Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 65(3) 2011 291-296

Compromise Mixed Allocation in Multivariate Stratified Sampling

Rahul Varshney* and M.J. Ahsan

Aligarh Muslim University, Aligarh

Received 09 March 2009; Revised 15 February 2011; Accepted 09 March 2011

SUMMARY

Ahsan *et al.* (2005) introduced the idea of "Mixed Allocation" in stratified sampling. In the present paper the authors worked out the "Compromise Mixed Allocation" for multivariate stratified sampling in which p > 1 characteristics are defined on each population unit. It is assumed that the properties of the strata on which the grouping scheme of Ahsan *et al.* (2005) is based are prevalent in the multivariate case also. A numerical example is also presented to illustrate the computational details.

Keywords: Stratified sampling, Optimum allocation, Mixed allocation, Multivariate stratified sampling, Compromise allocation, Compromise mixed allocation, Relative loss in efficiency.

1. INTRODUCTION

Equal, proportional, optimum and several other allocations are well known in stratified sampling literature. Usually any one type of allocation is selected according to the nature of the population, aim of the survey and available budget and is applied to all the strata. However, there are practical situations in which the strata differ significantly in one or the other respect such that using a single allocation to all strata is not advisable. In such situations, one can divide the strata into non-overlapping and exhaustive groups that are similar in nature internally. A particular type of allocation can then be applied to a particular group of strata depending on the nature of the group. Ahsan *et al.* (2005) worked out the allocation using the above criterion and named it as "Mixed Allocation".

They formulated the problem of finding a mixed allocation as the following nonlinear programming problem (NLPP)

Minimize
$$F(\alpha_j) = \sum_{j=1}^k \sum_{h \in I_j} \frac{W_h^2 S_h^2}{\alpha_j \beta_h}$$
 (1.1)

subject to $\sum_{j=1}^{k} \sum_{h \in I_{j}} \alpha_{j} c_{h} \beta_{h} \leq C_{0}$ (1.2)

and
$$\alpha_i \ge 0; j = 1, 2, ..., k$$
 (1.3)

where L strata are divided in k groups, the j^{th} group consists of L_j strata. The sample allocations are given by

$$n_h = \alpha_i \beta_h; h \in I_i, j = 1, 2, ..., k,$$
 (1.4)

where α_j ; j = 1, 2, ..., k are the solution to NLPP (1.1) - (1.3), I_j is the set of integers representing the strata numbers in the j^{th} group and β_h are fixed according to the particular allocation used. For example if proportional allocation is to be used in the q^{th} group then $\beta_h = W_h$; $h \in I_q$.

In multivariate stratified sampling where p > 1 characteristics are to measured on each selected unit of the sample, an allocation that minimizes the variance of one characteristic may result in significant losses of precision for other characteristics. For such situations Yates (1960) gave the idea to use an allocation that

*Corresponding author: Rahul Varshney
E-mail address: itsrahulvarshney@gmail.com

minimizes the weighted sum of variances of all the p-characteristics for a fixed budget. Since this type of allocation is based on a compromise criterion to have a combined objective instead of several objectives (minimizing the individual variances) it is named as a "Compromise allocation". Since then various authors suggested different compromise criteria or explored further the already existing criteria. Among them are Yates (1960), Aoyama (1963), Folks and Antle (1965), Kokan and Khan (1967), Chatterjee (1967, 1968), Arvanitis and Afonja (1971), Ahsan and Khan (1977, 1982), Melaku and Sadasivan (1987), Bankier (1988), Bethel (1989), Kreienbrock (1993), Jahan et al. (1994), Khan et al. (1997), Khan et al. (2003), Díaz-García and Cortez (2006, 2008), Ansari *et al.* (2009) and many others. Kozak (2006a) discussed five different compromise criteria to work out approximate optimum allocation in multivariate surveys and compared them using a simulation study. Kozak (2006b) gave three different compromise criteria and modified the random search method to develop an algorithm to obtain the compromise allocation for multivariate stratified populations.

Before Ahsan *et al.* (2005) no author used the term "Mixed Allocation" and thus no sampling literature is available on mixed allocation. However, Clark and Steel (2000) used a similar idea in univariate two-stage sampling design. In multivariate case instead of individual optimum allocations usually a compromise allocation is used.

