On Compromise Mixed Allocation in Multivariate Stratified Sampling with Random Parameters

Ummatul Fatima · Rahul Varshney · Najmussehar Siddiqui · Mohammad Jameel Ahsan

Received: 8 April 2013 / Accepted: 11 October 2013 / Published online: 3 December 2013
© Springer Science+Business Media Dordrecht 2013

Abstract For estimating the population mean Clark and Steel (Stat. 49, 1970–207 2000) worked out the optimum allocation of sample sizes to strata and stages with simple additional constraints to use different type of allocations in different strata. Ahsan et al. (Aligarh J. Statist. 25, 87–97 2005) used the same idea to work out optimum allocation in univariate stratified sampling and called it a ‘Mixed Allocation’. Later on Varshney and Ahsan (J. Indian Soc. Agr. Stat. 65(3), 291–296, 2011) extended this work for multivariate stratified sampling and called it a compromise mixed allocation. This article presents a more realistic approach to the compromise mixed allocation by formulating the problem as a Stochastic Nonlinear Programming Problem in which the stratum-wise measurement costs and the sample stratum standard deviations are independent random variables with known probability distributions. The application of this approach is exhibited through a numerical example with normal distributions of the random parameters. The proposed compromise mixed allocation is compared with some other well known compromise allocations available in multivariate stratified sampling literature. It is found that the author’s proposed compromise mixed allocation is the most efficient allocation among the discussed allocations. A simulation study is also carried out to support the claim made by the authors on the basis of the results of the numerical example.

Keywords Multivariate stratified surveys · Mixed allocation · Stochastic programming

U. Fatima (✉) · N. Siddiqui · M. J. Ahsan
Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh 202002, India
e-mail: fatimau2011@yahoo.com

N. Siddiqui
e-mail: gulgasht99@yahoo.com

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

R. Varshney
Department of Mathematics, Galgotias University, Greater Noida, India
e-mail: itsrahulvarshney@gmail.com
1 Introduction

In stratified sampling usually a single allocation is selected and applied to all the strata to select a representative sample of the population. But practical experience suggests that when the complete information about all the strata are not available a single allocation may not be advisable. For example the application of optimum allocation to a stratum requires the knowledge of stratum weight $W_h$, stratum variance $S^2_h$ and per unit measurement cost $c_h$. If any of these values are not known for a particular stratum optimum allocation cannot be applied to that stratum. Similarly if true values of $W_h$ are not known, proportional allocation can not be applied.

In such situations Ahsan et al. [1] proposed the use of “Mixed Allocation” in univariate stratified sampling using the idea of Clark and Steel [3]. Later on Varshney and Ahsan [14] extended their work for multivariate stratified populations. Usually, the true values of $S^2_h$ are unknown and their sample estimates $s^2_h$ are used instead. Being a sample statistic $s^2_h$ based on independent simple random samples are also independent random variables. Furthermore, the per unit measurement cost $c_h$ may also vary during the course of the survey because of the random causes hence it can also be treated as a random variable.

Taking the above facts into consideration in this paper the problem of finding a compromise mixed allocation in multivariate stratified sampling is viewed as a Stochastic Nonlinear Programming Problem (SNLPP). Where the per unit measurement costs $c_h$ and sample variances $s^2_h$ are random variables with known distributions.

In the following the approaches of Ahsan et al. [1] and Varshney and Ahsan [14] are summarised for the sake of continuity.

Ahsan et al. [1] divided the $L$ strata into $k$ disjoint groups according to the available information. (See Kozak [7, 8]). The sample allocations $n_h$ are defined as

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

(1.1)

where $I_j ; j = 1, 2, \ldots, k$ is the set of indices of the strata constituting the group $G_j$, $\beta_h$ ; $h \in I_j; \ j = 1, 2, \ldots, k$ are known constants depending upon the type of allocation to be used in the group $G_j$ and $\alpha_j$ ; $j = 1, 2, \ldots, k$ are the decision variables to be determined.

Ahsan et al. [1] formulated the problem as

Minimize $F(\alpha_j) = \sum_{j=1}^{k} \sum_{h \in I_j} \frac{W^2_h S^2_h}{\alpha_j \beta_h}$

(1.2)

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

(1.3)

and $\frac{2}{\beta_h} \leq \alpha_j \leq \frac{N_h}{\beta_h}; \ h \in I_j; \ j = 1, 2, \ldots, k$

(1.4)

where $C_0 = C - c_0$ and restrictions in Eq. 1.4 are equivalent to $2 \leq n_h \leq N_h$. 

Springer