Journal of Rajasthan Academy of Physical Sciences
ISSN: 0972-6306; URL: http://raops.org.in
International Conference on Mathematical and Statistical Computation (ICMSC-2022)
Swami Keshvanand Institute of Technology (SKIT), Jaipur, Rajasthan, (India) 3rd-5th March 2022
October, 2022, 89-98

COMPROMISE MIXED ALLOCATION IN MULTIVARIATE STRATIFIED SAMPLING WITH TRAVEL COST USING DYNAMIC PROGRAMMING TECHNIQUE

Abrar Ahmad[1], A.H. Ansari[2], Quazzafi Rabbani[3] and **Nisha Khanam[4]**
[1,3,4]Department of Mathematics & Statistics, Integral University, Lucknow, U.P.-226026 (India)
[2]Kejriwal Institute of Management and Development Studies, Namkum, Ranchi-834010, (India)
E-mail: [1]abrar.stats@gmail.com, [2]athar_hussain1@rediffmail.com, [3]quazzafi@gmail.com, [4]nishakaif24@gmail.com

Abstract: Proportional, Equal, Optimal and several other allocations are famous allocations in stratified random sampling (SRS) literature. Generally a specific allocation is used in every stratum. But practically it is not suggestive due to nature of strata, Ahsan *et al*. [3] present a new idea known as "Mixed Allocation", which is optimum in some sense means the special category or group of strata uses a specific allocation for different groups of strata. The concept of Mixed Allocation is discussed and explained by different authors. It is assumed that the concept of "Compromise Mixed Allocation" is applicable in multivariate case also. So, in this present manuscript the authors worked out the *p* characteristics as multivariate stratified sampling for multivariate in "Compromise Mixed Allocation" using Dynamic Programming Technique (DPT) having the quadratic travelling cost constraints because of practical nature of the cost imposed to get surveyed data from center of investigation to the sampling unit are imposed on every unit of universe (Population). An example is also mention to understand the computational details with the help of Lingo software.

AMS (2010) Subject Classification: 62D05

Keywords: Mixed Allocation, Multivariate Stratified Random Sampling, Dynamic Programming Technique, Quadratic Cost Constraints, Compromise Mixed Allocation.

1. Introduction

In sampling literature, stratified random sampling are having various types of allocation procedure present like Proportional, Equal, Optimum allocation etc. Literally any specific category of allocation is used based on the criteria of strata. But Ahsan *et al*. [3] worked on a practical situation where one type of allocation is not advisable to all the strata. They divided the M strata into t different mutually exclusive groups with similar nature. For particular group of strata they applied a special kind of allocation procedure based on the criteria of that group and termed it as Mixed Allocation. So, there is no literature available on mixed allocation prior to Ahsan *et al*. [3]

The following nonlinear programming problem (NLPP) is formulated to find the multivariate mixed allocation,

$$\left.\begin{aligned} &Min. && F(\beta_s)=\sum_{s=1}^{t}\sum_{p\in I_s}\frac{W_p^2 S_p^2}{\beta_s \alpha_p} \\ &Subject to && \sum_{s=1}^{t}\sum_{p\in I_s}\beta_s c_p \alpha_p \le C_0 \\ &and && \beta_s \ge 0;\ s=1,2,\dots,t \end{aligned}\right\} \tag{1}$$

where t groups are defined for M strata, the sth group consists of M_s strata. The allocation of sample are defined as,

$$n_p = \beta_s \alpha_p;\ p\in I_s,\ s=1,2,\dots,t, \tag{2}$$

where $\beta_s; s=1,2,3,\dots,t$, are calculated and get the result to NLPP (1), I_s is the integers set represents the number of in group s and α_p are defined constant depending upon the specific allocation procedure used. Suppose, if equal allocation is to be implemented in the q^{th} group then $\alpha_p = 1$; where $p \in I_q$.

