Balanced Sampling Plans excluding Adjacent Units - An Overview

B. N. Mandal, V. K. Gupta and Rajender Parsad
ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, Pusa, New Delhi, India

Abstract

Balanced sampling plans excluding adjacent units are those sampling plans in which second order inclusion probabilities are zero for pairs of adjacent units and constant for pairs of non-adjacent units. These plans are useful for sampling in situations where the contiguous units in a population are similar. In this article an overview of these sampling plans is given including some results on existence and construction.

***Keywords*:** Sampling plans, Balanced sampling plans, Adjacent units, Narain-Horvitz-Thompson estimator, Polygonal designs, Inclusion probability

1. Introduction

The purpose of sample survey is estimation of population parameters by observing a representative subset of units from a population consisting of distinctly identifiable units. This subset of sampling units is usually selected by probability sampling also termed as sampling scheme or a sampling design. Thus, selection of a suitable sampling scheme / sampling design for selecting samples from the population under study is a very important step to ensure the success of estimation process. Selection of a sampling design may depend upon a number of factors out of which an important factor is the nature of the population being surveyed. In this article, we consider sampling from a naturally ordered (in time or space) population where nearer units are expected to be similar. To make the exposition clearer, consider a situation where the investigator wants to estimate the potato production in a large field by observing potato production in a few sampled rows from total number of rows in the field. It is quite natural that the total produce of potato from nearer rows is expected to be similar because of similar microenvironments. Hence, if the investigator selects a sample using simple random sampling without replacement (SRSWOR), then s/he may get a sample which contains nearer rows as sampling units and thus, s/he may not get a good representation of the population. Thus, SRSWOR would not be a suitable sampling design in this situation. In this case, it would be better to select a sample in such a way that nearer units are avoided or given lesser chance of being included in a sample. A class of sampling plans called balanced sampling plans excluding contiguous units (BSEC plans) was introduced by Hedayat*et al.* (1988) for sampling in such scenario. Later on, a more generalized class of sampling plans called balanced sampling plans excluding adjacent units was introduced by Stufken (1993). From here onwards, we shall call these plans as BSA plans.

In this article, we give an overview of BSA plans. The remainder of the article is organized as follows. In Section 2, some preliminaries related to the terminologies in sample survey are given. BSA plans are defined and their properties are described in Section 3. The existence and construction of BSA plans is discussed in Section 4.

An illustration of efficiency of the plans with SRSWOR is provided for a real population data in Section 5. The article is concluded in Section 6 with some future directions for research.

2. Preliminaries

A finite population is a collection of known number N of distinct and identifiable sampling units. A population of size N may be represented by the set $U = \{U_1, U_2, \cdots, U_i, \cdots, U_N\}$, where U's denote the sampling units. The study variable is denoted by Y having value Y_i on unit i; i = 1,2,...,N.

We may represent by $Y = (Y_1, Y_2, \cdots, Y_i, \cdots, Y_N)'$ an N-component vector of the values of the study variable Y for the N population units. The vector $\boldsymbol{Y}$ is assumed fixed, though unknown. Sometimes auxiliary information is also available on some other characteristic X related with the study variable Y. The auxiliary information is generally available for all the population units. We may represent by $X = (X_1, X_2, \cdots, X_i, \cdots, X_N)'$ an N-component vector of the values of the auxiliary variable X for the N population units. The total $X_1 + X_2 + \cdots + X_i + \cdots + X_N$ is generally known.

A list of all the sampling units in the population along with their identity is known as sampling frame. The sampling frame is a basic requirement for sampling from finite populations. It is assumed that the sampling frame is available and it is perfect in the sense that it is free from under or over coverage and duplication.

The probability selection procedure selects the units from U with probability P_i, $i \in U$. We shall denote by $P = (P_1, P_2, \cdots, P_i, \cdots, P_N)'$ an N-component vector of the initial selection probabilities of the units such that $P'1 = 1$.. Generally $\mathbf{P} \sim g\ (n, N, \boldsymbol{X})$; *e.g.*, $P_i = 1/N\ \forall i \in U$; or $P_i = n/N\ \forall i = 1, 2, \ldots, k$, $k = N/n$; or $P_i = X_i/(X_1 + X_2 + \ldots + X_N)$, $\forall\ i \in U$.

