CSIRO PUBLISHING The South Pacific Journal of Natural and Applied Sciences, **30**, 65-71, 2012
www.publish.csiro.au/journals/spjnas 10.1071/SP12007

The problem of optimum allocation in multivariate stratified sampling

Ummatul Fatima, Shazia Ghufran and M.J. Ahsan

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

Abstract

Generally, sample surveys are multivariate in nature where multiple response are obtained on every unit selected in a sample, that is, more than one characteristics are defined on each and every unit of the population. While dealing with a multivariate stratified population, to workout an allocation that is optimum for all characteristics is almost impossible unless the characteristics are highly correlated. Some compromise must be allowed to obtain an allocation that is optimum, in some sense, for all the characteristics. Since such allocations are based on some compromise criteria they are known as compromise allocations. This paper deals with the problem of obtaining an optimum allocation in multivariate stratified sampling design.

Keywords: *Multivariate stratified sampling, Optimum allocation, Compromise allocation, Bonferroni inequality*

1. Introduction

Stratified sampling design is the most widely used sampling design for obvious reasons. The problem of obtaining optimum allocation of sample sizes to various strata with the aim to minimize the variance of the estimate of the population parameter under study for a fixed cost of survey or to minimize the cost of the survey for a given precision of the estimate for univariate stratified population is well known in the sampling literature due to the earlier works of Tschuprow (1923) and Neyman (1934).

Generally, in sample surveys on every unit of the population many characteristics are to be studied simultaneously. For every sampled unit the sampler gets a p-component response vector. If the characteristics are highly correlated then the individual optimum allocations may vary little from characteristic to characteristic. In such a situation Cochran (1977) suggested the use of the characteristic-wise average of the individual optimum allocations as a common allocation for all characteristics. For uncorrelated characteristics the individual optimum allocations may vary widely and there may be no obvious compromise. Therefore the sampler must use a compromise criterion to work out a common allocation for all characteristics. Such an allocation is called a compromise allocation.

Yates (1960) suggested two compromise criteria that are widely in use. For estimating population means or totals the first criterion suggests to obtain a compromise allocation that minimizes the weighted sum of the sampling variances of the estimators for fixed cost of the survey. The second criterion suggest to minimize the cost of the survey when the sampling variances of the estimators are subjected to pre-specified tolerance limits.

The problem of compromise allocation has been discussed by several authors namely Dalenius (1953, 1957), Ghosh (1958), Aoyama (1963), Chatterjee (1967, 1968), Kokan and Khan (1967), Huddleston *et al*. (1970), Ahsan and Khan (1977, 1982), Cochran (1977), Omule (1985), Bethel (1985, 1989), Khan *et al*. (1997), Jahan *et al*. (2001), Kozak (2004, 2006), Semiz (2004), Díaz-García and Cortez (2006), Miller *et al*. (2007), Khan *et al*. (2008), Ansari *et al*. (2009), Khan *et al*. (2010), Khowaja *et al*. (2011), Varshney *et al*. (2012) etc., are few of the many researchers who have worked on compromise allocation under different situations.

In the present paper two approaches to work out optimum allocation in multivariate stratified sample surveys are discussed.

2. First Approach

The sample size 'n' sample survey for estimating the population mean $\bar{Y}$ is usually given as

$$n \cong \frac{Z_{\alpha/2}^{2}\sigma^{2}}{d^{2}} \tag{1}$$

where $Z_{\alpha/2}$ is the value of the standard normal variate which cuts off an area $\alpha/2$ at the tails, σ^2 is the population variance and d is the pre-fixed margin of error in the estimated mean that is $\left|\bar{y}-\bar{Y}\right| \le d$. If σ^2 is unknown two-stage methods (Cochran (1977), Gillet (1989)) based on Stien (1945) may be used to estimate σ^2.

Let for a multivariate stratified sample survey with L strata and p independent characteristics the estimation of p overall population means $\bar{Y}_j$; j =1, 2, …, p is of interest.

