A multiple response stratified sampling design with travel cost

Shazia Ghufran¹, 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 $N_1, N_2, \ldots, N_h, \ldots, N_L$ with $\sum_{h=1}^{L} N_h = N$.

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

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

(1)

where $c_h; h=1,2,\ldots,L$ 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

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

(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 $V(\hat{\gamma}_{(s)})$ 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

$$\frac{n_h \sqrt{c_h}}{W_h s_h} = \text{Constant}; \ h=1,2,\ldots,L.$$  

This gives the sample size $n_h$ in $h$-th stratum as:

$$n_h = n \left( \frac{W_h s_h}{\sqrt{c_h}} \right); \ h=1,2,\ldots,L,$$  

(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. Be woodow et al. (1959) suggested that the cost of visiting the $n_k$ selected units in the $h$-th stratum may be taken as $t_h \sqrt{n_h}; h=1,2,\ldots,L$ 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:

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

(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):

\[
\text{Minimize } \quad V(\bar{y}, \lambda) = \sum_{h=1}^{L} \left( \frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_h^2.
\]

Subject to

\[
\sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0,
\]

and

\[
n_h \geq 0; \quad h = 1, 2, \ldots, L,
\]

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

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

where \( \lambda \) is the Lagrange Multiplier.

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

\[
\frac{\partial L(n_h, \lambda)}{\partial n_h} = \frac{W_h^2 S_h^2}{n_h^2} + \lambda \left( c_h + \frac{t_h}{2\sqrt{n_h}} \right) = 0; \quad h = 1, 2, \ldots, L.
\]

and

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

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 Ramphson’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 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), Khwaja 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 (2001), 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 \( (t_h) \) with in 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 (MNLP) discussed in the previous section may be expressed as:
\[
\begin{align*}
\text{Minimize} & \quad \left\{ \begin{array}{l}
V(\tilde{y}_{1st}) \\
V(\tilde{y}_{2nd}) \\
\vdots \\
V(\tilde{y}_{pst})
\end{array} \right. \\
\text{Subject to} & \quad \frac{L}{h=1} c_h n_h + \frac{L}{h=1} l_h \sqrt{n_h} \leq C_0; \\
& \quad 2 \leq n_h \leq N_h; \quad h=1,2,\ldots,L.
\end{align*}
\]

(9)

where \( V(\tilde{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_h} \), \( j = 1,2,\ldots,p \),
denote the sampling variance of the estimate

\[
\tilde{y}_{jst} = \sum_{h=1}^{L} W_h \tilde{y}_{jh}
\]

(11)

ignoring fpc of the overall population mean \( \tilde{y}_j \); \( j = 1,2,\ldots,p \) of the \( j \)-th characteristic.

\[
\tilde{y}_{jh} = \frac{1}{n_h} \sum_{k=1}^{n_h} y_{jkh}
\]

is the sample mean from the \( h \)-th stratum for the \( j \)-th characteristic and \( y_{jkh} \) is the value of the \( k \)-th selected unit of the sample from the \( h \)-th stratum for the \( j \)-th characteristic;

\[
k = 1,2,\ldots,n_h; \quad h = 1,2,\ldots,L; \quad j = 1,2,\ldots,p.
\]

The restrictions \( 2 \leq n_h \leq N_h \); \( h = 1,2,\ldots,L \) 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

Minimize \( V(\tilde{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_h} \)

Subject to \( \frac{L}{h=1} c_h n_h + \frac{L}{h=1} l_h \sqrt{n_h} \leq C_0; \)

\( j = 1,2,\ldots,p \)

and \( 2 \leq n_h \leq N_h; \quad h = 1,2,\ldots,L. \)

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

\[
V_j^* = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_{jh}^*}; \quad j = 1,2,\ldots,p.
\]

(13)

Further, let \( n_j^* = (n_{j1}^*, n_{j2}^*, \ldots, n_{jL}^*) \) be the vector of optimum compromise allocations with

\[
V_{cj} = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_{hc}}
\]

as the optimal value of the objective function for \( j \)-th characteristics under this allocation. Obviously,

\[
V_{cj}^* \geq V_j^* \quad \text{or} \quad V_{cj}^* - V_j^* \geq 0; \quad j = 1,2,\ldots,p.
\]

