CHAPTER 3

OPTIMUM MULTIVARIATE STRATIFIED SAMPLING DESIGN WITH TRAVEL COST: A MULTIOBJECTIVE INTEGER NONLINEAR PROGRAMMING APPROACH
CHAPTER - 3
OPTIMUM MULTIVARIATE STRATIFIED SAMPLING DESIGN WITH TRAVEL COST: A MULTIOBJECTIVE INTEGER NONLINEAR PROGRAMMING APPROACH

3.1 INTRODUCTION

An effective sampling technique within a population represents an appropriate extraction of useful data which provides meaningful knowledge of the important aspects of the population. Stratified sampling is one of the classical methods for obtaining such information. The efficiency advantage and the ability to generate stratum specific estimates are reason for considering stratified sampling design. Selecting the number of observations from each stratum is a primary decision in stratified random sampling design which can be computed by various procedures. Generally, allocation schemes aim to minimize or bound the variance associated while estimating some overall population parameter, subject to a limitation on sampling resources.

In multivariate surveys where more than one characteristics are defined on each unit of the population, an allocation criterion among various strata that is uniformly optimum does not exist unless the characteristics are highly correlated. If the characteristics are independent the individual
optimum allocations may be widely apart and there will be no obvious compromise. In such situations we need a compromise criterion that gives an allocation which is optimum for all characteristics in some sense. When the estimation of \( p \)-population means is of interest a most commonly used compromise criterion is to minimize the sum of the sampling variances of the estimators. But various authors objected on adding sampling variances because different characteristics are measured in different units and thus can not be added.

For a population the coefficient of variation (CV) is represented by the ratio of population standard deviation to the population mean. The CV is used to compare the precision of various estimates that are measured in different units. Ostle (1954) found that the population CV is an ideal device for comparing the variation in two series of data which are measured in two different units. In this chapter the squared CVs are used to build a compromise criterion to obtain a compromise allocation.

The simplest cost function used in a stratified sample survey is a linear function of sample sizes \( n_h \) given as

\[
C = c_0 + \sum_{h=1}^{L} c_h n_h
\]  

(3.1)
where \( c_h; \ h = 1, 2, ..., L \) denote the per unit cost of measurement in the \( h \)-th stratum and \( c_0 \) denotes the overhead cost. If the cost of traveling between the selected units of a stratum is significant then the linear cost function may not be a good approximation to the actual cost incurred. Beardwood et al. (1959) suggested that the cost of visiting the \( n_h \) randomly selected units in the \( h \)-th stratum may be taken as \( t_h \sqrt{n_h} \; ; \ h = 1, 2, ..., L \) approximately, where \( t_h \) is the travel cost per unit in the \( h \)-th stratum. This conjecture is based on the fact that the total shortest distance between \( k \) randomly scattered points is proportional to \( \sqrt{k} \).

Thus if the travel cost within stratum between units selected in the sample is substantial then the total cost \( C \) may be expressed more adequately as a sum of the overhead cost, the measurement cost and the traveling cost within strata as

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

(3.2)

RHS of expression (3.2) is a quadratic function in \( \sqrt{n_h} \).

In this chapter a method is developed to work out the compromise allocation in a multivariate stratified survey using the compromise criterion “Minimize the sum of squared coefficients of variation of the estimators”
subject to the quadratic cost constraint (3.2). The problem is formulated as a Multiobjective All Integer Non Linear Programming Problem (MAINLPP). Using the prior information about the population, which are classified as complete, partial or null, (see Ríos, Ríos Insua and Ríos Insua (1989), Miettinen (1999) and Steuer (1986)), different techniques are studied to solve the formulated MAINLPP. A numerical example is also worked out to illustrate the computational details of the methods. The numerical solutions are obtained through the optimization software LINGO (2001).


