International Journal of Mathematical Archive-6(7), 2015, 183-193
Available online through www.ijma.info ISSN 2229 – 5046

STOCHASTIC BIVARIATE STRATIFIED SAMPLE SURVEY PROBLEM IN PRESENCE OF PARTIAL RESPONSE

NEHA GUPTA*, IRFAN ALI

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

(Received On: 05-06-15; Revised & Accepted On: 22-07-15)

ABSTRACT

This article considers the problem of bivariate stratified sampling in presence of partial response. Each stratum of population is divided into complete non respondents, partial respondents and complete respondents. Here it is assumed that respondents of questions of category II is also the respondents to questions of category I, but the converse is not necessarily true. In sample survey problems uncertainty is inherent because only a sample is studied in place of whole population, therefore, the problem is formulated as a stochastic biobjective nonlinear programming problem where sampling variances are considered to be random. The formulated nondeterministic problem is converted into equivalent deterministic form by adopting the Modified E-model technique. To derive the compromise allocation four different techniques viz Distance based lexicographic programming, ε-constraint, Euclidean distance and Khuri and Cornell are adopted. A simulation study has been done to adjudge their significance from the point of precision. Data has been generated through R software and problems are solved by an optimization software LINGO.

Keywords: *Compromise allocation, Stochastic programming, Non response, Stratified sampling.*

1. INTRODUCTION

In sample surveys non response is the failure to obtain a valid response from a sampled unit. Non response has the potential to introduce bias into a survey estimates and reduce the precision of survey estimates. All types of non response result in missing data. Non response can be classified as follows: when a sampled unit fails to respond at all to the data collection efforts, such type of missing data is called unit non response (the failure of a sample unit to respond to the survey). Partial non response is another form of missing data that occurs when a unit responds to some of the data items in the survey but fails to answer one or more items or when only a portion of the survey is completed. Non response may occur for various reasons, but most non response may be classified into two broad categories (i) accessibility issues, refers to the ability to make contact with the sampled unit (ii) amenability issues, refers to the units willingness to cooperate with the survey request after contact has been made. A third and generally less significant cause of non response is loss due to administrative issues, such as mail questionnaires that are received too late or are lost in processing.

Firstly, the problem of non response has been discussed by Hansen and Hurwitz (1946) and in 1956 El-Badry extends this technique. After that several authors discuss the problem of complete non response in univariate as well as in multivariate case such as Khare (1987), Fabian and Hyunshik (2000), Najmussehar and Bari (2002) etc. Recently, problem of complete non response formulated as mathematical programming problem by some authors such as Khan et al. (2008), Varshney *et al.* (2011), Raghav *et al.* (2012), Gupta *et al.* (2012) etc. The second type of non response i.e. partial non response was first discussed by Tripathi and Khare (1997). They estimate the population mean in presence of partial response and after that Maqbool and Pirzada (2005) discuss it in two variate stratified sample surveys and find out the optimum sample size and sub-sampling fraction for a fixed budget.

Stochastic programming, as the name implies, is mathematical (i.e. linear, integer, mixed-integer, nonlinear) programming but with a stochastic element present in the data. Stochastic programming therefore deals with situations where some of the data incorporated into the objective or constraints is uncertain. Uncertainty is usually characterized by a probability distribution on the parameters. Several authors has been discussed the stochastic multivariate sample allocation problem in stratified sample surveys. Among them are Diaz-Garcia and Cortez (2006, 2008), Diaz-Garcia and Garay Tapia (2007), Khan *et al.* (2012), Ali *et al.* (2013), Gupta *et al.* (2013) and many others.

Corresponding Author: Neha Gupta*,
Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh, India.

In this article, we consider the problem of bivariate stratified sample survey in presence of partial response and formulate it as an Stochastic Biobjective Nonlinear Programming Problem (SBONLPP) in which the sampling variances are considered as random. Rest of the paper is organized as follows: Section 3 presents Modified E-model technique to convert the nondeterministic problem into equivalent deterministic form. In section 4, four different methods viz Distance based lexicographic programming, silon-constraint, Euclidean distance and Khuri and Cornell has been described to derive the compromise allocation of the BONLPP. In section 5 simulation study has been carried out for comparing the efficiency of the suggested methods. Finally section 6 gives the conclusion of the work.

2. PROBLEM SAMPLE SURVEYS IN PRESENCE OF PARTIAL RESPONSE

The sampling scheme used in formulation is as in Maqbool and Pirzada (2005). However, for the sake of continuity they are reproduced here.

Let $X_{hj1}, X_{hj2}, \ldots, X_{hjN_h},\ j = 1,2,\ldots,p;\ h = 1,2,\ldots,L$ be the measurement of N_h units who respond to j^{th} character in h^{th} stratum. Questionnaire is assumed to have the questions of two categories. Character I are measured by questions of category I and character II by those of category II.

First of all in phase one select a random sample from each stratum and send a mail questionnaire to all of the selected units in each stratum. After that identify the partial respondents (those who reply the questions of category I only) and the complete respondents (those who reply the questions of both the categories) in each stratum. Now by personnel interview or through some additional efforts collect data from the selected non-respondents and the partial respondents from each stratum in the sub sample. To make sure that a respondent to questions of category II always responds to questions of category I, it is assumed that the questions of category I are simple. Therefore the whole population is divided into three groups' viz. non response, partial response and complete response. In second attempt it is assumed that through extra efforts information from non respondents and partial respondents in each stratum are collected and each unit of the sub sample yields information on both the categories.

The suffix '*h*' denotes the stratum number; subscripts designate the attempts 1 and 2 while superscripts designate characters. The superscripts with bar will stand for the character under study corresponding to non respondents.

