

Three - Stage Stochastic Multivariate Stratified Sample Survey

Sanam Haseen

Department of Statistics & Operations Research
Aligarh Muslim University, Aligarh, India

Shafiullah

Department of Statistics & Operations Research
Aligarh Muslim University, Aligarh, India

Irfan Ali

Department of Statistics & Operations Research
Aligarh Muslim University, Aligarh, India
irfii.ali@gmail.com

Abdul Bari

Department of Statistics & Operations Research
Aligarh Muslim University, Aligarh, India

Abstract

In this paper, we have considered the problem of three-stage sample surveys. The problem of a three stage multivariate stratified sample survey has been formulated as a non-linear stochastic programming problem by considering survey cost and the variances as random variables. The stochastic programming problem has been converted into equivalent deterministic form using Chance constraint programming and modified E-model.

Keywords: Three Stage Sample Survey, Stochastic Programming, Modified E-Model, Chance Constraint Programming, Multivariate Stratified Sample Survey.

Introduction

The analysis of two-stage stratified sampling designs is well defined in the sampling literature. In two-stage stratified sampling designs the total population is subdivided into a number of strata and then two-stage stratified sampling procedure is applied for taking the samples. The two- stage stratified sampling designs generally specifies two- stages of selection: primary sampling units (*PSUs*) at the first stage and sub samples from each *PSUs* at second stage as a secondary sampling units (*SSUs*) units. The methods to obtain the optimum allocations of sampling units to each stage are readily available. Showkat *et al.* (2011) has used the geometric programming approach in multivariate two-stage sampling design for obtaining optimum sample sizes of each stage.

In the three–stage stratified sampling design the process of sub sampling of the population under study can be carried out by dividing the given population into a number of strata, instead of enumerating them completely. The use of three stage sampling

designs generally specifies three stages of selection: primary sampling units (*PSUs*) at the first stage, sub samples from each *PSUs* at second stage as secondary sampling units (*SSUs*) units and again sub samples from *SSUs* at third stage as tertiary sampling units (*TSUs*). For instance, in surveys to estimate crop production in India (Sukhatme, 1947), the village is a convenient sampling unit. Within a village, only some of the fields growing the crop in question are selected, so that the field is a sub-unit. When a field is selected, only certain parts of it are cut for the determination of yield per acre; thus the sub unit itself is sampled. Here we have to find the optimal sample sizes n , m and p for all the three stages with the minimum cost. The problem of optimum allocation in two-stage and three-stage sample surveys is described in standard text book on sampling such as W.G. Cochran (1977). Recently Shafiullah *et al.* (2013) has worked on three-stage sample surveys and applied geometric programming approach for finding optimum sample sizes of each stage.

In many real-life situations the decision makers have to optimize their objectives which they have decided under certain conditions. The parameters on which the decision makers have to optimize their objectives are not always certain. The mathematical programming problem which deals with the theory and methods of the unknown parameters where the variables are considered as random is called stochastic programming problem. Stochastic programming plays very important role for modeling optimization problems. Uncertainty is the root of the stochastic programming. The main target of using stochastic programming is for finding such solution where the feasibility occurs for all data and optimal in some cases. The stochastic programming is discussed by many authors in their text books such as Prékopa (1995), Charnes and Cooper (1959).

The stochastic programming approach is applied by many researchers in the field sample surveys. Some of them are Ali *et al.* (2011), Khan *et al.* (2011, 2012), Bakhshi *et al.* (2010), Javed *et al.* (2009), Kozak (2006), Diaz-Garcia and Tapia (2007), Diaz-Garcia and Cortez (2006, 2008) and many more.

In this paper, we have formulated the three-stage sample surveys problem as a stochastic programming problem. In three-stage sample surveys problem, we have considered that sampling variance and stratum costs has normally distributed random variable. The stochastic formulation of the problem has been converted into equivalent deterministic form by using chance constrained programming and modified E - model respectively.

2. Formulation of the problem in Three -Stage Stratified Sample Surveys

The population is considered to be a heterogeneous population; it is turned into a homogeneous population by dividing it into L homogeneous stratum. Let h^{th} strata have N_h population such that $N = \sum_{h=1}^L N_h$. Now, primary stage units (*PSU*) are selected from each strata taking into consideration the sizes to be constant within a stratum but may differ from stratum to stratum. As, is the case of third stage unit (*TSU*), an *SSU* is selected from *PSU* and further, a *TSU* is selected from *SSU* such that the h^{th} stratum contains N_h *PSUs* with M_h *SSUs* having P_h *TSUs*. Also, their corresponding sample sizes are n_h , m_h and p_h , $h = 1, 2, \dots, L$ with equal probability and without replacement at each stage.

Let the value in the population of k^{th} TSU in the j^{th} SSU in i^{th} PSU of h^{th} strata be Y_{ijk}^h such that $i = 1, 2, \dots, n_h, j = 1, 2, \dots, m_h, k = 1, 2, \dots, p_h, h = 1, 2, \dots, L$.

