J Math Model Algor (2013) 12:373–381
DOI 10.1007/s10852-012-9215-3

Minimum Cost Compromise Mixed Allocation

Rahul Varshney · A. H. Ansari · M. J. Ahsan

Received: 5 April 2012 / Accepted: 9 November 2012 / Published online: 30 December 2012

Abstract When more than one (say p) characteristics in multivariate stratified population are defined on each unit of the population, the individual optimum allocations may differ widely and can not be used practically. Moreover, there may be a situation such that no standard allocation is advisable to all the strata, for one reason or another. In such a situation, Clark and Steel (J R Stat Soc, Ser D Stat 49(2):197–207, 2000) suggested that different allocations may be used for different groups of strata having some common characteristics for double sampling in stratification. Later on, Ahsan et al. (Aligarh J Stat 25:87–97, 2005) used the same concept in univariate stratified sampling. They minimized the variance of the stratified sample mean for a fixed cost to obtain an allocation and called this allocation "mixed allocation". In the present paper, a "compromise mixed allocation" is worked out for the fixed precisions of the estimates of the p-population means of a multivariate stratified population. A numerical example is also presented.

Keywords Multivariate stratified sampling · Optimum allocation · Compromise mixed allocation

Mathematics Subject Classifications (2010) 62D05 · 62H12

R. Varshney (✉) · M. J. Ahsan
Department of Statistics and Operations Research,
Aligarh Muslim University, Aligarh, 202 002, Uttar Pradesh, India
e-mail: itsrahulvarshney@gmail.com

M. J. Ahsan
e-mail: mjahsan2001@yahoo.co.in

A. H. Ansari
Kejriwal Institute of Management and Development Studies,
Namkum, Ranchi, 834010, Jharkhand, India
e-mail: athar_hussain1@rediffmail.com

1 Introduction

In stratified sampling, the use of any particular type of allocation depends on the nature of the population, objectives of survey, the available budget, etc. Practically, there are situations in which all strata of a stratified population do not support the use of any single type of allocation. For example, when no information about some strata of the population is available, "equal" allocation may be used for a given total sample size for that strata. If the only information available for some strata is N_h, the size of the hth stratum, "proportional" allocation may be used, that is, $n_h \propto N_h$. If information about the variability S_h of some strata is available, an allocation $n_h \propto W_h S_h$ may be used. When there is evidence to believe that the relative standard error of the stratum mean $\bar{Y}_h$ based on one sample unit does not vary considerably over some strata, an allocation $n_h \propto W_h \bar{Y}_h$ may be used. As the range R_h of a stratum provides an approximation to the standard deviation in the absence of the knowledge of the stratum standard deviations S_h, the allocation may be taken as $n_h \propto W_h R_h$ (see [26]). When full information about the population is available, the obvious choice is the "optimum allocation." In the above scenario, Ahsan et al. [3] divided the strata into disjoint groups and used different allocations for different groups. They called their allocation as a mixed allocation. Later on, Varshney and Ahsan [28] extended the work of Ahsan et al. [3] for multivariate stratified sampling.

Unless specified otherwise, the notations of Cochran [13] are used in this manuscript.

Let a population, consisting of N units, be divided into L strata of sizes $N_1, N_2, \ldots, N_L; \sum_{h=1}^{L} N_h = N$. An unbiased estimate of the population mean $\bar{Y}$ is given by

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$$

with a variance

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} \quad (\text{ignoring fpc}).$$

Using a linear cost function, the total cost incurred in the survey may be given as

$$C = c_0 + \sum_{h=1}^{L} c_h n_h,$$

where c_h is the per-unit measurement cost of selected unit in the hth stratum, n_h is the sample size from the hth stratum, and c_0 is the overhead cost.

If "B" denotes the available budget of the survey, then we must have $C \leq B$

$$\text{or} \qquad c_0 + \sum_{h=1}^{L} c_h n_h \leq B,$$

$$\text{or} \qquad \sum_{h=1}^{L} c_h n_h \leq B - c_0,$$

$$\text{or} \qquad \sum_{h=1}^{L} c_h n_h \leq C_0,$$

where $C_0 = B - c_0$ denotes the available budget for measurement of the units selected in the stratified sample.

