OPTIMUM ALLOCATION IN STRATIFIED SURVEYS

José A. Díaz-García and Ma. Magdalena Garay-Tápia

Comunicación Técnica No I-05-14/07-12-2005
(PE/CIMAT)

Optimum allocation in stratified surveys

José A. Díaz-García and Ma. Magdalena Garay-Tápia

Universidad Autónoma Agraria Antonio Narro
Department of Statistics and Computation
25315 Buenavista, Saltillo
Coahuila, México

Abstract

In this work we consider the allocation problem in stratified surveys as a problem of non-linear stochastic programming of integers. An example is solved by the following techniques: Lagrange multipliers, modified E-model, E-model, V-model and chance constraints.

Key words: Stratified survey, E-model, stochastic programming, optimum allocation, integer programming, chance constraints, V-model.
PACS: 62D05.

1 Introduction

Following Sukhatme *et al.* (1984), the formulation of an optimum allocation problem in stratified surveys is stated as a non linear programming problem. In that case we want to minimize the cost function under a given precision V_0, which is assigned to the sample variance of the mean $\mathrm{Var}(\overline{y}_{st})$; or alternatively, we can minimize the sample variance of the mean $\mathrm{Var}(\overline{y}_{st})$, under a fixed cost given by the conditions of the cost function. Like this, suppose that a population of size N is divided in H sub-populations or strata, each one with size N_h, $h = 1, \dots, H$. Let n_h be the size of the sampling without replacement, which is obtained from the h-th stratum (independent samples among strata) and if we define $\mathbf{n} = (n_1, \dots, n_H)'$; then, the two above mentioned optimization

Email addresses: `jadiaz@uaaan.mx` (José A. Díaz-García), `magda@uaaan.mx` (Ma. Magdalena Garay-Tápia).

problems can be described by

$$\begin{aligned} &\min_{\mathbf{n}} \operatorname{Var}(\bar{y}_{st}) \\ &\text{subject to} \\ &\mathbf{c}'\mathbf{n} + c_0 = C \\ &\mathbf{n} \in \Re^H \end{aligned} \tag{1}$$

and

$$\begin{aligned} &\min_{\mathbf{n}} \mathbf{c}'\mathbf{n} + c_0 \\ &\text{subject to} \\ &\operatorname{Var}(\bar{y}_{st}) \leq V_0 \\ &\mathbf{n} \in \Re^H \end{aligned} \tag{2}$$

respectively.

Where

$$\operatorname{Var}(\bar{y}_{st}) = \sum_{h=1}^{H} \frac{W_h^2}{n_h} S_h^2 - \sum_{h=1}^{H} \frac{W_h}{N} S_h^2,$$

with $W_h = (N_h/N) > 0$, $\mathbf{c} = (c_1, \ldots, c_H)'$, $c_h > 0$ are the unitary cost in each stratum; $c_0 \geq 0$ is the fixed cost of the operation; C denotes the total cost and $\mathbf{S}^2 = (S_1^2, \ldots, S_H^2)'$ is the vector of true population variances.

When the sampling restrictions are due to number of hours per man or total time of survey, it is possible to determine or to fix the total sample size n; thus our problem can be stated as, see Arthanari and Dodge (1981)

$$\begin{aligned} &\min_{\mathbf{n}} \operatorname{Var}(\bar{y}_{st}) \\ &\text{subject to} \\ &\sum_{h=1}^{H} n_h = n \\ &\mathbf{n} \in \Re^H \end{aligned} \tag{3}$$

The explicit solution of these three problems, (1) - (3), is obtained via Lagrange multipliers, see Sukhatme *et al.* (1984) and Arthanari and Dodge (1981), among many others. However, note that in the solution $\mathbf{n}$, $n_h \in \Re$, but in our solution n_h's must be integers. This is, in the optimum allocation problems (1) - (3) we demand that the n_h's are integers, i e. $n_h \in \mathbb{N}$, $h = 1, 2, \ldots, H$, where $\mathbb{N}$ is the set of natural numbers, Arthanari and Dodge (1981, Chapter 5, p. 216).

For solving these problems, two ways have been proposed in literature: the first one approximates the solutions via Lagrange multipliers; or it rounds each n_h to the next integer (see Cochran (1977, p. 139)). The second one treats these problems as non linear optimum programming of integers (an algorithm for this method is proposed by Arthanari and Dodge (1981)).

