

# **An Optimum Multivariate-Multiobjective Stratified Sampling Design: Fuzzy Programming Approach**

Rahul Varshney

Department of Applied Statistics

Babasaheb Bhimrao Ambedkar University, Lucknow-226 025, India

Srikant Gupta

Department of Statistics and Operations Research

Aligarh Muslim University, Aligarh-202 002, India

Irfan Ali

Department of Statistics and Operations Research

Aligarh Muslim University, Aligarh-202 002, India

## **Abstract**

In stratified sampling design when the cost of measuring the units is not significant in each stratum, the estimation of population mean or total constructed from a selected sample according to Neyman allocation is advisable. In general the practical use of Neyman allocation suffers from a number of limitations, when there is no information about strata standard deviations except about the equality of standard deviations between some of the strata, then the precision of the estimate may be increased by pooling the strata with equal standard deviations as a single stratum and the problem of allocation is resolved by using Neyman and proportional allocations simultaneously. In this paper the case of multiple pooling of the standard deviations of the estimates in a multivariate stratified sampling for more than three strata. The problem is formulated as a Multiobjective Nonlinear Programming Problem and its solution procedure is suggested by using Fuzzy Programming approach.

**Keywords:** Multivariate Stratified Sampling, Compromise Allocation, Pooled Standard Deviations, Multiple Pooling, Multiobjective Nonlinear Programming, Fuzzy Programming.

## **1. Introduction**

In sampling literature the problem of determining the sample sizes of the units among strata that minimizes the sampling variance of the estimator of the population mean (or total) for a fixed cost or minimizes the total cost of the survey for a fixed precision of the estimator is termed as the problem of allocation. By using either of the above criteria the allocation, so obtained, is known as optimum allocation. When the measurement cost is constant and does not vary from stratum to stratum, minimizing the total cost of the survey is equivalent to minimize the total sample size (Cochran (1977)). The optimum allocation for univariate stratified populations was first suggested by Neyman (1934). Later on, Mahalonobis (1944) introduced the cost function. Stuart (1954) used Cauchy's inequality for sample allocation.

Suppose that a population of size  $N$  has stratified into  $L$  non-overlapping and exhaustive strata of sizes  $N_1, N_2, \dots, N_h, \dots, N_L$  with  $N = \sum_{h=1}^L N_h$ . The problem arises in determining the values of sample sizes  $n_h$ ,  $h = 1, 2, \dots, L$  of the units from each stratum of the population.

The optimum allocation for fixed total sample size  $n$  is termed as Neyman allocation and is given by

$$n_h^* = n \frac{W_h S_h}{\sum_{h=1}^L W_h S_h}, \quad h=1, 2, \dots, L, \quad (1.1)$$

where,  $W_h$  is the stratum weight and  $S_h$  is the stratum standard deviation for  $h^{th}$  stratum. In most of situations the true values of  $S_h$ ,  $h=1, 2, \dots, L$ , are unknown but their sample estimates  $s_h$ ,  $h=1, 2, \dots, L$  may be used to determine  $n_h$ ,  $h=1, 2, \dots, L$  which are given as

$$\hat{n}_h = n \frac{W_h s_h}{\sum_{h=1}^L W_h s_h}, \quad h=1, 2, \dots, L, \quad (1.2)$$

where  $\hat{n}_h$  are called as the Modified Neyman allocation (see Sukhatme *et al.* (1984)).

During stratification some strata variances are unknown but may be assumed with equal variances, as discussed by Park *et al.* (2007). They obtained an allocation by using estimated pooled standard deviations and proportional allocation for combined strata. First they obtained pooled standard deviation for a single stratum which comprises of strata with equal variances by pooling and worked out the modified Neyman allocation. The allocation, for the pooled stratum, is then reallocated among its constituent strata by the use of proportional allocation. They also showed that under certain conditions their allocation outperforms in comparison to modified Neyman allocation and proportional allocation.

Ansari *et al.* (2011) justified the assumption of the equality of some of the stratum variances as considered by the Park *et al.* (2007). Practically, there may be circumstances that allow this assumption. For example, consider a population with  $L$  strata and these strata are constructed with a view to make them internally homogeneous as far as possible. For administrative convenience there is a need of division of large homogeneous stratum into smaller strata for some reasons and also strata variances of the smaller strata are not significantly different. This can be ascertained by testing  $H_0 : S_h^2 = S_k^2; h < k; h = 1, 2, \dots, L-1, k = 2, 3, \dots, L$  pair-wise.

In the multivariate survey when  $p$ -characteristics are to be measured on each selected units of the sample. The optimum allocation for one characteristic may be far from optimality of other characteristics (Khan *et al.* (1997)) unless the characteristics are highly correlated. Thus to obtain an allocation which is optimum for all the characteristics, in some sense, we need a compromise criterion that suits well to all the characteristics in some respect. An allocation based on such compromise criterion is termed as a compromise allocation in sample surveys.

Various authors suggested different compromise criteria or explored further the already existing criteria. Among them are Dalenius (1957), Ghosh (1958), Yates (1960), Aoyama (1963), Folks and Antle (1965), Kokan and Khan (1967), Chatterjee (1967, 1968), Arvanitis and Afonja (1971), Ahsan and Khan (1977, 1982), Melaku and Sadasivan (1987), Bankier (1988), Bethel (1989), Kreienbrock (1993), Jahan *et al.* (1994), Khan *et al.* (1997), Khan *et al.* (2003), Ahsan *et al.* (2005), Díaz-García and Cortez (2006, 2008), Ansari *et al.* (2009), Varshney *et al.* (2012), Varshney *et al.* (2014), Varshney *et al.* (2015), and many others. Kozak (2006a) gave three different compromise criteria and modified the random search method to develop an algorithm to obtain the compromise allocation for multivariate stratified populations. Kozak (2006b) discussed five different criteria to work out approximate optimum allocation in multivariate surveys and compared them using a simulation study.

In the present paper the idea of pooling the standard deviations is extended to obtain a compromise allocation in a multivariate stratified population when the true values of the stratum standard deviations are unknown but the additional information about equality of standard deviations for a specified group of strata and the estimates of the strata standard deviations are available. The case of multiple pooling is also considered for the situation when there are more than one groups of strata that have equal stratum variances. It is assumed that the  $p$ -characteristics are independent and the estimation of population means  $\bar{Y}_j$ ;  $j=1,2,\dots,p$  is of interest for a fixed budget sample survey.

The problem has formulated as multiobjective nonlinear programming problem (MNLPP) to obtain a compromise allocation by minimizing the variances of the estimates of  $p$ -population means simultaneously for prefixed budget of the survey. The multiobjective formulation is converted into a single objective function by using Fuzzy Programming Technique. A simulation study, carried out by Ansari *et al.* (2011), is reconsidered to have two separate numerical examples for illustration and comparison with the proposed allocation.

## 2. Formulation of the Problem: The Univariate Case

In stratified random sampling, the stratified sample mean

$$\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h \tag{2.1}$$

is an unbiased estimator of the overall population mean

$$\bar{Y} = \sum_{h=1}^L W_h \bar{Y}_h \tag{2.2}$$

with a sampling variance

$$V(\bar{y}_{st}) = \sum_{h=1}^L \frac{W_h^2 S_h^2}{n_h} - \sum_{h=1}^L \frac{W_h^2 S_h^2}{N_h} \tag{2.3}$$

where,  $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$  is sample mean from  $h^{th}$  stratum, (2.4)

$$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi} \text{ is the stratum mean for the } h^{th} \text{ stratum} \quad (2.5)$$

and  $y_{hi}$  is the value of the  $i^{th}$  unit of the  $h^{th}$  stratum/sample from  $h^{th}$  stratum.

The problem of obtaining a Neyman allocation may be formulated as the following Nonlinear Programming Problem (NLPP)

$$\begin{aligned} \text{Minimize} \quad & V(\bar{y}_{st}) = \sum_{h=1}^L \frac{W_h^2 S_h^2}{n_h} - \sum_{h=1}^L \frac{W_h^2 S_h^2}{N_h} \\ \text{subject to} \quad & \sum_{h=1}^L n_h \leq n \\ \text{and} \quad & n_h \geq 0; h = 1, 2, \dots, L \end{aligned} \quad (2.6)$$

[The symbols used in this manuscript are as used in Cochran (1977) unless specified otherwise.]

In the objective function of NLPP (2.6) the second term in the expression of  $V(\bar{y}_{st})$  may be ignored because it is independent of  $n_h$ . Furthermore, the usual non-negativity restrictions  $n_h \geq 0$  may be taken as  $2 \leq n_h \leq N_h; h = 1, 2, \dots, L$  to estimate the strata variances and to avoid the problem of oversampling. After incorporating these, the NLLP (2.6) may be restated as

$$\begin{aligned} \text{Minimize} \quad & f(n_h) = \sum_{h=1}^L \frac{W_h^2 S_h^2}{n_h} \\ \text{subject to} \quad & \sum_{h=1}^L n_h \leq n \\ \text{and} \quad & 2 \leq n_h \leq N_h; h = 1, 2, \dots, L, \end{aligned} \quad (2.7)$$

where  $f(n_h)$  denote the variance of  $\bar{y}_{st}$  ignoring fpc.