Notations:

n_h = sample size drawn using SRSWOR

$\overline{n}_{h1}^{(1)}$ = number of respondents to questions of category I only at first phase,

$n_{h1}^{(1,2)}$ = number of complete respondents to questions of categories I and II both at first phase,

$\overline{n}_{h1}^{(1,2)}$ = number of complete non-respondents at first phase.

In second phase by personnel interview or through other extensive methods information is collected from the complete non-respondents and partial respondents for questions of both the categories.

$n_{h2}^{(1,2)}$ = sub-sample of $\overline{n}_{h1}^{(1,2)}$, all of which respond to questions of both the categories at second attempt.

$n_{h2}^{(2)}$ = sub-sample out of $\overline{n}_{h1}^{(1)}$, all of which respond to questions of category II at second attempt.

Here proportion of unit's viz. k_h selected for second attempt out of the partial respondents and the total non respondents are assumed to be same.

Further let

$$\overline{x}^{(1)} = \sum_{h=1}^{L} \frac{P_h}{n_h}\left[n_{h1}^{*}\overline{x}_{h1} + \overline{n}_{h1}^{(1,2)}\overline{x}_{h1}^{(1,2)*}\right] \text{(sample mean of character I)}$$

$$\overline{x}^{(2)} = \sum_{h=1}^{L} \frac{P_h}{n_h}\left[n_{h1}^{(1,2)}\overline{x}_{h2} + n_{h2}^{*}\overline{x}_{h2}^{(2)*}\right] \text{(sample mean of character II)}$$

where

$n_{h1}^{*} = n_{h1}^{(1,2)} + \overline{n}_{h1}^{(1)}$ number of units who respond to questions of category I at phase I

$n_{h2}^{*} = \overline{n}_{h1}^{(1,2)} + \overline{n}_{h1}^{(1)}$ number of respondents to questions of category II and non respondents only at phase I.

$\overline{x}_{h1}$ = mean of respondents to questions of category I for character I based on $\overline{n}_{h1}^{(1)} + n_{h1}^{(1,2)}$ units at first attempt.

$\bar{x}_{h1}^{(1,2)*}$ = sub-sample mean of respondents to questions of category I at second attempt based on $n_{h2}^{(1,2)}$ units taken out of $\bar{n}_{h1}^{(1,2)}$ non-respondents.

$\bar{x}_{h2}$ = mean of respondents to questions of category II (character II) based on $n_{h1}^{(1,2)}$ at first attempt.

$\bar{x}_{h2}^{(2)*}$ = sub-sample mean of respondents to questions of category II at second attempt based on $n_{h2}^{(1,2)}$ units.

With the variances of the two estimators $\bar{y}^{(1)}$ and $\bar{y}^{(2)}$ corresponding to the character I and II

$$V(\bar{x}^{(1)}) = \sum_{h=1}^{L}\left[\left(\frac{N_h - n_h}{N_h n_h}\right) + \left(\frac{k_h - 1}{n_h}\right) w_{h3}\right] P_h^2 s_{h1}^2 \tag{1}$$

$$V(\bar{x}^{(2)}) = \sum_{h=1}^{L}\left[\left(\frac{N_h - n_h}{N_h n_h}\right) + \left(\frac{k_h - 1}{n_h}\right) w_{h4}\right] P_h^2 s_{h2}^2 \tag{2}$$

where $k_h = \dfrac{\bar{n}_{h1}^{(1,2)}}{n_{h2}^{(1,2)}}$, $n_{h2}^{(1,2)} = \dfrac{\bar{n}_{h1}^{(1,2)}}{k_h}$, $P_h = N_h / N$ and s_{h1}^2, s_{h2}^2 are the variances of the non-response classes for the characters I and II respectively.

After ignoring the terms independent of n_h in variances of two estimators, expressions (1) and (2) can be written as:

$$V(\bar{x}^{(1)}) = \sum_{h=1}^{L}\left[\left(\frac{1}{n_h}\right) + \left(\frac{k_h - 1}{n_h}\right) w_{h3}\right] P_h^2 s_{h1}^2 \tag{3}$$

$$V(\bar{x}^{(2)}) = \sum_{h=1}^{L}\left[\left(\frac{1}{n_h}\right) + \left(\frac{k_h - 1}{n_h}\right) w_{h4}\right] P_h^2 s_{h2}^2 \tag{4}$$

The cost function is defined as:

$$C = c_0 + \sum_{h=1}^{L} c_h n_n + \sum_{h=1}^{L} c_{h1}^{(1)} + \sum_{h=1}^{L} c_{h1}^{(1)}\left(\bar{n}_{h1}^{(1)} + n_{h1}^{(1,2)}\right) + \sum_{h=1}^{L} c_{h1}^{(2)} n_{h1}^{(1,2)} + \sum_{h=1}^{L} c_{h2}\left(n_{h2}^{(2)} + n_{h2}^{(1,2)}\right) \tag{5}$$

$$C - c_0 = C_0 = \sum_{h=1}^{L} c_h n_n + \sum_{h=1}^{L} c_{h1}^{(1)} + \sum_{h=1}^{L} c_{h1}^{(1)}\left(\bar{n}_{h1}^{(1)} + n_{h1}^{(1,2)}\right) + \sum_{h=1}^{L} c_{h1}^{(2)} n_{h1}^{(1,2)} + \sum_{h=1}^{L} c_{h2}\left(n_{h2}^{(2)} + n_{h2}^{(1,2)}\right) \tag{6}$$

where c_0 = overhead cost

c_h = cost of including a unit in the sample in h^{th} stratum.

