

A multiple response stratified sampling design with travel cost

Shazia Ghufuran¹, Saman Khowaja¹, Najmussehar¹ and M. J. Ahsan¹

¹Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh, India.

Abstract

In developing the theory of stratified sampling usually the cost function is taken as a linear function of sample sizes n_h considering the measurement and the overhead costs only. In many practical situations the linear cost function does not approximate the actual cost incurred adequately. For example when the cost of traveling between the units selected in the sample within a stratum is significant, instead of linear cost function a cost function that is quadratic in $\sqrt{n_h}$ will be a more close approximation to the actual cost. In this paper the problem of finding a compromise allocation for a multiple response stratified sample survey with a significant travel cost within strata is formulated as a multiobjective non linear programming problem. A solution procedure is proposed using the goal programming approach. A numerical example is also presented to illustrate the computational details.

Keywords: Multiple response, Travel cost, Compromise allocation, Multiobjective programming, Goal programming.

1 Introduction

In stratified sampling the population of N units is first divided into L non-overlapping and exhaustive subpopulation called strata, of sizes $N_1, N_2, \dots, N_h, \dots, N_L$

with $\sum_{h=1}^L N_h = N$.

Consider a population of size N divided into L non-overlapping strata of sizes N_1, N_2, \dots, N_L . Let simple random samples of sizes n_1, n_2, \dots, n_L be drawn to construct the estimators of the unknown population parameters. The problem of determining sample sizes n_h ; $h = 1, 2, \dots, L$ is called the problem of allocation in stratified sampling literature. The total cost C incurred in a sample survey is a function of sample allocations n_h ; $h = 1, 2, \dots, L$. The simplest form of the cost function used in a stratified sample survey is a linear function of sample sizes n_h given as

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

where c_h ; $h = 1, 2, \dots, L$ denote the per unit cost of measurement in the h -th stratum and c_0 denotes the overhead cost. Other cost functions are also used for example Csenki (1997) used the cost function as

$$C = c_0 + \sum_{h=1}^L c_h n_h^\delta, \quad (2)$$

where $\delta > 0$ is a real number and c_h and c_0 are as defined above.

Usually the allocations n_h are worked out to minimize the variance $V(\bar{y}_{st})$ for a fixed total cost C of the survey or to minimize the total cost of the survey for a fixed precision of the estimate. An allocation obtained as above is called an optimum allocation.

Stuart (1954), using Cauchy-Schwarz inequality, showed for the optimum allocation with a linear cost

function of the form given in (1), we must have

$$\frac{n_h \sqrt{c_h}}{W_h S_h} = \text{Constant}; \quad h = 1, 2, \dots, L.$$

This gives the sample size n_h in h^{th} stratum as:

$$n_h = n \frac{W_h S_h / \sqrt{c_h}}{\sum_{h=1}^L W_h S_h / \sqrt{c_h}}; \quad h = 1, 2, \dots, L, \quad (3)$$

where $n = \sum_{h=1}^L n_h$ denote the total sample size (Cochran,

1977). The practical experience suggests that the linear cost function given in (1) may be used as a close approximation to the actual cost in most of the stratified sample surveys. It can also be noted that (1) is a special case of (2) with $\delta = 1$.

To collect the information from the units selected in the sample from a particular stratum the investigator has to travel from unit to unit. If the stratum consists of large geographical and difficult to travel area it may be costly to travel between the selected units. In this situation the linear cost function given in (1) will not be an adequate approximation to the actual cost incurred. The investigator will have to spend a significant amount on travel between the selected units. Beardwood *et al.* (1959) suggested that the cost of visiting the n_h selected units in the h -th stratum may be taken as $t_h \sqrt{n_h}$; $h = 1, 2, \dots, L$ approximately, where t_h is the travel cost per unit in the h -th stratum. This conjecture is based on the fact that the distance between k randomly scattered points is proportional to \sqrt{k} .

Under the above situation the total cost of a stratified sample survey will be the sum of (i) the overhead cost, (ii) the measurement cost, and (iii) the travel cost. This gives the total cost C as:

$$C = c_0 + \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h}, \quad (4)$$

which is quadratic in $\sqrt{n_h}$.

The study presented in this paper shows that the problem of optimum allocation with quadratic cost cannot be solved using the classical Lagrange Multipliers Technique as solved in case of linear cost. The problem is formulated as a neat Nonlinear Programming Problem that can be handled by available optimization softwares. Furthermore, this study presents a multivariate version of this problem also that is of great practical importance because in actual practice usually in sample surveys a large number of characteristics are measured on each unit selected in the sample. When the travel cost is significant and varies from stratum to stratum, that is the cost function is as given in (4) the problem of finding the optimum allocation may be given as the following Nonlinear Programming Problem (NLPP):

$$\left. \begin{aligned} \text{Minimize } V(\bar{y}_{st}) &= \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_h^2 \\ \text{Subject to } \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} &\leq C_0, \\ \text{and } n_h &\geq 0; \quad h = 1, 2, \dots, L, \end{aligned} \right\} \quad (5)$$

where $C_0 = C - c_0$, is the cost available to meet the travel and measurement expenses. For solving the NLPP (5) if Lagrange Multipliers Technique is used one has to take the cost constraint as an equation and to ignore the non-negativity restrictions. The Lagrangian function is defined as

$$L(n_h, \lambda) = \sum_{h=1}^L \frac{W_h^2 S_h^2}{n_h} + \lambda \left(\sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} - C_0 \right), \quad (6)$$

where λ is the Lagrange Multiplier.