Now, note that de population variances S_h^2 are in general unknown, then they are substituted in the problems (1) - (3) by the sample variances in each stratum, $\mathbf{s}^2 = (s_1^2, \dots, s_H^2)'$. Also, see that the sample variances s_h^2 are random variables, then the problems (1) - (3) define programs of non linear stochastic optimization for integers, Rao (1978) and Prékopa (1995). The stochastic optimization have been used in the solution of some problems in probability and statistics; see Prékopa (1978), among many others. In the context of response surface methodology, Díaz García and Ramos-Quiroga (2001), Díaz García and Ramos-Quiroga (2002) and Díaz García *et al.* (2005) give a detailed study of that problem under a number of stochastic optimization techniques, and also they propose some new alternative methods. In this work we study the problems of optimum allocation in stratified surveys as problems of stochastic optimization by using the following techniques: the modified E-model, the E-model, the V-model and the chance constraints, see Section 2. The results are applied to a simulated example and they are compared with the classical solutions obtained by the Lagrange multiplier method, see Section 3.

2 Optimum allocation via stochastic programming

Consider the following problem of optimization

$$\begin{gathered}\min_{\mathbf{x}} h(\mathbf{x}, \boldsymbol{\xi}_0) \\ \text{subject to} \\ g_1(\mathbf{x}, \boldsymbol{\xi}_1) \leq 0 \\ \vdots \\ g_m(\mathbf{x}, \boldsymbol{\xi}_m) \leq 0,\end{gathered}$$

where $\mathbf{x} \in \Re^n$ (variables of decision), and $\boldsymbol{\xi}_j \in \Re^{K_j}$ (shape parameters), $j = 0, 1, \dots, m$. If $\mathbf{x}$ and/or $\boldsymbol{\xi}_j$ are of random character, then this is a problem of stochastic optimization, Prékopa (1995, p. 234).

These stochastic problems can be solved by proposing an equivalent deterministic problem; where *equivalent* means that the solution of the deterministic problem is a solution of the stochastic problem. Using that idea, several techniques have been proposed in the literature, in particular the techniques of the modified E-model, the E-model, the V-model and the chance constraints, among many others, see Charnes and Cooper (1963), Rao (1978, pp. 598-599) and Prékopa (1995, p. 245). It is important to note that these techniques were proposed for linear programming, however as we will show below, these methods can be extended easily to non linear programming in the decision variables, but linear with respect to the shape parameters, which give the stochastic character to the optimization program.

2.1 Modified E-model

Consider the following stochastic programming for minimizing the estimated variance of the mean, subject to the cost function

$$\begin{gathered}\min_{\mathbf{n}} \widehat{\operatorname{Var}}(\bar{y}_{st})\\ \text{subject to}\\ \mathbf{c}'\mathbf{n} + c_0 = C\\ n_h \in \mathbb{N}, \text{ and } s_h^2 \text{ random variables }, h = 1, 2, \dots, H,\end{gathered} \tag{4}$$

where

$$\operatorname{Var}(\bar{y}_{st}) = \sum_{h=1}^{H} \frac{W_h^2}{n_h} s_h^2 - \sum_{h=1}^{H} \frac{W_h}{N} s_h^2.$$

By using the limiting distribution of the sample variances (see Melaku (1986)), consider the random variable ξ_h defined as

$$\xi_h = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{ih} - \bar{Y}_h)^2;$$

which has an asymptotical normal distribution with mean $\mathrm{E}(\xi_h)$ and variance $\operatorname{Var}(\xi_h)$, given by

$$\mathrm{E}(\xi_h) = \frac{n_h}{n_h - 1} S_h^2.$$

and

$$\operatorname{Var}(\xi_h) = \frac{n_h}{(n_h - 1)^2} [C_{y_h}^4 - (S_h^2)^2],$$

respectively; where $C_{y_h}^4$ is the fourth moment, and it is computed as

$$C_{y_h}^4 = \frac{1}{N_h} \sum_{i=1}^{N_h} (y_{ih} - \bar{Y}_h)^4.$$

Now, observe that

$$s_h^2 = \xi_h - \frac{n_h}{n_h - 1}(\bar{y}_h - \bar{Y}_h)^2,$$

where

$$\frac{n_h}{n_h - 1} \to 1$$

and

$$(\bar{y}_h - \bar{Y}_h)^2 \to 0 \quad \text{in probability.}$$

Then, the sample variances s_h^2 have an asymptotical normal distribution, moreover

$$s_h^2 \xrightarrow{a} \mathcal{N}(\mathrm{E}(\xi_h), \mathrm{Var}(\xi_h)), \quad h = 1, \ldots, H, \text{ independents} \tag{5}$$