$c_{h1}^{(1)}$ = cost incurred/unit in enumerating questions of category I in h^{th} stratum in first attempt.

$c_{h1}^{(2)}$ = cost incurred/unit in enumerating questions of category II in h^{th} stratum in first attempt.

c_{h2} = cost incurred/unit in h^{th} stratum in enumerating both the characters in second attempt.

It is understood that the values of $\bar{n}_{h1}^{(1)}$ and $n_{h1}^{(1,2)}$ are not known until the first attempt is made, therefore the expected cost is used in planning the sample as

$n_{h1}^{*} = n_h w_{h1}$, $n_{h1}^{(1,2)} = \dfrac{n_h w_{h3}}{k_h}$, and $n_{h2}^{(2)} = \dfrac{n_h w_{h4}}{k_h}$ and hence the total expected cost is given by

$$C_0 = \sum_{h=1}^{L} c_h n_n + \sum_{h=1}^{L} c_{h1}^{(1)} n_h w_{h1} + \sum_{h=1}^{L} c_{h1}^{(2)} n_h w_{h2} + \sum_{h=1}^{L} c_{h2} n_h\left(\frac{w_{h3}}{k_h} + \frac{w_{h4}}{k_h}\right) \tag{7}$$

where W_{hj} are the proportion of respondents and non respondents in h^{th} stratum to questions of both the categories such that

$$w_{h1} + w_{h3} = 1$$
$$w_{h2} + w_{h4} = 1$$

2.1. Stochastic Biobjective Nonlinear Programming Problem

Usually, a sample surveys problem is subject to certain uncertainty because only a sample has been studied instead of the whole population. There are several ways to deal with such uncertainty in data. Therefore, in this section a SBONLPP is formulated in which s_{h1}^2 and s_{h2}^2 are considered as random. The problem can be formulated as follows:

$$\left.\begin{array}{l} \text{Minimize } V(\overline{x}^{(1)}) = P\left(\sum_{h=1}^{L}\left[\left(\frac{1}{n_h}\right)+\left(\frac{k_h-1}{n_h}\right)w_{h3}\right]P_h^2 s_{h1}^2\right) \\ \text{Minimize } V(\overline{x}^{(2)}) = P\left(\sum_{h=1}^{L}\left[\left(\frac{1}{n_h}\right)+\left(\frac{k_h-1}{n_h}\right)w_{h4}\right]P_h^2 s_{h2}^2\right) \\ \text{Subject to } \sum_{h=1}^{L} c_h n_n + \sum_{h=1}^{L} c_{h1}^{(1)} n_h w_{h1} + \sum_{h=1}^{L} c_{h1}^{(2)} n_h w_{h2} + \sum_{h=1}^{L} c_{h2} n_h\left(\frac{w_{h3}}{k_h}+\frac{w_{h4}}{k_h}\right) \le C_0 \\ \qquad 2 \le n_h \le N_h;\ k_h \ge 0 \end{array}\right\} \tag{8}$$

To deal with the randomness in objective functions technique discussed by Diaz-Garcia and Tapia (2007) has been used and for the sake of continuity reproduced according to our sample survey problem in the next section.

3. Conversion of SBONLPP to equivalent DBONLPP

In this section probabilistic sampling variances converted into equivalent deterministic form by Modified E model technique as follows:

3.1. Modified E-model

Consider the stochastic programming problem in (9). By using the limiting distribution of the sample variances (see Melaku, 1986), consider the random variable ξ_h defined as

$$\xi_h = \frac{1}{n_h-1}\sum_{i=1}^{n_h}\left(x_{jhi} - \overline{X}^{(1)}\right)^2$$

where $\overline{X}^{(1)} = \frac{1}{N_h}\sum_{i=1}^{N_h} x_{jhi}$

which has an asymptotical normal distribution with mean $E(\xi_h)$ and variance $V(\xi_h)$, given by

$$E(\xi_h) = \frac{n_h}{n_h-1} S_{h1}^2$$

$$V(\xi_h) = \frac{n_h}{(n_h-1)^2}\left[C_{h1}^4 - (S_{h1}^2)^2\right]$$

where $C_{h1}^4 = \frac{1}{N_h}\sum_{i=1}^{N_h}\left(x_{jhi} - \overline{X}^{(1)}\right)^4$ *and* $S_{h1}^2 = \sum_{i=1}^{N_h}\left(x_{jhi} - \overline{X}^{(1)}\right)^2$

Now we see that

$$s_{h1}^2 = \xi_h - \frac{n_h}{n_h-1}\left(\overline{x}_{h1} - \overline{X}^{(1)}\right)^2,$$

where $\overline{x}_{h1} = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{jhi}$

$\frac{n_h}{n_h-1} \to 1$ and $\left(\overline{x}_{h1} - \overline{X}^{(1)}\right)^2 \to 0$ in probability.