Differentiating (6) with respect to n_h ; $h = 1, 2, \dots, L$ and λ partially and equating to zero, we get the following (L+1) equations as:

$$\frac{\partial L(n_h, \lambda)}{\partial n_h} = -\frac{W_h^2 S_h^2}{n_h^2} + \lambda \left(c_h + \frac{t_h}{2\sqrt{n_h}} \right) = 0; \quad h = 1, 2, \dots, L, \quad (7)$$

$$\frac{\partial L(n_h, \lambda)}{\partial \lambda} = \left(\sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} - C_0 \right) = 0. \quad (8)$$

Equations (7) are implicit equations in n_h , therefore, the exact solution of the system of equations (7) and (8) is not possible. However, an approximate solution may be obtained by using approximation methods like Newton Raphson's Method. In the absence of an explicit solution, when the numerical values of W_h, S_h, c_h, t_h, c_0 and C are available, we can use the software package 'LINGO' to solve the NLPP (5). LINGO is a user's friendly package for constrained optimization developed by LINDO Systems Inc. A user's guide-LINGO User's Guide (2001) is also available. For more information one can visit the site <http://www.lindo.com>.

In sample surveys usually several characteristics are to be measured on each selected unit of the sample. Such surveys are called "Multivariate or Multiple Response Surveys". The problem of allocation for a multivariate

stratified survey becomes complicated because an allocation that is optimal for one characteristic is usually far from optimal for other characteristics unless the characteristics are highly correlated. When the characteristics are highly correlated one may work out the characteristic-wise average of the individual optimum allocations for various strata and may use it for all characteristics. When the characteristics are uncorrelated there will be no obvious compromise.

In such situations the sampler may use an allocation based on some compromise criterion that is optimum for all characteristics in some sense. In sampling literature these allocations are called compromise allocations. Among the authors who gave new compromise criterion or explored further the already existing compromise criteria are Neyman (1934), Peter and Bucher (undated), Geary (1949), Dalenius (1957), Ghosh (1958), Yates (1960), Aoyama (1963), Folk and Antle (1965), Chatterjee (1967, 1968), Kokan and Khan (1967), Ahsan (1975-76, 1978), Ahsan and Khan (1977), Schittkowski (1985, 1986), Bethel (1985, 1989), Chromy (1987), Jahan *et al.* (1994, 2001), Jahan and Ahsan (1995), Khan *et al.* (1997), Bosch and Wildner (2003), Singh (2003), Khan *et al.* (2003, 2008), Khan *et al.* (2010), Khowaja *et al.* (2011).

With the advancement of Mathematical Programming Techniques, Multiobjective Programming emerged as a strong tool to deal with the simultaneous optimization of more than one objective functions. Authors like Kozak (2006), Díaz-García and Ulloa (2006, 2008) and some others discussed the problem of optimum allocation in multivariate stratified surveys as a multiobjective programming problem and suggested techniques to solve them.

Usually the travel cost within the strata to approach the selected units for measurement is ignored while constructing the cost function. There are practical situations where the travel cost is significant and thus cannot be ignored. In the present paper we assume that the characteristics are uncorrelated and the cost of traveling (t_h) within stratum to contact the selected units is significant. That is, the cost function is of the form as given in (4) that is quadratic in $\sqrt{n_h}$. The problem of allocation in multivariate stratified sample surveys with p -independent characteristics is formulated as a multiobjective NLPP. The ' p ' objectives are to minimize the individual variances of the estimates of the population means of p -characteristics simultaneously, subject to the cost constraint. The formulated multiobjective NLPP is solved by "Goal Programming Technique" using software package LINGO.

2 Formulation of the Problem

The Multiobjective Non-linear Programming Problem (MNLPP) discussed in the previous section may be expressed as:

$$\left. \begin{array}{l} \text{Minimize } \left\{ \begin{array}{l} V(\bar{y}_{1st}) \\ V(\bar{y}_{2st}) \\ \vdots \\ V(\bar{y}_{pst}) \end{array} \right\} \\ \text{Subject to } \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} \leq C_0, \\ \text{and } 2 \leq n_h \leq N_h; \quad h=1,2,\dots,L. \end{array} \right\} \quad (9)$$

$$\text{where } V(\bar{y}_{jst}) = \sum_{h=1}^L \frac{W_h^2 S_{jh}^2}{n_h}; \quad j=1,2,\dots,p, \quad (10)$$

denote the sampling variance of the estimate

$$\bar{y}_{jst} = \sum_{h=1}^L W_h \bar{y}_{jh} \quad (11)$$

ignoring fpc of the overall population mean \bar{Y}_j ; $j=1,2,\dots,p$ of the j -th characteristic.