3.2 NOTATIONS

Consider a population of size $N$, divided into $L$ non overlapping strata of sizes $N_h$; $h = 1,2,\ldots,L$, $N = \sum_{h=1}^{L} N_h$. Let $p \geq 2$, characteristics be defined on each population unit and the estimation of $p$-population means are of interest. Independent simple random samples of sizes
$n_h; \ h=1,2,...,L,$ are drawn from the $L$ strata. The symbols used in this chapter are as follows.

For the $h^{th}$ stratum

$N_h =$ Stratum size.

$n_h =$ Sample size.

$$n = \sum_{h=1}^{L} n_h = \text{Total sample size}$$

$x_{jhi} =$ Value of the $i-$th unit in the $h-$th stratum of the $j-$th characteristic.

$W_h = \frac{N_h}{N} =$ Stratum weight.

$$\bar{X}_{jh} = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{jhi} = \text{Stratum mean for the } j-\text{th characteristic}.$$  

$$\bar{x}_{jh} = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{jhi} = \text{Sample mean}.$$  

$$\bar{X}_j = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} x_{jhi} = \sum_{h=1}^{L} W_h \bar{X}_{jh} = \text{Overall population mean of the } j-\text{th characteristic}.$$
\[ \bar{x}_{sl, j} = \frac{1}{n} \sum_{h=1}^{L} \sum_{l=1}^{n_h} x_{jhi} = \sum_{h=1}^{L} W_h \bar{x}_{jh} \]  

Stratified sampling mean for the \( j \)-th characteristic.

\[ V(\bar{x}_{sl, j}) = \sum_{h=1}^{L} W_h^2 (1 - \frac{n_h}{N_h}) \frac{S^2_{jh}}{n_h} \]  

Sampling variance of \( \bar{x}_{sl, j} \).

\[ S^2_{jh} = \frac{1}{(N_h - 1)} \sum_{i=1}^{N_h} (x_{jhi} - \bar{X}_{jh})^2 \]  

Stratum variances.

\[ (CV)_j = CV(\bar{x}_{sl, j}) = \frac{SD(\bar{x}_{sl, j})}{\bar{X}_j}; \; j = 1, 2, \ldots, p \]  

Population Coefficient of Variation of the estimator \( \bar{x}_{sl, j} \).

### 3.3 FORMULATION OF THE PROBLEM

The squared Population Coefficient of Variation of the \( j \)-th characteristic

\[ (CV)_j^2 = \frac{V(\bar{x}_{sl, j})}{\bar{X}_j^2}; \; j = 1, 2, \ldots, p \]

\[ = \bar{X}_j^{-2} \left\{ \sum_{h=1}^{L} W_h^2 (1 - \frac{n_h}{N_h}) \frac{S^2_{jh}}{n_h} \right\} \]  

(3.3)

The MAINLPP for finding out the optimum compromise allocation with a quadratic cost constraint may now be expressed as:
Optimum Multivariate Stratified Sampling...

Minimize \[
\begin{pmatrix}
(CV)_1^2 \\
\vdots \\
(CV)_p^2
\end{pmatrix}
\]

Subject to \[
\sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0 \\
2 \leq n_h \leq N_h \\
\text{and} \quad n_h \text{ are integers; } h = 1, 2, \ldots, L
\]

The restrictions \(2 \leq n_h \leq N_h; \ h = 1, 2, \ldots, L\) are imposed to obtain the estimates of the stratum variances and to avoid the problem of oversampling.

Using the compromise criterion stated in section 3.1 the MAINLPP (3.4) may be expressed as the All Integer Nonlinear Programming Problem (AINLPP)

Minimize \[
\sum_{j=1}^{p} (CV)_j^2
\]

Subject to \[
\sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0 \\
2 \leq n_h \leq N_h \\
\text{and} \quad n_h \text{ are integers; } h = 1, 2, \ldots, L
\]

Substituting the values of \((CV)_j^2\) from (3.3) in the objective function of (3.5) we get the AINLPP as:

79
Optimum Multivariate Stratified Sampling...

\[
\text{Minimize} \quad Z = \sum_{j=1}^{p} \bar{X}_j^{-2} \left\{ \sum_{h=1}^{L} W_h^2 \left( 1 - \frac{n_h}{N_h} \right) S_{jh}^2 \right\}
\]

\[
\text{Subject to} \quad \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0
\]

\[
2 \leq n_h \leq N_h
\]

\[
\text{and} \quad n_h \text{ are integers; } h = 1, 2, ..., L.
\]

(3.6)

3.4 OPTIMIZATION METHODS FOR SOLVING MULTIOBJECTIVE PROGRAMMING PROBLEM

Methods to solve the multiobjective programming problems can be classified according to the available information about the population, that is, complete, partial or null information.