A nonempty set $\{s: s \subseteq U\}$, obtained by using probability selection procedure **P**, is called an unordered sample. The cardinality of s is n, which is also known as the (fixed) sample size. We shall generally assume a fixed sample size n throughout the article. A set of all possible samples is called sample space S. Given a probability selection procedure **P** which describes the probability of selection of units one by one, we define the probability of selection of a sample s as $p(s) = g(P_i): i \in s,\ s \in S$. We also denote by $p = \{p(1), p(2), \cdots, p(s), \cdots, p(\upsilon)\}'$ a $\upsilon = NC_n$-component vector of selection probabilities of the samples. Obviously, $p(s) \geq 0$ and $p'1 = 1$. It is well known that given a unit by unit selection procedure, there exists a unique mass selection procedure; the converse is also true.

After the sample is selected, data are collected from the sampled units. Let y_i be the value of study variable on the i^{th} unit selected in the sample s, $i \in s$ and $s \in S$. We shall denote by $y = (y_1, y_2, \cdots, y_i, \cdots, y_n)'$ an n-component vector of the sampled observations. It is assumed here that the observation vector **y** is measured without error and its elements are the true values of the sampled units.

The problem in sample surveys is to estimate some unknown population parameter $\theta = f(Y)$ or $\theta = f_1(Y, X)$. We shall focus our attention on the estimation of a particular choice of θ, say population total, $Y'1 = \sum_{i \in U} Y_i$, or population mean $\bar{Y}_N = N^{-1} Y'1 = N^{-1} \sum_{i \in U} Y_i$. An estimator e for a given sample s is a function such that its value depends on y_i, $i \in s$. In general $e = h(y, X)$ and the functional form h (. , .) would also depend upon the functional form of θ, besides being a function of the sampling design. We can also write $e_s = h\{\mathbf{y}, p(s)\}$.

A *sampling design* is defined as:

$$d = [\{s, p(s)\}: s \in S]. \quad (2.1)$$

Further $\sum_{s \in S} p(s) = 1$. (2.2)

The set S is also called the *support* of the sampling plan and $\upsilon = C_n$ is called the *support size*. A sampling plan is said to be a fixed-size sampling plan whenever $p(s) > 0$, the corresponding subsets of units are composed of the same number of units. The triplet (S, **p**, e_s) is called the *sampling strategy*.

A familiarity with the expectation and variance operators is assumed in the sequel. An estimator e_s is said to be *unbiased* for estimation of population parameter θ if

$E_d(e_s) = \theta$ with respect to a sampling design *d*,

where E_d denotes the expectation operator.

The *bias* of an estimator e_s for estimating θ, with respect to a sampling design *d*, is

$$B_d(e_s) = E_d(e_s) - \theta.$$

Variance of an unbiased estimator e_s for θ, with respect to sampling design *d*, is

$$V_d(e_s) = E_d\{e_s - E_d(e_s)\}^2 = E_d(e_s - \theta)^2 = E_d(e_s^2) - \theta^2.$$

The *mean square error* of a biased estimator e_s for θ is given by

$$MSE_d(e_s) = E_d(e_s - \theta)^2 = E_d\{e_s - E_d(e_d) + E_d(e_d) - \theta\}^2 = V_d(e) + \{B_d(e)\}^2.$$

3. BSA Plans and Their Properties

Assume that the population has N distinct and identifiable units and we are interested in drawing a sample of fixed size n. BSA plans are defined based on the adjacency of sampling units. Two distinct units are said to be adjacent if they are less than or equal to pre-specified distance m, where m is a positive integer. The distance between two units will depend on their arrangement. For simplicity and mathematical convenience, henceforth we assume that the population units are arranged in circular order, unless otherwise specified. In that case, the distance between two units i and j, $i \neq j=1, 2,\ldots\ldots, N$ is defined as

$$\delta(i,j) = \text{Min}\{|i-j|, N-|i-j|\}. \quad (3.1)$$

For example, if $N = 9$ then distance between units 1 and 6 is $\delta(1,6) = \text{Min}\{|1-6|, 9-|1-6|\} = 4$. It may be pointed out here that under circular ordering of units, (2.1) implies the distance between units 1 and N is 1, between units 1 and $N - 1$ is 2 and so on. It may be noted here that under circular ordering, maximum distance between two units can be $[N/2]$, where $[x]$ denote the largest integer in x. Now, we are in a position to define BSA plans.