$\bar{y}_{jh} = \frac{1}{n_h} \sum_{k=1}^{n_h} y_{jhk}$ is the sample mean from the h -th

stratum for the j -th characteristic and y_{jhk} is the value of the k -th selected unit of the sample from the h -th stratum for the j -th characteristic;

$k=1,2,\dots,n_h$; $h=1,2,\dots,L$; $j=1,2,\dots,p$.

The restrictions $2 \leq n_h \leq N_h$; $h=1,2,\dots,L$ are introduced to obtain the estimates of the stratum variances and to avoid the problem of oversampling.

It is assumed that the true values of S_{jh}^2 are known. In practice, if not known, some approximation of these parameters obtained in some recent or preliminary survey, may be substituted in their place.

3 The Goal Programming Approach

To solve the problem (9) using goal programming, we first solve the following p Non Linear Programming Problems (NLPPs) for all the ' p ' characteristics separately

$$\left. \begin{array}{l} \text{Minimize } V(\bar{y}_{jst}) = \sum_{h=1}^L \frac{W_h^2 S_{jh}^2}{n_h} \\ \text{Subject to } \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} \leq C_0, \\ \text{and } 2 \leq n_h \leq N_h; \quad h=1,2,\dots,L. \end{array} \right\}; j=1,2,\dots,p \quad (12)$$

Let $\underline{n}_j^* = (n_{j1}^*, n_{j2}^*, \dots, n_{jL}^*)$ denote the solution to the j -th NLPP in (12) with V_j^* as the value of the objective function given by

$$V_j^* = \sum_{h=1}^L \frac{W_h^2 S_{jh}^2}{n_{jh}^*}; \quad j=1,2,\dots,p. \quad (13)$$

Further, let $\underline{n}_c^* = (n_{1c}^*, n_{2c}^*, \dots, n_{Lc}^*)$ be the vector of optimum compromise allocations with

$$V_{cj}^* = \sum_{h=1}^L \frac{W_h^2 S_{jh}^2}{n_{hc}^*} \quad \text{as the optimal value of the objective}$$

function for j -th characteristics under this allocation. Obviously,

$$V_{cj}^* \geq V_j^* \quad \text{or} \quad V_{cj}^* - V_j^* \geq 0; \quad j=1,2,\dots,p. \quad (14)$$

A reasonable criterion to workout a compromise allocation may be to "Minimize the sum of increases in the variances V_j ; $j=1,2,\dots,p$ due to the use of the compromise allocation". We may express the multiobjective NLPP (9) using (14) with the above compromise criterion as the following single objective NLPP

$$\left. \begin{array}{l} \text{Minimize } \sum_{j=1}^p x_j \\ \text{Subject to } V_{cj} - V_j^* \leq x_j, \\ \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} \leq C_0, \\ x_j \geq 0 \\ \text{and } 2 \leq n_h \leq N_h; \quad h=1,2,\dots,L \end{array} \right\}; j=1,2,\dots,p \quad (15)$$

where $\underline{n}_c = (n_{1c}, n_{2c}, \dots, n_{Lc})$ denotes a compromise allocation with variance

$$V_{cj} = \sum_{h=1}^L \frac{W_h^2 S_{jh}^2}{n_{ch}}; \quad j=1,2,\dots,p \quad (16)$$

and $x_j \geq 0$; $j=1,2,\dots,p$ are called goal variables whose values are to be determined.

The 'Goal' is to "Find the compromise allocation $\underline{n}_c^* = (n_{1c}^*, n_{2c}^*, \dots, n_{Lc}^*)$ such that the increases in the j -th variance due to the use of compromise allocation should not exceed x_j ; $j=1,2,\dots,p$ and $\sum_{j=1}^p x_j$ is minimum."

NLPP (15) may be restated as

$$\left. \begin{array}{l} \text{Minimize } \sum_{j=1}^p x_j \\ \text{Subject to } \sum_{h=1}^L \frac{W_h^2 S_{jh}^2}{n_h} - x_j \leq V_j^*, \\ \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} \leq C_0, \\ x_j \geq 0 \\ \text{and } 2 \leq n_h \leq N_h; \quad h=1,2,\dots,L, \end{array} \right\} \quad j=1,2,\dots,p. \quad (17)$$

where the value of V_{cj} is substituted from (16) and the compromise allocation n_{hc} ; $h=1,2,\dots,L$ is replaced by n_h for simplicity. The optimal solution to the NLPP (17)

will be the required optimum compromise allocation n_c^* that minimizes the sum of deviations of the variances from their optimum values. NLPP (17) may be solved by using software package LINGO. In the next section a numerical example is given to illustrate the Goal Programming Approach.