Observe that the objective function in problem (4) is a linear function in the variables s_h^2, thus the objective function has also a normal distribution, with

$$\begin{aligned}
\mathrm{E}(\widehat{\mathrm{Var}}(\bar{y}_{st})) &= \mathrm{E}\left(\sum_{h=1}^{H} \frac{W_h^2}{n_h} s_h^2 - \sum_{h=1}^{H} \frac{W_h}{N} s_h^2\right) \\
&= \sum_{h=1}^{H} \frac{W_h^2}{n_h} \mathrm{E}(\xi_h) - \sum_{h=1}^{H} \frac{W_h}{N} \mathrm{E}(\xi_h) \\
&= \sum_{h=1}^{H} \frac{W_h^2}{(n_h - 1)} S_h^2 - \sum_{h=1}^{H} \frac{W_h}{N} \left(\frac{n_h}{n_h - 1}\right) S_h^2
\end{aligned} \tag{6}$$

and

$\mathrm{Var}(\widehat{\mathrm{Var}}(\bar{y}_{st})) =$

$$\begin{aligned}
&= \mathrm{Var}\left(\sum_{h=1}^{H} \frac{W_h^2}{n_h} s_h^2 - \sum_{h=1}^{H} \frac{W_h}{N} s_h^2\right) \\
&= \sum_{h=1}^{H} \frac{W_h^4}{n_h^2} \mathrm{Var}(\xi_h) - \sum_{h=1}^{H} \frac{W_h^2}{N^2} \mathrm{Var}(\xi_h) \\
&= \sum_{h=1}^{H} \frac{W_h^4}{n_h(n_h - 1)^2}(C_{y_h}^4 - (S_h^2)^2) - \sum_{h=1}^{H} \frac{W_h^2}{N^2}\left[\frac{n_h}{(n_h - 1)^2}(C_{y_h}^4 - (S_h^2)^2)\right].
\end{aligned} \tag{7}$$

Then, by applying the modified E-model technique, the new objective function is given by

$$f(\mathbf{n}) = k_1 \mathrm{E}(\widehat{\mathrm{Var}}(\bar{y}_{st})) + k_2\sqrt{\mathrm{Var}(\widehat{\mathrm{Var}}(\bar{y}_{st}))},$$

which is a deterministic objective function. Here k_1 and k_2 are non negative constants, and their values show the relative importance of the expectation and the variance of $\widehat{\text{Var}}(\bar{y}_{st})$. Some authors suggest that $k_1 + k_2 = 1$, see Rao (1978, p. 599). Thus, the equivalent deterministic problem to the stochastic problem (4) can be expressed as follows,

$$\begin{array}{c} \min\limits_{\mathbf{n}} f(\mathbf{n}) \\ \text{subject to} \\ \sum\limits_{h=1}^{H} c_h n_h + c_0 = C \\ n_h \in \mathbb{N}, \quad h = 1, 2, \dots, H, \end{array} \tag{8}$$

where

$$f(\mathbf{n}) = k_1 \left[\sum_{h=1}^{H} \frac{W_h^2}{(n_h - 1)} S_h^2 - \sum_{h=1}^{H} \frac{W_h}{N} \left(\frac{n_h}{n_h - 1} \right) S_h^2 \right]$$

$$+ k_2 \left[\sum_{h=1}^{H} \frac{W_h^4}{n_h (n_h - 1)^2} (C_{y_h}^4 - (S_h^2)^2) - \sum_{h=1}^{H} \frac{W_h^2}{N^2} \left(\frac{n_h}{(n_h - 1)^2} (C_{y_h}^4 - (S_h^2)^2) \right) \right]^{1/2}$$

