Available online at http://scik.org

J. Math. Comput. Sci. 11 (2021), No. 6, 8342-8353

https://doi.org/10.28919/jmcs/6774

ISSN: 1927-5307

OPTIMUM STRATIFICATION FOR STRATIFIED PPSWR SAMPLING DESIGN UNDER A MODEL BASED ALLOCATION

BHUWANESHWAR KUMAR GUPT[1,*], MANKUPAR SWER[2], MD. IRPHAN AHAMED[3], KH. HERACHANDRA SINGH[4]

[1]Department of Statistics, North Eastern Hill University, Shillong-793022, Meghalaya, India

[2]Department of Mathematics, North Eastern Regional Institute of Science and Technology, Nirjuli, Itanagar- 791109, Arunachal Pradesh, India

[3]Department of Mathematics, Umshyrpi College, Shillong-793004, Meghalaya, India

[4]Department of Mathematics, Manipur University, Canchipur-795003, Manipur, India

Abstract: In this article we consider the problem of finding optimum points of stratification (OPS), based on an auxiliary variable which is highly correlated with study variable, for a model-based allocation, under stratified probability proportional to size with replacement (PPSWR) sampling design. The known model-based allocation used here is an available allocation in earlier literature which was obtained by the authors of the same under a superpopulation model in PPSWR sampling design. Therefore, in this paper, we use the same sampling design and superpopulation model used by them in dealing with problems of finding OPS. Equations for obtaining the OPS have been obtained. It is fascinating to discover that of all hitherto stratification methods developed under PPSWR sampling design for stratifying heteroscedastic populations which are available in literature, our proposed method has come out to be the most brief and easiest to use as OPSs, by this proposed method, are given by geometric means of means of consecutive strata. Although this method is application friendly, it is implicit; therefore alternative methods of finding approximately optimum points of stratification (AOPS) have also been obtained. The efficiencies of all the proposed methods of stratification are examined by using two randomly chosen live populations. The proposed methods of stratification are found to be efficient and suitable for practical applications.

---

*Corresponding Author

E-mail Address: bhuwaneshwargupt@gmail.com

Received: September 9, 2021

Keywords: auxiliary variable; estimation variable; optimum points of stratification; probability proportional to size with replacement; superpopulation model.

2010 AMS Subject Classification: 62D05.

1. INTRODUCTION

One of the problems in stratified sampling is the determination of optimum points of stratification (OPS) by minimizing the variance of an estimate for a population parameter. For stratification based on estimation variable y, the problem of determining OPS was first worked out mainly by Dalenius [1], Dalenius and Gurney [2], and Hayashi, Maruyama and Ishida [3]. In practice information on the study variable y is not available but information on some highly correlated auxiliary variable x is available. Cochran [4] demonstrated that when information on an auxiliary variable which is highly correlated with the study variable is available, a superpopulation model could be constructed in which a finite population under consideration could be deemed as a random sample from an infinite superpopulation having the same characteristics as that of infinite superpopulation. Based on this, Hanurav [5] and Rao [6] were the first to use auxiliary information to allocate sample size to strata whereas Taga [7] and Singh and Sukhatme [8] were the first to deal with problem of optimum stratification based on the auxiliary variable. The minimal equations for obtaining OPS for various allocations based on auxiliary variable were worked out by many workers among which Taga's [7] work is considered to be the most noteworthy till that time. Singh and Sukhatme [8] obtained equations giving OPS and approximately OPS for Tschuprow [9] - Neyman [10] optimum allocation (TNOA) and proportional allocation for the construction of strata. Under simple random sampling scheme, Singh and Prakash [11] considered the problem of optimum stratification on auxiliary variable x for equal allocation. Yadava and Singh [12] too obtained equations giving OPS and approximately OPS for allocation proportional to stratum totals under the same sampling design. Singh and Sukhatme [13] considered the problem of optimum stratification on an auxiliary variable x for TNOA when the samples from different strata are selected with probability proportional to size with replacement (PPSWR); Singh [14] considered the same problem for proportional and equal allocations.