Ahsan et al. [3] formulated the problem of finding the mixed allocation as the following nonlinear programming problem (NLPP):

$$\text{minimize} \qquad F(\alpha_1, \alpha_2, \ldots, \alpha_k) = \sum_{j=1}^{k} \sum_{h \in I_j} \frac{W_h^2 S_h^2}{\alpha_j \beta_h} \tag{1.1}$$

$$\text{subject to} \qquad \sum_{j=1}^{k} \sum_{h \in I_j} \alpha_j c_h \beta_h \leq C_0 \tag{1.2}$$

$$\text{and} \qquad \alpha_j \geq 0; \;\; j = 1, 2, \ldots, k, \tag{1.3}$$

where k different types of allocations are to be used, so that L strata are divided into k groups, and the jth group consists of L_j strata. The sample allocations are given by

$$n_h = \alpha_j \beta_h; \;\; h \in I_j, \;\; j = 1, 2, \ldots, k \tag{1.4}$$

where α_j; $j = 1, 2, \ldots, k$ are the solution to the NLPP (1.1)–(1.3), I_j is the set of integers representing the strata numbers in the jth group and β_h; $h \in I_j$ are the constants depending upon the particular type of allocation used. For example, if equal allocation is to be used in any group, then $\beta_h = 1$ for proportional allocation, $\beta_h = W_h$, etc.

If the measurements of observations are costly and the tolerance limits on the variances of the estimates are available, then the use of an allocation that minimizes the cost of the survey for the fixed precisions of the estimates will be an acceptable compromise criterion for estimating the p population means $\bar{Y}^l$; $l = 1, 2, \ldots, p$.

Ahsan et al. [4] minimized the cost for the fixed precision for univariate case. They formulated the problem as an NLPP:

$$\text{minimize} \qquad F(\alpha_1, \alpha_2, \ldots, \alpha_k) = \sum_{j=1}^{k} \sum_{h \in I_j} \alpha_j c_h \beta_h \tag{1.5}$$

$$\text{subject to} \qquad \sum_{j=1}^{k} \sum_{h \in I_j} \frac{W_h^2 S_h^2}{\alpha_j \beta_h} \leq v \tag{1.6}$$

$$\text{and} \qquad \alpha_j \geq 0; \;\; j = 1, 2, \ldots, k, \tag{1.7}$$

where v is fixed according to the required precision of the estimate.

In this manuscript, the authors presented the multivariate extension of the same problem. It is assumed that the properties of the strata on which the grouping scheme is based are prevalent in the multivariate case also. Since the above-discussed allocation depends on a compromise criterion with the same grouping scheme as used by Ahsan et al. [3], to work out a mixed allocation, the proposed allocation may be called a "compromise mixed allocation".

Various other authors discussed different compromise criteria or explored the existing criteria further to obtain a compromise allocation, like Yates [29], Aoyama [6], Folks and Antle [15], Kokan and Khan [21], Chatterjee [10, 11], Arvanitis and Afonja [7], Ahsan and Khan [1, 2], Melaku and Sadasivan [25], Bankier [8], Bethel [9], Kreienbrock [23], Jahan et al. [16], Khan et al. [17–20], Kozak [22], Ansari et al. [5], and many others.

2 The Formulation

The problem of finding the compromise mixed allocation given in Eq. 1.5–1.7 for multivariate case with the compromise criterion discussed in Section 1 may be given as follows:

$$\text{minimize} \qquad F(\alpha_1, \alpha_2, \ldots, \alpha_k) = \sum_{j=1}^{k} \sum_{h \in I_j} \alpha_j c_h \beta_h \tag{2.1}$$

$$\text{subject to} \qquad \sum_{j=1}^{k} \sum_{h \in I_j} \frac{W_h^2 S_{lh}^2}{\alpha_j \beta_h} \le v_l; \; l = 1, 2, \ldots, p \tag{2.2}$$