It is important to see that the objective function is given in terms of the population variances S_h^2, which are unknown (by hypothesis), then we will use the sample variances s_h^2. Thus, the optimal equivalent deterministic problem to the stochastic programming (4) is given by

$$\begin{array}{c} \min\limits_{\mathbf{n}} \hat{f}(\mathbf{n}) \\ \text{subject to} \\ \sum\limits_{h=1}^{H} c_h n_h + c_0 = C. \\ n_h \in \mathbb{N}, \quad h = 1, 2, \dots, H, \end{array} \tag{9}$$

where, $\hat{f}(\mathbf{n})$ is given by

$$\hat{f}(\mathbf{n}) = k_1 \left[\sum_{h=1}^{H} \frac{W_h^2}{(n_h - 1)} s_h^2 - \sum_{h=1}^{H} \frac{W_h}{N} \left(\frac{n_h}{n_h - 1} \right) s_h^2 \right]$$

$$+ k_2 \left[\sum_{h=1}^{H} \frac{W_h^4}{n_h (n_h - 1)^2} (C_{y_h}^4 - (s_h^2)^2) - \sum_{h=1}^{H} \frac{W_h^2}{N^2} \left(\frac{n_h}{(n_h - 1)^2} (C_h^4 - (s_h^2)^2) \right) \right]^{1/2}$$

Remark 1 ***i)*** *If we take $k_1 = 1$ and $k_2 = 0$ in (9), the resulting method is known as the E-model. Alternatively, if $k_1 = 0$ and $k_2 = 1$, the method is called the V-model, see Charnes and Cooper (1963) and Prékopa (1995).*

ii) *Now, observe that the equivalent deterministic problem to the problem (3) comes from the solution of (1), established in (4), by substituting the constraint*

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

by the constraint

$$\sum_{h=1}^{H} n_h = n.$$

2.2 Chance constraints

Consider the stochastic programming for minimizing the cost function subject to a known bound for the estimated variance of the mean

$$\begin{array}{c}\min_{\mathbf{n}} \mathbf{c}^{'}\mathbf{n} + c_0\\ \text{subject to}\\ P\left[\widehat{\text{Var}}(\bar{y}_{st}) \leq V_0\right] \geq p_0\\ n_h \in \mathbb{N}, \text{ and} s_h^2 \text{ random variables }, h = 1, 2, \ldots, H,\end{array} \tag{10}$$

with

$$\widehat{\text{Var}}(\bar{y}_{st}) = \sum_{h=1}^{H} \frac{W_h^2}{n_h} s_h^2 - \sum_{h=1}^{H} \frac{W_h}{N} s_h^2,$$

where V_0 is a known non negative constant and p_0, $0 \leq p_0 \leq 1$ is an specified probability.

By (5), s_h^2 has an asymptotical normal distribution with mean $\text{E}(\xi_h)$ and variance $\text{Var}(\xi_h)$. Then the estimated variance $\widehat{\text{Var}}(\bar{y}_{st})$, in the stochastic problem (10), also has an asymptotical normal distribution with mean and variance given by (6) and (7), respectively.

By standardizing the function of the $\widehat{\text{Var}}(\bar{y}_{st})$, in the equation (10), we get

$$\text{P}\left[\frac{\widehat{\text{Var}}(\bar{y}_{st}) - \text{E}(\widehat{\text{Var}}(\bar{y}_{st}))}{\sqrt{\text{Var}(\widehat{\text{Var}}(\bar{y}_{st}))}} \leq \frac{V_0 - \text{E}(\widehat{\text{Var}}(\bar{y}_{st}))}{\sqrt{\text{Var}(\widehat{\text{Var}}(\bar{y}_{st}))}}\right] \geq p_0,$$

with

$$p_0 = \Phi\left(\frac{V_0 - \mathrm{E}(\widehat{\mathrm{Var}}(\bar{y}_{st}))}{\sqrt{\mathrm{Var}(\widehat{\mathrm{Var}}(\bar{y}_{st}))}}\right),$$

where $\Phi(\cdot)$, denotes the function of the standard normal distribution. Let e be the value of the standard normal random variable such that $\Phi(e) = p_0$, in such way that the inequality can be established as