Ahsan *et al.* (2005) worked out the mixed allocation for univariate stratified sampling. In this paper we extended the work of Ahsan *et al.* (2005) for the multivariate case. Thus the present paper presents a combination of Compromise and Mixed allocations. The allocation thus obtained, may be termed as "Compromise Mixed Allocation".

Section 2 of the manuscript gives the formulation and solution of the problem. Section 3 highlights the situation in which the compromise mixed allocation may be used in practice. Section 4 illustrates a numerical example to justify the use of compromise mixed allocation. Section 5 summarizes the comparative performance of the proposed allocation with some other compromise allocations. Section 6 gives the concluding remark on the basis of the results obtained in Sections 4 and 5.

A list of alphabetically arranged references is provided at the end of the manuscript.

2. THE COMPROMISE MIXED ALLOCATION IN MULTIVARIATE STRATIFIED SAMPLING

Using Yates (1960) criterion the problem of finding the mixed allocation given in (1.1) - (1.3) for multivariate case may be expressed as

Minimize
$$\sum_{l=1}^{p} a_l \sum_{j=1}^{k} \sum_{h \in I_j} \frac{W_h^2 S_{lh}^2}{\alpha_j \beta_h}$$
 (2.1)

subject to
$$\sum_{j=1}^{k} \sum_{h \in I_{j}} \alpha_{j} c_{h} \beta_{h} \leq C_{0}$$
 (2.2)

and
$$\alpha_i \ge 0; j = 1, 2, ..., k$$
 (2.3)

where $a_l > 0$ is the weight assigned to the variance of the l^{th} characteristic, S_{lh}^2 is the stratum variance for the l^{th} characteristic and $n_h = \alpha_j \beta_h$. Here-in-after c_h will denote the cost of measuring all the p characteristics

on a selected unit of
$$h^{th}$$
 stratum, that is $c_h = \sum_{l=1}^{p} c_{lh}$; h

= 1, 2, ..., L, c_{lh} denote the per unit cost of measurement for the l^{th} characteristic in the h^{th} stratum. Without loss

of generality we can assume that $\sum_{l=1}^{p} a_l = 1$.

Substituting
$$A_h = W_h^2 \sum_{l=1}^{p} a_l S_{lh}^2$$
; $h = 1, 2, ..., L$ and

rearranging the terms NLPP (2.1) - (2.3) may be restated as

Minimize
$$\sum_{j=1}^{k} \sum_{h \in I_{j}} \frac{A_{h}}{\alpha_{j} \beta_{h}}$$
 (2.4)

subject to
$$\sum_{j=1}^{k} \sum_{h \in I_{j}} \alpha_{j} c_{h} \beta_{h} \leq C_{0}$$
 (2.5)

and
$$\alpha_j \ge 0; j = 1, 2, ..., k$$
 (2.6)

Since the objective function is convex and the constraint is linear at the optimal point the constraint will be active (Ahsan 1976). Taking equality in (2.5) and ignoring restrictions in (2.6) the NLPP (2.4) - (2.6) may be solved by Lagrange multiplier technique.

Define the Lagrangian function ϕ as

$$\phi(\alpha_j, \lambda) = \sum_{j=1}^k \sum_{h \in I_j} \frac{A_h}{\alpha_j \beta_h} + \lambda \left(\sum_{j=1}^k \sum_{h \in I_j} \alpha_j c_h \beta_h - C_0 \right)$$

where λ is the Lagrange multiplier.

Differentiating ϕ with respect to α_j and λ partially and equating the partial derivatives to zero we get the following (k+1) simultaneous equations

$$\frac{\partial \phi}{\partial \alpha_j} = -\sum_{h \in I_j} \frac{A_h}{\alpha_j^2 \beta_h} + \lambda \sum_{h \in I_j} c_h \beta_h = 0;$$

$$j = 1, 2, ..., k \qquad (2.7)$$

$$\frac{\partial \phi}{\partial \lambda} = \sum_{j=1}^{k} \sum_{h \in I_j} \alpha_j c_h \beta_h - C_0 = 0 \qquad (2.8)$$

Solving (k + 1) equations in (2.7) and (2.8) as simultaneous equations we get

$$\alpha_{j} = C_{0} \frac{\sqrt{\left(\sum_{h \in I_{j}} \frac{A_{h}}{\beta_{h}}\right) / \left(\sum_{h \in I_{j}} c_{h} \beta_{h}\right)}}{\sum_{j=1}^{k} \sqrt{\left(\sum_{h \in I_{j}} \frac{A_{h}}{\beta_{h}}\right) \left(\sum_{h \in I_{j}} c_{h} \beta_{h}\right)}}$$
(2.9)