The word compromise has a special meaning in English literature is settlement of a dispute, but it has a different definition in SRS with multivariate literature where r characteristics are to be measured with condition $r \ge 2$ on each selected sample units, a solution that minimizes the one characteristic variance of may not minimize the others. For such situations Yates [27] provided the concept of Compromise Allocation for combined objective functions of several variables in place of different objective functions. All the r characteristics for a fixed cost having compromise criterion. Later on several authors modified pre-existing criteria or suggested some novel approaches for compromise allocation. Some of those are Yates [27], Aoyama [6], Folks and Antle [18], Kokan and Khan [22], Chatterjee [13][14], Arvanitis and Afonja [7], Ahsan and Khan [1] [2], Melaku and Sadasivan [26], Bankier [8], Bethel [10], Kreienbrock [25], Jahan *et al*. [19], Khan *et al*. [20] [21], Díaz-Garca and Cortez [16] [17], Ansari *et al*. [4] and many others. Kozak [23] discussed to work out of an approximate in multivariate optimal allocation with five different compromise criteria and also do the comparative study with a simulation study. Kozak [24] present a calculation procedure to get the multivariate compromise allocation for stratified universe in three different criteria of compromise and improve the search method. Clark and Steel [15] also introduced a concept in single variate sampling of two stages. Recent contributions on some efficient classes of estimators under stratified sampling design have been made by Bhusan *et al*. [11] [12].

In this manuscript we extended the work of Ansari *et al*. [5] for the quadratic travel cost constraints. Rest of the paper is arranged in the following way: Heading 2 of the manuscript provides the solution process with expression of the problem. Heading 3 presents an example to explain the utilization of compromise mixed allocation. Heading 4 gives the concluding remark on the basis of the results obtained in Section 3. At the end of this manuscript the arranged (alphabetically) list of references are mentioned.

2. Solution Process with Expression

Yates [27] used a univariate criterion to solve the problem to find a mixed allocation as shown in (1), is expressed as follows for multivariate case as,

$$\left.\begin{array}{ll} Min. & \sum_{u=1}^{r} a_u \sum_{s=1}^{t} \sum_{p\in I_s} \frac{W_p^2 S_p^2}{\beta_s \alpha_p} \\ Subjectto & \sum_{s=1}^{t} \sum_{p\in I_s} \beta_s c_p \alpha_p \le C_0 \\ and & \beta_s \ge 0\ ;\ s=1,2,\dots,t,\quad C_0 = C - c_0 \end{array}\right\} \tag{3}$$

where $a_u > 0$ is weights given to different variations of u^{th} characteristic, S_{up}^2 is u^{th} characteristic stratum variation, C is the allotted cost, c_0 is the fixed cost, C_0 is the allotted budget excluding the fixed cost and $n_p = \alpha_s \beta_p$. Afterward c_p will the r characteristic on a picked item of p^{th} stratum, denote cost of observation for all that is $c_p = \sum_{u=1}^{r} c_{up}$; $p=1,2,\dots,M$, c_{up} denote the cost of observation per unit for the u^{th} characteristic in p^{th} stratum. We can substitute $\sum_{u=1}^{r} a_u = 1$ as per the above explanation.

$A_p = W_p^2 \sum_{u=1}^{r} \beta_u S_p^2; p=1,2,3,\dots,M$ and reformulated the terms Non Linear Programming Problem (NLPP) (3) may be rewritten as,

$$\left.\begin{array}{lll} Min. & \sum_{s=1}^{t}\sum_{p\in I_s}\frac{A_p}{\beta_s\alpha_p} \\ Subjectto & \sum_{s=1}^{t}\sum_{p\in I_s}\beta_s c_p\alpha_p \le C_0 \\ and & \beta_s \ge 0;\ s=1,2,\ldots,t \end{array}\right\} \quad (4)$$