$$\text{and} \qquad \alpha_j \ge 0; \; j = 1, 2, \ldots, k, \tag{2.3}$$

where v_l is the tolerance limit for the $V\left(\bar{y}_{st}^l\right)$, $\bar{y}_{st}^l$ is the estimate of population mean $\bar{Y}^l$ of the lth characteristic, S_{lh}^2 is the stratum variance for the lth characteristic, c_h is the cost of measuring all the p characteristics on a selected unit of hth stratum, i.e., $c_h = \sum_{l=1}^{p} c_{lh}$; $h = 1, 2, \ldots, L$, c_{lh} denote the per unit cost of measuring the lth characteristic in the hth stratum. Furthermore, if S_{lh}^2 are unknown, they may be replaced by their sample estimates, s_{lh}^2.

The solution to the NLPP (2.1)–(2.3) may be obtained by using the optimization software LINGO [24]. Once α_j, $j = 1, 2, \ldots, k$ are known, the compromise mixed allocation may be obtained by using Eq. 1.4.

It is to be noted that if optimum allocation is to be used in any group (say kth), then the values of β_h will differ from characteristic to characteristic and is therefore denoted by

$$\beta_{lh} = \frac{W_h S_{lh}}{\sqrt{c_h}}; \; l = 1, 2, \ldots, p; \; h \in I_k. \tag{2.4}$$

To work out a compromise mixed allocation, we need a single value of β_h; $h \in I_k$ for all characteristics. As an arithmetic mean is an ideal average based upon all the observations and affected least by fluctuations of sampling, that is, it is a stable average, it is the most suitable single representative of a set of quantitative observations.

$$\text{Thus,} \qquad \bar{\beta}_h = \frac{1}{p}\sum_{l=1}^{p} \beta_{lh};\, h \in I_k \tag{2.5}$$

may be used as a single representative of the values of $\beta_h;\, h \in k$.

3 A Numerical Illustration

Ahsan et al. [3] gave a numerical illustration using artificial data. For the illustration of the proposed method, we have added another characteristic to that data with the corresponding estimated values of s_h for $l = 2$ as s_{2h}. We fixed the precision for the first and second characteristics as $v_1 = 0.65$ and $v_2 = 0.70$, respectively. For seven strata and two characteristics, the values of N_h, W_h, s_{1h}, s_{2h}, and c_h are given in Table 1.

It is assumed that

(a) Strata 1, 2, and 3 constitute group G_1 in which equal allocation is to be used, that is

$$\beta_h = 1;\;\; h \in I_1 \equiv \{1, 2, 3\}\,;$$

(b) Strata 4 and 5 constitute group G_2 in which proportional allocation is to be used, that is

$$\beta_h = W_h;\;\; h \in I_2 \equiv \{4,\; 5\}\,;\; \text{ and}$$

(c) Strata 6 and 7 constitute group G_3 in which optimum allocation is to be used, that is

$$\beta_{lh} = \frac{W_h s_{lh}}{\sqrt{c_h}};\;\; l = 1, 2;\;\; h \in I_3 \equiv \{6, 7\}\,,$$

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

It can be seen that $I_j;\, j = 1, 2, 3$ are mutually exclusive and exhaustive.

Tables 2 and 3 give the values of $\frac{W_h^2 s_{lh}^2}{\beta_h}$ and $c_h\beta_h$ for first and second characteristics.

Table 1 Data for seven strata and two characteristics

h	N_h	W_h	s_{1h}	s_{2h}	c_h
1	472	0.1888	5.237	7.815	6
2	559	0.2236	5.821	7.949	8
3	425	0.1700	5.238	7.725	7
4	218	0.0872	25.528	30.125	12
5	233	0.0932	22.232	32.231	11
6	328	0.1312	15.129	18.455	10
7	265	0.1060	40.125	45.358	15