Below are some of the usual notations that refer to h^{th} strata,

Sample mean of TSU that were selected,

$$\bar{y}_{ij}^h = \sum_{k=1}^{p_h} \frac{y_{ijk}^h}{p_h}.$$

Population mean of TSU that were selected,

$$\bar{Y}_{ij}^h = \sum_{k=1}^{P_h} \frac{Y_{ijk}^h}{P_h}.$$

Sample mean of SSU that were selected,

$$\bar{\bar{y}}_i^h = \sum_{j=1}^{m_h} \frac{\bar{y}_{ij}^h}{m_h}.$$

Population mean per SSU that were selected,

$$\bar{\bar{Y}}_i^h = \sum_{j=1}^{M_h} \frac{\bar{Y}_{ij}^h}{M_h}.$$

Sample mean of PSU that were selected,

$$\bar{\bar{\bar{y}}}^h = \sum_{i=1}^{n_h} \frac{\bar{\bar{y}}_i^h}{n_h}.$$

Population mean of PSU that were selected,

$$\bar{\bar{\bar{Y}}}^h = \sum_{i=1}^{N_h} \frac{\bar{\bar{Y}}_i^h}{N_h}.$$

Required variances are

Sampling variance among PSU means in h^{th} stratum,

$$s_{ph}^2 = \frac{1}{n_h-1} \sum_{i=1}^{n_h} \left(\bar{\bar{\bar{y}}}^h - \bar{\bar{\bar{Y}}}^h \right)^2.$$

Sampling variance among SSU's within PSU means in h^{th} stratum,

$$s_{sh}^2 = \frac{1}{n_h(m_h-1)} \sum_{i=1}^{n_h} \sum_{j=1}^{m_h} \left(\bar{y}_{ij}^h - \bar{\bar{y}}_i^h \right)^2.$$

Sampling variance among TSU's within SSU means in h^{th} stratum,

$$s_{th}^2 = \frac{1}{n_h m_h (p_h-1)} \sum_{i=1}^{n_h} \sum_{j=1}^{m_h} \sum_{k=1}^{p_h} \left(Y_{ijk}^h - \bar{y}_{ij}^h \right)^2.$$

Population variance among PSU means in h^{th} stratum

$$S_{ph}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} \left(\bar{\bar{\bar{y}}}^h - \bar{\bar{\bar{Y}}}^h \right)^2.$$

Population variance among SSU's within PSU means in h^{th} stratum,

$$S_{sh}^2 = \frac{1}{N_h(M_h-1)} \sum_{i=1}^{N_h} \sum_{j=1}^{M_h} \left(\bar{Y}_{ij}^h - \bar{\bar{Y}}_i^h \right)^2.$$

Population variance among TSU's within SSU means in h^{th} stratum,

$$S_{th}^2 = \frac{1}{N_h M_h (P_h-1)} \sum_{i=1}^{N_h} \sum_{j=1}^{M_h} \sum_{k=1}^{P_h} \left(Y_{ijk}^h - \bar{Y}_{ij}^h \right)^2.$$

An unbiased estimate of population mean, \bar{Y}_{ij} , per TSU may be written as

$$\bar{Y}_{st} = \sum_{h=1}^L \left(\frac{N_h M_h P_h \bar{y}^h}{\sum_{h=1}^L N_h M_h P_h} \right) = \sum_{h=1}^L W_h \bar{y}^h$$

where $W_h = \frac{N_h M_h P_h}{\sum_{h=1}^L N_h M_h P_h}$ is the relative size of the stratum in terms of the TSU's.

It is known that for stratified random sampling, WOR, with \bar{y}^h as the unbiased estimator of population mean \bar{Y}_{st} , the sampling variance is given by

$$V(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 \left(\frac{(1-f_{1h})}{n_h} S_{sh}^2 + \frac{(1-f_{2h})}{n_h m_h} S_{ph}^2 + \frac{(1-f_{3h})}{n_h m_h p_h} S_{th}^2 \right)$$

where $f_{1h} = \frac{n_h}{N_h}$, $f_{2h} = \frac{m_h}{M_h}$, $f_{3h} = \frac{p_h}{P_h}$ are the sample fraction at various stage and its estimated variance ignoring the fps is given by

$$\hat{V}(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 \left(\frac{S_{ph}^2}{n_h} + \frac{S_{sh}^2}{n_h m_h} + \frac{S_{th}^2}{n_h m_h p_h} \right)$$

Now, if the travel cost may be ignored, the total cost of survey can be written in the linear form given below

$$C_0 = \sum_{h=1}^L (n_h c_{1h} + n_h m_h c_{2h} + n_h m_h p_h c_{3h})$$

where

$$C_0 = C - c_0$$

C is the overall cost of sampling.

c_0 is the fixed cost in survey.

c_{1h} is the cost of obtaining information from the sampled FSU from the h^{th} stratum.

c_{2h} is the cost of obtaining information from the sampled SSU from the h^{th} stratum.

c_{3h} is the cost of obtaining information from the sampled TSU from the h^{th} stratum.

In practice, c_{1h} is likely to be larger than c_{2h} and c_{2h} is likely to be larger than c_{3h} . Hence, a unit increase in n_h increases the cost much more as compared to a unit increase in m_h similarly, a unit increase in m_h is much more compared to a unit increase in p_h . Thus, the third component of cost function will vary from sample to sample for given m_h .