Gupt and Rao [15] obtained a model based allocation of sample size to strata in sampling with PPSWR within each stratum, the allocation is given as

$$n_h \propto \sqrt{N_h(N_h - 1)\bar{X}_h} \tag{1}$$

under the following superpopulation model considered by them:

$$\begin{cases} \xi(y_i|x_i) & = & \beta x_i \\ \mathcal{V}(y_i|x_i) & = & \sigma^2 x_i \\ \zeta(y_i, y_j|x_i, x_j) & = & 0 \end{cases} \tag{2}$$

where β, σ^2 are superpopulation parameters with $\sigma^2 > 0$ and ξ, $\mathcal{V}$, ζ denote the conditional expectation, variance and covariance given x_i respectively. Gupt [16, 17] considered problem of sample size allocation in simple random sampling with and without replacement under a more general superpopulation model defined by $\xi(y_i|x_i) = \alpha + \beta x_i$, $\mathcal{V}(y_i|x_i) = \sigma^2 x_i^g$ and existence of intra-stratum correlation coefficient, and he obtained several allocations. Gupt and Ahamed [18] considered the problem of optimum stratification for a generalized auxiliary variable proportional allocation (GAVPA) obtained by Gupt [16, 17].

Finite population taken from a superpopulation is considered to be large, therefore, we assume $N_h \approx N_h - 1 \; \forall \; h = 1, 2, 3, \ldots, L$, where L is the number of strata into which the population is divided, and hence the allocation given by (1) reduces to

$$n_h \propto N_h\sqrt{\bar{X}_h}\ . \tag{3}$$

In this article we deal with the problem of obtaining OPS and approximately optimum points of stratification for the allocations given by (3) under stratified PPSWR sampling design. This article comprises of four sections. Equations giving OPS have been obtained in Section 2 and Sub-section 2.1. Derivation of the alternative method of finding approximately optimum points of stratification (AOPS) has been shown in Sub-section 2.2. Main results are given in Section 3. Numerical illustrations by using two randomly chosen live populations along with discussion on the efficiencies of the proposed methods of stratification are given in Sub-section 3.1 and conclusion is given in Section 4.

2. PRELIMINARIES

2.1. Equations Giving Optimum Points of Stratification. Considering a finite population of size N divided into L number of strata of sizes N_h, $h = 1, 2, \ldots, L$., the allocation given by (3) can be written as

(4) $$n_h = \frac{nN_h\sqrt{\bar{X}_h}}{\sum_{h=1}^{L} N_h\sqrt{\bar{X}_h}}.$$

Gupt and Rao [15] obtained the following expression under the superpopulation model (2):

(5) $$E\left[A_h^2(y)\middle|\underline{X_h}\right] = \sigma^2(N_h - 1)X_h$$

where $\underline{X_h}' = \left(X_{h1} \;\; X_{h2} \;\; X_{h3} \;\ldots\; X_{hN_h}\right)$ is a vector of x-values in the h^{th} stratum.

When $N_h \approx N_h - 1 \;\forall\; h$, we get

(6) $$E\left[A_h^2(y)\middle|\underline{X_h}\right] = \sigma^2 N_h X_h = \sigma^2 {N_h}^2 \bar{X}_h$$

The variance of estimator of population total for stratified PPSWR sampling design is given by

(7) $$V\left(\hat{Y}_{st}\right) = \sum_{h=1}^{L} \frac{A_h^2(y)}{n_h}$$

where $A_h^2(y) = \sum_{i=1}^{N_h} \frac{Y_{hi}^2}{P_{hi}} - Y_h^2$ and $P_{hi} = \frac{X_{hi}}{X_h}$.

