Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 65(3) 2011 297-304

Minimum Variance Optimal Controlled Nearest Proportional to Size Sampling Scheme using Multiple Objective Functions

Neeraj Tiwari[1] and U.C. Sud*[2]

[1]*Kumaun University, S.S.J. Campus, Almora*

[2]*Indian Agricultural Statistics Research Institute, New Delhi*

Received 10 July 2009; Revised 31 January 2011; Accepted 24 March 2011

SUMMARY

The optimal controlled nearest proportional to size sampling scheme suggested by Tiwari *et al*. (2007) uses only one objective function based programming approach for solving the controlled selection problems. In this article we apply the concept of multiple objective functions on optimal controlled nearest proportional to size sampling scheme to minimize the sampling variance of the Yates-Grundy form of the Horvitz-Thompson estimator. The proposed procedure minimizes the true sampling variance of the Horvitz-Thompson estimator while assigning zero probabilities to non-preferred samples. Empirical illustrations have been used to show that the true sample variance of the proposed procedure compares favorably with that of the existing optimal controlled and uncontrolled high entropy selection procedures.

Keywords: Controlled selection, Non-preferred samples, Quadratic programming, Variance estimation, High entropy variance, Multiple objective problem.

1. INTRODUCTION

In sample surveys, problems often arise when we do not allow certain samples because of cost limitations or other practical considerations such as organizational or other difficulties. Such samples are termed as non-preferred samples. Goodman and Kish (1950) introduced controlled selection as a method of sample selection that increases the probability of selecting a preferred sample and thereby decreases the selection probability of non-preferred samples, while conforming strictly to the requirements of probability sampling. Controlled selection has received considerable attention in recent years due to its practical importance. While Sande (1984), Causey *et al*. (1985), Carvalho *et al*. (1994), Cox (1995) and Fischetti and Salazar (2000) applied the concepts of controlled selection for 'statistical disclosure control' and 'cell supression', Matei and Skinner (2010) and Tiwari and Sud (2011) applied the method of controlled selection implemented with linear and quadratic programming to solve the 'sample coordination' problem.

Different researchers have developed various controlled selection procedures since the inception of the scheme of controlled selection by Goodman and Kish (1950). While the earlier authors used the various combinatorial properties of the sampling designs, which are time consuming and involve considerable amount of trial and error, Rao and Nigam (1990, 1992) were the first to use the linear programming approach to solve the one-dimensional controlled selection problem using simplex method. Their idea was extended to multi-dimensional controlled selection problems by Sitter and Skinner (1994) and Tiwari and Nigam (1998). These schemes attempted to minimize the selection probabilities of the non-preferred combinations of units subject to certain constraints. Recently, Tiwari *et al*.

*Corresponding author : U.C. Sud
E-mail address : ucsud@iasri.res.in

(2007) applied the idea of 'nearest proportional to size sampling design', originated by Gabler (1987), to obtain one-dimensional optimal controlled selection designs which fully exclude the non-preferred combinations of units from the selected samples by assigning them zero probabilities. They also compared the true sampling variances of the estimate for their design with those for alternative optimal controlled designs of Rao and Nigam (1990, 1992) and uncontrolled high entropy selection procedures of Goodman and Kish (1950) and Brewer and Donadio (2003). They observed that the values of the variance of Horvitz-Thompson (1952) [Hereafter we denote Horvitz-Thompson by HT] estimator for their plan were slightly higher than those obtained through the plans of Rao and Nigam (1990, 1992) and Brewer and Donadio (2003) for most of the examples and concluded that the increase in variance might be acceptable given the elimination of undesirable samples by their plan.

It may be noted that the earlier authors considered only one objective function based programming approach for solving the controlled selection problems. Many a times, programming problems involve optimization of more than one objective function subject to certain constraints. For example, a manufacturer may be interested in simultaneous maximization of revenue, total sales and exports. Such multiple objective problems can be solved by multiple objective simplex method. In this article, we introduce the idea of simultaneous minimization of two objective functions to minimize the true sampling variance of the HT estimator for one-dimensional controlled selection problems, which assign zero probability to non-preferred combinations. The paper derives its inspiration from the optimal controlled nearest proportional to size sampling scheme of Tiwari *et al*. (2007). Section 2 deals with some definitions and notations used in the paper. The proposed plan has been discussed in Section 3. In Section 4, some empirical illustrations have been used to demonstrate the utility of the proposed procedure.