If C_0 is considered as a finite limit on cost and the optimum size of n_h , m_h and p_h is required to be found so that the total survey variance can be minimized the allocation problem will be of the following Non Linear Programming Problem (NLPP) form with v characteristics can be given by

$$\left. \begin{aligned} \text{Minimize } V_r &= \sum_{h=1}^L W_h^2 \left(\frac{s_{p_{hr}}^2}{n_h} + \frac{s_{s_{hr}}^2}{n_h m_h} + \frac{s_{t_{hr}}^2}{n_h m_h p_h} \right) \\ \text{subject to } \sum_{h=1}^L (n_h c_{1h} + n_h m_h c_{2h} + n_h m_h p_h c_{3h}) &\leq C_o \\ n_h, m_h, p_h &\geq 2 \\ n_h \leq M_h, m_h \leq M_h, p_h \leq P_h; h = 1, 2, \dots, L; r = 1, \dots, v \end{aligned} \right\} \quad (1)$$

where $s_{p_{hr}}^2$, $s_{s_{hr}}^2$ and $s_{t_{hr}}^2$ are the sample variances at each stage with r^{th} characteristic, $r = 1, \dots, v$.

Now, let us assume that c_{1h} , c_{2h} and c_{3h} , $h = 1, 2, \dots, L$ to be independently normally distributed random variables. Further, sampling variance $s_{p_h}^2$, $s_{s_h}^2$ and $s_{t_h}^2$ in the h^{th} stratum are also random variables.

$$\left. \begin{aligned} \text{Minimize } \sum_{h=1}^L W_h^2 \left(\frac{s_{p_{hr}}^2}{n_h} + \frac{s_{s_{hr}}^2}{n_h m_h} + \frac{s_{t_{hr}}^2}{n_h m_h p_h} \right) & \quad (i) \\ \text{subject to } P \left[\sum_{h=1}^L (n_h c_{1h} + n_h m_h c_{2h} + n_h m_h p_h c_{3h}) \leq C_o \right] \geq P_o & \quad (ii) \\ n_h, m_h, p_h \geq 2 & \quad (iii) \\ n_h \leq M_h, m_h \leq M_h, p_h \leq P_h, h = 1, 2, \dots, L; r = 1, \dots, v & \quad (iv) \end{aligned} \right\} \quad (2)$$

4. Solution Using Modified E-technique

In objective function of Eq. 2 (i) $s_{p_{hr}}^2$, $s_{s_{hr}}^2$ and $s_{t_{hr}}^2$ are considered random variables with asymptotic normal distribution. Consider the random variable ξ_p defined as (See Melaku, 1986)

$$\xi_p^h = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(\bar{y}_i^h - \bar{\bar{Y}}^h \right)^2$$

which has an asymptotic normal distribution with mean

$$E(\xi_p^h) = \frac{n_h}{n_h - 1} S_{p_{hr}}^2$$

and

$$V(\xi_p^h) = \frac{n_h}{(n_h - 1)^2} \left(C_{p_{hr}}^4 - (S_{p_{hr}}^2)^2 \right),$$

where $C_{p_{hr}}^4$ is the fourth central moment and it is computed as

$$C_{p_{hr}}^4 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(\bar{y}_i^h - \bar{\bar{Y}}^h \right)^4$$

Observe that, $s_{p_h r}^2 = \xi_p^h - \frac{n_h}{n_h - 1} \left(\bar{y}_i^h - \bar{\bar{Y}}^h \right)^2$;

where $\frac{n_h}{n_h - 1} \rightarrow 1$ and $\left(\bar{y}_i^h - \bar{\bar{Y}}^h \right)^2 \rightarrow 0$ in terms of probability.

Consider the random variable ξ_s^h defined as

$$\xi_s^h = \frac{1}{n_h(m_h - 1)} \sum_{i=1}^{n_h} \sum_{j=1}^{m_h} \left(\bar{y}_{ij}^h - \bar{\bar{Y}}^h \right)^2$$

which has an asymptotic normal distribution with mean $E(\xi_s)$ and $V(\xi_s)$ given by

$$E(\xi_s^h) = \frac{m_h}{(m_h - 1)} S_{shr}^2$$

and

$$V(\xi_s^h) = \frac{m_h}{(m_h - 1)^2} (C_{shr}^4 - (S_{shr}^2)^2),$$

where C_{sh}^4 is the fourth central moment and it is computed as

$$C_{sh}^4 = \frac{1}{N_h(M_h - 1)} \sum_{j=1}^{M_h} \left(\bar{y}_{ij}^h - \bar{\bar{Y}}^h \right)^4$$

Observe that, $s_{shr}^2 = \xi_s^h - \frac{m_h}{m_h - 1} \left(\bar{y}_{ij}^h - \bar{\bar{Y}}^h \right)^2$;

where $\frac{m_h}{m_h - 1} \rightarrow 1$ and $\left(\bar{y}_{ij}^h - \bar{\bar{Y}}^h \right)^2 \rightarrow 0$ in terms of probability.