### 3. The Solution: The Univariate Case

The approach considered by Park *et al.* (2007) is summarized here for the sake of continuity in univariate case and to formulate the problem for its multivariate case.

In absence of the knowledge of the true values of the strata standard deviations their estimates are used to work out an optimum allocation. If the additional information about the equality of the strata standard deviations is available then Park *et al.* (2007) showed that under certain conditions this information could be used to improve the precision of the estimator  $\bar{y}_{st}$  of the population mean  $\bar{Y}$ .

First, the preliminary samples of sizes  $n'_h$  are drawn to work out the estimates  $s_h$  of the unknown  $S_h$ . The sizes of the preliminary samples may be worked out using the criterion

given in Sukhatme *et al.* (1984). The strata with equal  $S_h$  are then combined into a single stratum and the samples sizes are allocated by the modified Neyman allocation using the pooled variance obtained by pooling the equal strata variances. The sample size allocated to the combined stratum is then reallocated to its constituent strata using proportional allocation.

According to the above scheme if in a stratified population some of the strata (say  $k$ ) are known to have equal variances then without loss of generality it can be assumed that the first  $k$  ( $<L$ ) strata have equal variances, that is,  $S_1^2 = S_2^2 = \dots = S_k^2$ . These  $k$  strata when combined into a single stratum will have the pooled estimated standard deviation denoted by  $s_{pool}$  as

$$s_{pool} = \left[ \frac{\sum_{h=1}^k (n'_h - 1) s_h^2}{\sum_{h=1}^k (n'_h - 1)} \right]^{1/2} ; \quad h=1, 2, \dots, k \quad (3.1)$$

The simulation study carried out by the authors (Section 6) showed that the above assumption of equality of variances is not a rigid condition and if some of the strata variances are approximately equal ( $\pm 10\%$ ) even then the compromise allocation works well.

The Park's compromise allocation is then given by

$$n_h = n \frac{\left( \sum_{h=1}^k W_h \right) s_{pool}}{\left( \sum_{h=1}^k W_h \right) s_{pool} + \sum_{h=k+1}^L W_h S_h} \times \frac{W_h}{\sum_{h=1}^k W_h}$$

$$= n \frac{W_h s_{pool}}{\left( \sum_{h=1}^k W_h \right) s_{pool} + \sum_{h=k+1}^L W_h S_h} ; \quad h=1, 2, \dots, k \quad (3.2)$$

$$n_h = n \frac{W_h S_h}{\left( \sum_{h=1}^k W_h \right) s_{pool} + \sum_{h=k+1}^L W_h S_h} ; \quad h=k+1, k+2, \dots, L \quad (3.3)$$

where  $n$  denote the total sample size.

Park *et al.* (2007) also showed that if the differences between unequal strata standard deviations are large, then the estimator based on the suggested compromise allocation is more efficient than proportional allocation.

#### 4. The Problem: The Multivariate Case

Let us consider a stratified population with  $L$  strata and  $p$  characteristics. The estimation of  $p$  population means  $\bar{Y}_j; j=1,2,\dots,p$  is of interest, the NLPP (2.7) for the  $j^{th}$  characteristic can be expressed as:

$$\begin{aligned} \text{Minimize} \quad & f_j(n_{jh}) = \sum_{h=1}^L \frac{W_h^2 S_{jh}^2}{n_{jh}} \\ \text{subject to} \quad & \sum_{h=1}^L n_{jh} \leq n \quad (4.1) \\ \text{and} \quad & 2 \leq n_{jh} \leq N_h ; h = 1, 2, \dots, L. \end{aligned}$$

The suffix ‘ $j$ ’ has been introduced to represent the  $j^{th}$  characteristic (Ansari *et al.* (2011)). For a particular characteristic the strata, having equal or nearly equal stratum standard deviations, are combined into a single stratum. The pooled standard deviation is worked out using (3.1). The sample sizes are then allocated according to the modified Neyman allocation using the pooled standard deviations. The sample size allocated to the combined stratum is then reallocated to their constituents strata according to the proportional allocation. This gives Park’s compromise allocation for a particular characteristic as given in (3.2) and (3.3). Let  $V_j^*$  be the value of the variance of  $\bar{y}_{jst}$  (fpc ignored) under this compromise allocation.

To consider the multivariate case assumes that:

- (a) the stratum standard deviations  $S_{jh}; j=1,2,\dots,p; h=1,2,\dots,L$  be unknown but (i) their estimates are available from preliminary samples of sizes  $n'_{jh}$ .  
(ii) it is known that some of the strata have equal or nearly equal stratum variances.
- (b) for the  $j^{th}$  characteristics there are  $l_j$  groups  $G_{j1}, G_{j2}, \dots, G_{jl_j}$  having equal or nearly equal stratum variances.
- (c)  $g_{jk}$  be the number of strata in the group  $G_{jk}; k=1,2,\dots,l_j, \dots$ . The remaining strata with unequal stratum variance are treated as usual.
- (d) for  $j^{th}$  characteristic the number of strata having unequal stratum variances is equal to  $\left( L - \sum_{k=1}^{l_j} g_{jk} \right); j=1,2,\dots,p$ .

So that while working out the Modified Neyman allocation, the total number of strata,  $L_j$

(say), for  $j^{th}$  characteristic is given by 
$$L_j = \left( L + l_j - \sum_{k=1}^{l_j} g_{jk} \right); j=1,2,\dots,p.$$

For the  $j^{th}$  characteristic the number of poolings will be  $l_j$ . These  $l_j$  pooled standard deviations for equal or nearly equal stratum variances are given by

$$s_{j\ pool_k} = \left[ \frac{\sum_{h \in M_{jk}} (n'_{jh} - 1) s_{jh}^2}{\sum_{h \in M_{jk}} (n'_{jh} - 1)} \right]^{1/2}; \quad k=1,2,\dots,l_j \quad (4.2)$$

where  $M_{jk}; j=1,2,\dots,p, k=1,2,\dots,l_j$ , denote the set of  $g_{jk}$  indices of the strata constituting the group  $G_{jk}$  and  $n'_{jh}$  and  $s_{jh}$  are preliminary sample size and the estimated standard deviations respectively.

For the  $j^{th}$  characteristic the Modified Neyman allocation with pooled stratum variances will be the solution of the NLPP (4.1) re-expressed, incorporating the assumptions laid down earlier, as the NLPP:

$$\begin{aligned} &\text{Minimize} && f_j(m_{jk}, n_h) \\ &\text{subject to} && \sum_{k=1}^{l_j} m_{jk} + \sum_{h \notin M_{jk}} n_h \leq n \\ &&& 2 \leq n_h \leq N_h; h \notin M_{jk} \\ &\text{and} && 2l_j \leq m_{jk} \leq \sum_{h \in M_{jk}} N_h; j=1,2,\dots,p, k=1,2,\dots,l_j \end{aligned} \quad (4.3)$$

where  $m_{jk}; j=1,2,\dots,p, k=1,2,\dots,l_j$ , denote the combined sample sizes for the  $k^{th}$  group of strata for  $j^{th}$  characteristics.

The optimum values of  $m_{jk}$  (say  $m_{jk}^*$ ) are reallocated to their constituent strata according to proportional allocation. This will give the values of  $n_{jh}^*$  ( $j=1,2,\dots,p, h=1,2,\dots,L$ ).

Let  $f_j^*; j=1,2,\dots,p$  denote the optimal value of the objective function of NLPP (4.1) that is the values of  $f_j(n_{jh}^*); j=1,2,\dots,p$ .

$$\begin{aligned} &\text{Obviously} && f_j(n_{jh}^*) \leq f_j(n_h) \quad ; j=1,2,\dots,p, h=1,2,\dots,L \\ &\text{or} && f_j(n_h) - f_j(n_{jh}^*) \geq 0 \end{aligned} \quad (4.4)$$

where  $n_h; h=1,2,\dots,L$  denote a compromise allocation. The LHS of (4.4) denote the increase in the sampling variances of the estimate of  $\bar{Y}_j; j=1,2,\dots,p$  for using the compromise allocation instead of their individual optimum allocations.