3.4.1 The Value Function Approach

A function which represents the preferences of the decision maker among the objective vectors is called a value function. The value function is totally a decision maker's concept. Different decision makers have different value functions for the same problem. It offers a total (complete) ordering of the objective vector. The value functions are seldom explicitly used in solving multiobjective optimization problems but they are very important in the development of solution methods as a theoretical background. In many multiobjective optimization methods, the value function is assumed to be known implicitly and the decision maker is assumed to make selection on the basis of this knowledge (see Zionts (1997 a,b)).
Generally, the value function is assumed to be strongly decreasing, that is, the preference of the decision maker will increase if the value of an objective function decreases (for minimization problems) while all the other objective values remain unchanged (see Rosenthal (1985)).

This method is used when complete information about the population is available, that is, when the relative importance of each characteristic is known. In such cases relative weights can be assigned to them. The problem (3.6) under the value function technique may be expressed as

\[
\begin{align*}
\text{Minimize} & \quad \phi \left( \sum_{j=1}^{p} (CV_j)^2 \right) \\
\text{Subject to} & \quad \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0 \\
& \quad 2 \leq n_h \leq N_h \\
& \quad n_h \text{ are integers; } h = 1, 2, \ldots, L
\end{align*}
\] (3.7)

where \(\phi(\cdot)\) is a scalar function that summarizes the importance of each of the \(p\)-characteristics. For different problems the value function \(\phi(\cdot)\) takes a form appropriate to the nature of the optimization problem. In the present chapter \(\phi(\cdot)\) is define as the weighted sum of squared CV’s. Under this definition problem (3.7) becomes:
Minimize \[ \sum_{j=1}^{p} \lambda_j (CV)_j^2 \]

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

\[ 2 \leq n_h \leq N_h \]

and \[ n_h \text{ are integers; } h = 1, 2, \ldots, L \]

where \( \lambda_j \geq 0, j = 1, 2, \ldots, p, \) are the weights according to the relative importance of the characteristics.

When complete information is available the weights, may be decided according to some measures of the relative importance of the characteristics.

For example weights \( \lambda_j \) may be taken as:

\[ \lambda_j \propto \sum_{h=1}^{L} S_{jh}^2, j = 1, 2, \ldots, p \]

or

\[ \lambda_j = k \sum_{h=1}^{L} S_{jh}^2 \quad (3.9) \]

where \( k \) is the constant of proportionality.

Without loss of generality we can assume that \[ \sum_{j=1}^{p} \lambda_j = 1. \] Summing (3.9) for \( j = 1, 2, \ldots, p \) we get

\[ \sum_{j=1}^{p} \lambda_j = k \sum_{j=1}^{p} \sum_{h=1}^{L} S_{jh}^2 \]
or
\[
k = \frac{1}{\sum_{j=1}^{p} \sum_{h=1}^{L} S_{jh}^2}
\]

This gives,
\[
\lambda_j = \frac{\sum_{h=1}^{L} S_{jh}^2}{\sum_{j=1}^{p} \sum_{h=1}^{L} S_{jh}^2}; j = 1, 2, \ldots, p \quad \text{(See Khan et al. (2003))}.
\]

Using (3.5) AINLPP (3.8) may be restated as:
\[
\begin{cases}
\text{Minimize} & \sum_{j=1}^{p} \lambda_j \sqrt{X_j} \left( \sum_{h=1}^{L} W_h^2 \left( 1 - \frac{n_h}{N_h} \right) S_{jh}^2 \right) \\
\text{Subject to} & \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0 \\
& 2 \leq n_h \leq N_h \\
\text{and} & n_h \text{ are integers; } h = 1, 2, \ldots, L
\end{cases}
\]