Taking conditional expectation of (7) which is variance of estimator of the population total, we have

(8) $$E\left[V\left(\hat{Y}_{st}\right)|\boldsymbol{X}\right] = \sum_{h=1}^{L} \frac{E[A_h^2(y)|\underline{X_h}]}{n_h},$$

where $\boldsymbol{X} = (\; \underline{X}_1 \;\; \underline{X}_2 \ldots \underline{X}_r \ldots \underline{X}_L \;)$ is the matrix of x values in the population.

By taking (4), (6) and (8), we obtain

(9) $$E\left[V\left(\hat{Y}_{st}\right)|\boldsymbol{X}\right] = \frac{\sigma^2}{n}\left(\sum_{h=1}^{L} N_h\sqrt{\bar{X}_h}\right)^2.$$

We minimize $E\left[V\left(\hat{Y}_{st}|\; \boldsymbol{X}\right)\right]$ by differentiating (9) partially with respect to x_h, where x_h is the demarcation point between h^{th} and $(h+1)^{th}$ strata, and then equating to zero as follows:

(10) $$\frac{\sigma^2}{n}\frac{\partial}{\partial x_h}\left(\sum_{h=1}^{L} N_h\sqrt{\bar{X}_h}\right)^2 = 0.$$

While differentiating partially with respect to x_h, all other terms vanishes except the h^{th} and $(h+1)^{th}$ terms and, therefore, (10) reduces to

(11) $$\frac{\partial}{\partial x_h} N_h\sqrt{\bar{X}_h} + \frac{\partial}{\partial x_h} \; N_{h+1}\sqrt{\bar{X}_{h+1}} = 0.$$

If $f(x)$ is the density function of x and $W_h = \int_{x_{h-1}}^{x_h} f(x)dx$, then, we can obtain the following:

(12) $$W_h\bar{X}_h = \int_{x_{h-1}}^{x_h} xf(x)dx .$$

By taking the first term on the LHS of (11) and using (12), we get

(13) $$\frac{\partial}{\partial x_h} N_h\sqrt{\bar{X}_h} = \frac{N}{2}\frac{f(x_h)}{\sqrt{\bar{X}_h}}[x_h + \bar{X}_h] .$$

Similarly, taking the second term on the RHS of (11) and again using (12), we get

(14) $$\frac{\partial}{\partial x_h} N_{h+1}\sqrt{\bar{X}_{h+1}} = -\frac{N}{2}\frac{f(x_h)}{\sqrt{\bar{X}_{h+1}}}[x_h + \bar{X}_{h+1}].$$

Substituting (13) and (14) in (11) we obtain the solutions as

(15) $$\frac{x_h+\bar{X}_h}{\sqrt{\bar{X}_h}} = \frac{x_h+\bar{X}_{h+1}}{\sqrt{\bar{X}_{h+1}}}$$

(16) $$\Rightarrow x_h = \sqrt{\bar{X}_h.\bar{X}_{h+1}} , \quad h = 1,2,\dots,L-1.$$

Equations (16) give OPS of the study variable y based on the auxiliary variable x.

It is observed that the stratification method (16) gives OPS as the geometric mean of two consecutive strata means. It may be noted that in case of stratified simple random sampling design used with proportional allocation, the OPS is obtained as the arithmetic mean of two consecutive strata means (Murthy, [19]).

Hence we get the following theorem:

Theorem 2.1. *In stratified PPSWR sampling, for the allocation* $n_h \propto N_h\sqrt{\bar{X}_h}$ *under the superpopulation model (2), the optimum points of stratification for the study variable* y *based on an auxiliary variable* x *are given by the equations* $x_h = \sqrt{\bar{X}_h.\bar{X}_{h+1}}$ *where* $h = 1,2,\dots,L-1$ *and* L *is the number of strata.*

2.2. Derivation of Alternative Methods of Finding Approximately Optimum Points of Stratification. Although the equations (16) give the OPS, these equations comprise population parameters which are the functions of OPS. Because of the implicit nature of the equations, there is some difficulty in finding exact solutions. Therefore, we obtain alternative methods for finding approximate solutions to equations (16). We follow the technique of Singh and Sukhatme [8] of using Ekman's [20] identity in obtaining series expansion of conditional mean. We assume that the functions $f(x)$ possess various partial derivatives for all x in the range (a,b) with $(b-a) < \infty$.