Beardwood *et al*. [9] suggested that the n_h randomly selected units in the p^{th} stratum having cost of visiting may be treated as $t_p\sqrt{n_p}\,;\ p=1,2,3,\ldots.,M$ approx., where the travel cost per unit in the p^{th} stratum is t_p. With this conjecture, using Ansari *et al*. [5] to find the solution mixed allocation problem written in (3) the quadratic constraints problem in multivariate situation can now be expressed as,

$$\left.\begin{array}{lll} Min. & F(\beta_s)=\sum_{s=1}^{t}\sum_{p\in I_s}\frac{A_p}{\beta_s\alpha_p} \\ subject\ to & \sum_{s=1}^{t}\sum_{p\in I_s}\beta_s c_p\alpha_p+\sum_{s=1}^{t}\sum_{p\in I_s}t_p\sqrt{\beta_s\alpha_p}\le C_0 \\ and & \beta_s\ge 0\ ;\ s=1,2,\ldots,t \end{array}\right\} \quad (5)$$

where C_0 be the allotted budget and $n_p=\beta_s\alpha_p$. Afterward $c_p=\sum_{u=1}^{r}c_{up}\ ;\ p=1,2,\ldots,M$, be the observation cost of all the *r* characteristics on a selected unit of p^{th} stratum, c_{up} denotes the of observation cost of each item in p^{th} stratum for u^{th} characteristic is $t_p=\sum_{u=1}^{r}t_{up}\ ;\ p=1,2,\ldots,M$, denotes the travel cost per unit of *r* characteristics in the p^{th} stratum on a selected unit, t_{up} denotes the cost per unit of travel for u^{th} characteristic in p^{th} stratum.

The Non Linear Programming Problem (4) can be solved by using DPT.

Consider the r^{th} stage sub problem of MPP (5) for the first r(<k) groups,

$$\left.\begin{array}{lll} \textit{Min.} & \sum_{s=1}^{t} f_s(\beta_s) & \\ \textit{subject to} & \sum_{s=1}^{t} g_s(\beta_s) \le C_r & \\ \textit{and} & \beta_s \ge 0\ ;\ s=1,2,...,t & \end{array}\right\} \qquad (6)$$

where $f_s(\beta_s)=\sum_{p\in I_s}\frac{A_p}{\beta_s\alpha_p}$, $g_s(\beta_s)=\sum_{p\in I_s}c_p\alpha_p\beta_s+\sum_{p\in I_s}t_p\sqrt{\alpha_p\beta_s}\le C_0$; $s=1,2,...,t$.

$C_v < C_0$ be the total budget for observation of the selected items from the first *r* groups. We can use the defined C_v and can proceed as follows:

$$C_v = C_t \qquad if \qquad v=t$$

now $\quad C_v = g_1(\beta_1)+g_2(\beta_2)+...+g_v(\beta_v)$

$$C_{v-1} = g_1(\beta_1)+g_2(\beta_2)+...+g_{v-1}(\beta_{v-1}) = C_v - g_v(\beta_v)$$

$$C_2 = g_1(\beta_1)+g_2(\beta_2) = C_3 - g_3(\beta_3)$$

and $\quad C_1 = g_1(\beta_1) = C_2 - g_2(\beta_2)$

If $f\ (v, C_v)$ be the objective function value which is minimum of sub problem (6), then

$$f\ (v,C_v) = \underset{feasible\ \alpha_s}{Min}\left\{\sum_{s=1}^{v} f_s\ (\beta_s) : \sum_{s=1}^{v} g_s\ (\beta_s) = C_v\ \ and\ \beta_s \ge 0\ ; s=1,2,...,v\right\} \qquad (7)$$

For first stage (*v*=1)

$$f(1,C_1)=\frac{2\times\left(\sum_{p\in I_1}^{v}c_p\alpha_p\right)^2\times\left(\sum_{p\in I_p}\frac{A_p}{\alpha_p}\right)}{\left[2C_1\left(\sum_{p\in I_1}^{v}c_p\alpha_p\right)+\left(\sum_{p\in I_1}^{v}t_p\sqrt{\alpha_p}\right)^2\right]\pm\sqrt{\left[2C_1\left(\sum_{p\in I_1}^{v}c_p\alpha_p\right)+\left(\sum_{p\in I_1}^{v}t_p\sqrt{\alpha_p}\right)^2\right]^2-4\times\left(\sum_{p\in I_1}^{v}c_p\alpha_p\right)^2(C_1)^2}}$$