Therefore, the sample variance s_{h1}^2 follows an asymptotical normal distribution, i.e.

$$s_{h1}^2 \sim N\left(E(\xi_h), Var(\xi_h)\right)$$

From Diaz-Garcia and Tapia (2007) the objective function is given by

$$f(n) = k_1 E(V(\overline{x}^{(1)})) + k_2 \sqrt{Var(V(\overline{x}^{(1)}))} \tag{9}$$

Where

$$\begin{aligned} E(V(\overline{x}^{(1)})) &= E\left(\sum_{h=1}^{L} \left[\left(\frac{1}{n_h}\right) + \left(\frac{k_h - 1}{n_h}\right) w_{h3} \right] P_h^2 s_{h1}^2 \right) \\ &= \sum_{h=1}^{L} \left[\left(\frac{1}{n_h}\right) + \left(\frac{k_h - 1}{n_h}\right) w_{h3} \right] P_h^2 E(\xi_{h1}) \\ &= \sum_{h=1}^{L} \left[\left(\frac{1}{(n_h - 1)}\right) + \left(\frac{k_h - 1}{(n_h - 1)}\right) w_{h3} \right] P_h^2 S_{h1}^2 \end{aligned} \tag{10}$$

and

$$\begin{aligned} Var(V(\overline{x}^{(1)})) &= Var\left[\sum_{h=1}^{L} \left(\frac{1}{n_h}\right) P_h^2 s_{h1}^2 + \sum_{h=1}^{L} \left(\frac{k_h - 1}{n_h}\right) w_{h3} P_h^2 s_{h1}^2 \right] \\ &= \sum_{h=1}^{L} \left(\frac{1}{n_h}\right)^2 P_h^4 Var(\xi_{h1}) + \sum_{h=1}^{L} \left(\frac{k_h - 1}{n_h}\right)^2 w_{h3}^2 P_h^4 Var(\xi_{h1}) \\ &= \sum_{h=1}^{L} \frac{P_h^4 (C_{h1}^4 - (S_{h1}^2)^2)}{n_h (n_h - 1)^2} + \sum_{h=1}^{L} \frac{(k_h - 1)^2 w_{h3}^2 P_h^4 (C_{h1}^4 - (S_{h1}^2)^2)}{n_h (n_h - 1)^2} \\ &= \sum_{h=1}^{L} \frac{P_h^4 (C_{h1}^4 - (S_{h1}^2)^2)}{n_h (n_h - 1)^2} \left(1 + (k_h - 1)^2 w_{h3}^2\right) \end{aligned} \tag{11}$$

Now from eq. (10) the deterministic objective function can be written as:

$$f(n) = \beta_1 \sum_{h=1}^{L} \left[\left(\frac{1}{(n_h - 1)}\right) + \left(\frac{k_h - 1}{(n_h - 1)}\right) w_{h3} \right] P_h^2 S_{h1}^2 + \beta_2 \sqrt{\sum_{h=1}^{L} \frac{P_h^4 (C_{h1}^4 - (S_{h1}^2)^2)}{n_h (n_h - 1)^2} \left(1 + (k_h - 1)^2 w_{h3}^2\right)} \tag{12}$$

But the objective function is given in terms of the population variances s_{h1}^2, which are unknown (by hypothesis), therefore we will use the sample variances s_{h1}^2. Thus, the equivalent deterministic objective function will be:

$$f(n) = \beta_1 \sum_{h=1}^{L} \left[\left(\frac{1}{(n_h - 1)}\right) + \left(\frac{k_h - 1}{(n_h - 1)}\right) w_{h3} \right] P_h^2 s_{h1}^2 + \beta_2 \sqrt{\sum_{h=1}^{L} \frac{P_h^4 (C_{h1}^4 - (s_{h1}^2)^2)}{n_h (n_h - 1)^2} \left(1 + (k_h - 1)^2 w_{h3}^2\right)}$$

Similarly we can write the deterministic form of the second objective function as

$$f(n) = \beta_1 \sum_{h=1}^{L} \left[\left(\frac{1}{(n_h - 1)}\right) + \left(\frac{k_h - 1}{(n_h - 1)}\right) w_{h4} \right] P_h^2 s_{h2}^2 + \beta_2 \sqrt{\sum_{h=1}^{L} \frac{P_h^4 (C_{h1}^4 - (s_{h2}^2)^2)}{n_h (n_h - 1)^2} \left(1 + (k_h - 1)^2 w_{h4}^2\right)}$$

Here β_1 and β_2 are non-negative constants. Now the transformed BONLPP (8) can be presented as:

$$\left.\begin{aligned} &\text{Minimize } \beta_1 \sum_{h=1}^{L} \left[\left(\frac{1}{(n_h - 1)}\right) + \left(\frac{k_h - 1}{(n_h - 1)}\right) w_{h3} \right] P_h^2 s_{h1}^2 + \beta_2 \sqrt{\sum_{h=1}^{L} \frac{P_h^4 (C_{h1}^4 - (s_{h1}^2)^2)}{n_h (n_h - 1)^2} \left(1 + (k_h - 1)^2 w_{h3}^2\right)} \\ &\text{Minimize } \beta_1 \sum_{h=1}^{L} \left[\left(\frac{1}{(n_h - 1)}\right) + \left(\frac{k_h - 1}{(n_h - 1)}\right) w_{h4} \right] P_h^2 s_{h2}^2 + \beta_2 \sqrt{\sum_{h=1}^{L} \frac{P_h^4 (C_{h1}^4 - (s_{h2}^2)^2)}{n_h (n_h - 1)^2} \left(1 + (k_h - 1)^2 w_{h4}^2\right)} \\ &\text{Subject to } \sum_{h=1}^{L} c_h n_n + \sum_{h=1}^{L} c_{h1}^{(1)} n_h w_{h1} + \sum_{h=1}^{L} c_{h1}^{(2)} n_h w_{h2} + \sum_{h=1}^{L} c_{h2} n_h \left(\frac{w_{h3}}{k_h} + \frac{w_{h4}}{k_h} \right) \le C_0 \\ &\qquad 2 \le n_h \le N_h;\ k_h \ge 0 \end{aligned}\right\} \tag{13}$$

4. OPTIMIZATION TECHNIQUES

This section provides a detail description of various techniques used for obtaining a compromise allocation of DBONLPP (13).