Table 2 Values of $W_h^2 s_{1h}^2/\beta_h$ and $c_h\beta_h$ for the first characteristic

h	W_h	s_{1h}	c_h	β_h	$W_h^2 s_{1h}^2/\beta_h$	$c_h\beta_h$
1	0.1888	5.237	6	1	0.97762	6
2	0.2236	5.821	8	1	1.69410	8
3	0.1700	5.238	7	1	0.79292	7
				$\sum_{h\in I_1}$	3.46464	21
4	0.0872	25.528	12	0.08720	56.82639	1.04640
5	0.0932	22.232	11	0.09320	46.06520	1.02520
				$\sum_{h\in I_2}$	102.89159	2.07160
6	0.1312	15.129	10	0.62769	6.27687	6.27690
7	0.1060	40.125	15	1.09818	16.47283	16.47270
				$\sum_{h\in I_3}$	22.74970	22.74960

Since the optimum allocation is used in third group, G_3, the compromise value of $\bar{\beta}_h;\ h \in I_3$, obtained by using Eq. 2.5, are

$$\bar{\beta}_6 = \frac{\beta_{16}+\beta_{26}}{2} = \frac{0.62769+0.76568}{2} = 0.69669,$$

$$\bar{\beta}_7 = \frac{\beta_{17}+\beta_{27}}{2} = \frac{1.09818+1.24141}{2} = 1.16980,$$

$$\text{and} \quad \sum_{h\in I_3} c_h\bar{\beta}_h = (10\times 0.69669) + (15\times 1.16980) = 24.51390.$$

With the values obtained in Tables 2 and 3, the NLPP (2.1)–(2.3) may be expressed as follows:

$$\text{minimize} \qquad 21\,\alpha_1 + 2.07160\,\alpha_2 + 24.51390\,\alpha_3 \tag{3.1}$$

$$\text{subject to} \qquad \frac{3.46464}{\alpha_1} + \frac{102.89159}{\alpha_2} + \frac{22.74970}{\alpha_3} \le 0.65, \tag{3.2}$$

$$\frac{7.06079}{\alpha_1} + \frac{175.95500}{\alpha_2} + \frac{26.27788}{\alpha_3} \le 0.70, \tag{3.3}$$

$$\text{and} \qquad \alpha_1, \alpha_2, \alpha_3 \ge 0. \tag{3.4}$$

Table 3 Values of $W_h^2 s_{2h}^2/\beta_h$ and $c_h\beta_h$ for the second characteristic

h	W_h	s_{2h}	c_h	β_h	$W_h^2 s_{2h}^2/\beta_h$	$c_h\beta_h$
1	0.1888	7.815	6	1	2.17702	6
2	0.2236	7.949	8	1	3.15914	8
3	0.1700	7.725	7	1	1.72463	7
				$\sum_{h\in I_1}$	7.06079	21
4	0.0872	30.125	12	0.08720	79.13536	1.04640
5	0.0932	32.231	11	0.09320	96.81964	1.02520
				$\sum_{h\in I_2}$	175.95500	2.07160
6	0.1312	18.455	10	0.76568	7.65682	7.65680
7	0.1060	45.358	15	1.24141	18.62106	18.62115
				$\sum_{h\in I_3}$	26.27788	26.27795

Using the optimization software LINGO, the solution to NLPP (3.1)–(3.4) is obtained as $\alpha_1 = 46.92621$, $\alpha_2 = 745.84120$, and $\alpha_3 = 83.78909$ with an objective value of 4584.53200.

With these values of α_j; $j = 1, 2, 3$, the compromise mixed allocation is obtained as follows:

For $j = 1$, $n_{1(m)} = n_{2(m)} = n_{3(m)} = \alpha_1 = 46.92621 \approx 47$.

For $j = 2$, $n_{4(m)} = \alpha_2 \beta_4 = \alpha_2 W_4 = 745.84120 \times 0.0872 = 65.03735 \approx 65$ and

$$n_{5(m)} = \alpha_2 \beta_5 = \alpha_2 W_5 = 745.84120 \times 0.0932 = 69.51240 \approx 70.$$

For $j = 3$, $n_{6(m)} = \alpha_3 \bar{\beta}_6 = 83.78909 \times 0.69669 = 58.37502 \approx 58$ and

$$n_{7(m)} = \alpha_3 \bar{\beta}_7 = 83.78909 \times 1.16980 = 98.01648 \approx 98.$$

The total cost under this compromise mixed allocation is 4,587 units.