at

$$\beta_1 = \frac{\left[2C_1\left(\sum_{p\in I_1}^{v} c_p\alpha_p\right)+\left(\sum_{p\in I_1}^{v} t_p\sqrt{\alpha_p}\right)^2\right] \pm \sqrt{\left[2C_1\left(\sum_{p\in I_1}^{v} c_p\alpha_p\right)+\left(\sum_{p\in I_1}^{v} t_p\sqrt{\alpha_p}\right)^2\right]^2 - 4\times\left(\sum_{p\in I_1}^{v} c_p\alpha_p\right)^2 (C_1)^2}}{2\times\left(\sum_{p\in I_1}^{v} c_p\alpha_p\right)^2} \tag{8}$$

and for $v \geq 2$

$$f\ (v, C_v) = \min_{0\le g_v(\beta_v)\le C_v}\left\{\frac{\sum_{p\in I_v}\frac{A_p}{\alpha_p}}{\beta_v} + f\,(v-1, C_v - g_v(\beta_v))\right\} \tag{9}$$

Expression (9) provides the required relation of recurrence.

From $f\ (t, C)$, we can calculate the value of β_t is obtained from $f\ (t-1, C - g_t(\beta_t))$ to get the optimum value of β_{t-1} and so on until α_1 is found using recurrence relation calculation.

After obtaining $\beta_s; s = 1,2,...,t$ the different values of n_p are obtained by using (2), Now, we are in age of computer, so, the compromised mixed objective value may calculated accordingly by the help of Lingo software as,

$$V_{mixed} = \sum_{s=1}^{t}\frac{\sum_{p\in I_p}\frac{A_p}{\alpha_p}}{\beta_s} \tag{10}$$

Since the value of objective function is "compromised mixed" allocation, so, it is not as good as optimal objective. Therefore the authors are not interested to calculate relative loss of efficiency.

3. An Example

Ansari *et al*. [5] produced an example with arbitrary data. We added only travelling cost t_p, which is quadratic in nature. Thus the following situation arises.

In stratification two characteristics in seven strata are the values of N_p, s_{1p}, s_{2p}, c_p and t_p are shown in Table 1. Let us assume that $C = 7500$ units be the total budget available for the survey includes a fixed cost $c_0 = 500$ units. This gives $C_0 = C - c_0 = 7500 - 500 = 7000$ units be the total available amount for observation.

Table 1. Arbitrary Data

p	N_p	s_{1p}	s_{2p}	c_p	t_p	W_p
1	472	5.237	7.815	6	10	0.1888
2	559	5.821	7.949	8	5	0.2236
3	425	5.238	7.725	7	2	0.17
4	218	25.528	30.125	12	6	0.0872
5	233	22.232	32.231	11	3	0.0932
6	328	15.129	18.455	10	7	0.1312
7	265	40.125	45.358	15	8	0.106

In the above data, there are seven strata as 1, 2, 3, …. 7:

(i) First three Strata are combined in fist group named as G_1 where $\alpha_p = 1;\ p \in I_1 = \{1, 2, 3\}$, in this group we can use equal allocation,

(ii) Next two Strata are combined in second group named as G_2 where $\alpha_p = W_p;\ p \in I_2 = \{4, 5\}$, in this group we can use proportional allocation,

(iii) Next two Strata are combined in second group named as G_3 where $\alpha_p = \sqrt{A_p / c_p};\ p \in I_3 = \{6, 7\}$, in this group we can use optimum allocation,

Thus $I_1 = \{1, 2, 3\}$, $I_2 = \{4, 5\}$ and $I_3 = \{6, 7\}$.

It is shown in this example that $I_s; s = 1, 2, 3$ are different groups and not overlapping in nature.

For the sake of simplicity we assumed that no characteristics are more important therefore $a_1 = a_2 = 1/2$.