(3.10)

3.4.2 The \( \varepsilon \)-Constraint Approach

In the \( \varepsilon \)-constraint method, introduced by Haimes et al. (1971), one of the objective functions is selected to be optimized and all the other objective functions are placed into constraints by setting an upper bound to each of them. This method is used when only partial information are available. To apply this method the investigator needs to identify the most
important characteristics only. Thus the complete information about all the characteristics are not required.

Under this approach we may express the problem of obtaining the integer compromise allocation as

\[
\begin{align*}
\text{Minimize} & \quad (CV)^2_k \\
\text{Subject to} \quad & (CV)^2_r \leq (cv)^2_r, \quad r \neq k, r = 1, 2, \ldots, p \\
& \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0 \\
& 2 \leq n_h \leq N_h \\
& \quad \text{and} \quad n_h \text{ are integers; } h = 1, 2, \ldots, L
\end{align*}
\]

(3.11)

where, the \(k^{th}\) characteristic \(k \in \{1, 2, \ldots, p\}\), is assumed to be the most important one and \((cv)^2_r\) is a predetermined upper bound for the remaining \(p-1\) coefficients of variation.

In practice we can take \((cv)^2_r\) as the solutions to the following problems:

\[
\begin{align*}
\text{Minimize} & \quad (CV)^2_r \\
\text{Subject to} \quad & \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0 \\
& 2 \leq n_h \leq N_h \\
& \quad \text{and} \quad n_h \text{ are integers; } h = 1, 2, \ldots, L
\end{align*}
\]

(3.12)

Note that the choice of \(k^{th}\) characteristic and the upper limits \((cv)^2_r\) represent the evaluator's subjective preferences, and so if there were no
solution to problem (3.11), this would mean that one or more of the bounds on \((CV)^2\) have been set too low and they must be revised, (see Ríos, Ríos Insua and Ríos Insua (1989)).

3.4.3 The Distance Based Approach

In many situations, the investigator does not have sufficient information about the characteristics, or it is difficult to decide which is the most important characteristic. In such situations distance based method can be applied, as it requires only a vector of ideal goals which can be determined with the null information (see Ríos et al. (1989) and Steuer (1986)).

With distance based approach AINLPP (3.5) may be expressed as follows.

Define the \(p\)-component vector \(\alpha\) of targets as

\[
\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}
\]

where \(\alpha_j\) is the ideal point or goal for the objective \((CV)^2_j\); \(j = 1, 2, \ldots, p\).

This vector of targets \(\alpha\) can be computed by minimizing each objective \((CV)^2_j\); \(j = 1, 2, \ldots, p\) separately. Thus \(\alpha\) is the vector of individual
constrained minima, which can be obtained by solving the following $p-$AINLPP separately.

$$
\text{Minimize } (CV)_j^2
$$

$$
\text{Subject to } \begin{cases} 
\sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0 & j = 1, 2, \ldots, p \\
2 \leq n_h \leq N_h \\
\text{and } n_h \text{ are integers; } h = 1, 2, \ldots, L 
\end{cases} 
$$

Equation (3.13)

After computing $\alpha$, the problem of obtaining a compromise allocation may be given as:

$$
\text{Minimize } D[(CV)_j^2, \alpha_j]
$$

$$
\text{Subject to } \begin{cases} 
\sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0 & j = 1, 2, \ldots, p \\
2 \leq n_h \leq N_h \\
\text{and } n_h \text{ are integers; } h = 1, 2, \ldots, L 
\end{cases} 
$$