Definition 2.1. A BSA plan is a sampling plan in which

i) All pairs of units which are adjacent have zero second order inclusion probabilities, and
ii) All pairs of units which are non-adjacent have constant non-zero second order inclusion probabilities.

To be specific, if two units *i* and *j* are adjacent, then $\pi_{\text{ij}} = 0$, else $\pi_{\text{ij}} =$ constant, where π_{ij}denotes the second order inclusion probability of units *i* and *j*, $i \neq$ j=1,2,...,N. Since BSA plans imply constant second order inclusion probabilities for non-adjacent pairs, hence these are termed as '*balanced sampling plans avoiding adjacent units*'. It may be verified that the second order inclusion probabilities for non-adjacent units are given by:

$$\pi_{ij} = \frac{n(n-1)}{N(N-2m-1)} \forall \delta(\text{i,j}) > \text{m}. \quad (3.2)$$

From the relation, $\sum_{j(\neq i)=1}^{N} \pi_{\text{ij}} = (n-1)\pi_i$, we find that

$$\pi_i = \frac{n}{N}. \quad (3.3)$$

That is, the first order inclusion probabilities are constant for all the population units under a BSA plan. We give an example of a BSA plan for *N* = 9, *n* = 3 and *m* = 1 in Table 1. Under the plan, first order inclusion probabilities are $\pi_i = 3/9 = 1/3$and second order inclusion probabilities are: $\pi_{12}=\pi_{23}=\pi_{34}=\pi_{45}=\pi_{56}=\pi_{67}=\pi_{78}=\pi_{89}=\pi_{91} = 0$ and rest of the second order inclusion probabilities are $\pi_{\text{ij}} = 1/9$ for all non-adjacent pairs of units.

Table 1: A BSA plan for N = 9, n = 3 and m = 1

s	*p*(*s*)		*s*	*p*(*s*)
1, 3, 6	1/9		6, 8, 2	1/9
2, 4, 7	1/9		7, 9, 3	1/9
3, 5, 8	1/9		8, 1, 4	1/9
4, 6, 9	1/9		9, 2, 5	1/9
5, 7, 1	1/9			

There may appear some similarity between systematic sampling and BSA plan as systematic sampling plans also avoid nearer units. But systematic sampling is different from BSA plans in the sense that in systematic sampling, starting with a random unit *i*, units at a distance of exactly *k* and its multiple are selected and observed in the sample, while all other units are excluded from the sample. In other words, the second order inclusion probabilities are $\pi_{\text{ij}} = 1/k$ for *j* = *i*+ *k*, *i* + 2*k*, *i*+ 3*k*, ...*i*+(*n*-1)*k* and is zero otherwise. However, in BSA plan, second order inclusion probabilities of all non-adjacent pair of units are same non-zero constant.

Now, let us consider the estimation of population mean $\bar{Y}$using the Narain-Horvitz-Thompson estimator (Narain, 1951; Horvitz and Thompson, 1952), given by:

$$\hat{\bar{Y}}_{\text{NHT}} = \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i}. \quad (3.4)$$

The Sen-Yates-Grundy (Sen, 1953; Yates and Grundy, 1953) form of variance of Narain- Horvitz-Thompson estimator under BSA, using (3.2) and (3.3) can be obtained as:

$$V(\bar{\hat{Y}}_{\text{NHT}}) = \frac{\sigma^2}{n}\left(1 - \frac{(n-1)\left(1+2\sum_{j=1}^{m}\rho_j\right)}{N-2m-1}\right) \qquad (3.5)$$

where $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$ is the population variance and $\rho_j = \frac{1}{N\sigma^2}\sum_{i=1}^{N}(Y_i - \bar{Y})\left(Y_{i+j} - \bar{Y}\right)$ is the jth order circular serial correlation coefficient, $j = 1, 2, ..., m$.

Comparing with the variance of the Narain-Horvitz-Thompson estimator of population mean under SRSWOR, it can be easily seen that a BSA plan is more efficient than SRSWOR if $\sum_{j=1}^{m}\rho_j > -\frac{m}{N-1}$. For $m = 1$, the condition reduces to

$$\rho_1 > -\frac{1}{N-1}.$$

Since some of the second order inclusion probabilities are zero, an unbiased estimator of variance of Narain-Horvitz-Thompson estimator of population mean under BSA plan does not exist. An approximate variance estimator due to Wright and Stufken (2011) is

$$\hat{V}(\bar{\hat{Y}}_{\text{NHT}}) = \frac{N-(2m+1)n}{2Nn^2(n-1)}\sum_{i\in s}\sum_{j\in s}\left(y_i - y_j\right)^2. \qquad (3.6)$$

The approximate variance estimator (3.6) is non-negative whenever $N \geq (2m+1)n$.