Again, consider the random variable ξ_t^h defined as

$$\xi_t^h = \frac{1}{n_h m_h (p_h - 1)} \sum_{i=1}^{n_h} \sum_{j=1}^{m_h} \sum_{k=1}^{p_h} \left(Y_{ijk}^h - \bar{\bar{Y}}_{ij}^h \right)^2$$

which has an asymptotic normal distribution with mean $E(\xi_t)$ and variance $V(\xi_t)$

$$E(\xi_t^h) = \frac{p_h}{(p_h - 1)} S_{thr}^2$$

$$\text{and } V(\xi_t^h) = \frac{p_h}{(p_h - 1)^2} (C_{thr}^4 - (S_{thr}^2)^2)$$

where C_{th}^4 is the fourth central moment and it is computed as

$$C_{th}^4 = \frac{1}{N_h M_h (P_h - 1)} \sum_{k=1}^{P_h} \left(Y_{ijk}^h - \bar{\bar{Y}}_{ij}^h \right)^4$$

Observe that, $s_{thr}^2 = \xi_t^h - \frac{p_h}{p_h - 1} \left(Y_{ijk}^h - \bar{\bar{Y}}_{ij}^h \right)^2$;

where $\frac{p_h}{p_h - 1} \rightarrow 1$ and $\left(Y_{ijk}^h - \bar{\bar{Y}}_{ij}^h \right)^2 \rightarrow 0$ in terms of probability.

Now, modified E-model technique (Garcia-2007) is applied, so that the equivalent deterministic objective function of NLPP (2) can be written as

$$f(n, m, p) =$$

$$k_1 E \left(\sum_{h=1}^L W_h^2 \left(\frac{s_{p_h r}^2}{n_h} + \frac{s_{s_h r}^2}{n_h m_h} + \frac{s_{t_h r}^2}{n_h m_h p_h} \right) \right) + k_2 \sqrt{V \left(\sum_{h=1}^L W_h^2 \left(\frac{s_{p_h r}^2}{n_h} + \frac{s_{s_h r}^2}{n_h m_h} + \frac{s_{t_h r}^2}{n_h m_h p_h} \right) \right)}$$

$$\begin{aligned} \text{Now, } E\left(\sum_{h=1}^L W_h^2 \left(\frac{s_{p_{hr}}^2}{n_h} + \frac{s_{s_{hr}}^2}{n_h m_h} + \frac{s_{t_{hr}}^2}{n_h m_h p_h}\right)\right) &= \sum_{h=1}^L W_h^2 \left(\frac{E(s_{p_{hr}}^2)}{n_h} + \frac{E(s_{s_{hr}}^2)}{n_h m_h} + \frac{E(s_{t_{hr}}^2)}{n_h m_h p_h}\right) \\ \sum_{h=1}^L W_h^2 \left(\frac{E(\xi_p^h)}{n_h} + \frac{E(\xi_s^h)}{n_h m_h} + \frac{E(\xi_t^h)}{n_h m_h p_h}\right) &= \sum_{h=1}^L \left(\frac{W_h^2 S_{p_{hr}}^2}{(n_h-1)} + \frac{W_h^2 S_{s_{hr}}^2}{n_h(m_h-1)} + \frac{W_h^2 S_{t_{hr}}^2}{n_h m_h(p_h-1)}\right) \\ \text{Also, } V\left(\sum_{h=1}^L W_h^2 \left(\frac{s_{p_{hr}}^2}{n_h} + \frac{s_{s_{hr}}^2}{n_h m_h} + \frac{s_{t_{hr}}^2}{n_h m_h p_h}\right)\right) &= \sum_{h=1}^L W_h^2 \left(\frac{V(s_{p_{hr}}^2)}{n_h} + \frac{V(s_{s_{hr}}^2)}{n_h m_h} + \frac{V(s_{t_{hr}}^2)}{n_h m_h p_h}\right) \\ \sum_{h=1}^L W_h^2 \left(\frac{V(\xi_p^h)}{n_h} + \frac{V(\xi_s^h)}{n_h m_h} + \frac{V(\xi_t^h)}{n_h m_h p_h}\right) &= \sum_{h=1}^L \left(\frac{W_h^4 C_P}{n_h(n_h-1)^2} + \frac{W_h^4 C_S}{n_h^2 m_h(m_h-1)^2} + \frac{W_h^4 C_T}{n_h^2 m_h^2 p_h(p_h-1)^2}\right) \end{aligned}$$

where $C_P = C_{p_{hr}}^4 - (S_{p_{hr}}^2)^2$, $C_S = C_{s_{hr}}^4 - (S_{s_{hr}}^2)^2$ and $C_T = C_{t_{hr}}^4 - (S_{t_{hr}}^2)^2$

The equivalent deterministic form of Eq. 2(i) can be obtained by using modified E-model as

$$\begin{aligned} V_r = & k_1 \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{p_{hr}}^2}{(n_h-1)} + \frac{W_h^2 S_{s_{hr}}^2}{n_h(m_h-1)} + \frac{W_h^2 S_{t_{hr}}^2}{n_h m_h(p_h-1)} \right) \right) + \\ & k_2 \left(\sqrt{\sum_{h=1}^L \left(\frac{W_h^4 C_P}{n_h(n_h-1)^2} + \frac{W_h^4 C_S}{n_h^2 m_h(m_h-1)^2} + \frac{W_h^4 C_T}{n_h^2 m_h^2 p_h(p_h-1)^2} \right)} \right) \quad (3) \end{aligned}$$