Thus, we obtain

(17) $$\bar{X}_{h+1} = \mu_i(x) = x_h + \frac{k_i}{2} + \frac{f^{'}(x_h)}{12 f(x_h)} {k_i}^2 + \frac{f(x_h) f^{''}(x_h) - f^{'2}(x_h)}{24 f^2(x_h)} {k_i}^3 + O({k_i}^4)$$

where the function f and its derivatives are evaluated at $t = x_h$ in the interval $t \in [x_h, x_{h+1}]$ and $k_i = x_{h+1} - x_h$, $i = h + 1$.

Using (17), the RHS of equations (15) transform to

(18) $$\frac{1}{\sqrt{x_h}}\left[2x_h + \frac{1}{16 x_h} {k_i}^2 + \frac{2 x_h f^{'}(x_h) - 3 f(x_h)}{96 {x_h}^2 f(x_h)} {k_i}^3 + O({k_i}^4)\right].$$

Similarly, the LHS of equations (15) transform to

(19) $$\frac{1}{\sqrt{x_h}}\left[2x_h + \frac{1}{16 x_h} {k_h}^2 - \frac{2 x_h f^{'}(x_h) - 3 f(x_h)}{96 {x_h}^2 f(x_h)} {k_h}^3 + O({k_h}^4)\right],$$

where the function f and its derivatives are evaluated at $t = x_h$ in the interval $t \in [x_{h-1}, x_h]$ and $k_h = x_h - x_{h-1}$.

Considering a large number of strata whose strata width k_h are very small, the higher powers of k_h in the expansion can be neglected.

Putting (18) and (19) in equations (15) and taking $(t) = \frac{f(t)}{t^{3/2}}$, we get

(20) $${k_h}^3 g(t)\left[1 - \frac{g^{'}(t)}{g(t)}\frac{k_h}{2} + O\left({k_h}^2\right)\right] = {k_i}^3 g(t)\left[1 + \frac{g^{'}(t)}{g(t)}\frac{k_i}{2} + O\left({k_i}^2\right)\right].$$

Singh and Sukhatme [8] obtained the following result by expanding $\sqrt[\lambda]{f(t)}$ about the point $t = y$:

$$\left[\int_y^x \sqrt[\lambda]{f(t)}\, dt\right]^{\lambda} = k^{\lambda} f(y)\left[1 + \frac{k}{2}\frac{f^{'}(y)}{f(y)} + O(k^2)\right]$$

(21) $$= k^{\lambda - 1} \int_y^x f(t) dt\, [1 + O(k^2)]$$

Using (21) in (20), we obtain

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

$$\Rightarrow {k_h}^2 \int_{x_{h-1}}^{x_h} g(t) dt = {k_i}^2 \int_{x_h}^{x_{h+1}} g(t) dt$$

(23) $$\Rightarrow {k_h}^2 \int_{x_{h-1}}^{x_h} g(t) dt = C_1$$

(24) $$\Rightarrow \int_{x_{h-1}}^{x_h} \sqrt[3]{g(t)} dt = C_2$$

Using either the relation (23) or (24) we can obtain the AOPS corresponding to (16). To find the AOPS i.e. $x_h's$ by using (23) or (24), we fix the lower boundary x_{h-1} every time and the values of constants C_1 or C_2 are approximately evaluated by taking $C_1 = \frac{1}{L}(b - a)^2 \int_a^b g(t) dt$ or $C_2 =$

$\frac{1}{L}\int_a^b \sqrt[3]{g(t)}dt$ respectively, where b and a are upper and lower bounds of points of stratification x_h's, i.e., $a \le x_h \le b$.

Thus, we have established the following theorem.