4. Existence and Construction of BSA Plans

In this Section, we review some existence and construction results of BSA plans. For $n = 2$, a BSA plan always exist whenever $N \geq 2(m+1)$. The support of a BSA plan for such a population is obtained by taking all the possible pairs of non-adjacent units. For example, a BSA plan for sample size 2 for $N = 8$, $m = 2$ is given in Table 2.

Table 2: A BSA plan for N = 8, n = 2 and m = 2

S		**p(s)**
1	4	1/12
1	5	1/12
1	6	1/12
2	5	1/12
2	6	1/12
2	7	1/12
3	6	1/12
3	7	1/12
3	8	1/12
4	7	1/12
4	8	1/12

5	8	1/12

It may be checked that the plan gives $\pi_{ij} = 1/12$ for all non-adjacent pairs of units and $\pi_{ij} = 0$ for all adjacent pairs of units.

Establishing existence of the BSA plans for $n \geq 3$ is not trivial. Theorem 4.1 gives a necessary condition for existence of a BSA plan for $n \geq 3$.

Theorem 4.1: *A necessary condition for existence of a BSA plan for N, n and m is*

$N \geq (2m+1)n$ *for* $n \geq 3$ *and* $m \geq 1$ *and*
$N \geq (2m+1)n+1$

for the following combinations of (n, m): $\{(n,1); n \geq 5\}$, $\{(n,2); 6 \leq n \leq 12\}$, $\{(n,3); 5 \leq n \leq 9\}$, $\{(n,4); n = 6,7,8\}$, $\{(n,5); n = 6,7\}$.

For proof of Theorem 4.1, see Stufken (1993), Stufken*et al.* (1999) and Wright and Stufken (2008).

An interesting way of obtaining BSA plan from an existing BSA plan due to Stufken (1993) is given below.

Theorem 4.2: *If there exists a BSA plan for N, n and m, then there exists a BSA plan for N' = N +2m +1, n and m.*

For this, let $\{s, p(s)\}$ be an existing BSA plan with N, n and m. Change the sets $\{s\}$ by the elements $N - j$ to $N + m + 1 - 2j$ for $j = 1, 2, ..., m$. Let $\{s^*\}$ denote the sets obtained from $\{s\}$ by these relabeling of elements. Now, consider elements of each of the sets $\{s^*\}$ as residues modulo $N + 2m + 1$ and then develop each of these sets as modulo $N + 2m + 1$. The resulting support is a support of BSA plans for $N + 2m + 1$, n and m.

For example, consider the BSA plan for $N = 9$, $n = 3$ and $m = 1$ given in Table 1. Then, if we relabel the unit 8 as 9 in the support and develop the resulting sets modulo 12, then we get a support of BSA plan for $N = 12$, $n = 3$ and $m = 1$ as shown in Table 3. Giving probability of selection 1/108 to each of the sample in the support gives a BSA plan.

Table 3: The support of BSA plan for N = 12, n = 3 and m = 1

1	3	6	4	6	9	7	9	3
2	4	7	5	7	10	8	10	4
3	5	8	6	8	11	9	11	5
4	6	9	7	9	12	10	12	6
5	7	10	8	10	1	11	1	7
6	8	11	9	11	2	12	2	8
7	9	12	10	12	3	1	3	9
8	10	1	11	1	4	2	4	10
9	11	2	12	2	5	3	5	11
10	12	3	1	3	6	4	6	12
11	1	4	2	4	7	5	7	1

12	2	5	3	5	8	6	8	2
2	4	7	5	7	1	9	1	4
3	5	8	6	8	2	10	2	5
4	6	9	7	9	3	11	3	6
5	7	10	8	10	4	12	4	7
6	8	11	9	11	5	1	5	8
7	9	12	10	12	6	2	6	9
8	10	1	11	1	7	3	7	10
9	11	2	12	2	8	4	8	11
10	12	3	1	3	9	5	9	12
11	1	4	2	4	10	6	10	1
12	2	5	3	5	11	7	11	2
1	3	6	4	6	12	8	12	3
3	5	9	6	9	2	9	2	5
4	6	10	7	10	3	10	3	6
5	7	11	8	11	4	11	4	7
6	8	12	9	12	5	12	5	8
7	9	1	10	1	6	1	6	9
8	10	2	11	2	7	2	7	10
9	11	3	12	3	8	3	8	11
10	12	4	1	4	9	4	9	12
11	1	5	2	5	10	5	10	1
12	2	6	3	6	11	6	11	2
1	3	7	4	7	12	7	12	3
2	4	8	5	8	1	8	1	4

Combinatorial properties of block designs may be utilized to obtain BSA plans. For this purpose, Stufken*et al.* (1999) introduced polygonal designs which are defined below.