### 5. The Fuzzy Programming Approach

Obviously the best allocation will be the solution of the following MNLPP

$$\begin{aligned} &\text{Minimize} && [f_1(n_h), f_2(n_h), \dots, f_p(n_h)] \quad \text{simultaneously} \\ &\text{subject to} && \sum_{h=1}^L n_h \leq n \\ &\text{and} && 2 \leq n_h \leq N_h; h=1,2,\dots,L \end{aligned} \quad (5.1)$$

Since no algorithm is available to solve a multiobjective programming problem directly the problem is to be converted into a single objective problem by using some compromise criterion.

The solution is obtained by using Fuzzy programming approach to solve the problem (5.1) consists of the following steps:

**Step 1:** To obtain the solution of the multi-objective NLPP (MNLPP), consider the single objective problem using only one objective at a time and ignoring the other objective function and obtained the optimum solution for each characteristic as ideal solution.

**Step 2:** From the result of step-1, determine the corresponding values for every objective at each solution obtained. Let  $(n_{11}^*, \dots, n_{21}^*, \dots, n_{31}^*, \dots, n_{jh}^*)$  are the ideal solutions of the objectives functions  $f_1, f_2, \dots, f_j$ .

So  $U_j = \text{Max}\{f_1(n_{1h}), f_2(n_{2h}), \dots, f_p(n_{ph})\}$  and  $L_j = f_j(n_{jh})$   $j=1,2,K, p, h=1,2,\dots, L$ , where,  $U_j$  and  $L_j$  be the upper and lower bounds of the  $j^{\text{th}}$  objective function  $f_j(n_{jh})$ .

**Step 3:** The membership function for the given problem can be defined as:

$$\mu_j(f_j(n_{jh})) = \begin{cases} 0, & \text{if } f_j(n_{jh}) \geq U_j \\ \frac{U_j(n_{jh}) - f_j(n_{jh})}{U_j(n_{jh}) - L_j(n_{jh})}, & \text{if } L_j \leq f_j(n_{jh}) \leq U_j \\ 1, & \text{if } f_j(n_{jh}) \leq L_j \end{cases}$$

where  $\mu_j(f_j(n_{jh}))$  is a strictly monotonic decreasing function with respect to  $f_j(n_{jh})$ .

Therefore the general aggregation function can be defined as:

$$\mu_D(n_{jh}) = \mu_D\{\mu_1(f_1(n_{1h}), \mu_2 f_2(n_{2h}), K, \mu_j f_j(n_{jh})\}.$$

The fuzzy multi-objective formulation of the problem may be defined as

$$\begin{aligned} &\text{Maximize} && \mu_D(n_{jh}) \\ &\text{subject to} && \sum_{h=1}^L n_{jh} \leq n && (5.2) \\ &\text{and} && 2 \leq n_{jh} \leq N_h \quad j=1,2,K, p \text{ and } h=1,2,K, L. \end{aligned}$$

The problem is to find the optimal value of  $n_{jh}^*$  for this convex fuzzy decision based on addition operator (like Tiwari *et al.* (1987)). Therefore the problem (5.2) is rewritten, according to max-addition operator, as

$$\begin{aligned} &\text{Maximize} && \mu_D(n_{jh}^*) = \sum_{j=1}^p \mu_j(f_j(n_{jh})) = \sum_{j=1}^p \frac{U_j - (f_j(n_{jh}))}{U_j - L_j} \\ &\text{subject to} && \sum_{h=1}^L n_{jh} \leq n && (5.3) \\ &\text{and} && 2 \leq n_{jh} \leq N_h; \quad j=1,2,K, p; \quad h=1,2,K, L. \end{aligned}$$



The above problem (5.3) reduces to

$$\begin{aligned} \text{Maximize} \quad & \mu_D(n_{jh}^*) = \sum_{j=1}^p \mu_j(f_j(n_{jh})) = \sum_{j=1}^p \left\{ \frac{U_j}{U_j - L_j} - \frac{f_j(n_{jh})}{U_j - L_j} \right\} \\ \text{subject to} \quad & \sum_{h=1}^L n_{jh} \leq n \\ \text{and} \quad & 2 \leq n_{jh} \leq N_h; \quad j = 1, 2, K, p; \quad h = 1, 2, K, L. \end{aligned} \tag{5.4}$$

The problem (5.4) will attain its maxima if the function  $F_j(n_{jh}) = \left\{ \frac{f_j(n_{jh})}{U_j - L_j} \right\}$  is to be minimum. Therefore the problem (5.4) reduces into the following primal problem given as

**Model (1)**

$$\begin{aligned} \text{Minimize} \quad & Z = \sum_{j=1}^p F_j(n_{jh}) \\ \text{subject to} \quad & \sum_{h=1}^L n_{jh} \leq n \\ \text{and} \quad & 2 \leq n_{jh} \leq N_h; \quad j = 1, 2, K, p; \quad h = 1, 2, K, L. \end{aligned} \tag{5.5}$$

**Model (2):** A typical fuzzy programming using under and over deviational variables can be expressed as follows:

$$\begin{aligned} \text{Min} \quad & \sum_{j=1}^p (d_j^- + d_j^+) \\ \text{s.t.} \quad & \frac{U_j(n_{jh}) - f_j(n_{jh})}{U_j(n_{jh}) - L_j(n_{jh})} + d_j^- - d_j^+ = 1, \quad j = 1, 2, K, p; \quad h = 1, 2, K, L \\ & \sum_{h=1}^L n_{jh} \leq n \\ & 2 \leq n_{jh} \leq N_h; \quad j = 1, 2, K, p; \quad h = 1, 2, K, L \\ & d_j^-, d_j^+ \geq 0 \text{ \& } d_j^+ d_j^- = 0 \end{aligned} \tag{5.6}$$

Where  $d_j^-, d_j^+ \geq 0$ , with  $d_j^+ d_j^- = 0$  are respectively under and over derivations from target set.

**Model (3):** By introducing an auxiliary variable  $\lambda$ , the model can be reformulated as follows:

$$\begin{aligned} \text{Max} \quad & \lambda \\ \text{s.t.} \quad & \lambda \leq \frac{U_j(n_{jh}) - f_j(n_{jh})}{U_j(n_{jh}) - L_j(n_{jh})}, \quad j = 1, 2, K, p; \quad h = 1, 2, K, L \\ & \sum_{h=1}^L n_{jh} \leq n \\ & 2 \leq n_{jh} \leq N_h; \quad j = 1, 2, K, p; \quad h = 1, 2, K, L \\ & 0 \leq \lambda \leq 1 \end{aligned} \tag{5.7}$$

**Model (4):** By introduce auxiliary variables for each objective as  $\lambda_j$ , the model 3 can be formulated as follows:

$$\begin{aligned}
 & \text{Max } \lambda_j \\
 & \text{s.t.} \\
 & \lambda_j \leq \frac{U_j(n_{jh}) - f_j(n_{jh})}{U_j(n_{jh}) - L_j(n_{jh})}, \quad j = 1, 2, K, p; \quad h = 1, 2, K, L \\
 & \sum_{h=1}^L n_{jh} \leq n \tag{5.8} \\
 & 2 \leq n_{jh} \leq N_h; \quad j = 1, 2, K, p; \quad h = 1, 2, K, L \\
 & 0 \leq \lambda_j \leq 1
 \end{aligned}$$

The common value of  $\lambda$  may be termed a measure of the degree of satisfaction or the degree of compromise ( $0 \leq \lambda \leq 1$ ). If  $\lambda$  is close to 1, there is a high degree of satisfaction (compromise), and if  $\lambda$  is close to 0, there is a low degree of satisfaction.

The NLPP (5.5)-(5.8) may be solved by using a software package for solving constrained optimization problems. The software, developed by LINDO Systems Inc., is user's friendly and does not require much knowledge of computer programming or computer languages. A LINGO User's Guide (2001) is also available for reference.

### 6. Numerical Illustrations

**Example 1:** A simulation study has been carried out to illustrate the computational details of a multivariate population with multiple pooling of stratum variances. Consider a population with five strata ( $L=5$ ) in which three independent characteristics are defined on each unit of the population ( $p=3$ ). It is also assumed that the population of size  $N = 500$  is divided into five strata with stratum sizes  $N_h$  and stratum weights  $W_h$  as

$$\begin{aligned}
 & N_1 = 98, N_2 = 95, N_3 = 110, N_4 = 93 \text{ and } N_5 = 104 \text{ and} \\
 & W_1 = 0.196, W_2 = 0.190, W_3 = 0.220, W_4 = 0.186 \text{ and } W_5 = 0.208.
 \end{aligned}$$

The data for three independent normal populations with the specification of strata means  $\bar{Y}_{jh}$  and the strata standard deviations  $S_{jh}$  given in Tables 6.1 and 6.2 respectively are generated through the website "<http://www.alewand.de/stattabneu/stattab.htm>".