The resulting variance (ignoring fpc) is

$$V_{mixed} = \frac{1}{C_0} \left[\sum_{j=1}^{k} \sqrt{\left(\sum_{h \in I_j} \frac{A_h}{\beta_h} \right) \left(\sum_{h \in I_j} c_h \beta_h \right)} \right]^2$$
(2.10)

3. CRITERION FOR USING COMPROMISE MIXED ALLOCATION

The relative loss of efficiency (R.L.E.) by using different allocations in different groups of strata instead of optimum allocation is

R.L.E. =
$$L(\mathbf{n}) = \frac{V_{mixed} - V_{opt}}{V_{opt}} \times 100\%$$
 (3.1)

where $\mathbf{n} = (n_1, n_2, ..., n_L)$ is the vector of compromise mixed allocation.

 $L(\mathbf{n})$ given by (3.1) will be the sum of the losses of the efficiencies incurred due to various allocations

in different groups of strata. If any particular allocation results in a significant loss of efficiency then it may be replaced by any other more efficient allocation.

4. A NUMERICAL ILLUSTRATION

Ahsan *et al.* (2005) gave a numerical illustration using artificial data. We added another characteristic to that data with the corresponding values of s_h as s_{2h} . Thus we have the following situation

In stratification with seven strata and two characteristics the values of N_h , s_{1h} , s_{2h} and c_h are given in Table 1. It is assumed that the total budget available for measurements is 4000 units.

Table 1. Data for seven strata and two characteristics

h	N_h	s_{1h}	s_{2h}	c_h	W_h	
1	472	05.237	07.815	6	0.1888	
2	559	05.821	07.949	8	0.2236	
3	425	05.238	07.725	7	0.1700	
4	218	25.528	30.125	12	0.0872	
5	233	22.232	32.231	11	0.0932	
6	328	15.129	18.455	10	0.1312	
7	265	40.125	45.358	15	0.1060	

The strata are so numbered that:

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

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

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

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

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

$$\beta_h = \sqrt{A_h/c_h}; h \in I_3 = \{6, 7\}$$

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

11 (1, 2, 3), 1₂ (1, 3) and 1₃ (0, 7).

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

It is also assumed that both the characteristics are equally important that is $a_1 = a_2 = 0.5$.

The computations are shown in Tables 2 and 3.

	Table 2: Values of \mathcal{H}_h , \mathcal{H}_h \mathcal{P}_h and \mathcal{P}_h										
h	A_h c_h		$eta_{\!\scriptscriptstyle h}$	A_h/β_h	$c_h \beta_h$						
1	1.5773	6	1.0000	1.5773	6.0000						
2	2.4266	8	1.0000	2.4266	8.0000						
3	1.2588	7	1.0000	1.2588 7.000							
	h	∈ <i>I</i> ₁	5.2627	21.000							
4	5.9279	12	0.0872	67.9805	1.0464						
5	5 6.6584 11 0.0932		0.0932	71.4421	1.0252						
	h	∈ <i>I</i> ₂	139.4226	2.0716							
6	4.9013	10	0.7001	7.0009	7.0010						
7	7 20.6032 15 1.1720		1.1720	17.5795	17.5800						
	h	∈ I ₂	24.5804	24.5810							

Table 2. Values of A_h , A_h/β_h and $c_h\beta_h$

Table 3. Computation of α_i

Group	(A)	(B)	(C)	(D)	$\alpha_j = 4000 \times$	
No. j	$\sum_{h\in I_j} A_h/\beta_h$	$\sum_{h\in I_j}c_h\beta_h$	$\sqrt{(A)/(B)}$	$\sqrt{\left(A\right)\left(B\right)}$	$\frac{(C)}{\sum(D)}$	
1	5.2627	21.0000	0.5006	10.5127	038.4424	
2	139.4226	2.0716	8.2038	16.9949	629.9918	
3	24.5804	24.5810	1.0000	24.5807	076.7927	

With the values of α_j ; j = 1, 2, 3 given in last column of Table 3, the mixed allocation is obtained as:

For
$$j = 1$$
 $n_{1(m)} = n_{2(m)} = n_{3(m)} = \alpha_1 = 38.4424$
For $j = 2$ $n_{4(m)} = \alpha_2 \beta_4 = \alpha_2 W_4 = 629.9918 \times 0.0872$
 $= 54.9353$
 $n_{5(m)} = \alpha_2 \beta_5 = \alpha_2 W_5 = 629.9918 \times 0.0932$
 $= 58.7152$

For
$$j = 3$$
 $n_{6(m)} = \alpha_3 \beta_6 = \alpha_3 \sqrt{A_6/c_6}$
= $76.7927 \times 0.7001 = 53.7626$
 $n_{7(m)} = \alpha_3 \beta_7 = \alpha_3 \sqrt{A_7/c_7}$

The estimated variance under mixed allocation given by (2.10) is

 $= 76.7927 \times 1.1720 = 90.0010$

$$v_{mixed} = 0.6783$$

Table 4. Sample sizes under over all optimum allocation

h	W_h	A_h	c_h	$\sqrt{A_h/c_h}$	$\sqrt{A_h c_h}$	n _{h(opt)}
1	0.1888	1.5773	6	0.5127	3.0763	39.4205
2	0.2236	2.4266	8	0.5507	4.4060	42.3422
3	0.1700	1.2588	7	0.4241	2.9684	32.6082
4	0.0872	5.9279	12	0.7028	8.4341	54.0369
5	0.0932	6.6584	11	0.7780	8.5582	59.8189
6	0.1312	4.9013	10	0.7001	7.0009	53.8293
7	0.1060	20.6032	15	1.1720	17.5798	90.1128

Table 4 gives the sample sizes when overall optimum allocation is used. These values are required to work out R.L.E.

The estimated variance under optimum allocation is given by

$$v_{opt} = \frac{\left(\sum_{h=1}^{L} \sqrt{A_h c_h}\right)^2}{C_0} = \frac{(52.0237)^2}{4000} = 0.6766.$$

5. THE PERFORMANCE OF COMPROMISE MIXED ALLOCATION AS COMPARED TO SOME OTHER ALLOCATIONS

In this Section a comparative study of the proposed compromise mixed allocation has been made with three other well known compromise allocations viz. Cochran's Average Allocation (Cochran 1977), Chatterjee's Compromise Allocation (Chatterjee 1967) and Compromise Allocation for "Minimizing Trace" (Sukhatme *et al.* 1984). However, these compromise allocations assume that the values of W_h and S_h^2 are known for all strata.

The Cochran's Average Compromise Allocation (ACA)

The individual optimum allocations n_{lh}^* are given by

$$n_{lh}^* = \frac{C_0 W_h S_{lh} / \sqrt{c_h}}{\sum_{h=1}^L W_h S_{lh} \sqrt{c_h}}; l = 1, 2.$$

The average compromise allocation $n_{h(ACA)}$ is given by

$$n_{h(ACA)} = \frac{1}{p} \sum_{l=1}^{p} n_{lh}^*$$
; $h = 1, 2, ..., L$.

Chatterjee's Compromise Allocation (CCA)

Chatterjee's compromise allocation $n_{h(CCA)}$, obtained by minimizing the sum of the relative increases in the variances of the estimates is given as

$$n_{h(CCA)} = \frac{C_0 \sqrt{\sum_{l=1}^{p} n_{lh}^{*^2}}}{\sum_{h=1}^{L} c_h \sqrt{\sum_{l=1}^{p} n_{lh}^{*^2}}}; h = 1, 2, ..., L.$$

Sukhatme's Compromise Allocation (SCA)

This compromise allocation $n_{h(SCA)}$, obtained by minimizing the trace of the variance-covariance matrix is given by

$$n_{h(SCA)} = \frac{C_0 W_h \sqrt{\sum_{l=1}^p S_{lh}^2 / c_h}}{\sum_{h=1}^L W_h \sqrt{c_h \sum_{l=1}^p S_{lh}^2}}; h = 1, 2, ..., L.$$

These compromise allocations, for the data used in Section 4, are worked out and listed in Table 5.

Table 5 gives the rounded off values of the Cochran's, Chatterjee's, Sukhatme's and the Proposed compromise allocations, the variances v_1 and v_2 of the estimates of the two characteristics under study, the Trace $(v_1 + v_2)$ and the total cost incurred. It can be seen that the proposed allocation is almost as precise as the other allocations (in terms of the 'Trace') that assume the knowledge of the true values of W_h and S_h^2 for all strata. Whereas the proposed allocation may be used in

relaxed conditions as given elsewhere in this manuscript.