3. Solution Using Chance Constraint Programming

The costs c_{1h} , c_{2h} and c_{3h} in the constraint are assumed to be normally distributive random variables.

$$E(\sum_{h=1}^L c_{1h} n_h) = \sum_{h=1}^L n_h E(c_{1h}) = \sum_{h=1}^L n_h \mu_{1h} \quad (4)$$

$$V(\sum_{h=1}^L c_{1h} n_h) = \sum_{h=1}^L n_h^2 V(c_{1h}) = \sum_{h=1}^L n_h^2 \sigma_{1h}^2 \quad (5)$$

$$E(\sum_{h=1}^L c_{2h} n_h m_h) = \sum_{h=1}^L n_h m_h E(c_{2h}) = \sum_{h=1}^L n_h m_h \mu_{2h} \quad (6)$$

$$V(\sum_{h=1}^L c_{2h} n_h m_h) = \sum_{h=1}^L n_h^2 m_h^2 V(c_{2h}) = \sum_{h=1}^L n_h^2 m_h^2 \sigma_{2h}^2 \quad (7)$$

$$E(\sum_{h=1}^L c_{3h} n_h m_h p_h) = \sum_{h=1}^L n_h m_h p_h E(c_{3h}) = \sum_{h=1}^L n_h m_h p_h \mu_{3h} \quad (8)$$

$$V(\sum_{h=1}^L c_{3h} n_h m_h p_h) = \sum_{h=1}^L n_h^2 m_h^2 p_h^2 V(c_{3h}) = \sum_{h=1}^L n_h^2 m_h^2 p_h^2 \sigma_{3h}^2 \quad (9)$$

Finally, from Eq. (4), (6) and (8) mean of objective function with random cost will be

$$\begin{aligned} E(C) = & \sum_{h=1}^L (n_h \mu_{1h} + n_h m_h \mu_{2h} + \\ & n_h m_h p_h \mu_{3h}) \quad (10) \end{aligned}$$

and variance from Eq. (5), (7) and (9) with random cost is

$$\begin{aligned} V(C) = & \sum_{h=1}^L (n_h^2 \sigma_{1h}^2 + n_h^2 m_h^2 \sigma_{2h}^2 + \\ & n_h^2 m_h^2 p_h^2 \sigma_{3h}^2) \quad (11) \end{aligned}$$

Since, $\mu_{1h}, \mu_{2h}, \mu_{3h}, \sigma_{1h}^2, \sigma_{2h}^2$ and σ_{3h}^2 are unknown and therefore they are replaced by their estimators. The estimator of $E(C)$ is

$$\hat{E}(C) = \sum_{h=1}^L (n_h \bar{c}_{1h} + n_h m_h \bar{c}_{2h} + n_h m_h p_h \bar{c}_{3h}) \tag{12}$$

The estimator of $V(\sum_{h=1}^L n_h (c_{1h} + m_h (c_{2h} + p_h c_{3h})))$ is

$$\hat{V}(C) = \sum_{h=1}^L (n_h^2 s_{1h}^2 + n_h^2 m_h^2 s_{2h}^2 + n_h^2 m_h^2 p_h^2 s_{3h}^2) \tag{13}$$

where $C = \sum_{h=1}^L (n_h c_{1h} + n_h m_h c_{2h} + n_h m_h p_h c_{3h})$.

Again, $P(\hat{V}(C) \leq C_0) \geq P_0$.

$P(\hat{V}(C) \leq C_0)$ also imply

$$P\left(\frac{\hat{V}(C) - E(\hat{V}(C))}{\sqrt{V(\hat{V}(C))}} \leq \frac{C_0 - E(\hat{V}(C))}{\sqrt{V(\hat{V}(C))}}\right)$$

which in simplified form is given by

$$P(\hat{V}(C) \leq C_0) = \Phi\left(\frac{C_0 - E(\hat{V}(C))}{\sqrt{V(\hat{V}(C))}}\right)$$

where $\left(\frac{C_0 - E(\hat{V}(C))}{\sqrt{V(\hat{V}(C))}}\right)$ is a standard normal variate (SNV) with mean zero and variance one.

Now, $\Phi(z)$ represents the cumulative density function of the SNV evaluated at z . If K_α represents the value of SNV at which $(K_\alpha) = P_0$, then the constraint be stated as

$$\Phi\left(\frac{C_0 - E(\hat{V}(C))}{\sqrt{V(\hat{V}(C))}}\right) \geq \Phi(K_\alpha)$$

This inequality will satisfy only if $\frac{C_0 - E(\hat{V}(C))}{\sqrt{V(\hat{V}(C))}} \geq K_\alpha$

or $E(\hat{V}(C)) + K_\alpha \sqrt{V(\hat{V}(C))} \leq C_0$

$$\sum_{h=1}^L (n_h \bar{c}_{1h} + n_h m_h \bar{c}_{2h} + n_h m_h p_h \bar{c}_{3h}) + K_\alpha \sqrt{\sum_{h=1}^L (n_h^2 s_{1h}^2 + n_h^2 m_h^2 s_{2h}^2 + n_h^2 m_h^2 p_h^2 s_{3h}^2)} \leq C_0 \tag{14}$$