**Table 6.1: Strata means ( $\bar{Y}_{jh}$ )**

Stratum $\rightarrow$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
Characteristics $\downarrow$					
$j = 1$	10	12	08	06	14
$j = 2$	25	28	17	25	30
$j = 3$	45	40	50	38	52

**Table 6.2: Stratum standard deviations ( $S_{jh}$ ) for three characteristics and five strata**

Stratum $\rightarrow$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
Characteristics $\downarrow$					
$j = 1$	25	10	10	35	10
$j = 2$	22	05	15	07	05
$j = 3$	15	15	25	40	25

It should be noted that  $S_{12} = S_{13} = S_{15} = 10$ ,  $S_{22} = S_{25} = 5$ ,  $S_{31} = S_{32} = 15$  and  $S_{33} = S_{35} = 25$ .

In the above situation, we have, for the first characteristic ( $j = 1$ ),  $l_j = 1$  and there is only one group  $G_{11}$  with

$$g_{11} = 3, M_{11} = \{2, 3, 5\} \text{ and } L_1 = 3.$$

For the second characteristic ( $j = 2$ ),  $l_j = 1$  and there is only one group  $G_{21}$  with

$$g_{21} = 2, M_{21} = \{2, 5\} \text{ and } L_2 = 4.$$

For the third characteristic ( $j = 3$ ),  $l_j = 2$  and there are two groups  $G_{31}$  and  $G_{32}$ .

For  $G_{31}$ :  $g_{31} = 2, M_{31} = \{1, 2\}$ ,

and for  $G_{32}$ :  $g_{32} = 2, M_{32} = \{3, 5\}$  and  $L_3 = 3$ .

The preliminary samples sizes  $n'_{jh}$  used to estimate  $S_{jh}$  are given in Table 6.3.

**Table 6.3: Preliminary sample sizes ( $n'_{jh}$ )**

Stratum $\rightarrow$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
Characteristics $\downarrow$					
$j = 1$	24	13	15	33	15
$j = 2$	40	09	30	12	09
$j = 3$	15	14	22	28	21

**Table 6.4: Sample standard deviations ( $s_{jh}$ )**

Stratum $\rightarrow$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
Characteristics $\downarrow$					
$j = 1$	21.1413	8.6060	9.6656	30.1319	8.8185
$j = 2$	22.3244	4.6048	16.9006	5.7777	4.4331
$j = 3$	15.0988	10.0282	23.2814	37.4376	24.8519

The sample data are generated through a computer program using the model

$$y_{jhi} = S_{jh}Z_{hi} + \bar{Y}_{jh}; j=1, 2, K, , p; h=1, 2, K, , L; i=1, 2, K, , N_h,$$

where  $y_{jhi}$  denote the value of the  $i^{th}$  observation in  $h^{th}$  stratum for the  $j^{th}$  characteristic and  $Z_{hi}$  are the values of the randomly selected standard normal variate  $Z$ .

Table 6.4 gives the estimated strata standard deviations  $s_{jh}$ .

For the sake of comparisons the Averaged Neyman allocation for  $n = 100$  using true standard deviations  $S_{jh}$  are worked out and are given in Table 6.5.

**Table 6.5: Averaged Neyman allocation for  $n = 100$  (Using  $S_{jh}$ )**

Stratum $\rightarrow$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
Characteristics $\downarrow$					
$j = 1$	27.8567	10.8016	12.5071	37.0097	11.8249
$j = 2$	39.5451	8.7124	30.2641	11.9406	9.5378
$j = 3$	12.2858	11.9097	22.9837	31.0907	21.7300
Average	26.5626	10.4746	21.9183	26.6803	14.3642
Average (rounded off)	<b>27</b>	<b>10</b>	<b>22</b>	<b>27</b>	<b>14</b>

With the help of available data the sampling variances of the estimates of the population means of the three characteristics (fpc ignored) under Averaged Neyman allocation given in Table 6.5 are obtained as

$$V(\bar{y}_{1st}) = 3.348921164, V(\bar{y}_{2st}) = 1.413934847 \text{ and } V(\bar{y}_{3st}) = 6.488945238 \text{ respectively.}$$

Using the sample standard deviations the Averaged Modified Neyman allocation for  $n = 100$  are given in Table 6.6.

**Table 6.6: Averaged Modified Neyman allocation for  $n = 100$  (Using  $s_{jh}$ )**

Stratum $\rightarrow$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
Characteristics $\downarrow$					
$j = 1$	27.0052	10.6565	13.8584	36.5258	11.9541
$j = 2$	39.9037	7.9789	33.9080	9.8004	8.4091
$j = 3$	13.3792	8.6140	23.1559	31.4812	23.3697
Average	26.7627	9.0831	23.6407	25.9358	14.5776
Average (rounded off)	<b>27</b>	<b>9</b>	<b>24</b>	<b>26</b>	<b>14</b>

Also the sampling variances of the estimates of the population means of the three characteristics (fpc ignored) under the Averaged Modified Neyman allocation given in Table 6.6 are obtained as

$$V(\bar{y}_{1st}) = 2.569840623, \quad V(\bar{y}_{2st}) = 1.475324005 \quad \text{and} \quad V(\bar{y}_{3st}) = 5.594390484$$

respectively.

Assuming that the true strata standard deviations  $S_{jh}$  are unknown but the information about the equality of some of the stratum standard deviations for a particular characteristic are available as stated after Table 6.2.

The pooled standard deviations are worked out as follows.

$$\begin{aligned}
 s_{1pool} &= \left[ \frac{(n'_{12} - 1)s_{12}^2 + (n'_{13} - 1)s_{13}^2 + (n'_{15} - 1)s_{15}^2}{n'_{12} + n'_{13} + n'_{15} - 3} \right]^{1/2} \\
 &= \left[ \frac{(13-1)(8.6060)^2 + (15-1)(9.6656)^2 + (15-1)(8.8185)^2}{13 + 15 + 15 - 3} \right]^{1/2} \\
 &= 9.0629, \\
 s_{2pool} &= \left[ \frac{(n'_{22} - 1)s_{22}^2 + (n'_{25} - 1)s_{25}^2}{n'_{22} + n'_{25} - 2} \right]^{1/2} \\
 &= \left[ \frac{(9-1)(4.6048)^2 + (9-1)(4.4331)^2}{9 + 9 - 2} \right]^{1/2} = 4.5198, \\
 s_{3pool} &= \left[ \frac{(n'_{31} - 1)s_{31}^2 + (n'_{32} - 1)s_{32}^2}{n'_{31} + n'_{32} - 2} \right]^{1/2} \\
 &= \left[ \frac{(15-1)(15.0988)^2 + (14-1)(10.0282)^2}{15 + 14 - 2} \right]^{1/2} = 12.9085, \\
 s_{3pool} &= \left[ \frac{(n'_{33} - 1)s_{33}^2 + (n'_{35} - 1)s_{35}^2}{n'_{33} + n'_{35} - 2} \right]^{1/2}
 \end{aligned}$$

$$= \left[ \frac{(22-1)(23.2814)^2 + (21-1)(24.8519)^2}{22 + 21 - 2} \right]^{1/2} = 24.0603,$$

where  $s_{j\text{pooled}}$  are as defined in (4.2).

For the first characteristic ( $j=1$ ), the Modified Neyman allocation with pooled standard deviations will be the solution of the following NLPP

$$\begin{aligned} \text{Minimize} \quad &= \frac{(0.196)^2(21.1413)^2}{n_1} + \frac{(0.618)^2(9.0629)^2}{m_{11}} \\ &+ \frac{(0.186)^2(30.1319)^2}{n_4} \\ \text{subject to} \quad &n_1 + m_{11} + n_4 \leq 100 \\ \text{and} \quad &2 \leq n_1 \leq 98 \\ &6 \leq m_{11} \leq 309 \\ &2 \leq n_4 \leq 93. \end{aligned} \tag{6.1}$$

The values of  $n_1$ ,  $m_{11}$  and  $n_4$  according to the Modified Neyman allocation by using (1.2) for three strata are

$$n_1 = 26.9963 \approx 27, m_{11} = 36.4899 \approx 36 \quad \text{and} \quad n_4 = 36.5138 \approx 37.$$

These allocations already satisfy the limits of the sample sizes, thus they will solve the NLPP (6.1).