2. DEFINITIONS AND NOTATIONS

Consider a one-dimensional population of N units. Suppose a sample of n units is to be selected from this population. Let y be the characteristic under study, Y_i the y-value for the i^{th} unit in the population ($i = 1, 2,\ldots, N$) and y_i the y-value for the i^{th} unit in the sample ($i = 1, 2,\ldots, n$). We denote the first and second order inclusion probabilities by π_i and π_{ij}, respectively.

To estimate the population mean $\overline{Y}\left(= N^{-1}\sum_{i=1}^{N} y_i\right)$ based on a sample s of size n, we use the HT estimator defined as

$$\hat{\overline{Y}}_{HT} = \sum_{i \in s} \frac{Y_i}{N\pi_i} \tag{1}$$

Sen (1953) and Yates and Grundy (1953) showed independently that for fixed size sampling designs, $\hat{\overline{Y}}_{HT}$ has the variance

$$V(\hat{\overline{Y}}_{HT}) = \frac{1}{N^2}\sum_{i<j=1}^{N} (\pi_i\pi_j - \pi_{ij})\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2, \tag{2}$$

and an unbiased variance estimator

$$\hat{V}(\hat{\overline{Y}}_{HT}) = \frac{1}{N^2}\sum_{i<j=1}^{n} \frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 \tag{3}$$

The non-negativity of the variance estimator (3) is ensured when the condition $\pi_{ij} \le \pi_i\,\pi_j$ is satisfied.

The expression for variance of $\hat{\overline{Y}}_{HT}$ correct to $O(N^{-2})$ using the procedure of Goodman and Kish (1950) is given as

$$V(\hat{\overline{Y}}_{HT})_{GK} = \frac{1}{nN^2}\left[\sum_{i \in U} p_i A_i^2 - (n-1)\sum_{i \in U} p_i^2 A_i^2\right] - \frac{n-1}{nN^2}\left[2\sum_{i \in U} p_i^3 A_i^2 - \sum_{i \in U} p_i^2 \sum_{i \in U} p_i^2 A_i^2 - 2\left(\sum_{i \in U} p_i^2 A_i\right)^2\right], \tag{4}$$

where $A_i = \frac{Y_i}{p_i} - Y$, $Y = \sum_{i=1}^{N} Y_i$, $p_i = \pi_i / n$ and U denotes the finite population of N units.

Recently, Brewer and Donadio (2003) derived the π_{ij}-free formula for high entropy variance (meaning the absence of any detectable pattern or ordering in the selected sample units) of the HT estimator. Their expression for the variance of the HT estimator is given by

$$V(\hat{Y}_{HT})_{BD} = \frac{1}{N^2}\sum_{i\in U}\pi_i(1-c_i\pi_i)(Y_i\pi_i^{-1} - Y\,n^{-1})^2, \quad (5)$$

where $c_i = (n-1)/\{n-(2n-1)(n-1)^{-1}\,\pi_i + (n-1)^{-1}\sum_{k\in U}\pi_k^2\}$ for all $i \in U$.

To define 'nearest proportional to size sampling design', let us suppose that we are interested in a sampling design ($p_0(s)$) due to practical considerations whereas due to theoretical considerations an IPPS (inclusion probability proportional to size sampling) design would be desirable. Therefore, it becomes obvious to look for an IPPS design ($p_1(s)$) which is as near as possible to the original sampling design $p_0(s)$. Gabler (1987) suggested the idea of selecting the design $p_1(s)$ in such a manner that $p_1(s)$ minimizes the directed distance D from the sampling design $p_0(s)$ to the sampling design $p_1(s)$. This design $p_1(s)$ is known as the 'nearest proportional to size sampling design'.