Since, $\mu_{1h}, \mu_{2h}, \mu_{3h}, \sigma_{1h}^2, \sigma_{2h}^2$ and σ_{3h}^2 are unknown and therefore they are replaced by their estimators. The

Thus finally, the allocation problem will be using assumptions made for (see Melaku 1968) and using modified E-model (see Garcia 2007) the NLPP (3) will be formulated as

$$\text{Min } V_r = k_1 \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{phr}^2}{(n_h-1)} + \frac{W_h^2 S_{shr}^2}{n_h(m_h-1)} + \frac{W_h^2 S_{thr}^2}{n_h m_h (p_h-1)} \right) \right) + k_2 \left(\sqrt{\sum_{h=1}^L \left(\frac{W_h^4 C_P}{n_h(n_h-1)^2} + \frac{W_h^4 C_S}{n_h^2 m_h (m_h-1)^2} + \frac{W_h^4 C_{P_h}}{n_h^2 m_h^2 p_h (p_h-1)^2} \right)} \right)$$

subject to $\sum_{h=1}^L n_h \bar{c}_{1h} + n_h m_h \bar{c}_{2h} + n_h m_h p_h \bar{c}_{3h} + K_\alpha \sqrt{\sum_{h=1}^L n_h^2 s_{1h}^2 + n_h^2 m_h^2 s_{2h}^2 + n_h^2 m_h^2 p_h^2 s_{3h}^2} \leq C_0$
 $n_h, m_h, p_h \geq 2$
 $n_h \leq M_h, m_h \leq M_h, p_h \leq P_h; h = 1, 2, \dots, L ; r = 1, \dots, v$

Lexicographic Method

To solve the converted deterministic NLPP using lexicographic goal programming approach the with r characteristics arranged in lexicographic order of importance, at the first stage of the solution the NLPP with $j = 1$ has to be obtained. Let V_1^* be the optimal value of the objective function V_1 and $d_1 \geq 0$ is such that $V_1 - V_1^* \leq d_1$.

At the second stage of the solution the NLPP to be solved is given by

$$\text{Min } k_1 \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{ph2}^2}{(n_h-1)} + \frac{W_h^2 S_{sh2}^2}{n_h(m_h-1)} + \frac{W_h^2 S_{th2}^2}{n_h m_h (p_h-1)} \right) \right) + k_2 \left(\sqrt{\sum_{h=1}^L \left(\frac{W_h^4 C_P}{n_h(n_h-1)^2} + \frac{W_h^4 C_S}{n_h^2 m_h (m_h-1)^2} + \frac{W_h^4 C_{P_h}}{n_h^2 m_h^2 p_h (p_h-1)^2} \right)} \right)$$

subject to $k_1 \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{ph1}^2}{(n_h-1)} + \frac{W_h^2 S_{sh1}^2}{n_h(m_h-1)} + \frac{W_h^2 S_{th1}^2}{n_h m_h (p_h-1)} \right) \right) + k_2 \left(\sqrt{\sum_{h=1}^L \left(\frac{W_h^4 C_P}{n_h(n_h-1)^2} + \frac{W_h^4 C_S}{n_h^2 m_h (m_h-1)^2} + \frac{W_h^4 C_{P_h}}{n_h^2 m_h^2 p_h (p_h-1)^2} \right)} \right)$
 $\sum_{h=1}^L n_h \bar{c}_{1h} + n_h m_h \bar{c}_{2h} + n_h m_h p_h \bar{c}_{3h} + K_\alpha \sqrt{\sum_{h=1}^L n_h^2 s_{1h}^2 + n_h^2 m_h^2 s_{2h}^2 + n_h^2 m_h^2 p_h^2 s_{3h}^2} \leq C_0$
 $n_h, m_h, p_h \geq 2$
 $n_h \leq M_h, m_h \leq M_h, p_h \leq P_h; h = 1, 2, \dots, L ; r = 1, \dots, v$

Successively solving the problem at each stage, the NLPP at v^{th} stage will be given as

$$\text{Min } k_1 \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{phv}^2}{(n_h-1)} + \frac{W_h^2 S_{shv}^2}{n_h(m_h-1)} + \frac{W_h^2 S_{thv}^2}{n_h m_h (p_h-1)} \right) \right) + k_2 \left(\sqrt{\sum_{h=1}^L \left(\frac{W_h^4 C_P}{n_h(n_h-1)^2} + \frac{W_h^4 C_S}{n_h^2 m_h (m_h-1)^2} + \frac{W_h^4 C_{P_h}}{n_h^2 m_h^2 p_h (p_h-1)^2} \right)} \right)$$

subject to $k_1 \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{phr}^2}{(n_h-1)} + \frac{W_h^2 S_{shr}^2}{n_h(m_h-1)} + \frac{W_h^2 S_{thr}^2}{n_h m_h (p_h-1)} \right) \right) + k_2 \left(\sqrt{\sum_{h=1}^L \left(\frac{W_h^4 C_P}{n_h(n_h-1)^2} + \frac{W_h^4 C_S}{n_h^2 m_h (m_h-1)^2} + \frac{W_h^4 C_{P_h}}{n_h^2 m_h^2 p_h (p_h-1)^2} \right)} \right)$
 $\sum_{h=1}^L n_h \bar{c}_{1h} + n_h m_h \bar{c}_{2h} + n_h m_h p_h \bar{c}_{3h} + K_\alpha \sqrt{\sum_{h=1}^L n_h^2 s_{1h}^2 + n_h^2 m_h^2 s_{2h}^2 + n_h^2 m_h^2 p_h^2 s_{3h}^2} \leq C_0$
 $n_h, m_h, p_h \geq 2$
 $n_h \leq M_h, m_h \leq M_h, p_h \leq P_h; h = 1, 2, \dots, L ; r = 1, \dots, v - 1$