Further, let the total cost available be C out of which C_j units be allocated to the j^{th} characteristic (See Jahan *et al*. (1994)). In the notations of Cochran (1977) the optimum allocation that minimizes the sampling variance of the j^{th} stratified sample mean $\bar{y}_{jst}$ that is,

$$V(\bar{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_{jh}} \quad \text{(fpc ignored)} \tag{2}$$

for fixed cost

$$C_j = c_{j0} + \sum_{h=1}^{L} c_{jh}\, n_{jh} \tag{3}$$

is given by

$$n_{jh}^* = n \frac{\left(W_h S_{jh} \big/ \sqrt{c_{jh}} \right)}{\sum \left(W_h S_{jh} \big/ \sqrt{c_{jh}} \right)}; \; j=1,2,..., p; h=1,2,..., L. \tag{4}$$

where n denotes the total sample size for fixed cost C_j and is given by

$$n = \frac{(C_j - c_{j0}) \sum \left(W_h S_{jh} \big/ \sqrt{c_{jh}} \right)}{\sum W_h S_{jh} \sqrt{c_{jh}}} \tag{5}$$

c_{j0} and c_{jh}; $j = 1,2,..., p; h=1,2,..., L;$ denote the overhead cost and per unit measurement cost in h^{th} stratum for j^{th} characteristics respectively.

Formula (4) gives the $p \times L$ matrix of optimum allocations for '*p*' characteristics and '*L*' strata as

$$((n_{jh}^*)); h = 1,2,..., L; \; j = 1,2,..., p \tag{6}$$

Let M_h; $h=1,2,..., L$ denote the column maxima of $((n_{jh}^*))$, that is,

$$M_h = \underset{j}{Maximum}\, n_{jh}^*; \; h=1,2,..., L \tag{7}$$

It is suggested that by means of any randomizing device, L sets of M_h; $h=1,2,..., L$ random numbers are generated and recorded keeping the order of their generation unaltered. Since random numbers are generated independently, out of M_h random numbers the units corresponding to first n_{jh}^* random numbers will provide us the optimum allocation for the h^{th}; $h=1,2,..., L$ stratum for the j^{th}; $h=1,2,..., p$ characteristic. The j^{th} characteristic will thus be measured on these selected $(n_{j1}^*, n_{j2}^*, ...\, n_{jh}^*)$ sampled units. Since the cost of generating the random numbers is almost negligible the allocation will remain optimum for the fixed cost C_j; $j=1,2,..., p.$

The above procedure ensures that each characteristic is subjected to its individual optimum allocation.

The numerical example given in the next section will illustrate the numerical details.

3. Application of the First Approach

Jessen (1942) provided a multivariate data from a farm survey related to dairy farming in the state of Iowa, USA. The state is divided into five geographical regions (strata) and three characteristics are measured (Cows Milked, Gallons of Milk, Receipts of Diary Products) on each selected unit (farm) of the sample. The individual optimum allocations for a total sample size of 1000 farms is given in Table 1. Since the selections are independent and the order of selection is retained, out of M_h; $h=1,2,3,4,5$ the first n_{jh}^* random numbers will give the required optimum number of units of the sampling frame of h^{th} stratum on which the j^{th} characteristic is to be measured. The values of $M_h; h = 1,2,...,5$ as defined in (7) are recorded in column 5 of Table 1.

This simple procedure ensures the use of individual optimum allocations of various characteristics without any loss in precision of the estimates. Assuming a $10\,\%$ sample the strata sizes N_h may be taken as 1970, 1920, 2190, 1840, and 2080 for $h = 1,2,3,4,5$, respectively.

According to the proposed procedure five sets of 258, 246, 203, 145 and 228 four digit random numbers are selected out of random numbers 0000 to 9999 without repetition by some random number generating device and applying a method to minimize the number of rejections of the selected random numbers. The corresponding units of the strata are marked according to their order of selection. Now out of the 258 selected units of the first stratum, the first characteristic is measured for the first 254 units, the second characteristic is measured for all the 258 units and the third characteristic is measured for the first 236 units. The same procedure is repeated for the remaining four strata. Obviously, since individual optimum allocations are used for all the characteristics, comparison with other methods of obtaining compromise allocation to prove the

Table 1. Sample Sizes within strata ($n = 1000$).