Definition 4.1. A polygonal design in v treatments and b blocks with each block of size k is an incomplete block design such that

i) No treatment appears twice in a block,
ii) Every treatment appears in r blocks in the design, and
iii) Each pair of treatments which are at a distance of m or less do not appear together in any block and each pair of treatments which are at distance of more than m appear together in λ blocks.

The symbols v, b, r, k, λ and m are the parameters of the design and they satisfy the following necessary conditions.

i) $vr = bk$
ii) $\lambda(v-2m-1)=r(k-1)$.

The design given in Table 4 is a polygonal design with $v = 9$, $b = 9$, $r = 3$, $k = 3$, $\lambda=1$, m=1.

Table 4: A polygonal design for v = 9, b = 9, r = 3, k = 3, λ=1 ***and m = 1***

1	3	6
2	4	7
3	5	8
4	6	9
5	7	1
6	8	2
7	9	3
8	1	4
9	2	5

The polygonal designs have one to one correspondence with a BSA plan. With $N = v$ and $k = n$, if we consider the treatments as sampling units, the blocks as samples, the treatments in a block as the units in the sample and then if every block of a polygonal design is given probability of selection as $1/b$, then the polygonal design is equivalent to a BSA plan for population size N, sample size n and m. Thus, obtaining a polygonal design is equivalent to obtaining a BSA plan. We now present a result due to Mandal *et al.* (2008) and Stufken and Wright (2008) for constructing polygonal designs.

Theorem 4.3: Let $B_1, B_2, ..., B_t$ denote t initial blocks with k distinct treatments from the set {1, 2, ..., v}. Let $B_u = \{b_{u1}, b_{u2}, ..., b_{uk}\}$, u = 1, 2, ..., t. Then if in the $tk(k-1)$ pair wise distances of the elements from the t blocks, distances 1, 2, ..., m do not appear and distances $m + 1, m + 2, ...,[v/2]$ appear λ times then, a polygonal design is obtained by developing the t initial blocks modulo v. The parameters of the design are given by v, $b = tv$, $r = tk$, k, λ and m.

The design given in Table 4 is obtained making use of the Theorem 4.2 with initial block being (1, 3, and 6).

Theorem 4.3 was utilized to develop algorithms to obtain BSA plans by Stufken (2001) for $m = 1$ and by Mandal et al. (2008) and Stufken and Wright (2008) for $m \geq 1$. A catalogue of BSA plans is available in Mandal (2007) for $N \leq 40, n \leq 7, m \leq 4$. Note that Theorem 4.3 always gives polygonal designs which are cyclic in nature. In other words, the design is obtained by developing the generator blocks modulo v. Mandal *et al.* (2011) developed an integer linear programming formulation to identify the generator blocks, which when developed modulo v give a polygonal design. They constructed all the polygonal designs for $v \leq 100$, k=3 for all permissible m using the approach.

Another alternative approach to obtain BSA plans is through linear programming approach. Mandal *et al.* (2008) proposed such a linear programming approach to obtain BSA plans. The linear programming approach is presented below in brief.

Let S be the set of all possible samples of size n from the population and S_1be the set of non-preferred samples, which contain at least two adjacent units. Further let $p(s) \geq 0$ denote the selection probability of any sample s $\in S$. The linear programming problem demands obtaining $\{p(s)| \; s \in S\}$, minimizing the probability of selection of a non-preferred samples subject to constraints that the first order inclusion probability (π_i) of any unit i is n/N and the second order inclusion probability (π_{ij}) of any pair of units i and j is a constant whenever the two units i and j are not adjacent and π_{ij} is zero whenever two units i and j are adjacent, $i \neq j = 1, 2, \ldots, N$. The linear programming formulation for obtaining a BSA plan for population size N and sample size n for given m is:

Minimize $\varphi = \sum_{s \in S_1} p(s)$

Subject to constraints

i) $\sum_{s \ni i} p(s) = \frac{n}{N}$

ii) $\sum_{s \ni i,j} p(s) = 0$ if (i, j) are adjacent (4.1)

$= \frac{n(n-1)}{N(N-2m-1)}$ if (i, j) are non-adjacent

iii) $p(s) \geq 0 \forall s$

iv) $\sum_{s} p(s) = 1$

If there exists a feasible solution for given population size N, sample size n, an optimal solution is readily obtained by the simplex method. An optimum solution of linear programming gives the full support of the plan along with the probability of selections. If there is no feasible solution to the linear programming formulation (3.1) for a parameter set, then a BSA plan may not exist for that parameter set. One of the limitations of the proposed linear programming approach for construction of a BSA plans is that for large sample size n and population size N, number of possible samples becomes very large and linear programming becomes impractical to adopt.

Several other alternative approaches are available in literature to obtain BSA plans. Colbourn and Ling (1998) used partial triple system to solve the existence problem of BSA plans for $k = 3$ with $m = 1$. Tahir *et al.* (2010, 2012) used cyclic shift method to construct polygonal designs for $k = 3$ for particular settings of λ and m.

So far we have assumed that the population units have a circular ordering. This assumption is unrealistic and hence, BSA plans have been obtained considering that units have linear ordering and units have two-dimensional ordering.

Stufken and Wright (2008) presented results on existence of linear BSA plans. Mandal *et al.* (2008) presented a linear programming approach to obtain linear BSA plans. For two-dimensional populations, two-dimensional BSA plans were first introduced by Bryant *et al*. (2001). Later on, Wright (2008) introduced the concept of adjacency scheme to define several forms of adjacency between units in two-dimension. They also proposed one direct search algorithm for obtaining such plans. Gopinath (2014) formulated a linear program to construct BSA plans for two dimensional populations under different adjacency schemes given by Wright (2008).

5. Illustration with Real Data

In this section, we illustrate through a real data analysis that BSA plans can be efficiently used to obtain estimates of population means or totals. We consider daily evaporation data for the month of August, 2015 taken from the Indian Agricultural Research Institute, Pusa, New Delhi website http://www.iari.res.in/?option= com_content&id=402&Itemid=322. The data is displayed in Table

5.

Table 5: Evaporation data of August, 2015 in Delhi

Date	Evaporation (mm)	Date	Evaporation (mm)	Date	Evaporation (mm)
01-08-15	4	11-08-15	4	21-08-15	5.8
02-08-15	3.7	12-08-15	3.6	22-08-15	5
03-08-15	5	13-08-15	3.6	23-08-15	3.5
04-08-15	6.2	14-08-15	3.4	24-08-15	5.4
05-08-15	5.2	15-08-15	4.7	25-08-15	6
06-08-15	4.1	16-08-15	4.3	26-08-15	8
07-08-15	4.3	17-08-15	3.9	27-08-15	6.6
08-08-15	4.6	18-08-15	3.8	28-08-15	5.8
09-08-15	3.2	19-08-15	3.9	29-08-15	6
10-08-15	3.3	20-08-15	5.4	30-08-15	6.4
				31-08-15	7.2

We considered estimating the average evaporation for the month of August in Delhi using a sample of size n = 2, 3 and 4. The variance of the Narain-Horvitz-Thompson estimator of population mean is presented in Table 6 for these sample sizes under SRSWOR and BSA plans for m = 1, 2 and 3. It is clear from the Table that BSA plans are more efficient than SRSWOR plan.

Table 6: Variance of Narain-Horvitz-Thompson estimator of population mean under various sampling plans

Plan	n = 2	n = 3	n = 4
SRSWOR	0.734	0.472	0.341
BSA with m = 1	0.699	0.426	0.290
BSA with m = 2	0.679	0.400	0.260
BSA with m = 3	0.659	0.373	0.229

This is due to the fact that evaporation from nearer days are expected to be similar and are expected to be different for days far apart. Thus, a sampling plan which avoids nearer days in a sample is expected to be giving a better representation and thus, giving a more precise estimate of the average evaporation.

6. Concluding Remarks

Balanced sampling plans excluding adjacent units are useful for sampling from populations where the adjacent units provide similar measurements due to some natural ordering in time or space. These plans may be utilized for sampling in practical situations such as the estimation of yield of a certain crop in a region divided into geographical units like villages by two stage sampling, by choosing a number of villages first and then choosing a number of farms in each selected village.