The values of  $n_2$ ,  $n_3$  and  $n_5$  are then reallocated out of  $m_{11}$  to their constituent strata using proportional allocation. This gives

$$n_2 = 11.0680 \approx 11, n_3 = 12.8155 \approx 13 \quad \text{and} \quad n_5 = 12.1165 \approx 12.$$

Thus the optimum allocation with pooled strata variances for the first characteristic is

$$n_{11}^* = 27, n_{12}^* = 11, n_{13}^* = 13, n_{14}^* = 37 \text{ and } n_{15}^* = 12$$

with  $f_1^* = 2.356132272$

Similarly, for second and third characteristics ( $j=2$  and  $3$ ) the optimum allocations with pooled strata variances and the corresponding values of  $f_j^*$  are

$$j=2: n_{21}^* = 40, n_{22}^* = 8, n_{23}^* = 34, n_{24}^* = 10 \text{ and } n_{25}^* = 8$$

with  $f_2^* = 1.202697797$

$$j=3: n_{31}^* = 12, n_{32}^* = 11, n_{33}^* = 24, n_{34}^* = 31 \text{ and } n_{35}^* = 22$$

with  $f_3^* = 4.931664512$

Now the pay-off matrix of the above formulated problems is given as:

$$\begin{aligned}
 f_1(n_{1h}^*) &= 2.56132272 & f_2(n_{1h}^*) &= 1.944178835 & f_3(n_{1h}^*) &= 6.209619712 \\
 f_1(n_{2h}^*) &= 4.458094757 & f_2(n_{2h}^*) &= 1.202697797 & f_3(n_{2h}^*) &= 9.633288575 \\
 f_1(n_{3h}^*) &= 3.028499035 & f_2(n_{3h}^*) &= 2.316987515 & f_3(n_{3h}^*) &= 4.931664512
 \end{aligned}$$

The upper and lower bound of each objective functions can be expressed as:

$$\begin{aligned}
 f_1^u &= 4.458094757 & f_1^l &= 2.356132272 & f_2^u &= 2.316987515 & f_2^l &= 1.202697797 \\
 f_3^u &= 9.633288575 & f_1^l &= 4.931664512
 \end{aligned}$$

Let  $\mu_1(n_{jh}), \mu_2(n_{jh})$  and  $\mu_3(n_{jh})$  be the fuzzy membership function of the objective function  $f_j(n_{jh}), j = 1, 2, 3$  and they are defined as:

$$\begin{aligned}
 \mu_1(n_{1h}) &= \begin{cases} 1 & , \text{if } f_1(n_{1h}) \leq 2.356132272 \\ \frac{4.458094757 - f_1(n_{1h})}{2.101962485} & , \text{if } 2.356132272 \leq f_1(n_{1h}) \leq 4.458094757 \\ 0 & , \text{if } f_1(n_{1h}) \geq 4.458094757 \end{cases} \\
 \mu_2(n_{2h}) &= \begin{cases} 1 & , \text{if } f_2(n_{2h}) \leq 1.202697797 \\ \frac{2.316987515 - f_2(n_{2h})}{1.11428972} & , \text{if } 1.202697797 \leq f_2(n_{2h}) \leq 2.316987515 \\ 0 & , \text{if } f_2(n_{2h}) \geq 2.316987515 \end{cases} \\
 \mu_3(n_{3h}) &= \begin{cases} 1 & , \text{if } f_3(n_{3h}) \leq 4.931664512 \\ \frac{9.633288575 - f_3(n_{3h})}{4.70162406} & , \text{if } 4.931664512 \leq f_3(n_{3h}) \leq 9.633288575 \\ 0 & , \text{if } f_3(n_{3h}) \geq 9.633288575 \end{cases}
 \end{aligned}$$

On applying the **max-addition** operator, the MOSSD problem reduces to the problem as:

$$\begin{aligned}
 &\text{Maximize } \left\{ 6.09371222 - \left( \frac{f_1(n_{1h})}{2.101962485} + \frac{f_2(n_{2h})}{1.11428972} + \frac{f_3(n_{3h})}{4.70162406} \right) \right\} \\
 &\text{subject to } n_1 + n_2 + n_3 + n_4 + n_5 \leq 100 \tag{6.2} \\
 &\text{and } 2 \leq n_h \leq N_h, h=1, 2, K, 5.
 \end{aligned}$$

In order to maximize the above problem, we have to minimize  $\left( \frac{f_1(n_{1h})}{2.101962485} + \frac{f_2(n_{2h})}{1.11428972} + \frac{f_3(n_{3h})}{4.70162406} \right)$  subject to the constraints as described below:

$$\begin{aligned}
 & \text{Minimize } \left\{ \begin{aligned}
 & \frac{(0.196)^2(21.1413)^2}{2.101962485n_1} + \frac{(0.196)^2(22.3244)^2}{1.11428972n_1} + \frac{(0.196)^2(15.0988)^2}{4.70162406n_1} + \\
 & \frac{(0.190)^2(8.6060)^2}{2.101962485n_2} + \frac{(0.190)^2(4.6048)^2}{1.11428972n_2} + \frac{(0.190)^2(10.0282)^2}{4.70162406n_2} + \\
 & \frac{(0.220)^2(9.6656)^2}{2.101962485n_3} + \frac{(0.220)^2(16.9006)^2}{1.11428972n_3} + \frac{(0.220)^2(23.2814)^2}{4.70162406n_3} + \\
 & \frac{(0.186)^2(30.1319)^2}{2.101962485n_4} + \frac{(0.186)^2(5.7777)^2}{1.11428972n_4} + \frac{(0.186)^2(37.4376)^2}{4.70162406n_4} + \\
 & \frac{(0.208)^2(8.8185)^2}{2.101962485n_5} + \frac{(0.208)^2(4.4331)^2}{1.11428972n_5} + \frac{(0.208)^2(24.8519)^2}{4.70162406n_5}
 \end{aligned} \right\} \\
 & \text{subject to } n_1 + n_2 + n_3 + n_4 + n_5 \leq 100 \tag{6.3} \\
 & \text{and } 2 \leq n_h \leq N_h, h=1,2,K,5.
 \end{aligned}$$

Thus we get

$$\begin{aligned}
 & \text{Minimize } Z = \frac{27.21337752}{n_1} + \frac{2.73110869}{n_2} + \frac{20.13751097}{n_3} \\
 & \quad \quad \quad + \frac{26.29319393}{n_4} + \frac{8.046929969}{n_5} \\
 & \text{subject to } n_1 + n_2 + n_3 + n_4 + n_5 \leq 100 \\
 & \text{and } 2 \leq n_h \leq N_h, h=1,2,K,5.
 \end{aligned}$$

Using LINGO software package the solution to MOSSD problem is obtained as

$$n_1 = 26.99968, n_2 = 8.553362, n_3 = 23.22578, n_4 = 26.53927, n_5 = 14.68191$$

Rounding off  $n_h$ ;  $h=1,2,K,5$  to nearest integer values we get the compromise allocation as

$$n_1 = 27, n_2 = 9, n_3 = 23, n_4 = 27 \text{ and } n_5 = 15$$

with variances  $V(\bar{y}_{jst})$ ;  $j=1,2$  and  $3$  under compromise allocation ignoring fpc as

$$V_{1,comp} = 2.51726615, V_{2,comp} = 1.49467451 \text{ and } V_{3,comp} = 5.14560245$$

where,  $V_{j,comp} = V(\bar{y}_{jst})$ ,  $j=1,2$  and  $3$  under compromise allocation ignoring fpc.

**Example: 2** Consider a stratified population with six strata ( $L=6$ ) and three independent characteristics ( $p=3$ ). The data are generated via website “<http://www.alewand.de/stattabneu/stattab.htm>” for three independent normal populations with  $\bar{Y}_{jh}$  and  $S_{jh}$  as given in Table 6.7 and Table 6.8 respectively.



**Table 6.7: Strata Means ( $\bar{Y}_{jh}$ )**

Stratum → Characteristics ↓	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
$j = 1$	11	23	52	20	46	31
$j = 2$	22	42	15	34	36	18
$j = 3$	50	28	30	32	40	20

**Table 6.8: Stratum standard deviations ( $S_{jh}$ ) for three characteristics and six strata**

Stratum → Characteristics ↓	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
$j = 1$	35	10	9.5	15	9.8	21
$j = 2$	19	20	30.3	30	17	24
$j = 3$	32	25	12	40	24	18

It is assumed that the population of size  $N = 600$  is divided into six strata with stratum sizes  $N_h$  and stratum weights  $W_h$  as:

$N_1 = 93, N_2 = 99, N_3 = 105, N_4 = 96, N_5 = 132, N_6 = 75$  and  $W_1 = 0.155, W_2 = 0.165, W_3 = 0.175, W_4 = 0.160, W_5 = 0.220$  and  $W_6 = 0.125$  respectively.

Table 6.8 shows that for  $j=1, S_{12}, S_{13}$  and  $S_{15}$  are approximately equal, for  $j=2, S_{21} \approx S_{22}$  and  $S_{23} \approx S_{24}$  and for  $j=3, S_{32} \approx S_{35}$ . Thus:

For  $j = 1, l_j = 1$ , there is only one group  $G_{11}$  with  $g_{11} = 3, M_{11} = \{ 2, 3, 5 \}$  and  $L_1 = 4$ .