Stratum No. h	Allocation			Maximum M_h
	Optimum for			
(1)	Cows (2)	Gallons (3)	Receipts (4)	$(5) = Max\{(2),(3),(4)\}$
1	254	258	236	258
2	182	209	246	246
3	203	171	194	203
4	145	134	115	145
5	216	228	209	228

superiority of the discussed procedure is quite unnecessary.

4. Second Approach

Consider a multivariate stratified population with L strata. Let p -characteristics be defined on each unit of the population and the estimation of the p -population means be of interest.

Denote by

$$((n_{jh})); h = 1, 2, ..., L; \ j = 1, 2, ..., p$$

the $p \times L$ matrix of sample sizes. Let the j^{th} characteristic y_j have an unknown mean μ_j and known variance σ_j^2. If σ_j^2 is unknown, it is assumed that its estimate is available. Further, let $\mu' = (\mu_1, \mu_2, ..., \mu_j, ..., \mu_p)$ denote the p -component row vector of the unknown population means. The stratified sample mean

$\bar{y}_{jst} = \sum_{h=1}^{L} W_h \bar{y}_{jh}$ will give an unbiased estimate of

μ_j with a sampling variance

$$V(\bar{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_{jh}} \text{ (fpc ignored)} \quad (8)$$

where S_{jh}^2 denote the usual stratum variance for h^{th} stratum and j^{th} characteristic and $\bar{y}_{jh}$ is the sample mean from h^{th} stratum for j^{th} characteristic.

If d_j is the prefixed permissible margin of error for estimating μ_j such that $P\left(\left|\bar{y}_{jst} - \mu_j\right| \geq d_j\right) \leq \alpha_j$ then the at least $(1-\alpha_j) \times 100\%$ confidence interval A_j for μ_j is given by

$$A_j = (\bar{y}_{jst} - d_j, \bar{y}_{jst} + d_j); \ j = 1, 2, ..., p.$$

If $P(A_j)$ denote the probability that μ_j lies in the interval A_j then we have $P(A_j) \leq \alpha_j$ or equivalently $P(A_j) \geq (1-\alpha'); \ j = 1, 2, ..., p$. The overall confidence level is then given by $P\left(\bigcap_{j=1}^{p} A_j\right)$.

Assuming equal individual confidence levels for the p -population means that is, $\alpha_j = \alpha'$; for all j, and fixing the overall confidence level at $(1-\alpha) \times 100\%$ we get

$$P\left(\bigcap_{j=1}^{p} A_j\right) = 1 - \alpha .$$

By Bonferroni's inequality (Galambos (1977))

$$P\left(\bigcap_{j=1}^{p} A_j\right) \geq \sum_{j=1}^{p} P(A_j) - (p-1) . \quad (9)$$

Substituting the value of $P\left(\bigcap_{j=1}^{p} A_j\right) = 1 - \alpha$ and $P(A_j) \geq (1-\alpha')$ in the above inequality we get

$$1-\alpha \geq \sum_{j=1}^{p}(1-\alpha') - p + 1$$

or $$\alpha' \geq \frac{\alpha}{p} \tag{10}$$

In the above scenario using the Cui and Zhu (2007) approach the total sample size

$$n_j = \sum_{h=1}^{L} n_{jh};\ j = 1, 2, ..., p \tag{11}$$

for estimating the j^{th} population mean μ_j with a confidence level of α' the total sample size is given by

$$n_j = \frac{Z^2_{\alpha'/2}\,\sigma^2_j}{d^2_j};\ j = 1, 2, ..., p\,, \tag{12}$$

where σ^2_j is the overall population variance for the j^{th} characteristic. If the values of σ^2_j are not known they can be estimated from a pilot survey.
Define

$$n_{\max} = \max_j n_j;\ j = 1, 2, ..., p\,. \tag{13}$$

If the overall total sample size is taken as $n_{\max}$ given by (13) for all j then the actual probability of A_j, $P_{actual}(A_j);\ j = 1, 2, ..., p$ (say) will follow the inequality