Equation (3.14)

where $D[(CV)_j^2, \alpha_j] = \sqrt{\sum_{j=1}^{p}[(CV)_j^2 - \alpha_j]^2}$ is the Euclidean distance between vectors $(CV)_j^2$ and $(\alpha_j)$. Since minimizing $\sqrt{\sum_{j=1}^{p}[(CV)_j^2 - \alpha_j]^2}$ is equivalent to minimize $\sum_{j=1}^{p}[(CV)_j^2 - \alpha_j]^2$ this gives the AINLPP (3.14) as:
Optimum Multivariate Stratified Sampling...

\[
\begin{align*}
\text{Minimize} & \quad \sum_{j=1}^{p} [(CV)_j^2 - \alpha_j]^2 \\
\text{Subject to} & \quad \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0 \\
& \quad 2 \leq n_h \leq N_h \\
& \quad \text{and} \quad n_h \text{ are integers; } h=1,2,\ldots,L
\end{align*}
\] (3.15)

Alternatively, Khuri and Cornell (1986) proposed another distance given by

\[
\sum_{j=1}^{p} \left( \frac{(CV)_j^2 - \alpha_j}{\alpha_j} \right)^2.
\]

With this distance function the problem of obtaining the com22romise allocation may be given as the following AINLPP:

\[
\begin{align*}
\text{Minimize} & \quad \sum_{j=1}^{p} \frac{[(CV)_j^2 - \alpha_j]^2}{\alpha_j^2} \\
\text{Subject to} & \quad \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \leq C_0 \\
& \quad 2 \leq n_h \leq N_h \\
& \quad \text{and} \quad n_h \text{ are integers; } h=1,2,\ldots,L
\end{align*}
\] (3.16)

3.5 A NUMERICAL ILLUSTRATION

The following numerical example is presented to illustrate the practical utility and the computational details of the proposed allocation. The data are from Agricultural Censuses (2002) in Iowa State, USA conducted by National Agricultural Statistics Service, USDA, Washington D.C. (source: http://www.agcensus.usda.gov/). The 99 counties in Iowa State are
Optimum Multivariate Stratified Sampling...

divided into four strata. The relevant data with respect to two characteristics
(i) the quantity of corn harvested $X_1$, and
(ii) the quantity of oats harvested $X_2$

are given in Table 3.1

<table>
<thead>
<tr>
<th>$h$</th>
<th>$N_h$</th>
<th>$W_h$</th>
<th>$S^2_{1h}$</th>
<th>$S^2_{2h}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>0.0808</td>
<td>21601503189.8</td>
<td>1154134.2</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
<td>0.3434</td>
<td>19734615816.7</td>
<td>7056074.8</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>0.4545</td>
<td>27129658750.0</td>
<td>2082871.3</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>0.1212</td>
<td>17258237358.5</td>
<td>732004.9</td>
</tr>
</tbody>
</table>

Also $\bar{X}_1$ and $\bar{X}_2$ are assumed to be known as $\bar{X}_1 = 474973.90$ and $\bar{X}_2 = 1576.25$. It is of course untrue in real surveys. In practice some approximations of these parameters are used. They can be known from a recent or preliminary survey (Kozak (2006)).

We assumed the cost of measurement $c_h$ in the four strata as $c_1 = 15$, $c_2 = 7$, $c_3 = 5$ and $c_4 = 9$ units respectively and the total amount available for the measurements is $C_0 = (C - c_0) = 200$ units, where $c_0 = 50$ units, is the expected overhead cost and $C = 250$ units is the total budget of the survey. The travel cost per unit in the $h^{th}$ stratum $t_h$; $h = 1, 2, \ldots, 4$ are assumed as $t_1 = 10$, $t_2 = 5$, $t_3 = 2$ and $t_4 = 6$ units respectively.
3.5.1 Value Function Approach