(14)

A reasonable criterion to work out a compromise allocation may be to “Minimize the sum of increases in the variances \( V_j; \ j = 1,2,\ldots,p \) 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

Minimize \( \sum_{j=1}^{p} x_j. \)

Subject to \( V_{cj} - V_j^* \leq x_j; \)

\( \frac{L}{h=1} c_h n_h + \frac{L}{h=1} l_h \sqrt{n_h} \leq C_0; \)

\( j = 1,2,\ldots,p \)

and \( 2 \leq n_h \leq N_h; \quad h = 1,2,\ldots,L \)

(15)

where \( n_c = (n_{1c}, n_{2c}, \ldots, n_{Lc}) \) denotes a compromise allocation with variance

\[
V_{cj} = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_{hc}}; \quad j = 1,2,\ldots,p
\]

(16)

and \( x_j \geq 0; \quad j = 1,2,\ldots,p \) are called goal variables whose values are to be determined.

The ‘Goal’ is to “Find the compromise allocation \( n_c^* = (n_{1c}^*, n_{2c}^*, \ldots, n_{Lc}^*) \) such that the increases in the \( j \)-th variance due to the use of compromise allocation should not exceed \( x_j; \ j = 1,2,\ldots,p \) and \( \sum_{j=1}^{p} x_j \) is minimum.”

NLPP (15) may be restated as

Minimize \( \sum_{j=1}^{p} x_j. \)

Subject to \( \frac{L}{h=1} \frac{W_h^2 S_{jh}^2}{n_h} - x_j \leq V_j^*; \)

\( \frac{L}{h=1} c_h n_h + \frac{L}{h=1} l_h \sqrt{n_h} \leq C_0; \)

\( x_j \geq 0 \)

and \( 2 \leq n_h \leq N_h; \quad h = 1,2,\ldots,L. \)

(17)

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

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

or

$$n_h = kW_h; \quad h = 1, 2, ..., L$$

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

$$n_h = nW_h; \quad h = 1, 2, ..., L.$$  \hspace{1cm} (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

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

we get

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

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

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

in (20) we get a quadratic equation in $X$ as:

$$AX^2 + BX - C_0 = 0,$$  \hspace{1cm} (21)

with roots

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

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

$$X = \frac{-B + \sqrt{B^2 + 4AC_0}}{2A}$$  \hspace{1cm} (22)

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

$$\left(\sqrt{B^2 + 4AC_0} \right) > B$$

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 $j = 1, 2, ..., p$, over the characteristics.

Cochran’s compromise allocation is given by

$$n_h = \frac{1}{p} \sum_{j=1}^{p} n^*_{jh}.$$  \hspace{1cm} (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

$$a_j = \frac{p \sum_{h=1}^{L} S^2_{jh}}{\sum_{j=1}^{p} L S^2_{jh}}.$$  \hspace{1cm} (23)

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:

$$Z(n_1, n_2, ..., n_L) = \sum_{j=1}^{p} a_j V_j$$

$$= \sum_{j=1}^{p} \left( \frac{\sum_{h=1}^{L} W^2 h S^2_{jh}}{n_h} \right)$$

$$= \sum_{j=1}^{p} \sum_{h=1}^{L} W^2 h a_j S^2_{jh}$$

$$= \sum_{j=1}^{p} \sum_{h=1}^{L} W^2 h a^2_{jh}.$$

where $A^2_{jh} = \sum_{j=1}^{p} a_j S^2_{jh}; \quad h = 1, 2, ..., L.$

The NLPP for finding the optimum compromise allocation according to Khan et al. (2003) compromise criterion may be given as
Minimize $Z = \sum_{h=1}^{L} \frac{W_h^2 A_h^2}{n_h}$.