For  $j = 2, l_j = 2$ , there are two groups  $G_{21}$  and  $G_{22}$  with  $g_{21} = 2, M_{21} = \{ 1, 2 \}$  and  $g_{22} = 2, M_{22} = \{ 3, 4 \}$  respectively and  $L_2 = 4$ .

For  $j = 3, l_j = 1$ , there is only one group  $G_{31}$  with  $g_{31} = 2, M_{31} = \{ 2, 5 \}$  and  $L_3 = 5$ .

The preliminary samples sizes  $n'_{jh}$  used to estimate  $S_{jh}$  are given in Table 6.9.

**Table 6.9: Preliminary Sample Sizes ( $n'_{jh}$ )**

Stratum → Characteristics ↓	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
$j = 1$	13	15	13	37	16	26
$j = 2$	18	23	21	16	27	15
$j = 3$	24	19	16	25	14	22

The sample data are generated through a computer program using the same model as Example 1 that is,

$$y_{jhi} = S_{jh} Z_{hi} + \bar{Y}_{jh}; \quad j=1,2,\dots,p, \quad h=1,2,\dots,L \quad \text{and} \quad i=1,2,\dots,N_h$$

where  $y_{jhi}$  denote the value of the  $i^{th}$  observation in  $h^{th}$  stratum for the  $j^{th}$  characteristic and  $Z_{hi}$  are the values of the randomly selected standard normal variate  $Z$ .

The sample values of stratum standard deviations are summarized in Table 6.10.

**Table 6.10: Sample standard deviations ( $s_{jh}$ )**

Stratum $\rightarrow$ Characteristics $\downarrow$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
$j = 1$	15.9673	9.6105	10.2146	13.0483	11.3167	19.6771
$j = 2$	20.9850	20.3869	31.8779	31.8431	13.9237	16.6347
$j = 3$	38.1405	27.5064	10.7248	31.4545	24.7588	22.6623

Using the sample standard deviations the Averaged Modified Neyman allocation for  $n = 120$  are given in Table 6.11.

**Table 6.11: Averaged Modified Neyman allocation for  $n = 120$  (Using  $s_{jh}$ )**

Stratum $\rightarrow$ Characteristics $\downarrow$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
$j = 1$	23.0490	14.7679	16.6474	19.4429	23.1863	22.9065
$j = 2$	17.3997	17.9944	29.8421	27.2544	16.3862	11.1231
$j = 3$	27.6686	21.2416	8.7841	23.5544	25.4931	13.2582
Average	22.7058	18.0013	18.4245	23.4173	21.6885	15.7626
Average (rounded off)	<b>23</b>	<b>18</b>	<b>18</b>	<b>23</b>	<b>22</b>	<b>16</b>

From the available data the sampling variances of the estimates of the population means of the three characteristics (fpc ignored) under the Averaged Modified Neyman allocation are obtained as

$V_1(\bar{y}_{st}) = 1.432900652$ ,  $V_2(\bar{y}_{st}) = 4.642929222$  and  $V_3(\bar{y}_{st}) = 5.810951230$  respectively.

For the sake of comparisons the Averaged Neyman allocation for  $n = 120$  using true standard deviations  $S_{jh}$  are given in Table 6.12.

**Table 6.12: Averaged Neyman allocation for  $n = 120$  (Using  $S_{jh}$ )**

Stratum $\rightarrow$ Characteristics $\downarrow$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
$j = 1$	40.8958	12.4384	12.5326	18.0922	16.2528	19.7883
$j = 2$	15.3070	17.1521	27.5604	24.9486	19.4391	15.5929
$j = 3$	23.6990	19.7093	10.0338	30.5793	25.2280	10.7505
Average	26.6339	16.4333	16.7089	24.5400	20.3066	15.3772
Average (rounded off)	<b>27</b>	<b>16</b>	<b>17</b>	<b>25</b>	<b>20</b>	<b>15</b>

With the help of available data the sampling variances of the estimates of the population means of the three characteristics (fpc ignored) under Averaged Neyman allocation are obtained as

$$V_1(\bar{y}_{st}) = 2.344953919, V_2(\bar{y}_{st}) = 4.876740281 \text{ and } V_3(\bar{y}_{st}) = 5.603878698 \text{ respectively.}$$

Assuming that the true strata standard deviations  $S_{jh}$  are unknown but the information about the equality of some of the standard deviations for a particular characteristic are available as stated after Table 6.8, pooled standard deviations are worked out using (4.2) as

$$s_{1_{pool}} = 10.4370, s_{2_{pool}} = 20.6497, s_{2_{pool_2}} = 31.8630, s_{3_{pool}} = 26.3890.$$

For the characteristic  $j=1$ , the Modified Neyman allocation with pooled standard deviations will be the solution of the following NLPP:

$$\begin{aligned} \text{Minimize} \quad &= \frac{(0.155)^2(15.9673)^2}{n_1} + \frac{(0.560)^2(10.4370)^2}{m_{11}} + \\ &+ \frac{(0.160)^2(13.0483)^2}{n_4} + \frac{(0.125)^2(19.6771)^2}{n_6} \\ \text{subject to} \quad &n_1 + m_{11} + n_4 + n_6 \leq 120 \tag{6.4} \\ &2 \leq n_1 \leq 93 \\ &6 \leq m_{11} \leq 336 \\ \text{and} \quad &2 \leq n_4 \leq 96 \\ &2 \leq n_6 \leq 75 \end{aligned}$$

By using the formula (1.2) the modified Neyman allocation for four strata, are of  $n_1$ ,  $m_{11}$ ,  $n_4$  and  $n_6$  given as:

$$n_1 = 23.0817 \approx 23 \quad m_{11} = 54.5089 \approx 55 \quad n_4 = 19.4705 \approx 19 \quad n_6 = 22.9390 \approx 23$$

These allocations already satisfy the limits of sample sizes, thus they will solve the NLPP (6.4).

The sample size  $m_{11}$  to the combined stratum is reallocated to its constituent strata (2<sup>nd</sup>, 3<sup>rd</sup> and 5<sup>th</sup>) proportionally as:

$$n_2 = 16.2054 \approx 16 \quad n_3 = 17.1875 \approx 17 \quad \text{and} \quad n_5 = 21.6071 \approx 22$$

Thus the optimum values of the sample sizes to different strata for the first characteristic are:

$$n_{11}^* = 23, n_{12}^* = 16, n_{13}^* = 17, n_{14}^* = 19, n_{15}^* = 22 \text{ and } n_{16}^* = 23$$

with  $f_1^* = 1.385622811$

Similarly for the second and third characteristics ( $j=2$  and  $3$ ), the individual optimum values of  $n_{jh}$ ;  $j=2,3$ ,  $h=1,2,\dots,6$  are worked out as:

$$j=2: n_{21}^* = 17, n_{22}^* = 18, n_{23}^* = 30, n_{24}^* = 27, n_{25}^* = 17 \text{ and } n_{26}^* = 11$$

with  $f_2^* = 4.194774309$

$$j=3: n_{31}^* = 27, n_{32}^* = 20, n_{33}^* = 9, n_{34}^* = 24, n_{35}^* = 27 \text{ and } n_{36}^* = 13$$

with  $f_3^* = 5.487212235$

Now the pay-off matrix of the above problems is given below:

$$\begin{aligned} f_1(n_{1h}^*) &= 1.385622811 & f_2(n_{1h}^*) &= 4.978569804 & f_3(n_{1h}^*) &= 6.044702086 \\ f_1(n_{2h}^*) &= 2.389501246 & f_2(n_{2h}^*) &= 4.194774309 & f_3(n_{2h}^*) &= 6.730451175 \\ f_1(n_{3h}^*) &= 1.584181422 & f_2(n_{3h}^*) &= 6.172221739 & f_3(n_{3h}^*) &= 5.487212235 \end{aligned}$$

The upper and lower bound of each objective functions can be expressed as:

$$\begin{aligned} f_1^u &= 2.389501246 & f_1^l &= 1.385622811 & f_2^u &= 6.172221739 & f_2^l &= 4.194774309 \\ f_3^u &= 6.730451175 & f_3^l &= 5.487212235 \end{aligned}$$

Let  $\mu_1(n_{jh})$ ,  $\mu_2(n_{jh})$  and  $\mu_3(n_{jh})$  be the fuzzy membership function of the objective function  $f_j(n_{jh})$ ,  $j = 1, 2, 3$  and they are defined as:

$$\mu_1(n_{1h}) = \begin{cases} 1 & , \text{if } f_1(n_{1h}) \leq 1.385622811 \\ \frac{2.389501246 - f_1(n_{1h})}{1.00387844} & , \text{if } 1.385622811 \leq f_1(n_{1h}) \leq 2.389501246 \\ 0 & , \text{if } f_1(n_{1h}) \geq 2.389501246 \end{cases}$$

$$\mu_2(n_{2h}) = \begin{cases} 1 & , \text{if } f_2(n_{2h}) \leq 4.194774309 \\ \frac{6.172221739 - f_2(n_{2h})}{1.97744743} & , \text{if } 4.194774309 \leq f_2(n_{2h}) \leq 6.172221739 \\ 0 & , \text{if } f_2(n_{2h}) \geq 6.172221739 \end{cases}$$

$$\mu_3(n_{3h}) = \begin{cases} 1 & , \text{if } f_3(n_{3h}) \leq 5.487212235 \\ \frac{6.730451175 - f_3(n_{3h})}{1.24323894} & , \text{if } 5.487212235 \leq f_3(n_{3h}) \leq 6.730451175 \\ 0 & , \text{if } f_3(n_{3h}) \geq 6.730451175 \end{cases}$$