$$P_{actual}(A_j) \geq 1-\alpha' \tag{14}$$

Equations (9) and (14) implies

$$P\left(\bigcap_{j=1}^{p} A_j\right) \geq \sum_{j=1}^{p} P_{actual}(A_j) - (p-1)$$

or

$$P\left(\bigcap_{j=1}^{p} A_j\right) \geq p(1-\alpha') - (p-1)$$

$$= 1 - p\alpha' \tag{15}$$

Equations (10) and (15) implies

$$P\left(\bigcap_{j=1}^{p} A_j\right) \geq 1-\alpha\,. \tag{16}$$

Thus the overall confidence level of at least $(1-\alpha)\times 100\,\%$ is guaranteed if the total sample size is taken as $n_{\max}$ given by (13).

After fixing the total sample size as $n_{\max}$ the problem remains to find the sample size allocations $n_h;\ h = 1, 2, ..., L$ to various strata such that $\sum_{h=1}^{L} n_h = n_{\max}$. Note that unlike Procedure 1, in this approach all the p-characteristics are to be measured on each of the n_h units selected from the $h^{th};\ h = 1, 2, ..., L$ stratum. Such type of allocations are called compromise allocations in sampling literature (See Cochran (1977)). Several compromise criterion are available to work out a common allocation for all the p-characteristics.

In this manuscript the multivariate allocation problem is formulated as a problem of solving a system of $(p+1)$ simultaneous equations in L variables $n_h;\ h = 1, 2, ..., L$. Out of the $p+1$ equations one equation is linear and the remaining p-equations are nonlinear. Assume that v_j denotes the tolerance limit on the sampling variance of the stratified sample mean $\bar{y}_{jst}$. Now consider the following problem: Find

$$n_h \geq 0;\ h = 1, 2, ..., L \tag{17}$$

satisfying the system of equations

$$V(\bar{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_{jh}} = v_j;\ j = 1, 2, ..., p \tag{18}$$

$$\sum_{h=1}^{L} n_h = n_{\max}\,. \tag{19}$$

If the system (18)-(19) with restrictions (17) is consistent, its solution will provide a compromise allocation that estimates
(i) the p population means with at least $(1-\alpha)\times 100\,\%$ confidence that the error

in the estimate $\bar{y}_{jst}$ will not exceed the limit $d_j ; j = 1, 2, ..., p$

and

(ii) with the required precisions of the estimates.

4.1 The Solution Procedure

System of equations (18) and (19) with restrictions (17) may be solved by converting it into an equivalent Nonlinear Programming Problem (NLPP) using artificial variables.

Consider the following Nonlinear Problem 1 (NLP1):

$$Minimize \ \sum_{j=1}^{p+1} a_j \tag{20}$$

$$Subject \ to \ \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{n_h} + a_j = v_j ; \ j = 1, 2, ..., p \tag{21}$$

$$\sum_{h=1}^{L} n_h + a_{p+1} = n_{\max} \tag{22}$$

$$n_h, a_j \geq 0, h = 1, 2, ..., L \quad j = 1, 2, ..., p+1 \tag{23}$$

where $a_j \geq 0; j = 1, 2, ..., p+1$ are artificial variables.

The zero optimal objective value of the NLP1 implies that all the artificial variables $a_j = 0; j = 1, 2, ..., p+1$ and the optimal solution to the NLP1 will solve the system of equations (18)-(19) with restrictions (17). On the other hand a nonzero optimal value of the objective function implies that some of the artificial variables are > 0. This in turn implies that the system (17)-(19), in its present form, is inconsistent. In this case relaxations in some of the tolerance limits v_j or in the total sample size $n_{\max}$ or in both of these, as the case may be, are required to reach at an optimal solution to NLP1 which NLP1 can be solved by any suitable nonlinear programming technique.

However, the authors used the optimization software LINGO (2001) to solve NLP1 as illustrated through the numerical example in next section.