When mixed allocation is not used, the compromise allocation for fixed precision that minimizes the total cost of the survey will be the solution to the NLPP:

$$\text{minimize} \quad \sum_{h=1}^{L} c_h n_h \tag{3.5}$$

$$\text{subject to} \quad \sum_{h=1}^{L} \frac{W_h^2 s_{lh}^2}{n_h} \le v_l; \, l = 1, 2, \ldots, p \tag{3.6}$$

$$\text{and} \quad \begin{matrix} 2 \le n_h \le N_h \\ n_h \text{ integers}; \;\; h = 1, 2, \ldots, L \end{matrix}. \tag{3.7}$$

It is to be noted that the solution of the NLPP (3.5)–(3.7) is not the optimum allocation for fixed variances because the strata variances S_{lh}^2 are not known; hence, their estimates s_{lh}^2 are used so that this allocation is the modified optimum allocation [27].

For the illustrated example, the NLPP (3.5)–(3.7) will become

$$\text{minimize} \quad 6\,n_1 + 8\,n_2 + 7\,n_3 + 12\,n_4 + 11\,n_5 + 10\,n_6 + 15\,n_7 \tag{3.8}$$

$$\text{subject to} \quad \frac{0.97762}{n_1} + \frac{1.69410}{n_2} + \frac{0.79292}{n_3} + \frac{4.95526}{n_4} + \frac{4.29328}{n_5} + \frac{3.93993}{n_6} + \frac{18.09014}{n_7} \le 0.65 \tag{3.9}$$

$$\frac{2.17702}{n_1} + \frac{3.15914}{n_2} + \frac{1.72463}{n_3} + \frac{6.90060}{n_4} + \frac{9.02359}{n_5} + \frac{5.86267}{n_6} + \frac{23.11636}{n_7} \le 0.70 \tag{3.10}$$

$$\text{and} \quad \begin{matrix} 2 \le n_h \le N_h \\ n_h \text{ integers}; \;\; h = 1, 2, \ldots, 7 \end{matrix}. \tag{3.11}$$

Using LINGO, the solution to the NLPP (3.8)–(3.11) is obtained as $n_1 = 50$, $n_2 = 52$, $n_3 = 41$, $n_4 = 62$, $n_5 = 75$, $n_6 = 63$, and $n_7 = 101$ with a minimum cost of 4,717 units.

4 Conclusion

The numerical example illustrates how the compromise mixed allocation is used to minimize the total cost of the survey. A valid ground to put some of the strata in a particular group out of the three groups considered in this example is the availability of the information about the population parameters to the sampler. Díaz-García and Cortez [14] classified the availability of information as "full", "partial," and "zero". We assume the strata constituting G_1 as the zero information group. Thus, equal allocation is used in this group. G_2 is considered as the partial information group. Thus, proportional allocation is used in it. Finally, G_3 is considered as the full information group to use optimum allocation.

The study shows that when the availability of information in various groups of strata varies from group to group, the compromise mixed allocation gives better result than the modified optimum allocation for fixed precision. However, to assess the practicability of the proposed allocation, a large scale simulation study may be carried out.

Acknowledgements The authors are grateful to the editor and the learned reviewers for their valuable comments and suggestions that helped to revise the manuscript in its present form. The first author is also thankful to CSIR New Delhi, for their financial support for the work.