3. THE MINIMUM VARIANCE OPTIMAL CONTROLLED SAMPLING DESIGN

Suppose that in the one-dimensional population considered in Section 2, the initial selection probabilities (p_i's) of the N units of the population are known. Let S and S_1 denote respectively, the set of all possible samples and the set of non-preferred samples. In order to attain an optimal controlled sampling design ensuring zero selection probabilities to non-preferred samples, Tiwari *et al*. (2007) [Hereafter we denote the method suggested by Tiwari *et al*. (2007) by TNP] suggested applying the idea of Gabler (1987) to obtain the required IPPS design $p_1(s)$. In what follows, we extend the idea of TNP to suggest a minimum variance controlled sampling design.

Following the method suggested by TNP, we first obtain an appropriate uncontrolled inclusion probability proportional to size (IPPS) design $p(s)$, such as Midzuno-Sen (1952, 1953) or Sampford (1967) design, depending upon the conditions of these designs being satisfied by the initial selection probabilities (p_i's) of the units in the population. After obtaining the initial IPPS design $p(s)$, we get rid of the non-preferred samples (S_1) by restricting ourselves to the set S-S_1 by introducing a new design $p_0(s)$ that assigns zero probabilities of selection to all non-preferred samples, given by

$$p_0(s) = \begin{cases} \dfrac{p(s)}{1 - \sum_{s\in S_1} p(s)} & \text{for } s \in S - S_1 \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

where $p(s)$ is the initial uncontrolled IPPS sampling plan.

Consequently, $p_0(s)$ is no longer an IPPS design. Now we apply the idea of Gabler (1987) to obtain 'nearest proportional to size sampling design' $p_1(s)$ in the sense that $p_1(s)$ minimizes the directed distance D from the sampling design $p_0(s)$ to the sampling design $p_1(s)$, defined as

$$\phi_1 = D(p_0, p_1) = E_{p_0}\left[\frac{p_1}{p_0} - 1\right]^2 = \sum_{s\in S-S_1}\frac{p_1^2(s)}{p_0(s)} - 1 \quad (7)$$

To minimize the true sampling variance of the HT estimator, we include one more objective function to the above quadratic programming problem, given by

$$\phi_2 = Var(\hat{Y}_{HT}) = \frac{1}{N^2}\{np_i.np_j - \sum_{s\ni i,j} p_1(s)\}\left[\frac{Y_i}{np_i} - \frac{Y_j}{np_j}\right]^2 \quad (8)$$

Thus, the formulation of the problem is as follows:

Minimize the objective function $\phi_1 + \phi_2$

subject to the following constraints:

(i) $p_1(s) \geq 0$

(ii) $\sum_{s\in S-S_1} p_1(s) = 1$

(iii) $\sum_{s\ni i} p_1(s) = np_i$, $(i = 1,\ldots, N)$

(iv) $\sum_{s\ni i,j} p_1(s) > 0$ (9)

(v) $\sum_{s\ni i,j} p_1(s) \leq (np_i)(np_j)$, $(i < j = 1,\ldots, N)$

Constraints (i) and (ii) are necessary for any probability design. Constraint (iii) ensures the resultant design ($p_1(s)$) to be IPPS. Constraint (iv) ensures the unbiased estimation of variance and constraint (v) ensures the sufficient condition for the non-negativity of the Yates-Grundy estimator of variance.

The solution to the above two objective programming problem, viz., minimization of sum of (7) and (8) subject to the constraints (9), provides us with the minimum variance optimal controlled IPPS sampling plan that ensures zero probability of selection for the non-preferred samples and also minimizes the true sampling variance of the HT estimator. A unique feature of the proposed procedure is that besides utilizing all the advantages of the TNP plan, such as ensuring zero probability to non-preferred samples and obtaining a non-negative HT variance estimator, it also minimizes the true sampling variance of the HT estimator. The proposed method also provides an opportunity to add more objective functions to the controlled selection problem.

Multiple objective problems can be solved through multiple objective simplex method. In some situations, the multiple objective problem may not give a bounded solution. In such situations, one can use a weighted multiple objective function and may achieve a bounded solution. The weights can be estimated using the constraint of the dual problem through the phase-1 of the two-phase method.