Theorem 3.1. *If the function $g(x)$ is bounded and possesses at least the first two partial derivatives for all values of x in (a,b), for a given number of strata, taking equal intervals on the cumulative of $\sqrt[3]{g(x)}$ gives approximately optimum points of stratification of the variable x.*

3. MAIN RESULTS

3.1. Numerical Illustration for the Proposed Methods of Stratification. For examining the performances of the equations (16) and the methods of approximation (23) and (24) in obtaining optimum points of stratification, we illustrate the equations in the following randomly chosen live populations:

P1. The population is taken from Cochran [21], i.e., the number of inhabitants (in thousands) in each of a simple random sample of 49 cities drawn from the population of 196 large United States cities in 1920 and 1930 are taken as auxiliary variable X and estimation variable Y respectively.

P2. The population is taken from the Directorate of Economic and Statistics, Meghalaya, India [22] which consists the number of household and total population of 158 villages under the Baghmara Community and Rural Development Block of the South Garo Hills District of Meghalaya in India based on the population census 2011. The total population is taken as the estimation variable Y while the number of household is taken as the auxiliary variable X.

In sub-section 2.2 we have proved that the methods of approximation (23) and (24) are equivalent and therefore, we are free to use any of the two. Here we use (24) in our illustration. On applying linear least square regression technique between the response variable $\boldsymbol{y}$ and explanatory variable $\boldsymbol{x}$ in population P1, we get the coefficient of determination $\boldsymbol{R^2}$=0.963817, intercept, $\boldsymbol{\alpha}$ =8.384, β = 1.157. This shows that the auxiliary variable *x* is highly correlated with study variable *y* in population P1. Applying the same technique in population P2, we get the coefficient of determination $\boldsymbol{R^2}$=0.966078, intercept, $\boldsymbol{\alpha}$ =5.950, β = 5.874. This also shows that the auxiliary variable *x* is highly correlated with study variable *y* in population P2.

For stratifying population P1 using (24), it is required to determine a Probability Density Function (PDF) that the auxiliary variable x of the population P1 follows. To fit a suitable PDF in the population, we use the data of x variable. Using the fitdistrplus package in R-software, we fit a number of known PDFs in the data of x variable whose each of the values is divided by 100. The PDFs are fitted using the methods Maximum Likelihood Estimation (MLE), Moment Matching Estimation (MME) and Quantile Matching Estimation (QME) one after another.

By comparing the values of LL (log likelihood), AIC (Akaike Information Criteria), BIC (Bayesian Information Criteria) and standard errors ($s.e$) of parameters of the PDFs, when fitted various PDFs to the data, log-normal probability density function is found to be fitting best to the data of population P1.

Thus, the PDF followed by the x variable of population P1 is as follows:

(25) $$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], x > 0$$

In fitting the log-normal probability density function, the population P1 is characterized as follows:

$Mean\ log$, $\mu = -0.3477068$, $s.d\ log$, $\sigma = 0.9019393$, $s.e(\mu) = 0.128848$, $s.e(\sigma) = 0.091109$, $LL = -47.43316$, $AIC = 98.86631$, $BIC = 102.65$, which are estimated by MLE method.

For stratifying population P2 using (24), we proceed in the same manner as above and, coincidentally, we find that the log-normal probability density function fit best to the data of population P2 and thus the PDF followed by the x variable of the population P2 is also given by (25). In fitting the log-normal probability density function, the population P2 is characterized as follows:

$Mean\ log$, $\mu = -1.1385535$, $s.d\ log$, $\sigma = 0.6582057$, $s.e(\mu) = 0.052364$, $s.e(\sigma) = 0.0370266$, $LL = 21.78074$, $AIC = -39.56148$, $BIC = -33.43629$, which are estimated by MLE method.

For illustrating the approximation method (24) in stratifying each of the populations P1 and P2, we use the PDF (25) along with their respective estimates of parameters. Numerical integration and differentiation methods are used in working out the approximation method in stratifying each of the populations P1 and P2.