4.1. Distance based Lexicographic Goal Programming

This is a modified form of lexicographic goal programming, in this method a set of solutions is obtained by giving priorities to the objectives one after the other. Out of these solutions, an ideal solution is identified. A general procedure with two objectives is as follows:

If priority is given to the variance of first characteristic, then we have to solve the following lexicographic goal programming problem

$$\left.\begin{aligned}
&\text{Minimize } \sum_{j=1}^{2}\delta_{\mathrm{j}}\\
&\text{Subject to } V(\overline{x}^{(1)})-\delta_1\le V^*(\overline{x}^{(1)})\\
&\qquad V(\overline{x}^{(2)})-\delta_2\le V^*(\overline{x}^{(2)})\\
&\qquad \sum_{h=1}^{L}c_h n_n+\sum_{h=1}^{L}c_{h1}^{(1)}n_h w_{h1}+\sum_{h=1}^{L}c_{h1}^{(2)}n_h w_{h2}+\sum_{h=1}^{L}c_{h2}n_h\left(\frac{w_{h3}}{k_h}+\frac{w_{h4}}{k_h}\right)\le C_0\\
&\qquad 2\le n_h\le N_h;\ k_h\ge 0
\end{aligned}\right\} \tag{14}$$

Similarly if priority is given to variance of second characteristic, then we have to solve the following lexicographic goal programming problem

$$\left.\begin{aligned}
&\text{Minimize } \sum_{j=1}^{2}\delta_{\mathrm{j}}\\
&\text{Subject to } V(\overline{x}^{(2)})-\delta_2\le V^*(\overline{x}^{(2)})\\
&\qquad V(\overline{x}^{(1)})-\delta_1\le V^*(\overline{x}^{(1)})\\
&\qquad \sum_{h=1}^{L}c_h n_n+\sum_{h=1}^{L}c_{h1}^{(1)}n_h w_{h1}+\sum_{h=1}^{L}c_{h1}^{(2)}n_h w_{h2}+\sum_{h=1}^{L}c_{h2}n_h\left(\frac{w_{h3}}{k_h}+\frac{w_{h4}}{k_h}\right)\le C_0\\
&\qquad 2\le n_h\le N_h;\ k_h\ge 0
\end{aligned}\right\} \tag{15}$$

We will obtain P! (Factorial) different solutions by solving the P! problems arising for P! different priority structures as follows:

Table-1: Calculations for ideal solutions

Priority Structure	n_1	n_2	$\cdots$	n_h	$\cdots$	n_L	k_1	k_2	$\cdots$	k_h	$\cdots$	k_L
$V^{(1)}$	$n_1^{(1)}$	$n_2^{(1)}$	$\cdots$	$n_h^{(1)}$	$\cdots$	$n_L^{(1)}$	$k_1^{(1)}$	$k_2^{(1)}$	$\cdots$	$k_h^{(1)}$	$\cdots$	$k_L^{(1)}$
$V^{(2)}$	$n_1^{(2)}$	$n_2^{(2)}$	$\cdots$	$n_h^{(2)}$	$\cdots$	$n_L^{(2)}$	$k_1^{(2)}$	$k_2^{(2)}$	$\cdots$	$k_h^{(2)}$	$\cdots$	$k_L^{(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$V^{(r)}$	$n_1^{(r)}$	$n_2^{(r)}$	$\cdots$	$n_h^{(r)}$	$\cdots$	$n_L^{(r)}$	$k_1^{(r)}$	$k_2^{(r)}$	$\cdots$	$k_h^{(r)}$	$\cdots$	$k_L^{(r)}$
Ideal Solution	n_1^*	n_2^*	$\cdots$	n_h^*	$\cdots$	n_L^*	k_1^*	k_2^*	$\cdots$	k_h^*	$\cdots$	k_L^*

From the calculations done in Table 1 ideal solution is given as: Ideal so solution

$$=\left\{\begin{array}{l}\max(n_1^{(1)},\cdots,n_1^{(r)}),\max(n_2^{(1)},\cdots,\ n_2^{(r)}),...,\max(n_L^{(1)},\cdots,\ n_L^{(r)}),\min(k_1^{(1)},\cdots,k_1^{(r)}),\min(k_2^{(1)},\cdots,\ k_2^{(r)})\\ ,...,\min(k_L^{(1)},\cdots,\ k_L^{(r)})\end{array}\right\}$$

$$=\left\{n_1^*,n_2^*,...,n_L^*,k_1^*,k_2^*,...,k_L^*\right\}$$

Let $(n_h^{(r)}, k_h^{(r)}) = \{n_1^{(r)}, n_2^{(r)}, ..., n_L^{(r)}, k_1^{(r)}, k_2^{(r)}, ..., k_L^{(r)}\}, 1 \le r \le P!$ be the P! number of solutions obtained by giving priorities to P objective functions. Let $(n_1^*, n_2^*, ..., n_L^*, k_1^*, k_2^*, ..., k_L^*)$ be the ideal solution. But in practice ideal solution can never be achieved because it may violate the constraint. Therefore, the solution, which is closest to the ideal solution, is acceptable as the best compromise solution, and the corresponding priority structure is identified as most appropriate priority structure in the planning context. Distances of different solutions from the ideal solution defined in (16) below are then calculated. The solution corresponding to the minimum distance gives the best compromise solution.
Now,

$$(D_1)^r = \sum_{h=1}^{L} \left| n_h^* - n_h^{(r)} + k_h^* - k_h^{(r)} \right| \tag{16}$$

is defined as the D_1-distance from the ideal solution $(n_1^*, n_2^*, ..., n_L^*, k_1^*, k_2^*, ..., k_L^*)$, to the r^{th} solution (shown in Table 2)

$$\{n_1^{(r)}, n_2^{(r)}, ..., n_L^{(r)}, k_1^{(r)}, k_2^{(r)}, ..., k_L^{(r)}\}, 1 \le r \le P!$$