To demonstrate the utility of the proposed procedure, we use some empirical illustrations to compare the true sampling variance of the HT estimator for the proposed procedure obtained through (3) with variances of the HT estimator using the optimal controlled plan of Tiwari *et al*. (2007), Rao and Nigam (1990, 1992) and those of two uncontrolled high entropy procedures of Goodman and Kish (1950) and Brewer and Donadio (2003).

4. EMPIRICAL EVALUATIONS

In this section we shall present some empirical illustrations to demonstrate that the true sample variance of the proposed procedure compares favorably with that of the existing optimal controlled and uncontrolled high entropy selection procedures.

Let us consider a population consisting of six villages (numbered as 1, 2, …, 6), borrowed from Hedayat and Lin (1980). The set S of all possible samples consists of 20 samples each of size $n = 3$. Due to the considerations of travel, organization of fieldwork and cost considerations, Rao and Nigam (1990) identified the following 7 samples as non-preferred samples

123; 126; 136; 146; 234; 236; 246

Here, 123 denotes the sample consisting of units 1, 2 and 3; 126 denotes the sample consisting of units 1, 2 and 6; and so on.

Set 1(a). The Y_i and p_i values associated with the six villages of the population are

Y_i :	12	15	17	24	17	19
p_i :	0.14	0.14	0.15	0.16	0.22	0.19

Since the p_i values satisfy the condition

$$\frac{1}{n}\frac{(n-1)}{(N-1)} \le p_i \le \frac{1}{n}, \tag{10}$$

that is, the p_i values lie between 2/15 and 1/3, we apply the Midzuno-Sen (MS) scheme to get an IPPS plan with the revised normal size measures (p_i*s) given by

$$p_i^* = np_i\frac{(N-1)}{(N-n)} - \frac{(n-1)}{(N-n)}, \tag{11}$$

and the probability of including n units $i_1, i_2,…,i_n$ in the s^{th} sample, given by

$$p(s) = \pi_{i_1 i_2 \ldots i_n} = \frac{1}{\binom{N-1}{n-1}} \; (p_{i_1}^* + p_{i_2}^* + \ldots + p_{i_n}^*) \tag{12}$$

Applying the method discussed in Section 2 and solving the resulting two objective problem through the SAS 9.1 windows version package, we obtain the controlled IPPS plan given in Table 1.

While minimizing the true sampling variance of the estimate, this plan also matches the original π_i values, satisfies the condition $\pi_{ij} \le \pi_i \pi_j$, and ensures that the probability of selecting non-preferred samples is exactly equal to zero. Obviously, due to the fulfillment of the condition $\pi_{ij} \le \pi_i \pi_j$, we can apply the Yates-Grundy form of the HT variance estimator for estimating the variance of the proposed plan.

The values of the true sampling variance of the HT estimator [$V(\hat{Y}_{HT})$] for the proposed plan, the TNP plan

Table 1. Optimal controlled IPPS plan corresponding to Midzuno-Sen (MS) and Sampford's (SAMP) schemes for Example1

Sample (s)	$p_1(s)$ [MS]	$p_1(s)$ [SAMP]	Sample (s)	$p_1(s)$ [MS]	$p_1(s)$ [SAMP]
124	0.1462	0.0874	245	0.0048	0.1413
125	0.0301	0.0064	256	0.1500	0.0877
134	0	0	345	0	0.0108
135	0.1046	0	346	0.1938	0.1026
145	0.0554	0.0869	356	0.0627	0.0594
156	0.0837	0.1193	456	0.0798	0.1710
235	0.0889	0.1272			

[Tiwari *et al*. (2007, p. 89)], the plan (3) of Rao and Nigam (1990, p. 809) with specified π_{ij}'s taken from the Sampford's plan [to be denoted by RN3] and their plan (4) [to be denoted by RN4], the Randomized Systematic IPPS sampling plan of Goodman and Kish (1950) [to be denoted by GK] and the uncontrolled high entropy sampling plan of Brewer and Donadio (2003) [to be denoted by BD] are produced in the first row of Table 2. It is clear from Table 2 that the proposed plan yields 7.39% reduction in sampling variance in comparison to that obtained through TNP plan. The variance obtained through the proposed plan is even less than that obtained through RN4 plan.