For simplicity we assumed that both the characteristics are equally important that is, \( \lambda_1 = \lambda_2 = 0.5 \). For the values given in Table 3.1 the AINLPP (3.10) becomes

\[
\begin{align*}
\text{Minimize} & \quad (0.5) \left[ \left( \frac{1}{225600205681.21} \right) \left( \frac{1}{8} \right) \frac{190177129.163709}{n_1} \\
& \quad \left( 1 - \frac{n_2}{34} \right) \frac{3075358778.397211}{n_2} \left( 1 - \frac{n_3}{45} \right) \frac{8750902957.838071}{n_3} \left( 1 - \frac{n_4}{12} \right) \frac{451381009.970472}{n_4} \\
& \quad \left( 1 - \frac{n_2}{34} \right) \frac{588180.653800}{n_2} \left( 1 - \frac{n_3}{45} \right) \frac{221961.923300}{n_3} \left( 1 - \frac{n_4}{12} \right) \frac{5705.062673}{n_4} \right] \\
\text{Subject to} & \quad 15n_1 + 7n_2 + 5n_3 + 9n_4 + 10\sqrt{n_1} + 5\sqrt{n_2} + 2\sqrt{n_3} + 6\sqrt{n_4} \leq 200 \\
& \quad 2 \leq n_1 \leq 8 \\
& \quad 2 \leq n_2 \leq 34 \\
& \quad 2 \leq n_3 \leq 45 \\
& \quad 2 \leq n_4 \leq 12 \\
\text{and} & \quad n_h \text{ integers; } h = 1,2,...,4 
\end{align*}
\]

(3.17)

The compromise allocation under value function approach \( n_{\text{VF}}^* = (n_1^*,n_2^*,n_3^*,n_4^*) \) is obtained as \( n_1^* = 2, n_2^* = 9, n_3^* = 9, n_4^* = 2 \), where the suffix ‘VF’ stands for ‘value function’. The corresponding value of the
objective function is 0.01735876. The solution is obtained using the
optimization software LINGO.

3.5.2 $\varepsilon$-Constraint Approach

Using the values given in Table 3.1 the AINLIPP (3.6) and their
optimal solutions $n_j^*$; $j = 1, 2$ with the corresponding values of objective
functions are listed below.

For $j=1$

$$
\begin{align*}
\text{Minimize} & \quad (1 - \frac{n_1}{8}) \frac{190177129.163709}{n_1} + (1 - \frac{n_2}{34}) \frac{3075358778.397211}{n_2} + (1 - \frac{n_3}{45}) \frac{8750902957.838071}{n_3} \\
& \quad + (1 - \frac{n_4}{12}) \frac{451381009.970472}{n_4} \\
\text{Subject to} \quad & \quad 15n_1 + 7n_2 + 5n_3 + 9n_4 + 10\sqrt{n_1} + 5\sqrt{n_2} + 2\sqrt{n_3} + 6\sqrt{n_4} \leq 200 \\
& \quad 2 \leq n_1 \leq 8 \\
& \quad 2 \leq n_2 \leq 34 \\
& \quad 2 \leq n_3 \leq 45 \\
& \quad 2 \leq n_4 \leq 12 \\
\text{and} & \quad n_h \text{ integers; } h = 1, 2, ..., 4
\end{align*}
$$

(3.18)

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

$n_{11}^* = 2$, $n_{12}^* = 7$, $n_{13}^* = 12$, $n_{14}^* = 2$. The corresponding value of the objective
function is 0.005068222. This gives $(cv)^2 = 0.005068222$. 

90
For \( j = 2 \)

Minimize \[ \frac{1}{2484564.0625} \left[ (1 - \frac{n_1}{8}) \frac{5073.889916}{n_1} + (1 - \frac{n_2}{34}) \frac{588180.653800}{n_2} + (1 - \frac{n_3}{45}) \frac{221961.923300}{n_3} + (1 - \frac{n_4}{12}) \frac{5705.062673}{n_4} \right] \]

Subject to \[ 15n_1 + 7n_2 + 5n_3 + 9n_4 + 10\sqrt{n_1} + 5\sqrt{n_2} + 2\sqrt{n_3} + 6\sqrt{n_4} \leq 200 \]

\[ 2 \leq n_1 \leq 8 \]
\[ 2 \leq n_2 \leq 34 \]
\[ 2 \leq n_3 \leq 45 \]
\[ 2 \leq n_4 \leq 12 \] (3.19)

and \( n_h \) integers; \( h = 1, 2, \ldots, 4 \)

The optimum allocation \( n^*_2 = (n^*_{21}, n^*_{22}, n^*_{23}, n^*_{24}) \) is \( n^*_{21} = 2, n^*_{22} = 9, n^*_{23} = 9, n^*_{24} = 2 \). The corresponding value of the objective function is 0.0200458. This gives \((cv)^2_2 = 0.0200458\).