4 Some Other Compromise Allocations with Quadratic Cost

In this section three other compromise allocations are discussed for the sake of comparison with the proposed allocation.

4.1 The Proportional Allocation for Fixed Quadratic Cost

Because of its simplicity the proportional allocation is the most commonly used allocation in stratified sample surveys. In the proportional allocation the sample size from the h -th stratum is proportional to its stratum weights that is

$$n_h \propto W_h; \quad h = 1, 2, \dots, L$$

or

$$n_h = kW_h; \quad h = 1, 2, \dots, L \quad (18)$$

where $k > 0$ is the constant of proportionality.

From (18)

$\sum_{h=1}^L n_h = k \sum_{h=1}^L W_h$ or $n = k$, where n is the total sample size. Thus (18) gives

$$n_h = nW_h; \quad h = 1, 2, \dots, L. \quad (19)$$

To work out the value of the total sample size n for fixed cost we proceed as follows. Substitution of the values of n_h from (18) in the cost function (4) with $C_0 = C - c_0$ that is, in

$$C_0 = \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h}, \quad \text{we get}$$

$$C_0 = n \sum_{h=1}^L c_h W_h + \sqrt{n} \sum_{h=1}^L t_h \sqrt{W_h}, \quad (20)$$

The RHS of (20) is quadratic in \sqrt{n} . Putting

$$\sum_{h=1}^L c_h W_h = A > 0, \quad \sum_{h=1}^L t_h \sqrt{W_h} = B > 0 \quad \text{and} \quad \sqrt{n} = X > 0 \quad \text{in (20)}$$

we get a quadratic equation in X as:

$$AX^2 + BX - C_0 = 0, \quad (21)$$

with roots

$$X = \frac{-B \pm \sqrt{B^2 + 4AC_0}}{2A}$$

As $X > 0$ we have the only usable root as

$$X = \frac{-B + \sqrt{B^2 + 4AC_0}}{2A} \quad (22)$$

The RHS of (22) will be positive if and only if

$$\left(\sqrt{B^2 + 4AC_0}\right) > B \quad \text{which is true because } 4AC_0 > 0.$$

When the numerical values of A, B and C_0 are available we can easily compute the value X . X^2 will

give the total sample size n . Substitution of the value of n in (19) gives the proportional allocation.

4.2 Cochran's Compromise Allocation with Quadratic Cost

Cochran (1977) gave the compromise criteria by averaging the individual optimum allocations n_{jh}^* that are solutions to NLPP (17) for $j = 1, 2, \dots, p$, over the characteristics.

Cochran's compromise allocation is given by

$$n_h = \frac{1}{p} \sum_{j=1}^p n_{jh}^*. \quad (23)$$

4.3 Minimizing Weighted Sum of Variances with Quadratic Cost

To work out a compromise allocation Khan *et al.* (2003) used the compromise criteria as

“Minimize $\sum_{j=1}^p a_j V_j$, where $a_j > 0$ is the weights assigned to V_j ”.

The above compromise criterion was first used by Yates (1960). Khan *et al.* (2003) conjectured that

$$a_j = \frac{\sum_{h=1}^L S_{jh}^2}{\sum_{j=1}^p \sum_{h=1}^L S_{jh}^2}. \quad \text{It can be seen that } \sum_{j=1}^p a_j = 1.$$

To compare their allocation with the proposed allocation the Khan *et al.* (2003) compromise allocation is worked out with a quadratic cost function in the following. The objective function of this problem may be expressed as:

$$\begin{aligned} Z(n_1, n_2, \dots, n_L) &= \sum_{j=1}^p a_j V_j \\ &= \sum_{j=1}^p a_j \left(\sum_{h=1}^L \frac{W_h^2 S_{jh}^2}{n_h} \right) \\ &= \sum_{h=1}^L \sum_{j=1}^p \frac{W_h^2 (a_j S_{jh}^2)}{n_h} \\ &= \sum_{h=1}^L \frac{W_h^2}{n_h} \sum_{j=1}^p a_j S_{jh}^2 \\ &= \sum_{h=1}^L \frac{W_h^2 A_h^2}{n_h}, \end{aligned}$$

where $A_h^2 = \sum_{j=1}^p a_j S_{jh}^2; \quad h = 1, 2, \dots, L.$

The NLPP for finding the optimum compromise allocation according to Khan *et al.* (2003) compromise criterion may be given as

$$\left. \begin{aligned}
 & \text{Minimize } Z = \sum_{h=1}^L \frac{W_h^2 A_h^2}{n_h} \\
 & \text{Subject to } \sum_{h=1}^L c_h n_h + \sum_{h=1}^L t_h \sqrt{n_h} \leq C_0, \\
 & \text{and } 2 \leq n_h \leq N_h; \quad h=1, 2, \dots, L.
 \end{aligned} \right\} j=1, 2, \dots, p$$

(24)

When the numerical values of W_h, A_h, c_h, t_h, C_0 and N_h are available, LINGO optimization software may be used to obtain a solution.