$$\Phi\left(\frac{V_0 - \mathrm{E}(\widehat{\mathrm{Var}}(\bar{y}_{st}))}{\sqrt{\mathrm{Var}(\widehat{\mathrm{Var}}(\bar{y}_{st}))}}\right) \geq \Phi(e),$$

which holds only if

$$\frac{V_0 - \mathrm{E}(\widehat{\mathrm{Var}}(\bar{y}_{st}))}{\sqrt{\mathrm{Var}(\widehat{\mathrm{Var}}(\bar{y}_{st}))}} \geq e,$$

or equivalently

$$\mathrm{E}(\widehat{\mathrm{Var}}(\bar{y}_{st})) + e\sqrt{\mathrm{Var}(\widehat{\mathrm{Var}}(\bar{y}_{st}))} - V_0 \leq 0 \tag{11}$$

Then the equivalent deterministic problem to the stochastic programming (10), is given by

$$\begin{array}{c}\min\limits_{\mathbf{n}} \mathbf{c}'\mathbf{n} + c_o\\ \text{subject to}\\ \mathrm{E}(\widehat{\mathrm{Var}}(\bar{y}_{st})) + e\sqrt{\mathrm{Var}(\widehat{\mathrm{Var}}(\bar{y}_{st}))} - V_0 \leq 0\\ n_h \in \mathbb{N}, \quad h = 1, 2, \ldots, H,\end{array} \tag{12}$$

where

$$\mathrm{E}(\widehat{\mathrm{Var}}(\bar{y}_{st})) + e\sqrt{\mathrm{Var}(\widehat{\mathrm{Var}}(\bar{y}_{st}))} = \sum_{h=1}^{H} \frac{W_h^2}{(n_h - 1)} S_h^2 - \sum_{h=1}^{H} \frac{W_h}{N}\left(\frac{n_h}{n_h - 1}\right) S_h^2$$

$$+ e\left[\sum_{h=1}^{H} \frac{W_h^4}{n_h(n_h - 1)^2}(C_{y_h}^4 - (S_h^2)^2) - \sum_{h=1}^{H} \frac{W_h^2}{N^2}\left(\frac{n_h}{(n_h - 1)^2}(C_{y_h}^4 - (S_h^2)^2)\right)\right]^{1/2}.$$

Again, note that the function for the constraint in (12) is given in terms of the population variances S_h^2, then, by using the estimator of the variances s_h^2 in place of S_h^2, the following equivalent deterministic problem to (10) is obtained

$$\begin{array}{c} \min_{\mathbf{n}} \mathbf{c}^{'}\mathbf{n} + c_o \\ \text{subject to} \\ \widehat{\mathrm{E}}(\widehat{\mathrm{Var}}(\bar{y}_{st})) + e\sqrt{\widehat{\mathrm{Var}}(\widehat{\mathrm{Var}}(\bar{y}_{st}))} - V_0 \leq 0 \\ n_h \in \mathbb{N}, \quad h = 1, 2, \ldots, H, \end{array} \tag{13}$$

with

$$\widehat{\mathrm{E}}(\widehat{\mathrm{Var}}(\bar{y}_{st})) + e\sqrt{\widehat{\mathrm{Var}}(\widehat{\mathrm{Var}}(\bar{y}_{st}))} = \sum_{h=1}^{H} \frac{W_h^2}{(n_h - 1)} s_h^2 - \sum_{h=1}^{H} \frac{W_h}{N} \left(\frac{n_h}{n_h - 1} \right) s_h^2$$

$$+ e \left[\sum_{h=1}^{H} \frac{W_h^4}{n_h(n_h - 1)^2} (C_{y_h}^4 - (s_h^2)^2) - \sum_{h=1}^{H} \frac{W_h^2}{N^2} \left(\frac{n_h}{(n_h - 1)^2} (C_{y_h}^4 - (s_h^2)^2) \right) \right]^{1/2}.$$

Remark 2 *Finally, even though the problems for finding integer solutions were solved by considering the problems (1)-(3) as optimum programming of integers, two additional problems appear: the oversampling ($n_h > N_h$) and the case $n_h = 1$, because the variance in the stratum h cannot be estimated. Thus, adding the constraints*

$$2 \leq n_h \leq N_h, \quad h = 1, \ldots, H$$

these problems are solved.

