38 4.14	0.117	2.068 2.848 3.350 3.884	0.111	106	1.996 2.602 3.085 3.793	0.112	104	2.063 2.847 3.350 3.884	0.111	106	2.069 2.848 3.350 3.884	0.111	106
6	1.733 2.366 3.00 3.633 4.266	0.107	1.629 2.239 2.921 3.539 4.095	0.103	104	1.602 2.215 2.917 3.536 4.091	0.0959	111	1.626 2.239 2.921 3.539 4.095	0.103	104	1.629 2.239 2.921 3.539 4.095	0.101	106

In the population of Normal distribution, the proposed methods perform with higher efficiencies than that of equal interval stratification. At the same time, approximation methods are performing almost same as or slightly better than that of equations (15) although approximation method (16) is performing better than approximation methods (17) and (18) for numbers of strata 4, 5 and 6.

Table 4: Comparison with equal interval stratification, Right Triangular distribution

No. of Strata (L)	Equal Interval Stratification		Stratification due to equations (15)			Stratification due to approx. method (16)			Stratification due to approx. method (17)			Stratification due to approx. method (18)		
	Points of stratification	$nV(\overline{y_{st}})$	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency
2	1.50	0.028	1.373	0.022	127	1.355	0.022	127	1.373	0.022	127	1.394	0.022	127
3	1.334 1.667	0.014	1.273 1.587	0.013	108	1.269 1.585	0.013	108	1.272 1.587	0.013	108	1.272 1.587	0.013	108
4	1.25 1.50 1.75	0.012	1.180 1.386 1.629	0.009	133	1.178 1.390 1.628	0.009	133	1.179 1.386 1.631	0.010	120	1.179 1.387 1.629	0.010	120
5	1.20 1.40 1.60 1.80	0.010	1.170 1.370 1.579 1.736	0.010	100	1.178 1.380 1.578 1.736	0.009	111	1.179 1.379 1.579 1.736	0.010	100	1.179 1.379 1.579 1.736	0.010	100
6	1.167 1.334 1.499 1.667 1.833	0.009	1.145 1.289 1.435 1.589 1.736	0.009	100	1.121 1.270 1.430 1.588 1.736	0.008	113	1.145 1.290 1.435 1.589 1.736	0.009	100	1.145 1.289 1.435 1.589 1.736	0.009	100

In the population of Right-triangular distribution, all the proposed methods of stratification are found to be having higher efficiencies than that of equal interval stratification except at numbers of strata 5 and 6 at which the proposed equations (15) and approximation methods (17) and (18) have same efficiencies with that of equal interval stratification. But, proposed method (16) performs better than equal interval stratification.

Table 5: Comparison with equal interval stratification, Standard Gamma distribution

No. of Strata (L)	Equal Interval Stratification		Stratification due to equations (15)			Stratification due to approx. method (16)			Stratification due to approx. method (17)			Stratification due to approx. method (18)		
	Points of stratification	$nV(\overline{y_{st}})$	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency
2	3.00	0.405	2.614	0.373	109	2.539	0.368	110	2.614	0.373	109	2.616	0.373	109
3	2.33 3.66	0.268	1.882 3.044	0.205	131	1.902 3.076	0.211	127	1.882 3.043	0.205	131	1.896 3.051	0.205	131
4	2.00 3.00 4.00	0.169	1.793 2.780 3.986	0.161	105	1.749 2.673 3.796	0.161	105	1.793 2.780 3.984	0.161	105	1.794 2.781 3.986	0.161	105
5	1.80 2.60 3.40 4.20	0.145	1.753 2.534 3.268 4.153	0.143	102	1.695 2.396 3.105 4.056	0.137	106	1.752 2.534 3.268 4.152	0.143	102	1.753 2.534 3.268 4.153	0.143	102
6	1.66 2.32 2.98 3.64 4.30	0.129	1.660 2.221 2.862 3.625 4.327	0.129	100	1.630 2.170 2.833 3.620 4.326	0.128	102	1.644 2.188 2.845 3.624 4.327	0.131	99	1.645 2.188 2.846 3.625 4.327	0.131	99

In the population of Standard Gamma distribution, all the proposed methods of stratification are stratifying the population more efficiently than equal interval stratification except for number of strata 6 at which approximation methods (17) and (18) are having slightly less efficiencies than that of equal interval stratification and the proposed equations (15) perform with same efficiency as that of equal interval stratification. In this population too, the proposed method of approximation (16) is found to be performing with higher efficiency than that of equal interval stratification as well as all other proposed methods.