5 A Numerical Illustration

In the table below the stratum sizes, stratum weights, stratum standard deviations, measurement costs, and the travel costs within stratum are given for four different characteristics under study in a population stratified in five strata. The data are mainly from Chatterjee (1968). The values of strata sizes are added assuming the population size as 6000. The traveling cost t_h is also assumed for the five strata by the authors.

Table 1. Values of N_h, W_h, c_h, t_h and S_{jh} for five strata and four characteristics.

h	N_h	W_h	c_h	t_h	S_{jh}			
					S_{1h}	S_{2h}	S_{3h}	S_{4h}
1	1500	0.25	1	0.5	28	206	38	120
2	1920	0.32	1	0.5	24	133	26	184
3	1260	0.21	1.5	1	32	48	44	173
4	480	0.08	1.5	1	54	37	78	92
5	840	0.14	2	1.5	67	9	76	117

The total budget of the survey is assumed to be 1500 units with an overhead cost $c_0 = 300$ units. Thus $C_0 = C - c_0 = 1500 - 300 = 1200$ units are available for measurement and travel within strata for approaching the selected units for measurement.

5.1 The Proposed Compromise Allocation with Quadratic Cost

Using the values given in Table 1 the NLPP (12) and their optimal solutions \underline{n}_j^* ; $j=1, 2, 3, 4$ with the

corresponding values of V_j^* are listed below. These values are obtained by software LINGO.

For $j=1$

$$\left. \begin{aligned}
 & \text{Minimize } \frac{49}{n_1} + \frac{58.9824}{n_2} + \frac{45.1584}{n_3} + \frac{18.6624}{n_4} + \frac{87.9844}{n_5} \\
 & \text{Subject to } \ln_1 + \ln_2 + 1.5n_3 + 1.5n_4 + 2n_5 + 0.5\sqrt{n_1} + \\
 & \quad 0.5\sqrt{n_2} + 1\sqrt{n_3} + 1\sqrt{n_4} + 1.5\sqrt{n_5} \leq 1200, \\
 & \quad 2 \leq n_1 \leq 1500, \\
 & \quad 2 \leq n_2 \leq 1920, \\
 & \quad 2 \leq n_3 \leq 1260, \\
 & \quad 2 \leq n_4 \leq 480, \\
 & \text{and } 2 \leq n_5 \leq 840.
 \end{aligned} \right\}$$

(25)

The optimum allocation $\underline{n}_1^* = (n_{11}^*, n_{12}^*, n_{13}^*, n_{14}^*, n_{15}^*)$ is

$n_{11}^* = 193.6535, n_{12}^* = 212.5509, n_{13}^* = 151.1150,$
 $n_{14}^* = 96.82734, n_{15}^* = 182.6156.$ The corresponding value

of the variance ignoring fpc is $V_1^* = 1.503902$.

Similarly for $j = 2, 3$ & 4 the results are

$j=2; n_{21}^* = 535.7324, n_{22}^* = 442.4963, n_{23}^* = 84.55834,$
 $n_{24}^* = 24.46572, n_{25}^* = 8.779119.$ The corresponding value

of the variance ignoring fpc is $V_2^* = 10.78476$.

$j=3; n_{31}^* = 210.3325, n_{32}^* = 184.0994, n_{33}^* = 166.3332,$
 $n_{34}^* = 112.0209, n_{35}^* = 165.6085.$ The corresponding value of

the variance ignoring fpc is $V_3^* = 2.349571$.

$j=4; n_{41}^* = 208.6400, n_{42}^* = 410.4944, n_{43}^* = 205.6995,$
 $n_{44}^* = 41.09868, n_{45}^* = 79.59035.$ The corresponding value

of the variance ignoring fpc is $V_4^* = 23.86480$.

Using the computed values of V_j^* ; $j=1, 2, 3, 4$ and the compromise criterion conjectured in section 3, the Goal Programming Problem given in (17) may be expressed as:

$$\begin{aligned}
 & \text{Minimize } \sum_{j=1}^4 x_j . \\
 & \text{Subject to } \frac{49}{n_1} + \frac{58.9824}{n_2} + \frac{45.1584}{n_3} + \frac{18.6624}{n_4} + \frac{87.9844}{n_5} - x_1 \leq 1.503902, \\
 & \frac{2652.25}{n_1} + \frac{1811.3536}{n_2} + \frac{101.6064}{n_3} + \frac{8.7616}{n_4} + \frac{1.5876}{n_5} - x_2 \leq 10.78476, \\
 & \frac{90.25}{n_1} + \frac{69.2224}{n_2} + \frac{85.3776}{n_3} + \frac{38.9376}{n_4} + \frac{113.2096}{n_5} - x_3 \leq 2.349571, \\
 & \frac{900}{n_1} + \frac{3466.8544}{n_2} + \frac{1319.8689}{n_3} + \frac{54.1696}{n_4} + \frac{268.3044}{n_5} - x_4 \leq 23.86480, \\
 & 1n_1 + 1n_2 + 1.5n_3 + 1.5n_4 + 2n_5 + \\
 & 0.5\sqrt{n_1} + 0.5\sqrt{n_2} + 1\sqrt{n_3} + 1\sqrt{n_4} + 1.5\sqrt{n_5} \leq 1200, \\
 & \qquad \qquad \qquad 2 \leq n_1 \leq 1500, \\
 & \qquad \qquad \qquad 2 \leq n_2 \leq 1920, \\
 & \qquad \qquad \qquad 2 \leq n_3 \leq 1260, \\
 & \qquad \qquad \qquad 2 \leq n_4 \leq 480, \\
 & \qquad \qquad \qquad 2 \leq n_5 \leq 840, \\
 & \text{and } x_j \geq 0 ; j = 1, 2, 3, 4.
 \end{aligned} \tag{26}$$

The optimum compromise allocation which is the solution to the NLPP (26) given by the optimization software LINGO is:

$$\begin{aligned}
 n_{1c}^* &= 309.1612, n_{2c}^* = 374.3831, n_{3c}^* = 162.7233, \\
 n_{4c}^* &= 44.8336, n_{5c}^* = 77.01911.
 \end{aligned}$$

After rounding off to the nearest integer value we get the optimum compromise allocation as

$$n_{1c}^* = 309, n_{2c}^* = 374, n_{3c}^* = 162, n_{4c}^* = 45, n_{5c}^* = 77 \text{ The}$$

variances $V(\bar{y}_{jst})$ under compromise allocations denoted by $V(\bar{y}_{jst})_{comp}$ are:

$$\begin{aligned}
 V(\bar{y}_{1st})_{comp} &= 2.152413105, V(\bar{y}_{2st})_{comp} = 14.26904518, \\
 V(\bar{y}_{3st})_{comp} &= 3.33971459 \text{ \& } V(\bar{y}_{4st})_{comp} = 25.0178660
 \end{aligned}$$

with increases in the variances for the individual characteristics as:

$$\begin{aligned}
 x_1 &= 0.6482838, x_2 = 3.472783, \\
 x_3 &= 0.9903062 \text{ and } x_4 = 1.109451
 \end{aligned}$$

5.2 Proportional Allocation with Quadratic Cost

Using the values of c_h, t_h and W_h as given in Table 1 with $C_0 = 1200$ the numerical values of A and B are obtained as

$$A = 1.2850 \text{ and } B = 1.8353. \text{ This gives}$$

$$X = \sqrt{n} = 29.8532.$$

$$\text{or } n = X^2 = 891.2136.$$

Substituting this value of n in (18) the proportional allocation is obtained as:

$$n_1 = 222.8034, n_2 = 285.1884, n_3 = 187.1549,$$

$$n_4 = 71.2971, n_5 = 124.7699.$$

After rounding off to the nearest integer value we get

$$n_1 = 223, n_2 = 285, n_3 = 187, n_4 = 71, n_5 = 125.$$

The variances of $V(\bar{y}_{jst})$ under proportional allocation (ignoring fpc) may be obtained by substituting the above values of n_h in variance formula

$$V(\bar{y}_{jst}) = \sum_{h=1}^L \frac{W_h^2 S_{jh}^2}{n_h}; \quad j = 1, 2, \dots, 4, \tag{27}$$

which gives $V(\bar{y}_{jst})$ for $j = 1, 2, \dots, 4$ as:

$$V_1 = V(\bar{y}_{1st})_{prop} = 1.6350, \quad V_2 = V(\bar{y}_{2st})_{prop} = 18.9285$$

$$V_3 = V(\bar{y}_{3st})_{prop} = 2.5583 \text{ \& } V_4 = V(\bar{y}_{4st})_{prop} = 26.1678$$

Table 2. R. E. as compared to proportional allocation.

S.No (1)	Compromise allocations (2)	Values of $V_j ; j = 1,2,3 \& 4$ under various compromise allocation				Trace (3)+(4)+(5)+(6) (7)	R.E. $= T_{Prop.}/T_{Comp.}$ (8)
		V_1 (3)	V_2 (4)	V_3 (5)	V_4 (6)		
1	Proportional	1.6350	18.9285	2.5583	26.1678	49.2896	1.00
2	Cochran's Compromise allocation	1.7345	15.8570	2.7010	26.1775	46.4700	1.06
3	Khan's Compromise allocation	3.0100	13.6979	4.4648	25.8342	47.0061	1.05
4	Proposed	2.1524	14.2690	3.3397	25.0179	44.7790	1.10

5.3 Cochran's Compromise Allocation with Quadratic Cost

For the present example Cochran's compromise allocations given by (23) are

$$n_1 = 287.0896, n_2 = 312.4103, n_3 = 151.9265, n_4 = 68.6031, n_5 = 109.148$$

After rounding off to the nearest integer value we get $n_1 = 287, n_2 = 312, n_3 = 152, n_4 = 69, n_5 = 109$.