References

1. Ahsan, M.J., Khan, S.U.: Optimum allocation in multivariate stratified random sampling using prior information. J. Indian Stat. Assoc. **15**, 57–67 (1977)
2. Ahsan, M.J., Khan, S.U.: Optimum allocation in multivariate stratified random sampling with overhead cost. Metrika **29**, 71–78 (1982)
3. Ahsan, M.J., Najmussehar, Khan, M.G.M.: Mixed allocation in stratified sampling. Aligarh J. Stat. **25**, 87–97 (2005)
4. Ahsan, M.J., Najmussehar, Khan, M.G.M.: Mixed allocation for fixed precision. Pure Appl. Math. Sci. **LXIX**(1–2), 93–103 (2009)
5. Ansari, A.H., Najmussehar, Ahsan, M.J.: On multiple response stratified random sampling design. J. Stat. Sci., Kolkata, India **1**(1), 45–54 (2009)
6. Aoyama, H.: Stratified random sampling with optimum allocation for multivariate populations. Ann. Inst. Stat. Math. **14**, 251–258 (1962)
7. Arvanitis, L.G., Afonja, B.: Use of the generalized variance and the gradient projection method in multivariate stratified random sampling. Biometrics **27**, 119–127 (1971)
8. Bankier, M.D.: Power allocations: determining sample sizes for subnational areas. Am. Stat. **42**, 174–177 (1988)
9. Bethel, J.: Sample allocation in multivariate surveys. Surv. Methodol. **15**(1), 47–57 (1989)
10. Chatterjee, S.: A note on optimum allocation. Scand. Actuar. J. **50**, 40–44 (1967)
11. Chatterjee, S.: Multivariate stratified surveys. J. Am. Stat. Assoc. **63**, 530–534 (1968)
12. Clark, R.G., Steel, D.G.: Optimum allocation of sample to strata and stages with simple additional constraints. J. R. Stat. Soc., Ser. D Stat. **49**(2), 197–207 (2000)
13. Cochran, W.G.: Sampling Techniques. Wiley, New York (1977)
14. Díaz-García, J.A., Cortez, L.U.: Optimum Allocation in Multivariate Stratified Sampling: Multi-Objective Programming. Comunicación Técnica No. I-06-07/28-03-2006 (PE/CIMAT), México (2006)

15. Folks, J.L., Antle, C.E.: Optimum allocation of sampling units to the strata when there are R responses of interest. J. Am. Stat. Assoc. **60**, 225–233 (1965)
16. Jahan, N., Khan, M.G.M., Ahsan, M.J.: A generalized compromise allocation. J. Indian Stat. Assoc. **32**, 95–101 (1994)
17. Khan, M.G.M., Ahsan, M.J., Jahan, N.: Compromise allocation in multivariate stratified sampling: an integer solution. Nav. Res. Logist. **44**, 69–79 (1997)
18. Khan, M.G.M., Khan, E.A., Ahsan, M.J.: An optimal multivariate stratified sampling design using dynamic programming. Aust. N. Z. J. Stat. **45**(1), 107–113 (2003)
19. Khan, M.G.M., Khowaja, S., Ghufran, S., Varshney, R., Ahsan, M.J.: A comparative study of various compromise criteria in multiple quantitative response stratified sampling. J. Stat. Sci. **2**(1), 1–12 (2010)
20. Khan, M.G.M., Maiti, T., Ahsan, M.J.: An optimal multivariate stratified sampling design using auxiliary information: an integer solution using goal programming approach. J. Off. Stat. **26**(4), 695–708 (2010)
21. Kokan, A.R., Khan, S.: Optimum allocation in multivariate surveys: an analytical solution. J. R. Stat. Soc., B **29**, 115–125 (1967)
22. Kozak, M.: On sample allocation in multivariate surveys. Commun. Stat., Simul. Comput. **35**(4), 901–910 (2006)
23. Kreienbrock, L.: Generalized measures of dispersion to solve the allocation problem in multivariate stratified random sampling. Commun. Stat., Theory Methods **22**(1), 219–239 (1993)
24. LINDO System Inc.: LINGO-User's Guide. LINDO System Inc., Chicago (2001)
25. Melaku, A., Sadasivan, G.: L_1-norm and other methods for sample allocation in multivariate stratified surveys. Comput. Stat. Data Anal. **5**(4), 415–423 (1987)
26. Murthy, M.N.: Sampling Theory and Methods. Statistical Publishing Society, Kolkata (1967)
27. Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S., Asok, C.: Sampling Theory of Surveys with Applications. Iowa State University Press, Ames. and Indian Society of Agricultural Statistics, New Delhi (1984)
28. Varshney, R., Ahsan, M.J.: Compromise mixed allocation in multivariate stratified sampling. J. Indian Soc. Agric. Stat. **65**(3), 291–296 (2011)
29. Yates, F.: Sampling Methods for Censuses and Surveys. Charles Griffin & Co. Ltd., London (1960)