Table 6: Comparison with equal interval stratification, Uniform distribution

No. of Strata (L)	Equal Interval Stratification		Stratification due to equations (15)			Stratification due to approx. method (16)			Stratification due to approx. method (17)			Stratification due to approx. method (18)		
	Points of stratification	$nV(\overline{y_{st}})$	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency	Points of stratification	$nV(\overline{y_{st}})$	Relative Efficiency
2	1.50	0.028	1.504	0.028	100	1.498	0.028	100	1.504	0.028	100	1.504	0.028	100
3	1.334 1.667	0.016	1.323 1.667	0.016	100	1.318 1.658	0.016	100	1.323 1.666	0.016	100	1.323 1.666	0.016	100
4	1.25 1.50 1.75	0.013	1.252 1.475 1.732	0.013	100	1.248 1.469 1.727	0.0126	101	1.252 1.475 1.732	0.013	100	1.252 1.475 1.732	0.013	100
5	1.20 1.40 1.60 1.80	0.0124	1.212 1.369 1.529 1.743	0.0119	104	1.211 1.368 1.529 1.742	0.0119	104	1.212 1.369 1.529 1.743	0.0119	104	1.212 1.369 1.529 1.743	0.0119	104
6	1.167 1.334 1.499 1.667 1.833	0.0107	1.169 1.318 1.504 1.716 1.873	0.0108	99	1.162 1.303 1.488 1.708 1.873	0.0107	100	1.169 1.318 1.504 1.716 1.873	0.0108	99	1.169 1.318 1.504 1.716 1.873	0.0108	99

For the population of Uniform distribution, equal interval stratification is generally considered to be the best stratification method. We have observed that all the proposed methods of stratification work with same efficiencies as that of equal interval stratification in almost all the considered numbers of strata except at number of strata 6, the proposed equations (15) and methods of approximation (17) and (18) are performing with slightly less efficiencies than that of equal interval stratification. But at stratum number 5, our proposed methods are performing with slightly higher efficiencies than that of equal interval stratification. Hence it is seen that all our proposed methods of stratification are efficient in stratifying the population. The proposed approximation method (16) is found to be most credibly stratifying the population to give OPS.

On the other hand, as Yadava and Singh (1984) did not perform numerical illustration, we compare our proposed equations and methods of approximation with that of Gupt and Ahamed (2020) whose special case gives Yadava and Singh's (1984) methods. The comparisons are shown in Tables 7-9. Gupt and Ahamed (2020) illustrated their proposed methods in three generated populations of Exponential, Right-triangular and Uniform distribution in the same range of x variable as taken in this paper too, and therefore the comparison of our methods with respect to their methods are made in these three populations only.

Table 7: Comparison with the methods proposed by Gupt and Ahamed (2020), Exponential distribution

No. of Strata (L)	Equations for OPS proposed by Gupt and Ahamed(2020)	Proposed equations (15)	Comparison of (15) *w.r.t* equations for OPS proposed by Gupt and Ahamed (2020)	Approximation methods proposed by Gupt and Ahamed (2020)	$nV(\overline{y_{st}})$ for proposed Approximation Methods			Comparison of methods (16), (17) and (18) *w.r.t* approximation method proposed by Gupt and Ahamed (2020)		
	$nV(\overline{y_{st}})$	$nV(\overline{y_{st}})$	Relative efficiency	$nV(\overline{y_{st}})$	Method (16)	Method (17)	Method (18)	Relative efficiency of (16)	Relative efficiency of (17)	Relative efficiency of (18)
2	0.1673	0.167	100	0.1683	0.168	0.167	0.167	100	101	101
3	0.1029	0.100	103	0.1045	0.100	0.100	0.100	105	105	105
4	0.0855	0.085	101	0.0729	0.075	0.086	0.086	97	85	85
5	0.0659	0.066	100	0.0605	0.062	0.066	0.066	98	92	92
6	0.0535	0.053	101	0.053	0.052	0.053	0.053	102	100	100

In Exponential distribution, our proposed equations (15) are found to be having same or slightly higher efficiencies than that of equations giving OPS proposed by Gupt and Ahamed (2020). When we compare our proposed methods of approximation with the method of approximation proposed by Gupt and Ahamed (2020), for number of strata 2, 3 and 6 our approximation methods are slightly better or equal to that of their method, but for number of strata 4 and 5, their approximation method is slightly better. But our proposed approximation (16) is having almost same efficiency as that of theirs for numbers of strata 4 and 5.