3 Application

Consider the simulation of 187 observations, selected by random, as a previous study. The data correspond to the length of a leaf of certain species of flowers cultivated in four regions. The 187 data were obtained in the 4 regions, 48 in the first, 49 in the second, 47 in the third and 43 fourth. The region is the variable considered for the stratification, in such way that the population is divided in 4 strata (without overlapping), one for each region. Then, we obtain the following results:

TABLE 1: Data: stratum h, stratum size N_h, unitary cost c_h, sampling variance s_h^2 and fourth moment C_h^4.

h	N_h	c_h	s_h^2	C_h^4
1	2500	25	0.1694	0.0884
2	2300	30	8.4317	330.4106
3	2800	15	0.0972	0.0319
4	3000	25	3.8590	34.1001

The total population is $N = 10600$. The solution of all non linear programming of this application (including the deterministic, stochastic and integer programs) were obtained by the use of LINGO software.

Modified E-model

Minimization of the variance subject to a cost function

In this first example we want to minimize the variance restricted to a fixed cost which is given by the cost function. Then, by applying the modified E-model technique and considering a cost of $C = 5000$ we have the following equivalent deterministic problem

$$\begin{gathered}\min_{\mathbf{n}} f(\mathbf{n})\\ \text{subject to}\\ \sum_{h=1}^{H} c_h n_h = 5000\\ 2 \leq n_h \leq N_h,\ n_h \in \mathbb{N} \ \ h = 1, 2, \ldots, H,\end{gathered}$$

where

$$f(\mathbf{n}) = k_1 \left[\sum_{h=1}^{4} \frac{W_h^2}{(n_h - 1)} s_h^2 - \sum_{h=1}^{4} \frac{W_h}{10600} \left(\frac{n_h}{n_h - 1}\right) s_h^2\right]$$

$$+ k_2 \left[\sum_{h=1}^{4} \frac{W_h^4}{n_h(n_h - 1)^2} (C_{y_h}^4 - (s_h^2)^2) - \sum_{h=1}^{4} \frac{W_h^2}{(10600)^2} \left(\frac{n_h}{(n_h - 1)^2} (C_{y_h}^4 - (s_h^2)^2)\right)\right]^{1/2}$$

The solutions using mathematical programming are given in Table 2 and the solution via stochastic programming are showed in Table 3.

In Table 2, the rows 2 and 3 show the results obtained by the classical method of Lagrange multipliers. In row 4 we give the solution by a non linear programming of integers. Observe that the sample size (190 sample units) is the same for both methods (fitted Lagrange multipliers and non linear programming of integers), however, the allocation varies by a unit in the strata 3 and 4. The generated allocation by a non linear programming of integers gives 16 units to the stratum 3 and 78 to the stratum 4, which is a more convenient allocation because, the cost by sample unit in the stratum 3 is less than the cost by sample unit in the stratum 4. As we expected, the greater amount of sample units was assigned to the strata 2 and 4, because these strata have the greatest variance.

Table 2: Minimization of the variance restricted to a cost function. Solutions of the optimum allocation problem in stratified surveys under the different deterministic optimization approaches.

Solution	n_1	n_2	n_3	n_4	n
Lagrange	13.795	81.758	15.113	79.026	189.69
Rounded [a]	14	82	15	79	190
Integer [b]	14	82	16	78	190

[a] This solution was obtained using Lagrange method and rounding each n_i, see Cochran (1977, pag.139). Alternatively, in the statistical literature this can be fitted also to the next integer.

[b] This solution was found by a non linear programming of integers with the Branch cut method.

TABLE 3: Minimization of the variance restricted to a cost function. Solutions of the optimum allocation problem in stratified surveys under different stochastic optimization approaches. .

Solution	n_1	n_2	n_3	n_4	n
Modified E-model (Lagrange) [a]	14.534	81.211	15.828	78.515	190.08
Modified E-model (rounded) [b]	15	81	16	79	191
Modified E-model (integer) [c]	14	82	16	78	190
V-model (Lagrange)	12.080	103.986	16.508	53.231	185.8
V-model (rounded)	12	104	17	53	186
V-model (integer)	13	94	17	64	188

[a] Where $k_1 = k_2 = 0.5$

[b] This solution was obtained by using Lagrange method and rounding each n_i, see Cochran (1977, pag.139). Alternatively, in the statistical literature this can be fitted also to the next integer.