Since nearer villages and nearer farms may provide similar observations, it would be desirable to select the samples at both stages in such a way that nearer units are given less chance of inclusion and units at farther distance are given more chance of inclusion in the sample. For several other applications of Balanced Sampling plans excluding adjacent units and further advances, one may refer to Gupta et al. (2012).

Balanced sampling plans excluding adjacent units are a very interesting class of sampling plans in terms of combinatorics. There is lot of scope of research for existence and construction of such plans for circular ordering of populations for sample size 5 or more. Recently, the plans have been extended for linear ordering of populations and two dimensional ordering of populations. However, there exist very challenging problems in construction and existence of such plans. Further research efforts are needed in this direction.

References

Bryant, D., Chang, Y., Rodger, C. A., and Wei, R. (2002). *Two-dimensional balanced sampling plans excluding contiguous units. Comm. Statist. Th. & Meth.,**31**(8), 1441-1455.*

Colbourn, C. J., and Ling, A. C. (1998). *A class of partial triple systems with applications in survey sampling. Comm. Statist. Th. & Meth.,**27**(4), 1009-1018.*

Colbourn, C. J. and Ling, A. C. H. (1999). *Balanced sampling plans with block Size four excluding contiguous units. Austr. J. Combi., **20**, 37-46.*

Gopinath, P. P. (2014). *Two dimensional balanced sampling plans excluding adjacent units. Unpublished M.Sc. Thesis, Indian Agricultural Research Institute, New Delhi.*

Gupta, V.K., Mandal, B.N. and Parsad, R. (2012). *Combinatorics in Sample Surveys vis-a-vis Controlled Selection. LAP Lambert Academic Publishing GmbH & Co. KG, Verlag, Germany.*

Hedayat, A. S., Rao, C. R., and Stufken, J. (1988). *Sampling plans excluding contiguous units. J. Statist. Plann. Inf., 19(2), 159-170.*

Horvitz, D. G. and Thompson, D. J. (1952). *A generalization of sampling without replacement from a finite universe. J. Amer. Statist. Assoc., **47**(260), 663–685.*

Mandal, B. N. (2007). *Combinatorics and its applications with special reference to sample surveys. Unpublished Ph.D. Thesis, Indian Agricultural Research Institute, New Delhi.*

Mandal, B. N., Gupta, V. K. and Parsad, R. (2011). *Construction of polygonal designs using linear integer programming. Comm. Statist. Th. & Meth., 40 (10), 1787-1794.*

Mandal, B. N., Parsad, R., and Gupta, V.K. (2008). *Computer-aided constructions of balanced sampling plans excluding adjacent units. J. Statist. Appl., 3, 59-85.*

Sen, A. R. (1953). *On the estimate of the variance in sampling with varying probabilities, J. Ind. Soc. Agril. Statist. 5(1194), 127-.*

Stufken, J. (1993). *Combinatorial and statistical aspects of sampling plans to avoid the selection of adjacent units. J. Combi. Inf. Sys. Science, 18, 81-92.*

Stufken, J. and Wright, J. H. (2008). *New balanced sampling plans excluding adjacent units. J. Statist. Plann. Inf., 138, 3326 – 3335.*

Stufken, J. and Wright, J.H. (2001). *Polygonal designs with blocks of size k ≤10. Metrika, 54, 179-184.*

Stufken, J., Song, S.Y., See, K. and Driessel, K.R. (1999). *Polygonal Designs: some existence and non-existence results. J. Statist. Plann. Inf., 77, 155-166.*

Tahir, M. H., Iqbal, I. and Shabbir J. (2012). *Polygonal designs with block size 3 and single distance. Hacettepe J. Math. Statist, 41(4), 587-604.*

Tahir, M. H., Iqbal, I., Akhtar, M. and Shabbir, J. (2010). *Cyclic polygonal designs with block size 3 and λ = 1 for joint distance α = 6 to 16. J. Statist. Th. Pr., 4(2), 203-220.*

Wright, J. H. (2008). *Two-dimensional balanced sampling plans excluding adjacent units. J. Statist. Plann. Inf.,138(1), 145-153.*

Wright, J. H., and Stufken, J. (2011). *Variance approximation under balanced sampling plans excluding adjacent units. J. Statist. Th. Pr., 5(1), 147-160.*