5. Application of the Second Approach

The following data are from Chatterjee (1968). Table 2 shows the stratum weights W_h, stratum standard deviations S_{jh} for four characteristics under study. The tolerance limits v_j for, the sampling variances of the four estimates of the population means $\mu_j ; j = 1, 2, 3, 4$ are assumed by authors as 3, 20, 5 and 25 respectively. It is further assumed that the margins of permissible error d_j and the population standard deviations σ_j for $j = 1, 2, 3, 4$ are:

$$d_1 = 3 \ \ d_2 = 9 \ \ d_3 = 4 \ \ d_4 = 10$$
$$\sigma_1 = 40 \ \ \sigma_2 = 125 \ \ \sigma_3 = 55 \ \ \sigma_4 = 160$$

respectively.

The overall confidence level $(1-\alpha)\times 100\%$ is fixed at $\alpha = 0.1$. Thus by (10)

$$\alpha' = \alpha_j \geq \frac{0.1}{4} = 0.025 \Rightarrow \frac{\alpha'}{2} = 0.0125.$$

From the area table of standard normal curve $Z_{0.0125} = 2.24$.

Formula (12), with the above values gives the characterwise total sample size n_j for $j = 1, 2, 3, 4$ as

$$n_1 = 892, \ n_2 = 968, \ n_3 = 949, \ n_4 = 1285.$$

By (13)

$$n_{\max} = Maximum \{892, \ 968, \ 949, \ 1285\} = 1285.$$

For the data given in Table 2 the values of $W_h^2 S_{jh}^2$ are computed and recorded in Table 3. Substitution of the computed values in (18)-(19) we get the following system of equations

$$\frac{49.00}{n_1} + \frac{58.98}{n_2} + \frac{45.16}{n_3} + \frac{18.66}{n_4} + \frac{87.98}{n_5} = 3.0 \tag{24}$$

$$\frac{2652.25}{n_1} + \frac{1811.35}{n_2} + \frac{101.61}{n_3} + \frac{8.76}{n_4} + \frac{1.59}{n_5} = 20.0 \tag{25}$$

$$\frac{90.25}{n_1} + \frac{69.22}{n_2} + \frac{85.38}{n_3} + \frac{38.94}{n_4} + \frac{113.21}{n_5} = 5.0 \tag{26}$$

$$\frac{900.00}{n_1} + \frac{3466.85}{n_2} + \frac{1319.87}{n_3} + \frac{54.17}{n_4} + \frac{268.30}{n_5} = 25.0 \tag{27}$$

$$n_1 + n_2 + n_3 + n_4 + n_5 + a_5 = 1285 \tag{28}$$

$$with \quad n_h \geq 0; h = 1, 2, 3, 4, 5 \tag{29}$$

Table 2. Stratum Weight and Stratum standard Deviations for five strata and four characteristics.

h	W_h	S_{1h}	S_{2h}	S_{3h}	S_{4h}
1	0.25	28	206	38	120
2	0.32	24	133	26	184
3	0.21	32	48	44	173
4	0.08	54	37	78	92
5	0.14	67	9	76	117

Table 3. Values of $W_h^2 S_{jh}^2$.

h	$W_h^2 S_{1h}^2$	$W_h^2 S_{2h}^2$	$W_h^2 S_{3h}^2$	$W_h^2 S_{4h}^2$
1	49.00	2652.25	90.25	900.00
2	58.98	1811.35	69.22	3466.85
3	45.16	101.61	85.38	1319.87
4	18.66	8.76	38.94	54.17
5	87.98	1.59	113.21	268.30