We examine the methods of stratification in each of the populations P1 and P2 by finding the OPS due to the equations (16) and AOPS due to the alternative approximation method (24). For each of the populations P1 and P2, variance of estimator of population mean at the optimum points of stratification for each number of strata $L = 2, 3, 4, 5, 6$ are calculated for both the methods. The calculations for finding OPS and AOPS involve successive iterations. We also calculate the variance of estimator of population mean for equal interval stratification for each of the number of strata $L = 2, 3, 4, 5, 6$. The efficiencies of the equations (16) and the alternative approximation method (24) are compared with that of equal interval stratification in stratifying the live populations. The OPS, AOPS, variances and relative efficiencies obtained due to the two methods for the two live populations P1 and P2 are given in Table 1 and Table 2 respectively.

TABLE 1. Population P1

No. of strata (L)	Equal interval stratification		Stratification by equations (16)		Relative efficiency	Stratification by alternative method (24)		Relative efficiency
	Points	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		Points	$nV(\bar{y}_{st})$	
2	255	2668.23	122.54	2383.98	**112**	180.94	2578.17	**103**
3	170.67 339.33	2476.42	77.30 207.42	2082.97	**119**	115.22 263.05	2214.34	**112**
4	128.50 255 381.50	2338.87	79.37 184.99 336.02	2075.46	**113**	86.86 180.94 312.24	2075.46	**113**
5	103.2 204.4 305.6 406.8	2181.22	79.37 184.99 319.40 441.24	2071.95	**105**	70.92 139.88 227.86 344.97	1901.49	**115**
6	86.33 170.67 254.99 339.33 423.67	2054.34	56.46 107.66 197.43 319.40 441.24	1726.74	**119**	60.62 115.22 180.94 263.05 368.29	1775.16	**116**

From the above table, we find that the equations (16) work with higher efficiency for $L = 5$ and with much higher efficiency for $L = 2, 3, 4$ and 6 when compared with that of equal interval stratification. Likewise, the approximation method (24) performs with higher efficiency for $L = 2$ and with much higher efficiency for $L = 3, 4, 5$ and 6 when compared with that of equal interval stratification. It is also seen that the equations (16) perform well than the alternative method (24) for $L = 2, 3$ and 6, same as the alternative method (24) for $L = 4$. However for $L = 5$, the alternative method (24) performs better than the equations (16).

TABLE 2. Population P2

No. of strata (L)	Equal interval stratification		Stratification by equations (16)		Relative efficiency	Stratification by alternative method (24)		Relative efficiency
	Points	$nV(\bar{y}_{st})$	Points	$nV(\bar{y}_{st})$		Points	$nV(\bar{y}_{st})$	
2	110	747.69	72.26	708.29	**106**	67.62	711.24	**105**
3	74.67 145.33	731.73	33.03 78.57	700.07	**105**	44.96 97.70	707.01	**103**
4	57 110 163	759.32	32.49 69.55 142.90	699.14	**109**	35.28 67.62 117.31	705.33	**108**
5	46.4 88.8 131.2 173.6	716.01	38.71 76.69 110.80 163.66	685.77	**104**	29.78 53.38 84.47 131.24	675.99	**106**
6	39.33 74.67 110 145.33 180.67	682.11	28.40 53.81 81.55 110.80 163.66	633.71	**108**	26.18 44.96 67.62 97.70 141.67	637.95	**107**

From the above table, we find that both the equations (16) and alternative method (24) work with higher efficiency for $L = 2, 3, 4, 5$ and 6 when compared with that of equal interval stratification. It is also seen that the equations (16) perform slightly better than the alternative approximation method (24) for $L = 2, 3, 4$ and 6. However for $L = 5$, the alternative approximation method (24) performs better than the equations (16).

Overall, we find that the equations (16) as well as the approximation method (24) perform well in stratifying the live populations.