Assuming the second characteristic as more important, that is, \( k = 2 \) the problem (3.11) becomes
Minimize \( \left( \frac{1}{2484564.0625} \right) [(1 - \frac{n_1}{8}) \frac{5073.889916}{n_1} + (1 - \frac{n_2}{34}) \frac{588180.653800}{n_2} + (1 - \frac{n_3}{45}) \frac{221961.923300}{n_3} + (1 - \frac{n_4}{12}) \frac{5705.062673}{n_4}] \).

Subject to \( 15n_1 + 7n_2 + 5n_3 + 9n_4 + 10\sqrt{n_1} + 5\sqrt{n_2} + 2\sqrt{n_3} + 6\sqrt{n_4} \leq 200 \)

\( \left( \frac{1}{225600205681.21} \right) [(1 - \frac{n_1}{8}) \frac{190177129.163709}{n_1} + (1 - \frac{n_2}{34}) \frac{3075358778.397211}{n_2} + (1 - \frac{n_3}{45}) \frac{875902957.838071}{n_3} + (1 - \frac{n_4}{12}) \frac{451381009.970472}{n_4} ] \leq 0.005068222 \)

\( 2 \leq n_1 \leq 8 \)
\( 2 \leq n_2 \leq 34 \)
\( 2 \leq n_3 \leq 45 \)
\( 2 \leq n_4 \leq 12 \)

and \( n_h \) integers; \( h = 1, 2, ..., 4 \)

\( (3.20) \)

The compromise allocation under \( \epsilon \)-constraint approach \( n^*_\epsilon = (n_1^*, n_2^*, n_3^*, n_4^*) \) is obtained as \( n_1^* = 2, n_2^* = 7, n_3^* = 12, n_4^* = 2 \). The corresponding value of the objective function is 0.03403838.

3.5.3 Distance Based Approach

The vector of targets obtained from (3.18) and (3.19) is

\[ \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} (cv)^2 \\ (cv)^2 \end{pmatrix} = \begin{pmatrix} 0.005068222 \\ 0.020045800 \end{pmatrix} \]

as worked out under \( \epsilon \)-constraint approach.

92
The problem (3.15) can now be stated as:

\[
\begin{align*}
\text{Minimize} \quad & \left[ \frac{1}{225600205681.21} \left( (1 - \frac{n_1}{8})^{190177129.163799} + 
\right. \\
& \left. \left(1 - \frac{n_2}{34}\right)^{3075358778.397211} + (1 - \frac{n_3}{45})^{8750902957.838071} + 
\right. \\
& \left. \left(1 - \frac{n_4}{12}\right)^{451381009.970472} - 0.0050068222 \right] \right] \quad + \\
& \left[ \frac{1}{2484564.0625} \left( (1 - \frac{n_1}{8})^{5073.889916} + 
\right. \\
& \left. \left(1 - \frac{n_2}{34}\right)^{588180.653800} + (1 - \frac{n_3}{45})^{221961.923300} + 
\right. \\
& \left. \left(1 - \frac{n_4}{12}\right)^{5705.062673} - 0.0200458 \right] \right)^2
\end{align*}
\]

Subject to \( 15n_1 + 7n_2 + 5n_3 + 9n_4 + 10\sqrt{n_1} + 5\sqrt{n_2} + 2\sqrt{n_3} + 6\sqrt{n_4} \leq 200 \\
2 \leq n_1 \leq 8 \\
2 \leq n_2 \leq 34 \\
2 \leq n_3 \leq 45 \\
2 \leq n_4 \leq 12 \\
\text{and} \\
n_h \text{ integers; } h = 1,2,...,4
\]