Introducing $a_j \geq 0;\ j = 1, 2, 3, 4, 5$ as artificial variables we can have an instance of NLP1 as NLP2:

$$\text{Minimize } \sum_{j=1}^{5} a_j \tag{30}$$

$$\text{Subject to } \frac{49.00}{n_1}+\frac{58.98}{n_2}+\frac{45.16}{n_3}+\frac{18.66}{n_4}+\frac{87.98}{n_5}+a_1=3.0 \tag{31}$$

$$\frac{2652.25}{n_1}+\frac{1811.35}{n_2}+\frac{101.61}{n_3}+\frac{8.76}{n_4}+\frac{1.59}{n_5}+a_2=20.0 \tag{32}$$

$$\frac{90.25}{n_1}+\frac{69.22}{n_2}+\frac{85.38}{n_3}+\frac{38.94}{n_4}+\frac{113.21}{n_5}+a_3=5.0 \tag{33}$$

$$\frac{900.00}{n_1}+\frac{3466.85}{n_2}+\frac{1319.87}{n_3}+\frac{54.17}{n_4}+\frac{268.30}{n_5}+a_4=25.0 \tag{34}$$

$$n_1+n_2+n_3+n_4+n_5+a_5=1285 \tag{35}$$

$$\left.\begin{array}{l} n_h \geq 0; h=1,2,3,4,5 \\ a_j \geq 0; j=1,2,3,4,5 \end{array}\right\} \tag{36}$$

The optimization software LINGO gives the optimal solution to the NLP2 as:

$$a_1^* = a_2^* = a_3^* = a_4^* = a_5^* = 0$$

$$n_1^* = 157.7458 \cong 158, n_2^* = 885.6352 \cong 886, n_3^* = 157.0229 \cong 157,$$
$$n_4^* = 18.6344 \cong 19 \text{ and } n_5^* = 65.9615 \cong 66.$$

Since all the artificial variable are zero the above values of $n_h^*; h = 1, 2, 3, 4, 5$ will provide a solution to the system of equation (24)-(28) and hence the required common or compromise allocation.

6. Discussion

This paper presented two approaches to work out sample size allocations for a multivariate stratified sample survey for estimating the overall population means. In the first approach a simple method is described in which the individual optimum allocations may be used. The second approach uses 'Bonferroni's Inequality' from probability theory to work out a compromise allocation that can be used for all the characteristics such that the overall confidence level of at least $(1-\alpha)\times 100\%$ is maintained for fixed margins of errors in the estimates of the population means of the various characteristics. This total sample size is then divided among the various strata such that the variances of the estimates of the p -population means will remain within their prefixed tolerance limits.

The second approach may be called 'Biobjective Approach', because we achieve two objectives simultaneously. The required confidence levels of the estimates are attained and their individual precision requirements are met.

Acknowledgement

The authors are grateful to the Editor and the referee for their valuable comments, suggestions and corrections that helped to improve the manuscript to its present form.

References

Ahsan, M.J. and Khan, S.U. 1977. Optimum allocation in multivariate stratified random sampling using prior information. *Journal of the Indian Statistical Association* **15**, 57-67.

Ahsan, M.J. and Khan, S.U. 1982. Optimum allocation in multivariate stratified random sampling with overhead cost. *Metrika* **29**, 71-78.

Ansari, A.H., Najmussehar and Ahsan, M.J. 2009. On multiple response stratified random sampling design. *International Journal of Statistical Sciences, Kolkata* **1**, 45-54.

Aoyama, H. 1963. Stratified random sampling with optimum allocation for multivariate populations. *Annals of the Institute of Statistical Mathematics* **14**, 251-258.

Bethel, J. 1985. An optimum allocation algorithm for multivariate surveys. *Proceedings of the Survey Research Methods section*. American Statistical Association, 209-212.

Bethel, J. 1989. Sample allocation in multivariate surveys. *Survey Methodology* **15**, 47-57.

Chatterjee, S. 1967. A note on optimum allocation. *Scandinavian Actuarial Journal* **50**, 40-44.

Chatterjee, S. 1968. Multivariate stratified surveys. *Journal of American Statistical Association* **63**, 530-534.

Cochran, W.G. 1977. *Sampling Techniques*. 3rd ed. New York: John Wiley.

Cui, W. and Zhu, F. 2007. A note on sample size determination for the estimation of the mean vector of a multivariate population. *Communications in Statistics-Theory and Methods* **36**, 1607-1610.

Dalenius, T. 1953. The multivariate sampling problem. *Skandinavisk Actuarietidskrift* **36**, 92-102.