The expressions for $V(\hat{Y}_{HT})$ for GK and BD plans are given in equations (4) and (5) in Section 2, respectively. For the other plans viz. proposed plan, TNP, RN3 and RN4 the value of $V(\hat{Y}_{HT})$ has been calculated using equation (2) in Section 2, with probabilities calculated by their respective methods.

Set 1(b). Now suppose that the p_i values for the above population of 6 units are as follows

$$p_i: \quad .10 \quad .15 \quad .10 \quad .20 \quad .27 \quad .18$$

Since these values of p_i do not satisfy the condition (10) of the MS plan, we apply the Sampford (1967) plan to get the initial IPPS design $p(s)$, given by

$$p(s) = \pi_{i_1 i_2 \ldots i_n} = n\,K_n\,\lambda_{i_1}\lambda_{i_2}\ldots\lambda_{i_n}\,\Big(1-\sum_{u=1}^{n} p_{i_u}\Big), \qquad (13)$$

where $K_n = \left(\sum_{t=1}^{n} \frac{tL_{n-t}}{n^t}\right)^{-1}$, $\lambda_t = p_i/(1-p_i)$ and for a set $S(m)$ of $m \le N$ different units, $i_1, i_2, \ldots, i_m$, L_m is defined as

$$L_0 = 1,\ L_m = \sum_{S(m)} \lambda_{i_1}\lambda_{i_2}\ldots\lambda_{i_m} \quad (1 \le m \le N).$$

Applying the method discussed in Section 2 and solving the resultant two objective problem, we obtain the controlled IPPS plan given in Table 1. This plan again ensures zero probability to non-preferred samples, satisfies the non-negativity condition for the Yates-Grundy form of the HT variance estimator and ensures zero probability to non-preferred samples. Moreover, the proposed plan also minimizes the true sampling

Table 2. Values of the true sampling variance of the HT estimator [$V(\hat{Y}_{HT})$] for the Proposed, RN3, RN4, GK and BD plans

$V(\hat{Y}_{HT})$	RN3	RN4	GK	BD	TNP	PROPOSED
Set 1(a) : N = 6, n = 3	2.930	4.024	3.034	2.919	4.057	3.762
Set 1(b) : N = 6, n = 3	4.757	5.069	4.894	4.154	4.784	4.191
Set 2(a) : N = 7, n = 3	4.476	5.008	4.610	4.447	3.564	2.300
Set 2(b) : N = 7, n = 3	11.967	14.520	12.250	11.443	9.489	4.832
Set 3(a) : N = 8, n = 3	4.854	4.289	4.957	4.836	3.902	1.981
Set 3(b) : N = 8, n = 3	7.292	8.429	7.735	7.372	8.168	5.793
Set 4(a) : N = 8, n = 4	3.185	3.463	3.227	3.154	3.745	2.479
Set 4(b) : N = 8, n = 4	2.409	2.527	2.544	2.385	2.254	1.575
Set 5 : N = 7, n = 4	3.076	3.929	3.122	3.075	5.100	2.759

variance of the HT estimator. The values of $V(\hat{Y}_{HT})$ for the proposed plan, the TNP plan, the RN3 plan, the RN4 plan, the GK plan and the BD plan are produced in the second row of Table 2. The proposed plan appears to perform better than the TNP, RN3, RN4 and GK plans and quite close to BD plan. The proposed plan ensures 12.34% reduction in true sampling variance in comparison to TNP plan.