Table 2. Values of A_p, A_p/α_p and $c_p\alpha_p$

p	N_p	s_{1p}	s_{2p}	c_p	t_p	W_p	A_p	α_p	A_p/α_p	$c_p\alpha_p$
1	472	5.237	7.815	6	10	0.1888	1.577	1	1.57732	6
2	559	5.821	7.949	8	5	0.2236	2.427	1	2.42662	8
3	425	5.238	7.725	7	2	0.17	1.259	1	1.25877	7
$p \in I_1$									5.26271	21
4	218	25.528	30.125	12	6	0.0872	5.928	0.0872	67.9809	1.046
5	233	22.232	32.231	11	3	0.0932	6.658	0.0932	71.4424	1.025
$p \in I_2$									139.423	2.072
6	328	15.129	18.455	10	7	0.1312	4.901	0.7001	7.00086	7.001
7	265	40.125	45.358	15	8	0.106	20.6	1.172	17.5796	17.58
$p \in I_3$									24.5804	24.58

$$\left.\begin{array}{ll} \textit{Minimize} & \dfrac{5.2627}{\beta_1} + \dfrac{139.423}{\beta_2} + \dfrac{24.5804}{\beta_3} \\ \textit{subject to} & 21\beta_1 + 2.072\beta_2 + 24.58\beta_3 + 17\sqrt{\beta_1} + 2.6876\sqrt{\beta_2} + 14.5177\sqrt{\beta_3} \le 7000 \\ \textit{and} & \beta_s \ge 0;\, s = 1,2,3 \end{array}\right\}$$

Using (8)

$$f(1, C_1) = \frac{2 \times 441 \times 5.2627}{(289 + 42C_1) \pm 17\sqrt{289 + 84 \times C_1}} \quad \text{at}$$

$$\beta_1 = \frac{(289 + 42C_1) \pm 17\sqrt{289 + 84 \times C_1}}{2 \times 441}$$

The problems is solved using Lingo software and get the solution as,

$$\beta_1 = 62.9021 \qquad \beta_2 = 1046.2620 \quad \beta_3 = 127.1654$$

Using (2) the values of n_p are obtained as

$$n_1 = \beta_1 \alpha_1 = 62.9021 \times 1.00 \qquad \approx 63$$

$$n_2 = \beta_1 \alpha_2 = 62.9021 \times 1.00 \qquad \approx 63$$

$$n_3 = \beta_1 \alpha_3 = 62.9021 \times 1.00 \qquad \approx 63$$

$$n_4 = \beta_2 \alpha_4 = 1046.2620 \times 0.0872 \quad = 91.2340 \quad \approx 91$$

$$n_5 = \beta_2 \alpha_5 = 1046.2620 \times 0.0932 \quad = 97.5116 \quad \approx 98$$

$$n_6 = \beta_3 \alpha_6 = 127.1654 \times 0.7001 \quad = 89.0285 \quad \approx 89$$

$$n_7 = \beta_3 \alpha_7 = 127.1654 \times 1.172 \quad = 149.0378 \quad \approx 149$$

Now $V_{mixed} = \sum_{s=1}^{t} \dfrac{\sum_{p \in I_s} \dfrac{A_p}{\alpha_p}}{\beta_s}$ or $V_{mixed} = 0.4102$

4. Conclusion

In literature, several authors studied the problem of allocation but no one has taken care of travelling cost in multivariate stratified mixed allocation sampling although it plays an important role in cost optimization. The proposed compromise mixed allocation technique used efficiently in practical situation as mentioned in Ahsan *et al.* [3]. The proposed model enables the decision maker to consider the quadratic travel cost when dealing with non-linear multivariate stratified random sampling with more than two characteristics.

Acknowledgement: The authors are thankful to the Referee for valuable comments and suggestions.

References

[1] Ahsan, M.J. and Khan, S.U. (1977). Optimum allocation in multivariate stratified random sampling using prior information. J. Indian Statist. Assoc., **15**, 57-67.

[2] Ahsan, M.J. and Khan, S.U. (1982). Optimum allocation in multivariate stratified random sampling with overhead cost. Metrika, **29**, 71-78.

[3] Ahsan, M.J., Najmussehar and Khan, M.G.M. (2005). Mixed allocation in stratified sampling. Aligarh J. Statist., **25**, 87-97.

[4] Ansari, A. H., Najmussehar and Ahsan, M. J. (2009). On multiple response stratified random sampling design. International Journal of Statistical Sciences, **1**(1), 1-11.