Subject to $\sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h n_h \leq C_0$, \hspace{1cm} (j = 1, 2, \ldots, p)$

and $2 \leq n_h \leq N_h$; \hspace{1cm} (h = 1, 2, \ldots, L) \hspace{1cm} (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>
<thead>
<tr>
<th>$h$</th>
<th>$N_h$</th>
<th>$W_h$</th>
<th>$c_h$</th>
<th>$t_h$</th>
<th>$S_{jh}$</th>
<th>$S_{2h}$</th>
<th>$S_{3h}$</th>
<th>$S_{4h}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1500</td>
<td>0.25</td>
<td>1</td>
<td>0.5</td>
<td>28</td>
<td>206</td>
<td>38</td>
<td>120</td>
</tr>
<tr>
<td>2</td>
<td>1920</td>
<td>0.32</td>
<td>1</td>
<td>0.5</td>
<td>24</td>
<td>133</td>
<td>26</td>
<td>184</td>
</tr>
<tr>
<td>3</td>
<td>1260</td>
<td>0.21</td>
<td>1.5</td>
<td>1</td>
<td>32</td>
<td>48</td>
<td>44</td>
<td>173</td>
</tr>
<tr>
<td>4</td>
<td>480</td>
<td>0.08</td>
<td>1.5</td>
<td>1</td>
<td>54</td>
<td>37</td>
<td>78</td>
<td>92</td>
</tr>
<tr>
<td>5</td>
<td>840</td>
<td>0.14</td>
<td>2</td>
<td>1.5</td>
<td>67</td>
<td>9</td>
<td>76</td>
<td>117</td>
</tr>
</tbody>
</table>

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 $n^*_j$: $j = 1, 2, 3, 4$ with the corresponding values of $V^*_j$ are listed below. These values are obtained by software LINGO.

For $j = 1$

Minimize $\frac{49}{n_1} + \frac{58.9824}{n_2} + \frac{45.1584}{n_3} + \frac{18.6624}{n_4} + \frac{87.9844}{n_5}$

Subject to $l_1 + l_2 + 1.5 n_3 + 1.5 n_4 + 2 n_5 + 0.5 \sqrt{n_1} + 0.5 \sqrt{n_2} + 1/\sqrt{n_3} + 1/\sqrt{n_4} + 1.5/\sqrt{n_5} \leq 1200$, \hspace{1cm} (25)$

and $2 \leq n_1 \leq 1500$, \hspace{1cm} $2 \leq n_2 \leq 1920$, \hspace{1cm} $2 \leq n_3 \leq 1260$, \hspace{1cm} $2 \leq n_4 \leq 480$, \hspace{1cm} $2 \leq n_5 \leq 840$. \hspace{1cm} (25)$

The optimum allocation $n^*_j = (n^*_{11}, n^*_{12}, n^*_{13}, n^*_{14}, n^*_{15})$ 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:
Minimize \[ \sum_{j=1}^{4} x_j. \]

Subject to

\[
\begin{align*}
\frac{49}{n_1} + \frac{58.9824}{n_2} + \frac{45.1584}{n_3} + \frac{18.6624}{n_4} + \frac{87.9844}{n_5} - x_1 & \leq 1.503902, \\
\frac{2652.25}{n_1} + \frac{1811.3536}{n_2} + \frac{101.6064}{n_3} + \frac{8.7616}{n_4} + \frac{1.5876}{n_5} - x_2 & \leq 10.78476, \\
\frac{90.25}{n_1} + \frac{69.2224}{n_2} + \frac{85.3776}{n_3} + \frac{38.9376}{n_4} + \frac{113.2096}{n_5} - x_3 & \leq 2.349571, \\
\frac{900}{n_1} + \frac{3466.8544}{n_2} + \frac{1319.8689}{n_3} + \frac{54.1696}{n_4} + \frac{268.3044}{n_5} - x_4 & \leq 23.86480.
\end{align*}
\]

\[ \ln n_1 + \ln n_2 + 1.5 n_3 + 1.5 n_4 + 2 n_5 + 0.5 \sqrt{n_1} + 0.5 \sqrt{n_2} + 1 \sqrt{n_3} + 1 \sqrt{n_4} + 1.5 \sqrt{n_5} \leq 1200, \]

\[ \begin{align*}
2 \leq n_1 \leq 1500, \\
2 \leq n_2 \leq 1920, \\
2 \leq n_3 \leq 1260, \\
2 \leq n_4 \leq 480, \\
2 \leq n_5 \leq 840, \\
x_j \geq 0; j = 1, 2, 3, 4.
\end{align*}\]