Table-2: D_1-Distances from the ideal solution as

Priority Structure	n_1	$\dots$	n_L	k_1	$\dots$	n_h	$(D_1)^r$
$V^{(1)}$	$\lvert n_1^* - n_1^{(1)} \rvert$	$\dots$	$\lvert n_L^* - n_L^{(1)} \rvert$	$\lvert k_1^* - k_1^{(1)} \rvert$	$\dots$	$\lvert n_h^* - n_h^{(1)} \rvert$	$\sum_{h=1}^{L} \lvert n_h^* - n_h^{(1)} + k_h^* - k_h^{(1)} \rvert$
$V^{(2)}$	$\lvert n_1^* - n_1^{(2)} \rvert$	$\dots$	$\lvert n_L^* - n_L^{(2)} \rvert$	$\lvert k_1^* - k_1^{(2)} \rvert$	$\dots$	$\lvert n_2^* - n_2^{(2)} \rvert$	$\sum_{h=1}^{L} \lvert n_h^* - n_h^{(2)} + k_h^* - k_h^{(2)} \rvert$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$V^{(r)}$	$\lvert n_1^* - n_1^{(r)} \rvert$	$\dots$	$\lvert n_L^* - n_L^{(r)} \rvert$	$\lvert n_1^* - n_1^{(r)} \rvert$	$\dots$	$\lvert n_2^* - n_2^{(r)} \rvert$	$\sum_{h=1}^{L} \lvert n_h^* - n_h^{(r)} + k_h^* - k_h^{(r)} \rvert$

Therefore, the optimal distance from the ideal solution is given as

$$(D_1)_{opt} = \min_{1 \le r \le P!} (D_1)^r = \min_{1 \le r \le P!} \sum_{i=1}^{m} \left| n_h^* - n_h^{(r)} \right| + \sum_{i=1}^{m} \left| k_h^* - k_h^{(r)} \right|$$

Let the minimum be attained for *r* = *t*.

Then $(n_1^{(t)}, n_2^{(t)}, ..., n_L^{(t)}, k_1^{(t)}, k_2^{(t)}, ..., k_L^{(t)})$ is the best compromise solution of the problem.

4.2. ε-Constraint Method

In the ε-constraint method, investigator needs to identify the most important characteristic and the objective functions corresponding to that characteristic is selected to be optimized and the other entire objective functions are converted into constraints by setting an upper bound to each of them (see Miettinen, 1999).

Let the l^{th} characteristic $l \in \{1, 2, \dots, p\}$, be the most important. Then the problem to be solved is of the form

$$\left.\begin{array}{l} \text{Minimize } \; V(\overline{x}^{(l)}) \\ \text{Subject to } \; V(\overline{x}^{(j)}) \le \varepsilon_j \quad \forall \; j = 1, 2, \dots, p, \; j \ne l \\ \qquad \sum_{h-1}^{L} c_h n_h + \sum_{h=1}^{L} c_{h1}^{(1)} n_h w_{h1} + \sum_{h=1}^{L} c_{h1}^{(2)} n_h w_{h2} + \sum_{h=1}^{L} c_{h2} n_h \left(\frac{w_{h3}}{k_h} + \frac{w_{h4}}{k_h} \right) \le C_0 \\ \qquad 2 \le n_h \le N_h; \quad k_h \ge 0 \end{array}\right\} \tag{18}$$

where $l \in \{1, 2, \dots, p\}$. Problem (18) is called an ε-constraint problem.

4.3. Euclidean Distance Method

In the situations where it is difficult to decide which is the most important characteristic of the survey distance based method can be used. It requires only a vector of ideal goals that can be determined with the null information expressed in the problem (see Rios et al., 1989 and Steuer, 1986).

Using this method compromise allocation is obtained by minimizing the distances between the vector of individual optimum variances and their vector of targets.

Consider the vector of target α as

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}$$

where $\alpha_j;\ j = 1,2,\ldots,p$ be the individual optimum value for the j^{th} objective which is computed by minimizing $V(\overline{x}^{(j)});\ j = 1,2,\ldots,p$ separately for each characteristic.

After computing α, under Euclidean distance method the optimization problem can be formulated as:

$$\left.\begin{array}{l} \text{Minimize} \quad \sum_{j=1}^{p}\left[V(\overline{x}^{(j)}) - \alpha_j\right]^2 \\ \text{Subject to} \ \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} c_{h1}^{(1)} n_h w_{h1} + \sum_{h=1}^{L} c_{h1}^{(2)} n_h w_{h2} + \sum_{h=1}^{L} c_{h2} n_h \left(\frac{w_{h3}}{k_h} + \frac{w_{h4}}{k_h}\right) \le C_0 \\ \qquad 2 \le n_h \le N_h;\quad k_h \ge 0 \end{array}\right\} \tag{19}$$

where the objective function represents the squared distance between the vectors

$$\underline{V} = \begin{pmatrix} V(\overline{x}^{(1)}) \\ \vdots \\ V(\overline{x}^{(j)}) \end{pmatrix} \text{ and } \alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}$$

belongs to n-dimensional Euclidean space.