Other Allocation Methods

A Comparative Study

1. Proportional allocation

The proportional

2. Cochran's Allocation

The compromise criterion of Cochran's allocation is to average the individual optimum allocations of n_{vh}^* , m_{vh}^* , p_{vh}^* ; $h = 1, \dots, L$ and $r = 1, \dots, v$ that are solutions to the NLPP for all the p characteristics separately.

3. Minimizing Weighted Sum of Variances

Khan et al. (2003) conjectured that

$$a_r = \frac{\sum_{h=1}^L S_{rh}^2}{\sum_{r=1}^v \sum_{h=1}^L S_{rh}^2}, \quad \sum_{r=1}^v a_r = 1$$

$$Z(n_1, n_2, \dots, n_L) = \sum_{r=1}^v a_r V_r$$

$$= \sum_{r=1}^v a_r \left(\sum_{h=1}^L W_h^2 \left(\frac{S_{p hr}^2}{n_h} + \frac{S_{s hr}^2}{n_h m_h} + \frac{S_{t hr}^2}{n_h m_h p_h} \right) \right)$$

$$= \sum_{h=1}^L W_h^2 \left(\frac{A_{1h}^2}{n_h} + \frac{A_{2h}^2}{n_h m_h} + \frac{A_{3h}^2}{n_h m_h p_h} \right)$$

where, $A_{1h}^2 = \sum_{r=1}^v a_r S_{p hr}^2$, $h = 1, 2, \dots, L$

$A_{2h}^2 = \sum_{r=1}^v a_r S_{s hr}^2$, $h = 1, 2, \dots, L$

$A_{3h}^2 = \sum_{r=1}^v a_r S_{t hr}^2$, $h = 1, 2, \dots, L$

4. Sukhatme's Allocation

Sukhatme et al. [16] obtained the compromise allocation by minimizing the sum of the variances for the p characteristics under linear cost constraints. The NLPP for this allocation is given as

$$\text{Min } \sum_{r=1}^v k_1 \left(\sum_{h=1}^L \left(\frac{W_h^2 S_{p hr}^2}{(n_h-1)} + \frac{W_h^2 S_{s hr}^2}{n_h(m_h-1)} + \frac{W_h^2 S_{t hr}^2}{n_h m_h (p_h-1)} \right) \right) + k_2 \left(\sqrt{\sum_{h=1}^L \left(\frac{W_h^4 C_P}{n_h(n_h-1)^2} + \frac{W_h^4 C_S}{n_h^2 m_h (m_h-1)^2} + \frac{W_h^4 C_T}{n_h^2 m_h^2 p_h (p_h-1)^2} \right)} \right)$$

subject to $\sum_{h=1}^L n_h \bar{c}_{1h} + n_h m_h \bar{c}_{2h} + n_h m_h p_h \bar{c}_{3h} + K_\alpha \sqrt{\sum_{h=1}^L n_h^2 s_{1h}^2 + n_h^2 m_h^2 s_{2h}^2 + n_h^2 m_h^2 p_h^2 s_{3h}^2} \leq C_0$

$n_h, m_h, p_h \geq 2$

$n_h \leq M_h, m_h \leq M_h, p_h \leq P_h; h = 1, 2, \dots, L; r = 1, \dots, v$

Simulation Study

To illustrate the theory developed in previous section a simulation study has been done. Considering the population to possess two characteristics, randomized data with normal probabilities have been generated at each stage with total population being divided into four stratum. The data for simulation of three stage sampling is obtained through R-Software.

For the population of first characteristic at primary stage, normal random variables with specified mean and variance are generated through R-software. Later, the data is divided into four assumed stratum and mean variance and fourth moment are obtained for each

strata. The normal random variables are regenerated with different mean and standard deviation for second characteristic.

Similarly, different populations at each stage are generated and regenerated for second characteristic for pre-assumed means and variances. The required data generated through the R-software for characteristic one and two are shown in table 1 and 2 respectively.