[c] This solution was found by a non linear stochastic programming of integers with the Branch cut method.

Note that in Table 3, the stochastic solutions under the Lagrange and the rounding methods differ each other when they are compared with the deterministic solutions; this occurs mainly in the strata 1 and 3 (which have minor variance). Although the sample size are almost the same for all the solutions (190 sample units); the stochastic solution assigns one unit more to the stratum with minor variance. If we compare the stochastic optimization programming of integers in the modified E-model for $k_1 = 0.25$ and $k_2 = 0.75$, $k_1 = 0.75$ and $k_2 = 0.25$, with the E-model, the allocations to the strata are the same that the solutions for modified E-model with $k_1 = 0.5$ and $k_2 = 0.5$, but the solutions under the Lagrange and the rounding methods differ a little bit each other. Finally, observe that under the V-model as the sample size as the allocations to the stratum vary considerably with respect to the E-model, the modified E-model and the deterministic solution. Moreover, the allocation to the strata of greater variance (strata 2 and 4) drastically vary when we consider the stochastic programming of integers.

Minimization of the variance subject to a fixed sample size

Now we want to minimize the sample variance of the mean constrained to a fixed sample size. In this case we propose the value $n = 190$. Thus the problem for minimizing is

$$\min_{\mathbf{n}} f(\mathbf{n})$$
$$\text{subject to}$$
$$\sum_{h=1}^{H} n_h = 190$$
$$2 \leq n_h \leq N_h,\ n_h \in \mathbb{N} \quad h = 1, 2, \ldots, H,$$

where

$$f(\mathbf{n}) = k_1 \left[\sum_{h=1}^{4} \frac{W_h^2}{(n_h - 1)} s_h^2 - \sum_{h=1}^{4} \frac{W_h}{10600} \left(\frac{n_h}{n_h - 1} \right) s_h^2 \right]$$

$$+ k_2 \left[\sum_{h=1}^{4} \frac{W_h^4}{n_h (n_h - 1)^2} (C_{y_h}^4 - (s_h^2)^2) - \sum_{h=1}^{4} \frac{W_h^2}{(10600)^2} \left(\frac{n_h}{(n_h - 1)^2} (C_{y_h}^4 - (s_h^2)^2) \right) \right]^{1/2}$$

Table 4 shows the results when the problem is considered as a deterministic optimization programming of integers.

TABLE 4: Minimization of the variance restricted to a sample size. Solutions of the optimum allocation problem in the stratified survey under different deterministic optimization approaches.

Solution	n_1	n_2	n_3	n_4
Lagrange	13.507	87.677	11.461	77.354
Rounded	14	88	11	77
Integer	14	88	11	77

In this case, the rounded solution coincides with the solution of integer optimization. As we expect, both assign the greater sample size to the strata 2 and 4 which have greater variances.

When the stochastic programming is considered, the solutions of this problem are given in Table 5 . Even though the solutions via Lagrange multipliers vary a few for different combinations of the values of k_1 and k_2 ($k_1 = 0.5$, $k_2 = 0.5$; $k_1 = 0.25$, $k_2 = 0.75$ and $k_1 = 0.75$, $k_2 = 0.25$). The rounded solutions and the the integer solutions coincide. In such cases one unit less is assigned to the stratum 2, which is assigned to the stratum 3, when we compare with the solution given by the deterministic programming. For the V-model solution, again a drastic contrast occurs: first, between this solution and the around deterministic solution; and second, between the rounded solution and the optimization of integers solution.

TABLE 5: Minimization of the variance restricted to a sample size. Solutions of the optimum allocation problem in the stratified survey under different stochastic optimization approaches.