4.4. Khuri and Cornell Method

In 1986, Khuri and Cornell proposed an alternative distance model. Formulated problem according to his method is given as:

$$\left.\begin{array}{l} \text{Minimize} \quad \sum_{j=1}^{p}\frac{\left[V(\overline{x}^{(j)}) - \alpha_j\right]^2}{\alpha_j^2} \\ \text{Subject to} \ \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} c_{h1}^{(1)} n_h w_{h1} + \sum_{h=1}^{L} c_{h1}^{(2)} n_h w_{h2} + \sum_{h=1}^{L} c_{h2} n_h \left(\frac{w_{h3}}{k_h} + \frac{w_{h4}}{k_h}\right) \le C_0 2 \le n_h \le N_h;\quad k_h \ge 0 \end{array}\right\} \tag{20}$$

5. SIMULATION STUDY AND DISCUSSION

For comparing the efficiency of the suggested techniques, a simulation study has been carried out. The R language (2004) has been used to perform the simulations and data analysis. We have generated two populations with sizes N=100 and 1050. From these populations four strata are randomly generated. The characteristics for the two populations have generated in the following way:

$$\overline{X}^{(1)} \sim N(500,150) \text{ and } \overline{X}^{(2)} \sim N(300,95)$$

Data obtained by simulation study for the two populations are divided into four strata and categorized into two characteristic shown in Table (3) and Table (4):

Table-3: Data for two characteristics and four strata

h	N_h	P_h	S_{h1}^2	S_{h2}^2	c_h	$c_{h1}^{(1)}$	$c_{h1}^{(2)}$	c_{h2}	C_{h1}^4	C_{h2}^4
1	26	0.26	5395.387	2485.661	0.5	8.5	8.7	25	89713790	23062924
2	35	0.35	2821.256	776.6979	0.7	7.4	7.6	20	11196593	1145433
3	16	0.16	109.6188	294.5205	0.4	7	7.2	18	21423.18	150959
4	23	0.23	5016.379	1028.69	0.6	9	9.2	25	54215891	2280157

In addition to the above, it is assumed that the relative value of the variances of the non-respondents and respondents, that is, $S_{jh2}^2 \big/ S_{jh}^2 = 0.25$ for $j = 1,2$ *and* $h = 1,2,\ldots,4$. Further, let the total amount available for the survey be C_0=1700 units for the problem (8). The proportions of respondents are $W_{h1} = 0.4$ and $W_{h2} = 0.3$ and the proportions of non-respondents are $W_{h3} = 0.6$ and $W_{h4} = 0.7$ for the character I and II respectively.

Table-4: Data for two characteristics and four strata

h	N_h	P_h	S_{h1}^2	S_{h2}^2	C_{h1}^4	C_{h2}^4
1	306	0.291429	5800.17	2385.23	144345206	24762269
2	205	0.195238	555.248	160.637	552045.7	48203.06
3	353	0.33619	1592.11	684.948	4712352	911132.7
4	186	0.177143	5464.03	1749.39	115612562	12573051

5.1. Compromise allocation for the small population

The compromise allocation $(n_h^*, k_h^*);\ h = 1,2,\ldots,L$ corresponding to different methods discuss in section 4 are summarized into the Table (5).

Table-5: Compromise allocations

Methods	Allocations							
	n_1	n_2	n_3	n_4	k_1	k_2	k_3	k_4
D_1-distance	26	29	5	23	1.9021	2.2553	2.3927	2.1725
ε-Constraint	26	25	8	19	1.6292	2.0332	2.1274	2.1993
Eucliedian distance	26	28	6	23	1.8561	2.2182	2.4117	2.2616
Khuri n Cornell	26	27	7	23	1.8078	2.1735	2.4965	2.3599

Using the allocations given in Table (5), values of variances $V(\overline{x}^{(j)});\ j = 1,2$ and trace values are calculated and summarized into the Table (6).

Table-6: Variances & Trace value

Methods	Variances		Trace
	$V(\overline{x}^{(1)})$	$V(\overline{x}^{(2)})$	$V = V(\overline{x}^{(1)}) + V(\overline{x}^{(2)})$
D_1-distance	9.239585	5.032199	14.27178
ε -Constraint	9.844362	4.694511	14.53887
Eucliedian distance	9.316274	4.871827	14.18810
Khuri n Cornell	9.413643	4.790833	14.20448

5.2. Compromise allocation for the large population

The compromise allocation $(n_h^*, k_h^*);\ h = 1,2,\ldots,L$ corresponding to different methods discuss in section 4 are summarized into the Table (7).

Table-7: Compromise allocations

Methods	Allocations							
	n_1	n_2	n_3	n_4	k_1	k_2	k_3	k_4
D_1-distance	41	8	23	23	2.6635	2.4695	2.3518	2.5879
ε-Constraint	40	7	23	21	2.4682	2.3293	2.1669	2.4784
Eucliedian distance	40	8	23	23	2.5588	2.5331	2.2957	2.6310
Khuri n Cornell	40	8	23	22	2.5369	2.5491	2.2662	2.5162

Using the allocations given in Table (7), values of variances $V(\overline{x}^{(j)});\ j=1,2$ and trace values are calculated and summarized into the Table (8).

Table-8: Variances & Trace value

Methods	Variances		Trace
	$V(\overline{x}^{(1)})$	$V(\overline{x}^{(2)})$	$V=V(\overline{x}^{(1)})+V(\overline{x}^{(2)})$
D_1-distance	11.49681	4.642227	16.13904
ε-Constraint	11.60296	4.615816	16.21878
Eucliedian distance	11.50543	4.633290	16.13872
Khuri n Cornell	11.51934	4.625682	16.14502

6. CONCLUSION

In this article, we consider a bivariate stratified population in presence of partial response. In real world problems, several uncertainties are present in data, so an SBONLPP has been formulated to deal with these uncertainties. In SBONLPP sampling variances are considered as random variables. SBONLPP has been converted into equivalent deterministic form by Modified E-model technique. Since the problem is bivariate it is not necessary the optimum allocation for one characteristic is also optimum for second characteristic, so we have to obtain a compromise allocation which is optimum for both the characteristics in some sense. Therefore, we obtain the compromise allocations by using four different methods such as Distance based lexicographic goal programming, **ε**-Constraint, Euclidean distance and Khuri and Cornell. To check the efficiency of the methods two different populations have been generated through simulation study by R-software (2004). All the formulated problems have been solved by an optimizing software LINGO (2013).