Table 8: Comparison with the methods proposed by Gupt and Ahamed (2020), Right-triangular distribution

No. of Strata (L)	Equations for OPS proposed by Gupt and Ahamed (2020)	Proposed equations (15)	Comparison of (15) *w.r.t* equations for OPS proposed by Gupt and Ahamed (2020)	Approximation methods proposed by Gupt and Ahamed (2020)	$nV(\overline{y_{st}})$ for proposed Approximation Methods			Comparison of methods (16), (17) and (18) *w.r.t* approximation method proposed by Gupt and Ahamed (2020)		
	$nV(\overline{y_{st}})$	$nV(\overline{y_{st}})$	Relative efficiency	$nV(\overline{y_{st}})$	Method (16)	Method (17)	Method (18)	Relative efficiency of (16)	Relative efficiency of (17)	Relative efficiency of (18)
2	0.0224	0.022	102	0.0234	0.022	0.022	0.022	106	106	106
3	0.0130	0.013	100	0.0130	0.013	0.013	0.013	100	100	100
4	0.0109	0.009	121	0.0102	0.009	0.010	0.010	113	102	102
5	0.0098	0.010	98	0.0103	0.009	0.010	0.010	114	103	103
6	0.0088	0.009	98	0.0087	0.008	0.009	0.009	109	97	97

In the population of Right-triangular distribution, the proposed equations (15) are having slightly more efficiencies than that of their proposed equations giving OPS for numbers of strata 2, 3 and 4. For numbers of strata 5 and 6, the proposed equations (15) are having slightly less efficiencies than that of the equations proposed by them. But in the case of comparison of the methods of approximations (16), (17) and (18) with their proposed method of approximation, our methods are having higher efficiencies except at number of strata 6, our proposed methods (17) and (18) are having slightly less efficiencies than that of their method of approximation. But, the proposed approximation method (16) is performing best of all approximation methods proposed in this paper as well as proposed by them.

Table 9: Comparison with the methods proposed by Gupt and Ahamed (2020), Uniform distribution

No. of Strata (L)	Equations for OPS proposed by Gupt and Ahamed (2020)	Proposed equations (15)	Comparison of (15) *w.r.t* equations for OPS proposed by Gupt and Ahamed (2020)	Approximation methods of Gupt and Ahamed (2020)	$nV(\overline{y_{st}})$ for proposed Approximation Methods			Comparison of methods (16), (17) and (18) *w.r.t* approximation method proposed by Gupt and Ahamed (2020)		
	$nV(\overline{y_{st}})$	$nV(\overline{y_{st}})$	Relative efficiency	$nV(\overline{y_{st}})$	Method (16)	Method (17)	Method (18)	Relative efficiency of (16)	Relative efficiency of (17)	Relative efficiency of (18)
2	0.0278	0.028	99	0.0283	0.0278	0.028	0.0278	102	101	102
3	0.0164	0.0164	100	0.0168	0.0162	0.0164	0.0164	104	102	102
4	0.0127	0.0127	100	0.0128	0.0126	0.0127	0.0127	102	101	101
5	0.0119	0.0119	100	0.0117	0.0119	0.0119	0.0119	98	98	98
6	0.0108	0.0108	100	0.0113	0.0107	0.0108	0.0108	106	105	105

In the population of Uniform distribution, the proposed equations (15) perform with almost same efficiencies as that of their proposed equations giving OPS. The proposed methods of approximation (16), (17) and (18) are performing with higher efficiencies than that of approximation method proposed by them except for number of strata 5 at which these proposed methods are performing with slightly less efficiencies than that of approximation method proposed by them.