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

\[ n_{1c} = 309.1612, \quad n_{2c} = 374.3831, \quad n_{3c} = 162.7233, \]
\[ n_{4c} = 44.8336, \quad n_{5c} = 77.01911. \]

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

\[ n_{1c} = 309, \quad n_{2c} = 374, \quad n_{3c} = 162, \quad n_{4c} = 45, \quad n_{5c} = 77. \]

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

\[ \begin{align*}
V(\bar{y}_{1st})_{comp} & = 2.152413105, \quad V(\bar{y}_{2st})_{comp} = 14.26904518, \\
V(\bar{y}_{3st})_{comp} & = 3.33971459 & V(\bar{y}_{4st})_{comp} & = 25.0178660.
\end{align*}\]

with increases in the variances for the individual characteristics as:

\[ \begin{align*}
x_1 = 0.6482838, \quad x_2 = 3.472783, \\
x_3 = 0.9903062 & \quad and \quad x_4 = 1.109451.
\end{align*}\]

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 \( A = 1.2850 \) and \( B = 1.8353 \). This gives \( X = \sqrt{n} = 29.8532 \).

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

\[ \begin{align*}
V(\bar{y}_{jst}) & = \sum_{k=1}^{L} \frac{W_h^2 y_{jk}^2}{n_h}; \quad j = 1, 2, ..., 4, \quad (27)
\end{align*}\]

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

\[ \begin{align*}
V_1 = V(\bar{y}_{1st})_{prop} = 1.6350, \quad V_2 = V(\bar{y}_{2st})_{prop} = 18.9285, \\
V_3 = V(\bar{y}_{3st})_{prop} = 2.5583 & \quad and \quad V_4 = V(\bar{y}_{4st})_{prop} = 26.1678.
\end{align*}\]
5.3 Cochran’s Compromise Allocation with Quadratic Cost
For the present example Cochran’s compromise allocations given by (23) are
\[ n_1 = 287.0896, \quad n_2 = 312.4103, \quad n_3 = 151.9265, \quad n_4 = 68.6031, \quad n_5 = 109.148 \]
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
\[
V_C^1 = V(\bar{y}_{1st})_C = 1.7345, \quad V_C^2 = V(\bar{y}_{2st})_C = 15.8570 \\
V_C^3 = V(\bar{y}_{3st})_C = 2.7010 \quad \text{and} \quad V_C^4 = V(\bar{y}_{4st})_C = 26.1775.
\]

5.4 Khan’s Compromise Allocation with Quadratic Cost
Using the values given in Table 1 the NLPP (24) becomes
\[
\text{Minimize } Z = \frac{1377.5081}{n_1} + \frac{2449.5358}{n_2} + \frac{740.9373}{n_3} + \frac{35.7884}{n_4} + \frac{56.2973}{n_5}
\]
Subject to \( \ln n_1 + \ln n_2 + 1.5n_3 + 1.5n_4 + 2n_5 + 0.5\sqrt{n_1} + 0.5\sqrt{n_2} + 1.5\sqrt{n_3} + 1.5\sqrt{n_4} + 1.5\sqrt{n_5} \leq 1200, \)
\[
2 \leq n_1 \leq 1500, \quad 2 \leq n_2 \leq 1920, \quad 2 \leq n_3 \leq 1260, \quad 2 \leq n_4 \leq 480, \quad 2 \leq n_5 \leq 840.
\]
and
\[
(28)
\]

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
\[ V_{k1} = 3.0100, \quad V_{k2} = 13.6979, \quad V_{k3} = 4.4648 \quad \text{and} \quad V_{k4} = 25.8342, \] 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 \( T_{Prop.} = \text{Sum of the variances under proportional allocation} \) and \( T_{Comp.} = \text{Sum of the variances under the given compromise allocations} \). 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

Peter, J.H. and Bucher (Undated). The 1940 section sample survey of crop aggregates in Indiana and Iowa. U.S., Dept. of Agriculture.


Correspondence to: Shazia Ghufran
E. mail: itsshaziaghufran@gmail.com

http://www.publish.csiro.au/journals/spinas