(3.21)

The compromise allocation under distance based method \( u_D^* = (n_1^*, n_2^*, n_3^*, n_4^*) \) is obtained as \( n_1^* = 2, n_2^* = 9, n_3^* = 9, n_4^* = 2 \). Where the suffix ‘D’ stands for ‘Distance Based’. The corresponding value of the objective function is 0.00008067544.
3.5.4 Khuri and Cornell's Approach

With the distance proposed by Khuri and Cornell (1986), the problem (3.16) can be given as:

\[
\text{Minimize } \left[ \left( \frac{1}{225600205681.21} \right) \left( 1 - \frac{n_1}{8} \right) \frac{190177129.163709}{n_1} + \left( 1 - \frac{n_2}{34} \right) \frac{3075358778.397211}{n_2} + \left( 1 - \frac{n_3}{45} \right) \frac{8750902957.838071}{n_3} + \left( 1 - \frac{n_4}{12} \right) \frac{451381009.970472}{n_4} - 0.0050068222)^2 / (0.0050068222)^2 \right] \left( \frac{1}{2484564.0625} \right) \left( 1 - \frac{n_1}{8} \right) \frac{5073.889916}{n_1} + \left( 1 - \frac{n_2}{34} \right) \frac{588180.653800}{n_2} + \left( 1 - \frac{n_3}{45} \right) \frac{221961.923300}{n_3} + \left( 1 - \frac{n_4}{12} \right) \frac{5705.062673}{n_4} - 0.0200458)^2 / (0.0200458)^2 \right] \]

Subject to \( 15n_1 + 7n_2 + 5n_3 + 9n_4 + 10\sqrt{n_1} + 5\sqrt{n_2} + 2\sqrt{n_3} + 6\sqrt{n_4} \leq 200 \)
\( 2 \leq n_1 \leq 8 \)
\( 2 \leq n_2 \leq 34 \)
\( 2 \leq n_3 \leq 45 \)
\( 2 \leq n_4 \leq 12 \)

and \( n_h \) integers; \( h = 1, 2, ..., 4 \)

(3.22)

The compromise allocation under distance based method \( n^*_KC = \left( n^*_1, n^*_2, n^*_3, n^*_4 \right) \) is obtained as \( n^*_1 = 2, n^*_2 = 9, n^*_3 = 9, n^*_4 = 2 \). Where
the suffix ‘KC’ stands for ‘Khuri and Cornell’. The corresponding value of
the objective function is 0.2159159.

3.6 DISCUSSION

Table 3.2 shows the summary of the results obtained by the four
different methods of allocations discussed in this chapter.

<table>
<thead>
<tr>
<th>S.No</th>
<th>Approach</th>
<th>$n_1$</th>
<th>$n_2$</th>
<th>$n_3$</th>
<th>$n_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Value function</td>
<td>2</td>
<td>9</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>$\epsilon$-Constraint</td>
<td>2</td>
<td>7</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Distance based</td>
<td>2</td>
<td>9</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Khuri &amp; Cornell</td>
<td>2</td>
<td>9</td>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>

While dealing with the multiobjective optimization problem we may
have complete, partial or null information about the population. In the
problem of optimum allocation in multivariate stratified sampling, it is first
necessary to establish in which of the three scenarios the problem fits in. The
technique to be applied is decided on the basis of the available information.
The solution for a particular allocation problem should be achieved by the
implementation of a single method. Therefore, the results obtained by
applying different approaches can not be compared, see Díaz-García and
It can be seen that allocations in the second and third strata are greater than the allocations in first and fourth strata. It is because $W_h S_{jh}$ for $h=2$ and $3$ is greater than $W_h S_{jh}$ for $h=1$ and $4$ for $j=1$ and $j=2$ both, (See Cochran (1977) p.89).

It can also be seen that for the given example the first, third and fourth approaches give identical results. The second approach gives slightly different result for second and third strata.