Table 1: Characteristic one i.e. $r = 1$

h	N_h	M_h	P_h	S_{ph}^2	S_{sh}^2	S_{th}^2	C_{ph}^4	C_{sh}^4	C_{th}^4
1	28	120	240	101.4907	149.3986	338.1822	16744.84	76925.87	328265.00
2	35	88	238	127.7659	144.2444	420.0881	30352.46	58804.94	454358.00
3	25	116	256	89.4826	131.6389	434.8655	21610.17	57372.1	603104.40
4	32	96	266	106.1326	114.8018	422.8810	23220.17	39431.28	590399.00

Table 2: Characteristic two i.e. $r = 2$

h	N_h	M_h	P_h	S_{ph}^2	S_{sh}^2	S_{th}^2	C_{ph}^4	C_{sh}^4	C_{th}^4
1	28	120	240	75.0717	99.2154	198.4202	10692.33	26751.71	109388.30
2	35	88	238	78.6746	88.9380	166.7440	12815.83	31685.18	101540.80
3	25	116	256	37.6336	101.4011	152.4509	3076.868	23863.5	78502.63
4	32	96	266	62.6065	102.4223	188.7107	10507.18	27047.83	128872.50

The per unit cost for measurement in various strata are independently normally distributed with assumed means and variances as shown in table below

$E(c_{1h})$	$E(c_{2h})$	$E(c_{3h})$	s_1^2	s_2^2	s_3^2
3	2	1	0.75	0.50	0.25
4	3	1	1.00	0.75	0.25
5	4	1	1.25	1.00	0.25
6	5	1	1.50	1.25	0.25

Results

Trace Values and Cost from different methods are given in the form of table below

	Lexicographic	Proportional	Cochran	Weighted Average Sum of Variances	Sukhatme
<i>Trace Value</i>	1.0669	1.1086	1.08645	1.0669	1.0669
Cost	9954.429	9109.81	9387.852	9954.429	9954.429

Conclusions

This paper has provided comprehensive study of an optimum allocation in three-stage multivariate stratified sample surveys with costs and the variances as random parameters. The problem is formulated as a non-linear stochastic programming problem by considering survey cost and the variances as random variables. The stochastic problem of

three-stage multivariate stratified sample surveys is converted into equivalent deterministic form by using Chance constraint programming and modified E-model. Furthermore the researchers can use these formulations for obtaining optimum allocation for three-stage sample surveys whenever their costs are needed to be optimized with a limitation on variance.

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*, 83(5), pp. 960-974 (Taylor & Francis, DOI:10.1080/00949655.2011.643890).
2. Bakhshi, Z.H. Khan, M.F. and Ahmad, Q.S. (2010). Optimal sample numbers in multivariate stratified sampling with a probabilistic cost constraint, *Int. J. Math. Appl. Statist.* 1(2), pp. 111–120.
3. Charnes, A. and Cooper, W.W. (1959). Chance constrained programming, *Management Science* 5, pp. 73–79.
4. Cochran, W.G. (1977): *Sampling Techniques*, 3rd ed., Wiley, New York.
5. Díaz-García J. A. and M.M. Garay Tapia, (2007). Optimum allocation in stratified surveys: Stochastic programming, *Comput. Statist. Data Anal.* 51, pp. 3016–3026.
6. Díaz-García J.A. and Cortez, L.U. (2006). Optimum allocation in multivariate stratified sampling: Multiobjective Programming. Comunicación Técnica No. I-06-07/28-03-2006 (PE/CIMAT), México.
7. Díaz-García, J.A. and Cortez, L.U. (2008). Multi-objective Optimisation for optimum allocation in multivariate stratified sampling. *Surv. Methodol.* 34(2), 215–222
8. Ignizio, J.P., *Goal Programming and Extensions*, Lexington, Mass: Health, Lexington Books, Lexington, MA, 1976.
9. Ijiri, Y., *Management goals and Accounting for control*, North Holland, Amsterdam, 1965.
10. Javed S., Bakhshi Z.H. and Khalid M.M. (2009). Optimum allocation in stratified sampling with random costs, *Int. Rev. Pure Appl. Math.* 5(2), pp. 363–370.
11. Khan, M.F., 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 (USA)*, 2(1), pp. 122-125.
12. Khan, M.F., Ali, I. and Ahmad, Q.S. (2011). Chebyshev Approximate Solution to Allocation Problem in Multiple Objective Surveys with Random Costs 1, pp. 247-251 (doi:10.4236/ajcm.2011.14029).
13. Kozak, M. (2006). On Sample Allocation in Multivariate Surveys, *Communications in Statistics - Simulation and Computation*, 35:4, 901-910.
14. Lee, S.N., *Goal Programming for Decision Analysis*, Aurebach, Philadelphia, 1972.

15. Maqbool, S., Mir, A. H. and Mir, S. A. (2011). Geometric Programming Approach to Optimum Allocation in Multivariate Two-Stage Sampling Design. *Electronic Journal of Applied Statistical Analysis*, 4(1), pp. 71 – 82
16. Melaku, A., (1986). Asymptotic normality of the optimal allocation in multivariate stratified random sampling. *Indian J. Statist. 48 (Series B)*, 224–232.
17. Pandit, S.N.N. and Srinivas, K. (1962). A Lexisearch algorithm for traveling Salesman problem, *IEEE*, 2521-2527.
18. Prékopa, A. (1995). *Stochastic Programming, Series in Mathematics and its Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands.
19. Rao, S.S., (1979). *Optimization Theory and Applications*. *Wiley Eastern Limited*.
20. Shafiullah, Ali, I. and Bari, A. (2013): Geometric Programming Approach in Three – Stage Sampling Design, *International Journal of Scientific & Engineering Research (France)*, 4(6), pp. 2229-5518.
21. Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S., Asok, C. (1984). *Sampling Theory of Surveys with Applications*. Iowa State University Press, Iowa, U.S.A. and Indian Society of Agricultural Statistics, New Delhi.