The variances $V(\bar{y}_{jst})_C$ under the Cochran's compromise allocation are

$$V_1^C = V(\bar{y}_{1st})_C = 1.7345, \quad V_2^C = V(\bar{y}_{2st})_C = 15.8570$$

$$V_3^C = V(\bar{y}_{3st})_C = 2.7010 \text{ and } V_4^C = V(\bar{y}_{4st})_C = 26.1775.$$

5.4 Khan's Compromise Allocation with Quadratic Cost

Using the values given in Table 1 the NLPP (24) becomes

$$\text{Minimize } Z = \left(\begin{array}{l} \frac{1377.5081}{n_1} + \frac{2449.5358}{n_2} + \frac{740.9373}{n_3} \\ + \frac{35.7884}{n_4} + \frac{56.2973}{n_5} \end{array} \right)$$

$$\text{Subject to } \left. \begin{array}{l} 1n_1 + 1n_2 + 1.5n_3 + 1.5n_4 + 2n_5 + 0.5\sqrt{n_1} \\ + 0.5\sqrt{n_2} + 1\sqrt{n_3} + 1\sqrt{n_4} + 1.5\sqrt{n_5} \leq 1200, \\ 2 \leq n_1 \leq 1500, \\ 2 \leq n_2 \leq 1920, \\ 2 \leq n_3 \leq 1260, \\ 2 \leq n_4 \leq 480, \\ \text{and } 2 \leq n_5 \leq 840. \end{array} \right\} \quad (28)$$

The optimal solution to NLPP (28) using optimization software LINGO is obtained as:

$$n_1 = 310.9282, n_2 = 415.0147, n_3 = 185.2513, n_4 = 40.1666, n_5 = 43.5408.$$

After rounding off to the nearest integer value we get $n_1 = 311, n_2 = 415, n_3 = 185, n_4 = 40, n_5 = 44$.

The variances $V(\bar{y}_{jst})$ using the Khan *et al.* (2003) compromise criterion are

$$V_{1K} = 3.0100, \quad V_{2K} = 13.6979 \quad V_{3K} = 4.4648 \quad \text{and} \\ V_{4K} = 25.8342, \text{ where 'K' stands for Khan's compromise allocation.}$$

6 Results and Discussion

In this section a comparative study of the four compromise allocations discussed in this paper has been made. The basis of comparison is the traces of the variance-covariance matrices of the estimates under various compromise allocations. Since the characteristics under study are assumed as independent, the covariances are zero. The traces are the sum of the diagonal elements of the variances-covariance matrices that are the variances of the estimates of the population means of the different characteristics. Sukhatme *et al.* (1984) define the relative efficiency (R. E.) of a compromise allocation with respect to proportional allocation as

$$\text{R.E.} := \frac{T_{Prop.}}{T_{Comp.}} \quad (29)$$

where $T_{Prop.}$ = Sum of the variances under proportional allocation. and $T_{Comp.}$ = Sum of the variances under the given compromise allocations. Column (8) of Table 2 gives the R.E. of the three compromise allocation discussed in this article as compared to the proportional allocation.

7 Conclusion

The results summarized in Table 2 indicate that the proposed compromise allocation compares favorably with the other studied allocations when the travel costs within the strata are significant.

Acknowledgements

The authors are grateful to the learned referees and the editors for their valuable suggestions and comments that helped us to improve the paper in its present form.