Further illustrations were constructed to analyze the performance of the proposed plan. The Y_i, p_i values along with the set of non-preferred samples for each population are summarized in the Appendix. The p_i values for Sets 2(a), 3(a) and 4(a) and 5 satisfy the condition (10) of MS plan and hence for these examples the MS IPPS plan is used to obtain the initial IPPS design $p(s)$. Sets 2(b), 3(b) and 4(b) do not satisfy the MS condition. Therefore we apply the Sampford IPPS plan to obtain the initial IPPS design. For Set 5, the non-negativity condition [$\pi_{ij} \leq \pi_i \pi_j$] of the Yates-Grundy form of the HT variance estimator is not satisfied. Thus the constraint (v) in (9) has been dropped to solve this problem. The values of $V(\hat{Y}_{HT})$ for the proposed, the TNP, the RN3, the RN4, the GK and the BD plans for the population summarized in the Appendix are given in Table 2. From this Table we conclude that for all the empirical illustrations [Set 2 to Set 5] considered above, the proposed plan appears to perform better than the TNP, RN3, RN4, GK and BD plans. This may be a great advantage of the proposed plan that even after completely eliminating non-preferred samples from the selected samples, the true sampling variances are less than the uncontrolled high entropy selection procedures.

5. CONCLUSION

In this paper the multiple objective programming approach has been proposed to minimize the true sampling variance of the HT estimator while assigning zero probabilities to non-preferred samples. One of the drawbacks of the TNP plan, that is the increase in sampling variance in comparison to existing optimal controlled and uncontrolled high entropy selection procedures, has been successfully removed in the proposed method by introducing one more objective function in the form of variance of HT estimator and minimizing it. The true sampling variance of the HT estimator is empirically compared with that of existing controlled sampling plans and uncontrolled high entropy selection procedures. The proposed plan appears to perform suitably well.

One limitation of the proposed plan is that it becomes impractical when N is very large, as the process of enumeration of all possible samples and formation of objective function as well as constraints becomes quite tedious. However, with the advent of faster computing techniques and modern statistical packages, there may not be much difficulty in using the proposed procedure for moderately large populations. Another limitation of the proposed procedure is that as in the case of linear programming, there is no guarantee of convergence of a quadratic programming problem. Moreover, due to consideration of two objective functions, there may be a minor loss of efficiency in minimization of directed distance D given by equation (7). However, with the help of different illustrations considered by us, it was found that this loss was negligible and does not affect the performance of the proposed procedure. This may also be noted that the above multiple programming problem may also be solved by the methods other than simplex method. The reason behind using the simplex method to solve the above problem was its wide acceptability and availability in SAS 9.1 windows version package used by us to solve the multiple programming problem.

ACKNOWLEDGEMENTS

The authors are thankful to the Editor and one reviewer for their valuable suggestions and constructive comments, which led to considerable improvement in the presentation of this work.

REFERENCES

Avadhani, M.S. and Sukhatme, B.V. (1973). Controlled sampling with equal probabilities and without replacement. *Inter. Statist. Rev.*, **41**, 175-182.

Brewer, K.R.W. and Donadio, M.E. (2003). The high-entropy variance of the Hortivz-Thompson estimator. *Survey Methodology,* **29**, 189-196.

Carvalho, F.D., Dellaert, N.P. and Osorio, M.S. (1994). Statistical disclosures in two-dimensional tables: General tables. *J. Amer. Statist. Assoc.*, **89**, 1547-1557.

Causey, B.D., Cox, L.H. and Ernst, L.R. (1985). Application of transportation theory to statistical problems. *J. Amer. Statist. Assoc.*, **80**, 903-909.

Cox. L.H. (1995). Network models for complementary cell suppression. *J. Amer. Statist. Assoc.*, **90**, 1453-1462.

Fischetti, M. and Salazar, J.J. (2000). Models and algorithms for optimizing cell suppression in tabular data with linear constraints. *J. Amer. Statist. Assoc.*, **95**, 916-928.

Gabler, S. (1987). The nearest proportional to size sampling design. *Comm. Statist.-Theory Methods*, **16(4)**, 1117-1131.

Goodman, R. and Kish, L. (1950). Controlled selection-a technique in probability sampling. *J. Amer. Statist. Assoc.*, **45**, 350-372.

Hedayat, A. and Lin, B.Y. (1980). Controlled probability proportional to size sampling designs. *Technical Report,* University of Illinois at Chicago.

Hedayat, A., Lin, B.Y. and Stukken, J. (1989). The construction of IPPS sampling designs through a method of emptying boxes. *Ann. Statist.*, **17**, 1886-1905.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from finite universes. *J. Amer. Statist. Assoc.*, **47**, 663-685.