Dalenius, T. 1957. *Sampling in Sweden*. Contibutions to the Methods and Theories of Sample Survey Practice, Stockholm: Almqvist and Wicksell.

Díaz-García, J.A. and Cortez, L.U. 2006. Optimum allocation in multivariate stratified sampling: multi-objective programming. *Comunicacion Tecnica No. I-06-07/28-03-2006 (PE/CIMAT)*, Mexico.

Galambos, J. 1977. Bonferroni Inequalities. *The Annals of Probability* **5**, 577-581.

Ghosh, S.P. 1958. A note on stratified random sampling with multiple characters. *Calcutta Statistical Association Bulletin* **8**, 81-89.

Gillet, R. 1989. Confidence interval construction by Stein's method: a practical and economical approach to sample size determination. *Journal of Market Research* **26**, 237-240.

Huddleston , H.F., Claypool, P.L. and Hocking, R.R. 1970. Optimum sample allocation to strata using convex programming. *Applied Statistics* **19**, 273-278.

Jahan, N., Khan, M.G.M. and Ahsan, M.J. 1994. A generalized compromise allocation. *Journal of the Indian Statistical Association* **32**, 95-101.

Jahan, N., Khan, M.G.M. and Ahsan, M.J. 2001. Optimum compromise allocation using dynamic programming. *Dhaka University Journal of Science* **49**, 197-202.

Jessen, R.J. 1942. *Statistical Investigation of A Sample Survey for Obtaining Farm Facts*. Research Bulletin 304, Iowa State College of Agriculture and Mechanic Arts.

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.

Khan, M.G.M., Khan, E.A. and Ahsan, M.J. 2008. Optimum allocation in multivariate stratified sampling in presence of non-response. *Journal of the Indian Society of Agricultural Statistics* **62**, 42-48.

Khan, M.G.M., Maiti, T. and Ahsan, M.J. 2010. An optimal multivariate stratified sampling design using auxiliary information: an integer solution using goal programming approach. *Journal of Official Statistics* **26**, 695-708.

Khowaja, S., Ghufran, S. ansd Ahsan, M.J. 2011. Estimation of population means in multivariate stratified random sampling. *Communication in Statistics-Simulation and Computation* **40**, 710-718.

Kokan, A.R. and Khan, S.U. 1967. Optimum allocation in multivariate surveys: An analytical solution. *Journal of Royal Statistical Society* **B29**, 115-125.

Kozak, M. 2004. Multivariate sample allocation problem in two schemes of two-stage sampling. *Statistics in Transition* **6**, 1047-1054.

Kozak, M. 2006. On sample allocation in multivariate surveys. *Communication in Statistics-Simulation and Computation* **35**, 901-910.

LINGO 2001. *LINGO-User's Guide*. Lindo System Inc., 1415, North Dayton Street, Chicago, Illinois, 60622, USA.

Miller, T.J., Skalski, J.R. and Lanelli, J.N. 2007. Optimizing a stratified sampling design when faced with multiple objectives-ICES. *Journal of Marine Science* **64**, 97-109.

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 the Royal Statistical Society* **97**, 558-625.

Omule, S.A.Y. 1985. Optimum design in multivariate stratified sampling. *Biometrical Journal* **27**, 907-912.

Semiz, M. 2004. Determination of compromise integer strata sample sizes using goal programming. *Hacettepe Journal of Mathematics and Statistics* **33**, 91-96.

Stein, C. 1945. A two-sample test for a linear hypothesis whose power is independent of the variance. *Annals of Mathematical Statistics* **16**, 243-258.

Tschruprow, A.A. 1923. On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron* **2**, 461-493.

Varshney, R., Najmussehar and Ahsan, M.J. 2012. An optimum multivariate stratified double sampling design in presence of non-response. *Optimization Letters* **6**, 993-1008.

Yates, F. 1960. *Sampling Methods for Censuses and Surveys*. Charles Griffin and Co., London, 3rd ed.

Correspondence to: Ummatul Fatima
E-mail: fatimau2011@yahoo.com