[5] Ansari, A. H., Varshney, R. and Ahsan, M. J. (2018). Compromise Mixed Allocation in Multivariate Stratified Sampling Using Dynamic Programming Technique. Journal of Advance Statistics, **3**(4), 45-89

[6] Aoyama, H. (1963). Stratified random sampling with optimum allocation for multivariate populations. Ann. Inst. Statist. Math., **14**, 251-258.

[7] Arvanitis, L.G. and Afonja, B. (1971). Use of the generalized variance and the gradient projection method in multivariate stratified random sampling. Biometrics, **27**, 119-127.

[8] Bankier, M.D. (1988). Power allocations: Determining sample sizes for sub national areas. Amer. Statist., **42**, 174-177.

[9] Beardwood, J., Halton, J.H., Hammersley, J. M. (1959). The shortest path through many points. Mathematical proceedings of the Cambridge Philosophical Society, **55**: 299-327

[10] Bethel, J. (1989). Sample allocation in multivariate surveys. Survey Methodology, **15**(1), 47-57.

[11] Bhushan, S., Kumar, A. and Singh, S. (2021). Some efficient classes of estimators under stratified sampling. Communications in Statistics - Theory and Methods, DOI: 10.1080/03610926.2021.1939052

[12] Bhushan, S., Kumar, A., Tyagi, D. and Singh, S. (2022). On some improved classes of estimators under stratified sampling using attributes. Journal of Reliability and Statistical Studies, **15**(1), 187-210. DOI: 10.13052/jrss0974-8024.1518

[13] Chatterjee, S. (1967). A note on optimum allocation. Scand. Actuar. J., **50**, 40-44.

[14] Chatterjee, S. (1968). Multivariate stratified surveys. J. Amer. Statist. Assoc., **63**, 530-534.

[15] Clark, R.G. and Steel, D.G. (2000). Optimum allocation of sample to strata and stages with simple additional constraints. The Statistician, **49**(2), 197-207.

[16] Díaz-García, J.A. and Cortez, L.U. (2006). Optimum allocation in multivariate stratified sampling: Multi-objective programming. Comunicación Técnica No. I-06-07/28-03-2006 (PE/CIMAT).

[17] Díaz-García, J.A. and Cortez, L.U. (2008). Multi-objective optimisation for optimum allocation in multivariate stratified sampling. Survey Methodology, **34**(2), 215-222.

[18] Folks, J.L. and Antle, C.E. (1965). Optimum allocation of sampling units to the strata when there are R responses of interest. J. Amer. Statist. Assoc., **60**, 225-233.

[19] Jahan, N., Khan, M.G.M. and Ahsan, M.J. (1994). A generalized compromise allocation. J. Indian Statist. Assoc., **32**, 95-101.

[20] Khan, M.G.M., Ahsan, M.J. and Jahan, N. (1997). Compromise allocation in multivariate stratified sampling: An integer solution. Naval Res. Logist., **44**, 69-79.

[21] Khan, M.G.M., Khan, E.A. and Ahsan, M.J. (2003). An optimal multivariate stratified sampling design using dynamic programming. Aust. N.Z.J. Stat., **45**(1), 107-113.

[22] Kokan, A.R. and Khan, S. (1967). Optimum allocation in multivariate surveys: An analytical solution. J. Roy. Statist. Soc., **B29**(1), 115-125.

[23] Kozak, M. (2006a). On sample allocation in multivariate surveys. Comm. Statist. Simulation Comput., **35**, 901-910.

[24] Kozak, M. (2006b). Multivariate sample allocation: Application of random search method. Statistics in Transition, **7**(4), 889-900.

[25] Kreienbrock, L. (1993). Generalized measures of dispersion to solve the allocation problem in multivariate stratified random sampling. Comm. Statist. Theory Methods, **22**(1), 219-239.

[26] Melaku, A. and Sadasivan, G. (1987). L_1-norm and other methods for sample allocation in multivariate stratified surveys. Comput. Statist. Data Anal., **5**(4), 415-423.

[27] Yates, F. (1960). Sampling methods for Censuses and Surveys. 3rd ed., Charles Griffin and Co. Ltd., London.