Solution	n_1	n_2	n_3	n_4
Modified E-model (Lagrange) [a]	14.239	86.819	12.245	76.695
Modified E-model (rounded)	14	87	12	77
Modified E-model (integer)	14	87	12	77
V-model (Lagrange)	10.461	122.475	19.133	37.931
V-model (rounded)	10	122	19	38
V-model (integer)	16	92	18	64

[a] Where $k_1 = k_2 = 0.5$

Chance constraints

The present second problem minimizes the cost function restricted to a specified variance tolerance $V_0 = .015$ and a $p_0 = 0.99$. Under these conditions, the equivalent deterministic problem is given by

$$\min_{\mathbf{n}} \mathbf{c}'\mathbf{n}$$
$$\text{subject to}$$

$$\sum_{h=1}^{4}\frac{W_h^2}{(n_h-1)}s_h^2 - \sum_{h=1}^{4}\frac{W_h}{10600}\left(\frac{n_h}{n_h-1}\right)s_h^2 + 2.3263$$
$$\times\left[\sum_{h=1}^{4}\frac{W_h^4}{n_h(n_h-1)^2}(C_{y_h}^4-(s_h^2)^2) - \sum_{h=1}^{4}\frac{W_h^2}{(10600)^2}\left(\frac{n_h}{(n_h-1)^2}(C_{y_h}^4-(s_h^2)^2)\right)\right]^{1/2} \leq 0.015$$
$$2 \leq n_h \leq N_h,\ n_h \in \mathbb{N}\quad h = 1, 2, \ldots, H,$$

The results are summarized in Table 6.

Table 6: Minimization of the cost function restricted to the variance. Solutions to the optimum allocation problem in the stratified survey under the different deterministic and stochastic approaches of optimization.

Solution	n_1	n_2	n_3	n_4	n
Lagrange	8.936	52.951	9.787	51.175	122.849
Rounded	9	53	10	51	123
Integer	9	53	10	51	123
Stochastic (Lagrange)	9.934	53.965	10.784	52.160	126.84
Stochastic (rounded)	10	54	11	52	127
Stochastic (integer)	10	54	11	52	127

Rows 2 and 3 show the results under Lagrange multiplier method and row 4 gives the results when we consider a non linear programming problem of integers. Note that the sample size increases only one unit when the last method is used (124 sample units), but the allocation of sample units basically changes in all the strata (by one unit). This allocation is more convenient because the cost by sample unit is small where one unit is increased. Perhaps, the difference is not radical for this case, because the cost variation is not extremal, but if the cost for sampling a unit is high, then that difference would represent an important economic benefit.

For this problem, the stochastic solutions differ almost 4 units from the deterministic optimization solutions, and although the budget is increased, the difference in sample units would represent a greater contribution in terms of precision.

Acknowledgments

This study was carried out with the assistance of CONACYT-45974-F and MTM2005-09209-Ministerio de Educación y Ciencia, España.

References

Arthanari, T. S., and Dodge Y. (1981). *Mathematical Programming in Statistics*. John Wilwy & Sons, New York.

Charnes, A., and Cooper, W. W., (1963). Deterministic equivalents for optimizing and satisficing under chance constraints. *Operat. Res.*, **11**, 18-39.

Cochran, W. G. 1977. *Sampling techniques.* Wiley, New York.

Díaz García, J. A., and Ramos-Quiroga, R., 2001. An approach to optimization in response surfaces. *Comm. in Stat., Part A- Theory and Methods*, **30**(5), 827-835.

Díaz García, J. A., and Ramos-Quiroga, R., (2002). Erratum. An approach to optimization in response surfaces. *Comm. in Stat., Part A- Theory and Methods*, **31**(1), 161.

Díaz García, J. A., Ramos-Quiroga, R., and Cabrera-Vicencio, E., (2005). Stochastic programming methods in the response surface methodology. *Com. Statist. Data Analy.*, **49**, 837-848.

Melaku, A. (1986). Asymptotic Normality of the Optimal Allocation in Multivariate Stratified Randon Sampling. *The Indian Journal of Statistics*, Vol. **48**, Series B, pág. 224-232.

Prékopa A. (1978). The use of stochastic programing for the solution of the some ploblems in statistics and probability. Technical Summary report 1983. University of Wisconsin-Madison.

Prékopa A. (1995). *Stochastic Programming.* Kluwer Academic Publishers, Serie Mathematics and its Applications.

Rao, S. S. 1978-79. *Optimization theory and applications.* Wiley Eastern Limited.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Asok, C. (1984). *Sampling theory of surveys with applications.* Third edition. Ames, Iowa: Iowa State University Press.