On applying the **max-addition** operator, the MOSSD problem reduces to the problem as:

$$\begin{aligned} \text{Maximize} \quad & \left\{ 10.91521965 - \left( \frac{f_1(n_{1h})}{1.00387844} + \frac{f_2(n_{2h})}{1.97744743} + \frac{f_3(n_{3h})}{1.24323894} \right) \right\} \\ \text{subject to} \quad & n_1 + n_2 + n_3 + n_4 + n_5 + n_6 \leq 120 \quad (6.5) \\ \text{and} \quad & 2 \leq n_h \leq N_h, \quad h=1,2,K,6. \end{aligned}$$

In order to maximize the above problem, we have to minimize  $\left( \frac{f_1(n_{1h})}{1.00387844} + \frac{f_2(n_{2h})}{1.97744743} + \frac{f_3(n_{3h})}{1.24323894} \right)$  subject to the constraints as described below:

$$\begin{aligned} \text{Minimize} \quad & \left\{ \begin{aligned} & \frac{(0.155)^2 (15.9673)^2}{1.00387844 n_1} + \frac{(0.155)^2 (20.9850)^2}{1.97744743 n_1} + \frac{(0.155)^2 (38.1405)^2}{1.24323894 n_1} + \\ & \frac{(0.165)^2 (9.6105)^2}{1.00387844 n_2} + \frac{(0.165)^2 (20.3869)^2}{1.97744743 n_2} + \frac{(0.165)^2 (27.5064)^2}{1.24323894 n_2} + \\ & \frac{(0.175)^2 (10.2146)^2}{1.00387844 n_3} + \frac{(0.175)^2 (31.8779)^2}{1.97744743 n_3} + \frac{(0.175)^2 (10.7248)^2}{1.24323894 n_3} + \\ & \frac{(0.160)^2 (13.0483)^2}{1.00387844 n_4} + \frac{(0.160)^2 (31.8431)^2}{1.97744743 n_4} + \frac{(0.160)^2 (31.4545)^2}{1.24323894 n_4} + \\ & \frac{(0.220)^2 (11.3167)^2}{1.00387844 n_5} + \frac{(0.220)^2 (13.9237)^2}{1.97744743 n_5} + \frac{(0.220)^2 (24.7588)^2}{1.24323894 n_5} + \\ & \frac{(0.125)^2 (19.6771)^2}{1.00387844 n_6} + \frac{(0.125)^2 (16.6347)^2}{1.97744743 n_6} + \frac{(0.125)^2 (22.6623)^2}{1.24323894 n_6} \end{aligned} \right\} \\ \text{subject to} \quad & n_1 + n_2 + n_3 + n_4 + n_5 + n_6 \leq 120, \quad (6.6) \\ \text{and} \quad & 2 \leq n_h \leq N_h, \quad h=1,2,K,6. \end{aligned}$$

On simplifying we get

$$\begin{aligned} \text{Minimize } Z &= \frac{39.56324029}{n_1} + \frac{24.79547140}{n_2} + \frac{21.75439250}{n_3} \\ &\quad + \frac{37.84158531}{n_4} + \frac{34.78404490}{n_5} + \frac{14.66758129}{n_6} \\ \text{subject to } &n_1 + n_2 + n_3 + n_4 + n_5 + n_6 \leq 120, \\ \text{and } &2 \leq n_h \leq N_h, h=1,2,K,6. \end{aligned}$$

Using LINGO software package the solution to MOSSD problem is obtained as

$$n_1 = 23.72606, n_2 = 18.78304, n_3 = 17.59354, n_4 = 23.20408, n_5 = 22.24691, n_6 = 14.44638$$

Rounding off  $n_h$ ;  $h=1,2,K,5$  to nearest integer values we get the compromise allocation as

$$n_1 = 24, n_2 = 19, n_3 = 18, n_4 = 23, n_5 = 22 \text{ and } n_6 = 14$$

with variances  $V(\bar{y}_{jst})$ ;  $j=1,2$  and  $3$  under compromise allocation ignoring fpc as

$$V_{1,comp} = 1.46846786, V_{2,comp} = 4.52928072 \text{ and } V_{3,comp} = 5.75905709$$

where,  $V_{j,comp} = V(\bar{y}_{jst})$ ,  $j=1,2$  and  $3$  under compromise allocation ignoring fpc.

## 7. Conclusion

To validate the proposed compromise allocation, it is compared with some other existing compromise allocations and proportional allocation as well. Tables 7.1 and 7.2, for Examples 1 and 2 respectively, explore the performance of the proposed allocation and other comparative allocations.

The proportional allocation is worked out by

$$n_h = nW_h : h = 1,2,\dots,L, \tag{7.1}$$

and its variance (ignoring fpc) is computed directly using the formula

$$V_{j,prop} = V(\bar{y}_{j,prop}) = \sum_{h=1}^L \frac{W_h^2 S_{jh}^2}{n_h}; j = 1,2,3.$$

The following averaged compromise allocations are selected for comparison.

- (i) Averaged Neyman allocation,
- (ii) Averaged Modified Neyman allocation,
- (iii) Averaged Allocation with Pooled Standard Deviations.

These allocations are averaged over characteristics and rounded off to the nearest integers.

Kozak (2006b) considered five methods for working out the compromise allocation in multivariate stratified surveys. In the fifth method, Kozak minimized the sum of relative increases in the variances due to not using the individual optimum allocations. Thus the problem of allocation may be stated as the following NLPP

$$\begin{aligned} &\text{Minimize } f(n_1, n_2, \dots, n_L) = \sum_{j=1}^p \frac{V_j}{V_j^*} \\ &\text{subject to } \sum_{h=1}^L n_h = n \\ &\text{and } n_h \geq 0; h = 1, 2, \dots, L. \end{aligned} \tag{7.2}$$

Since the true  $S_{jh}$  are assumed to be unknown the objective function may be expressed as  $\sum_{j=1}^p \frac{\hat{V}_j}{\hat{V}_j^*}$ , where  $\hat{V}_j$  and  $\hat{V}_j^*$  are sample estimates of  $V_j$  and  $V_j^*$ .

The solution to the NLPP (7.2), with objective as “Minimize  $\sum_{j=1}^p \frac{\hat{V}_j}{\hat{V}_j^*}$ ” is obtained by using Lagrange multiplier technique after ignoring non-negativity restrictions, is given as

$$n_h = n \frac{W_h b_h}{\sum_{h=1}^L W_h b_h}; h = 1, 2, \dots, L, \tag{7.3}$$

$$\text{where } b_h = \sqrt{\sum_{j=1}^p s_{jh}^2 / \hat{V}_j^*} \text{ and } \hat{V}_j^* = \frac{1}{n} \left( \sum_{h=1}^L W_h s_{jh} \right)^2. \tag{7.4}$$

The allocations given by (7.3) are termed as “Kozak’s allocation” are placed for comparison in Tables 7.1 and 7.2.

The basis of comparison is the ‘TRACE’ (the sum of principal diagonal elements =  $\sum_{j=1}^p V_j(\bar{y}_{st})$ ) of the variance-covariance matrix of the estimator of the  $j^{\text{th}}$  population means  $\bar{Y}_j; j = 1, 2, 3$ . Since all the characteristics are assumed to be independent, the covariances are zero. The relative efficiency (R. E.) of a compromise allocation with respect to the proportional allocation is defined as

$$R.E. = T_{prop} / T_{comp} \quad (\text{Sukhatme et al. (1984)}),$$

where  $T_{prop}$  represents the trace under proportional allocation and  $T_{comp}$  represents the trace under a compromise allocation.

The last columns of Tables 7.1 and 7.2 explain relative efficiency of all allocations, as discussed, with respect to proportional allocation.

Thus it can be concluded about the proposed approach, using fuzzy programming for a specific model as discussed, may be considered as usable compromise criterion to solve the problems of allocation in multivariate surveys.