References

- Ahsan, M.J. 1975-76. A procedure for the problem of optimum allocation in multivariate stratified random sampling. *The Aligarh Bull. of Math.* **5-6**, 37-42.
- Ahsan, M.J. and Khan, S.U. 1977. Optimum allocation in multivariate stratified random sampling using prior information. *Journal of Indian Statistical Association* **15**, 57-67.
- Ahsan, M.J. 1978. Allocation problem in multivariate stratified random sampling. *Journal of the Indian Statistical Association* **16**, 1-5.
- Aoyama, H. 1963. Stratified random sampling with optimum allocation for multivariate populations. *Annals of the Institute of Statistical Mathematics* **14**, 251-258.
- Beardwood, J., Halton, J.H. and Hammersley, J.M. 1959. The shortest path through many points, *Proc. Cambridge Phil. Soc.*, **55**, 299-327.
- Bethel, J. 1985. An optimum allocation algorithm for multivariate surveys. *Proceedings of the Survey Research Methods section, American Statistical Association* 209-212.
- Bethel, J. 1989. Sample allocation in multivariate surveys. *Survey Methodology* **15**, 47-57.
- Bosch, V. and Wildner, R. 2003. Optimum allocation of stratified random samples designed for multiple mean estimates and multiple observed variables. *Communications in Statistics- Theory and Methods* **32**, 1897-1909.
- Chatterjee, S. 1967. A note on optimum allocation, *Scandinavian Actuarial Journal* **50**, 40-44.
- Chatterjee, S. 1968. Multivariate stratified surveys, *Journal of American Statistica. Association* **63**, 530-534.
- Chromy, J.R. 1987. Design optimization with multiple objectives. *Proceedings of the Survey Research Methods section, American Statistical Association* 194-199.
- Cochran, W.G. 1977. *Sampling Techniques*, 3rd ed., John Wiley, New York.
- Csenki, A. 1997. Optimum Allocation in stratified random sampling via Holder's inequality, *Statistician* **46**, 439-441.
- Dalenius, T. 1957. *Sampling in Sweden*. Contributions to the Methods and Theories of Sample Survey Practice, Almqvist and Wicksell, Stockholm.
- Díaz-García, J.A. & Cortez, L.U. 2006. Optimum allocation in multivariate stratified sampling: multi-objective programming, *Comunicación Técnica No. 1-06-07/28-03-2006 (PE/CIMAT)*, México.
- Díaz-García, J.A. and Cortez, L.U. 2008. Multi-objective optimisation for optimum allocation in multivariate stratified sampling, *Survey Methodology* **34**, 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. *Journal of American Statistical Association* **60**, 225-233.
- Geary, R.C. 1949. Sampling methods applied to Irish agricultural statistics. *Technical Series*, Central Statistical office, Dublin.
- Ghosh, S.P. 1958. A note on stratified random sampling with multiple characters. *Calcutta Statistical Association Bulletin* **8**, 81-89.
- Jahan, N., Khan, M.G.M. and Ahsan, M.J. 1994. A generalized compromise allocation. *Journal of the Indian Statistical Association* **32**, 95-101.
- Jahan, N. & Ahsan, M.J. 1995. Optimum allocation using separable programming. *Dhaka University Journal of Science* **43**, 157-164.
- Jahan, N., Khan, M.G.M. and Ahsan, M.J. 2001. Optimum compromise allocation using dynamic programming. *Dhaka University Journal of Science* **49**, 197-202.
- Khan, M.G.M., Ahsan, M.J. and Jahan, N. 1997. Compromise Allocation in Multivariate Stratified Sampling: An Integer Solution. *Naval Research Logistics* **44**, 69-79.
- Khan, M.G.M., Khan, E.A. and Ahsan, M.J. 2003. An optimal multivariate stratified sampling design using dynamic programming. *Australian & New Zealand J. Statist* **45**, 107-113.
- Khan, M.G.M., Khan, E.A. and Ahsan, M.J. 2008. Optimum allocation in multivariate stratified sampling in presence of non-response. *Journal of Indian Society of Agricultural Statistics* **62**, 42-48.
- Khan, M.G.M., Maiti, T. and Ahsan, M.J. 2010. An optimal multivariate stratified sampling design using auxiliary information: an integer solution using goal programming approach. *Journal of Official Statistics* **26**, 695-708.
- Khowaja, S., Ghufuran, S. and Ahsan, M.J. 2011. Estimation of population means in multivariate stratified random sampling. *Communication in Statistics-Simulation and Computation* **40**, 710-718.
- Kokan, A.R. and Khan, S.U. 1967. Optimum allocation n multivariate surveys: An analytical solution. *Journal of Royal Statistical Society B* **29**, 115-125.
- Kozak, M. 2006. On sample allocation in multivariate surveys. *Communication in Statistics-Simulation and Computation* **35**, 901-910.
- LINGO User's Guide 2001. published by Lindo Systems Inc., 1415 North Dayton Street, Chicago, Illinois-60622 (USA).
- Neyman, J. 1934. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of Royal Statistical Society* **97**, 558-625.
- Peter, J.H. and Bucher (Undated). The 1940 section sample survey of crop aggregates in Indiana and Iowa. *U.S., Dept. of Agriculture*.
- Schittkowski, K. 1985-86. NLPQL: A FORTRAN subroutine solving constrained nonlinear programming problems. *Annals of Operations Research* **5**, 485-500.
- Singh S. 2003 *Advanced Sampling Theory with Application*, Vol.II, Netherland: Kluwer Academic Publishers, Dordrecht.
- Stuart, A. 1954. A simple presentation of optimum sampling results, *Journal of Royal Statistical Society, B* **16**, 239-241.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Ashok, C. 1984. *Sampling Theory of Surveys with Applications*, 3rd ed., Iowa State University Press, Ames, Iowa and Indian Society of Agricultural Statistics, New Delhi.

Yates, F. 1960. *Sampling Methods for Censuses and Surveys*. Charles Griffin and Co., London, third edition.

Correspondence to : Shazia Ghufuran
E. mail: itsshaziaghufuran@gmail.com