In using the approximation methods, Gupt and Ahamed's (2020) method is restricted to use in stratifying population in which auxiliary variable, based on which stratification is to be done, of the study variable follows a known probability density function, but our three proposed methods of approximation are free from such restriction.

4. Conclusion

In this paper, it is observed that the proposed equations (15) giving OPS and methods of approximation (16), (17) and (18) giving AOPS are found to be different from the equations and methods proposed by Yadava and Singh (1984), which could be obtained as particular case of the generalised methods proposed by Gupt and Ahamed (2020), although allocation used is same in all the cases. The use of different techniques and procedure in the same problem in this paper has yielded a distinct set of equations for giving OPS and a few methods of obtaining AOPS which are very efficient and fairly suitable for practical applications in stratifying various heteroscedastic populations. In all the empirical illustrations, the proposed equations and methods of approximation are found to be performing better than or as good as methods of stratification proposed by other authors considered in this paper. It is also fascinating to learn that all the approximation methods proposed in this paper are found almost as efficient as exact equations (15) giving OPS. The method of approximation (16) - AOPS between any two consecutive strata are given by geometric mean of means of two consecutive strata of the auxiliary variable - is found to be the best of all the proposed methods in terms of efficiently stratifying population of all types and suitability for practical applications. Therefore, this method is recommended for practical application in stratifying populations of the considered level of heteroscedasticity optimally.

Acknowledgements

We, the authors of this paper, express our sincere thankfulness to the anonymous reviewer whose enormous suggestion has led to the elaborate improvement of the paper. We are also thankful to the Chief Editor for his valuable suggestion during the revision of the paper.

References

Dalenius, T. (1950). The problem of optimum stratification-I.*Skandinavisk Aktuaritidskrift,* **33**, 203-213.

Dalenius, T. and Gurney, M. (1951). The problem of optimum stratification-II. *Skandinavisk Aktuaritidskrift,* **34**, 133-148.

Dalenius, T., and Hodges, J. L. (1957). The choice of stratification points. *Skandinavisk Aktuaritidskrift,* **3(4)**, 198-203.

Gupt, B. K. (2003). Sample size allocation for stratified sampling under a correlated superpopulation model. *METRON -International Journal of Statistics, LXI, n,1:*35-52.

Gupt, B. K. (2012). *Allocation of sample size in stratified sampling under superpopulation models.* Saarbrucken, Deutschland/Germany: LAP LAMBERTAcademic Publishing AV Akademikerverlag GmbH & Co.KG.

Gupt, B. K. and Ahamed, M. I. (2020). 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

Neyman, J. (1934). 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*, **97(4)**, 558-625.

Rao, T. J. (1968). On the allocation of sample size in stratified sampling. *Annals of the Institute of Statistical Mathematics*, **20**, 159-166.

Singh, R. and Sukhatme, B. V. (1969). Optimum stratification. *Annals of the Institute of Statistical Mathematics*, **21**, 515-528.

Singh, R. and Sukhatme, B. V. (1972). Optimum stratification in sampling with varying probabilities. *Annals of the Institute of Statistical Mathematics*, **24**, 485-494.

Singh, R. (1971). Approximately optimum stratification on the auxiliary variable. *Journal of the American Statistical Association*, **66**, 829-833.

Singh, R. (1975a). An alternate method of stratification on the auxiliary variable. *Sankhya* **C37,** 100-108.

Singh, R. (1975b). On optimum stratification for proportional allocation. *Sankhya*, C**37**, Pt.I, 109-115.

Singh, R. (1975c). A note on optimum stratification in sampling with varying probabilities. *Australian Journal of Statistics*, **27(1)**, 12–21.

Singh, R. and Prakash, D. (1975). Optimum stratification for equal allocation. *Annals of the Institute of Statistical Mathematics*, **27**, 273-280.

Taga, Y. (1967). On optimum stratification for the objective variable based onconcomittant variable. *Annals of the Institute of Statistical Mathematics*, **19**, 101-130.

Tschuprow, A. A. (1923). On mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron,* **2**, 461-493.

Yadava, S. S. and Singh, R. (1984). Optimum stratification for allocation proportional to strata totals for simple random sampling scheme. *Communications in Statistics: Theory and Methods*, **13(22)**, 2793-2806.