**Table 7.1: Relative efficiencies of different compromise allocations as compared to the proportional allocation.**

(Example 1)

Allocations	$n_1$	$n_2$	$n_3$	$n_4$	$n_5$	Trace	Relative Efficiency ( $T_{prop}/T_{comp}$ ) with respect to proportional allocation
Proportional	20	19	22	18	21	10.628278360	1.000000000
Compromise Allocations:							
(i) Averaged Neyman	27	10	22	27	14	11.251801250	0.944584616
(ii) Averaged Modified Neyman	27	9	24	26	14	9.639555112	1.102569386
(iii) Averaged Allocation with Pooled Standard Deviations	26	10	24	26	14	9.625212282	1.104212359
(iv) Kozak's Allocation	27	9	23	26	15	9.573005390	1.110234240
(v) Ansari et al. (2011)	22	9	22	29	18	9.374293805	1.133768429
(vi) Proposed Allocation							
a) Method 1	27	9	23	26	15	9.57305	1.110229066
b) Method 2	27	9	23	26	15	9.57305	1.110229066
c) Method 3	28	8	27	23	14	9.878360	1.075915269
d) Method 4	27	9	23	26	15	9.57305	1.110229066

**Table 7.2: Relative efficiencies of different compromise allocations as compared to the proportional allocation.**

(Example 2)

Allocations	$n_1$	$n_2$	$n_3$	$n_4$	$n_5$	$n_6$	Trace	Relative Efficiency ( $T_{prop}/T_{comp}$ ) with respect to proportional allocation
Proportional	19	20	21	19	26	15	12.137564700	1.000000000
Compromise Allocations:								
(i) Averaged Neyman	27	16	17	25	20	15	12.825572900	0.946356533
(ii) Averaged Modified Neyman	23	18	18	23	22	16	11.886781100	1.021097688
(iii) Averaged Allocation with Pooled Standard Deviations	22	18	19	23	22	16	11.878224270	1.021833266
(iv) Kozak's Allocation	22	18	20	23	21	16	11.856593500	1.021973573
(v) Ansari et al. (2011)	23	19	19	24	21	14	11.836881600	1.025402222
(vi) Proposed Allocation								
a) Method 1	24	19	18	23	22	14	11.85681	1.023678772
b) Method 2	24	19	18	23	22	14	11.85681	1.023678772
c) Method 3	24	19	18	24	23	12	11.88560	1.021199157
d) Method 4	24	19	18	23	22	14	11.85681	1.023678772



## **Acknowledgments**

The authors are grateful to the Chief Editor and the learned Referees for their valuable comments. The first author is also grateful to the UGC Start up grant to carry out the research.

## **References**

1. Ahsan, M. J. and Khan, S. U. (1977). Optimum allocation in multivariate stratified random sampling using prior information, *Journal of Indian Statistical Association*, 15, 57-67.
2. Ahsan, M. J. and Khan, S. U. (1982). Optimum allocation in multivariate stratified random sampling with overhead cost, *Metrika*, 29, 71-78.
3. Ahsan, M. J., Najmussehar and Khan, M. G. M. (2005). Mixed allocation in stratified sampling, *Aligarh Journal of Statistics*, 25, 87-97.
4. Ansari, A. H., Najmussehar and Ahsan, M. J. (2009). On multiple response stratified random sampling design, *International Journal of Statistical Sciences*, Kolkata, India, 1 (1), 1-11.
5. Ansari, A. H., Varshney, R., Najmussehar and Ahsan, M. J. (2011). An optimum multivariate-multiobjective stratified sampling design, *Metron*, Vol. LXIX (3), 227 – 250.
6. Aoyama, H. (1963). Stratified random sampling with optimum allocation for multivariate populations, *Annals of the Institute of Statistical Mathematics*, 14, 251-258.
7. Arvanitis, L. G. and Afonja, B. (1971). Use of the generalized variance and the gradient projection method in multivariate stratified random sampling, *Biometrics*, 27, 119-127.
8. Bankier, M. D. (1988). Power allocations: Determining sample sizes for sub national areas, *The American Statistician*, 42, 174-177.
9. Bethel, J. (1989). Sample allocation in multivariate surveys, *Survey Methodology*, 15 (1), 47-57.
10. Chatterjee, S. (1967) A note on optimum allocation, *Scandinavian Actuarial Journal*, 50, 40-44.
11. Chatterjee, S. (1968). Multivariate stratified surveys, *Journal of American Statistical Association*, 63, 530-534.
12. Cochran, W. G. (1977). *Sampling Techniques*, 3<sup>rd</sup> ed., John Wiley and Sons, New York.
13. Dalenius, T. (1957). *Sampling in Sweden: Contributions to the Methods and Theories of Sample Survey Practice*, Almqvist and Wiksell, Stockholm.
14. Díaz-García, J. A. and Cortez, L. U. (2006). Optimum allocation in multivariate stratified sampling: multi-objective programming, *Comunicación Técnica No. I-06-07/28-03-2006 (PE/CIMAT)*.

15. Díaz-García, J. A. and Cortez, L. U. (2008). Multi-objective optimisation for optimum allocation in multivariate stratified sampling, *Survey Methodology*, 34 (2), 215-222.
16. Fatima, U., Varshney, R., Najmussehar and Ahsan, M.J. (2014). On Compromise Mixed Allocation in Multivariate Stratified Sampling with Random Parameters, *Journal of Mathematical Modelling and Algorithms in Operations Research*, Vol. 13(4), 523 – 536.
17. Folks, J. L. and Antle, C. E. (1965). Optimum allocation of sampling units to the strata when there are  $R$  responses of interest, *Journal of the American Statistical Association*, 60, 225-233.
18. Ghosh, S. P. (1958). A note on stratified random sampling with multiple characters, *Calcutta Statistical Association Bulletin*, 8, 81-89.
19. Jahan, N., Khan, M. G. M. and Ahsan, M. J. (1994). A generalized compromise allocation, *Journal of the Indian Statistical Association*, 32, 95-101.
20. Khan, M. G. M., Ahsan, M. J. and Jahan, N. (1997). Compromise allocation in multivariate stratified sampling: an integer solution, *Naval Research Logistics*, 44, 69-79.
21. Khan, M. G. M., Khan, E. A. and Ahsan, M. J. (2003). An optimal multivariate stratified sampling design using dynamic programming, *Australian & New Zealand Journal of Statistics*, 45 (1), 107-113.
22. Kokan, A. R. and Khan, S. (1967). Optimum allocation in multivariate surveys: An analytical solution, *Journal of the Royal Statistical Society, Series B*, 29 (1), 115-125.
23. Kozak, M. (2006a). Multivariate sample allocation: application of random search method, *Statistics in Transition*, 7 (4), 889-900.
24. Kozak, M. (2006b). On sample allocation in multivariate surveys, *Communications in statistics-Simulation and Computation*, 35, 901-910.
25. Kreienbrock, L. (1993). Generalized measures of dispersion to solve the allocation problem in multivariate stratified random sampling, *Communications in Statistics-Theory and Methods*, 22 (1), 219-239.
26. Lingo User's Guide (2001). *LINGO-User's Guide*, Published by LINDO SYSTEM INC., 1415, North Dayton Street, Chicago, Illinois, 60622, USA.
27. Mahalanobis, P. C. (1944). On large-scale sample surveys, *Philosophical Transactions of the Royal Society, Series B*, 231, 329-451.
28. Melaku, A. and Sadasivan, G. (1987).  $L_1$ -norm and other methods for sample allocation in multivariate stratified surveys, *Computational Statistics & Data Analysis*, 5 (4), 415-423.
29. Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection, *Journal of Royal Statistical Society*, 97 (4), 558-625.

30. Park, H., NA, S. and Jeon, J. (2007). Compromise allocation in univariate stratified sampling, *Communications in Statistics-Theory and Methods*, 36, 265-271.
31. Stuart, A. (1954). A simple presentation of optimum sampling results, *Journal of Royal Statistical Society, Series B*, 16 (2), 239-241.
32. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Asok, C. (1984). *Sampling Theory of Surveys with Applications*, 3<sup>rd</sup> ed., Iowa State University Press, Ames, Iowa and Indian Society of Agricultural Statistics, New Delhi.
33. Tiwari, R.N., Dharman, S. and Rao, J.R. (1987). Fuzzy Goal Programming - An Additive Model, *Fuzzy Sets and Systems*, 24, 27-34.
34. Varshney, R., Khan, M.G.M., Fatima, U. and Ahsan, M.J. (2015). Integer Compromise Allocation in Multivariate Stratified Surveys, *Annals of Operations Research*, Vol. 226 (1), 659 – 668.
35. Varshney, R., Najmussehar and Ahsan, M. J. (2012). An optimum multivariate stratified double sampling design in presence of non-response, *Optimization Letters*, Vol. 6(5), 993-1008.
36. Yates, F. (1960). *Sampling methods for Censuses and Surveys*, 3<sup>rd</sup> ed., Charles Griffin and Co., London.