4. CONCLUSION

In this paper, our proposed methods of stratification under PPSWR design are found to stratify population with high efficiencies when illustrated in randomly chosen live populations. The equations (16) for obtaining OPS and the alternative method (24) for obtaining AOPS perform with higher efficiency for all strata L when compared with that of equal interval stratification. It is, therefore, empirically as well as analytically justified that the both the methods can be applied to efficiently stratify populations. The specialty of the proposed stratification method (16) is that it gives OPS by geometric means of means of consecutive strata. Both the methods are user friendly although method (16) is implicit. The methods are convenient for practical applications in stratifying heteroscedastic populations effectively when samples are to be selected from strata with probability proportional to size with replacement. It is also worthy to mention that of all stratification methods under PPSWR sampling design available in literature so far, our proposed methods are the easiest, shortest and user friendly, besides their good efficiencies.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] T. Dalenius, The Problem of optimum stratification-I, Scandinavian Actuarial J. 33 (1950), 203-213.

[2] T. Dalenius, M. Gurney, The problem of optimum stratification II, Scandinavian Actuarial J. 34 (1951), 133-148.

[3] C. Hayashi, F. Maruyama, M.D. Ishida, On some criteria for Stratification, Ann. Inst. Stat. Math. 2 (1950), 76-86.

[4] W.G. Cochran, Relative accuracy of systematic and stratified random samples for a certain class of populations, Ann. Math. Stat. 17 (1946), 164-177.

[5] T.V. Hanurav, Optimum sampling strategies and some related problems, PhD Thesis, Indian Statistical Institute, (1965).

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

[7] Y. Taga, On optimum stratification for the objective variable based on concomittant variable, Ann. Inst. Stat. Math. 19 (1967), 101-130.

[8] R. Singh, B.V. Sukhatme, Optimum stratification, Ann. Inst. Stat. Math. 21 (1969), 515-528.

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

[10] J. Neyman, On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection, J. R. Stat. Soc. 97(4) (1934), 558-625.

[11] R. Singh, D. Prakash, Optimum stratification for equal allocation, Ann. Inst. Stat. Math. 27 (1975), 273-280.

[12] S.S. Yadava, R. Singh, Optimum stratification for allocation proportional to strata totals for simple random sampling scheme, Commun. Stat.: Theory Meth. 13(22) (1984), 2793-2806.

[13] R. Singh, B.V. Sukhatme, Optimum stratification in sampling with varying probabilities, Ann. Inst. Stat. Math. 24 (1972), 485-494.

[14] R. Singh, A note on optimum stratification in sampling with varying probabilities, Australian J. Stat. 17 (1975), 12-21.

[15] B.K. Gupt, T.J. Rao, Stratified PPS sampling and allocation of Sample size, J. Indian Soc. Agric. Stat. 50(2) (1997), 199-208.

[16] B.K. Gupt, Sample size allocation for stratified sampling under a correlated superpopulation model, METRON- Int. J. Stat. LXI(1) (2003), 35-52.

[17] B.K. Gupt, Allocation of sample size in stratified sampling under superpopulation models, LAP Lambert Academic Publishing, Saarbrucken, Deutschland/Germany, (2012).

[18] B.K. Gupt, Md.I. Ahamed, Optimum stratification for a generalized auxiliary variable proportional allocation under a superpopulation model, Communications in Statistics - Theory and Methods. (2020) 1–16. https://doi.org/10.1080/03610926.2020.1793203.

[19] M.N. Murthy, Sampling Theory and Methods, Statistical Publishing Society, Calcutta, (1967).

[20] G. Ekman, Approximate expressions for the conditional mean and variance over small intervals of a continuous distribution, Ann. Math. Stat. 30(4) (1959), 1131-1134.

[21] W.G. Cochran, Sampling Technique, 3rd Edition, Wiley, New York, 152 p (1977).

[22] Directorate of Economics and Statistic Meghalaya, A Handbook on the Block-Wise Demographic Profile of Meghalaya – 2018, Government of Meghalaya, Shillong, (2018), http://megplanning.gov.in/handbook.htm