6. CONCLUSION

Since the estimated relative loss in efficiency of the compromise mixed allocation as compared to the overall optimum allocation is

(R.L.E.)_{mixed} =
$$\frac{v_{mixed} - v_{opt}}{v_{opt}} \times 100\%$$

= $\frac{0.6783 - 0.6766}{0.6766} \times 100\%$
= 0.2513%

and is very small, the proposed compromise mixed allocation may be used without any significant loss in the efficiency. In addition to the above fact the compromise mixed allocation also works well in comparison with other compromise allocations discussed in Section 5.

ACKNOWLEDGEMENTS

The authors are grateful to the Editor JISAS and the learned referee for their encouraging remarks and valuable suggestions on the basis of which the manuscript has been revised thoroughly to make it publishable.

REFERENCES

Ahsan, M.J. (1976). A procedure for the problem of optimum allocation in multivariate stratified random sampling. *Aligarh Bull. Math.*, **5-6**, 37-42.

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.

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

Table 5. Various rounded off compromise allocations and the corresponding variances

h	Allocations	n_1	n_2	n_3	n_4	n_5	n_6	n_7	v_1	v_2	Trace = $v_1 + v_2$	Cost incurred
1	Cochran's	39	42	32	54	59	54	91	0.52647	0.82826	1.35473	3996
2	Chatterjee's	39	42	32	54	59	54	91	0.52647	0.82826	1.35473	3996
3	Minimizing Trace	39	42	33	54	60	54	90	0.52671	0.82689	1.35360	3999
4	Proposed	38	38	38	55	59	54	90	0.52800	0.82963	1.35763	3997

- Ahsan, M.J., Najmussehar and Khan, M.G.M. (2005). Mixed allocation in stratified sampling. *Aligarh J. Statist.*, **25**, 87-97.
- Ansari, A.H., Najmussehar and Ahsan, M.J. (2009). On multiple response stratified random sampling design. *J. Statist. Sci.*, Kolkata, India, **1(1)**, 45-54.
- Aoyama, H. (1963). Stratified random sampling with optimum allocation for multivariate populations. *Ann. Inst. Statist. Math.*, **14**, 251-258.
- 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.
- Bankier, M.D. (1988). Power allocations: Determining sample sizes for sub national areas. *Amer. Statist.*, **42**, 174-177.
- Bethel, J. (1989). Sample allocation in multivariate surveys. *Survey Methodology*, **15(1)**, 47-57.
- Chatterjee, S. (1967). A note on optimum allocation. *Scand. Actuar. J.*, **50**, 40-44.
- Chatterjee, S. (1968). Multivariate stratified surveys. J. Amer. Statist. Assoc., 63, 530-534.
- 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.
- Cochran, W.G. (1977). Sampling Techniques. 3rd ed., John Wiley, New York.
- Díaz-García, J.A. and Cortez, L.U. (2006). Optimum allocation in multivariate stratified sampling: Multiobjective programming. *Comunicación Técnica No. I-*06-07/28-03-2006 (PE/CIMAT).
- 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.

- 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.
- Jahan, N., Khan, M.G.M. and Ahsan, M.J. (1994). A generalized compromise allocation. *J. Indian Statist. Assoc.*, **32**, 95-101.
- 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.
- 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.
- Kokan, A.R. and Khan, S. (1967). Optimum allocation in multivariate surveys: An analytical solution. *J. Roy. Statist. Soc.*, **B29(1)**, 115-125.
- Kozak, M. (2006a). On sample allocation in multivariate surveys. *Comm. Statist. Simulation Comput.*, **35**, 901-910.
- Kozak, M. (2006b). Multivariate sample allocation: Application of random search method. *Stat. Trans.*, **7(4)**, 889-900.
- 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.
- Melaku, A. and Sadasivan, G. (1987). L₁-norm and other methods for sample allocation in multivariate stratified surveys. *Comput. Statist. Data Anal.*, **5(4)**, 415-423.
- Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1984). *Sampling Theory of Surveys with Applications*. 3rd ed., Iowa State University Press, Iowa, U.S.A. and Indian Society of Agricultural Statistics, New Delhi, India.
- Yates, F. (1960). Sampling Methods for Censuses and Surveys. 3rd ed., Charles Griffin and Co. Ltd., London.