REFERENCES

1. Ali, I., Raghav, Y. S., and Bari, A. (2011), "Compromise allocation in multivariate stratified surveys with stochastic quadratic cost function", *Journal of Statistical computation and Simulation*, DOI:10.1080/00949655. 2011.643890.
2. Diaz-Garcia, J. A. and Cortez, L. U. (2006), "Optimum allocation in multivariate stratified sampling: multi-objective programming", *Communication Tecnica No.*, I-06-07/28-03-2006 (PE/CIMAT), Mexico
3. Diaz-Garcia, J. A. and Cortez, L. U. (2008), "Multi-objective optimization for optimum allocation in multivariate stratified sampling", *Surv. Methodol.*, 34(2), 215-222
4. Diaz-Garcia, J. A. and Garay-Tapia, M. M. (2007), "Optimum allocation in stratified surveys: Stochastic programming", *Computational Statistics & Data Analysis*, 51, 3016–3026.
5. El-Badry, M. A. (1956), "A sampling procedure for mailed questionnaire", *Jour. Amer. Stat. Assoc*, 51, 209-277.
6. Fabian C. Okafor and Hyonshil Lee (2000), "Double sampling for ratio and regression estimation with sub sampling the non respondents", *Survey Methodology*, 26(2), 183-188.
7. Gupta, N., Shafiullah, Iftekhar, S. and Bari, A. (2012), "Fuzzy Goal Programming Approach to solve Non-Linear Bi-Level Programming Problem in stratified double sampling design in presence of non-response", *International Journal of Scientific & Engineering research*, 3(10), ISSN 2229-5518.
8. Gupta, N., Ali, I., and Bari, A. (2013), "An Optimal Chance Constraint Multivariate Stratified Sampling Design Using Auxiliary Information", *J Math Model Algor*, DOI 10.1007/s10852-013-9237-5.
9. Gupta, N., Ali, I., and Bari, A. (2013), "A compromise solution for multi-objective chance constraint capacitated transportation problem", *ProbStat Forum*, 06, 60-67.
10. Hansen, M. H. and Hurwitz, W. N.(1946), "The problem of non response in sample surveys", *J. Amer. Statist. Assoc.*, 41, 517-529.
11. Khan, M. G. M., Khan, E. A., and Ahsan, M. J. (2008), "Optimum allocation in multivariate stratified sampling in presence of non-response", *J. Indian Soc. Agricultural Statist.*, 62(1), 42-48
12. Khan, M., Ali, I., Raghav, Y. S. and Bari A. (2012), "Allocation in Multivariate Stratified Surveys with Non-Linear Random Cost Function", *American Journal of Operations Research*, 2 (1), 100-105
13. Khare, B. B. (1987), "Allocation in stratified sampling in presence of non response", *Metron*, 45(1-2), 213-221.
14. Khuri, A. I., and Cornell, J. A. (1986), "Response Surface: Designs Analysis", *Marcel Dekker, Inc*., New York.
15. Kozak, M. (2006), "On sample allocation in multivariate surveys", *Comm. Statist. Simulation Comput.*, 35(4), 901-910.
16. LINGO-User's Guide (2013), "LINGO-User's Guide", Published by *LINDO SYSTEM INC.*, 1415, North Dayton Street, Chicago, Illinois, 60622 (USA).
17. Miettinen, K. M. (1999), "Non linear multi objective optimization", *Kluwer Academic Publishers*, Boston.
18. Maqbool, S. and Pirzada, S. (2005), "Allocation to response and non-response groups in two character stratified sampling", *Nihonkai Math. J.*, 16, 135-143 USA.

19. Melaku, A. (1986), "Asymptotic normality of the optimal allocation in multivariate stratified random sampling", *Indian J. Statist.*, 48 (Series B), 224–232.
20. Najmussehar, Bari, A. (2006), "Double sampling for stratification with sub-sampling the non-respondents: a dynamic programming approach", *Aligarh J. Statist.*, 22, 27-41.
21. R Development Core Team. (2004), "R: A language and environment for statistical computing", *Vienna, Austria: R foundation for statistical computing*, http://www.R-project.org.
22. Raghav, Y. S., Ali, I., and Bari, A. (2010), "Optimum Sample Sizes in Case of Stratified Sampling for Non-Respondents: An Integer Solution", *International Journal of Operations Research and Optimization*, 1(1), 149-160.
23. Ravindran, A., Phillips Don, T., Solberg James, J.(1987), "Operations Research Principles & Practice", second ed., *John Wiley*, New York
24. Rios, S., Rios Insua, S., and Rios Insua, M. J. (1989), "Procesos de decision multicriterio", *EUDEMA*, Madrid (in Spanish)
25. Steuer, R. E. (1986), "Multiple criteria optimization: Theory, computation and applications", New York, *John Wiley & Sons*, Inc
26. Varshney, R., Najmussehar, and Ahsan, M. J. (2011), "An optimum multivariate stratified double sampling design in non-response", *Springer*, *Optimization letters*, Verlag.

Source of support: Nil, Conflict of interest: None Declared