Lu, W. and Sitter, R.R. (2002). Multi-way stratification by linear programming made practical. *Survey Methodology,* **28(2)**, 199-207.

Matei, A. and Skinner, C. (2009). Optimal sample coordination using controlled selection. *J. Statist. Plann. Inf.*, **139**, 3112-3121.

Midzuno, H. (1952). On the sampling system with probability proportional to sums of sizes. *Ann. Instt. Statist. Maths.*, **3**, 99-107.

Rao, J.N.K. and Nigam, A.K. (1990). Optimal controlled sampling designs. *Biometrika*, **77**, 807-814.

Rao, J.N.K. and Nigam, A.K. (1992). 'Optimal' controlled sampling: A unified approach. *Inter. Statist. Rev.*, **60**, 89-98.

Sampford, M.R. (1967). On sampling with replacement with unequal probabilities of selection. *Biometrika*, **54**, 499-513.

Sande, G. (1984). Automated cell suppression to preserve confidentiality of business statistics. *Statist. J. United Nations ECE*, **2**, 33-41.

Sen, A.R. (1953). On the estimation of variance in sampling with varying probabilities. *J. Ind. Soc. Agril. Statist.*, **5**, 119-127.

Sitter, R.R. and Skinner, C.J. (1994). Multi-way stratification by linear programming. *Survey Methodology*, **20**, 65-73.

Tiwari, N. and Nigam, A.K. (1998). On two-dimensional optimal controlled selection. *J. Statist. Plann. Inf.*, **69**, 89-100.

Tiwari, N., Nigam, A.K. and Pant, I. (2007). On an optimal controlled nearest proportional to size sampling scheme. *Survey Methodology,* **33**, 87-94.

Tiwari, N. and Sud, U.C. (2011). An optimal procedure for sample coordination using multiple objective functions and nearest proportional to IPPS size sampling designs. *Comm. Statist.-Theory Methods*. In Press.

Wynn, H.P. (1977). Convex sets of finite population plans. *Ann. Statist.*, **5**, 414-418.

Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *J. Roy. Statist. Soc.*, **B15**, 253-261.

APPENDIX

The populations for Set 2-4 with Y_i and p_i values and the set of non-preferred samples

Set 2. N = 7, n = 3

Non-preferred samples :

123; 126; 136; 146; 234; 236; 246;

137; 147; 167; 237; 247; 347; 467.

Y_i:	12	15	17	24	17	19	25
(a). p_i:	0.12	0.12	0.13	0.14	0.20	0.15	0.14
(b). p_i:	0.08	0.08	0.16	0.11	0.24	0.20	0.13

Set 3. N = 8, n = 3

Non-preferred samples :

123; 126; 136; 146; 234; 236; 246;

137; 147; 167; 237; 247; 347; 467;

128; 178; 248; 458; 468; 478; 578.

Y_i:	12	15	17	24	17	19	25	18
(a). p_i:	0.10	0.10	0.11	0.12	0.18	0.13	0.12	0.14
(b). p_i:	0.05	0.09	0.20	0.15	0.10	0.11	0.12	0.18

Set 4. N = 8, n = 4

Non-preferred samples :

1234; 1236; 1238; 1246; 1248; 1268; 1346; 1348; 1357; 1456; 1468; 1567; 1568; 1678; 2345; 2346; 2456; 2468; 2567; 2568; 2678; 3456; 3468; 3567; 3678; 4567; 4678; 5678.

Y_i:	12	15	17	24	17	19	25	18
(a). p_i:	0.11	0.11	0.12	0.13	0.17	0.12	0.11	0.13
(b). p_i:	0.09	0.09	0.18	0.11	0.12	0.14	0.17	0.10

Set 5. N = 7, n = 4 (when the non-negativity condition $\pi_{ij} \leq \pi_i \pi_j$ is not satisfied)

Non-preferred samples :

1234; 1236; 1246; 1346; 1357; 1456; 1567; 2345; 2346; 2456; 2567; 3456; 3567; 4567.

Y_i:	12	15	17	24	17	19	25
p_i:	0.14	0.13	0.15	0.13	0.16	0.15	0.14